THE DEVELOPMENT AND APPLICATION OF MULTI-OMICS INTEGRATION APPROACHES TO DISSECTING COMPLEX TRAITS IN PLANTS

EDITED BY: Shang-Qian Xie, Jiang Libo, Lidan Sun and Yuehua Cui
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics



frontiers

Frontiers eBook Copyright Statement

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-88976-135-7
DOI 10.3389/978-2-88976-135-7

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

THE DEVELOPMENT AND APPLICATION OF MULTI-OMICS INTEGRATION APPROACHES TO DISSECTING COMPLEX TRAITS IN PLANTS

Topic Editors:
Shang-Qian Xie, Hainan University, China
Jiang Libo, Beijing Forestry University, China
Lidan Sun, Beijing Forestry University, China
Yuehua Cui, Michigan State University, United States

Citation: Xie, S.-Q., Libo, J., Sun, L., Cui, Y., eds. (2022). The Development and Application of Multi-Omics Integration Approaches to Dissecting Complex Traits in Plants. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-135-7

Table of Contents

06 ***Transcriptome and Co-expression Network Analyses Reveal Differential Gene Expression and Pathways in Response to Severe Drought Stress in Peanut (*Arachis hypogaea *L.)***
Nannan Zhao, Shunli Cui, Xiukun Li, Bokuan Liu, Hongtao Deng, Yingru Liu, Mingyu Hou, Xinlei Yang, Guojun Mu and Lifeng Liu

22 ***Genome-Wide Analysis of Coding and Non-coding RNA Reveals a Conserved miR164–NAC–mRNA Regulatory Pathway for Disease Defense in* Populus**
Sisi Chen, Jiadong Wu, Yanfeng Zhang, Yiyang Zhao, Weijie Xu, Yue Li and Jianbo Xie

34 ***Organ-Specific Transcriptome Analysis Identifies Candidate Genes Involved in the Stem Specialization of Bermudagrass (*Cynodon dactylon *L.)***
Si Chen, Xin Xu, Ziyan Ma, Jianxiu Liu and Bing Zhang

48 ***Transcriptomics and Metabolomics Reveal Purine and Phenylpropanoid Metabolism Response to Drought Stress in* Dendrobium sinense*, an Endemic Orchid Species in Hainan Island***
Cuili Zhang, Jinhui Chen, Weixia Huang, Xiqiang Song and Jun Niu

60 ***Integrative Identification of Crucial Genes Associated With Plant Hormone-Mediated Bud Dormancy in* Prunus mume**
Ping Li, Tangchun Zheng, Zhiyong Zhang, Weichao Liu, Like Qiu, Jia Wang, Tangren Cheng and Qixiang Zhang

72 ***Combined Analysis of MicroRNAs and Target Genes Revealed miR156-SPLs and miR172-AP2 are Involved in a Delayed Flowering Phenomenon After Chromosome Doubling in Black Goji (*Lycium ruthencium*)***
Shupei Rao, Yue Li and Jinhuan Chen

83 ***Comparative Transcriptome and Weighted Gene Co-expression Network Analysis Identify Key Transcription Factors of* Rosa chinensis *'Old Blush' After Exposure to a Gradual Drought Stress Followed by Recovery***
Xin Jia, Hui Feng, Yanhua Bu, Naizhe Ji, Yingmin Lyu and Shiwei Zhao

104 ***Identification of Candidate Genes Regulating the Seed Coat Color Trait in Sesame (*Sesamum indicum *L.) Using an Integrated Approach of QTL Mapping and Transcriptome Analysis***
Chun Li, Yinghui Duan, Hongmei Miao, Ming Ju, Libin Wei and Haiyang Zhang

116 ***Dissection of Allelic Variation Underlying Floral and Fruit Traits in Flare Tree Peony (*Paeonia rockii*) Using Association Mapping***
Xin Guo, Chunyan He, Fangyun Cheng, Yuan Zhong, Xinyun Cheng and Xiwen Tao

132 ***Comparative Analysis of Genomic and Transcriptome Sequences Reveals Divergent Patterns of Codon Bias in Wheat and Its Ancestor Species***
Chenkang Yang, Qi Zhao, Ying Wang, Jiajia Zhao, Ling Qiao, Bangbang Wu, Suxian Yan, Jun Zheng and Xingwei Zheng

145 ***Integrated Analysis of mRNA and Non-coding RNA Transcriptome in Pepper (*Capsicum chinense*) Hybrid at Seedling and Flowering Stages***
Huang-ying Shu, He Zhou, Hai-ling Mu, Shu-hua Wu, Yi-li Jiang, Zhuang Yang, Yuan-yuan Hao, Jie Zhu, Wen-long Bao, Shan-han Cheng, Guo-peng Zhu and Zhi-wei Wang

159 ***Transcriptome and Metabolome Analyses of the Flowers and Leaves of* Chrysanthemum dichrum**
Hua Liu, Xiaoxi Chen, Haixia Chen, Jie Lu, Dongliang Chen, Chang Luo, Xi Cheng, Yin Jia and Conglin Huang

175 ***Insights Into Walnut Lipid Metabolism From Metabolome and Transcriptome Analysis***
Suxian Yan, Xingsu Wang, Chenkang Yang, Junyou Wang, Ying Wang, Bangbang Wu, Ling Qiao, Jiajia Zhao, Pourkheirandish Mohammad, Xingwei Zheng, Jianguo Xu, Huming Zhi and Jun Zheng

186 ***Transcriptomic Time-Series Analyses of Gene Expression Profile During Zygotic Embryo Development in* Picea mongolica**
Jia Yan, Ha buer, Ya ping Wang, Gegen zhula and Yu´e Bai

199 ***Genome-Wide Association Mapping of Late Blight Tolerance Trait in Potato (*Solanum tuberosum *L.)***
Fang Wang, Meiling Zou, Long Zhao, Zhiqiang Xia and Jian Wang

210 ***Transcriptomic and Lipidomic Analysis of Lipids in* Forsythia suspensa**
Bei Wu, Yinping Li, Wenjia Zhao, Zhiqiang Meng, Wen Ji and Chen Wang

225 ***Unraveling the Genetic Architecture of Two Complex, Stomata-Related Drought-Responsive Traits by High-Throughput Physiological Phenotyping and GWAS in Cowpea (*Vigna. Unguiculata *L. Walp)***
Xinyi Wu, Ting Sun, Wenzhao Xu, Yudong Sun, Baogen Wang, Ying Wang, Yanwei Li, Jian Wang, Xiaohua Wu, Zhongfu Lu, Pei Xu and Guojing Li

236 ***RNA Sequencing Reveals Phenylpropanoid Biosynthesis Genes and Transcription Factors for* Hevea brasiliensis *Reaction Wood Formation***
Xiangxu Meng, Yue Wang, Jia Li, Nanbo Jiao, Xiujie Zhang, Yuanyuan Zhang, Jinhui Chen and Zhihua Tu

247 ***Asymmetric Divergence in Transmitted SNPs of DNA Replication/Transcription and Their Impact on Gene Expression in Polyploid* Brassica napus**
Minqiang Tang, Juanling Li, Xu Hu, Lu Sun, MMU Helal, Jianguo Chen and Yuanyuan Zhang

258 ***A Multilayer Interactome Network Constructed in a Forest Poplar Population Mediates the Pleiotropic Control of Complex Traits***
Huiying Gong, Sheng Zhu, Xuli Zhu, Qing Fang, Xiao-Yu Zhang and Rongling Wu

273 ***Integrated Lipidomic and Transcriptomic Analysis Reveals Lipid Metabolism in Foxtail Millet (*Setaria italica*)***
Haiying Zhang, Junyou Wang, Jing Zhao, Changqing Sun, Jin Wang, Qian Wang, Fei Qu, Xiaodong Yun and Zhiwei Feng

286 ***A Bivariate Mapping Model Identifies Major Covariation QTLs for Biomass Allocation Between Leaf and Stem Growth of* Catalpa bungei**
Miaomiao Zhang, Nan Lu, Tianqing Zhu, Guijuan Yang, Guanzheng Qu, Chaozhong Shi, Yue Fei, Bingyang Liu, Wenjun Ma and Junhui Wang

300 ***Analysis of Wheat Wax Regulation Mechanism by Liposome and Transcriptome***

Hongwei Wen, Ying Wang, Bangbang Wu, Yanru Feng, Yifei Dang, Bin Yang, Xiaofei Ma and Ling Qiao

314 ***Integration of Transcriptome and Methylome Analyses Provides Insight Into the Pathway of Floral Scent Biosynthesis in*** **Prunus mume**

Xi Yuan, Kaifeng Ma, Man Zhang, Jia Wang and Qixiang Zhang

327 ***Transcript and Protein Profiling Provides Insights Into the Molecular Mechanisms of Harvesting-Induced Latex Production in Rubber Tree***

Yujie Fan, Jiyan Qi, Xiaohu Xiao, Heping Li, Jixian Lan, Yacheng Huang, Jianghua Yang, Yi Zhang, Shengmin Zhang, Jun Tao and Chaorong Tang

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 April 2021
doi: 10.3389/fgene.2021.672884

Transcriptome and Co-expression Network Analyses Reveal Differential Gene Expression and Pathways in Response to Severe Drought Stress in Peanut (*Arachis hypogaea* L.)

*Nannan Zhao, Shunli Cui, Xiukun Li, Bokuan Liu, Hongtao Deng, Yingru Liu, Mingyu Hou, Xinlei Yang, Guojun Mu and Lifeng Liu**

State Key Laboratory for Crop Improvement and Regulation, Hebei Agricultural University, Baoding, China

Drought is one of the major abiotic stress factors limiting peanut production. It causes the loss of pod yield during the pod formation stage. Here, one previously identified drought-tolerant cultivar, "L422" of peanut, was stressed by drought (35 ± 5%) at pod formation stage for 5, 7, and 9 days. To analyze the drought effects on peanut, we conducted physiological and transcriptome analysis in leaves under well-watered (CK1, CK2, and CK3) and drought-stress conditions (T1, T2, and T3). By transcriptome analysis, 3,586, 6,730, and 8,054 differentially expressed genes (DEGs) were identified in "L422" at 5 days (CK1 vs T1), 7 days (CK2 vs T2), and 9 days (CK3 vs T3) of drought stress, respectively, and 2,846 genes were common DEGs among the three-time points. Furthermore, the result of weighted gene co-expression network analysis (WGCNA) revealed one significant module that was closely correlated between drought stress and physiological data. A total of 1,313 significantly up-/down-regulated genes, including 61 transcription factors, were identified in the module at three-time points throughout the drought stress stage. Additionally, six vital metabolic pathways, namely, "MAPK signaling pathway-plant," "flavonoid biosynthesis," "starch and sucrose metabolism," "phenylpropanoid biosynthesis," "glutathione metabolism," and "plant hormone signal transduction" were enriched in "L422" under severe drought stress. Nine genes responding to drought tolerance were selected for quantitative real-time PCR (qRT-PCR) verification and the results agreed with transcriptional profile data, which reveals the reliability and accuracy of transcriptome data. Taken together, these findings could lead to a better understanding of drought tolerance and facilitate the breeding of drought-resistant peanut cultivars.

Keywords: peanut (*Arachis hypogaea* L.), RNA-seq, WGCNA, drought stress, differentially expressed genes

OPEN ACCESS

Edited by:
Jiang Libo,
Beijing Forestry University, China

Reviewed by:
Yuning Chen,
Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, China
Ana Laura Furlan,
National University of Río Cuarto, Argentina

***Correspondence:**
Lifeng Liu
liulifeng@hebau.edu.cn

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

Received: *26 February 2021*
Accepted: *06 April 2021*
Published: *30 April 2021*

Citation:
Zhao N, Cui S, Li X, Liu B, Deng H, Liu Y, Hou M, Yang X, Mu G and Liu L (2021) Transcriptome and Co-expression Network Analyses Reveal Differential Gene Expression and Pathways in Response to Severe Drought Stress in Peanut (Arachis hypogaea L.). Front. Genet. 12:672884. doi: 10.3389/fgene.2021.672884

INTRODUCTION

Peanut (*Arachis hypogaea* L.) is one of the most important oil crops and economic crops in the world. It is a vital vegetable oil and protein source and is widely distributed in the tropical and subtropical regions. Drought is one of the most severe abiotic stresses that affects plant growth and development and causes constraint to agricultural productivity

(Shao et al., 2009; Osakabe et al., 2014). Drought not only severely limits the growth and production of peanuts, but also causes higher levels of aflatoxin infection (Girdthai et al., 2010; Jeyaramraja et al., 2018). It has become an important limiting factor to improve the yield and quality of peanuts. Therefore, improving the drought resistance of varieties has become an important goal of peanut breeding.

Timing, duration, and severity of drought are important factors affecting peanut yield and quality (Rao et al., 1989; Dang et al., 2013). In general, the form of peanut is slightly drought-resistant, but in some specific periods, water shortage seriously affects the yield of peanut. In the pod formation stage, drought can severely reduce yield of peanut because it can largely decrease the number and fullness of pods (Rao et al., 1985; Koolachart et al., 2013; Yang et al., 2019). Therefore, understanding the molecular basis of drought response at pod formation stage is essential in peanut breeding programs to improve pod yield.

Plants have evolved complex molecular, physiological, and biochemical processes to cope with the effects of drought (Ramachandra Reddy et al., 2004; Shinozaki and Yamaguchi-Shinozaki, 2007; Shao et al., 2009; Osakabe et al., 2014). For example, drought stress causes the production of reactive oxygen species (ROS), and excessive ROS would lead to oxidative stress, inhibit plant growth, and even cause cell death. The key enzymes in plants can change under stress conditions, including superoxide dismutase (SOD), catalase (CAT), peroxidase (POD), ascorbate peroxidase (APX), glutathione reductase (GR), and so on, to involve themselves in the detoxification of ROS (Reddy et al., 2004). Also, plant response to drought stress includes osmotic regulation and hormone regulation (Shao et al., 2009; Gong et al., 2020). For example, Malondialdehyde (MDA) is the product of lipid peroxidation, and its dynamic accumulation in plant cells indicates the degree of membrane damage (Deng et al., 2019). Soluble sugar and protein act as osmoregulatory substances to protect plants from stress (Shao et al., 2009; Ozturk et al., 2020). In the molecular process, numerous functional genes and regulatory genes have been discovered under drought stress (Harb, 2016). For example, late embryogenesis abundant (LEA) proteins play a crucial role in protecting cells during dehydration (Hanin et al., 2011). Overexpression of *TaSnRK2.9* enhanced tobacco tolerance to drought and salt stresses through improved ROS scavenging ability (Feng et al., 2019). Transcription factors (TFs) also play a vital role in the response to drought stress, such as heat shock factor (HSF), basic helix-loop-helix (bHLH), NAC, and WRKY transcription factor families (Castilhos et al., 2014; Joshi et al., 2016; Duan et al., 2017; Guo et al., 2019; Manna et al., 2020; Zeng et al., 2020). Although there are many studies on drought resistance in plants, drought resistance is a complex trait controlled by a large number of genes, which has still not been fully elucidated and needs more investigation (Budak et al., 2015; Kumar et al., 2017).

Transcriptomic analysis is a highly efficient way to investigate genome function and the related important pathways (Mia et al., 2020). Many studies have been carried out using transcriptome analysis for drought stress in numerous crops (Ma et al., 2017; Zhu et al., 2019; Hao et al., 2020; Tiwari et al., 2021). A few studies have revealed many genes involved in drought stress in peanuts using transcriptome analysis (Zhao et al., 2018; Bhogireddy et al., 2020; Huang et al., 2020; Jiang et al., 2021). Nevertheless, the drought-related networks need to be further explained using transcriptome analysis due to the complexity of the relevant genetic pathways. With the recent development of bioinformatics, weighted gene co-expression network analysis (WGCNA) can be used for identifying genes with similar expression patterns that may participate in specific biological functions (Langfelder and Horvath, 2008). Our previous study has shown that "L422" is a drought tolerant cultivar (Zhao N. et al., 2020). Plants can preserve water through various anatomical features when subjected to drought, such as reducing leaf surface area by leaf rolling, folding, or shedding (Goche et al., 2020). Here, the transcriptional response of the leaves of "L422" to severe drought was analyzed at the pod formation stage by using RNA sequencing (RNA-seq). Further, differential gene expression in multiple crucial signaling pathways involved in plant drought stress was analyzed from the module that was strongly correlated with drought stress and physiological data using WGCNA. These findings will provide a valuable resource for the study of drought resistance in peanut and lay a foundation for further targeted research on drought resistance genes.

MATERIALS AND METHODS

Plant Materials and Drought Treatment

Peanut cultivar "L422" was a drought tolerant cultivar based on a previous study (Zhao N. et al., 2020). "L422" was planted in rainout shelters in Baoding, China (115°E, 38°N) in 2019, and confirmed again to be drought tolerant. In brief, "L422" (drought-resistant) (Zhao N. et al., 2020), "Huayu 23" (drought-sensitive) (Ding et al., 2017), "Huayu 25" (drought-resistant) (Zhang et al., 2011), and "L632" (drought-sensitive; data not shown) cultivars with different drought tolerances were planted in environmentally controlled rainout shelters (6 m × 8 m), with two water treatments (well-watered and drought) and three replicates. The relative soil water content (RSWC) was maintained at 70–75% until the plants reached the reproductive phase (pod formation stage). At the pod formation stage, the control group continued to be under well-watered conditions while the treatment group stopped irrigation until the RSWC of the soil decreased to 35%.

In the current study, we performed a transcriptomic analysis of "L422" at the pod formation stage under drought stress. Seeds were surface sterilized with 70% ethanol followed by thorough washing with sterile distilled water. After sterilizing, the seeds were soaked in deionized water at room temperature for 12 h. Subsequently, the seeds were placed in two layers of damp filter paper for 24 h in the dark to induce germination. Germinated seeds were planted in plastic pots (one seedling for each pot) in rainout shelters (Baoding, China) under well-watered conditions at 70–75% RSWC. The pots were 29.5 cm in diameter, 26.0 cm in ground diameter, and 23.5 cm in height.

Then peanut seedlings under the same cultivation conditions were divided into control group and treatment group. The control group was well-watered continually, and irrigation was interrupted for the treatment group when peanuts entered the pod formation stage, which was 75 days after planting (DAP). Based on the RSWC (35 ± 5%) and phenotypic changes of the treatment group, fully expanded leaves from the main stem (Third nodal) of control (CK) and treatment (T) plants were sampled after 5 (80 DAP), 7 (82 DAP), and 9 (84 DAP) days of drought treatment, and then were immediately frozen in liquid nitrogen and stored at −80°C for subsequent analyses. Each treatment was replicated three times.

Physiological Index Measurements

Phenotypic and physiological characterizations were determined for "L422" under well-watered and drought-stress conditions. The relative water content (RWC) was determined based on the method described by Galmes et al. (2007). Similarly, the relative electric conductivity of the peanut leaves was measured according to the method of Zhang et al. (2018). The MDA content, POD activity, soluble sugar, and soluble protein content of samples were measured using physiological assay kits (Suzhou Grace Biotechnolgy Co., Ltd, Jiangsu, China) referring to the manufacturers' recommendations based on the methods of thiobarbituric acid-reactive-substances (TBARs), guaiacol colorimetric, anthrone colorimetric, and bicinchoninic acid (BCA), respectively. All processes were biologically and temporally repeated in three independent and parallel experiments. Student's *t*-test was performed to calculate the *p*-values using GraphPad Prism software, version 8.01 (GraphPad Software, Inc., San Diego, CA, United States).

Transcriptome Sequencing and *de novo* Assembly Analysis

The isolation of total RNA from non-stressed and stressed leaves of "L422" was optimized according to the instruction manual of the Trizol Reagent (Invitrogen, Carlsbad, CA, United States). RNA degradation and contamination were monitored on 1% agarose gels. The quality of the RNA was evaluated using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States), and 18 qualified RNA samples were used for RNA-seq analysis. A library was constructed using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, United States). The cDNA library construction and sequencing were carried out on the Illumina HiseqTM 2500 platform by Gene Denovo Biotechnology Co. (Guangzhou, China).

Sequencing Reads Processing and Mapping

The quality of raw data (raw reads) was firstly processed by fastp (version 0.18.0) (Chen et al., 2018). In this step, clean data (clean reads) was obtained by removing reads containing adapters, more than 10% of unknown nucleotides (N), and more than 50% of low quality (*Q*-value ≤ 20) bases. Meanwhile, Q20 (99% base call accuracy), Q30 (99.9% base call accuracy), GC-content, and sequence duplication levels of the clean data were calculated. Qualified clean reads were then mapped to the peanut reference genome sequence (Tifrunner.gnm1.ann1.CCJH) using a spliced aligner HISAT2 software (version 2.2.4) (Kim et al., 2015). The mapped reads of each sample were assembled using StringTie (version 1.3.1) (Pertea et al., 2015). The gene expression level was normalized using the FPKM (Fragments per Kilobase of transcript per Million mapped reads) method. All the downstream analyses were based on high-quality clean data.

Differential Expression Analysis

Differential gene expression analysis of the two groups was performed by DESeq2 (Love et al., 2014). The corrected *p*-values were used to control the false discovery rate (FDR). Genes were considered to be differentially expressed when the value of log2 Fold Change was >2 or <-2 with an FDR value below 0.01 between two groups. The Gene Ontology (GO) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis of differentially expressed genes (DEGs) were conducted using the hypergeometric test by comparing with the whole genome background. GO terms and KEGG pathways with FDR-corrected *p*-value ≤ 0.05 were regarded as significantly enriched in DEGs.

Weighted Gene Co-expression Network Analysis

Weighted gene co-expression network analysis is a systems biology method for describing the correlation patterns among genes across multiple samples. This method aims to find clusters (modules) of highly correlated genes and relating modules to external sample traits (Zhang and Horvath, 2005). Co-expression networks were constructed using WGCNA (version 1.47) package in R (Langfelder and Horvath, 2008). After filtering non-varying or low-abundance (FPKM < 2) genes of samples (>70%), gene expression values were imported into WGCNA to construct co-expression modules using the automatic network construction function blockwise modules with default settings, except that the power is 10, TOMType is unsigned, mergeCutHeight is 0.75, and minModuleSize is 50. Genes were clustered into nine correlated modules.

Gene Expression Validation

Nine genes with different expression profiles obtained by Illumina RNA-seq were randomly selected for validation by qPCR. Gene-specific primers were designed by Wcgene Biotech (Shanghai, China) (**Supplementary Table 1**). The Actin gene was used as housekeeping gene. Three biological and technical repetitions were used for each sample. The quantitative real-time PCR (qRT-PCR) was run on the ABI StepOnePlus instrument using Fast Super EvaGreen®qPCR Master Mix (US Everbright®Inc., China) according to the manufacturer's instructions. The amplification program was set as follows: 95°C for 2 min followed by 45 cycles of 95°C for 5 s and 60°C for 1 min. All data from qRT-PCR amplification were calculated with $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

RESULTS

Physiological and Phenotypic Changes of Peanuts Under Drought Stress

To investigate the physiological responses of peanuts to water deficit, the physiological indexes were evaluated at the pod formation stage, including leaf RWC and relative electrical conductivity (**Supplementary Figure 1**). As shown in **Supplementary Figure 1A**, there were significant phenotypic changes in four peanut varieties. In terms of leaves, peanuts shriveled up under drought stress in "Huayu 23" and "Huayu 25," but "Huayu 23" withered intensely. Although the leaves of "L632" did not wither like "Huayu 23" and "Huayu 25," they turned yellow. However, the phenotypic change was not obvious in "L422." The RWC of "L422," "Huayu 23," "Huayu 25," and "L632" decreased to 32.3, 47.7, 34.6, and 44.0% under drought stress, respectively, as compared with the control (**Supplementary Figure 1B**). Relative electrical conductivity is widely used to measure the ability of plants to avoid or repair membrane damage. The relative electrical conductivity of "L422," "Huayu 23," "Huayu 25," and "L632" increased by 107.6, 208.6, 108.8, and 167.8%, respectively (**Supplementary Figure 1C**).

"L422" were planted in plastic pots and treated with drought at the pod formation stage. After 5 days of drought treatment, the leaves of "L422" began to shrivel up. After 7 and 9 days of drought treatment, "L422" showed distinct wilting (**Figure 1A**). Some physiological indicators' response to drought stress were then measured. The leaf RWC decreased significantly ($p < 0.01$) with the increasing days of stress exposure (**Figure 1B**). Compared with the control, the RWCs of drought-treated leaves decreased to 51.8, 57.0, and 58.2% at 5, 7, and 9 days after drought treatment, respectively. The relative electrical conductivity of "L422" increased by 122.8, 383.6, and 460.9% in 5, 7, and 9 days, respectively (**Figure 1C**). Results of MDA content showed that the stressed group was significantly ($p < 0.01$) higher than the non-stressed group by 58.2, 80.8, and 59.9% under 5, 7, and 9 days with drought stress, respectively (**Figure 1D**). These data suggested that the leaves of "L422" were damaged under severe drought stress. As shown in **Figure 1G**, POD activity increased under severe drought stress but did not increase significantly at 9 days of drought stress. Compared with the control group, the soluble sugar content showed a trend of increase by 56.15, 93.5, and 102.3% at 5, 7, and 9 days with drought stress (**Figure 1E**). Additionally, the soluble protein content exhibited a greater increase at 7 days than 5 and 9 days of drought stress (**Figure 1F**).

RNA Sequencing Analysis of "L422" Under Drought Stress

To investigate the key genes of peanuts in response to drought, the treated leaves of "L422" were sequenced. A total of approximately 94.04 million raw reads were generated from the 18 cDNA libraries (six samples × three replications) by RNA sequencing. The raw sequencing data had been deposited in NCBI under the accession number PRJNA706902. After deleting 0.60% of adapter sequences, and filtering 0.33% of low-quality reads and 0.00% of n-containing reads, 93.16 million high-quality clean reads were finally confirmed (**Supplementary Table 2**). The percentage of high-quality clean reads mapped to the peanut reference genome arahy.Tifrunner.gnm1.KYV3 ranged from 89.24 to 93.78% (**Table 1**). These results showed that the transcriptome sequencing quality was sufficient for further analyses.

Differentially Expressed Genes and qRT-PCR Validation

Generally, a stringent threshold absolute log2 FC $\geq$ 2 and FDR < 0.01 was used to screen out DEGs. The number of DEGs after 5, 7, and 9 days of drought stress were 3,586, 6,730, and 8,054, respectively (CK1 vs T1, CK2 vs T2, and CK3 vs T3), and 2,846 genes were common DEGs among the three time points in "L422" (**Figures 2A,B**). After 5 days of drought treatment, 1,800 up-regulated DEGs and 1,786 down-regulated DEGs were identified. Of these DEGs after 7 days of drought treatment, 3,398 were up-regulated and 3,332 were down-regulated. After 9 days of continuous stress, the number of DEGs was the largest in the three treatment time points with 3,954 up-regulated genes and 4,100 down-regulated genes. On the whole, the number of up-regulated DEGs is higher than down-regulated DEGs, except for drought stress for 9 days. Together, the results revealed that the number of induced DEGs greatly increased with the continuation of drought stress time. All these DEGs were selected for further analysis.

To experimentally confirm the results of RNA-Seq data, nine DEGs were randomly selected to perform qRT-PCR. As shown in **Figure 3A**, the selected DEGs had consistent expression patterns between RNA-Seq and qRT-PCR. The results showed a good correlation between the qRT-PCR results and the RNA-Seq results ($r = 0.99$, $p < 2.2e\text{-}16$, **Figure 3B**). This signifies that the RNA-seq data was of high-quality.

Weighted Gene Co-expression Network Analysis Under Drought Stress

To identify the expression of genes related to drought stress in peanut, a gene co-expression network was constructed using WGCNA. The 26,409 selected genes were assigned to nine merged co-expression modules (with various colors) (**Figure 4A**). As shown in **Figure 4B**, we successfully identified two modules significantly associated with drought stress for "L422" ($p < 0.05$). The MM.darkred module ($r = 0.98$, $p = 5e\text{-}04$) was positively correlated with resistance throughout the severe drought stress period, while the MM.black module ($r = -0.84$, $p = 0.04$) was negatively correlated with drought stress. Additionally, a module-trait relationships analysis was performed using module eigengenes and physiological data at each time point. As shown in **Figure 4C**, the MM.darkred module ($r = -0.98$, $p = 2e\text{-}12$) was negatively correlated with RWC under drought stress. In contrast, the MM.darkred module had a significant positive correlation with the RWC ($r = 0.97$, $p = 5e\text{-}11$), relative electrical conductivity($r = 0.96$, $p = 2e\text{-}10$), soluble sugar ($r = 0.81$, $p = 4e\text{-}05$), POD ($r = 0.51$, $p = 0.03$), soluble protein($r = 0.71$, $p = 0.001$), and MDA ($r = 0.81$, $p = 3e\text{-}6$). The identification of peanut genotype-specific modules in severe drought stress

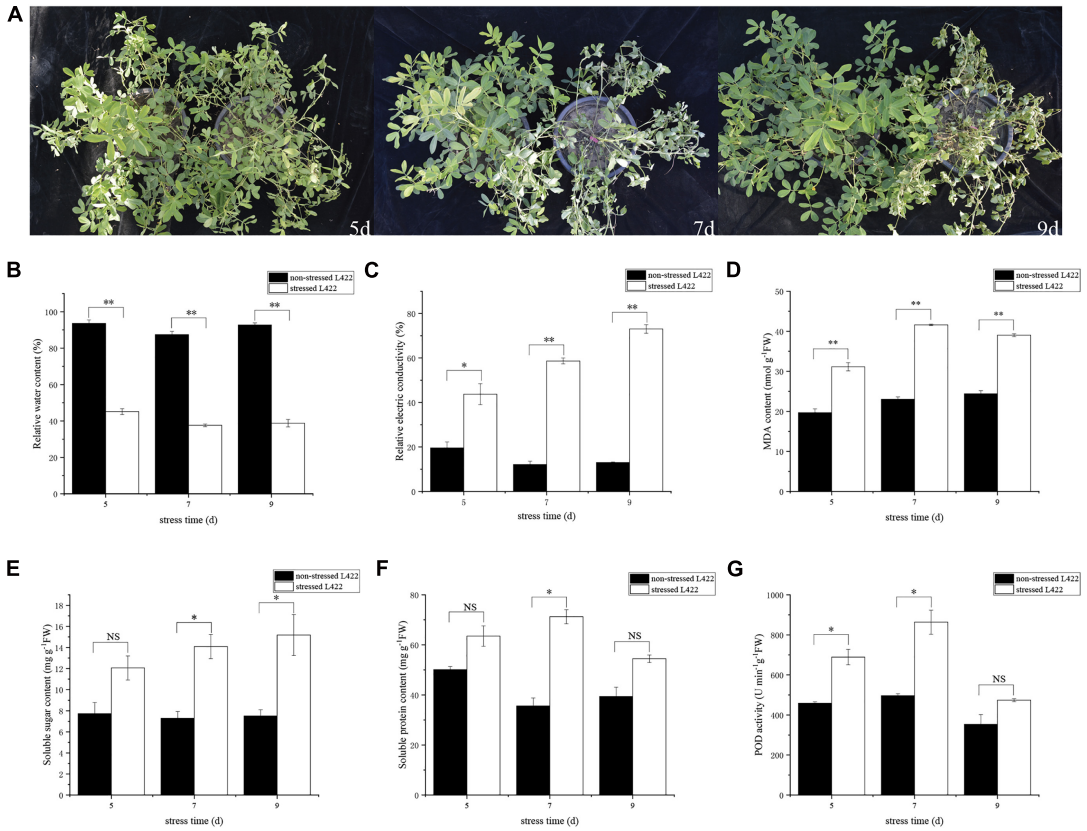


FIGURE 1 | Phenotypic and physiological changes of "L422" under drought stress. **(A)** Phenotypic responses of "L422" to severe drought stress at three-time points (5, 7, and 9 days). The pots on the right and left correspond to the drought-treated and well-watered control peanut, respectively. The changes of the relative water content **(B)**, the relative electric conductivity **(C)**, MDA content **(D)**, soluble sugar content **(E)**, soluble protein content **(F)**, and POD activity **(G)** in leaves of "L422" under well-watered and drought conditions. Values are the mean ± standard deviation of three biological replicates. * and ** indicate the significant difference at 5% level and 1% level, respectively, "NS" indicates non-significant.

was particularly important. Based on the above results, the MM.darkred module was related to drought response, and was selected for further analysis.

Enrichment Analysis of the Detected Co-expressed Modules

As shown in **Figure 4D**, these genes in the MM.darkred module had a similar preponderant expression stage based on the gene expression heatmaps and eigengene histograms. In accordance with the condition of FPKM $\geq$ 9 for at least one sample, we screened 1,313 common DEGs based on the MM.darkred module and 2,846 common DEGs at three time points under severe drought stress to perform KEGG analysis (**Supplementary Figure 2**). Multiple crucial pathways involved in plant drought stress were determined, which included "Mitogen activated protein kinase (MAPK) signaling pathway-plant," "flavonoid biosynthesis," "starch and sucrose metabolism," "phenylpropanoid biosynthesis," "glutathione metabolism," and "plant hormone signal transduction," and summarized for analysis (**Supplementary Table 3**). In MAPK signaling pathway, all seven DEGs were up-regulated throughout the severe drought stress containing two protein kinase superfamily proteins, four protein phosphatase 2C family proteins, and one chitinase family protein (**Figure 5**). The genes annotated as starch and sucrose metabolism were three down-regulated and three up-regulated (**Figure 6A**). In total, three down-regulated genes and six up-regulated genes were found in the flavonoid biosynthesis and phenylpropanoid biosynthesis pathways (**Figure 6B**). The genes annotated as glutathione metabolism exposed to drought stress had three down-regulated genes and 20 up-regulated genes, which were mainly annotated as glutathione S-transferase family proteins (**Supplementary Table 3** and **Figure 6C**). In addition, many genes involved in hormone biosynthesis were detected to be differentially expressed under severe drought stress. A total of 18 DEGs involved in plant hormone signal transduction of auxin (IAA), jasmonic acid (JA), gibberellin (GA), brassinosteroid (BR), and abscisic acid (ABA) metabolism were identified in this study (**Figure 7** and **Supplementary Table 3**). In IAA signal pathway, two AUX1 genes and four AUX/IAA genes were down-regulated under drought at three time points, but only one SAUR gene was up-regulated (**Figure 7A**). Six up-regulated genes were identified in ABA signal pathway, including four PP2C genes and two SnPK2 genes (**Figure 7B**). However, only one TF gene

TABLE 1 | Summary of sequencing data for different samples.

Sample	RawData (bp)	CleanData (bp)	Q20 (%)	Q30 (%)	N (%)	GC (%)	Unmapped reads	Unique mapped reads	Total_Mapped (%)
T1-1	7443549900	7133966630	6948467942 (97.40%)	6642569327 (93.11%)	109097 (0.00%)	3178907415 (44.56%)	4133261 (8.54%)	33739797 (69.67%)	44291299 (91.46%)
T1-2	7478063700	7192087091	6990017761 (97.19%)	6680180461 (92.88%)	110950 (0.00%)	3213723611 (44.68%)	4183864 (8.62%)	33366382 (68.74%)	44353176 (91.38%)
T1-3	7402878300	7157768553	6963430531 (97.28%)	6662999707 (93.09%)	109804 (0.00%)	3202300203 (44.74%)	3607567 (7.50%)	33639294 (69.93%)	44493441 (92.50%)
T2-1	8447553300	8117926339	7905605355 (97.38%)	7574245766 (93.30%)	124397 (0.00%)	3653219812 (45.00%)	4953862 (9.04%)	37313003 (68.12%)	49820414 (90.96%)
T2-2	7438557600	7151027498	6957420272 (97.29%)	6655615104 (93.07%)	109852 (0.00%)	3205202420 (44.82%)	4642050 (9.62%)	32773043 (67.95%)	43588052 (90.38%)
T2-3	7893744900	7604749442	7407754289 (97.41%)	7095644448 (93.31%)	115653 (0.00%)	3397292779 (44.67%)	4724289 (9.20%)	35577300 (69.28%)	46630185 (90.80%)
T3-1	7752862500	7492206686	7286261792 (97.25%)	6960459720 (92.90%)	115313 (0.00%)	3344979970 (44.65%)	4628456 (9.20%)	34583523 (68.77%)	45661100 (90.80%)
T3-2	6797084400	6560746525	6379007254 (97.23%)	6102273974 (93.01%)	101097 (0.00%)	2903234585 (44.25%)	3928108 (8.95%)	30967665 (70.57%)	39952456 (91.05%)
T3-3	7047023400	6784517938	6600157401 (97.28%)	6311771245 (93.03%)	104954 (0.00%)	3050939041 (44.97%)	3824863 (8.37%)	31321183 (68.58%)	41849305 (91.63%)
CK1-1	8348989800	8055064749	7842051364 (97.36%)	7506055348 (93.18%)	123426 (0.00%)	3588914604 (44.55%)	4348583 (7.95%)	39021323 (71.38%)	50316235 (92.05%)
CK1-2	7285225200	7017227422	6813863910 (97.10%)	6494308099 (92.55%)	107792 (0.00%)	3143924964 (44.80%)	3916454 (8.25%)	33069245 (69.62%)	43581134 (91.75%)
CK1-3	7498422000	7201743917	7009836007 (97.34%)	6711677567 (93.20%)	111270 (0.00%)	3226974407 (44.81%)	3072703 (6.28%)	34751895 (71.03%)	45854949 (93.72%)
CK2-1	6220695600	6146941251	5605972285 (91.20%)	4903839074 (79.78%)	63350 (0.00%)	2753172797 (44.79%)	4392239 (10.76%)	30121830 (73.80%)	36424409 (89.24%)
CK2-2	7348115100	7030908860	6815994172 (96.94%)	6449075104 (91.72%)	12606 (0.00%)	3178348608 (45.21%)	3169138 (6.60%)	34231533 (71.28%)	44855020 (93.40%)
CK2-3	10131380700	9762527954	9504451596 (97.36%)	9096260361 (93.18%)	150432 (0.00%)	4368690192 (44.75%)	4127084 (6.22%)	47372258 (71.36%)	62261582 (93.78%)
CK3-1	8406291000	8118055249	7903244807 (97.35%)	7559991795 (93.13%)	125012 (0.00%)	3659974797 (45.08%)	3786157 (6.91%)	37898670 (69.16%)	51008683 (93.09%)
CK3-2	9171919200	8839433152	8615603298 (97.47%)	8257129252 (93.41%)	134137 (0.00%)	3972660094 (44.94%)	4132115 (6.91%)	41561446 (69.48%)	55688533 (93.09%)
CK3-3	8942134200	8616413360	8393324512 (97.41%)	8033373375 (93.23%)	132093 (0.00%)	3834640592 (44.50%)	3699547 (6.31%)	41809441 (71.36%)	54886911 (93.69%)

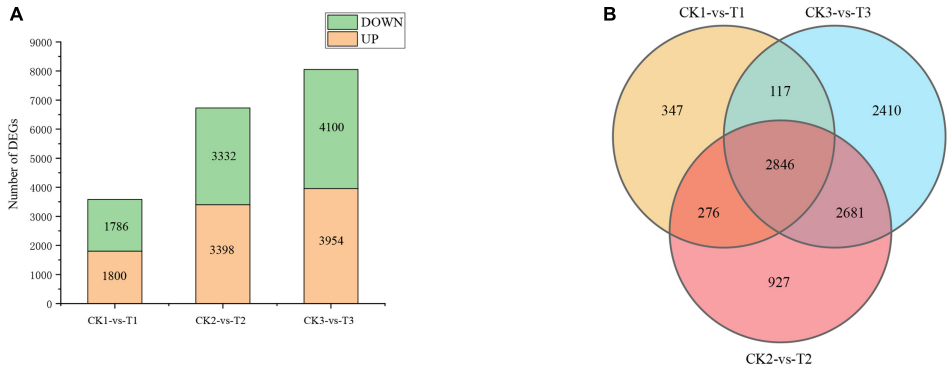


FIGURE 2 | Difference analysis of gene expression by pairwise comparisons. **(A)** The number of DEGs induced by drought. **(B)** Venn diagram analysis of DEGs at the three-time points under severe drought stress. CK1-vs-T1: comparison between 5 days of drought and 5 days of well-watered condition; CK2-vs-T2: comparison between 7 days of drought and 7 days of well-watered condition; CK3-vs-T3: comparison between 9 days of drought and 9 days of well-watered condition.

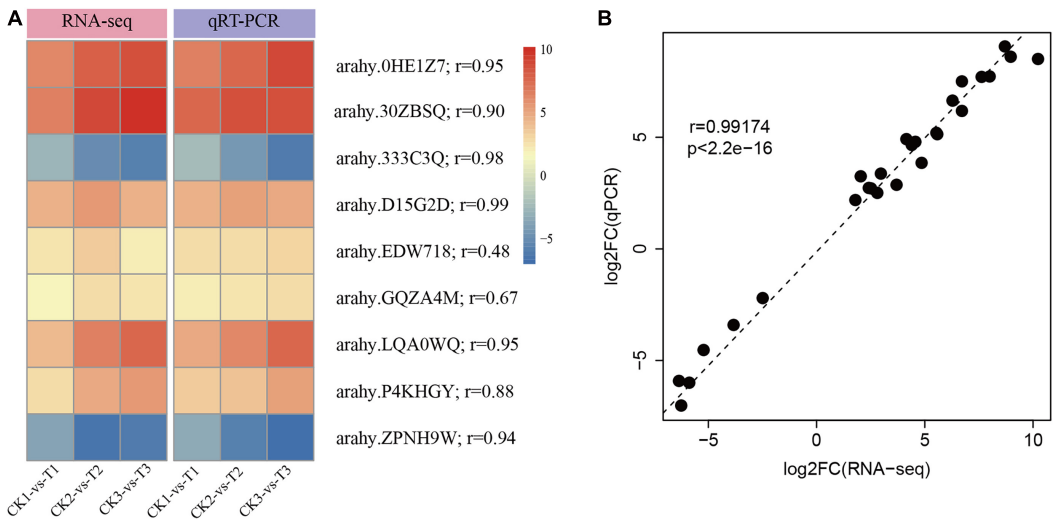


FIGURE 3 | Confirmation of RNA-Seq results by quantitative real-time PCR (qRT-PCR). **(A)** The heatmap presentation of fold changes of nine DEGs obtained from RNA-seq analysis and qRT-PCR results. **(B)** Correlation between RNA-Seq expression profile and qRT-PCR results.

and TCH4 gene were down-regulated in the GA and BR signal pathways, respectively (**Figure 7C,D**). We also found two up-regulated GAZ genes in the SA pathway (**Figure 7E**). Moreover, the GO terms related to drought response were also identified (**Supplementary Figure 2**).

Transcription Factors in Response to Drought Stress

A total of 902 TFs were identified in the MM.darkred module (**Figure 8A**). One-hundred fifty-two differentially expressed TF genes were obtained according to the condition of FPKM $\geq$ 9 with at least one sample throughout the severe stress stage (**Figure 8C** and **Supplementary Table 4**), of which the TF families of bHLH, NAC, and WRKY were the top three families (**Figure 8B**). Sixty-one TFs were screened based on the MM.darkred module and 2,846 common DEGs at three time points under severe drought stress (**Supplementary Table 4**). The key TF genes included *arahy.Tifrunner.gnm1.ann1.PPQG6E* (*BHLH 72*), *arahy.Tifrunner.gnm1.ann1.02IZMF* (*PIL15*), *arahy.Tifrunner.gnm1.ann1.72Q128* (*NAC029*), *arahy.Tifrunner.gnm1.ann1.D15G2D* (*WRKY71*), *arahy.Tifrunner.gnm1.ann1.30ZBSQ* (*WRKY75*), and *arahy.Tifrunner.gnm1.ann1.EDW718* (*WRKY15*). Overall, the high expression of these TF families in "L422" may play a vital role under severe drought stress.

DISCUSSION

The cultivated peanut is an allotetraploid (amphidiploid with 2n = 4x = 40) and is relatively drought-tolerant to a certain extent. However, water deficit stress in pod formation stage would seriously affect the yield and productivity of peanuts (Haro et al., 2011; Koolachart et al., 2013). Therefore, improving drought tolerance of peanuts is very important and more research

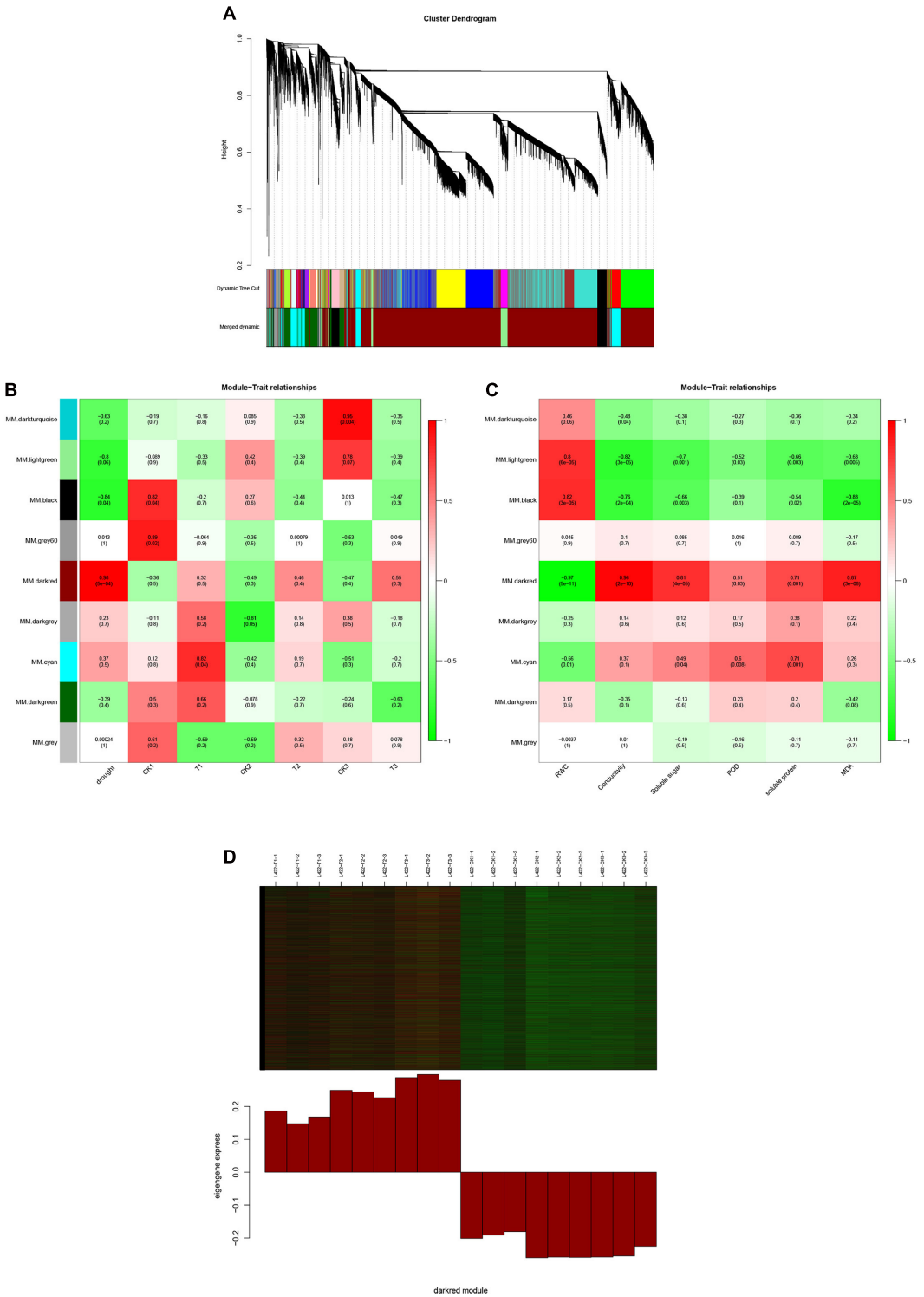


FIGURE 4 | WGCNA of the transcripts changes in "L422." **(A)** Hierarchical cluster tree shows nine modules of co-expressed genes in "L422." Different modules are marked with different colors. Each leaf in the tree represents one gene. **(B)** Correlations of drought degree and samples with WGCNA modules. **(C)** Correlations of physiological indicators with WGCNA modules. The right color scale corresponds to module-trait correlation. Each row represents a specific module. The numbers in each cell represent the correlation coefficients and correlation significance levels (in parentheses). **(D)** Expression pattern of the genes and eigengenes of MM.darkred module. The heatmap was plotted using the log10 FPKM values.

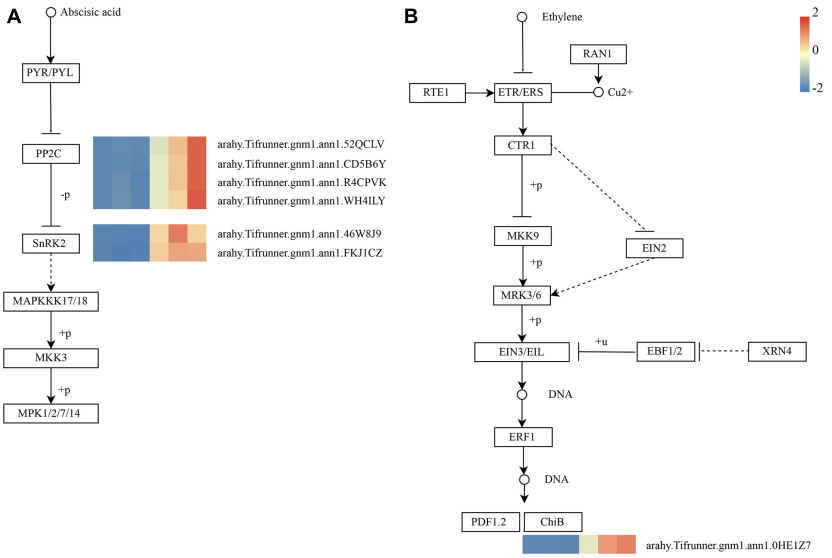


FIGURE 5 | Drought-responsive genes of MM.darkred module in MAPK signaling pathway-plant. **(A)** ABA and **(B)** Ethylene signal transduction pathways. Relative expression levels are normalized based on the Z-score and shown as a color gradient from low (blue) to high (red). The columns in heat map are 5, 7, and 9 days of well-watered condition, and 5, 7, and 9 days of drought-treated condition under severe drought from left to right, respectively.

is needed to explore and understand drought stress. Here we performed the physiological and transcriptomic analysis of "L422" at the pod formation stage under drought and well-watered conditions. Then, bioinformatics methods were utilized to analyze differential gene expression in multiple signaling pathways that were potentially associated with drought. The results provided informative clues for the elucidation of drought stress tolerance in peanuts, as well as providing a basis for the identification of drought resistance candidate genes.

Physiological and Phenotypic Changes of Peanuts in Response to Drought Stress

Plant drought stress response and adaptation are extremely complex, including physiological changes. In this study, we assessed physiological changes of peanuts at the pod formation stage under severe drought conditions. Compared with the control, the phenotypic and physiological changes of "L422" under severe drought stress were obvious (**Figure 1**). We observed greater leaf rolling with decreasing water content and increasing relative electrical conductivity in drought treated leaves, respectively (Figure, 1B,C). The excessive accumulation of free radicals in cells leads to membrane lipid peroxidation. MDA is the main product of cytoplasmic membrane peroxidation, which is an important index to evaluate plant tolerance to drought stress (Deng et al., 2019). In this study, the MDA content showed that the stressed group was significantly higher than the non-stressed group under severe drought stress. This observation may suggest that the cell membrane of leaves is damaged, which leads to the release of cell membrane lipid and the destruction of membrane structure (**Figure 1D**). The measurement of RWC, relative electrical conductivity, and MDA confirmed that "L422" suffered physiological damage under severe drought stress. It is well known that drought stress can lead to the accumulation of ROS in plants and its over-accumulation is harmful to plant cells (Blokhina et al., 2003). The scavenging system comprising antioxidants plays important roles in scavenging the ROS. Peroxidase, as an important antioxidant, can minimize cellular damage by scavenging and detoxifying ROS-generated H_2O_2 (Sharma et al., 2012). However, we found that POD activity did not change significantly at 9 days of drought stress, which may mean that antioxidant enzymes cannot effectively scavenge ROS under long-term drought conditions (**Figure 1G**). Various osmoregulatory substances such as soluble sugar and soluble protein can increase the osmotic potential at the cellular level to prevent loss of moisture and enhance plant stress resistance (Shao et al., 2009; Ozturk et al., 2020). In our study, the soluble sugar and soluble protein content increased compared with the control group under severe drought conditions (**Figures 1E,F**). The result was consistent with former studies (Fu et al., 2011; Ozturk et al., 2020). Also, the soluble protein content showed a greater increase at 7 days of drought stress, but the subsequent changes were not significant. The greater increase may be due to the expression of new stress proteins, and then the increase was not significant because of the serious decline in photosynthesis. These physiological and phenotypic changes suggest that severe drought stress has a serious effect on the pod formation stage of peanuts.

Analysis of Starch and Sucrose Metabolism in Response to Drought Stress

Based on KEGG pathway enrichment analysis, starch and sucrose metabolism pathway was identified in the MM.darkred module.

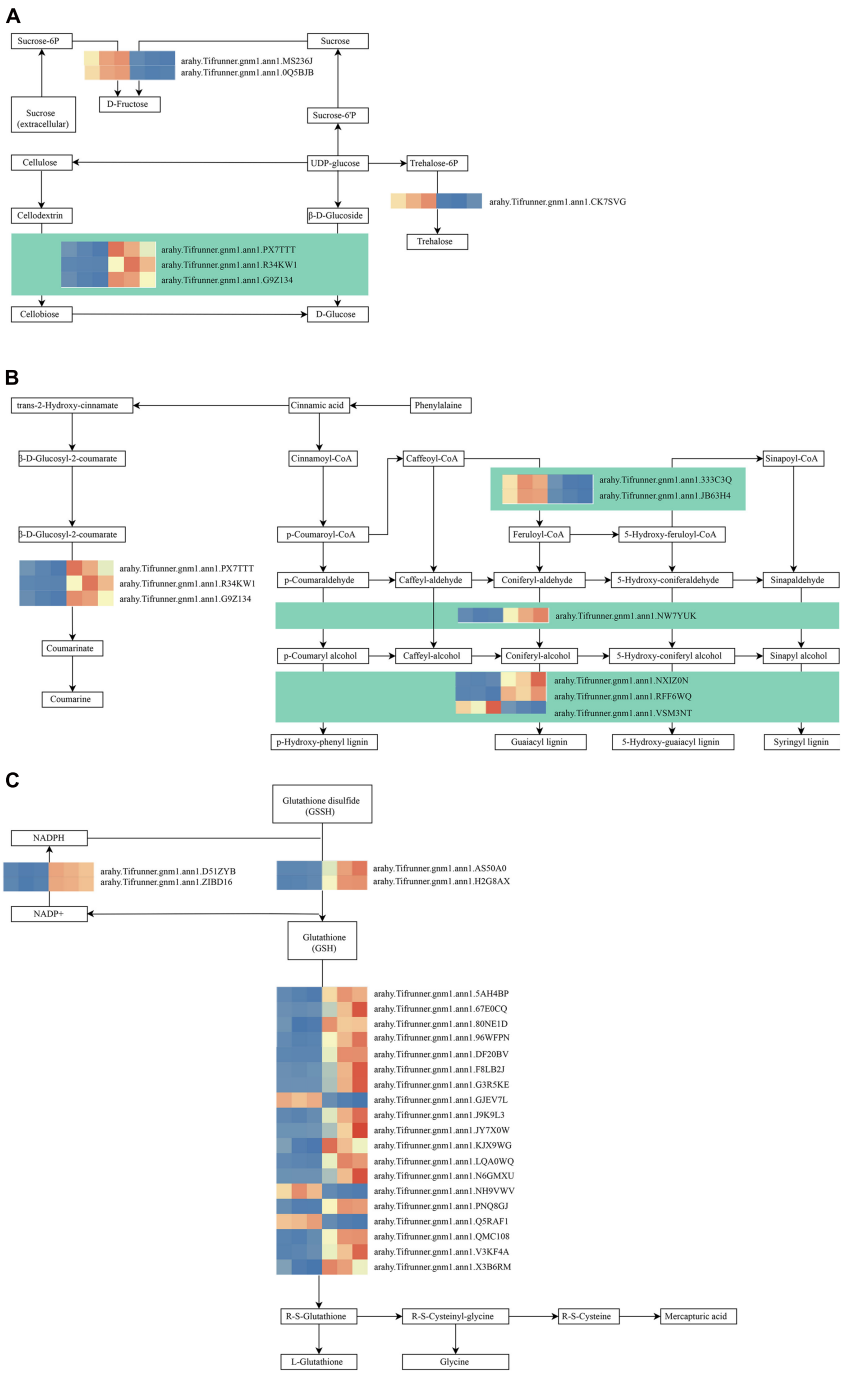


FIGURE 6 | Drought responsive genes in starch and sucrose metabolism **(A)**, flavonoid biosynthesis **(B)**, phenylpropanoid biosynthesis **(B)**, and Glutathione metabolism **(C)** pathways. Relative expression levels are normalized based on the Z-score and shown as a color gradient from low (blue) to high (red). The columns in heat map are 5, 7, and 9 days of well-watered condition, and 5, 7, and 9 days of drought-treated condition under severe drought from left to right, respectively.

The starch and sucrose metabolic process is widely identified in many plants under drought stress (Min et al., 2016; Ma et al., 2017; Khan et al., 2019; Hong et al., 2020). Trehalose, glucose, and sucrose are important soluble sugars to maintain cell osmotic potential (Konigshofer and Loppert, 2015; Han et al., 2016; Cui et al., 2019). Based on the analysis of soluble sugar content in the present study (**Figure 1E**), we hypothesized that these DEGs were involved in the biosynthesis of trehalose, glucose, and sucrose to maintain cell osmotic potential under severe drought stress. Several studies have highlighted the role of related genes in drought stress. For instance, overexpression of *OsTPS1* (trehalose-6-phosphate synthase) increased the amount

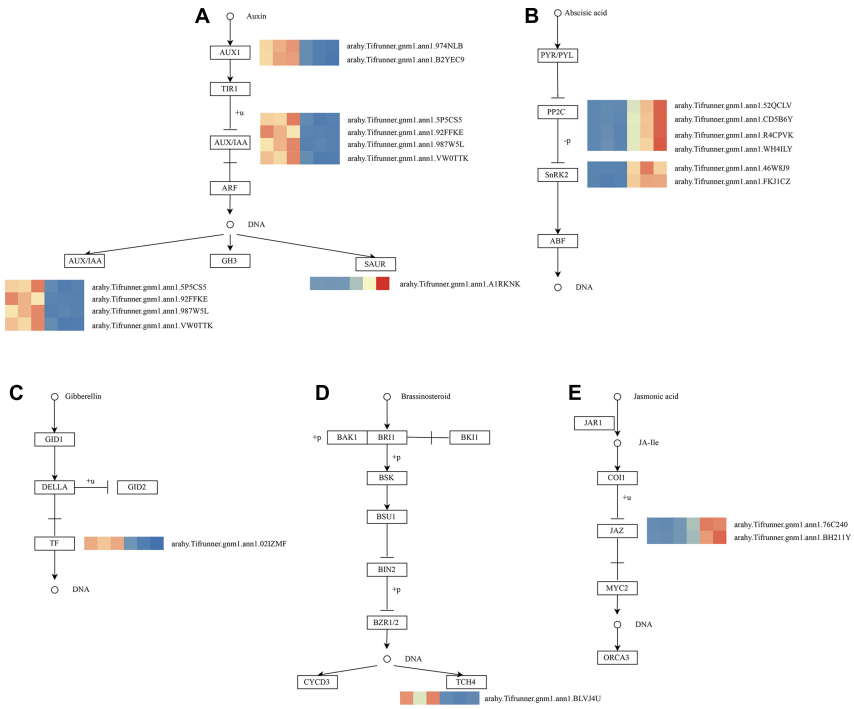


FIGURE 7 | Drought responsive genes in plant hormone signal transduction pathway. **(A)** IAA signal transduction pathway; **(B)** ABA signal transduction pathway; **(C)** GA signal transduction pathway; **(D)** BR signal transduction pathway; **(E)** JA signal transduction pathway. Relative expression levels are normalized based on the Z-score and shown as a color gradient from low (blue) to high (red). The columns in heat map are 5, 7, and 9 days of well-watered condition, and 5, 7 and 9 days of drought-treated condition under severe drought from left to right, respectively.

of trehalose and proline, and enhanced abiotic stress tolerance in plants (Li et al., 2011). Simultaneously, Invertases (INVs) plays an important role in primary metabolism and plant development, which can hydrolyze sucrose into glucose and fructose (Ruan et al., 2010; Min et al., 2016) and contribute to osmotic adjustment under water deficit conditions (Konigshofer and Loppert, 2015). A similar result has shown that starch and sucrose metabolism was significantly affected by drought stress in peanut (Gundaraniya et al., 2020). These results showed genes involved in the regulation of starch and sucrose metabolism may play an important role in drought stress.

Analysis of Secondary Metabolites Biosynthesis and Glutathione Metabolism in Response to Drought Stress

Biosynthesis of secondary metabolites such as phenylpropanoids and flavonoids is essential for a plant's response to stresses (Hernandez et al., 2009; Nakabayashi et al., 2014; Deng and Lu, 2017; Sharma et al., 2019). In our study, two caffeoyl-CoA 3-O-methyltransferase (CCoAOMT) related genes were down-regulated. *CCoAOMT1* and *COMT1* have a vital role in the biosynthesis of lignin, flavonoids, and sinapoyl malate in *Arabidopsis* (Do et al., 2007). Moreover, POD-encoding genes were also induced under severe drought stress. We speculated that these genes may participate in the regulation of peanut response to drought by combining the result of POD activity. Glutathione metabolism plays a key role in cellular defense (Noctor et al., 2002; Ball et al., 2004). In glutathione metabolism, glutathione can be oxidized to glutathione disulfide, and glutathione disulfide is again reduced to glutathione by glutathione reductase (Gong et al., 2018; Borgohain et al., 2019). And, dehydroascorbate (DHA) is reduced to ascorbic acid (Asc) by dehydroascorbate reductase (DHAR) in the presence of reduced glutathione, which in turn is regenerated by glutathione reductase (Chang et al., 2017). In our study, several key enzymes involved in glutathione metabolism were identified, which were also reported to be involved in drought tolerance regulation in *Oudneya* (Talbi et al., 2015). Consistently, Overexpression of *JcDHAR* can effectively enhance the tolerance to oxidative stress in plants (Chang et al., 2017). Therefore, the regulation of glutathione metabolism might contribute to drought tolerance in peanut under severe drought stress.

Analysis of Plant Hormone Signal Transduction and Protein Kinases in Response to Drought Stress

Multiple hormone-related pathways have been reported to be involved in the drought tolerance of plants. In this study, seven genes encoding protein involved in IAA signaling pathway were differentially expressed, including AUX1s, AUX/IAAs, and SAUR, under severe drought stress. A previous study has

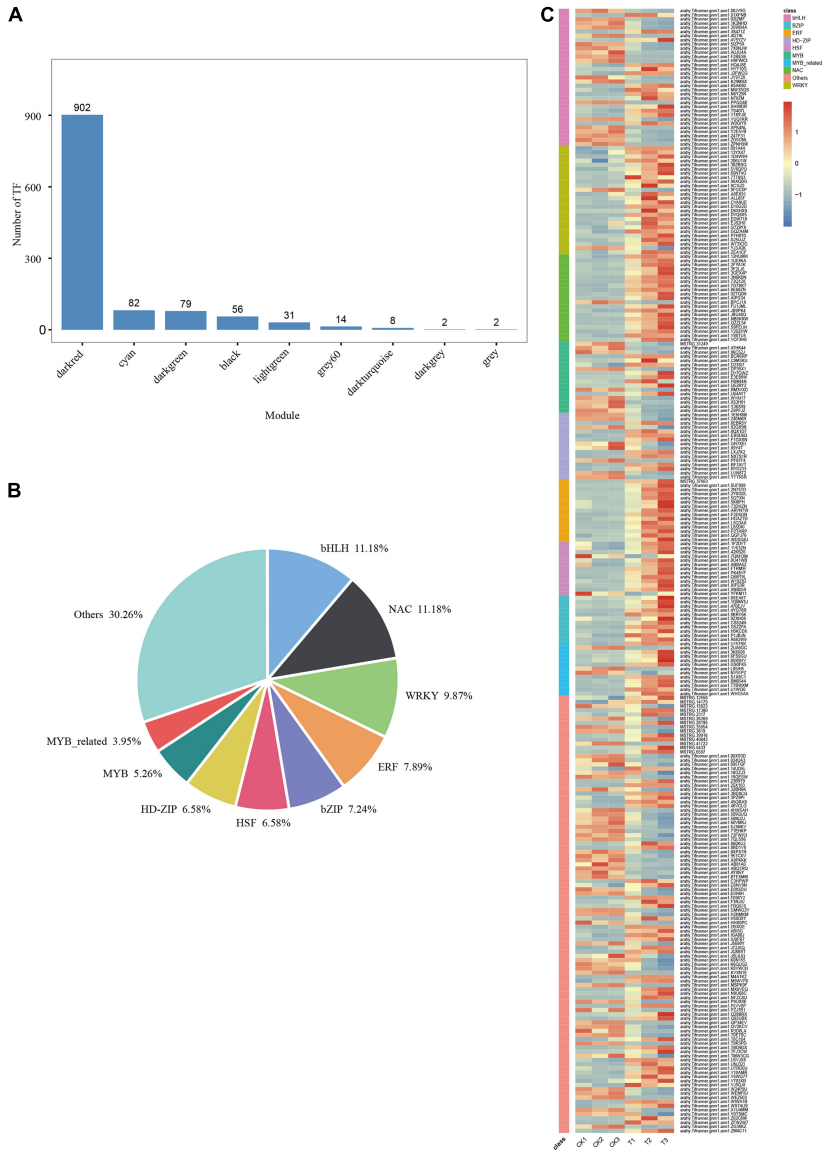


FIGURE 8 | WGCNA of transcription factors analysis in "L422." **(A)** The number of TFs for each module. **(B)** The proportion of genes in the top 9 abundant TF families in MM.darkred module. **(C)** Heat map of TF gene expression in MM.darkred module. Relative expression levels are normalized based on the Z-score and shown as a color gradient from low (blue) to high (red).

also demonstrated that overexpression of *OsIAA6* increased in transgenic rice drought tolerance (Jung et al., 2015). In another case, *TaSAUR75* transgenic Arabidopsis showed higher root length and survival rate under salt and drought stress (Guo et al., 2018). Typically, environmental stress is known to trigger changes in ABA levels and ABA regulates plant defense to drought stress (Verma et al., 2016). The central signaling complex PYR/PYL (Pyracbactin Resistance/Pyracbactin Resistance-like)-PP2Cs (Protein Phosphatase 2C)-SnRK2s (SNF1-Related Protein Kinases type 2) of ABA signaling pathway was activated in "L422." Among them, the genes encoding protein PP2Cs and SnRK2s were up-regulated but did not affect PYR/PYL. JA signaling pathway is associated with the alleviation of drought stress (Ali and Baek et al., 2020; Wang et al., 2020). Here, we found that two JAZ genes associated with JA signal transduction were predominantly expressed in "L422." A study found that OsJAZ1 could act as a transcriptional regulator of the OsbHLH148-related JA signaling pathway, leading to drought tolerance (Seo et al., 2011). Additionally, we found down-regulated *XTH23* and *PIL* genes in GA and BR signaling pathways, respectively. Interestingly, a previous study has reported that *XTH23* was induced from seed priming with BR on peanut under drought condition (Huang et al., 2020). Based on the analysis results, these DEGs may play a vital role via hormonal crosstalk in response to drought. Moreover, signaling pathways are induced under environmental stresses, in which

one of the major pathways is MAPK cascade in plant. MAPK cascade can convert environmental signals into molecular and cellular responses (Kaur and Gupta, 2005; Cheong and Kim, 2010; Sinha et al., 2011; Kumar et al., 2020). Previous findings clearly demonstrated that MAPK cascades were implicated in ABA and ethylene (ET) signaling (Zwerger and Hirt, 2001; Kar, 2011; de Zelicourt et al., 2016; Jagodzik et al., 2018). Interestingly, our study found that seven DEGs were detected in the ET and ABA signaling pathways. These genes will provide important implications for further research on the drought tolerance of peanuts.

Major TFs Involved in the Drought Response of Peanuts

Transcription factors as key regulators of transcription are important in plant responses to drought stress. In the present study, the bHLH family contained the most members, followed by NAC and WRKY families, indicating that they played an important role in coping with drought stress. Many transcription factors have been demonstrated to play an important role under drought stress in many crops. For example, a previous study reported that MdbHLH130 acts as a positive regulator of drought stress responses through modulating stomatal closure and ROS-scavenging in tobacco (Zhao Q. et al., 2020). Further, a recent study has shown that some *NAC* genes were induced under salt and drought stresses via RNA-seq and RT-qPCR analysis in peanut (Yuan et al., 2020). Interestingly, *AhNAC 65*, *AhNAC* 87, and *AhNAC 102* were induced in both drought and salt stresses, which were up-regulated in our result. We speculated that the three genes may play an important role in stress resistance. Although *NAC* 18 was only induced in salt stress, it was up-regulated in our result. Therefore, further studies of key NACs in our result will help to reveal the role of NACs in drought resistance in peanut. In soybean plant, GmWRKY54 conferred drought tolerance in transgenic soybean enhancing ABA/Ca^{2+} signaling pathways for stomatal closure and activating the expression of large numbers of stress-related TFs (Wei et al., 2019). Here, we found that 15 differentially expressed WRKYs were all up-regulated in the present study. Additionally, three (arahy.Tifrunner.gnm1.ann1.D15G2D, arahy.Tifrunner.gnm1.ann1.30ZBSQ, and arahy.Tifrunner.gnm1.ann1.P7H0T0) of 15 WRKYs were also induced and the rest of the WRKYs did not change in our previous study (**Supplementary Table 4**) (Zhao Q. et al., 2020). The roles of these WRKYs need to be elucidated in further investigation. Taken together, these differentially expressed TFs might be involved in response to drought stress, and they would provide important information for the study of drought tolerance in peanut.

CONCLUSION

In this study, we performed the physiological and transcriptomic analysis of "L422" at the pod formation stage under drought and well-watered conditions. Many DEGs were identified between well-watered and drought conditions by using RNA-Seq and WGCNA. The DEGs related to "MAPK signaling pathway-plant," "flavonoid biosynthesis," "starch and sucrose metabolism," "phenylpropanoid biosynthesis," "Glutathione metabolism," and "plant hormone signal transduction" were enriched in drought-tolerant cultivar. And numerous TF genes participated in the regulation networks under drought stress. The results provided a basis for further research on drought resistance genes in peanut.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories (BioProject ID: PRJNA706902).

AUTHOR CONTRIBUTIONS

LL and NZ designed the study and wrote the manuscript. NZ carried out the experiments and analyzed data. LL revised the manuscript. SC, BL, and HD carried out the field experiments for screening material. XL, YL, MH, XY, and GM assisted in writing the manuscript. All authors contributed to the manuscript and approved the final manuscript to publish.

FUNDING

This work was financially sponsored by the National Natural Science Foundation of China (32072097 and 31771833), the China Agriculture Research System (CARS-13), the Earmarked Fund for Hebei Oil Innovation Team of Modern Agroindustry Technology Research System (HBCT2018090202), the Startup Fund of State Key Laboratory for Crop Improvement and Regulation in North China (NCCIR2020RC-2), and the Startup Fund of Hebei Agricultural University (YJ2020044 and YJ2020058).

ACKNOWLEDGMENTS

We sincerely thank Dr. Meijing He (Hebei North University) and Dr. Oluwamosope Adeniji (Auburn University) for critically reviewing the manuscript. We thank Mr. Zhenyue Yuan (Gene Denovo Biotechnology Co., Guangzhou) and Gene Denovo Company (Guangzhou) for helping with transcriptome sequencing and valuable suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.672884/full#supplementary-material

Supplementary Figure 1 | Phenotypic and physiological changes of four cultivars with different drought tolerance under drought stress. **(A)** Phenotypic responses of four cultivars to drought stress. The changes of the relative water content **(B)** and relative electric conductivity **(C)** in leaves of four cultivars under well-watered and

drought conditions. Values are the mean ± standard deviation of three biological replicates.

Supplementary Figure 2 | Q-value heatmap of the GO significant pathway enrichment of the three main ontology for the 1,313 common DEGs in the MM.darkred module. The color scale indicates the Q-value.

Supplementary Table 1 | Sequences of the primers used in this study.

Supplementary Table 2 | Detailed information on the obtained reads *via* RNA-Seq.

Supplementary Table 3 | DEGs of six vital metabolic pathways in the MM.darkred module.

Supplementary Table 4 | Differentially expressed TFs in the MM.darkred module and differentially expressed WRKY TFs in previous study.

REFERENCES

Ali, M. S., and Baek, K. H. (2020). Jasmonic acid signaling pathway in response to abiotic stresses in plants. *Intern. J. Mol. Sci.* 21:621. doi: 10.3390/ijms21020621

Ball, L., Accotto, G. P., Bechtold, U., Creissen, G., Funck, D., Jimenez, A., et al. (2004). Evidence for a direct link between glutathione biosynthesis and stress fefense gene expression in *Arabidopsis*. *Plant Cell* 16, 2448–2462. doi: 10.1105/tpc.104.022608

Bhogireddy, S., Xavier, A., Garg, V., Layland, N., Arias, R., Payton, P., et al. (2020). Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. *Sci. Rep.* 10:4071. doi: 10.1038/s41598-020-60187-z

Blokhina, O., Virolainen, E., and Fagerstedt, K. V. (2003). Antioxidants, oxidative damage and oxygen deprivation stress: a review. *Ann. Bot.* 91, 179–194. doi: 10.1093/aob/mcf118

Borgohain, P., Saha, B., Agrahari, R., Chowardhara, B., Sahoo, S., van der Vyver, C., et al. (2019). SlNAC2 overexpression in *Arabidopsis* results in enhanced abiotic stress tolerance with alteration in glutathione metabolism. *Protoplasma* 256, 1065–1077. doi: 10.1007/s00709-019-01368-0

Budak, H., Hussain, B., Khan, Z., Ozturk, N. Z., and Ullah, N. (2015). From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. *Front. Plant Sci.* 6:1012. doi: 10.3389/fpls.2015.01012

Castilhos, G., Lazzarotto, F., Spagnolo-Fonini, L., Bodanese-Zanettini, M. H., and Margis-Pinheiro, M. (2014). Possible roles of basic helix-loop-helix transcription factors in adaptation to drought. *Plant Sci.* 223, 1–7. doi: 10.1016/j.plantsci.2014.02.010

Chang, L. M., Sun, H., Yang, H., Wang, X. H., Su, Z. Z., Chen, F., et al. (2017). Over-expression of dehydroascorbate reductase enhances oxidative stress tolerance in tobacco. *Electron. J. Biotechnol.* 25, 1–8. doi: 10.1016/j.ejbt.2016.10.009

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cheong, Y. H., and Kim, M. C. (2010). Functions of MAPK cascade pathways in plant defense signaling. *Plant Pathol. J.* 26, 101–109. doi: 10.5423/Ppj.2010.26.2.101

Cui, G. C., Zhang, Y., Zhang, W. J., Lang, D. Y., Zhang, X. J., Li, Z. X., et al. (2019). Response of carbon and nitrogen metabolism and secondary metabolites to drought stress and salt stress in plants. *J. Plant Biol.* 62, 387–399. doi: 10.1007/s12374-019-0257-1

Dang, P. M., Chen, C. Y., and Holbrook, C. C. (2013). Evaluation of five peanut (*Arachis hypogaea*) genotypes to identify drought responsive mechanisms utilising candidate-gene approach. *Funct. Plant Biol.* 40, 1323–1333. doi: 10.1071/Fp13116

de Zelicourt, A., Colcombet, J., and Hirt, H. (2016). The role of mapk modules and ABA during abiotic stress signaling. *Trends Plant Sci.* 21, 677–685. doi: 10.1016/j.tplants.2016.04.004

Deng, Q. Q., Dou, Z. M., Chen, J. W., Wang, K. L., and Shen, W. K. (2019). Drought tolerance evaluation of intergeneric hybrids of BC3F1 lines of *Saccharum officinarumx Erianthus arundinaceus*. *Euphytica* 215:207. doi: 10.1007/s10681-019-2513-3

Deng, Y. X., and Lu, S. F. (2017). Biosynthesis and regulation of phenylpropanoids in plants. *Crit. Rev. Plant Sci.* 36, 257–290. doi: 10.1080/07352689.2017.1402852

Ding, H., Zhang, Z., Kang, T., Dai, L., Ci, D., Qin, F., et al. (2017). Rooting traits of peanut genotypes differing in drought tolerance under drought stress. *Intern. J. Plant Product.* 11, 349–360. doi: 10.22069/ijpp.2017.3544

Do, C. T., Pollet, B., Thevenin, J., Sibout, R., Denoue, D., Barriere, Y., et al. (2007). Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in *Arabidopsis*. *Planta* 226, 1117–1129. doi: 10.1007/s00425-007-0558-3

Duan, M., Zhang, R. X., Zhu, F. G., Zhang, Z. Q., Gou, L. M., Wen, J. Q., et al. (2017). A lipid-anchored NAC transcription factor is translocated into the nucleus and activates Glyoxalase I expression during drought stress. *Plant Cell* 29, 1748–1772. doi: 10.1105/tpc.17.00044

Feng, J. L., Wang, L. Z., Wu, Y. N., Luo, Q. C., Zhang, Y., Qiu, D., et al. (2019). TaSnRK2.9, a sucrose non-fermenting 1-related protein kinase gene, positively regulates plant response to drought and salt stress in transgenic tobacco. *Front. Plant Sci.* 9:2003. doi: 10.3389/fpls.2018.02003

Fu, G. F., Song, J., Xiong, J., Li, Y. R., Chen, H. Z., Le, M. K., et al. (2011). Changes of oxidative stress and soluble sugar in anthers involve in rice pollen abortion under drought stress. *Agric. Sci. China* 10, 1016–1025. doi: 10.1016/S1671-2927(11)60089-8

Galmes, J., Flexas, J., Save, R., and Medrano, H. (2007). Water relations and stomatal characteristics of Mediterranean plants with different growth forms and leaf habits: responses to water stress and recovery. *Plant Soil* 290, 139–155. doi: 10.1007/s11104-006-9148-6

Girdthai, T., Jogloy, S., Vorasoot, N., Akkasaeng, C., Wongkaew, S., Holbrook, C. C., et al. (2010). Heritability of, and genotypic correlations between, aflatoxin traits and physiological traits for drought tolerance under end of season drought in peanut (*Arachis hypogaea* L.). *Field Crops Res.* 118, 169–176. doi: 10.1016/j.fcr.2010.05.007

Goche, T., Shargie, N. G., Cummins, I., Brown, A. P., Chivasa, S., and Ngara, R. (2020). Comparative physiological and root proteome analyses of two sorghum varieties responding to water limitation. *Sci. Rep.* 10:11835. doi: 10.1038/s41598-020-68735-3

Gong, B., Sun, S., Yan, Y., Jing, X., and Shi, Q. (2018). "Glutathione metabolism and its function in higher plants adapting to stress," in *Antioxidants and Antioxidant Enzymes in Higher Plants*, eds D. K. Gupta, J. M. Palma, and F. J. Corpas (Cham: Springer International Publishing), 181–205.

Gong, Z. Z., Xiong, L. M., Shi, H. Z., Yang, S. H., Herrera-Estrella, L. R., Xu, G. H., et al. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci. China Life Sci.* 63, 635–674. doi: 10.1007/s11427-020-1683-x

Gundaraniya, S. A., Ambalam, P. S., and Tomar, R. S. (2020). Metabolomic profiling of drought-tolerant and susceptible peanut (*Arachis hypogaea* L.) genotypes in response to drought stress. *ACS Omega* 5, 31209–31219. doi: 10.1021/acsomega.0c04601

Guo, Y., Jiang, Q. Y., Hu, Z., Sun, X. J., Fan, S. J., and Zhang, H. (2018). Function of the auxin-responsive gene TaSAUR75 under salt and drought stress. *Crop J.* 6, 181–190. doi: 10.1016/j.cj.2017.08.005

Guo, Y., Ping, W. J., Chen, J. T., Zhu, L. Y., Zhao, Y. F., Guo, J. J., et al. (2019). Meta-analysis of the effects of overexpression of WRKY transcription factors on plant responses to drought stress. *BMC Genet.* 20:63. doi: 10.1186/s12863-019-0766-4

Han, B. Y., Fu, L. L., Zhang, D., He, X. Q., Chen, Q., Peng, M., et al. (2016). Interspecies and intraspecies analysis of trehalose contents and the biosynthesis pathway gene family reveals crucial roles of trehalose in osmotic-stress tolerance in cassava. *Intern. J. Mol. Sci.* 17:1077. doi: 10.3390/ijms17071077

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal. Behav.* 6, 1503–1509. doi: 10.4161/psb.6.10.17088

Hao, L. Y., Liu, X. Y., Zhang, X. J., Sun, B. C., Liu, C., Zhang, D. F., et al. (2020). Genome-wide identification and comparative analysis of drought related genes in roots of two maize inbred lines with contrasting drought tolerance by RNA sequencing. *J. Integrat. Agric.* 19, 449–464. doi: 10.1016/S2095-3119(19)62660-2

Harb, A. (2016). "Identification of candidate genes for drought stress tolerance," in *Drought Stress Tolerance in Plants, Vol 2: Molecular and Genetic Perspectives*, eds M. A. Hossain, S. H. Wani, S. Bhattacharjee, D. J. Burritt, and L.-S. P. Tran (Cham: Springer International Publishing), 385–414.

Haro, R. J., Mantese, A., and Otegui, M. E. (2011). Peg viability and pod set in peanut: response to impaired pegging and water deficit. *Flora* 206, 865–871. doi: 10.1016/j.flora.2011.05.003

Hernandez, I., Alegre, L., Van Breusegem, F., and Munne-Bosch, S. (2009). How relevant are flavonoids as antioxidants in plants? *Trends Plant Sci.* 14, 125–132. doi: 10.1016/j.tplants.2008.12.003

Hong, Y., Ni, S. J., and Zhang, G. P. (2020). Transcriptome and metabolome analysis reveals regulatory networks and key genes controlling barley malting quality in responses to drought stress. *Plant Physiol. Biochem.* 152, 1–11. doi: 10.1016/j.plaphy.2020.04.029

Huang, L. P., Zhang, L., Zeng, R. E., Wang, X. Y., Zhang, H. J., Wang, L. D., et al. (2020). Brassinosteroid priming improves peanut drought tolerance via eliminating inhibition on genes in photosynthesis and hormone signaling. *Genes* 11:919. doi: 10.3390/genes11080919

Jagodzik, P., Tajdel-Zielinska, M., Ciesla, A., Marczak, M., and Ludwikow, A. (2018). Mitogen-activated protein kinase cascades in plant hormone signaling. *Front. Plant Sci.* 9:1387. doi: 10.3389/fpls.2018.01387

Jeyaramraja, P. R., Meenakshi, S. N., and Woldesenbet, F. (2018). Relationship between drought and preharvest aflatoxin contamination in groundnut (*Arachis hypogaea* L.). *World Mycotoxin J.* 11, 187–199.

Jiang, C., Li, X., Zou, J., Ren, J., Jin, C., Zhang, H., et al. (2021). Comparative transcriptome analysis of genes involved in the drought stress response of two peanut (*Arachis hypogaea* L.) varieties. *BMC Plant Biol.* 21:64. doi: 10.1186/s12870-020-02761-1

Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., et al. (2016). Transcription factors and plants response to drought stress: current understanding and future directions. *Front. Plant Sci.* 7:1029. doi: 10.3389/fpls.2016.01029

Jung, H., Lee, D. K., Choi, Y. D., and Kim, J. K. (2015). OsIAA6, a member of the rice Aux/IAA gene family, is involved in drought tolerance and tiller outgrowth. *Plant Sci.* 236, 304–312. doi: 10.1016/j.plantsci.2015.04.018

Kar, R. K. (2011). Plant responses to water stress: role of reactive oxygen species. *Plant Signal Behav.* 6, 1741–1745. doi: 10.4161/psb.6.11.17729

Kaur, N., and Gupta, A. K. (2005). Signal transduction pathways under abiotic stresses in plants. *Curr. Sci.* 88, 1771–1780.

Khan, R., Zhou, P., Ma, X., Zhou, L., Wu, Y., Ullah, Z., et al. (2019). Transcriptome profiling, biochemical and physiological analyses provide new insights towards drought tolerance in *Nicotiana tabacum* L. *Genes* 10:1041. doi: 10.3390/genes10121041

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

Konigshofer, H., and Loppert, H. G. (2015). Regulation of invertase activity in different root zones of wheat (*Triticum aestivum* L.) seedlings in the course of osmotic adjustment under water deficit conditions. *J. Plant Physiol.* 183, 130–137. doi: 10.1016/j.jplph.2015.06.005

Koolachart, R., Jogloy, S., Vorasoot, N., Wongkaew, S., Holbrook, C. C., Jongrungklang, N., et al. (2013). Rooting traits of peanut genotypes with different yield responses to terminal drought. *Field Crops Res.* 149, 366–378. doi: 10.1016/j.fcr.2013.05.024

Kumar, K., Raina, S. K., and Sultan, S. M. (2020). *Arabidopsis* MAPK signaling pathways and their cross talks in abiotic stress response. *J. Plant Biochem. Biotechnol.* 29, 700–714.

Kumar, P., Rouphael, Y., Cardarelli, M., and Colla, G. (2017). Vegetable grafting as a tool to improve drought resistance and water use efficiency. *Front. Plant Sci.* 8:1130. doi: 10.3389/fpls.2017.01130

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

Li, H. W., Zang, B. S., Deng, X. W., and Wang, X. P. (2011). Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice. *Planta* 234, 1007–1018. doi: 10.1007/s00425-011-1458-0

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

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Ma, J., Li, R. Q., Wang, H. G., Li, D. X., Wang, X. Y., Zhang, Y. C., et al. (2017). Transcriptomics analyses reveal wheat responses to drought stress during reproductive stages under field conditions. *Front. Plant Sci.* 8:592. doi: 10.3389/fpls.2017.00592

Manna, M., Thakur, T., Chirom, O., Mandlik, R., Deshmukh, R., and Salvi, P. (2020). Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. *Physiol. Plant.* 1–22. doi: 10.1111/ppl.13268

Mia, M. S., Liu, H., Wang, X. Y., Zhang, C., and Yan, G. J. (2020). Root transcriptome profiling of contrasting wheat genotypes provides an insight to their adaptive strategies to water deficit. *Sci. Rep.* 10:4854. doi: 10.1038/s41598-020-61680-1

Min, H. W., Chen, C. X., Wei, S. W., Shang, X. L., Sun, M. Y., Xia, R., et al. (2016). Identification of drought tolerant mechanisms in maize seedlings based on transcriptome analysis of recombination inbred lines. *Front. Plant Sci.* 7:1080. doi: 10.3389/fpls.2016.01080

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., et al. (2014). Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. *Plant J.* 77, 367–379. doi: 10.1111/tpj.12388

Noctor, G., Gomez, L., Vanacker, H., and Foyer, C. H. (2002). Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. *J. Exper. Bot.* 53, 1283–1304. doi: 10.1093/jexbot/53.372.1283

Osakabe, Y., Osakabe, K., Shinozaki, K., and Tran, L. S. (2014). Response of plants to water stress. *Front. Plant Sci.* 5:86. doi: 10.3389/fpls.2014.00086

Ozturk, M., Turkyilmaz Unal, B., Garcia-Caparros, P., Khursheed, A., Gul, A., and Hasanuzzaman, M. (2020). Osmoregulation and its actions during the drought stress in plants. *Physiol. Plant.* 1–15. doi: 10.1111/ppl.13297

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Ramachandra Reddy, A., Chaitanya, K. V., and Vivekanandan, M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* 161, 1189–1202. doi: 10.1016/j.jplph.2004.01.013

Rao, R. C. N., Singh, S., Sivakumar, M. V. K., Srivastava, K. L., and Williams, J. H. (1985). Effect of water deficit at different growth phases of peanut. I. Yield responses1. *Agron. J.* 77, 782–786.

Rao, R. C. N., Williams, J. H., and Singh, M. (1989). Genotypic sensitivity to drought and yield potential of peanut. *Agron. J.* 81, 887–893.

Reddy, A. R., Chaitanya, K. V., and Vivekanandan, M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* 161, 1189–1202. doi: 10.1016/j.jplph.2004.01.013

Ruan, Y. L., Jin, Y., Yang, Y. J., Li, G. J., and Boyer, J. S. (2010). Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Mol. Plant* 3, 942–955. doi: 10.1093/mp/ssq044

Seo, J. S., Joo, J., Kim, M. J., Kim, Y. K., Nahm, B. H., Song, S. I., et al. (2011). OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. *Plant J.* 65, 907–921. doi: 10.1111/j.1365-313X.2010.04477.x

Shao, H. B., Chu, L. Y., Jaleel, C. A., Manivannan, P., Panneerselvam, R., and Shao, M. A. (2009). Understanding water deficit stress-induced changes in the basic metabolism of higher plants - biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. *Crit. Rev. Biotechnol.* 29, 131–151. doi: 10.1080/07388550902869792

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. S. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules* 24, 2452. doi: 10.3390/molecules24132452

Sharma, P., Jha, A. B., Dubey, R. S., and Pessarakli, M. (2012). Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *J. Bot.* 2012:217037.

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exper. Bot.* 58, 221–227. doi: 10.1093/jxb/erl164

Sinha, A. K., Jaggi, M., Raghuram, B., and Tuteja, N. (2011). Mitogen-activated protein kinase signaling in plants under abiotic stress. *Plant Signal. Behav.* 6, 196–203. doi: 10.4161/psb.6.2.14701

Talbi, S., Romero-Puertas, M. C., Hernandez, A., Terron, L., Ferchichi, A., and Sandalio, L. M. (2015). Drought tolerance in a Saharian plant *Oudneya africana*: role of antioxidant defences. *Environ. Exper. Bot.* 111, 114–126. doi: 10.1016/j.envexpbot.2014.11.004

Tiwari, P., Srivastava, D., Chauhan, A. S., Indoliya, Y., Singh, P. K., Tiwari, S., et al. (2021). Root system architecture, physiological analysis and dynamic transcriptomics unravel the drought-responsive traits in rice genotypes. *Ecotoxicol. Environ. Saf.* 207:111252. doi: 10.1016/j.ecoenv.2020.111252

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

Wang, J., Song, L., Gong, X., Xu, J. F., and Li, M. H. (2020). Functions of Jasmonic Acid in Plant regulation and response to abiotic stress. *Intern. J. Mol. Sci.* 21:1446. doi: 10.3390/ijms21041446

Wei, W., Liang, D. W., Bian, X. H., Shen, M., Xiao, J. H., Zhang, W. K., et al. (2019). GmWRKY54 improves drought tolerance through activating genes in Abscisic acid and Ca2+ signaling pathways in transgenic soybean. *Plant J.* 100, 384–398. doi: 10.1111/tpj.14449

Yang, X., Luo, L., Yu, W., Mo, B., and Liu, L. (2019). Recent advances in the acclimation mechanisms and genetic improvement of peanut for drought tolerance. *Agric. Sci.* 10, 1178–1193.

Yuan, C. L., Li, C. J., Lu, X. D., Zhao, X. B., Yan, C. X., Wang, J., et al. (2020). Comprehensive genomic characterization of NAC transcription factor family and their response to salt and drought stress in peanut. *BMC Plant Biol.* 20:454. doi: 10.1186/s12870-020-02678-9

Zeng, J., Wu, C., Wang, C., Liao, F., Mo, J., Ding, Z., et al. (2020). Genomic analyses of heat stress transcription factors (HSFs) in simulated drought stress response and storage root deterioration after harvest in cassava. *Mol. Biol. Rep.* 47, 5997–6007. doi: 10.1007/s11033-020-05673-3

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17. doi: 10.2202/1544-6115.1128

Zhang, S. H., Xu, X. F., Sun, Y. M., Zhang, J. L., and Li, C. Z. (2018). Influence of drought hardening on the resistance physiology of potato seedlings under drought stress. *J. Integrat. Agric.* 17, 336–347. doi: 10.1016/S2095-3119(17)61758-1

Zhang, Z.-M., Wan, S.-B., Dai, L.-X., Song, W.-W., Chen, J., and Shi, Y.-Q. (2011). Estimating and screening of drought resistance indexes of peanut. *Chin. J. Plant Ecol.* 35:100. doi: 10.3724/SP.J.1258.2011.00100

Zhao, N., He, M., Li, L., Cui, S., Hou, M., Wang, L., et al. (2020). Identification and expression analysis of WRKY gene family under drought stress in peanut (*Arachis hypogaea* L.). *PLoS One* 15:e0231396. doi: 10.1371/journal.pone.0231396

Zhao, Q., Fan, Z. H., Qiu, L. N., Che, Q. Q., Wang, T., Li, Y. Y., et al. (2020). MdbHLH130, an Apple bHLH transcription factor, confers water stress resistance by regulating stomatal closure and ROS homeostasis in transgenic tobacco. *Front. Plant Sci.* 11:543696. doi: 10.3389/fpls.2020.543696

Zhao, X. B., Li, C. J., Wan, S. B., Zhang, T. T., Yan, C. X., and Shan, S. H. (2018). Transcriptomic analysis and discovery of genes in the response of *Arachis hypogaea* to drought stress. *Mol. Biol. Rep.* 45, 119–131. doi: 10.1007/s11033-018-4145-4

Zhu, H., Zhou, Y. Y., Zhai, H., He, S. Z., Zhao, N., and Liu, Q. C. (2019). Transcriptome profiling reveals insights into the molecular mechanism of drought tolerance in sweetpotato. *J. Integr. Agric.* 18, 9–23. doi: 10.1016/S2095-3119(18)61934-3

Zwerger, K., and Hirt, H. (2001). Recent advances in plant MAP kinase signalling. *Biol. Chem.* 382, 1123–1131. doi: 10.1515/BC.2001.142

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.

Copyright © 2021 Zhao, Cui, Li, Liu, Deng, Liu, Hou, Yang, Mu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fgene.2021.668940

Genome-Wide Analysis of Coding and Non-coding RNA Reveals a Conserved miR164–NAC–mRNA Regulatory Pathway for Disease Defense in *Populus*

*Sisi Chen[1,2,3], Jiadong Wu[1,2,3], Yanfeng Zhang[2,3], Yiyang Zhao[1,2,3], Weijie Xu[1,2,3], Yue Li[1,2,3] and Jianbo Xie[1,2,3]**

[1] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [2] National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [3] Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

OPEN ACCESS

Edited by:
Shang-Qian Xie, Hainan University, China

Reviewed by:
Chaoling Wei, Anhui Agriculture University, China
Jinhui Chen, Hainan University, China

***Correspondence:**
Jianbo Xie
jbxie@bjfu.edu.cn

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

Received: *17 February 2021*
Accepted: *29 March 2021*
Published: *28 May 2021*

Citation:
Chen S, Wu J, Zhang Y, Zhao Y, Xu W, Li Y and Xie J (2021) Genome-Wide Analysis of Coding and Non-coding RNA Reveals a Conserved miR164–NAC–mRNA Regulatory Pathway for Disease Defense in Populus. Front. Genet. 12:668940. doi: 10.3389/fgene.2021.668940

MicroRNAs (miRNAs) contribute to plant defense responses by increasing the overall genetic diversity; however, their origins and functional importance in plant defense remain unclear. Here, we employed Illumina sequencing technology to assess how miRNA and messenger RNA (mRNA) populations vary in the Chinese white poplar (*Populus tomentosa*) during a leaf black spot fungus (*Marssonina brunnea*) infection. We sampled RNAs from infective leaves at conidia germinated stage [12 h post-inoculation (hpi)], infective vesicles stage (24 hpi), and intercellular infective hyphae stage (48 hpi), three essential stages associated with plant colonization and biotrophic growth in *M. brunnea* fungi. In total, 8,938 conserved miRNA-target gene pairs and 3,901 *Populus*-specific miRNA-target gene pairs were detected. The result showed that *Populus*-specific miRNAs (66%) were more involved in the regulation of the disease resistance genes. By contrast, conserved miRNAs (>80%) target more whole-genome duplication (WGD)-derived transcription factors (TFs). Among the 1,023 WGD-derived TF pairs, 44.9% TF pairs had only one paralog being targeted by a miRNA that could be due to either gain or loss of a miRNA binding site after the WGD. A conserved hierarchical regulatory network combining promoter analyses and hierarchical clustering approach uncovered a miR164–NAM, ATAF, and CUC (NAC) transcription factor–mRNA regulatory module that has potential in *Marssonina* defense responses. Furthermore, analyses of the locations of miRNA precursor sequences reveal that pseudogenes and transposon contributed a certain proportion (∼30%) of the miRNA origin. Together, these observations provide evolutionary insights into the origin and potential roles of miRNAs in plant defense and functional innovation.

Keywords: microRNA, defense response, infection, poplar, *Marssonina brunnea*

INTRODUCTION

MicroRNAs (miRNAs) are ∼21 to 24-nucleotide (nt) non-coding endogenous small RNAs that can regulate gene expression, maintain genome integrity and chromatin structure, and influence plant development and stress response (Carrington and Ambros, 2003; Jones-Rhoades et al., 2006; Voinnet, 2009; Sunkar et al., 2012; Meyers and Axtell, 2019). Under pathogen stress, basal defense and resistance gene-mediated resistance are the two well-defined defense responses carried out by plants. Innate immunity is an evolutionarily ancient mechanism that protects plants from a wide range of pathogens (Peláez and Sanchez, 2013). Many lines of evidence have confirmed that miRNAs contribute to plant defenses against pathogens (Li et al., 2012; Pumplin and Voinnet, 2013; Thiebaut et al., 2015; Yang and Huang, 2015). Evolutionary analyses revealed that a miRNA superfamily composed of the miR482 and miR2118 families targets the plant nucleotide-binding leucine-rich-repeat (NB-LRR) defense genes (Zhao et al., 2015). Moreover, miRNAs–transcription factor (TFs) regulation module was proposed to be ubiquitous in plant defense and plays key roles in regulation networks controlling many biological processes, including responses to biotic and abiotic stresses (Seo et al., 2015; Thiebaut et al., 2015).

As we know, the *Populus* genus consists of many important woody species, such as the western balsam poplar (*Populus trichocarpa*) (Tuskan et al., 2006), the desert poplar (*Populus euphratica*) (Ma et al., 2013), and the Chinese white poplar (*Populus tomentosa*) (Du et al., 2014). Several species have been selected as model tree species for their small genome size and rapid growth. The availability of reference genome sequences for *Populus* species thus makes them important model systems for the investigation of miRNA functions during pathogen infections. To date, hundreds of miRNAs have been identified in *Populus*, and the function of several well-known miRNAs has been clarified in literatures (Zhang et al., 2010; Kozomara and Griffiths-Jones, 2014). During the infection of bacterial or fungal pathogens, the transcription patterns of poplar miRNAs were highly associated with the disease resistance (DR) response (Zhao et al., 2012; Li et al., 2016). An economically important group of poplar pathogens, *Marssonina brunnea*, is a typical hemibiotrophic fungal pathogen, which can cause disease *Marssonina* leaf spot of poplars (MLSP) (Zhang et al., 2018). Although MLSP has been studied for over 30 years, the key non-coding RNAs that function during *M. brunnea* infection and their effects on plant defense are poorly understood. Therefore, increasing molecular understanding of the plant-*M. brunnea* interaction will be helpful for the development of control strategies against MLSP.

In the present study, we combined transcriptome and genomic analyses to explore the origin, evolution, functional innovation, and plant defense effects of the poplar miRNAs during the three essential stages of MLSP fungus (*M. brunnea*) infection, including conidia germinated stage [12 h post-inoculation (hpi)], infective vesicle stage (24 hpi), and intercellular infective hyphae stage (48 hpi), as described in Chen et al. (2020). By exploring the origin and evolutionary patterns of poplar miRNAs, our study provides new insight into the feedback regulation mechanism of miRNAs. Besides, our study attempts to compare the regulation mechanism between conserved and *Populus*-specific miRNA to achieve a better understanding of the coevolution between miRNAs and target genes. Finally, a conserved hierarchical regulatory network combining promoter analyses and hierarchical clustering approach reveals a miR164–NAM, ATAF, and CUC (NAC) transcription factor–messenger RNA (mRNA) regulatory module that has potential in *Marssonina* defense responses. Overall, this study reveals the evolutionary patterns and illustrates the functional novelty of lineage-specific miRNAs in DR processes.

MATERIALS AND METHODS

Plant Materials and Fungal Treatments

The Chinese white poplar (*P. tomentosa* cv. "LM50" clones) was planted in pots under natural light conditions (12 h of 1,250 μmol m^{-2} s^{-1} photosynthetically active radiation) at 25°C ± 2°C (day and night) and 50% ± 1% relative humidity (day and night) in an air-conditioned glasshouse using soilless culture technology. The infection experiments were performed using *M. brunnea* f. sp. *Monogermtubi* strain bj01. Six inoculation spots per poplar leaf were each inoculated with 5 μl of 10^5 condia ml^{-1}, with three biological replicates per treatment. The spore suspension was sprayed onto the abaxial surfaces of the leaves *in vitro*, and the inoculated leaves were incubated in an artificial climate incubator (LT36VL, Percival Scientific, Inc., Perry, United States) under 25°C and 95% relative humidity and harvested at 12, 24, and 48 hpi. In the control group, the leaves were sprayed with sterile tap water and harvested at 12 hpi. The samples from leaves exposed to the same treatment were pooled and treated as one biological repeat, and two independent experimental repeats were performed for each treatment (CK, 12, 24, and 48 hpi). All samples were immediately frozen in liquid nitrogen and stored at −80°C for RNA extraction.

MicroRNA Library Construction

Total RNA was extracted from inoculated leaf samples using a TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions. Additional on-column DNase digestion was performed during the RNA purification using RNase-Free DNase (Qiagen). The total RNAs were ligated with 3′ and 5′ adapters using a Small RNA Sample Prep Kit (Illumina). sRNAs with adapters on both ends were used as templates to create cDNA constructs using reverse transcription PCR. After being purified and quantified using a Qubit dsDNA HS (Qubit 2.0 Fluorometer) and Agilent 2100, the PCR products were used for cluster generation and sequencing on an Illumina HiSeq 4000 according to the cBot and Hiseq 4000 user guides, respectively.

Messenger RNA Sequencing, Alignment, and Normalization

mRNA reads were aligned using TopHat2 (Kim et al., 2013) using—*read-mismatches* 2-*p*-*G* to generate read alignments for

each sample. Up to two mismatches were permitted in each read alignment. The transcript abundance was calculated, and differential transcript expression was computed using CuffDiff with the parameters-*p*-*b* (Ghosh and Chan, 2016).

Small RNAome Analysis

The sequences generated from the leaves exposed to the three infection treatments were used to detect the transcript abundance of mature sRNAs. All sRNA reads, referred to as raw reads, were processed to remove adaptors, low-quality tags, and contaminants. Clean reads were then mapped to version 3.0 of *P. trichocarpa* genome with no more than one mismatch. These perfectly aligned sequences were annotated by BLAST-searching them against the GenBank and Rfam databases (version 13[1]), allowing one mismatch. The tRNAs, rRNAs, snRNAs, snoRNAs, and scRNAs were removed from the sequencing reads. The remaining unannotated sRNAs were searched against the known miRNAs from miRBase version 22.1[2], allowing a maximum of two mismatches. Then, the remaining unannotated unique sequences were mapped to the *P. trichocarpa* genome to uncover novel miRNAs from poplar, according to the established criteria (Meyers et al., 2008), using Mireap software[3]. Finally, only miRNAs with high expression levels (actual count of reads exceed 10 in at least one sample) and loci that could produce both mature miRNAs and antisense miRNA (miRNA*) sequences were kept in our study.

Genomic Locations of MicroRNAs in the *P. trichocarpa* Genome

The pre-miRNAs were screened for their localization within the transposons, introns, exons, pseudogenes, intergenic regions, 5′ untranslated region (UTR), 3′ UTR, coding sequence (CDS), and promoter (upstream 2 kb of coding genes) region of the *P. trichocarpa* genome v3.0 (Phytozome version 12), with an overlapping rate of above 80%.

Target Prediction and Functional Annotation

Targets of each miRNA were predicted using Web server psRNATarget[4], with *P. trichocarpa* transcript (phytozome v10.0, genome V3.0, internal number 210) as target gene search scope, the expectation is less than 5, and other parameters as default. The functional annotation and categorization of candidate miRNA targets were performed using the AgriGO software suite v2.0[5] with default parameter (Tian et al., 2017). Regulatory networks were drawn using Cytoscape version 3.8.0 (Shannon, 2003).

Transposon Element Annotation

The transposon element (TE) annotations used in this study were obtained from the outputs of the RepeatMasker (RM) software version 4.0.7 combined with the database (Dfam_Consensus-20170127, RepBase-20170127; -species parameter: *Populus*). These RM outputs were filtered to remove non-TE elements, such as satellites, simple repeats, low complexity sequences, and rRNA.

Pseudogene Annotation

The intergenic sequences of the *P. trichocarpa* genome were used to identify the putative pseudogenes. The overall pipeline used for this identification was generally based on the PlantPseudo workflow (Xie et al., 2019) and consisted of four major steps: (1) identify the masked intergenic regions with sequence similarity to known proteins using BLAST; (2) eliminate redundant and overlapping BLAST hits in places where a given chromosomal segment has multiple hits; (3) link homologous segments into contigs; and (4) realign sequences using tfasty to identify features that disrupt contiguous protein sequences.

Real-Time Quantitative PCR

RT-qPCR was performed on a 7,500 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA, United States) using the SYBR Green *Premix Ex* Taq II (TaKaRa). All primer pairs for the candidate genes were designed by an online tool provided by Integrated DNA Technologies[6], as shown in **Supplementary Data 1**. Poplar 18S rRNA was used as an internal control for gene expression measurements for target genes. The Mir-X^{TM} miRNA qRT-PCR SYBR Kit (Clontech, Mountain View, CA) was used. The relative expression level of each miRNA was measured and standardized to 5.8S rRNA. The relative expression of miRNAs and target genes was calculated using the $2^{-\Delta\Delta Ct}$ method.

RESULTS

Expression Dynamics of Poplar MicroRNAs During the *Marssonina* Infection

To study the posttranscriptional regulation associated with poplar defense to *Marssonina*, we inoculated the leaves of *P. tomentosa* LM50 clones with *M. brunnea* f. sp. *Monogermtubi* bj01 conidial suspension, and the expression pattern of miRNAs was investigated by small RNA sequencing. Specifically, we prepared a library of RNAs of 18–30 nucleotides (nt) from each sample (CK, 12, 24, and 48, two biological repeats), generating 17.5 million reads in total; 83–86.7% of the total reads could be aligned perfectly (no more than one mismatch) to the *P. trichocarpa* genome (version 3.0) (**Supplementary Data 2**; Tuskan et al., 2006). In total, we identified 131 miRNA precursors by alignment to miRbase v22.1, which were grouped into 37 miRNA families containing an average of 3.5 genes per family (**Supplementary Data 3**). A total of 21 sequences were identified as potential novel miRNAs, with average minimum free energy (MFE) of -58.4 kal/mol (**Supplementary Figure 1** and **Supplementary Data 3**). Compared to known miRNAs, the expression of novel miRNAs was generally low (**Figure 1A** and

[1]http://rfam.xfam.org/

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

[3]https://sourceforge.net/projects/mireap/

[4]http://plantgrn.noble.org/psRNATarget/

[5]http://systemsbiology.cau.edu.cn/agriGOv2

[6]https://sg.idtdna.com/Scitools/Applications/RealTimePCR/

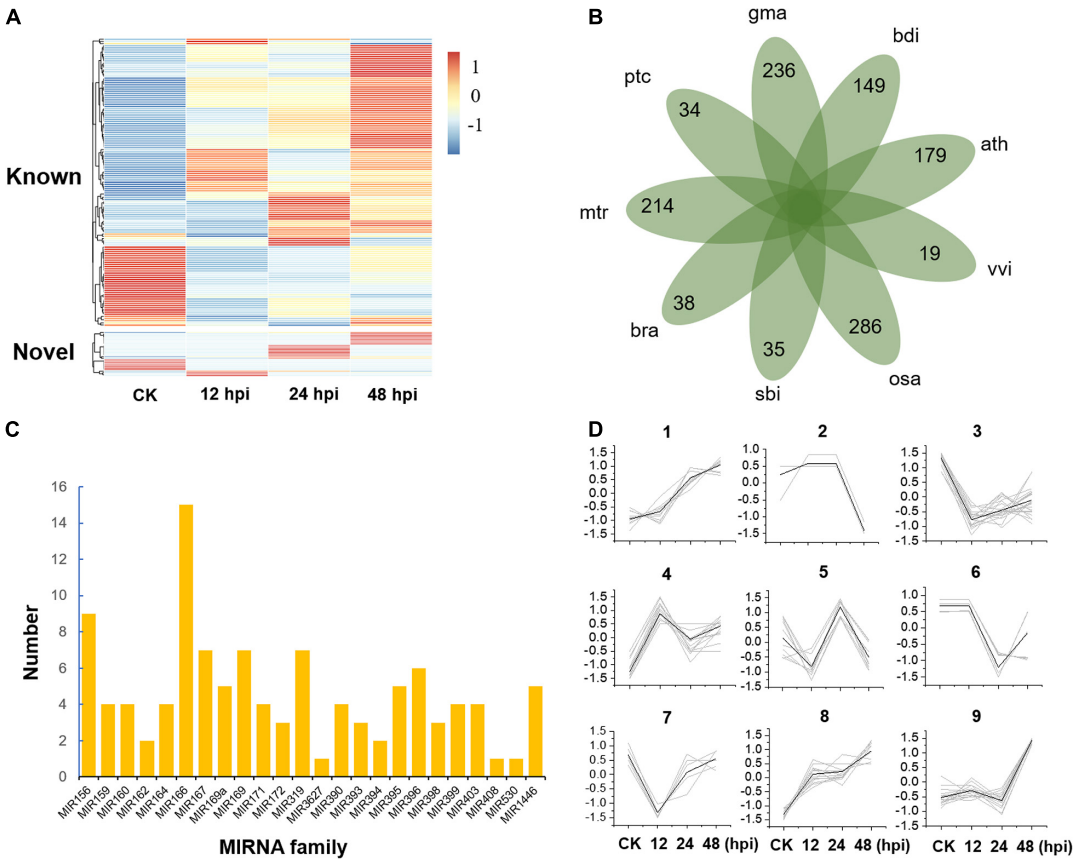


FIGURE 1 | Expression and classification of microRNAs (miRNAs). **(A)** Expression of known and novel miRNAs. hpi, hours post-inoculation. **(B)** The species-specific miRNA families of nine plant species. ptc, *Populus trichocarpa*; mtr, *Medicago truncatula*; bra, *Brassica rapa*; sbi, *Sorghum bicolor*; osa, *Oryza sativa*; vvi, *Vitis vinifera*; ath, *Arabidopsis thaliana*; bdi, *Brachypodium distachyon*; gma, *Glycine max*. **(C)** The distribution of 110 conserved miRNA members in 23 miRNA families. **(D)** Nine clusters were obtained by *K*-means clustering with Euclidean distance as the distance metric.

Supplementary Data 4). Some novel miRNAs were discovered in only one of the libraries partly because the sequencing depth provided insufficient coverage of all the miRNAs or some miRNA expressions are specifically turned on or turned off by pathogen stress. A total of 110 conserved miRNAs belonging to 23 miRNA families and 42 *Populus*-specific miRNAs belonging to 34 miRNA families were identified (**Figures 1B,C** and **Supplementary Data 5**).

To study the global expression patterns, we also performed k-means clustering to describe the expression of miRNA during poplar response to *M. brunnea*. We detected nine co-expression clusters (**Figure 1D** and **Supplementary Data 6**). Cluster 1 and cluster 8 contained 18 and 25 miRNAs, respectively, showing a consistent increase during defense response. While cluster 3 showed a reverse trend. Cluster 9 contained 21 miRNAs, which show their expression peaks at 48 hpi. MiRNAs of cluster 4 and cluster 5 showed their expression peaks at 12 and 24 hpi, respectively (**Figure 1D** and **Supplementary Data 6**). Gene Ontology (GO) enrichment analysis showed that most of the target genes of cluster 1 and cluster 8 are involved in cell adenyl ribonucleotide binding, ATP binding, and protein kinase activity. Targets from miRNAs in Cluster 3 were significantly enriched in the biological process of regulating the primary metabolism, biosynthesis, and transcription ($P < 1 \times 10^{-4}$). Overall, miRNAs play essential roles in stress response by activating or suppressing the expression of their target genes.

Pseudogenes and Transposons Act as Catalysts for the Formation of MicroRNA

To elucidate the underlying mechanism of emergence of (conserved and *Populus*-specific) miRNAs, we examined the locations of miRNA precursor sequences (*MIRs*) in the regions of *P. trichocarpa* genome, including intragenic regions (exons, introns, CDS, and UTRs) and intergenic regions (**Supplementary Data 7**). These conserved or *Populus*-specific miRNAs were extensively distributed in poplar genomes (**Figure 2A**). Of these miRNAs, 52 (34.21%) were located within protein-encoding genes (PEGs), five (3.29%) were in unclassified sequences (scaffold), and 95 (62.50%) were in the intergenic region (**Figure 2B**). Of the 52 miRNAs within PEGs, 11 (7.24%) were in intron regions, 20 (13.16%) were in CDS regions, 18 (11.84%) were in 5′ UTR regions, three (1.97%) were in 3′ UTR regions (**Figure 2B**). Notably, nearly half of these intragenic-derived miRNAs (24 out of 52) showed the same transcriptional

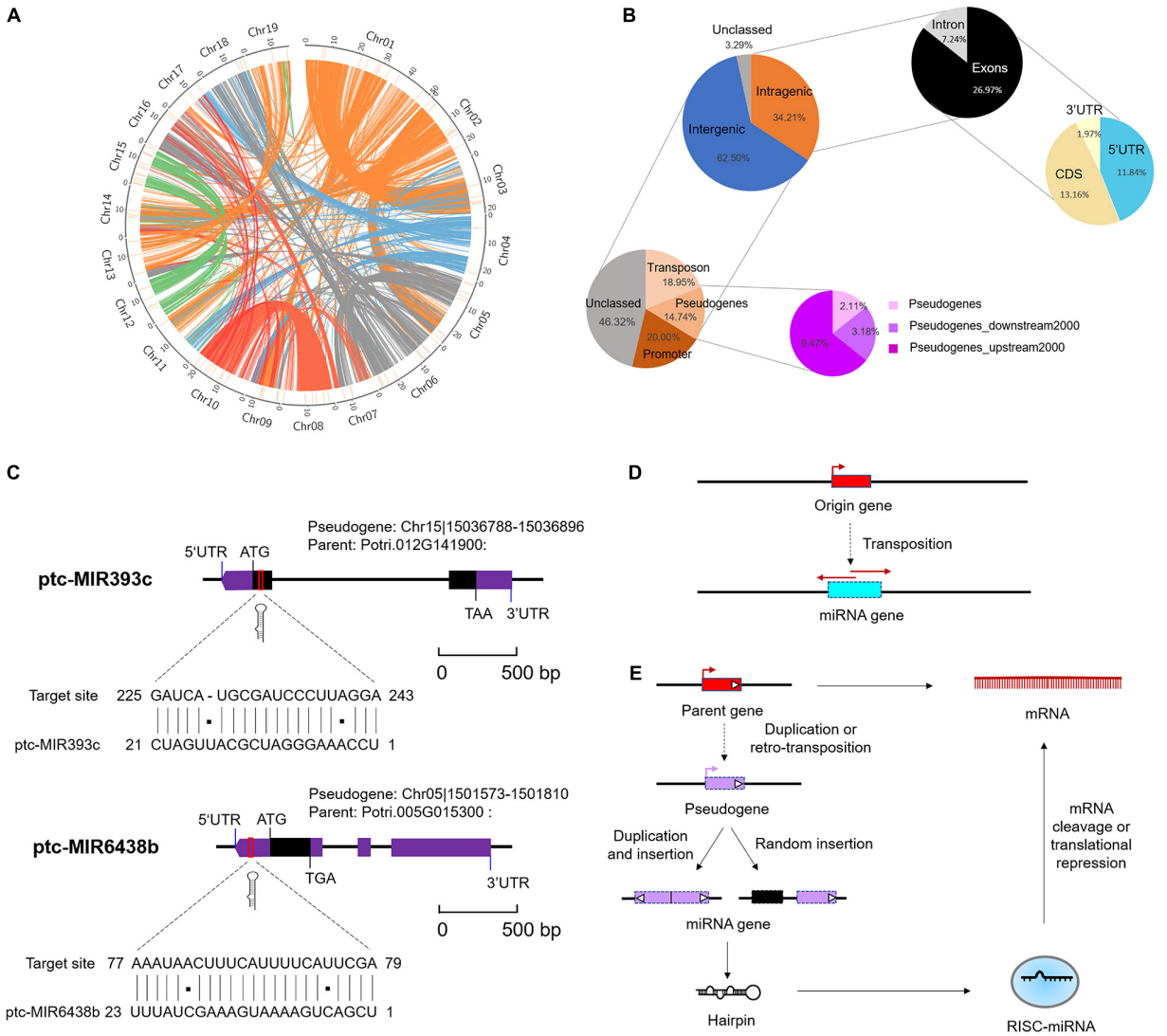


FIGURE 2 | Genomic locations of microRNAs (miRNAs) in *Populus trichocarpa* genome. **(A)** Genomic distribution of 152 *Marssonina brunnea* responsive miRNAs. Chromosomes are represented by the circle, and the inner circles (short orange lines) represent the location of miRNAs at the genome. The central colorful lines represent lines that connect syntenic block across chromosomes. **(B)** miRNA locations at the intergenic and protein-encoding genes (PEGs) region. The two charts on the top right indicate the locations of miRNA precursor sequences (*MIRs*) in the regions of PEGs. The two charts on the bottom left indicate the detailed classifications of *MIRs* overlapping with transposon, pseudogenes, and promoter region. **(C)** miRNA origin from pseudogenes and targeted parent gene. miR393c and miR6438b were selected as representative miRNAs. **(D)** Mechanism of miRNA gene origin from transposon. **(E)** Mechanism of miRNA gene origin from pseudogenes and feedback regulated the parent gene.

orientation as their host genes, indicating that the transcription of the miRNAs may associate with the host genes.

Moreover, we further examined the location of *MIRNA* precursors in relation to transposon region, pseudogenes, and promoter region (2-kb sequences of genes upstream) in poplar genome. As a result, we detected 24 (15.79%) miRNAs in the transposon region, 22 (14.47%) were in pseudogenes, and 23 (15.13%) were in the promoter region (**Figure 2B** and **Supplementary Data 7**). Notably, the proportions of miRNAs located in pseudogenes and transposons were significantly larger than expected by chance ($P < 1 \times 10^{-3}$, one-sided z test). This suggests that pseudogenes and TEs may contribute to the origin of *Populus* miRNAs. A careful examination of their precursor sequences revealed that *Populus*-specific miR478e and miR6427 were transposons-derived miRNAs; conserved miR393c and *Populus*-specific miR6438b were pseudogenes-derived miRNAs (**Figure 2C**). The *Populus*-specific miR6438b was derived from upstream 192–215 bp of pseudogene Chr13| 15188492-15190233 and targeted the 5′ UTR of its parent gene (Potri.005G015300); and conserved miR393c were derived from 2 to 23 bp of pseudogene Chr04| 22273397-22273670 and targeted the CDS of its parent gene (Potri.012G141900) (**Figure 2C**). Together, the strong association of miRNAs with pseudogenes and TE provides an important mechanism

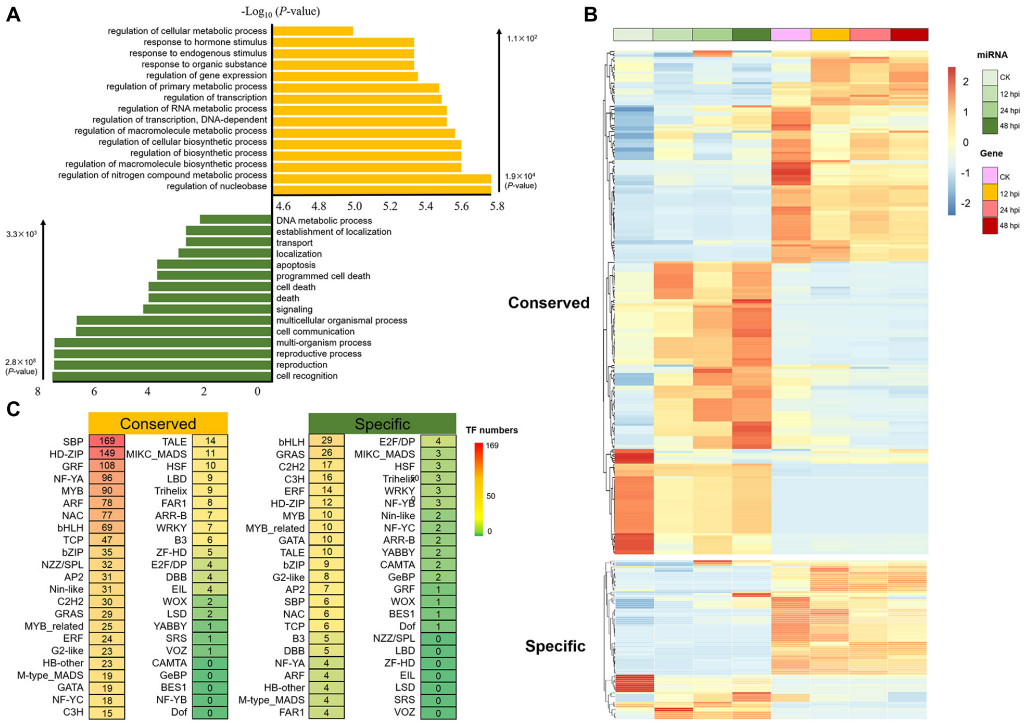


FIGURE 3 | Analysis of the target of conserved and *Populus*-specific miRNAs. **(A)** Significantly enriched Gene Ontology biological processes for target genes of conserved and *Populus*-specific miRNAs, respectively (Top 15; $P < 1 \times 10^{-2}$). **(B)** Expression of the miRNA/target genes with negatively correlated expression patterns ($r < -9 \times 10^{-1}$, $P < 5 \times 10^{-2}$; Pearson correlation). **(C)** The number of transcription factor (TF) targets of conserved and *Populus*-specific miRNAs, respectively.

for the origin and posttranscriptional regulation of miRNAs (**Figures 2D,E**).

The *Populus* MicroRNAs Fine-Tune the Expression of Disease Resistance Genes

To investigate the regulatory networks associated with these *M. brunnea*-responsive miRNAs, 12,839 predicted miRNA-target gene pairs were identified by using psRNATarget (Dai and Zhao, 2011). It contains 8,938 conserved miRNA-target gene pairs and 3,901 *Populus*-specific miRNA-target gene pairs (**Supplementary Data 8**). The results of GO functional enrichment showed that target genes of conserved miRNA were mainly concentrated on regulating metabolism and transcription, whereas *Populus*-specific miRNAs were significantly enriched in the process of signaling and programmed cell death ($P < 1 \times 10^{-2}$; **Figure 3A**). An examination of the expression of the miRNA/target genes revealed 321 pairs with negatively correlated expression patterns ($r < -9 \times 10^{-1}$, $P < 5 \times 10^{-2}$; Pearson correlation; **Figure 3B** and **Supplementary Data 9**). Using the publicly available degradome library (Xie et al., 2017), we identified 247 miRNA/target pairs (**Supplementary Data 10**), of which several conserved miRNA/target pairs (miR156-*SPL*, miR164-*NAC*, and miR172-*RAP*) were observed in our dataset (Wang et al., 2009, 2015, 2020).

Next, we performed target prediction analyses. As a result, a total of 114 DR genes and 123 DR genes were predicted to be the targets of the conserved and *Populus*-specific miRNAs, respectively. Notably, a larger proportion of *Populus*-specific miRNAs (28 of 42) was found to target DR genes than that of conserved miRNAs (60 of 110). The results suggest that *Populus*-specific miRNAs were more involved in the regulation of the DR genes ($P = 1.07 \times 10^{-11}$; Fisher's exact test; **Supplementary Data 8**). For instance, a 22-nt *Populus*-specific ptc-miRN11 and a 23-nt *Populus*-specific ptc-miR6478 were predicted to target 35 and 11 DR genes, respectively (**Figure 4**). To gain a better understanding of the functional roles of miRNAs, we next performed a pfam domain analysis of DR targets. A total of 48 TIR-NBS-LRR (TNL), 19 CC-NBS-LRR (CNL), 99 NBS (N), 61 NBS-LRR (NL), one TIR, and two LRR family proteins were detected. Among them, two 24-nt *Populus*-specific families, ptc-miR6445 and ptc-miR1447, were predicted to target a TNL (Potri.019G069200) and an N (Potri.012G123000) DR gene, respectively. Notably, an examination of the expression patterns of the DR genes showed that these genes were expressed at a very low level at all-time points. Thus, these results indicated that susceptibility genotype increased the resistance partly by fine-tuning the DR genes.

The Conserved MicroRNA–Transcription Factor Model Supports the Gene Dosage Balance Hypothesis

To explore the role of miRNA–TF pairs in fungi pathogen-stress response, we next analyzed the posttranscriptional regulation of

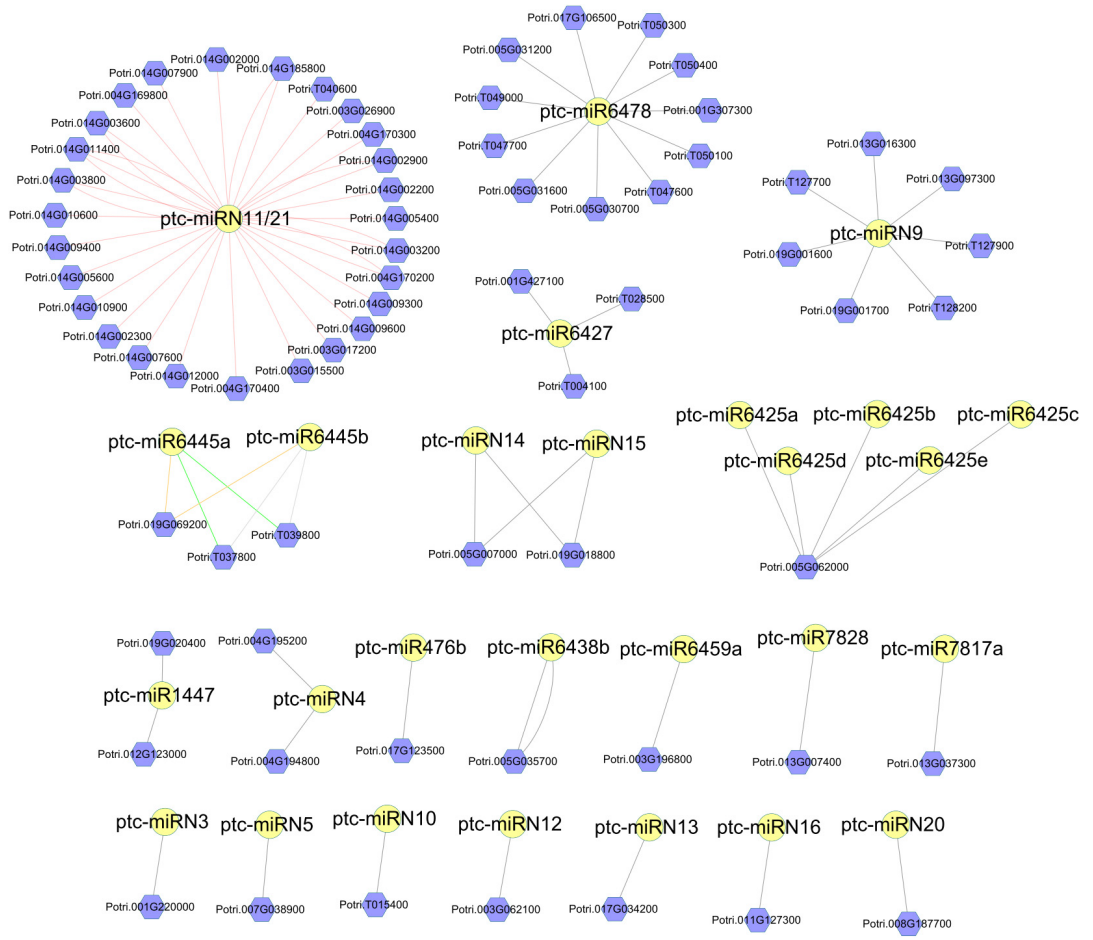


FIGURE 4 | Network of *Populus*-specific miRNA/disease resistance (DR) target gene pairs. The nodes with yellow circles are miRNAs. The nodes with violet hexagon are DR genes. The lines between miRNAs and DR genes represent the targeting relationship.

152 miRNAs. A total of 1,603 miRNA–TF pairs were detected, including 1,384 conserved miRNA–TF pairs and 219 *Populus*-specific miRNA–TF pairs (**Supplementary Data 11**). Here, we observed that miRNAs target numerous TFs that associated with defense response during the *M. brunnea* infection, such as *SBP*, *NAC*, *NF-YA*, *ARF*, and *MYB* (**Figure 3C**). SBP is a well-known TF family in plants, which is involved in various stress response networks (Liu et al., 2019). In total, 152 conserved miRNA–*SBP* regulation pairs were detected; 83 miRNA–NAC regulation pairs (78 conserved and five *Populus*-specific) were detected, for example, *NAC1* and *NAC100* (*NAC1*: Potri.007G065400; *NAC100*: Potri.012G001400) were negatively correlated with miR164a-d; 82 miRNA–ARF regulation pairs were detected; 110 miRNA-MYB regulation pairs were detected. Thus, this implied that miRNAs might play important roles in regulating a wide range of molecular events during the *M. brunnea* infection.

To study the evolutionary effects of polyploidy on a transcriptional network, we reanalyzed the functional genomic and transcriptome data for numerous duplicated gene pairs formed by ancient polyploidy events in poplar (Rodgers-Melnick et al., 2012). A total of 5,931 "salicoid duplications" targeted by miRNAs were detected, including 1,023 WGD-derived TF pairs (**Supplementary Datas 12**, **13**). Notably, of these TF pairs, 459 (∼44.9%) have only one paralog being targeted by a miRNA. This could be due to either gain or loss of the miRNA binding sites of one of the duplicates after WGD. Furthermore, detailed analysis revealed that 83.7% of the TF WGDs were targeted by conserved miRNAs. This could be explained by gene balance hypothesis. Under the hypothesis, we expected that more conserved miRNAs would target genes of central roles in networks such as functional TFs. Overall, miRNAs play important roles in biological regulating network, with conserved miRNAs regulating central biological nodes, supporting the gene balance hypothesis.

MiR164–NAC–mRNA Regulatory Network in Response to Biotic Stress in *Populus*

To further study the complexity of the transcription regulatory network in response to biotic stress, we carried out motif occurrence analysis of five groups of disease-resistant related genes, including signaling cascades, TFs, reactive oxygen,

pathogen-related, and NBS. As a result, 52,164 TF binding sites were enriched in the promoters of detected genes with a frequency that exceeds 85% (**Supplementary Data 14**). Many DR-related genes were predicted to be the upstream TFs in our data set, such as NAC, MYB, WRKY, ERF, and bZIP. In particular, plant NAC domain protein may serve as a convergent node in developmental processes and stress response. For instance, NAC was found to increase necrotrophic/biotrophic pathogen tolerance, which could be induced by wounding and defense-related hormones (Galle et al., 2013). Also, overexpression of NAC4 (ANAC079/080) in *Arabidopsis* could increase the pathogen stress tolerance (Lee et al., 2017).

To further study the regulatory network of miRNAs involved in DR, we next performed hierarchical clustering analysis on TFs (targets of conserved miRNAs) and DR genes. We thus constructed a three-layer network uncovering the module of miR164–*NAC*–mRNA, with an important role in the fungal pathogen infection (**Figure 5A**). This module includes three *NAC* genes (*NAC1*: Potri.007G065400; *NAC100*: Potri.012G001400; *NAC1*: Potri.005G098200) whose expressions were negatively correlated with miRNA164a (**Figures 5A,B**). This was also supported by the real-time quantitative PCR analyses ($P < 0.05$; **Supplementary Figure 2**). Sequence conservation analyses showed that the mature regions of miR164a were completely conserved in *Arabidopsis*, rice, maize, *Medicago*, *Brassica*, *Sorghum*, *Vitis*, *Brachypodium*, and *Glycine*, and the precursor sequences of miR164a show an extensive similarity (41.76%) in eight plants (**Figure 5C**). Moreover, analysis of the genomic and protein sequences of the three *NAC* genes showed that all of them were composed of three exons and two introns and evolutionarily conserved (**Figures 5D–H**). In total, 134 genes were predicted to be the downstream targets of the three NAC (**Supplementary Data 14**). Functional enrichment analysis showed that these genes were mainly involved in biological processes such as apoptosis and innate immune response ($P < 1e\text{-}56$). Taken together, a multilayered hierarchical gene regulation network provides opportunities to investigate transcriptome dynamics and identifies key genes involved in specific pathways.

DISCUSSION

Pseudogenes and Posttranscriptional Regulation

The origins of miRNA genes have attracted wide attention in recent years. In plants, there are at least four hypotheses, for instance, according to sequence homology between *MIR* genes and target genes, Allen et al. (2004) proposed the inverted duplication hypothesis. Under the hypothesis, these young miRNA genes were supposedly generated from inverted duplication events of one of their target genes by forming two adjacent gene segments in either convergent or divergent orientation. Genome-wide analysis of miRNA genes in *A. thaliana* further revealed that some genomic repeats (including WGDs, tandem duplications, and segmental duplications) and following dispersal and diversification were also an essential pathway for the origin of miRNAs (Smalheiser and Torvik, 2005). Moreover, another potential source of miRNAs is random sequences and spontaneously formed from foldback sequences (Fenselau et al., 2008). As a large proportion of miRNA genes were laying within TEs or pseudogenes, the hypothesis of miRNA originating from TEs or pseudogenes has been proposed by researchers recently (Piriyapongsa and Jordan, 2008; Sasidharan and Gerstein, 2008).

Despite previously being referred to as junk DNA (Zhang et al., 2003), pseudogenes are now known to be essential elements of most eukaryotic genomes, making important contributions to their structure, diversity, capacity, and adaptation (Balasubramanian et al., 2009; Poliseno et al., 2015). The widely distributed pseudogenes are a rapidly evolving part of the genome because they have the potential for incorporating new functions into DNA sequences by mutant alleles (Balakirev and Ayala, 2003; Zhang et al., 2003). Here, when exploring the distribution of poplar miRNAs in different parts of genome regions, including the 5′ UTRs, CDS, 3′ UTRs, introns, exons, promoters, transposons, pseudogenes, and intergenic regions, we determined that pseudogenes contributed a certain proportion (14.47%) of the miRNAs. Owing to their origin as gene copies, pseudogenes typically exhibit a high sequence homology to their parent gene. Consequently, it is possible that some pseudogene-derived miRNA may be implicated in repressing transcription of their parental gene. This strong association of miRNAs with pseudogenes provides an important mechanism for the origin and posttranscriptional regulation of miRNAs (Guo et al., 2009; Xie et al., 2019).

MicroRNA Mediated Defense Against Pathogen Stress

The plant NB-LRR genes mediate effector-triggered immunity by acting as key receptors during the innate immunity response against a wide range of pests and diseases. The NB-LRR genes are generally grouped into two subclasses: the toll/interleukin-1 receptor-like group (TIR-NB-LRRs) and a coiled-coil domain-containing group (CC-NB-LRRs) (Jones and Dangl, 2006). Both classes can be triggered by miRNAs to generate phasiRNAs, which can reduce the levels of the transcripts of their targets in *cis* and trans (Zhai et al., 2011). In contrast to low-copy genes, many NB-LRR genes have undergone dramatic duplications and losses, domain architecture variations, the partitioning of subfamilies, and copy number variation among species (Karasov et al., 2014). Thus, the NB-LRR genes are highly variable, lineage-specific, and associated with the plant immune response, providing material to allow rapid adaptative evolution.

The target sites of conserved miRNAs are often located within the highly conserved domains of the target genes (Rhoades et al., 2002). Unlike conserved miRNAs, newly evolved miRNAs tend not to target these conserved functional domains and may instead target mRNAs simply by chance (Chen and Rajewsky, 2007). Indeed, the target genes of the newly emerged miRNAs in *Populus* were found to have various functions, including numerous DR *NB-LRR* genes and few TFs. Our analysis demonstrated that more *Populus*-specific miRNAs target the NB-LRR genes than

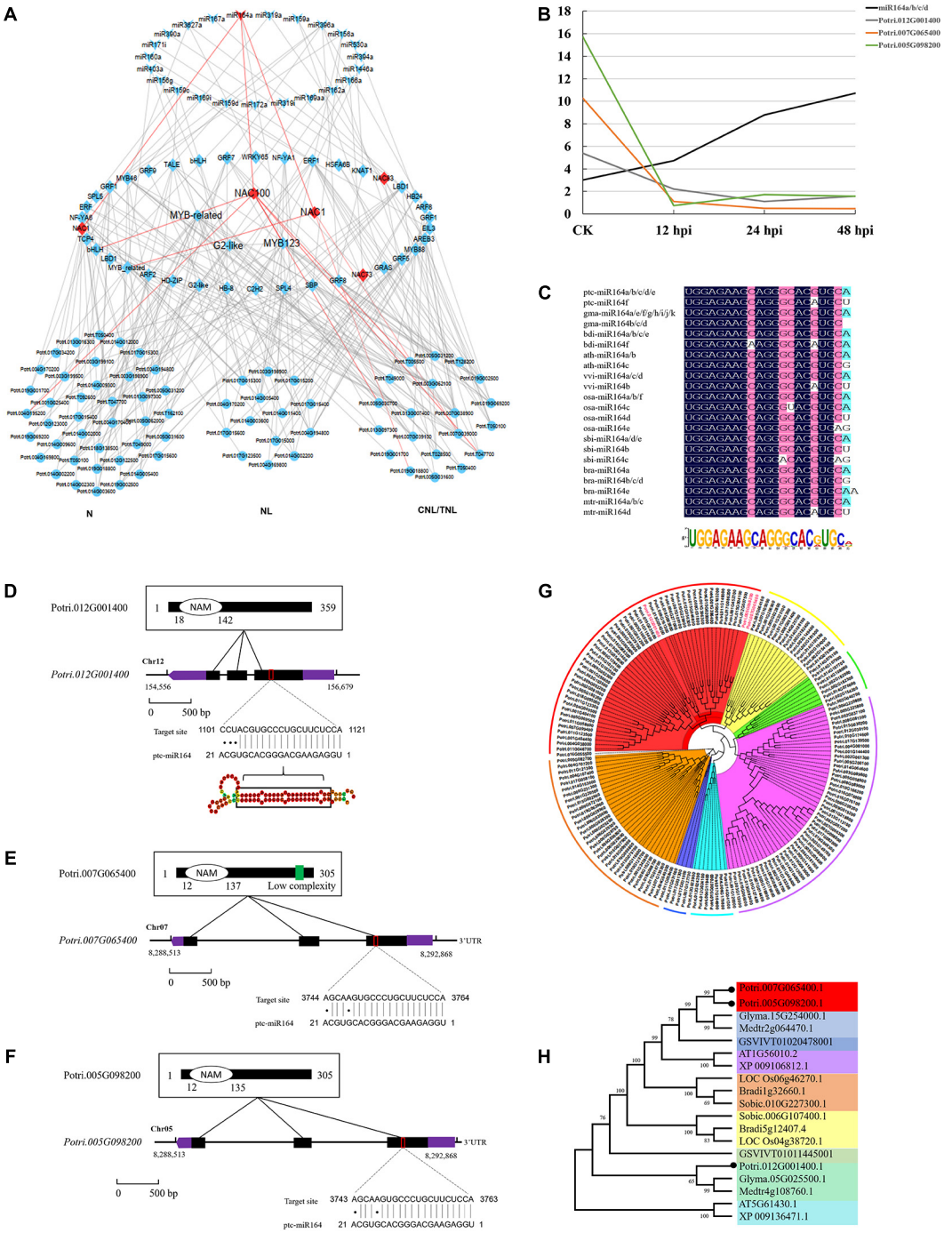


FIGURE 5 | Conserved miR164–NAC–mRNA regulatory network in response to fungi pathogen stress in *Populus*. **(A)** The three-layered gene regulatory network (GRN) was constructed with the backward elimination random forest (BWERF) algorithm. The nodes with red color highlighted the key regulatory transcription factors (TFs). **(B)** The expression value of miR164 and *NAC1/100* genes. **(C)** Sequence logo view of the mature miR164 sequence. **(D–F)** Conserved domains of ptc-NAC1/100 protein sequence, gene structure of *ptc-NAC1/100*, and predicted base-pairing interaction between ptc-miR164 and *ptc-NAC1/100*. Exons are shown as *black boxes* and introns as *lines*. The 5′ UTR and 3′ UTR are shown as *purple boxes*. **(G,H)** Phylogenetic analysis of *NAC* targets of miR164 in *Populus*. Phylogenetic analysis of *ptc-NAC1/NAC100* homologous genes in eight other plant species.

conserved miRNAs. Considering that this *de novo* diversity may be associated with plant defense, the *Populus*-specific miRNAs have a greater potential to target newly evolved plant DR genes, further contributing to the phenotypic innovation of the host. Once the newly emerged miRNAs become fixed in the regulatory modules, they may gradually evolve to target more genes

linked to their specific function (Chen and Rajewsky, 2007; Xie et al., 2017).

Populus MicroRNA/Target Patterns Support the Gene Dosage Balance Hypothesis

Gene duplication is one of the primary driving forces in the evolution of genomes and genetic systems (Moore and Purugganan, 2003). Duplicated genes were classified into five types, including WGD, proximal duplication, tandem duplication, transposed duplication, and dispersed duplication. As an extreme gene replication mechanism, WGD results in a sudden increase in the size of the genome and entire gene set. In contrast to small-scale duplicates, duplicates created by WGD (also called homologs) tend to be retained at much higher fractions (Rodgers-Melnick et al., 2012). Also, gene duplicability or the ability of genes to be retained following duplication is often biased. As we all know, three WGD events occurred during *Populus* evolution: an ancient duplication event, a middle event shared among the Eurosids, and a recent event shared among the Salicaceae (Tuskan et al., 2006). The modern poplar genome began to diverge around 6 million years after the "Salicoid" duplication, and retained WGD genes are biased toward more central roles in networks, such as members of signal transduction cascades and TFs (Freeling, 2009; Rodgers-Melnick et al., 2012). Therefore, genes retained as duplicate pairs following WGDs are disproportionately likely to encode TFs and components of multi-protein complexes, with a potential explanation for this phenomenon given by the gene balance hypothesis (Birchler and Veitia, 2007, 2012; Edger and Pires, 2009; Liang and Schnable, 2018).

The role of miRNAs was potentially important in terms of modulating the expression of TFs because miRNAs can operate in a dosage-sensitive manner (Guo et al., 2010). Besides, the target sites of conserved miRNAs are often located within the highly conserved domains of the target genes (Rhoades et al., 2002). Following WGDs, many of these duplicated TFs evolved separate functions in divergent ways, such as non-functionalization (Ohno, 1971), subfunctionalization (Lynch and Force, 2000) or neofunctionalization, to adapt growth/development and stress response. In this case, only one of the duplicates is targeted by miRNA, indicating a gain or loss of miRNA target site after the WGD event. Also, the evolution of miRNA binding sites suggests a coevolution between miRNAs and their targets tending to preserve core duplicates in adapting to the change of environment. Together, our study provides insights into the regulation of miRNAs and target functional evolution in the defense process.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Genome Sequence Archive in the Beijing Institute of Genomics BIG Data Center, Chinese Academy of Sciences (https://bigd.big.ac.cn/gsa) under the accession number CRA003506 (Small RNA sequencing data) and CRA001647 (RNA sequencing data).

AUTHOR CONTRIBUTIONS

JX designed the research. SC performed the research, analyzed the data, and wrote the manuscript. JW, YfZ, YyZ, WX, and JX revised the manuscript. YL provided valuable suggestions to the manuscript. JX obtained funding and is responsible for this article. All authors read and approved the manuscript.

FUNDING

This work was supported by the Project of the National Natural Science Foundation of China (Nos. 32022057 and 31972954), Young Elite Scientists Sponsorship Program by CAST (No. 2018QNRC001), and Forestry and Grassland Science and Technology Innovation Youth Top Talent Project of China (No. 2020132607).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.668940/full#supplementary-material

Supplementary Figure 1 | Hairpin structure of 21 novel miRNAs.

Supplementary Figure 2 | Validation of identified expression of miRNAs and target genes.

Supplementary Data 1 | Primer sequences used for RT-qPCR.

Supplementary Data 2 | Mapping results statistics.

Supplementary Data 3 | Known and novel miRNAs identified in this study.

Supplementary Data 4 | The expression value of each miRNA (TPM) and target genes (FPKM).

Supplementary Data 5 | Conserved and Populus-specific miRNA identified in this study.

Supplementary Data 6 | miRNA list in each cluster by K-means analysis.

Supplementary Data 7 | The position of miRNAs in Populus genome.

Supplementary Data 8 | The predicted targets of miRNAs.

Supplementary Data 9 | Correlation of miRNA/target pairs ($r < -9 \times 10^{-1}$, $P < 5 \times 10^{-2}$).

Supplementary Data 10 | The miRNA/target pairs having degradome evidence.

Supplementary Data 11 | miRNA–TF pairs.

Supplementary Data 12 | WGD pairs in TF targets.

Supplementary Data 13 | WGD pairs in all target genes.

Supplementary Data 14 | Transcription factor binding sites of disease-resistance related genes.

REFERENCES

Allen, E., Xie, Z., Gustafson, A. M., Sung, G. H., Spatafora, J. W., and Carrington, J. C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* 36, 1282–1290. doi: 10.1038/ng1478

Balakirev, E. S., and Ayala, F. J. (2003). Pseudogenes: are they "Junk" or functional DNA? *Annu. Rev. Genet.* 37, 123–151. doi: 10.1146/annurev.genet.37.040103.103949

Balasubramanian, S., Zheng, D., Liu, Y. J., and Gang, F. (2009). Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. *Genome Bio.* 10:R2. doi: 10.1186/gb-2009-10-1-r2

Birchler, J. A., and Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395–402. doi: 10.1105/tpc.106.049338

Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109

Carrington, J. C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science* 301, 336–338. doi: 10.1126/science.1085242

Chen, K., and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8:93. doi: 10.1038/nrg1990

Chen, S., Zhang, Y., Zhao, Y., Xu, W., and Zhang, D. (2020). Key genes and genetic interactions of plant-pathogen functional modules in poplar infected by *Marssonina brunnea*. *Mol. Plant Microbe Interact.* 33, 1080–1090. doi: 10.1094/MPMI-11-19-0325-R

Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39:W155. doi: 10.1093/nar/gkr319

Du, Q., Tian, J., Yang, X., Pan, W., Xu, B., Li, B., et al. (2014). Identification of additive, dominant, and epistatic variation conferred by key genes in cellulose biosynthesis pathway in *Populus tomentosa*. *DNA Res.* 22, 53–67. doi: 10.1093/dnares/dsu040

Edger, P. P., and Pires, J. C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* 17, 699–717. doi: 10.1007/s10577-009-9055-9

Fenselau, d. F. F, Schneeberger, K., Dezulian, T., Huson, D. H., and Weigel, D. (2008). Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *Rna* 14, 2455–2459. doi: 10.1261/rna.1149408

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Galle, L. H., Camille, P., Barbara, C., Fanja, R., Clémentine, G., Christophe, C., et al. (2013). Grapevine NAC1 transcription factor as a convergent node in developmental processes, abiotic stresses, and necrotrophic/biotrophic pathogen tolerance. *J. Exp. Bot.* 64, 4877–4893. doi: 10.1093/jxb/ert277

Ghosh, S., and Chan, C. K. K. (2016). Analysis of RNA-Seq data using TopHat and Cufflinks. *Plant Bioinform.* 1374, 339–361. doi: 10.1007/978-1-4939-3167-5_18

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartei, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840. doi: 10.1038/nature09267

Guo, X., Zhang, Z., Gerstein, M. B., Zheng, D., and Zhang, W. (2009). Small RNAs originated from pseudogenes: cis- or trans-acting? *PLoS Comput. Biol.* 5:e1000449. doi: 10.1371/journal.pcbi.1000449

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53. doi: 10.1038/nrg1379

Karasov, T. L., Kniskern, J. M., Gao, L., Deyoung, B. J., Ding, J., Dubiella, U., et al. (2014). The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512, 436–440. doi: 10.1038/nature13439

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., and Kelley, R. (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

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Lee, M.-H., Jeon, H. S., Kim, H. G., and Park, O. K. (2017). An *Arabidopsis* NAC transcription factor NAC4 promotes pathogen-induced cell death under negative regulation by microRNA164. *New Phytol.* 214, 343–360. doi: 10.1111/nph.14371

Li, D., Wang, F., Wang, C., Zou, L., Wang, Z., Chen, Q., et al. (2016). MicroRNA-mediated susceptible poplar gene expression regulation associated with the infection of virulent *Melampsora larici-populina*. *BMC Genomics* 17:59. doi: 10.1186/s12864-015-2286-6

Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., et al. (2012). MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1790–1795. doi: 10.1073/pnas.1118282109

Liang, Z., and Schnable, J. C. (2018). Functional divergence between subgenomes and gene pairs after whole genome duplications. *Mol. Plant* 11, 388–397. doi: 10.1016/j.molp.2017.12.010

Liu, M., Shi, Z., Zhang, X., Wang, M., Zhang, L., Zheng, K., et al. (2019). Inducible overexpression of ideal plant architecture1 improves both yield and disease resistance in rice. *Nat. Plants* 5, 389–400. doi: 10.1038/s41477-019-0383-2

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473. doi: 10.1023/A:1004191625603

Ma, J.-X., Huang, X., Li, W.-H., and Zhu, C.-G. (2013). Sap flow and trunk maximum daily shrinkage (MDS) measurements for diagnosing water status of *Populus euphratica* in an inland river basin of Northwest China. *Ecohydrology* 6, 994–1000. doi: 10.1002/eco.1439

Meyers, B. C., and Axtell, M. J. (2019). MicroRNAs in plants: key findings from the early years. *Plant Cell* 31, 1206–1207. doi: 10.1105/tpc.19.00310

Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., and Zhu, J. K. (2008). Criteria for annotation of plant MicroRNAs. *Plant Cell Online* 20, 3186–3190. doi: 10.1105/tpc.108.064311

Moore, R. C., and Purugganan, M. D. (2003). The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15682–15687. doi: 10.1073/pnas.2535513100

Ohno, S. (1971). Evolution by gene duplication. *Am. J. Hum. Genet.* 23:541. doi: 10.2307/1530208

Peláez, P., and Sanchez, F. (2013). Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Front. Plant Sci.* 4:343. doi: 10.3389/fpls.2013.00343

Piriyapongsa, J., and Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14, 814–821. doi: 10.1111/j.1752-1688.2000.tb04278.x

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2015). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.7554/eLife.08245

Pumplin, N., and Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* 11, 745–760. doi: 10.1038/nrmicro3120

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

Rodgers-Melnick, E., Mane, S. P., Dharmawardhana, P., Slavov, G. T., Crasta, O. R., Strauss, S. H., et al. (2012). Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Res.* 22, 95–105. doi: 10.1101/gr.125146.111

Sasidharan, R., and Gerstein, M. (2008). Genomics: protein fossils live on as RNA. *Nature* 453, 729–731. doi: 10.1038/453729a

Seo, E., Choi, D., and Choi. (2015). Functional studies of transcription factors involved in plant defenses in the genomics era. *Brief. Funct. Genomics* 14, 260–267. doi: 10.1093/bfgp/elv011

Shannon, P. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Smalheiser, N. R., and Torvik, V. I. (2005). Mammalian microRNAs derived from genomic repeats. *Trends Genet.* 21, 322–326. doi: 10.1016/j.tig.2005.04.008

Sunkar, R., Li, Y. F., and Jagadeeswaran, G. (2012). Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17, 196–203. doi: 10.1016/j.tplants.2012.01.010

Thiebaut, F., Grativol, C., Hemerly, A. S., and Ferreira, P. C. G. (2015). MicroRNA networks in plant-microorganism interactions. *Trop. Plant Biol.* 8, 40–50. doi: 10.1007/s12042-015-9149-9

Tian, T., Yue, L., Hengyu, Y., Qi, Y., Xin, Y., Zhou, D., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046

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., Sun, S., Jin, J., Fu, D., and Zhang, Q. (2015). Coordinated regulation of vegetative and reproductive branching in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15504–15509. doi: 10.1073/pnas.1521949112

Wang, W. Q., Wang, J., Wu, Y. Y., Li, D. W., Allan, A. C., and Yin, X. R. (2020). Genome−wide analysis of coding and non−coding RNA reveals a conserved miR164−NAC regulatory pathway for fruit ripening. *New Phytol.* 225, 1618–1634. doi: 10.1111/nph.16233

Xie, J., Li, Y., Liu, X., Zhao, Y., Li, B., Ingvarsson, P. K., et al. (2019). Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. *Plant Cell* 31, 563–578. doi: 10.1105/tpc.18.00601

Xie, J., Yang, X., Song, Y., Du, Q., Li, Y., Chen, J., et al. (2017). Adaptive evolution and functional innovation of *Populus*-specific recently evolved microRNAs. *New Phytol.* 213, 206–219. doi: 10.1111/nph.14046

Yang, L., and Huang, H. (2015). Roles of small RNAs in plant disease resistance. *J. Integr. Plant Biol.* 56, 962–970. doi: 10.1111/jipb.12200

Zhai, J. X., Jeong, D. H., De Paoli, E., Park, S., Rosen, B. D., Li, Y., et al. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* 25, 2540–2553. doi: 10.1101/gad.177527.111

Zhang, Y., He, W., and Yan, D.-H. (2018). Histopathologic characterization of the process of *Marssonina brunnea* infection in poplar leaves. *Can. J. Forest Res.* 48, 1302–1310. doi: 10.1139/cjfr-2018-0152

Zhang, Z., Harrison, P. M., Liu, Y., and Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541–2558. doi: 10.1101/gr.1429003

Zhang, Z., Yu, J., Li, D., Zhang, Z., Liu, F., Zhou, X., et al. (2010). PMRD: plant microRNA database. *Nucleic Acids Res.* 38, D806. doi: 10.1093/nar/gkp818

Zhao, J. P., Jiang, X. L., Zhang, B. Y., and Su, X. H. (2012). Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. *PLoS One* 7:e44968. doi: 10.1371/journal.pone.0044968

Zhao, M., Cai, C., Zhai, J., Lin, F., Li, L., Shreve, J., et al. (2015). Coordination of MicroRNAs, PhasiRNAs, and NB-LRR genes in response to a plant pathogen: insights from analyses of a set of soybean Rps gene near-isogenic lines. *Plant Genome* 8, 1–13. doi: 10.3835/plantgenome2014.09.0044

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.

Copyright © 2021 Chen, Wu, Zhang, Zhao, Xu, Li and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 June 2021
doi: 10.3389/fgene.2021.678673

Organ-Specific Transcriptome Analysis Identifies Candidate Genes Involved in the Stem Specialization of Bermudagrass (*Cynodon dactylon* L.)

*Si Chen[1], Xin Xu[1], Ziyan Ma[1], Jianxiu Liu[2] and Bing Zhang[1,3]**

[1] College of Animal Science and Technology, Yangzhou University, Yangzhou, China, [2] Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, [3] Joint International Research Laboratory of Agriculture and Agri-Product Safety, Ministry of Education of China, Yangzhou University, Yangzhou, China

As an important warm-season turfgrass and forage grass species with wide applications, bermudagrass (*Cynodon dactylon* L.) simultaneously has shoot, stolon and rhizome, three types of stems with different physiological functions. To better understand how the three types of stems differentiate and specialize, we generated an organ-specific transcriptome dataset of bermudagrass encompassing 114,169 unigenes, among which 100,878 and 65,901 could be assigned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Gene Ontology (GO) terms, respectively. Using the dataset, we comprehensively analyzed the gene expression of different organs, especially the shoot, stolon and rhizome. The results indicated that six organs of bermudagrass all contained more than 52,000 significantly expressed unigenes, however, only 3,028 unigenes were enrich-expressed in different organs. Paired comparison analyses further indicated that 11,762 unigenes were differentially expressed in the three types of stems. Gene enrichment analysis revealed that 39 KEGG pathways were enriched with the differentially expressed unigenes (DEGs). Specifically, 401 DEGs were involved in plant hormone signal transduction, whereas 1,978 DEGs were transcription factors involved in gene expression regulation. Furthermore, in agreement with the starch content and starch synthase assay results, DEGs encoding starch synthesis-related enzymes all showed the highest expression level in the rhizome. These results not only provided new insights into the specialization of stems in bermudagrass but also made solid foundation for future gene functional studies in this important grass species and other stoloniferous/rhizomatous plants.

Keywords: *Cynodon dactylon*, organogenesis, transcriptomics, stolon, rhizome, starch

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Cheng Qin,
Zunyi Vocational and Technical College, China
Youlu Yuan,
Institute of Cotton Research (CAAS), China

***Correspondence:**
Bing Zhang
bingzhang@yzu.edu.cn

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

Received: *10 March 2021*
Accepted: *01 June 2021*
Published: *23 June 2021*

Citation:
Chen S, Xu X, Ma Z, Liu J and Zhang B (2021) Organ-Specific Transcriptome Analysis Identifies Candidate Genes Involved in the Stem Specialization of Bermudagrass (Cynodon dactylon L.). Front. Genet. 12:678673. doi: 10.3389/fgene.2021.678673

INTRODUCTION

Bermudagrass (*Cynodon dactylon* L.) is an important warm-season turfgrass and forage grass species with wide applications. In warm regions around the world, different bermudagrass cultivars and their hybrid progenies with other *Cynodon* species are frequently used to generate high-quality turfs for multiple purposes (Reasor et al., 2016). With fast growth rate and special nutrition value,

bermudagrass cultivars of forage type are highly valuable for cattle and dairy industry (Hill et al., 2001). In some countries, bermudagrass is also used as a traditional medical plant to cure several diseases, including anasarca, diarrhea and hemorrhage (Nagori and Solanki, 2011). On the other hand, bermudagrass is also a notorious weed in farmlands of warm regions, leading to a yield reduction of major crop plants (Jiménez-Brenes et al., 2019).

Unlike many well-known cereal grass species including rice, wheat, maize, and sorghum, bermudagrass has unique plant architectural characteristics that its stems are differentiated into shoot, stolon and rhizome (Dong and de Kroon, 1994). Among the three types of stems, shoot is an erect growing stem and widely seen in other plants, whereas stolon and rhizome are two kinds of specialized stems that grow horizontally aboveground and underground, respectively (Guo et al., 2020). With stolon and rhizome, bermudagrass can fast propagate by asexually clonal growth, which is exactly the reason why bermudagrass is both a useful turfgrass and a harmful weed (Horowitz, 1972; Fernandez, 2003). In the past several years, a few studies have successfully analyzed the functions of shoot, stolon and rhizome of bermudagrass and pointed out the difference among the three types of stems (Pornaro et al., 2019; Zhang et al., 2019). However, the molecular mechanism underlying the differentiation and development of the three stems in bermudagrass remains unclear.

Quantifying the differential expression of genes in various plant organs and tissues is vital to understand organogenesis and histogenesis of plants (Klepikova and Penin, 2019). In recent years, high-throughput comparative transcriptome analyses were successfully employed to identify candidate genes related to the differentiation and development of different organs and tissues in many plant species according to their temporal and spatial expression profiles. For example, a gene expression atlas encompassing 79 tissue samples was constructed to explore root development in maize (Stelpflug et al., 2016), whereas RNA sequencing of maize embryo, endosperm and seed resulted in the identification of novel transcripts possibly involved in embryo and seed development (Chen et al., 2014; Yi et al., 2019). Comparative transcriptome analysis of achenes and receptacles at four stages of fruit ripening revealed the possible important role of ethylene in receptacle ripening of strawberry (Sánchez-Sevilla et al., 2017). Transcriptome analysis of five organs identified putative 1,077 genes involved in rhizome development of *Miscanthus lutarioriparius* (Hu et al., 2017), whereas transcriptome analysis of the *Zingiber zerumbet* flower at two stages revealed 2,075 transcription factors (TFs) possibly involved in the flower development (Zhao et al., 2020). Transcriptome analyses were also conducted in bermudagrass to characterize the cold-resistance, salt-tolerance and other stress-responsive mechanisms (Chen L. et al., 2015; Hu et al., 2015; Melmaiee et al., 2015; Shi et al., 2015; Zhu et al., 2015; Fan et al., 2019). However, similar transcriptomics studies about the growth and development of bermudagrass are deficient.

In this study, we integrated the PacBio full-length transcriptome data and organ-specific Illumina sequencing results to characterize gene expression in the six organs of bermudagrass cultivar Yangjiang, especially the three types of stems: shoot, stolon and rhizome. The results not only provided new insights into the stem specialization of bermudagrass at the transcriptome level but also established a high-confidence database for future gene functional studies in this important grass species.

MATERIALS AND METHODS

Plant Material

Bermudagrass (*C. dactylon*) cultivar Yangjiang was used in this study. The bermudagrass turf were grown in turfgrass plots of Yangzhou University (32°35′N, 119°40′E; 5 m a.s.l.; average annual temperatures: 22.4°C; average annual precipitation: 1,106 mm; annual average sunshine hours: 1,960 h; soil type: 80% river sand mixed with 20% peat soil) under routine management conditions (irrigation: as required to keep the soil moist; fertilization: four times/year; mowing: two times/month) for 2 years before the experiments were conducted. Approximately 15 g of leaves, shoots, stolons, rhizomes, roots and inflorescences were randomly collected from different plants (0.1 g per plant; 5 g sample collected from 50 plants as a replicate; three replicates) of the bermudagrass turf at flowering stages. The collected organ samples were frozen in liquid nitrogen and then stored at −80°C for RNA extraction, starch and soluble sugar content measurement, and enzyme activity assay.

RNA Extraction, cDNA Library Construction, and Illumina Sequencing

Total RNA was extracted from 0.5 g of frozen organ samples using RNAprep pure Plant kit (Tiangen, Beijing, China). RNA integrity and concentration were determined by gel electrophoresis and Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, United States), respectively. Only samples of high-quality RNA (RNA integrity number ≥ 7) were used for following library preparation. cDNA libraries were prepared using Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, United States) according to the manufacturer's instructions. Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, United States) was used to check the quality of the libraries. The qualified cDNA libraries were paired-end sequenced using Illumina HiSeq^{TM} 3000 (Illumina). For each organs, three sample replicates were sequenced to represent three biological replicates.

Correction of the PacBio Sequencing Reads Using Illumina Sequencing Data

Adapter sequences, short reads (length <50 bp) and low quality reads (Q-value $\leq$ 30) were removed from the raw Illumina sequencing reads using Cutadapt software with default parameters (Martin, 2011). The obtained high-quality clean reads was used by LoRDEC software with default parameters (Salmela and Rivals, 2014) to construct de Bruijn graph and compared with the long consensus reads of the previously reported PacBio transcriptome of bermudagrass (Zhang et al., 2018). Corrected long consensus reads were further compared and clustered to

non-redundant gene sequences, the unigenes, using the CD-HIT software with default parameters (-c 0.99 -aL 0.90 -AL 100 -aS 0.99 -AS 30) (Fu et al., 2012).

Functional Annotation

The unigene sequences and their translated amino acid sequences were BLAST searched against the NCBI nucleotide (NT) database[1], the NCBI non-redundant protein (NR) database (see text footnote 1), the Eukaryotic Ortholog Groups (KOG) database (see text footnote 1), the Gene Ontology (GO) database[2], and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[3] with an *E*-value threshold $\leq 10^{-5}$ to obtain the annotation information.

Gene Expression Analysis

The Illumina sequenced clean reads were mapped to the unigene sequences using Bowtie2 software of the RSEM package with default parameters (Li and Dewey, 2011). The numbers of mapped reads were converted to FPKM (fragments per kilobase of transcript per million mapped fragments) values. The $\log_2$ transformed FPKM values were applied to perform Hierarchical clustering using Pearson's correlation distance in the Pvclust software package with default settings (Suzuki and Shimodaira, 2006). The significantly expressed unigenes were defined as FPKM value ≥ 1 (Nautiyal et al., 2020). The stably expressed unigenes were defined as the FPKM ratio (minimal versus maximal FPKM value in different organ samples) > 0.8 and the coefficient of variation (C.V., standard deviation divided by the average FPKM value) < 0.3 (Klepikova et al., 2016). The organ-enhanced unigenes were defined as FPKM value is five-fold above the average FPKM values of other organs, whereas organ-enriched unigenes were defined as FPKM value is five-fold above the FPKM values of any other organs (Uhlén et al., 2016). The DEGs were determined through comparison of FPKM values between two organs (three biological replicates) using DESeq2 software with the criterion of fold-change > 2.0 and p-value < 0.05 (Love et al., 2014).

Pathway Enrichment Analysis and Transcription Factor Identification

The KEGG pathway enrichment analysis of DEGs was performed using KEGG Orthology Based Annotation System (KOBAS) web server with the significance cutoff of p-value < 0.05 (Xie et al., 2011). The TF families were identified through BLASTx search against known plant TFs recorded in PlantTFDB database[4] with an *E*-value threshold $\leq 10^{-5}$.

Quantitative PCR

cDNA was synthesized using the PrimeScript RT reagent kit (Takara, Dalian, China). RT-qPCR reactions were performed on a Mini Opticon Real-Time PCR System (Bio-Rad, Hercules, United States) using the SYBR Premix ExTaq (TaKaRa). PCR primer pairs were designed using the PrimerSelect software of the DNASTAR package (v7.1.0) and manually checked to ensure a high-efficient and accurate amplification (**Supplementary Table 1**). Relative gene expression level was quantified using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Soluble Sugar and Starch Content Determination

Soluble sugar and starch content were determined as previously described (Zhang et al., 2019). Briefly, frozen organ samples were baking dried to remove water and 0.1 g dry organ samples were ground to a fine powder using mortar and pestle. After washing with 100% acetone to remove interfering pigments, the powder was dissolved in 5 ml of 80% ethanol, incubated at 80°C in a water bath for 30 min and centrifuged at 8,000 × *g* for 10 min. For the soluble sugar content assay, the supernatants were mixed with a five-fold volume of 1% (m/v) anthrone dissolved in H_2SO_4. The mixture was held in a 100°C water bath for 10 min. The absorbance at 625 nm was determined using an Ultrospec 3300 Pro spectrophotometer (Amersham Biosciences, Uppsala, Sweden). The sugar content was calculated using the standard curve method. For the starch content assay, 3 ml of water was first added to redissolve the centrifuged pellet in a 100°C water bath for 10 min and then 2 ml of 1.1% (v/v) HCl was added to promote the degradation of starch to soluble sugar. After centrifugation at 8,000 × *g* for 10 min, the same procedures were performed to determine the sugar content of the supernatants, which represents the starch content.

Starch Synthase Activity Assay

The starch synthase activity was assayed according to the previously described method with minor modifications (Zhou et al., 2018). Briefly, 0.5 g frozen organ samples were homogenized using mortar and pestle with 5 mL ice-cold HEPES-NaOH buffer (pH 7.5) containing 0.1% (w/v) PMSF. The homogenates were centrifuged at 5,000 × *g* for 15 min, and 100 μL supernatants were immediately transferred to 500 μL reaction solutions containing 50 mM HEPES-NaOH (pH 7.5), 1.6 mM adenosine diphosphate glucose, 15 mM DTT and 1 mg amylopectin. The reaction lasted for 20 min at 30°C and was stopped by boiling in a water bath for 2 min. After cooling, the produced ADP was converted to ATP by adding 100 μL of 40 mM phosphoenolpyruvate (PEP), 50 μL of 100 mM $MgCl_2$ and 1 U pyruvate kinase (Sigma, Shanghai, China), followed by another 20 min incubation at 30°C. The produced ATP was determined using the ATP Assay System Bioluminescence Detection Kit (Promega, Madison, WI, United States) with a GloMax-Multi luminescence reader (Promega).

Statistical Analyses

Unless otherwise specified, all the experiments were at least repeated for three biological and technical replicates. Tukey's multiple comparison test was used for variation analyses among

[1] http://www.ncbi.nlm.nih.gov
[2] http://www.geneontology.org
[3] http://www.genome.jp/kegg
[4] http://planttfdb.gao-lab.org/

the different samples. The statistical analyses were performed with SPSS 16.0 statistical software package.

RESULTS

Generation of a Bermudagrass Organ-Specific Transcriptome Dataset

As a typical turf-type cultivar, bermudagrass cultivar Yangjiang forms uniform turfs with a high plant density (**Figure 1A**). Single mature bermudagrass plant is comprised of leaf, root, inflorescence, shoot, stolon, and rhizome (**Figure 1B**). In our previous study, we successfully used PacBio single-molecule long-read sequencing technology to obtain a full-length reference transcriptome of bermudagrass cultivar Yangjiang (Zhang et al., 2018). Here, we resampled the six organs of bermudagrass cultivar Yangjiang and sequenced the total mRNA of each organs using accurate Illumina sequencing technology. The short reads (average read length of 150 bp; average sequencing depth of 25.96) obtained from Illumina sequencing were compared with the long reads of PacBio sequencing to polish and correct the transcripts, which were further clustered together to remove redundant sequences (**Figure 1C**). After these steps, 114,169 unigenes with an average gene length of 2,237 bp were finally obtained. Notably, 82.12% of the unigenes (93,755) have a gene length between 1,000 and 4,000 bp (**Figure 2A**). Among the 114,169 unigenes, 92.37% (105,453) were successfully annotated by at least one database and 89.41% (102,081) could find homolog genes in NR database (**Figure 2B** and **Supplementary Table 2**). Notably, as many as 100,878 and 65,901 unigenes could be assigned to KEGG and GO terms, respectively.

Through mapping the Illumina short reads with the reference transcriptome and normalization through FPKM algorithm, we obtained the overall status of gene expression in the six organs (**Supplementary Table 3**). The six organs all had approximately 94,000 unigenes expressed with detectable reads (**Figure 2C**), however, the number of unigenes significantly expressed (FPKM $\geq$ 1) in at least one organ was only 71,653 (**Supplementary Table 4**). Notably, shoot and inflorescence both had more than 58,000 significantly expressed unigenes, whereas

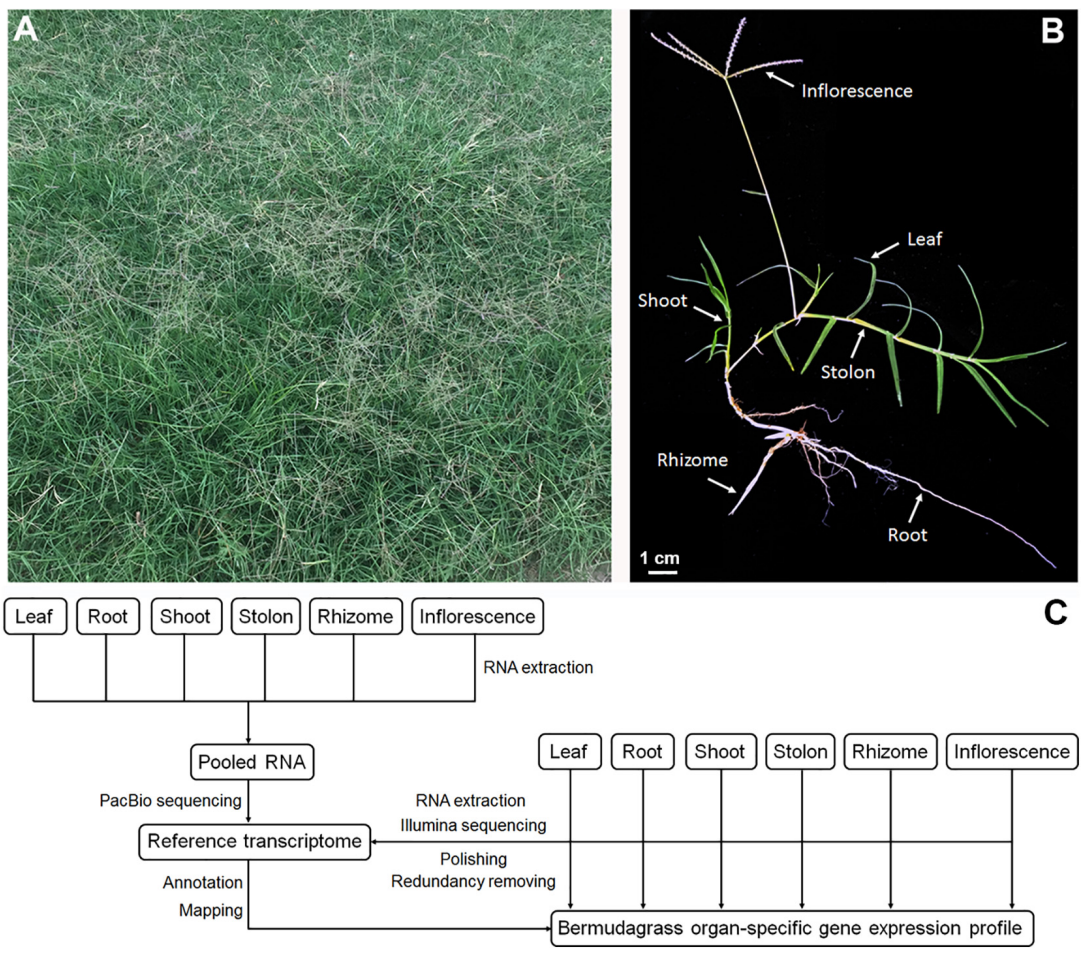


FIGURE 1 | Experimental design for the bermudagrass organ-specific transcriptome dataset. **(A)** Photograph of the bermudagrass cultivar Yangjiang turf at flowering stage. **(B)** Photograph of the bermudagrass cultivar Yangjiang plant with six different organs. **(C)** Pipeline for the Illumina sequencing, reference transcriptome assembly and annotation, and gene expression profiling.

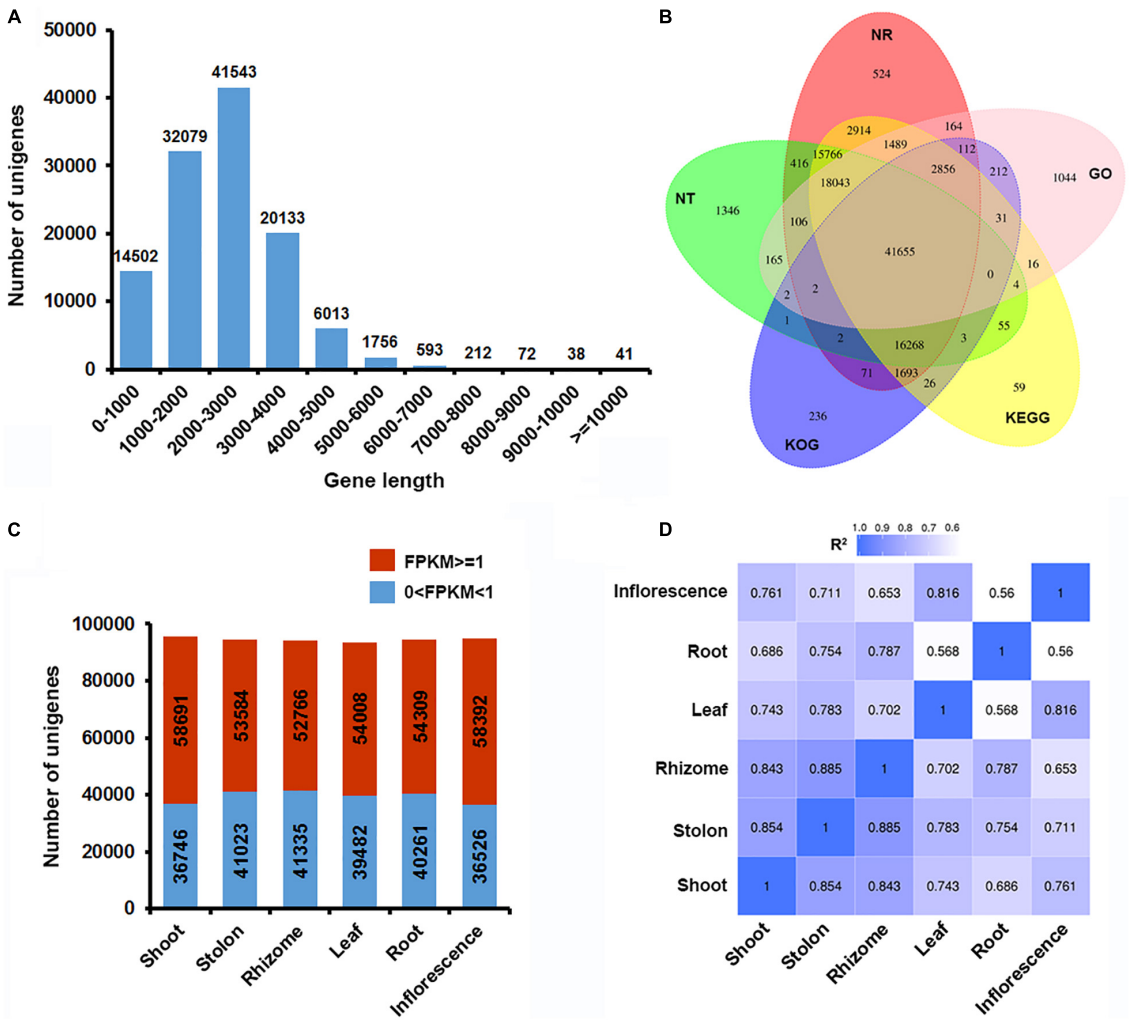


FIGURE 2 | Overview of the bermudagrass organ-specific transcriptome dataset. **(A)** Unigene length distribution. **(B)** Venn diagram of NR, NT, GO, KOG, and KEGG annotation results of the bermudagrass reference transcriptome. **(C)** Distribution of significantly expressed unigenes (FKPM ≥ 1) in each organs. **(D)** Pearson's correlation coefficients (R^2) in pairwise comparisons of the bermudagrass organ-specific transcriptome dataset.

the other four organs only had about 52,000–54,000 significantly expressed unigenes (**Figure 2C**). Pearson correlation R^2 values for all sequencing samples were between 0.82 and 1.0, with a mean value of 0.97, suggesting a high congruence of the biological replicates (**Supplementary Figure 1**). On the other hand, the expression profiles of different organs were highly divergent that the 1-R^2 value were between 0.15 and 0.44, with a mean value of 0.25 (**Figure 2D**). Specifically, the R^2 values of shoot, stolon and rhizome were similar and larger than that of other three organs, which is in accordance with their identities as different types of stems.

RT-qPCR Validation of the Bermudagrass Organ-Specific Transcriptome Dataset

RT-qPCR was routinely used to validate the reliability of transcriptome dataset (Ortiz-Ramírez et al., 2016), however, the reference gene for RT-qPCR must be correctly selected to obtain confident results (Liu et al., 2017). Using the organ-specific transcriptome dataset, we firstly evaluated the previously reported reference genes for quantitative analyses of gene expression in bermudagrass, including *alpha tubulin* (TUB), *actin* (ACT), *glyceraldehyde 3-phosphate dehydrogenase* (GAPDH), *elongation factor-1a* (EF1α), *TIP41-like family protein* (TIP41), *protein phosphatase 2A* (PP2A), *clathrin adaptor complex subunit* (CACS), and *E3 ubiquitin-protein ligase* (UPL7) (Chen Y. et al., 2015). We observed that the eight genes were not stably expressed in the six organs, showing divergent FPKM values in the organ-specific transcriptome dataset (**Figure 3A** and **Supplementary Table 5**). On the other hand, we found that ten unigenes, including *protein EMSY-LIKE 3 isoform X1* (EMSY3-1), *E3 ubiquitin ligase SUD1* (SUD1), *CCR4-NOT transcription complex subunit 11* (CNOT11), *transcription initiation factor TFIID subunit 12b* (TAF12B), *cyclic nucleotide-gated ion channel 17* (CNGC17), *chaperone protein dnaJ 49* (dnaJ49), *calmodulin-binding receptor-like cytoplasmic kinase 3* (CRCK3), *Spastin*

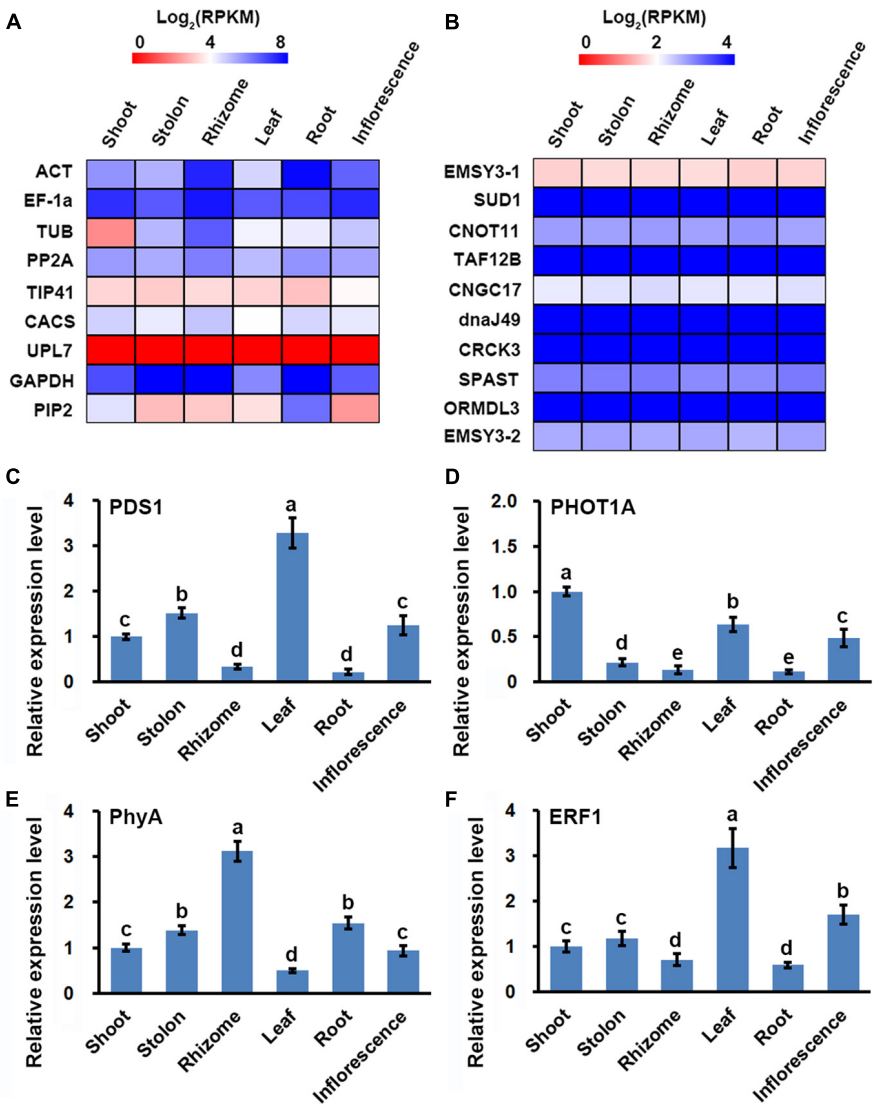


FIGURE 3 | Identification of the stably expressed unigenes and its application in RT-qPCR assay of gene expression in six bermudagrass organs. Heatmap of the expression of **(A)** classical reference genes used for RT-qPCR analysis, and **(B)** stably expressed unigenes identified from the bermudagrass organ-specific transcriptome dataset. RT-qPCR analysis of the expression of **(C)** PDS1, **(D)** PHOT1A, **(E)** PhyA, and **(F)** ERF1 in six bermudagrass organs using the stably expressed SUD1 as a reference. Error bars represent SE. Different letters indicate significant differences determined by Tukey's multiple comparison test.

(SPAST), *ORM1-like protein 3* (ORMDL3), and *protein EMSY-LIKE 3 isoform X2* (EMSY3-2), all showed relatively stable expression in the six organs with high ratios of minimal versus maximal expression level (>0.9) and small C.V. values (<0.1) (**Figure 3B** and **Supplementary Table 6**).

Using the stably expressed unigene SUD1 as a reference, we performed RT-qPCR assay of four randomly selected genes to test the validity of the organ-specific transcriptome dataset. The results indicated that *phytoene desaturase 1* (PDS1) and *ethylene responsive factor 1* (ERF1) genes were both highly expressed in leaf and weakly expressed in root, whereas *phototropin-1A* (PHOT1A) and *phytochrome A* (PhyA) were highly expressed in shoot and rhizome, respectively (**Figures 3C–F**). These results were highly correlated with the variations of FPKM values in the organ-specific transcriptome dataset since the Pearson correlation *R* value of the two results were all higher than 0.85 (**Supplementary Table 7**). Similar expression quantification result of ERF1 was also obtained in a study analyzing the function of ERF1 in cold tolerance of bermudagrass (Hu et al., 2020), which further verified the correctness of our results.

Identification of Unigenes Differentially Expressed in Three Types of Stems of Bermudagrass

To obtain an overview of the transcriptional diversity of the six organs, we plotted the average FPKM value of the significantly expressed 71,653 unigenes (FPKM ≥ 1) in the six organs

as a heatmap and analyzed the holistic expression pattern using hierarchical clustering. The results indicated that leaf and inflorescence were clustered in one group, whereas root, shoot, stolon, and rhizome were clustered in another group (**Figure 4A**). To further explore how different unigenes show an organ-specific expression profile, we analyzed the organ-enhanced and organ-enriched expression of the unigenes according to their FPKM values (Uhlén et al., 2016). The results indicated that there were 8,441 organ-enhanced unigenes and 3,028 organ-enriched unigenes in bermudagrass (**Figure 4B** and **Supplementary Tables 8 9**). It's noteworthy that root contained the largest proportion of enhance-expressed unigenes (40.90%) and enrich-expressed unigenes (64.04%), whereas unigenes enhance- or enrich-expressed in shoot, stolon and rhizome were much fewer. Specifically, only 22 and 19 unigenes were enrich-expressed in stolon and rhizome, respectively (**Supplementary Table 9**).

The small number of enhance- and enrich-expressed unigenes in shoot, stolon and rhizome compared with other organs suggested that stem specialization of bermudagrass is a complex process entailing the participation of multiple genes rather than a few unique genes. In order to identify unigenes possibly involved in stem specialization of bermudagrass, we further analyzed the gene expression profile of shoot, stolon and rhizome in a paired comparison manner and identified the differentially expressed unigenes among the three types of stems. The results indicated that 8,443 and 3,338 unigenes were differentially expressed between shoot and stolon, 5,179 and 4,361 unigenes were differentially expressed between stolon and rhizome, whereas 8,901 and 2,978 unigenes were differentially expressed between shoot and rhizome, respectively (**Figure 4C**). Furthermore, 5,375, 1,653, and 1,293 unigenes were preferentially expressed in shoot, stolon and rhizome, respectively (**Figure 4C**).

Sucrose Metabolism Was Differentially Regulated in Three Types of Stems of Bermudagrass

To better understand how the DEGs coordinate in the specialization of shoot, stolon, and rhizome in bermudagrass, KOBAS analyses were conducted to explore biological pathways enriched with DEGs in the three types of stems. The results indicated that as many as twenty pathways were significantly enriched with DEGs in shoot and stolon comparison and the top five pathways with largest number of DEGs were plant hormone signal transduction, glycolysis/gluconeogenesis, plant-pathogen interaction, starch and sucrose metabolism, and alpha-linolenic acid metabolism (**Figure 5A**). In stolon and rhizome comparison, there were ten pathways significantly enriched with DEGs, including ribosome, photosynthesis, fructose and mannose metabolism, carotenoid biosynthesis, and DNA replication (**Figure 5B**). In shoot and rhizome comparison, nine pathways were significantly enriched with DEGs and the top five pathways were plant hormone signal transduction, ribosome, starch and sucrose metabolism, glycolysis/gluconeogenesis, and photosynthesis (**Figure 5C**).

In plants, sucrose is degraded by glycolysis to provide energy and carbon skeleton for amino acid and other secondary metabolite biosynthesis, and also can be used for starch synthesis (Ruan, 2014). KOBAS analysis results indicated that the two routes of sucrose metabolism were both significantly

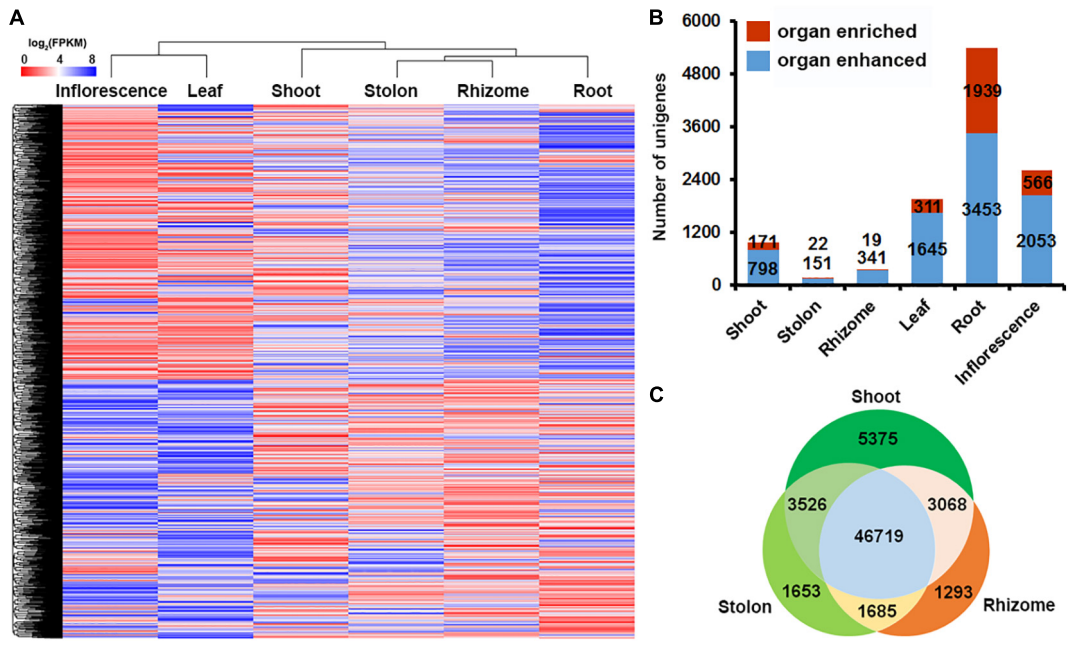


FIGURE 4 | Gene expression profiling in six bermudagrass organs. **(A)** Hierarchical clustering of all the significantly expressed unigenes across six bermudagrass organs based on $\log_2$ transformed FPKM values. **(B)** Distribution of organ-enhanced and organ-enriched unigenes in each organs. **(C)** Venn diagram of the DEGs in pairwise comparisons of the three types of stems.

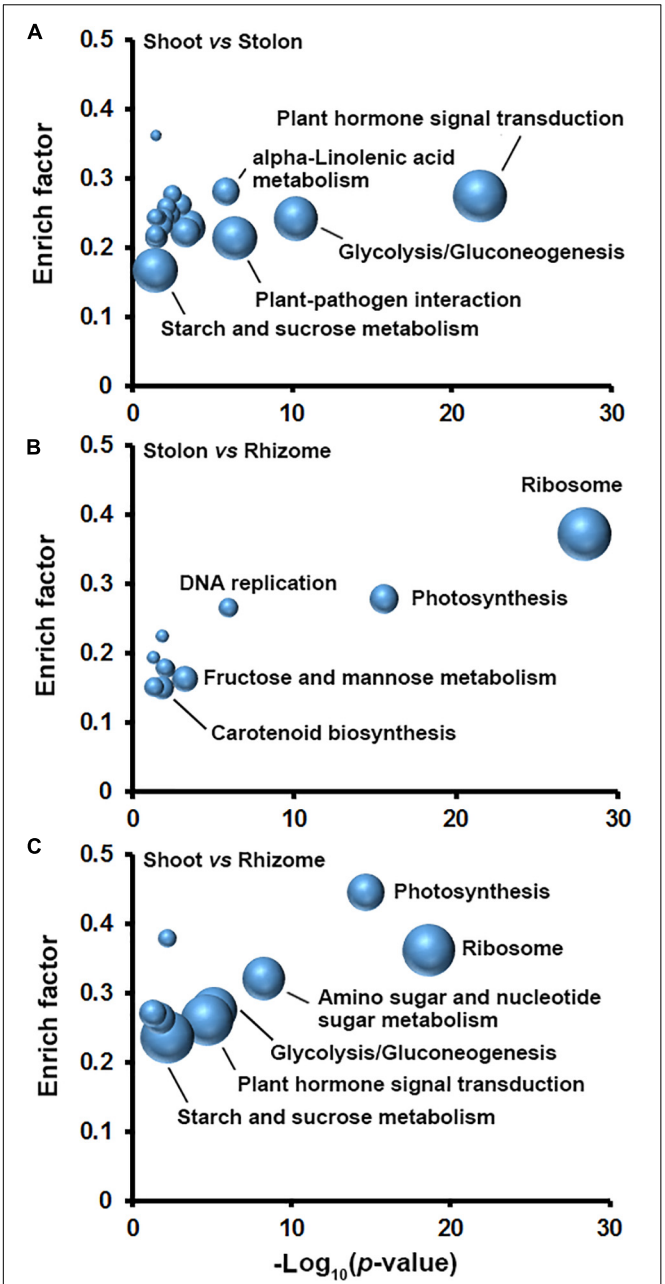


FIGURE 5 | Pathways enriched with DEGs in three types of stems of bermudagrass. DEGs identified in pairwise comparisons of **(A)** shoot and stolon, **(B)** stolon and rhizome, and **(C)** shoot and rhizome were analyzed in KOBAS to obtain the biochemical pathways enriched with DEGs. Fold enrichment (*y* axis) is plotted as a function of statistical significance (minus $\log_{10}$ transformed *p*-value, *x* axis). Node size corresponds to the number of DEGs.

enriched with DEGs in the three stems (**Figure 5**). Functional annotation indicated that there are 449 DEGs encoding different isoforms of 41 enzymes to catalyze multiple reactions of glycolysis/gluconeogenesis, and starch and sucrose metabolism (**Supplementary Table 10**). Interestingly, expression profiling indicated that 55 DEGs encoding starch synthesis-related enzymes, including glucose-1-phosphate adenylyltransferase (AGPase), starch synthase (SS), and starch branching enzyme (GBE), all showed relatively higher expression level in stolon and rhizome (**Figure 6A**). Furthermore, DEGs encoding SS and GBE all showed the highest expression level in the rhizome. By contrast, enzymes catalyzing the degradation of sucrose by glycolysis, especially fructose-bisphosphate aldolase (ADO), showed relatively higher mRNA expression level in the shoot and the lowest expression level in the rhizome (**Figure 6A**). In agreement with the transcriptomics analysis results, soluble sugar measurement results revealed that leaf has the highest soluble sugar content (>70 mg/g), shoot and inflorescence both have a medium soluble sugar content (about 40 mg/g), whereas stolon, rhizome and root all showed the lowest soluble sugar content (<30 mg/g) (**Figure 6B**). In contrast, starch measurement results indicated that rhizome has the highest starch content (>200 mg/g). Stolon also accumulate substantial starch (about 80 mg/g), whereas other four organs all showed relatively low starch content (<40 mg/g) (**Figure 6C**). Furthermore, enzyme activity assay indicated that rhizome and leaf has the highest and lowest SS enzyme activity, respectively (**Figure 6D**). These results collectively suggested that sucrose assimilated in the leaf by photosynthesis was mainly transported to the rhizome and stored in the form of starch.

Phytohormone and Transcription Regulatory Networks in Three Types of Stems of Bermudagrass

Considering that plant hormone signal transduction were also significantly enriched with DEGs in three types of stems (**Figure 5**), in combination with the important regulatory roles of phytohormone in stem growth and development (Gamuyao et al., 2017; Ge et al., 2019), we thoroughly analyzed the expression profile of 401 DEGs participating in plant hormone signal transduction. Among the 401 DEGs, 89, 41, 33, 71, 30, 31, 59, and 47 DEGs were annotated as auxin, cytokinine, gibberellin, abscisic acid, ethylene, brassinosteroid, jasmonic acid (JA), and salicylic acid signal transduction-related genes, respectively. Furthermore, the 401 DEGs were distributed in 35 of the 42 gene families constituting the different plant hormone signal transduction pathways (**Figure 7A** and **Supplementary Table 11**). Interestingly, expression profiling of the 35 gene families indicated that DEGs belonging to Gretchen Hagen 3 (GH3) family all showed relatively lower expression levels in shoot and stolon, whereas DEGs assigned as jasmonate resistant 1 (JAR1) all showed highest expression levels in the shoot and lowest expression levels in the rhizome (**Figure 7A**). RT-qPCR analyses of two unigenes, i2_LQ_mixture_c21936/f1p17/2403 and i2_LQ_mixture_c18254/f1p2/2233, which were annotated as *GH3.8* and *JAR1*, respectively, obtained similar results (**Figures 7B,C**). These results strongly suggested that auxin and JA signals were differently responded in three types of stems of bermudagrass.

Transcription factors play essential regulatory roles in plant organ differentiation and development (Durgaprasad et al., 2019;

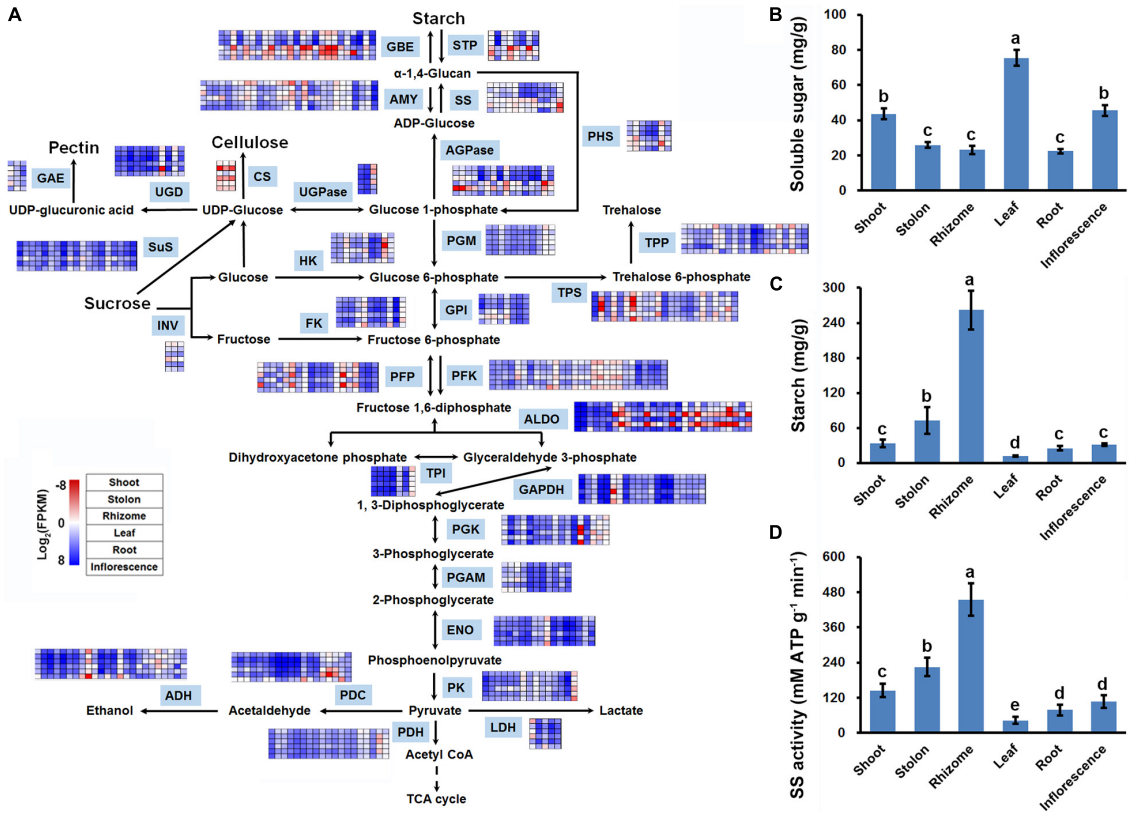


FIGURE 6 | Sucrose metabolic activity analysis in three types of stems of bermudagrass. **(A)** DEGs identified in pairwise comparisons of three stems were shown in the sucrose metabolic reactions. **(B)** Soluble sugar content, **(C)** Starch content, and **(D)** Activities of starch synthase in the six organs of bermudagrass. Error bars represent SE. Different letters indicate significant differences determined by Tukey's multiple comparison test.

Soyano et al., 2019). To provide insights into the transcription regulatory mechanism underlying shoot, stolon and rhizome specialization in bermudagrass, we further analyzed the dynamic expression of TFs in the three types of stems. Totally, 1,978 TFs belonging to 81 different families were found to be differentially expressed in the three types of stems, whereas 5,099 TFs were not differentially expressed (**Figure 8A** and **Supplementary Table 12**). Among the 81 TF families, bHLH, AP2/ERF, bZIP, WRKY, C2H2, MYB, and NAC all contained more than 70 DEGs, implying their in-depth participation in the shaping of the three stems. Notably, many TF families, including ARF, TIFY, were also classified as plant hormone signal transduction gene families (**Figure 7A** and **Supplementary Table 11**). Lastly, we also analyzed the expression pattern of the 1,978 TFs using hierarchical clustering. The results indicated that root, rhizome and stolon were clustered in one group, whereas shoot, leaf and inflorescence were clustered in another group (**Figure 8B**).

DISCUSSION

PacBio sequencing technique is an efficient approach to sequence long DNA molecules, leading to its widely application in full-length reference transcriptome construction of many plant species (Teng et al., 2019; Zhou et al., 2019; Xie et al., 2020). However, due to the inherent high error rate, transcriptome analyses based on PacBio sequencing is error prone and requires adjustment (Mahmoud et al., 2019). In this study, we successfully used the accurate Illumina sequencing data of six bermudagrass organs to correct the reference transcriptome obtained by PacBio sequencing (Zhang et al., 2018). After polishing and correction, the unigene number of the bermudagrass reference transcriptome was increased from 78,192 to 114,169, whereas the average unigene length was decreased from 2,317 bp to 2,237 bp (**Figure 2A**). The more accurate transcript information provided solid foundation for future molecular and breeding studies of this important grass species.

Identification of genes stably expressed in different tissues, organs, and growth conditions is important for the normalization of gene expression among samples in gene functional studies (Liu et al., 2017). In this study, we found that eight classical reference genes showed divergent expression pattern in different organs, whereas other ten unigenes were stably expressed in the six organs (**Figure 3**). Interestingly, although the ten unigenes participated in some essential cellular processes, including regulation of chromatin states (EMSY3-1 and EMSY3-2), protein degradation (SUD1), sphingolipid synthesis (ORMDL3), and transcription regulation (CNOT11 and TAF12B), they were not classical housekeeping genes (Robles et al., 2007;

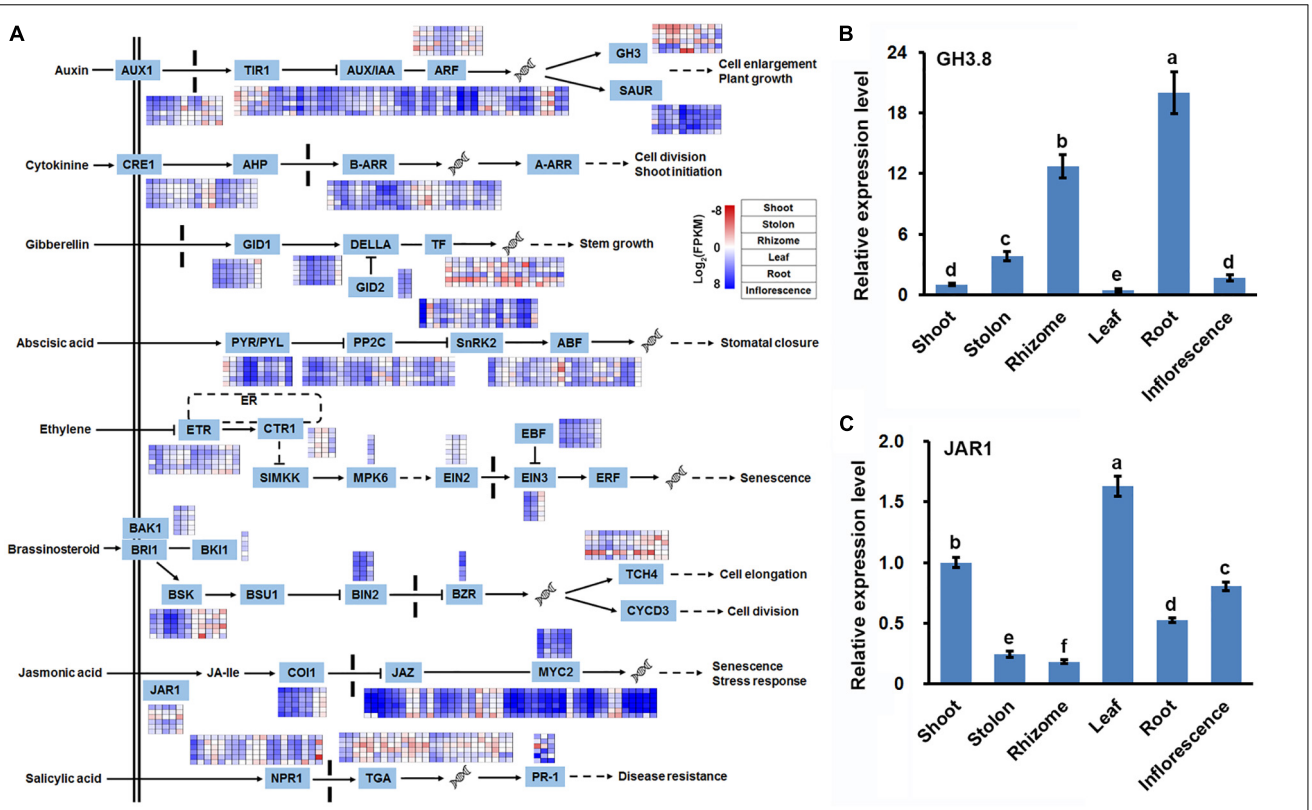


FIGURE 7 | Expression profile of unigenes involved in plant hormone signal transduction. **(A)** DEGs identified in pairwise comparisons of three stems were shown in the eight phytohormone signaling transduction pathways. RT-qPCR analysis of the expression of **(B)** GH3.8, and **(C)** JAR1 in six bermudagrass organs using the stably expressed SUD1 as a reference. Error bars represent SE. Different letters indicate significant differences determined by Tukey's multiple comparison test.

Breslow et al., 2010; Tsuchiya and Eulgem, 2011; Doblas et al., 2013; Laribee et al., 2015). These results were in line with the findings of transcriptomics survey of reference genes in other plants, which all identified many novel stably expressed genes (Mu et al., 2019; Wang et al., 2019; Long et al., 2020).

Using the organ-specific transcriptome dataset, we successfully analyzed the holistic gene expression of different organs in bermudagrass. Clustering analysis indicated that leaf and inflorescence showed similar gene expression profiles, whereas gene expression profiles in root, shoot, stolon, and rhizome are highly similar (**Figure 4A**). It was widely recognized that flowers are attractive and reproductive sexual leaf-like organs (Ruelens et al., 2017), thus it's not surprising that leaf and inflorescence of bermudagrass have similar gene expression profiles. The clustering of root with the three types of stems in another group is also understandable because root and stem have similar vascular tissue composition and both function in solute transport and mechanical support (Ruonala et al., 2017). Interestingly, clustering analysis of TFs in the six organs obtained a different result that root, rhizome and stolon were clustered in one group, whereas shoot, leaf and inflorescence were clustered in another group (**Figure 8B**). Considering that shoot, leaf and inflorescence are three aboveground organs growing in the sunshine, whereas root, rhizome and stolon are grown underground or shaded by the canopy, this result implied that light might be an important factor to induce the differential expression of TFs in these organs (Xie et al., 2016; Anna et al., 2019).

Our study also provided many new insights into the specialization of stems in bermudagrass. Firstly, gene expression profiling, sugar and starch content assay as well as SS enzyme activity analysis strongly suggested that sucrose was differently metabolized in the three types of stems (**Figure 6**). In shoot, sucrose was mainly degraded through glycolysis, whereas in stolon and rhizome, especially rhizome, sucrose was efficiently transformed to starch. This result was in line with the previous suspicions that rhizome of bermudagrass functions as a storage organ (Dong and de Kroon, 1994; Pornaro et al., 2019) and shoot has relatively higher glycolysis activity (Zhang et al., 2019). Secondly, expression profiling and RT-qPCR both indicated that *GH3* and *JAR1* were preferentially expressed in rhizome and shoot, respectively (**Figure 7**). *GH3* encodes an auxin-amido synthetase protein to promote the inactivation of auxin, thereby inhibiting the auxin signaling transduction (Aoi et al., 2020). On the other hand, *JAR1* encodes a JA-conjugating enzyme to catalyze the transformation of JA to active JA-isolecucine (Chen et al., 2018). The synergistic regulation of *GH3* and *JAR1* expression could lead to variant auxin and JA contents in different types of stems to promote their morphogenesis and functional specialization.

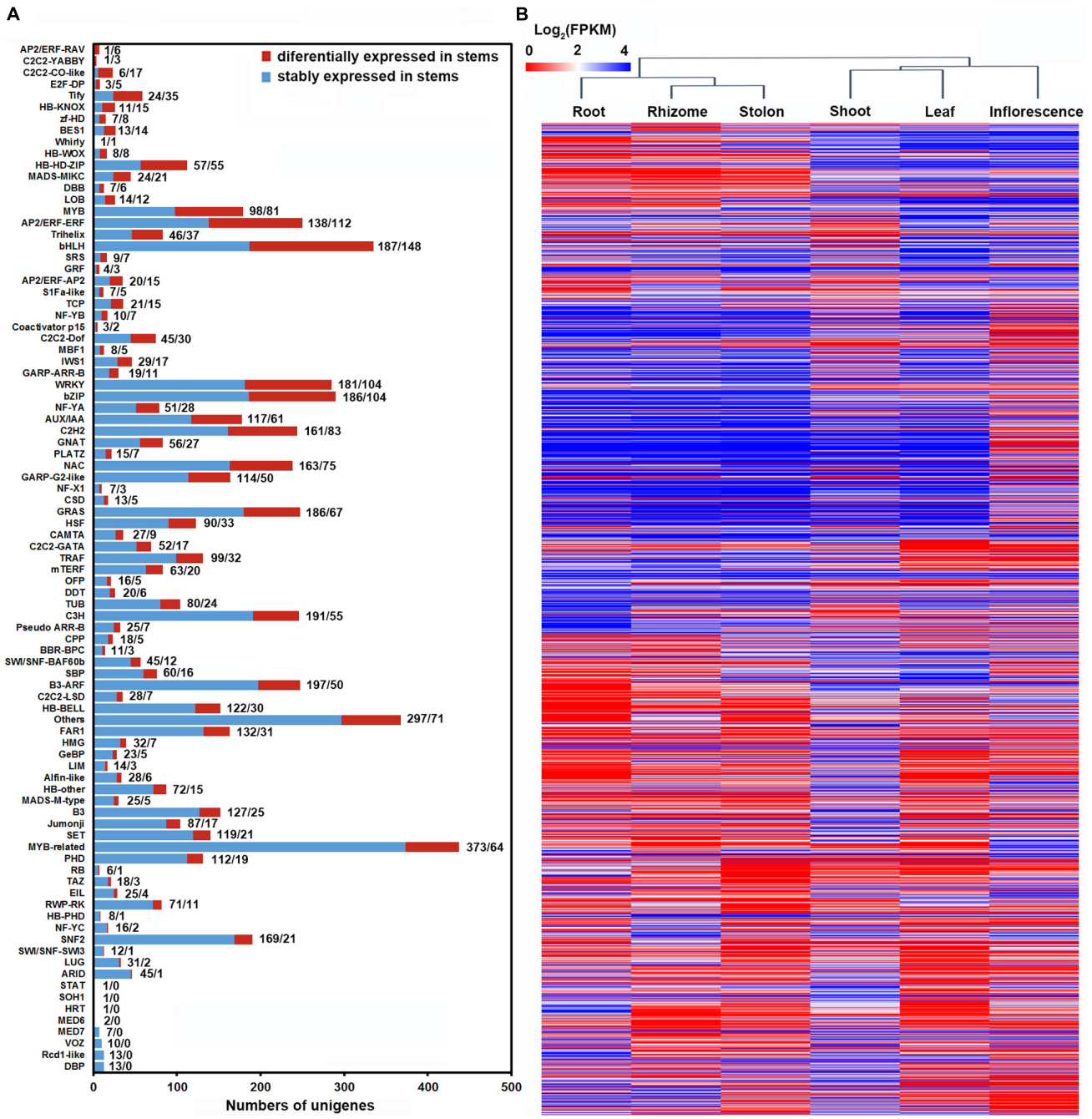


FIGURE 8 | Differential expression of TFs in three types of stems of bermudagrass. **(A)** Distribution of TFs differentially expressed in three types of stems across each TF families. **(B)** Hierarchical clustering of the differentially expressed TFs across six bermudagrass organs based on $\log_2$ transformed FPKM values.

CONCLUSION

In summary, an organ-specific transcriptome dataset of bermudagrass cultivar Yangjiang was successfully constructed in the current study. Comprehensive gene expression in six organs of bermudagrass, including leaf, root, inflorescence, shoot, stolon and rhizome, were analyzed using the transcriptome dataset. Gene expression profiling indicated that the expression of 8,441 and 3,028 unigenes were enhanced and enriched in different organs, respectively. Paired comparison further revealed that totally 11,762 unigenes were differentially expressed among shoot, stolon and rhizome. Notably, 401 unigenes participating in plant hormone signal transduction, 449 unigenes encoding enzymes involved in glycolysis/gluconeogenesis and starch/sucrose metabolism, and 1,978 TFs belonging to 81 families were all identified to be differentially expressed in the

three types of stems, implying the involvement of phytohormone, sucrose metabolism and TFs in the specialization of stem organs in bermudagrass. These results provided essential information for future functional studies of specific genes regulating the growth and development of different stems in bermudagrass and other plants.

DATA AVAILABILITY STATEMENT

RNA-seq data generated in this work has been submitted to the Bioproject database in NCBI under the accession number PRJNA685207.

AUTHOR CONTRIBUTIONS

BZ conceived and designed the study. SC, XX, and ZM conducted the experiments and analyzed the data. BZ and JL wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (32072613 and 31601786) and Open Project Program of Joint International Research Laboratory of Agriculture and Agri-Product Safety, Ministry of Education of China, and Yangzhou University (JILAR-KF202003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.678673/full#supplementary-material

Supplementary Figure 1 | Pearson's correlations for all the sequencing samples.

Supplementary Table 1 | Primers used for RT-qPCR.

Supplementary Table 2 | Annotation information of the unigenes in NR database.

Supplementary Table 3 | Sequencing and mapping statistics for each organ samples.

Supplementary Table 4 | Expression profiles of the significantly expressed unigenes.

Supplementary Table 5 | Expression profiles of the unigenes previously used as reference genes in RT-qPCR.

Supplementary Table 6 | Unigenes stably expressed in different organs.

Supplementary Table 7 | Expression profiles of four selected unigenes in different organs.

Supplementary Table 8 | Unigenes with enhanced expression in specific organs.

Supplementary Table 9 | Unigenes with enriched expression in specific organs.

Supplementary Table 10 | Expression profiles of the unigenes involved in sucrose metabolism and differentially expressed in three types of stems.

Supplementary Table 11 | Expression profiles of the unigenes involved in plant hormone signal transduction and differentially expressed in three types of stems.

Supplementary Table 12 | Expression profiles of the transcription factors possibly involved in stem specialization.

REFERENCES

Anna, B. B., Grzegorz, B., Marek, K., Piotr, G., and Marcin, F. (2019). Exposure to high-intensity light systemically induces micro-transcriptomic changes in *Arabidopsis thaliana* roots. *Int. J. Mol. Sci.* 20:5131. doi: 10.3390/ijms20205131

Aoi, Y., Tanaka, K., Cook, S. D., Hayashi, K. I., and Kasahara, H. (2020). GH3 auxin-amido synthetases alter the ratio of indole-3-acetic acid and phenylacetic acid in *Arabidopsis*. *Plant Cell Physiol.* 61, 596–605. doi: 10.1093/pcp/pcz223

Breslow, D. K., Collins, S. R., Bodenmiller, B., Aebersold, R., Simons, K., Shevchenko, A., et al. (2010). Orm family proteins mediate sphingolipid homeostasis. *Nature* 463, 1048–1053. doi: 10.1038/nature08787

Chen, H. J., Fu, T. Y., Yang, S. L., and Hsieh, H. L. (2018). FIN219/JAR1 and cryptochrome1 antagonize each other to modulate photomorphogenesis under blue light in *Arabidopsis*. *PLoS Genet.* 14:e1007248. doi: 10.1371/journal.pgen.1007248

Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A., et al. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol.* 166, 252–264. doi: 10.1104/pp.114.240689

Chen, L., Fan, J., Hu, L., Hu, Z., Xie, Y., Zhang, Y., et al. (2015). A transcriptomic analysis of bermudagrass (*Cynodon dactylon*) provides novel insights into the basis of low temperature tolerance. *BMC Plant Biol.* 15:216. doi: 10.1186/s12870-015-0598-y

Chen, Y., Tan, Z., Hu, B., Yang, Z., Xu, B., Zhuang, L., et al. (2015). Selection and validation of reference genes for target gene analysis with quantitative RT-PCR in leaves and roots of bermudagrass under four different abiotic stresses. *Physiol. Plant* 155, 138–148. doi: 10.1111/ppl.12302

Doblas, V. G., Amorim-Silva, V., Posé, D., Rosado, A., Esteban, A., Arró, M., et al. (2013). The SUD1 gene encodes a putative E3 ubiquitin ligase and is a positive regulator of 3-hydroxy-3-methylglutaryl coenzyme a reductase activity in *Arabidopsis*. *Plant Cell* 25, 728–743. doi: 10.1105/tpc.112.108696

Dong, M., and de Kroon, H. (1994). Plasticity in morphology and biomass allocation in *Cynodon dactylon*, a grass species forming stolons and rhizomes. *Oikos* 70, 99–106. doi: 10.2307/3545704

Durgaprasad, K., Roy, M. V., Venugopal, M. A., Kareem, A., Raj, K., Willemsen, V., et al. (2019). Gradient expression of transcription factor imposes a boundary on organ regeneration potential in plants. *Cell Rep.* 29, 453–463. doi: 10.1016/j.celrep.2019.08.099

Fan, J., Lou, Y., Shi, H., Chen, L., and Cao, L. (2019). Transcriptomic analysis of dark-induced senescence in bermudagrass (*Cynodon dactylon*). *Plants* 8:614. doi: 10.3390/plants8120614

Fernandez, O. N. (2003). Establishment of *Cynodon dactylon* from stolon and rhizome fragments. *Weed Res.* 43, 130–138. doi: 10.1046/j.1365-3180.2003.00324.x

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gamuyao, R., Nagai, K., Ayano, M., Mori, Y., Minami, A., Kojima, M., et al. (2017). Hormone distribution and transcriptome profiles in bamboo shoots provide insights on bamboo stem emergence and growth. *Plant Cell Physiol.* 58, 702–716. doi: 10.1093/pcp/pcx023

Ge, Y., Lai, Q., Luo, P., Liu, X., and Chen, W. (2019). Transcriptome profiling of *Gerbera hybrida* reveals that stem bending is caused by water stress and regulation of abscisic acid. *BMC Genom.* 20:600. doi: 10.1186/s12864-019-5961-1

Guo, L., Plunkert, M., Luo, X., and Liu, Z. (2020). Developmental regulation of stolon and rhizome. *Curr. Opin. Plant Biol.* 59:101970. doi: 10.1016/j.pbi.2020.10.003

Hill, G. M., Gates, R. N., and West, J. W. (2001). Advances in bermudagrass research involving new cultivars for beef and dairy production. *J. Anim. Sci.* 79, E48–E58.

Horowitz, M. (1972). Spatial growth of *Cynodon dactylon* (L.) Pers. *Weed Res.* 12, 373–383. doi: 10.1111/j.1365-3180.1972.tb01231.x

Hu, L., Li, H., Chen, L., Lou, Y., Amombo, E., and Fu, J. (2015). RNA-seq for gene identification and transcript profiling in relation to root growth of bermudagrass (*Cynodon dactylon*) under salinity stress. *BMC Genom.* 16:575. doi: 10.1186/s12864-015-1799-3

Hu, R., Yu, C., Wang, X., Jia, C., Pei, S., He, K., et al. (2017). De novo transcriptome analysis of *Miscanthus Lutarioriparius* identifies candidate genes in rhizome development. *Front. Plant Sci.* 8:492. doi: 10.3389/fpls.2017.00492

Hu, Z., Huang, X., Amombo, E., Liu, A., Fan, J., Bi, A., et al. (2020). The ethylene responsive factor CdERF1 from bermudagrass (*Cynodon dactylon*) positively regulates cold tolerance. *Plant Sci.* 294:110432. doi: 10.1016/j.plantsci.2020.110432

Jiménez-Brenes, F. M., López-Granados, F., Torres-Sánchez, J., Peña, J. M., Ramírez, P., Castillejo-González, I. L., et al. (2019). Automatic UAV-based detection of *Cynodon dactylon* for site-specific vineyard management. *PLoS One* 14:e0218132. doi: 10.1371/journal.pone.0218132

Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. doi: 10.1111/tpj.13312

Klepikova, A. V., and Penin, A. A. (2019). Gene expression maps in plants: current state and prospects. *Plants* 8:309. doi: 10.3390/plants8090309

Laribee, R. N., Hosni-Ahmed, A., Workman, J. J., and Chen, H. (2015). Ccr4-not regulates RNA polymerase I transcription and couples nutrient signaling to the control of ribosomal RNA biogenesis. *PLoS Genet.* 11:e1005113. doi: 10.1371/journal.pgen.1005113

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

Liu, Y., Liu, J., Xu, L., Lai, H., Chen, Y., Yang, Z., et al. (2017). Identification and validation of reference genes for *Seashore Paspalum* response to abiotic stresses. *Int. J. Mol. Sci.* 18:1322. doi: 10.3390/ijms18061322

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

Long, X., Lu, J., Kav, N. N. V., Qin, Y., and Fang, Y. (2020). Identification and evaluation of suitable reference genes for gene expression analysis in rubber tree leaf. *Mol. Biol. Rep.* 47, 1921–1933. doi: 10.1007/s11033-020-05288-8

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.

Mahmoud, M., Zywicki, M., Twardowski, T., and Karlowski, W. M. (2019). Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics* 111, 43–49. doi: 10.1016/j.ygeno.2017.12.011

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Melmaiee, K., Anderson, M., Elavarthi, S., Guenzi, A., and Canaan, P. (2015). Transcriptional analysis of resistance to low temperatures in bermudagrass crown tissues. *PLoS One* 10:e0136433. doi: 10.1371/journal.pone.0136433

Mu, J., Chen, L., Gu, Y., Duan, L., Han, S., Li, Y., et al. (2019). Genome-wide identification of internal reference genes for normalization of gene expression values during endosperm development in wheat. *J. Appl. Genet.* 60, 233–241. doi: 10.1007/s13353-019-00503-0

Nagori, B. P., and Solanki, R. (2011). *Cynodon dactylon* (L.) Pers.: a valuable medicinal plant. *Res. J. Med. Plant* 5, 508–514. doi: 10.3923/rjmp.2011.508.514

Nautiyal, A. K., Gani, U., Sharma, P., Kundan, M., Fayaz, M., Lattoo, S. K., et al. (2020). Comprehensive transcriptome analysis provides insights into metabolic and gene regulatory networks in trichomes of *Nicotiana tabacum*. *Plant Mol. Biol.* 102, 625–644. doi: 10.1007/s11103-020-00968-2

Ortiz-Ramírez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., et al. (2016). A transcriptome atlas of Physcomitrella patens provides insights into the evolution and development of land plants. *Mol. Plant* 9, 205–220. doi: 10.1016/j.molp.2015.12.002

Pornaro, C., Macolino, S., and Richardson, M. D. (2019). Rhizome and stolon development of bermudagrass cultivars in a transition-zone environment. *Acta Agric. Scand. B SP* 69, 657–666. doi: 10.1080/09064710.2019.1639805

Reasor, E. H., Brosnan, J. T., Trigiano, R. N., Elsner, J. E., Henry, G. M., and Schwartz, B. M. (2016). The genetic and phenotypic variability of interspecific hybrid bermudagrasses (*Cynodon dactylon* (L.) Pers. × *C. transvaalensis* Burtt-Davy) used on golf course putting greens. *Planta* 244, 761–773. doi: 10.1007/s00425-016-2573-8

Robles, L. M., Wampole, J. S., Christians, M. J., and Larsen, P. B. (2007). Arabidopsis enhanced ethylene response 4 encodes an EIN3-interacting TFIID transcription factor required for proper ethylene response, including ERF1 induction. *J. Exp. Bot.* 58, 2627–2639. doi: 10.1093/jxb/erm080

Ruan, Y. L. (2014). Sucrose metabolism: gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* 65, 33–67. doi: 10.1146/annurev-arplant-050213-040251

Ruelens, P., Zhang, Z., van Mourik, H., Maere, S., Kaufmann, K., and Geuten, K. (2017). The origin of floral organ identity quartets. *Plant Cell* 29, 229–242. doi: 10.1105/tpc.16.00366

Ruonala, R., Ko, D., and Helariutta, Y. (2017). Genetic networks in plant vascular development. *Annu. Rev. Genet.* 51, 335–359. doi: 10.1146/annurev-genet-120116-024525

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Sánchez-Sevilla, J. F., Vallarino, J. G., Osorio, S., Bombarely, A., Posé, D., Merchante, C., et al. (2017). Gene expression atlas of fruit ripening and transcriptome assembly from RNA-seq data in octoploid strawberry (*Fragaria*×*ananassa*). *Sci. Rep.* 7:13737.

Shi, H., Jiang, C., Ye, T., Tan, D. X., Reiter, R. J., Zhang, H., et al. (2015). Comparative physiological, metabolomic, and transcriptomic analyses reveal mechanisms of improved abiotic stress resistance in bermudagrass [*Cynodon dactylon* (L). Pers.] by exogenous melatonin. *J. Exp. Bot.* 66, 681–694. doi: 10.1093/jxb/eru373

Soyano, T., Shimoda, Y., Kawaguchi, M., and Hayashi, M. (2019). A shared gene drives lateral root development and root nodule symbiosis pathways in Lotus. *Science* 366, 1021–1023. doi: 10.1126/science.aax2153

Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., et al. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genom.* 9, 1–16.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117

Teng, K., Teng, W., Wen, H., Yue, Y., Guo, W., Wu, J., et al. (2019). PacBio single-molecule long-read sequencing shed new light on the complexity of the *Carex breviculmis* transcriptome. *BMC Genom.* 20:789. doi: 10.1186/s12864-019-6163-6

Tsuchiya, T., and Eulgem, T. (2011). EMSY-like genes are required for full RPP7-mediated race-specific immunity and basal defense in *Arabidopsis*. *Mol. Plant Microb. Interact.* 24, 1573–1581. doi: 10.1094/mpmi-05-11-0123

Uhlén, M., Hallström, B. M., Lindskog, C., Mardinoglu, A., Pontén, F., and Nielsen, J. (2016). Transcriptomics resources of human tissues and organs. *Mol. Syst. Biol.* 12:862. doi: 10.15252/msb.20155865

Wang, X., Wu, Z., Bao, W., Hu, H., Chen, M., Chai, T., et al. (2019). Identification and evaluation of reference genes for quantitative real-time PCR analysis in *Polygonum cuspidatum* based on transcriptome data. *BMC Plant Biol.* 19:498. doi: 10.1186/s12870-019-2108-0

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322.

Xie, L., Teng, K., Tan, P., Chao, Y., Li, Y., Guo, W., et al. (2020). PacBio single-molecule long-read sequencing shed new light on the transcripts and splice isoforms of the perennial ryegrass. *Mol. Genet. Genom.* 295, 475–489. doi: 10.1007/s00438-019-01635-y

Xie, X. F., Hu, Y. K., Pan, X., Liu, F. H., Song, Y. B., and Dong, M. (2016). Biomass allocation of stoloniferous and rhizomatous plant in response to resource availability: a phylogenetic meta-analysis. *Front. Plant Sci.* 7:603. doi: 10.3389/fpls.2016.00603

Yi, F., Gu, W., Chen, J., Song, N., Gao, X., Zhang, X., et al. (2019). High temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell* 31, 974–992. doi: 10.1105/tpc.18.00961

Zhang, B., Fan, J., and Liu, J. (2019). Comparative proteomic analysis provides new insights into the specialization of shoots and stolons in bermudagrass (*Cynodon dactylon* L.). *BMC Genom.* 20:708. doi: 10.1186/s12864-019-6077-3

Zhang, B., Liu, J., Wang, X., and Wei, Z. (2018). Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol. Biochem.* 132, 95–103. doi: 10.1016/j.plaphy.2018.08.039

Zhao, T., Specht, C. D., Dong, Z., Ye, Y., Liu, H., and Liao, J. (2020). Transcriptome analyses provide insights into development of the *Zingiber zerumbet* flower, revealing potential genes related to floral organ formation and patterning. *Plant Growth Regul.* 90, 331–345. doi: 10.1007/s10725-020-00575-7

Zhou, Q., Huang, M., Huang, X., Liu, J., Wang, X., Cai, J., et al. (2018). Effect of post-anthesis waterlogging on biosynthesis and granule size distribution of starch in wheat grains. *Plant Physiol. Biochem.* 132, 222–228. doi: 10.1016/j.plaphy.2018.08.035

Zhou, S., Zhang, J., Han, H., Zhang, J., Ma, H., Zhang, Z., et al. (2019). Full-length transcriptome sequences of *Agropyron cristatum* facilitate the prediction of putative genes for thousand-grain weight in a wheat-*A. cristatum* translocation line. *BMC Genom.* 20:1025. doi: 10.1186/s12864-019-6416-4

Zhu, H., Yu, X., Xu, T., Wang, T., Du, L., Ren, G., et al. (2015). Transcriptome profiling of cold acclimation in bermudagrass (*Cynodon dactylon*). *Sci. Hortic.* 194, 230–236. doi: 10.1016/j.scienta.2015.08.024

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.

Copyright © 2021 Chen, Xu, Ma, Liu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fgene.2021.692702

Transcriptomics and Metabolomics Reveal Purine and Phenylpropanoid Metabolism Response to Drought Stress in *Dendrobium sinense*, an Endemic Orchid Species in Hainan Island

Cuili Zhang[1†], Jinhui Chen[1,2†], Weixia Huang[1], Xiqiang Song[1] and Jun Niu[1*]*

[1] Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, College of Forestry, Hainan University, Haikou, Hainan, China, [2] Engineering Research Center of Rare and Precious Tree Species in Hainan Province, College of Forestry, Hainan University, Haikou, Hainan, China

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Chunyu Zhang,
Beijing Forestry University, China
Quanxin Bi,
Chinese Academy of Forestry, China

***Correspondence:**
Xiqiang Song
songstrong@hainanu.edu.cn
Jun Niu
niujun@hainanu.edu.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: 09 April 2021
Accepted: 24 May 2021
Published: 02 July 2021

Citation:
Zhang C, Chen J, Huang W, Song X and Niu J (2021) Transcriptomics and Metabolomics Reveal Purine and Phenylpropanoid Metabolism Response to Drought Stress in Dendrobium sinense, an Endemic Orchid Species in Hainan Island. Front. Genet. 12:692702. doi: 10.3389/fgene.2021.692702

Drought stress is a bottleneck factor for plant growth and development, especially in epiphytic orchids that absorb moisture mainly from the air. Recent studies have suggested that there are complex transcriptional regulatory networks related to drought stress in *Dendrobium sinense.* In this study, the transcription and metabolite alterations involved in drought stress response in *D. sinense* were investigated through RNA-seq and metabolomics. A total of 856 metabolites were identified from stressed and control samples, with 391 metabolites showing significant differences. With PacBio and Illumina RNA sequencing, 72,969 genes were obtained with a mean length of 2,486 bp, and 622 differentially expressed genes (DEGs) were identified. Correlation analysis showed 7 differential genes, and 39 differential metabolites were involved in interaction networks. The network analysis of differential genes and metabolites suggested that the pathways of purine metabolism and phenylpropanoid biosynthesis may play an important role in drought response in *D. sinense*. These results provide new insights and reference data for culturally important medicinal plants and the protection of endangered orchids.

Keywords: *Dendrobium sinense*, drought stress, metabolomics, transcriptomics, third-generation sequencing, purine metabolism, phenylpropanoid biosynthesis

INTRODUCTION

Through evolutionary processes, a series of response mechanisms has developed in plants that allow for chemical, physiological, and developmental responses to changes in the external environment (Claeys and Inze, 2013). Drought is one abiotic factor that strongly affects plant growth and development (Singh and Laxmi, 2015; Zhu, 2016). Several studies have elucidated important genes involved in drought stress response, mainly by studying gene expression patterns of drought-stressed plants (Joshi et al., 2016). For example, drought resistance in *Triticum aestivum* can be improved by *TaFBA* overexpression that affects the accumulation of sucrose and starch (Zhou et al., 2014). Additionally, aldehyde dehydrogenases (ALDHs) have been considered as general detoxifying enzymes that eliminate abiotic stress in a variety of organisms. Transgenic *Arabidopsis* plants expressing *CsALDH12A1* showed enhanced tolerance to drought stress during

plant development (Duan et al., 2015). Under drought stress, the overexpressions of *allene oxide synthase* (*AOS*) and *allene oxide cyclase* (*AOC*) genes were positively correlated with the levels of major oxylipin metabolites in the *AOS* branch of the pathway, ultimately leading to the synthesis of jasmonates (Stenzel et al., 2003; Domenico et al., 2012). In CBL-interacting protein kinases (CIPKs) signaling, TaCIPK23 plays an important role in drought stress response in wheat (Cui et al., 2018).

Previous studies have reported that drought stress can induce many metabolites, such as carbohydrates and amino acids (Maruyama et al., 2009, 2014). Sugars serve not only as a solute to regulate cell osmotic pressure but also as primary messengers for signal transduction (Baena-González et al., 2007). Recent research in *Arabidopsis* indicated that glucose passes through hexokinase-dependent pathways and phytohormone response pathways, which has been proven to associate with participation in adversity (Sharma et al., 2019). *OsTF1L* overexpression in rice increased drought tolerance by regulating the gene expression involved in lignin biosynthesis (Bang et al., 2018). In addition, drought stress in rice seeds could induce the expression of *OsNCED3*, promoting the biosynthesis of abscisic acid (ABA) and inhibiting seed germination (Chen et al., 2019). Thus, abundant evidence shows that plants clearly regulate the accumulation of metabolites and gene transcription to respond to drought stress (Morimoto et al., 2017).

In recent years, Illumina sequencing has accelerated research of the transcriptome, which has made great achievements in functional gene verification (Du et al., 2016) and gene mining (Zhang et al., 2019). Although the short reads produced by this second-generation sequencing technology has high accuracy, the short-read length affects assembly and mapping accuracy in the absence of a reference genome, resulting in short transcript splicing and incomplete transcripts (Du and Sun, 2016). In contrast, third-generation long-read sequencing, such as PacBio, can generate full-length transcript sequences but has high error rates. Therefore, the depth of second-generation sequencing can complement the full-length transcript sequencing of third-generation sequencing, identifying subtypes with higher accuracy in transcriptome sequencing, especially for plants without reference genomes (Du and Sun, 2016).

Dendrobium sinense is an endangered epiphytic orchid plant with high medicinal value (Sun et al., 2014). It is well known that epiphytic orchids usually attach to smooth rocks and tree trunks that have low water retention capacity. Compared with terrestrial plants, epiphytes are subjected to moderate drought stress for most of the year (Zhang et al., 2016). As a consequence, moisture stress is the most important abiotic factor limiting the growth and development of epiphytic orchids (Zotz and Bader, 2009). However, few studies have combined the analysis of gene transcripts and metabolites to explore how epiphytic orchids will response to water deficit. In this present work, differential metabolites of *D. sinense* under drought stress were obtained by metabolome analysis. In addition to greatly increasing molecular resources devoted to *D. sinense* that lack genome sequence information, the combination of PacBio and Illumina sequencing illustrates gene expression patterns in drought stress. Combined with transcriptome and metabolome data, the importantly metabolic pathways related to drought response were identified in *D. sinense*. Our data provide new insights and reference data for improving drought resistance in these endangered orchids.

MATERIALS AND METHODS

Plant Materials and Drought Stress Management

Wild *D. sinense* plants were collected at the national nature reserves of Bawangling, Changjiang County, Hainan Province, China. Individuals were cultured at room conditions [25°C, relative humidity (RH) $\geq$ 95%] for 2 months. Then, we simulated a range of drought conditions by adjusting relative air humidity (RH) in separate growing containers for 7 days. Based on our previous investigation (Yang, 2009), *D. sinense* was assigned to grow under conditions for group A (RH $\geq$ 95%), group B (45% $\leq$ RH $\leq$ 50%), or group C (RH $\leq$ 5%), which were named *D. sinense* A, B, and C (DSA, DSB, and DSC), respectively. After 7 days of growth in these conditions, leaves and pseudo-bulbs were collected for further metabolome and transcriptome studies, immediately frozen in liquid nitrogen and stored at –80°C.

For each treatment experiment, 100 mg sample was fully ground in liquid nitrogen for physiological test. The contents of catalase (CAT), peroxidase isoenzyme (POD), superoxide dismutase (SOD), and ascorbate peroxidase (APX) were determined using the respective Solarbio detection kits. The experimental data were analyzed by SPSS software.

Metabolome Detection and Analysis

DSA, DSB, and DSC samples were crushed using a mixer mill (MM 400, Retsch, Shanghai, China) with a zirconia bead for 1.5 min at 30 Hz. One hundred milligrams powder was dissolved in 1.0 ml 70% aqueous methanol and incubated overnight at 4°C. Following centrifugation (10,000*g* for 10 min), the extracts were absorbed by CNWBOND Carbon-GCB SPE Cartridge (ANPEL, Shanghai, China) and filtered by SCAA-104 of 0.22 μm pore size (ANPEL, Shanghai, China). The data of high-performance liquid chromatography (HPLC) was analyzed by Shim-pack UFLC SHIMADZU CBM30A system (Applied Biosystems 6500 QTRAP, Thermo Fisher Scientific, Waltham, United States).

The extracted samples were analyzed using a liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) system (UPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS/MS, Applied Biosystems 6500 QTRAP). The conditions for liquid phase analysis were as follows: chromatographic column, Waters ACQUITY UPLC HSS T3 C18 (1.8 μm, 2.1 mm × 100 mm); solvent system, water (0.04% acetic acid)/acetonitrile (0.04% acetic acid); gradient program, 95:5 V/V at 0 min, 5:95 V/V at 11.0 min, 5:95 V/V at 12.0 min, 95:5 V/V at 12.1 min, 95:5 V/V at 15.0 min; flowrate, 0.40 ml/min; temperature, 40°C; and injection volume, 2 μl. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS.

The ESI source operation parameters were as follows: electrospray ionization, source temperature 500°C; ion spray voltage (IS), 5,500 V; curtain gas (CUR) at 25.0 psi;

and the collision gas (CAD), high. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired as multiple reaction monitoring (MRM) experiments with collision gas (nitrogen) set to 5 psi. Declustering potential (DP) and collision energy (CE) for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period (Chen et al., 2013).

Metabolome Data

The software *analyst 1.6.3* was used to process the mass spectrometry data, and the metabolites of the samples were qualitatively and quantitatively analyzed based on Met Ware metabolic database (Wuhan, China). Orthogonal partial least-squares discriminant analysis (OPLS-DA) with supervised mode was used to analyze the differences within and among experimental groups. The metabolic contents were normalized by R software[1], through which an accumulation-mode clustering analysis [hierarchical clustering analysis (HCA)] of metabolites was performed to test for differences among samples (Chong and Xia, 2018). The pathway-based Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used for the functional annotation of differential metabolites (Kanehisa and Goto, 2000).

RNA Isolation and RNA-Seq Library Construction

Total RNA was extracted from mature non-aging leaves and pseudo-bulbs with the Quick RNA Isolation Kit (Waryong, Beijing, China). RNA quality and concentration of each sample were measured using agarose gel electrophoresis, Nanodrop 2000 (Thermo Fisher Scientific, Waltham, United States), and Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, United States). Subsequently, an Iso-Seq library was prepared using the Clontech SMARTer PCR cDNA Synthesis Kit and sequenced on a PacBio Sequel (PN 100-092-800-03).

Transcriptome Data Analysis

SMRTlink 6.0 software was used to process raw sequence data, and circular consensus sequences (CCS) were obtained by subread BAM files. The CCS BAM files were then classified as full-length non-chimera (FLNC) or non-full length (nFL) according to 5′-primer, 3′-primer and poly-A. Consensus sequences were obtained by clustering full-length sequences with isoform-level clustering (ICE), and additional nucleotide errors in consensus reads were corrected using the Illumina RNA-seq data using the LoRDEC software (Salmela and Rivals, 2014). The consensus sequences were corrected with non-full-length and non-chimeric sequences, resulting in polished consensus sequences (Fu et al., 2012). Finally, any redundancy in corrected consensus reads was removed by CD-HIT to obtain transcripts for subsequent analyses and as reference sequences (Fu et al., 2012). The RSEM software was used to estimate gene expression levels for each sample by mapping the transcript sequences to obtain the read count of each transcript (Li and Dewey, 2011). Considering the sequence depth and gene length of fragments, the gene expression level was calculated with the fragments per kilobase million (FPKM) method.

The Analysis of Differential Expression and Functional Annotation

DESeq-R package (1.10.1) was used to analyze the differential expression between two groups. The differentially expressed genes (DEGs) were selected by $|\log2\ (\text{fold change})| > 1$ and false discovery rate (FDR) < 0.05 (Storey and Tibshirani, 2003). Gene annotations and gene ontology (GO) enrichment analyses were implemented with the Go seq-R package. GO terms with FDR ≤ 0.05 were considered significantly enriched by DEGs. The KEGG database was used as a resource for understanding high-level functions and utilities of the biological system[2]. The KOBAS (3.0) software was used to test the statistical enrichment of DEGs in KEGG pathways.

Statistical Analysis

The Pearson coefficient between metabolome and transcriptome data were calculated using R (cor test) software. Correlations corresponding to $R^2 > 0.8$ and FDR < 0.05 were selected. The correlated data were visualized using Cytoscape software and KEGG database. Differences between samples were tested for statistical significance using the Duncan MRT method. Statistical analysis was implemented by SPSS software (version 19.0).

RESULTS

Effects of Drought Stress on Protective Enzymes in Pseudobulbs of *D. sinense*

In order to explore the physiological mechanism of *D. sinense* response to drought stress, three treated experiments were conducted with different RH gradients sustained for 7 days. The color of DSA leaves was healthy green, while DSB and DSC leaves showed a noticeably yellow color with the increase in stress degree (**Supplementary Figure 1**). There are protective enzymes to scavenging reactive oxygen species (ROS) in plants, and the high activity of the enzyme is closely related to the degree of stress. Catalase and POD activities displayed a significant gradual increase ($p < 0.05$) with the intensification of drought stress (**Figure 1**). Superoxide dismutase and APX protective enzyme activities were only significantly different ($p < 0.05$) between DSA and DSC (**Figure 1**). The results showed that the activity of protective enzymes in *D. sinense* was increased, and the scavenging ability of reactive oxygen species was enhanced after drought stress.

Identification and Annotation of Metabolites in *D.sinense*

A total of 856 metabolites were detected, belonging to 16 categories (**Supplementary Table 1**). Among them, the most

[1]www.r-project.org/

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

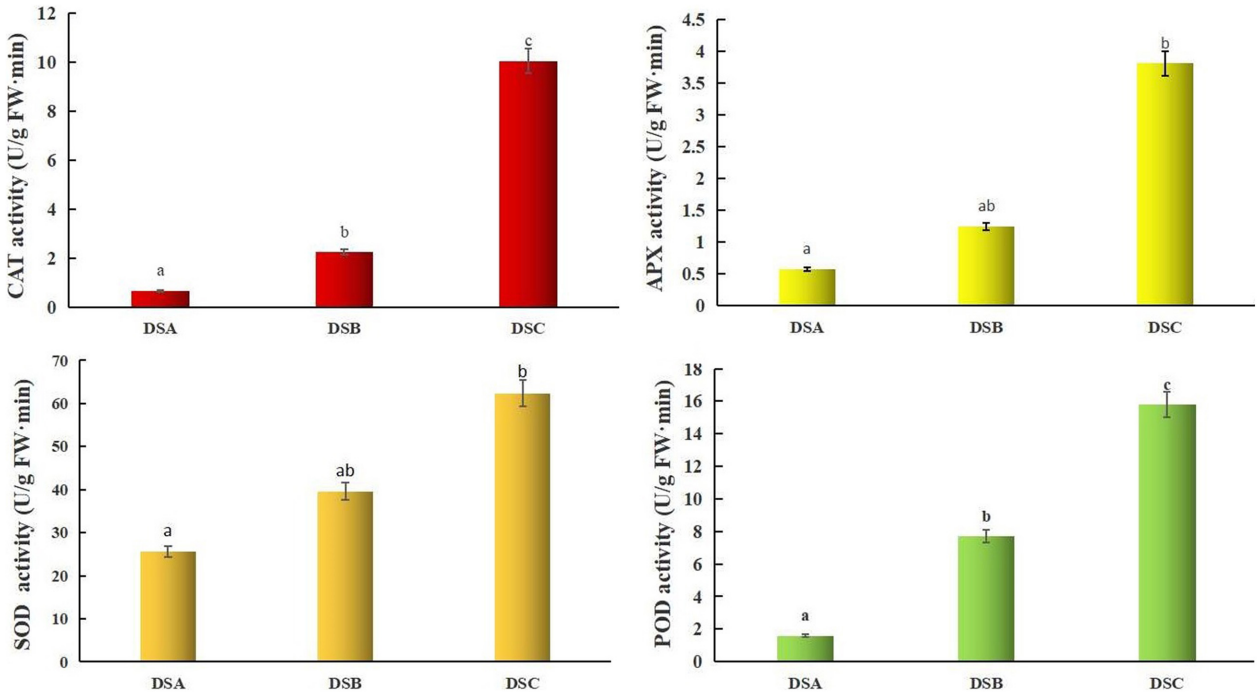


FIGURE 1 | Effects of drought stress on physiological and biochemical parameters in pseudobulb of *D. sinense*. Data are means ± SE of three separate measurements. The significance of differences was analyzed using *Duncan* method.

abundant were flavonoids (240; 28.04%), organic acids and derivatives (106; 12.38%), and amino acids and derivatives (93; 10.86%). Differential abundance of metabolites indicated 391 differential metabolites related to drought stress (**Figure 2**), mainly including flavonoids (116, 29.67%), phenylpropanoids (38, 9.72%), amino acids and derivatives (37, 9.46%), and lipids (35, 8.95%) (**Supplementary Table 1**). Interestingly, upregulated metabolites were more common than downregulated metabolites under drought stress, as observed in the comparisons of DSA vs. DSB and DSA vs. DSC groups (**Supplementary Figure 2**). Additionally, when comparing DSA (control) to DSB (moderate drought stress), 32 of the 151 identified upregulated metabolites were increased by more than 2.5-fold with a variable importance in project (VIP) score > 1.5. Of these, acetosyringone, 4-methylcatechol, pyrocatechol, apigenin 6-C-hexosyl-8-C-hexosyl-O-hexoside, and parthenolide increased most dramatically by more than >2,000-fold change and VIP > 1.7 (**Supplementary Table 2**). DSA vs. DSC (extreme drought stress) also included 32 upregulated metabolites with more than 2.5-fold change and VIP > 1.5. Of these, c-hexosyl-apigenin c-pentoside, acetosyringone, magnolol, 4-methylcatechol, pyrocatechol, and catechol increased by more than 2,000-fold (**Supplementary Table 2**). Intriguingly, 13 of the same metabolites with significant differences were detected in DSA vs. DSB and DSA vs. DSC, such as, L-tryptamine, S-(5′-adenosy)-L-homocysteine, *p*-coumaric acid, and 4-methylcatechol (**Supplementary Table 2**).

To further study the response to drought stress, KEGG enrichment pathway analysis was performed (**Supplementary Table 3**). The differential metabolites in the significant enrichment pathway were analyzed by cluster analysis. Under moderate drought stress, organic acids and their derivatives and procyanidin metabolites were significantly enriched. Under more extreme drought stress (RH ≤ 5%), flavonoids and carbohydrates were significantly enriched (**Figure 3** and **Supplementary Table 3**).

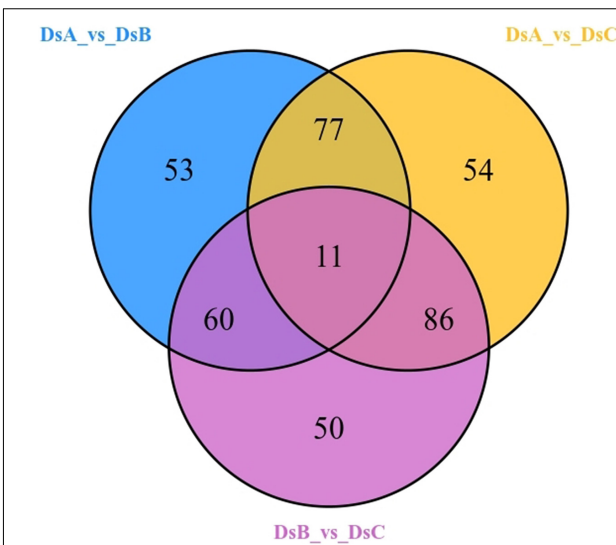


FIGURE 2 | Differential metabolites by Venn diagram.

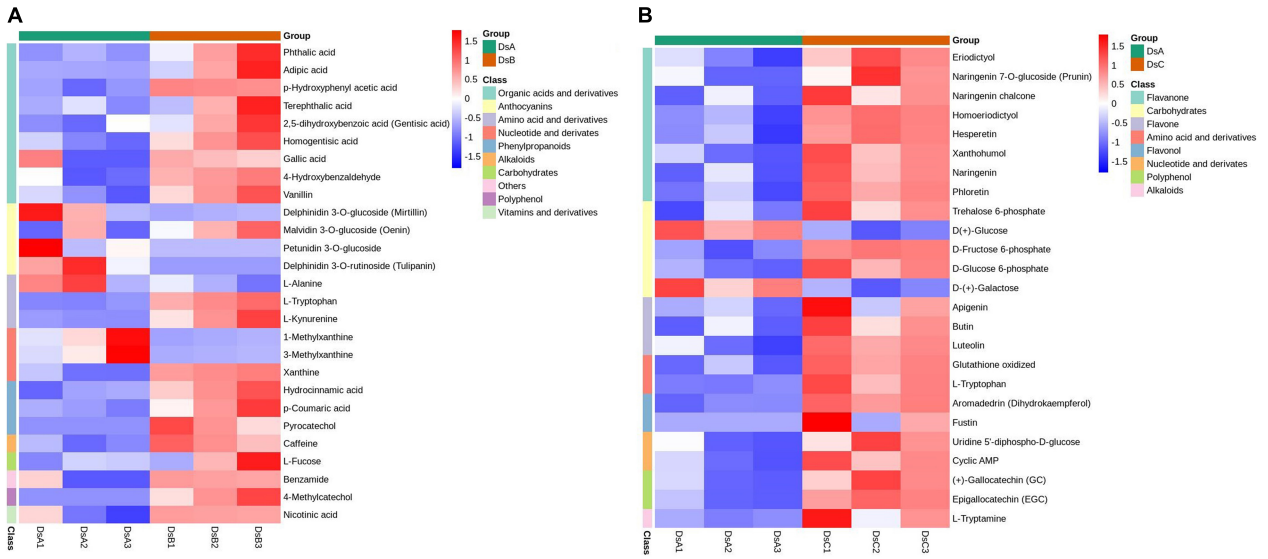


FIGURE 3 | Metabolites on metabolic pathways with significantly different enrichment. **(A)** Cluster analysis of DSB vs. DSA differential metabolites. **(B)** Cluster analysis of DSC vs. DSA differential metabolites. The analysis of hierarchical clustering for differential metabolism calculated by complete linkage approach. These metabolites were derived from metabolic pathways ($p < 0.05$) in DSA vs. DSB and DSA vs. DSC. Red denotes metabolites with high content, and blue means low content. The agglomeration method to be used by hclust function.

Transcriptome Sequencing of *D. sinense*

To obtain a more reliable transcriptome library, different treatments and organs of *D. sinense* were prepared for third-generation sequencing. 18.41G polymerase read bases (504,082 polymerase read) were obtained on the PacBio Sequel platform. Additionally, 84.47G clean bases were obtained on an Illumina platform (**Supplementary Table 4**). A total of 72,797 unigenes with a mean length of 2,486 bp were obtained (**Table 1**). We found that 62,038 (85.02%) unigenes had been previously annotated at least in one database (**Supplementary Figure 3**). Through GO analysis, we categorized these unigenes into 56 pathways, with "metabolic process," "cell," and "catalytic activity" having high ranking in biological processes, cellular components, and molecular functions, respectively (**Supplementary Figure 3**). A total of 38,037 unigenes were divided into 25 groups using the KOG database. "General function prediction only" was the most significant category, follow by "Posttranslational modification protein turnover chaperones" and "Signal transduction mechanisms" (**Supplementary Figure 3**).

Drought Inducible DEGs in *D. sinense*

We compared transcript counts across treatments to identify DEGs. Three hundred seventy-four DEGs were identified in DSA vs. DSB (151 upregulated and 223 downregulated genes), 130 DEGs in DSA vs. DSC (53 upregulated and 77 downregulated genes), and 256 DEGs in DSB vs. DSC (129 upregulated and 127 downregulated genes) (**Supplementary Figure 4**). Statistically, a total of 622 DEGs were characterized in differently treated *D. sinense* (**Supplementary Figure 4**). Surprisingly, DEGs were not found to be shared among different comparison groups (**Supplementary Figure 4**). To further explore the expression patterns of DEGs in response to drought stress, 622 DEGs were grouped into six clusters by k-means (**Figure 4**). Compared with control (DSA), the DEGs from class 1 showed a positive response, while those of class 6 showed a negative response (**Figure 4**). In DSC, the class 2 DEGS were upregulated, and the class 4 DEGs were downregulated (**Figure 4**). On the other hand, in DSB, the class 3 DEGs were upregulated, but the class 5 DEGs were downregulated (**Figure 4**). In the GO analysis, the functions of class 1 were related to "carbohydrate derivative binding," "adenyl nucleotide binding," "ATPase activity," and "nucleotide binding" (**Supplementary Table 5**). "Transcription factor activity," "oxidoreductase activity," and "pectin metabolic process" were identified from class 2 of DEGs (**Supplementary Table 5**).

Identification of Transcription Factors in Candidate Genes

Transcription factors (TFs) are involved in regulating the biosynthesis of secondary metabolites in plants. According to the plant transcription family database, 3,284 TFs were identified in our data and distributed across 89 families using iTAK software (**Figure 5A**). Importantly, 27 differentially expressed TFs, classified into 21 families, were identified (**Figure 5B**). When combined with the analysis of expression trends of DEGs, the TFs were distributed across different classes. Class 3 contained the most differentially expressed TFs (ARID4, BIM2, C3H50, RR22, MYB-like, BT1, and BPM3). Class 5 contained the second most differentially expressed TFs (IDD4, C3H17, FRS6, HMG-like, PIE1, and pyrH). Class 6 consisted of five differentially expressed TFs (BBRD, ABI5, GATA26, BLH1, and NLP3). The ABI5 TF is involved in ABA signal transduction through a negative

TABLE 1 | Summary of non-redundant transcriptional sequences.

Polymerase read bases (G)	Polymerase reads	CCS number	Mean length	Consensus number	Mean length	Unigene number	Mean length
18.41	504,082	424,187	1,955	140,872	2,244	72,969	2,486

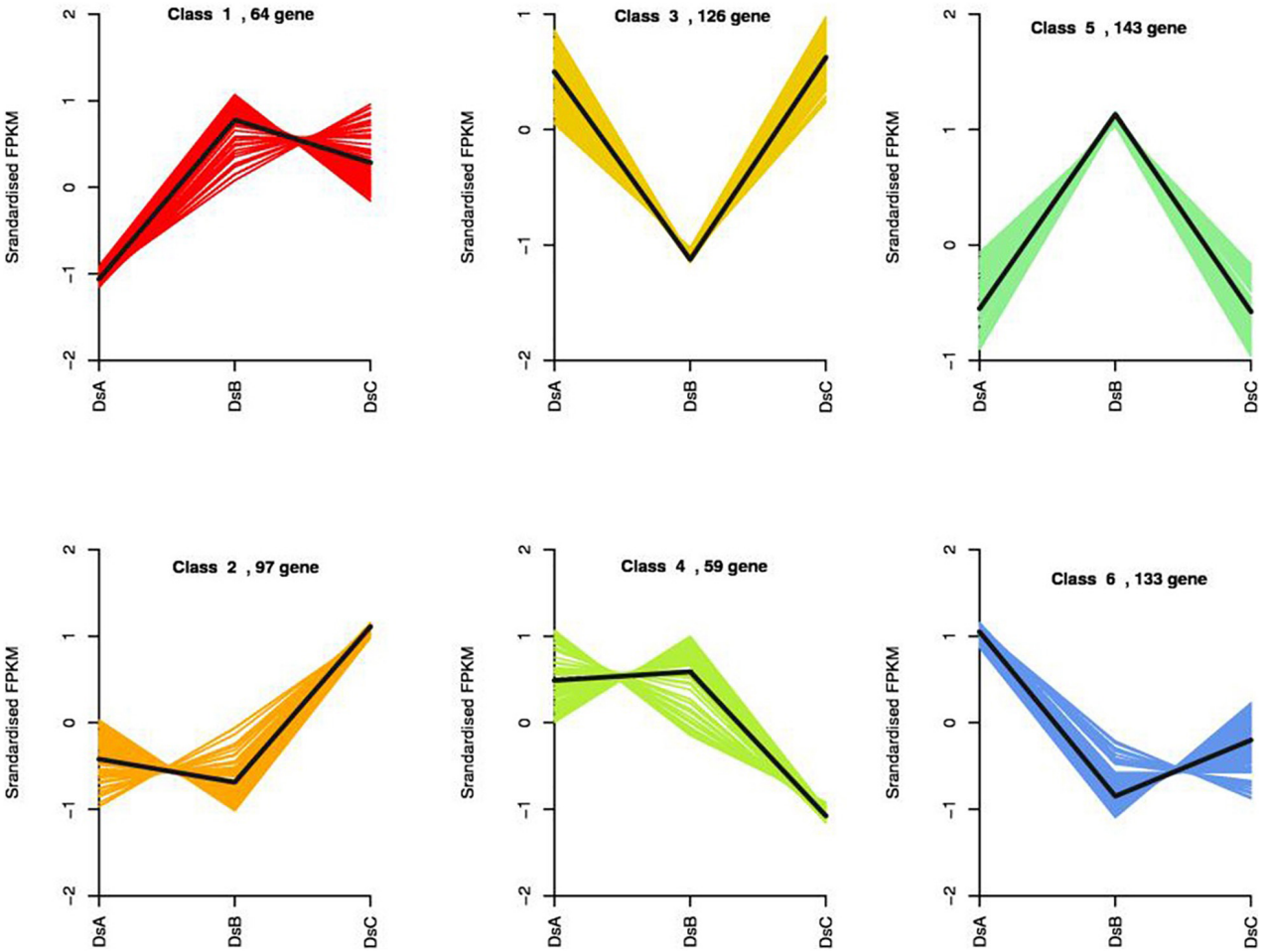


FIGURE 4 | Results of the k-means clustering of 622 DEGs identified by PacBio Iso-Seq based on their expression patterns.

regulatory relationship. Classes 1 and 4 have the same number of differentially expressed TFs: FRS3, ATXR7, and PIE1 from class 1 and FRS6, NAC017, and CIA2 from class 4. Additionally, the TFs of PRR95 and ALY3 were upregulated after extreme drought stress (RH $\leq$ 5%). These results suggested that TFs may regulate the metabolite biosynthesis pathway of *D. sinense* in response to drought stress.

Coexpression Networks of *D. sinense* in Response to Drought Stress

We then explored correlations between the DEG and metabolites by calculating the Pearson correlation coefficients between all DEGs and differential metabolites. One hundred twenty-one DEGs and 239 differential metabolites had a strong correlation (R^2 > 0.8, FDR < 0.05) (**Supplementary Table 6**). We then screened for extremely significantly DEGs (p < 0.01), which resulted in a set of DEGs that were enriched for genes relating to redox process, metabolic process, and phospholipid biosynthesis process (**Supplementary Table 7**). Nine extremely significantly DEGs were involved in microtubule organizing center and nuclear cell components, and 24 extremely significantly DEGs related to molecular functions were involved in molecular processes such as 6-phosphofructokinase activity, methionine adenosyltransferase, and GTP binding (**Supplementary Table 7**).

We also constructed a coexpression network of 33 extremely significantly DEGs and 11 differential metabolites (**Figure 6**). The result showed that extremely significant DEGs may play an important role in the response of *D. sinense* to drought stress. Further analysis on the relationship between genes and metabolites found five metabolic pathways in the coexpression network, such as glycerophospholipid

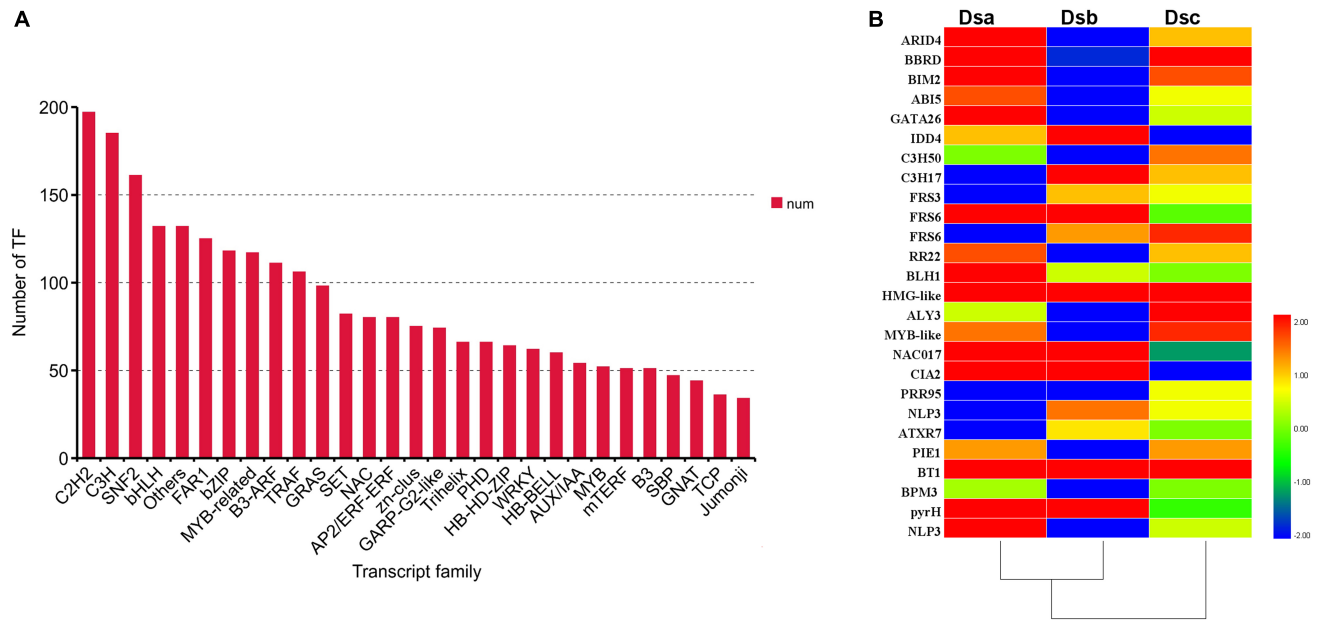


FIGURE 5 | TFs involved in response to drought stress. **(A)** Number of TFs in different families. **(B)** Heat map presenting the expression patterns of different TFs in response to drought treatments. Red denotes TFs with high expression levels, and blue means low expression. The color ranges from red to blue represents the $\log_2$ (FPKM + 0.01) value from large too small.

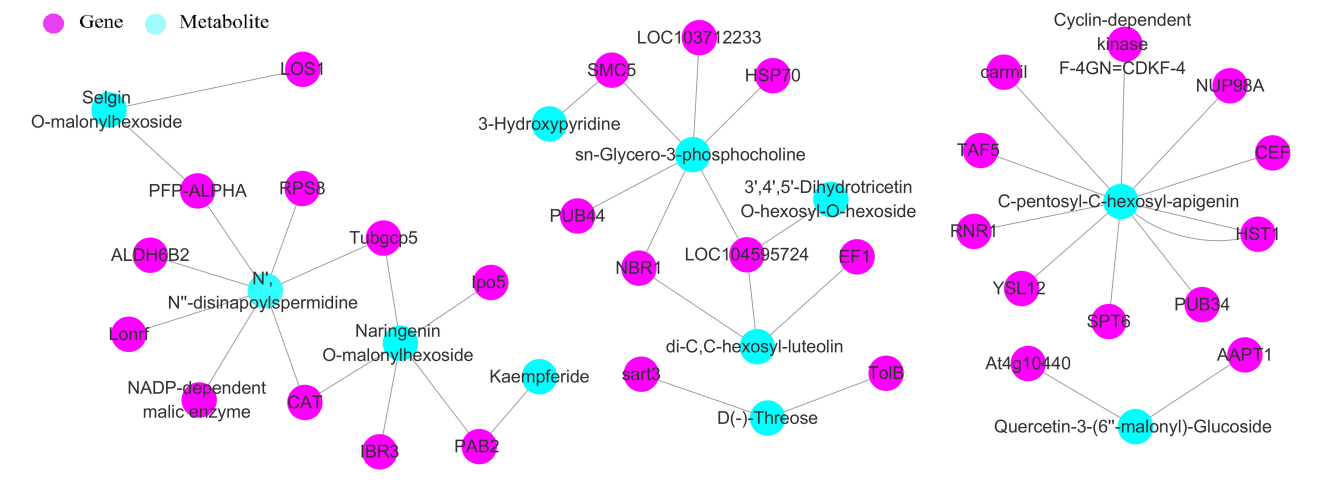


FIGURE 6 | Coexpression network of DEGs and differentially expressed metabolites. These DEGs were correlated with metabolites ($R > 0.8$). The networks were visualized with the Cytoscape software. The purple denotes represent DEGs, and blue means differential metabolites.

metabolism pathway (ko0564), arginine and proline metabolism pathway (ko0330), purine metabolism pathway (ko0230), tryptophan metabolism pathway (ko0380), and phenylpropanoid biosynthesis pathway (ko0940).

Purine Metabolism and Phenylpropanol Biosynthesis of *D. sinense* in Response to Drought

By combining the correlation analysis of transcriptome and metabolome in *D. sinense*, we found that purine metabolism and phenylpropanol biosynthesis were the main metabolic pathways in response to drought stress. The analysis found 7 DEGs and 39 differential metabolites from the purine metabolic pathway, such as *phosphatidylglycamine synthase* (*purL*), *adenylate kinase* (*AK*), and *ribonucleoside diphosphate reductase subunit M1* (*RRM1*). In addition, the expression of *inosine-5′-monophosphate dehydrogenase* (*IMPDH*) gene was upregulated specifically in DSB (the moderate drought conditions) (**Figure 7**). Compared with DSA, the metabolite accumulation of 3,4-dihydroxy-DL-phenylalanine in DSB and DSC was downregulated, and the contents of guanine, xanthine,

and hypoxanthine were upregulated. For example, xanthines accumulated significantly in DSB (DSB > DSA > DSC). Therefore, the biomass accumulation of xanthine, hypoxanthine, and guanine was consistent with the expression trend of *IMPDH* gene under drought stress. In addition, five *PAL*, four *4CL*, seven *CAD*, two *CCR*, and two *cytochrome P450* genes were found in the phenylpropanoid biosynthesis pathway. Compared with DSA, these genes did not show significant differences in DSB and DSC. However, the *CAD* gene was downregulated in DSB vs. DSC. Surprisingly, the phenylpropanol biosynthetic pathway of downstream metabolites [coumaric acid, conibenol, (E)-p-coumaric acid, and 4-hydroxy-3-methoxy cinnamaldehyde] was upregulated compared with DSA. In addition, the glycolysis pathway was also involved in response to drought stress (**Figure 7**). These results indicate that *D. sinense* improves its ability to resist drought stress by regulating the metabolism of genetic material and promoting the accumulation of antioxidant products and cytoplasmic material under drought stress.

DISCUSSION

Third-Generation Sequencing Reveals the Transcriptome Complexity in *D. sinense*

Compared to genomic DNA-seq, RNA-seq is often faster, simpler, and more scalable and can reveal important information about gene transcription and expression (Trapnell et al., 2010; José and Marta, 2014). Extensive transcriptomic studies have shown that RNA-seq technology is a useful tool for studying adaption of plants to different environmental stress (Brenner and Schmülling, 2012; José and Marta, 2014; Begara-Morales et al., 2014).

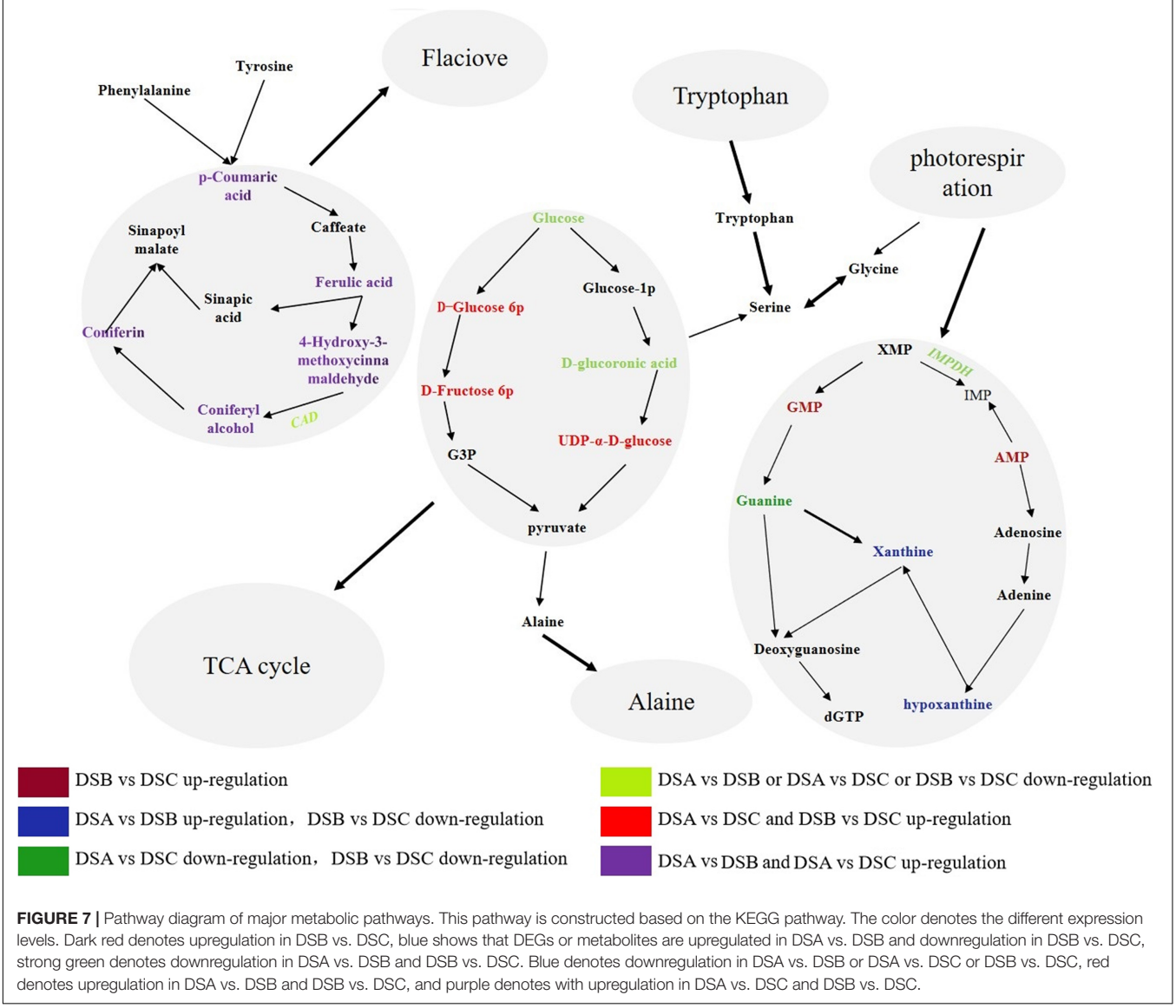


FIGURE 7 | Pathway diagram of major metabolic pathways. This pathway is constructed based on the KEGG pathway. The color denotes the different expression levels. Dark red denotes upregulation in DSB vs. DSC, blue shows that DEGs or metabolites are upregulated in DSA vs. DSB and downregulation in DSB vs. DSC, strong green denotes downregulation in DSA vs. DSB and DSB vs. DSC. Blue denotes downregulation in DSA vs. DSB or DSA vs. DSC or DSB vs. DSC, red denotes upregulation in DSA vs. DSB and DSB vs. DSC, and purple denotes with upregulation in DSA vs. DSC and DSB vs. DSC.

However, many studies had found that second-generation RNA-seq technology is imperfect, and short reads can reduce sequence and assembly accuracy (Haas and Zody, 2010; Grabherr et al., 2013). With the development of newer sequencing technologies, full-length transcript sequences can be obtained using the PacBio Iso-Seq technology. This technology solves the inherent limitations of Illumina high-throughput sequencing and identifies new gene subtypes with alternatively spliced transcript variants (Roberts et al., 2013; Du and Sun, 2016; An et al., 2018). Previous Illumina-based RNA-seq studies in orchids found the average transcript length to be 764.3 bp in *Dendrobium huoshanense* (Yuan et al., 2018) and 505 bp in *Dendrobium officinale* (He et al., 2017). In this present work, the full-length transcripts had an average length of 2,244 bp, similar to previous results of *Dendrobium candidum* (Xu et al., 2017). These results support the viewpoint that the third-generation sequencing technique can obtain longer transcripts than Illumina-based RNA-seq. To obtain a complete and accurate transcriptome library, short-reads generated from Illumina sequencing were used to correct long reads from third-generation sequencing. Previous studies have used this strategy to generate high-throughput sequencing data, providing reference transcriptomes for plant species that lack genome sequence information (Sharon et al., 2013; Grabherr et al., 2013; Huddleston et al., 2014). To our knowledge, this is the first report of transcriptome sequencing using RNA-seq and PacBio SMRT in *D. sinense*. This full-length reference transcriptome will be helpful to future genetic studies of *D. sinense*.

Metabolic Responses to Drought Stress in *D. sinense*

It is well known that multiple adaptative mechanisms have evolved in plants to reduce the effects of stress. Metabolites appear to be the key to these adaptations, with many studies revealing significant differences in the accumulation of metabolites under drought stress (Georgii et al., 2017; Navarro-Cerrillo et al., 2019). In this study, flavonoids, amino acids and derivatives, organic acids and derivatives, and phenylpropanoids were identified as accumulating at significantly different rates across drought conditions, showing upregulation under drought stress in both DSA vs. DSB and DSA vs. DSC. These results are similar to previous studies that also found that the content of flavonoids, amino acids, and other metabolites were increased in drought stress (Georgii et al., 2017; Navarro-Cerrillo et al., 2019). A previous study in *Dendrobium moniliforme* also indicated that the accumulation of flavonoids and alkaloid was induced by drought stress (Wu et al., 2016). Physiologically, these secondary metabolites in plants can prevent the accumulation of reactive oxygen species from damaging proteins and cell membranes, thereby improving drought resistance (Mutwakil et al., 2017; Šircelj et al., 2017). Specifically, organic acids and small molecule compounds may be involved in carbon metabolism (López-Bucio et al., 2000). From a physiological and biochemical perspective, the increase in flavonoid content in the pseudobulbs of *D. sinense* appears to be related to adaptability to water (Zhu, 2010; Šircelj et al., 2017). The extremely significant accumulations of amino acids, organic acids, and derivatives in drought-stressed *D. sinense* further suggest that these are important metabolites in response to drought stress (López-Bucio et al., 2000). In addition, only a few secondary metabolites showed a strong response to drought stress, such as four phenylpropanoids, two flavone, one polyphenol, and one parthenolide. Importantly, acyl syringone, 4-methylcatechol, catechol, apigenin6-C-hexosyl-8-C-hexosyl-O-hexoside, parthenolide, magnolol, and pyrocatechol may be key molecular components of drought stress response in *D. sinense*. Importantly, strong drought tolerance in *D. sinense* may in fact be related to its high flavonoid content in the body. This strongly suggests that drought stress induces the accumulation of metabolites that may improve drought tolerance of *D. sinense*.

Functional Clustering of DEGs and Metabolites Response to Drought Stress in *D. sinense*

Previous work has shown that plants adapt to drought stress by responding to the accumulation of metabolites and altering gene expression (Georgii et al., 2017; Acevedo et al., 2019). Although many drought-resistant genes have been found in model plants, few investigations have been done on *Dendrobium*. Our study presents a systematic analysis of drought-resistance genes, shedding light on the mechanisms of drought response in *Dendrobium*. In *D. sinense*, 622 DEGs were identified as potentially involved in drought stress. Through GO and KEGG enrichment analyses, we found that these DEGs were significantly enriched in ATPase activity, hydrolase activity, acting on acid anhydrides, pyrophosphatase activity, ion binding, and RNA helicase activity. Moreover, many DEGs encoding protein kinases, lipoxygenase, phospholipase, antioxidant enzymes, transporters, ubiquitin-protein ligase, and TFs were identified, such as *FBA1*, *CIPK32*, *PBL27*, *LOX2*, *PLC4*, *GAD*, *ABCF1*, *KEG*, *GATA26*, and *C3H*. Notably, *LOX2* and *KEG* are involved in the biosynthesis and signal transduction of plant hormones (Wasternack, 2007; Wasternack and Hause, 2013, 2019). *ABI5* and *EGS* are related to ABA biosynthesis (Joshi et al., 2016; Zhu, 2016). These results clearly suggest that hormone signaling may play an important role in the response of *D. sinense* to drought stress.

Underlying the Responses of Phenylpropanoid Biosynthesis and Purine Metabolism in *D. sinense* to Drought Stress

Studies in model plants have revealed that nucleotide biosynthesis and degradation are two of the metabolic pathways involved in drought response in *Arabidopsis* (Watanabe et al., 2014) and rice (Boo and Jung, 1998). Recent work indicated that the nucleotide metabolism pathway was an important pathway for *D. wangliangii* to adapt to drought stress (Zhao et al., 2019). The *IMPHD* gene was reported as the rate-limiting enzyme of guanine nucleoside biosynthesis, directly affecting the biosynthesis of guanine, xanthine, and hypoxanthine (Mercati et al., 2019). In DSA vs. DSB, the *IMPHD* gene

was upregulated, and the content of guanine, xanthine, and hypoxanthine increased. In DSB vs. DSC, the expression level of *IMPHD* gene was downregulated and the metabolites guanine, xanthine, and hypoxanthine were also downregulated. These data indicated that drought stress affected the biosynthesis of purines pathway.

Previous work showed that when guanine is deficient, guanine is replaced by xanthine and hypoxanthine in RNA and DNA synthesis, leading to disorders of the metabolic system (Zoref-Shani et al., 2010). Our results showed that guanine biosynthesis was promoted in response to drought stress in *D. sinense*, while xanthine and hypoxanthine bioaccumulation was reduced under severe drought stress. This allowed for the maintenance of normal metabolic processes to reduce stress-induced damage. In addition, previous studies have reported that nucleotide metabolism may increase drought tolerance in *Hylocereus undatus* (Fan et al., 2014) and *Triticum boeoticum* (Liu et al., 2015). This suggested that purine metabolism may increased the drought tolerance of *D. sinense*. This result therefore supports findings from previous studies and provides a new insight to explore drought tolerance of epiphytes.

PAL, *C4H*, and *4CL* are key genes in the early stages of lignin and flavonoid biosynthesis (Martens and Mithöfer, 2005). In *Fagopyrum esculentum* Moench, lignin increases cell wall thickness to prevent water loss and improve drought resistance (Hou et al., 2019). In our study, we found that coumaric acid, coniferol, (E)-p-coumaric acid, and 4-hydroxy-3-methoxy cinnamaldehyde were upregulated in DSB vs. DSC, and five genes related to lignin were also identified. For these genes, only the *CAD* gene was downregulated in DSA vs. DSB, and no other genes were significantly different in DSB vs. DSC. This negative correlation between lignin biosynthesis and gene expression under drought stress is consistent with results from previous studies (Hu, 2008; Liu, 2019). Reports have also suggested that the root tissue thickness of *D. sinense* roots increased under drought stress (Wang, 2017). Our results suggest that when *D. sinense* experience drought stress, lignin accumulation is promoted, thereby increasing cell wall thickness to prevent water loss. In addition, most flavonoids were significantly upregulated in DSB vs. DSC even though we did not identify DEGs associated with flavonoid biosynthesis. However, previous research has shown that genes related to flavonoid biosynthesis were upregulated in buckwheat under drought stress (Hou et al., 2019). It is also suspected that flavonoids have antioxidant properties (Sarker and Oba, 2018). In light of this, our results suggest that *D. sinense* responds to water deficit in the air by increasing its antioxidant capacity, but the pathway for regulating flavonoid biosynthesis under drought stress needs further exploration.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI BioProject Accession PRJNA723915.

AUTHOR CONTRIBUTIONS

XS, JN, and CZ conceived the project and designed the experiments. CZ, WH, and JC collected the datasets. CZ and JC plotted the figures. CZ, JC, XS, and JN wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Innovative Research Team Program of Hainan Natural Science Fund (2018CXTD331) and Hainan University [KYQD(ZR)1701].

ACKNOWLEDGMENTS

We would like to thank Stephen Gaughran at Princeton University for his assistance with English language and grammatical editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.692702/full#supplementary-material

Supplementary Figure 1 | The phenotypes of *D. sinense* under drought stress. **(A,D)** were growth in RH ≥ 95%, **(B,E)** were growth in 45% ≤ RH ≤ 50%, **(C,F)** were growth in RH ≤ 5%

Supplementary Figure 2 | Different metabolites up- or down-regulated in different groups. **(A)** DSA vs. DSB, **(B)** DSA vs. DSC, **(C)** DSB vs. DSC.

Supplementary Figure 3 | Functional prediction and annotation of transcriptome genes in *D. sinense. (A) D. sinense* GO enrichment function analysis. The abscissa is the GO term of the next level in the three major categories of GO, and the ordinate is the number of genes annotated to the term (including its sub-term). The three different classifications represent the three basic classifications of GO term from left to right (i.e., biological processes, cell components, and molecular functions). **(B)** *D. sinense* gene function prediction. The abscis-axis is the names of the 26 groups of KOG, and the ordinate is the number of genes annotated to the group. **(C)** *D. sinense* gene functional annotation Venn diagram. The sum of the numbers in each large circle represents the number of transcripts for the database annotation, and the overlapping parts of the circles represent the transcript annotation results that are common between databases.

Supplementary Figure 4 | Number and distribution of up-regulated and down-regulated genes. **(A)** The distribution of DEGs in different sample, **(B)** Summary of *D. sinense* DEGs under drought stress.

Supplementary Table 1 | Detection of all metabolic quantities and species.

Supplementary Table 2 | Significantly different metabolites between stressed and control samples.

Supplementary Table 3 | Significantly enriched KEGG pathways.

Supplementary Table 4 | Summary of RNA-seq for DSA, DSB, and DSC.

Supplementary Table 5 | GO enrichment result in DEGs.

Supplementary Table 6 | Differential genes and differential metabolites with a correlation coefficient $R^2 \geq 0.8$.

Supplementary Table 7 | Extremely significantly DEGs GO function annotation.

REFERENCES

Acevedo, R. M., Avicol, E. H., Gonzalez, S., Salvador, A. R., Rivarola, M., Paniego, N., et al. (2019). Transcript and metabolic adjustments triggered by drought in *Ilex paraguariensis* leaves. *Planta* 250, 445–462. doi: 10.1007/s00425-019-03178-3

An, D., Cao, H., Li, C., Humbeck, K., and Wang, W. (2018). Isoform sequencing and state-of-art applications for unravelling complexity of plant transcriptomes. *Genes* 9:43. doi: 10.3390/genes9010043

Baena-González, E., Rolland, F., Thevelein, J. M., and Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signaling. *Nature* 448, 938–942. doi: 10.1038/nature06069

Bang, S. W., Lee, D. K., Jung, H., Chung, P. J., Kim, Y. S., Choi, Y. D., et al. (2018). Overexpression of *OsTF1L*, a rice HD-Zip transcription factor, promotes lignin biosynthesis and stomatal closure that improves drought tolerance. *Plant Biotechnol. J.* 17, 118–131. doi: 10.1111/pbi.12951

Begara-Morales, J. C., Sanchez-Calvo, B., Luque, F., Leyva-Perez, M. O., Leterrier, M., Corpas, F. J., et al. (2014). Differential transcriptomic analysis by RNA-Seq of GSNO-responsive genes between *Arabidopsis* roots and leaves. *Plant Cell Physiol.* 55, 1080–1095. doi: 10.1093/pcp/pcu044

Boo, Y. C., and Jung, J. (1998). Alteration in pyridine nucleotide status in cells as an adaptive response to water stress in Rice (Oryza sativa L.) seedlings. *Appl. Biol. Chem.* 41, 228–234.

Brenner, W. G., and Schmülling, T. (2012). Transcript profiling of cytokinin action in *Arabidopsis* roots and shoots discovers largely similar but also organ-specific responses. *BMC Plant Biol.* 12:112. doi: 10.1186/1471-2229-12-112

Chen, B. X., Peng, Y. X., Gao, J. D., Zhang, Q., Liu, Q. J., Fu, H., et al. (2019). Coumarin-induced delay of rice seed germination is mediated by suppression of abscisic acid catabolism and reactive oxygen species production. *Front. Plant Sci.* 10:828. doi: 10.3389/fpls.2019.00828

Chen, W., Gong, L., Guo, Z., Wang, W. S., Zhang, H. Y., Liu, X. Q., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

Chong, J., and Xia, J. G. (2018). MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics* 34, 4313–4314. doi: 10.1093/bioinformatics/bty528

Claeys, H., and Inze, D. (2013). The agony of choice: how plants balance growth and survival under water-limiting conditions. *Plant Physiol.* 162, 1768–1779. doi: 10.1104/pp.113.220921

Cui, X. Y., Du, Y. T., Fu, J. D., Yu, T. F., Wang, C. T., Chen, M., et al. (2018). Wheat CBL-interacting protein kinase 23 positively regulates drought stress and ABA responses. *BMC Plant Biol.* 18:93. doi: 10.1186/s12870-018-1306-5

Domenico, S. D., Bonsegna, S., Horres, R., Pastor, V., Taurino, M., Poltronieri, P., et al. (2012). Transcriptomic analysis of oxylipin biosynthesis genes and chemical profiling reveal an early induction of jasmonates in chickpea roots under drought stress. *Plant Physiol. Biochem.* 61, 115–122. doi: 10.1016/j.plaphy.2012.09.009

Du, N., and Sun, Y. (2016). Improve homology search sensitivity of PacBio data by correcting frameshifts. *Bioinformatics* 32, 529–537.

Du, X. L., Zhang, X. F., Nie, H., Liu, M. L., and Cheng, J. L. (2016). Transcript profiling analysis reveals crucial genes regulating main metabolism during adventitious root formation in cuttings of *Morus alba* L. *Plant Growth Regul.* 79, 251–262. doi: 10.1007/s10725-015-0130-2

Duan, Z., Zhang, D. Y., Zhang, J. Q., Di, H. Y., Wu, F., Hu, X. W., et al. (2015). Co-transforming bar and *CsALDH* genes enhanced resistance to herbicide and drought and salt stress in transgenic alfalfa (*Medicago sativa* L.). *Front. Plant Sci.* 6:1115. doi: 10.3389/fpls.2015.01115

Fan, J. Q., Yan, X. F., Qiao, G., Zhang, X. B., and Wen, P. X. (2014). Identification of differentially-expressed genes potentially implicated in drought response in pitaya (*Hylocereus undatus*) by suppression subtractive hybridization and cDNA microarray analysis. *Gene* 553, 322–331. doi: 10.1016/j.gene.2013.08.098

Fu, L. M., Niu, B. F., Zhu, Z. W., Wu, S. T., and Li, W. Z. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Georgii, E., Jin, M., Zhao, J., Kanawati, B., Schmitt-Kopplin, P., Albert, A., et al. (2017). Relationships between drought, heat and air humidity responses revealed by transcriptome-metabolome co-analysis. *BMC Plant Biol.* 17:120. doi: 10.1186/s12870-017-1062-y

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., and Amit, I. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652.

Haas, B. J., and Zody, M. C. (2010). Advancing RNA-Seq analysis. *Nat. Biotechnol.* 28, 421–423. doi: 10.1038/nbt0510-421

He, L., Fu, S., Xu, Z., Yan, J., Xu, J., Hong, Z., et al. (2017). Hybrid sequencing of full-length cDNA transcripts of stems and leaves in *Dendrobium officinale*. *Genes* 8:257. doi: 10.3390/genes8100257

Hou, Z. H., Yin, J. L., Lu, Y. F., Song, J. H., Wang, S. P., Wei, S. D., et al. (2019). Transcriptomic analysis reveals the temporal and spatial changes in physiological process and gene expression in common buckwheat (*Fagopyrum esculentum* Moench) grown under drought stress. *Agronomy* 9:569. doi: 10.3390/agronomy9100569

Hu, Y. (2008). *Differential Expression of Candidate Genes for Lignin Biosynthesis Under Drought Stress in Maize*. Ya'an: Sichuan Agricultural University.

Huddleston, J., Ranade, S., Malig, M., Antonacci, F., Chaisson, M., Hon, L., et al. (2014). Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* 24, 688–696. doi: 10.1101/gr.168450.113

José, G., and Marta, F. (2014). RNA-Seq applications for plant transcriptome analysis in response to cold acclimation by Ion Torrent technology. *Technol. Health Care* 12, 125–127.

Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., et al. (2016). Transcription factors and plants response to drought stress: current understanding and future directions. *Front. Plant Sci.* 7:1029. doi: 10.3389/fpls.2016.01029

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

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

Liu, H., Sultan, M. A. R. F., Liu, X. L., Zhang, J., Yu, F., and Zhao, H. X. (2015). Physiological and comparative proteomic analysis reveals different drought responses in roots and leaves of drought-tolerant wild wheat (*Triticum boeoticum*). *PLoS One* 10:e0121852. doi: 10.1371/journal.pone.0121852

Liu, W. (2019). *Functional Research of CmCADs in Lignin Biosynthesis in Oriental Melon (Cucumis melo L.) Under Abiotic Stresses*. Shenyang: Shenyang Agricultural University.

López-Bucio, J., Nieto-Jacobo, M. F., Ramirez-Rodriguez, V. V., and Herrera-Estrella, L. (2000). Organic acid metabolism in plants: from adaptive physiology to transgenic varieties for cultivation in extreme soils. *Plant Sci.* 160, 1–13. doi: 10.1016/s0168-9452(00)00347-2

Martens, S., and Mithöfer, A. (2005). Flavones and flavone synthases. *Phytochemistry* 66, 2399–2407. doi: 10.1016/j.phytochem.2005.07.013

Maruyama, K., Takeda, M., Kidokoro, S., Yamada, K., Sakuma, Y., Urano, K., et al. (2009). Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by *DREB1A* and *DREB2A*. *Plant Physiol.* 150, 1972–1980. doi: 10.1104/pp.109.135327

Maruyama, K., Urano, K., Yoshiwara, K., Morishita, Y., Sakurai, N., Suzuki, H., et al. (2014). Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts. *Plant Physiol.* 164, 1759–1771. doi: 10.1104/pp.113.231720

Mercati, F., Fontana, I., Gristina, A. S., Martorana, A., El Nagar, M., De Michele, R., et al. (2019). Transcriptome analysis and codominant markers development in caper, a drought tolerant orphan crop with medicinal value. *Sci. REP U. K.* 9:10411. doi: 10.1038/s41598-019-46613-x

Morimoto, K., Ohama, N., Kidokoro, S., Mizoi, J., Takahashi, J., Todaka, D., et al. (2017). BPM-CUL3 E3 ligase modulates thermotolerance by facilitating negative regulatory domain-mediated degradation of *DREB2A* in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8528–8536.

Mutwakil, M. Z., Hajrah, N. H., Atef, A., Edris, S., Sabir, M. J., Al-Ghamdi, A. K., et al. (2017). Transcriptomic and metabolic responses of *Calotropis procera* to salt and drought stress. *BMC Plant Biol.* 17:231. doi: 10.1186/s12870-017-1155-7

Navarro-Cerrillo, R. M., Sanchez-Salguero, R., Rodriguez, C., Lazo, J. D., Moreno-Rojas, J. M., Palacios-Rodriguez, G., et al. (2019). Is thinning an alternative when trees could die in response to drought? The case of planted *Pinus nigra*

and *P. Sylvestris* stands in southern Spain. *For. Ecol. Manag.* 433, 313–324. doi: 10.1016/j.foreco.2018.11.006

Roberts, R. J., Carneiro, M. O., and Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol.* 14:405.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Sarker, U., and Oba, S. (2018). Drought stress effects on growth, ROS markers, compatible solutes, phenolics, flavonoids, and antioxidant activity in *Amaranthus tricolor*. *Appl. Biochem. Biotech.* 186, 999–1016. doi: 10.1007/s12010-018-2784-5

Sharma, M., Banday, Z. Z., Shukla, B. N., and Laxmi, A. (2019). Glucose-regulated HLP1 acts as a key molecule in governing thermomemory. *Plant Physiol.* 180, 1081–1100. doi: 10.1104/pp.18.01371

Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* 31, 1009–1014. doi: 10.1038/nbt.2705

Singh, D., and Laxmi, A. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front. Plant Sci.* 6:895. doi: 10.3389/fpls.2015.00895

Šircelj, H., Tausz, M., Grill, D., and Batiča, F. (2017). Detecting different levels of drought stress in apple trees (*Malus domestica* Borkh.) with selected biochemical and physiological parameters. *Sci. Hortic.* 113, 362–369. doi: 10.1016/j.scienta.2007.04.012

Stenzel, I., Hause, B., Miersch, O., Kurz, T., Maucher, H., Weichert, H., et al. (2003). Jasmonate biosynthesis and the allene oxide cyclase family of *Arabidopsis thaliana*. *Plant Mol. Biol.* 51, 895–911.

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genome wide studies. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9440–9445.

Sun, M., Yang, S. J., Zhang, J. L., Bartlett, M., and Zhang, S. B. (2014). Correlated evolution in traits influencing leaf water balance in Dendrobium (Orchidaceae). *Plant Ecol.* 215, 1255–1267. doi: 10.1007/s11258-014-0383-2

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M. J. V., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Wang, X. M. (2017). *Mycorrhiza Biology of Dendrobium sinense (Orchidaceae), An Endemic Species in Hannan Island*. Hai Kou: Han Nan University.

Wasternack, C. (2007). Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* 100, 681–697. doi: 10.1093/aob/mcm079

Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. an update to the 2007 review in Annals of Botany. *Ann. Bot.* 111, 1021–1058. doi: 10.1093/aob/mct067

Wasternack, C., and Hause, B. (2019). The missing link in jasmonic acid biosynthesis. *Nat. Plants* 5, 776–777. doi: 10.1038/s41477-019-0492-y

Watanabe, S., Kounosu, Y., Shimada, H., and Sakamoto, A. (2014). Arabidopsis xanthine dehydrogenase mutants defective in purine degradation show a compromised protective response to drought and oxidative stress. Plant Biotechnol. 31, 173–178. doi: 10.5511/plantbiotechnology.14.0117a

Wu, X., Yuan, J., Luo, A., Chen, Y., and Fan, Y. (2016). Drought stress and re-watering increase secondary metabolites and enzyme activity in *Dendrobium moniliforme*. *Ind. Crops Prod.* 94, 385–393. doi: 10.1016/j.indcrop.2016.08.041

Xu, M., Liu, X., Wang, J. W., Teng, S. Y., Shi, J. Q., Li, Y. Y., et al. (2017). Transcriptome sequencing and development of novel genic SSR markers for *Dendrobium officinale*. *Mol. Breed.* 37:18.

Yang, F. S. (2009). *Conservation Biology on the Base of Endophytic Mycorrhizal Fungi of an Endangered Orchid Dendrobium sinense*. Matsubara: Han Nan University.

Yuan, Y. D., Yu, M. Y., Jia, Z. H., Song, X. E., Liang, Y. Q., and Zhang, J. C. (2018). Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC Genomics* 19:978. doi: 10.1186/s12864-018-5305-6

Zhang, W., Hu, H., and Zhang, S. B. (2016). Divergent adaptive strategies by two co-occurring epiphytic Orchids to water stress: escape or avoidance? *Front. Plant Sci.* 7:588. doi: 10.3389/fpls.2016.00588

Zhang, X. Y., Yang, Z. R., Li, Z., Zhang, F. L., and Hao, L. Z. (2019). *De novo* transcriptome assembly and co-expression network analysis of *Cynanchum thesioides*: identification of genes involved in resistance to drought stress. *Gene* 710, 375–386. doi: 10.1016/j.gene.2019.05.055

Zhao, D. K., Shi, Y. N., Senthilkumar, H. A., Qiao, Q., Wang, Q. X., Shen, Y., et al. (2019). Enriched networks 'nucleoside/nucleotide and ribonucleoside/ribonucleotide metabolic processes' and 'response to stimulus' potentially conferred to drought adaptation of the epiphytic orchid *Dendrobium wangliangii*. *Physiol. Mol. Biol. Plants* 25, 31–45. doi: 10.1007/s12298-018-0607-3

Zhou, S. M., Sun, X. D., Yin, S. H., Kong, X. Z., Zhou, S., Xu, Y., et al. (2014). The role of the F-box gene *TaFBA1* from wheat (*Triticum aestivum* L.) in drought tolerance. *Plant Physiol. Biochem.* 84, 213–223. doi: 10.1016/j.plaphy.2014.09.017

Zhu, C. C. (2010). *Effects of Environmental Factors on the Growth and Principal Medicinal Compositions of Ginkgo Biloba Seedlings*. Nanjing: Nanjing Forestry University.

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

Zoref-Shani, E., Bromberg, Y., Brosh, S., Sidi, Y., and Sperling, O. (2010). Characterization of the alterations in purine nucleotide metabolism in hypoxanthine-guanine phosphoribosyltransferase-deficient rat neuroma cell line. *J. Neurochem.* 61, 457–463. doi: 10.1111/j.1471-4159.1993.tb02146.x

Zotz, G., and Bader, M. Y. (2009). Epiphytic plants in a changing world-global: change effects on vascular and non-vascular epiphytes. *Prog. Bot.* 7, 147–170. doi: 10.1007/978-3-540-68421-3_7

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.

Copyright © 2021 Zhang, Chen, Huang, Song and Niu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 July 2021
doi: 10.3389/fgene.2021.698598

Integrative Identification of Crucial Genes Associated With Plant Hormone-Mediated Bud Dormancy in *Prunus mume*

Ping Li[1,2], Tangchun Zheng[1,2], Zhiyong Zhang[3], Weichao Liu[1,2], Like Qiu[1,2], Jia Wang[1,2], Tangren Cheng[1,2] and Qixiang Zhang[1,2]**

[1] Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding; National Engineering Research Center for Floriculture; Beijing Laboratory of Urban and Rural Ecological Environment; Engineering Research Center of Landscape Environment of Ministry of Education; Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, Beijing Forestry University, Beijing, China, [2] School of Landscape Architecture, Beijing Forestry University, Beijing, China, [3] Department of Hematology, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Yuehua Cui, Michigan State University, United States

Reviewed by:
Bahman Panahi, Agricultural Biotechnology Research Institute of Iran, Iran
Kai Shu, Northwestern Polytechnical University, China
Guiyan Yang, Northwest A&F University, China

***Correspondence:**
Tangchun Zheng
zhengtangchun@bjfu.edu.cn
Qixiang Zhang
zqxbjfu@126.com

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

Received: *21 April 2021*
Accepted: *07 June 2021*
Published: *06 July 2021*

Citation:
Li P, Zheng T, Zhang Z, Liu W, Qiu L, Wang J, Cheng T and Zhang Q (2021) Integrative Identification of Crucial Genes Associated With Plant Hormone-Mediated Bud Dormancy in Prunus mume. Front. Genet. 12:698598. doi: 10.3389/fgene.2021.698598

Prunus mume is an important ornamental woody plant with winter-flowering property, which is closely related to bud dormancy. Despite recent scientific headway in deciphering the mechanism of bud dormancy in *P. mume*, the overall picture of gene co-expression regulating *P. mume* bud dormancy is still unclear. Here a total of 23 modules were screened by weighted gene co-expression network analysis (WGCNA), of which 12 modules were significantly associated with heteroauxin, abscisic acid (ABA), and gibberellin (GA), including GA1, GA3, and GA4. The yellow module, which was positively correlated with the content of ABA and negatively correlated with the content of GA, was composed of 1,426 genes, among which 156 transcription factors (TFs) were annotated with transcriptional regulation function. An enrichment analysis revealed that these genes are related to the dormancy process and plant hormone signal transduction. Interestingly, the expression trends of *PmABF2* and *PmABF4* genes, the core members of ABA signal transduction, were positively correlated with *P. mume* bud dormancy. Additionally, the *PmSVP* gene had attracted lots of attention because of its co-expression, function enrichment, and expression level. *PmABF2*, *PmABF4*, and *PmSVP* were the genes with a high degree of expression in the co-expression network, which was upregulated by ABA treatment. Our results provide insights into the underlying molecular mechanism of plant hormone-regulated dormancy and screen the hub genes involved in bud dormancy in *P. mume*.

Keywords: *Prunus mume*, dormancy, co-expression network, ABA, expression and function

INTRODUCTION

Dormancy is a strategy in which higher plants survive under adverse conditions by suspending growth and development (Holdsworth et al., 2008). When the seasons change—for instance, from summer to autumn—plants will face environmental changes such as lower temperature, shorter days, and lower ratios of red (R) and far-red (FR) light (de Wit et al., 2013; Kegge et al., 2015). By integrating various environmental stimuli and *in vivo* signal responses to balance growth and

dormancy, plants show seasonal adaptability for long-term evolution. The two main signals that plants depend on to respond to seasonal changes were photoperiod and temperature, which play a key role in the growth–dormancy cycle of trees (Maurya and Bhalerao, 2017). Plenty of studies have reported that photoperiod and temperature play crucial but opposite roles in the induction and release of dormancy, but the thresholds and combined effects of these environmental factors remain to be determined. Studies on peach showed that a short photoperiod can induce bud dormancy under non-low-temperature conditions, whereas low temperature promotes bud dormancy under a long-photoperiod condition (Li S. et al., 2018). If the photoperiod is shorter than the critical threshold for plant growth (short day), the growth will stop and the terminal bud will eventually form, surrounding the shoot tip meristem, while short day will cause the bud to transit to dormancy (Kainer et al., 1991; Barrero et al., 2012).

In previous studies, the core components of ABA biosynthesis and signal transduction have been identified by molecular genetics and biochemical and pharmacological methods (Chen et al., 2020). ABA-mediated dormancy is usually associated with different degrees and types of seed dormancy, while bud dormancy is scarce (Shu et al., 2016; Chen et al., 2020; Zhou et al., 2020). Meanwhile, ABA-mediated bud dormancy has been explored more in woody plants than in herbaceous plants, such as poplar, grape, and peach (Holdsworth et al., 2008; Li and Dami, 2015; Li S. et al., 2018; Tylewicz et al., 2018). Light signals are closely related to plant hormones, some of which are involved in the induction or release of winter dormancy, thus regulating the adaptability of plants to the environment. Short day induced the expression of abscisic acid (ABA) receptors in hybrid poplar and increased the level of ABA in bud, suggesting that ABA responded to short-photoperiod-mediated plasmodesmata closure and further induced bud dormancy (Lau and Deng, 2010; Singh et al., 2019). At the same time, short day induced plasmodesmata D closure by enhancing the ABA response, prevented the transmission of growth signals, and maintained bud dormancy (Singh et al., 2019). The photoperiodic control of ABA-mediated dormancy reveals that *SHORT VEGETATIVE PHASE-LIKE* (*SVL*) induces *CALLOSE SYNTHASE* expression to further induce plasmodesmatal closure and negatively regulates the gibberellin (GA) pathway to promote dormancy (Singh et al., 2018, 2019). The ectopic expression of *SVP* affects the dormancy duration of cold-tolerant kiwifruit, has a minimal effect on the dormancy duration of cold-sensitive kiwifruit, and suggests a complementary role with ABA (Wu et al., 2017). In grape, *VvSnRK1* regulates the dormancy induction and regulation of metabolic changes through both ABA synthesis and activities (Parada et al., 2016). During leafy spurge dormancy, the content of ABA and PA decreased, while the level of heteroauxin (IAA) remained unchanged (Chao et al., 2017). Ethylene and ABA regulate the biosynthesis of ascorbic acid by regulating EIN3 and ABI4 (Bossi et al., 2009; An et al., 2010). The opposing roles of ABA and GA in dormancy result in a balanced control mechanism, and other plant hormone signal transduction and synthesis also contribute to this balance (Liu and Hou, 2018; Chen et al., 2020).

The gene expression profiles are effectively and comprehensively described using high-throughput transcriptome sequencing technologies. The RNA-seq technology has been widely applied on a range of systems in many studies of plants, such as stress, growth, and development (Moenga et al., 2020; Swift et al., 2020; Sun et al., 2021). However, previous transcriptome studies have usually focused on the identification and screening of differentially expressed genes, while the degree of interconnection between related genes has not yet been considered. Since genes with similar expression patterns may have the same function, the identification of these genes can provide more information about their corresponding possible molecular regulatory mechanisms (Yuan et al., 2018; Farhadian et al., 2021). WGCNA can be used to construct the co-expression networks by gene expression profiles (Langfelder and Horvath, 2008; Presson et al., 2008). WGCNA has been reported to be used to investigate the co-expression network and hub genes involved mainly in plants under abiotic stress, such as salt, drought and cold stress (Cheng et al., 2020; Moenga et al., 2020; Panahi and Hejazi, 2021). More recently, the effectiveness of this approach for deciphering and disentangling the complex process has been proven, thereby highlighting the power of the co-expression networks to provide deep insights into these complex processes (Panahi et al., 2020; Farhadian et al., 2021; Panahi and Hejazi, 2021).

Prunus mume is a very important ornamental woody plant, providing quality material for ecotope-based landscape design, and its contribution to the public environment and urban landscaping continues to increase daily. *P. mume* is an excellent material for studying plant bud dormancy because of its winter-flowering property. In our previous study, *PmDAM* and *PmCBF* genes are involved in the molecular regulation of *P. mume* bud dormancy, while PmCBF1 transcription factor can form homodimer and heterodimer with PmCBF6 and heterodimer with three members of the PmDAM family (PmDAM1/2/6). PmCBFs are involved in the regulation of *PmDAM6* gene expression by binding to *cis*-acting elements (Xu et al., 2014; Zhao et al., 2018a,b,c). Based on the transcriptome profile that reveals the key roles of hormones and sugars, a hypothetical model was proposed to understand the molecular mechanisms of dormancy in *P. mume* (Zhang et al., 2018). Despite recent scientific headway in deciphering the mechanism of bud dormancy in *P. mume*, the overall picture of gene co-expression regulating *P. mume* bud dormancy is still unclear. In this study, we established the co-expression network in *P. mume* bud dormancy and aided in clarifying the mechanisms regulating dormancy in *P. mume*. Moreover, the genes identified here can serve as valuable genetic resources or selection targets for further molecular breeding of *P. mume*.

MATERIALS AND METHODS

Data Collection

To further investigate the correlation between gene expression and plant hormones, we collected transcriptome data and plant hormone content during *P. mume* bud dormancy. The dormancy

process was divided into four stages according to the percentage bud break. All samples from *P. mume* var. "Zaohua Lve," including the four stages of dormancy, were three biological replicates from independent individuals. The RNA-seq datasets of 12 samples were obtained using the Illumina HiSeq^{TM}4000 platform. Meanwhile, we collected the plant hormone content at these four stages from *P. mume* var. "Zaohua Lve" (Zhang et al., 2018). The plant hormonal quantification, including NAA, ABA, GA1, GA3 and GA4, was measured using HPLC–ESI–MS/MS with a standard measure. *P. mume* genome was obtained from the *P. mume* genome project (Zhang et al., 2012) for gene annotation.

Co-expression Network Construction

First, the samples and genes were filtered according to the gene expression profiles. Here we removed genes and samples with the absent rate greater than or equal to 10%. Co-expression network and module identification were analyzed for filtered dataset using the WGCNA package (Langfelder and Horvath, 2008). In order to maintain the gene connectivity and greater weight to strongest correlations, the soft threshold of the correlation matrix (β) was selected. The β value was calculated using the "pickSoftThreshold" function in the WGCNA package. The "blockwiseModules" function is one-step network construction and module detection. The adjacency matrix was converted to topological overlap matrix, and the topological overlap measure was calculated. The topological overlap matrix used dissimilarity between genes to cluster genes, and then it used dynamic tree cutting algorithm to shear the trees into different modules. Then, the co-expression network was constructed, and 400 genes were randomly selected to draw topologically overlapping heat maps.

Correlation Analysis of Gene Expression Level and Plant Hormone Content

The content of plant hormonal quantification was used as trait data. The correlation coefficients between each module eigengene and phenotype were calculated using the "cor" and "corPvalueStudent" functions. Module–trait associations were estimated based on the correlation between the module eigengene and the phenotype, and the associated heat map was drawn using the "labeledHeatmap" function. Modules with a correlation greater than or equal to 0.9 and a minimum *P*-value were selected in the module–trait associations.

Enrichment Analysis of Genes Within Network Modules

Genome-wide Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation files were constructed using *P. mume* genomes, respectively. Based on the GO database (Gene Ontology, 2021), the genes within the module were annotated through the GO online tool[1], including molecular function, cellular component, and biological process. Based on the KEGG database (Kanehisa et al., 2021), the genes within the module were enriched through the KEGG online tool[2]. R language was used for statistics and visualization of enrichment results.

Expression Trend Analysis and TF Enrichment

To investigate patterned differences in expression profiles among modules, genes within the module were identified for all possible expression trends at four stages. The enriched sequences were extracted using TBtool software (Chen et al., 2018). Meanwhile, the interaction network of enriched TFs was constructed according to the orthologs of *Arabidopsis* using STRING (Szklarczyk et al., 2017). Cytoscape software was used for visualization of the network (Smoot et al., 2011). The symbol and description of genes were annotated through UniProtKB and NCBI databases, respectively (Johnson et al., 2008; UniProt, 2019).

ABA Treatment and Relative Gene Expression Analysis

Prunus mume var. "Zaohua Lve" from grafted plants with the same genotype was cultivated in the greenhouse. For ABA treatment, the plants were sprayed with 50 μM ABA, and water was used as the control. The samples were collected 24 h after the treatment, frozen in liquid nitrogen, and stored at −80°C before use.

Total RNA extraction and cDNA synthesis were performed according to previous methods (Zhuo et al., 2018). We designed specific primers for qRT-PCR using INTEGRATED DNA TECHNOLOIES tool[3]. The detailed information of the primers are shown in **Supplementary Table 1**. The gene expression level was measured using the CFX96 Real-Time PCR Detection System with TB Green^{TM} Premix Ex *Taq*^{TM} II (TaKaRa, Bejing, China). The relative gene expression level was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

RESULTS

Construction of Co-expression Network and Correlation Between Modules and Hormone Content

To assess how the change of plant hormone content contributes to *P. mume* bud dormancy, WGCNA was applied to investigate gene sets that were related to plant hormone content using the RNA-seq during the dormancy period. After filtering, a total of 14,983 genes were selected based on their expression levels throughout the 16 samples. A scale-free network was constructed with the soft-threshold β = 12 based on the cutoff of R^2 = 0.9 (**Figure 1A** and **Supplementary Figure 1**). We randomly selected 400 genes to construct an interactive relationship network, which further verified the reliability of the module (**Supplementary Figure 2**). We identified 23 distinct co-expression modules using the dynamic tree cutting algorithm, each consisting of 35 to

[1]https://www.omicshare.com/tools/Home/Soft/gogseasenior

[2]https://www.omicshare.com/tools/Home/Soft/pathwaygseasenior

[3]https://sg.idtdna.com/scitools/Applications/RealTimePCR/

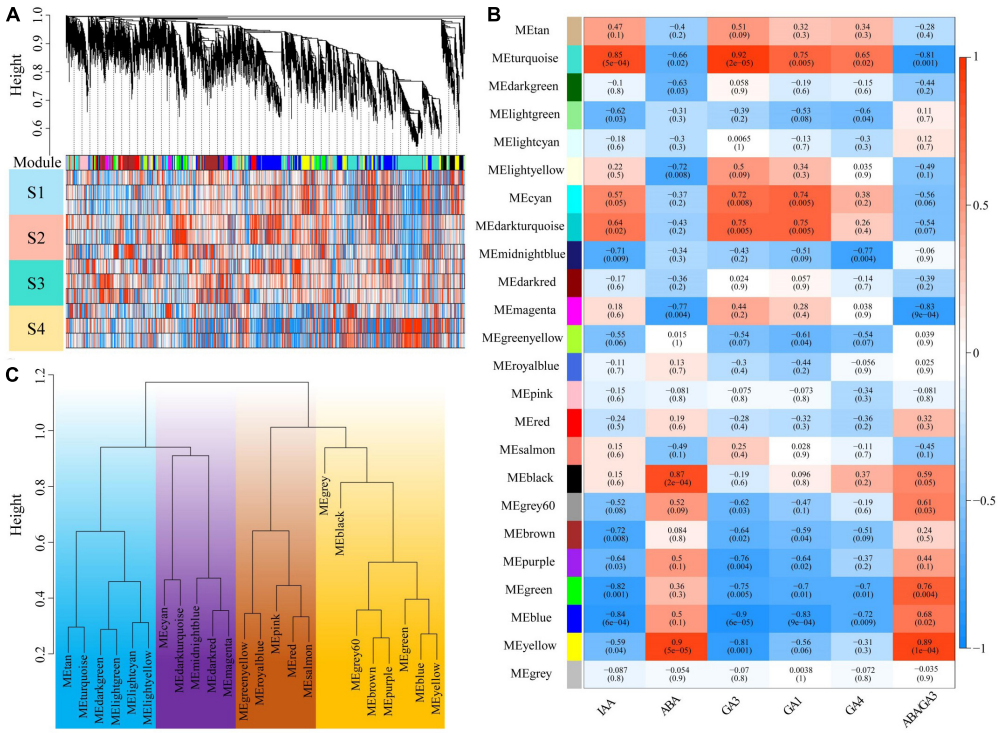


FIGURE 1 | Identification of key modules correlated with plant hormone content in the RNA-seq dataset through weighted gene co-expression network analysis. **(A)** The gene dendrogram and the corresponding module colors. The clustering was based on the RNA-seq data in four stages of dormancy. **(B)** A heat map showing the relationship between the modules and plant hormones. The corresponding correlation and *p*-value were presented in the first line and second line of each cell, respectively. **(C)** Clustering of module eigengenes. Clustering was distinguished by different colors.

3,825 correlated genes, of which 12 modules were significantly associated with IAA, ABA, GA1, GA3, and GA4. Most of the gene co-expression modules showed an inverse relationship between ABA and GA, but there were also some hormone specific co-expression gene modules—for example, the yellow module, including 1,426 genes, was most positively related to ABA (r = 0.9, P = 5e-5), whereas it was negatively correlated with the GA3 (r = −0.8, P = 0.001). Nevertheless, the black module was only significantly correlated with ABA (r = 0.87, P = 2e-4) (**Figure 1B**). The GA significantly correlated modules had a correlation relationship similar to that of the co-expression modules, in which GA3 had a stronger correlation among these modules than GA1 and GA4, suggesting that the biosynthetic genes and expression regulation genes of GA3 might play a more important role in *P. mume* bud dormancy. Furthermore, the eigengene dendrogram showed that these modules, including the gray module without the correlated genes, could be divided into four major branches (**Figure 1C**), indicating that co-expression patterns with varying functions were present in the networks.

Functional Enrichment of the Gene Module Analysis

We focused on the yellow and turquoise modules that were significantly associated with ABA and GA for functional enrichment analysis. In the yellow and turquoise modules, 111 and 199 GO terms were identified in three GO groups (cellular component, molecular function, and biological process), respectively (**Figure 2A** and **Supplementary Tables 2–7**). In three types of GO terms, more than half of the gene enrichment belonged to the biological process (**Figure 2A**), while genes from the yellow module did not have significantly enriched GO terms in the cellular component (**Figure 2B**). The molecular function of the selected genes in the yellow modules revealed an enrichment of the cell protein phosphorylation state, including phosphoprotein phosphatase activity, phosphoric ester hydrolase activity, and phosphatase activity, suggesting that these genes were involved in regulating cell activity. Meanwhile, the molecular function of the turquoise module was mostly enriched in catalytic activity, oxidoreductase activity, and nucleoside phosphate binding (**Figure 2C**). In biological process, the yellow and turquoise module presented obvious differences in enriched GO terms but also contained the same enriched GO terms, such as response to biotic stimulus (**Figure 2D**). The biological process of the yellow module was mostly enriched in nucleobase-containing compound biosynthetic process, organic substance biosynthetic process, and heterocycle biosynthetic process. Additionally, GO terms associated with plant dormancy were significantly enriched, including dormancy process and seed dormancy process (**Supplementary Tables 6**, **7**).

Furthermore, the pathway enrichment analysis of genes from the yellow and turquoise modules, respectively, was conducted

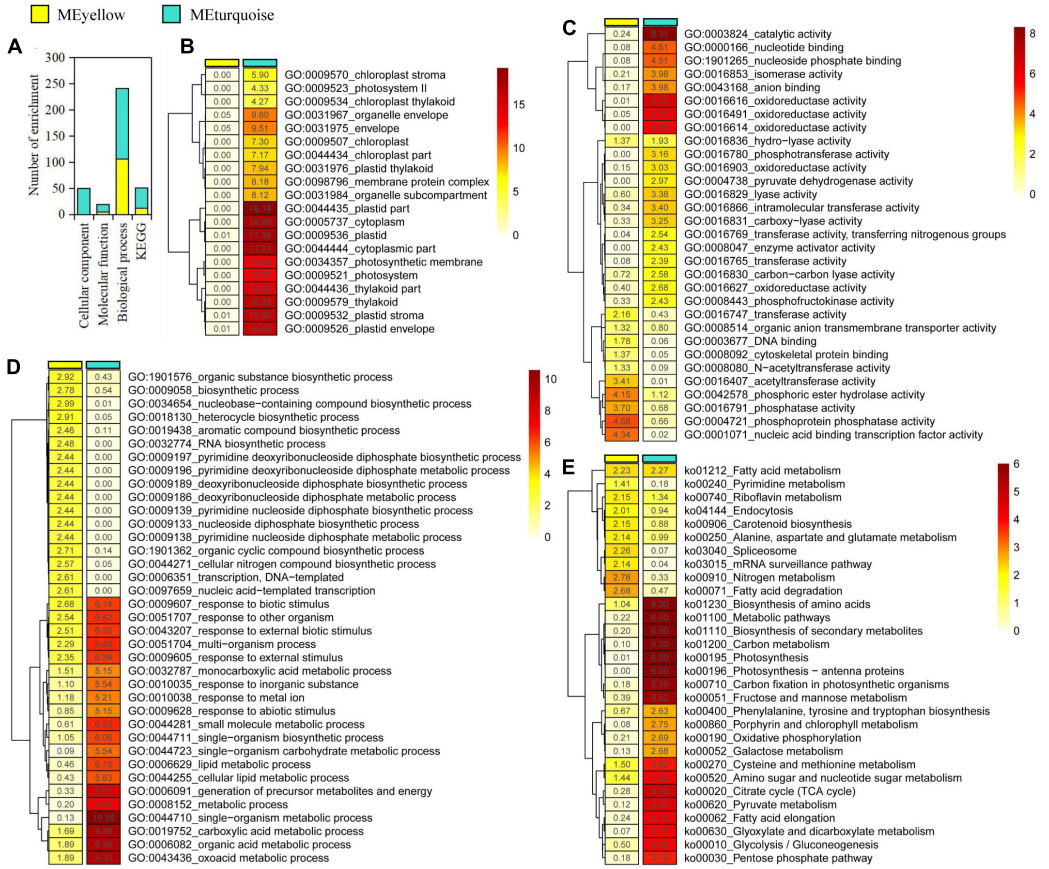


FIGURE 2 | Gene enrichment analysis in yellow and turquoise modules. **(A)** Statistics of significantly enriched genes. **(B–E)** Heat maps of -log10 enrichment *p*-values for the 20 most-enriched Gene Ontology terms and Kyoto Encyclopedia of Genes and Genomes pathway in the shared genes.

using KEGG database. As a result, we discovered a number of very important biosynthetic and metabolic pathways, such as carbon metabolism and fatty acid biosynthesis (**Figure 2E**). Notably, plant hormone signal transduction was enriched in the modules, which contained 43 genes in the yellow module and 107 genes in the turquoise module, respectively (**Supplementary Tables 8**, **9**). In plant hormone signal transduction, the enriched genes in the yellow module included abscisic acid receptor PYR/PYL family (*PYL*), protein phosphatase 2C (*PP2C*), and ABA-responsive element binding factor (*ABF*) genes involved in ABA signal transduction pathway. Meanwhile, the enriched genes in the turquoise module included gibberellin receptor GID1 (*GID1*), F-box protein GID2 (*GID2*), and DELLA protein (*DELLA*) genes, which are involved in GA signal transduction pathway. In addition, some genes that involved cytokinin, gibberellin, jasmonic acid, brassinolide, and auxin signal transduction pathways were identified in these enriched genes, suggesting that there were elaborate signaling networks and complex cross-talking among different hormone signal transduction pathways. Based on GO and KEGG enrichment results, we speculated that ABA- and GA3-related genes played important regulatory roles but presented different regulatory mechanisms in *P. mume* bud dormancy.

Expression Trend Analysis

To explore the possible expression pattern of the module-related genes in bud dormancy, expression trend analysis was performed to further divide the module-related genes into 26 profiles. Ten profiles (colored block) were divided into three clusters with clear and distinct expression profiles (**Figure 3A**). On the whole, the gene expression pattern between the yellow and turquoise modules showed an opposite trend. In the yellow module, 75.8% of the genes showed a downregulated trend, attributed to the red cluster, including profiles 0, 3, 9, and 12. However, 76.1% of the genes were upregulated in the turquoise module (**Figure 3B**). The expression trends of 136 genes, belonging to profiles 1 and 3, were specifically found in the yellow module. Similarly, there were specific trend profiles in the turquoise module, among which profile 13 contained most genes. Based on gene expression trends, we speculated that these downregulated genes from the yellow module might be actively involved in the regulation of *P. mume* bud dormancy. Therefore, as was also evident in our analysis of gene enrichment, the enrichment of functional differences in these two important modules was probably associated with expression pattern and represented plant-hormone-specific regulatory mechanisms in dormancy.

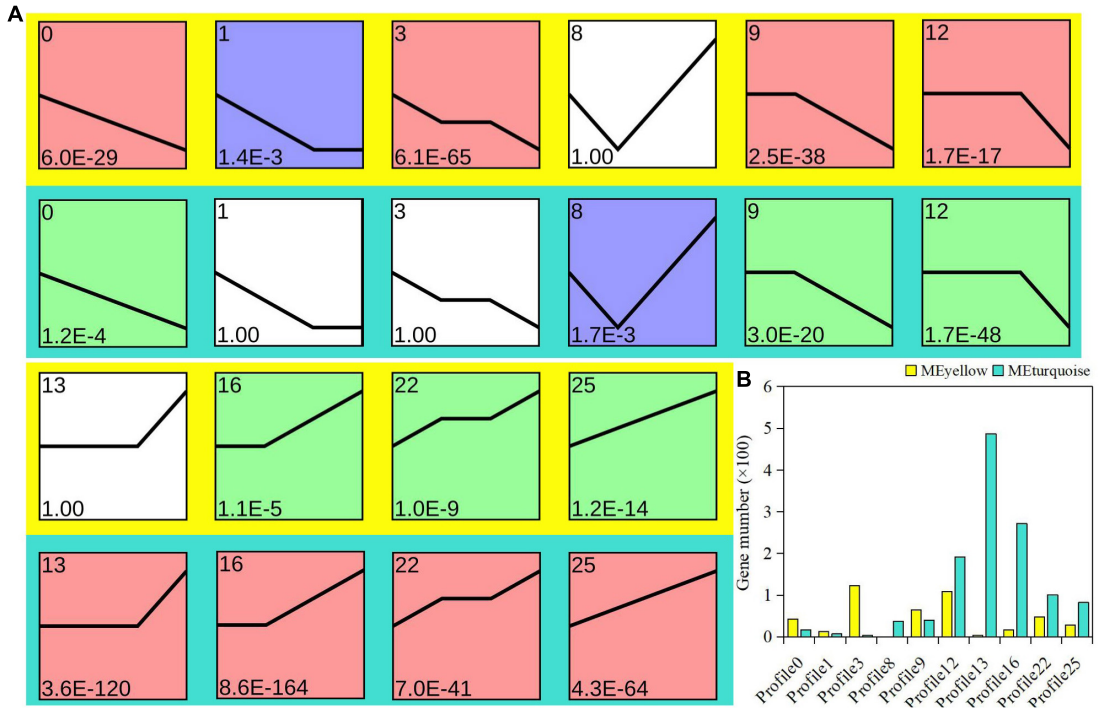


FIGURE 3 | Analysis of gene expression trend in yellow and turquoise modules. **(A)** Each plot shows the median expression value of all genes with a similar expression profile during *Prunus mume* bud dormancy. **(B)** Statistics of gene quantity in profiles.

A Network of Expression and Transcriptional Regulation in Modules

To estimate the crucial genes by weighted gene co-expression network and to better rank expression and regulatory interactions, a total 5,251 genes from the yellow and turquoise modules were annotated according to the orthologous genes of *Arabidopsis thaliana*. TFs played a more extensive role in gene expression, which was mainly controlled *via* the interactions with promoters and chromatin-modifying machinery (Liu and Stewart, 2016; Moenga et al., 2020). In this study, we obtained 156 and 76 TFs in the yellow and turquoise modules, respectively, which were co-expressed with all genes in the modules. The TF interaction network was predicted by the string database and hid disconnected nodes. Therefore, a network of expression and transcriptional regulation about hub TFs was constructed, combining co-expression relationships and interaction networks (**Figure 4**). The network showed that TFs from different modules with different expression modes also had close regulatory relationships. Profile X, with no significant enrichment in the gene expression pattern, contained the most TFs and interacted with other TFs, suggesting that these TFs might not be involved in bud dormancy regulation but could interact with TFs involved in bud dormancy regulation and participate in other life activities. We found that most transcription factors were associated with abiotic stress, growth and development, and hormonal responses, such as *WRKY*, *bZIP*, *AP2/ERE*, and *MYB* genes. Interestingly, the TFs of different modules in the same profile (profile 12) had different molecular functions, with the TF function in the yellow module preferring dormancy and abiotic stress, while the TF function in the blue module preferred transition to germination, flowering, and shoot development. In profile 3, the functional annotation of each TF was most closely related, mainly including cold stress response (*CBF*, *CRF*, and *WRKY*), plant hormone (*ABF*, *ERF*, and *MYB*), and light response (*SVP* and *LSH*). These results indicated that plant-hormone-related TFs were involved in the regulation of plant dormancy and that they play an important role in the network of expression and transcriptional regulation.

Identification of Hub Genes and Expression Analysis

Considering the expression trend and intramodular connectivity for genes in bud dormancy, three genes (*PmABF2*, *PmABF4*, and *PmSVP*) closely related to ABA were selected; they belong to profile 3 in the yellow module. *PmABF2*, *PmABF4*, and *PmSVP*, as hub genes, were co-expressed with 1,301, 1,197, and 1,305 genes in the yellow module by the WGCNA algorithm, respectively. Here the top 30 genes co-expressed with *PmABF2*, *PmABF4*, and *PmSVP* were presented based on the intramodular connectivity through a network (**Figure 5A**). We found that more than half of the genes were collectively co-expressed with these three genes; however, only two genes were specific in the *ABF2* gene co-expression network. The expression patterns of 47 genes in this network were divided into two categories, among which 22 genes were upregulated and 25 genes were downregulated during bud dormancy (**Figure 5B**), suggesting that *PmABF2*,

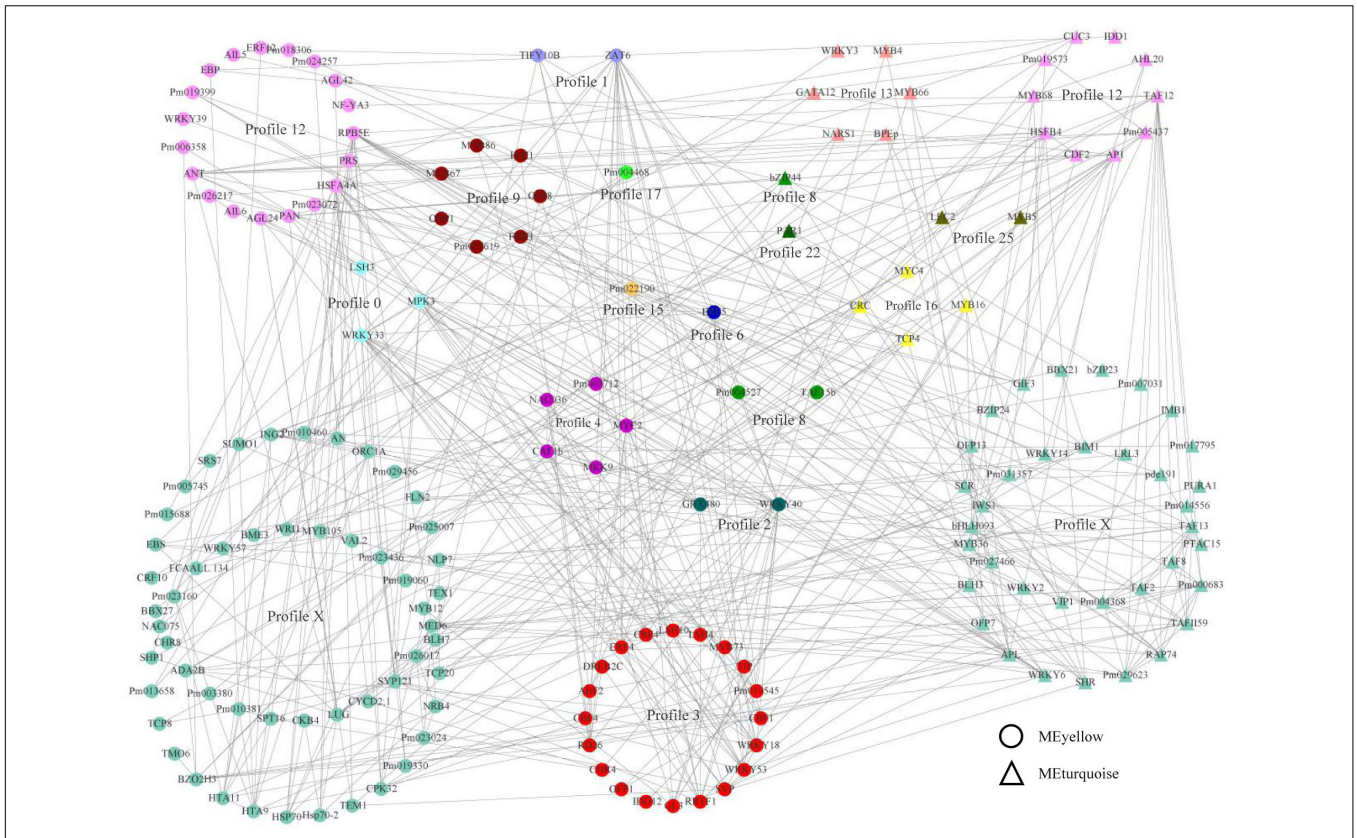


FIGURE 4 | A network of expression and transcriptional regulation. The transcription factors in the turquoise module were noted with triangles, and those in the yellow module were noted with circles. The colors indicated the trends in which the genes were expressed.

PmABF4, and *PmSVP* might negatively regulate the upregulated genes. In ABA signal transduction pathway, *PmPYL* (Pm019775) and two *PmPP2C* (Pm000664 and Pm019734) genes had similar expression patterns with *PmABFs* (**Figure 5C**). However, PYL inhibits the activity of PP2Cs in an ABA-independent manner but more efficiently when activated by ABA (Park et al., 2009; Hao et al., 2011). These results suggested that another upregulated *PmPP2C* gene (Pm022723) might play a role in this pathway. The two clusters were composed of transcripts that were specific dormancy stages in the expression trends, and the associated *PmABF2*, *PmABF4*, and *PmSVP* provided clues about the regulatory networks.

Although the intensity of expression varies among different tissue, 65.5% of these genes were upregulated with ABA treatment, compared with 34.5% downregulated transcripts (**Figure 6**). The expression of both *PmPYL* and *PmABF* homologs, which were involved in the first and final steps of the ABA signal transduction pathway, respectively, was induced by ABA treatment (**Figure 6A**). Interestingly, the ABA signal transduction pathway led to a higher expression level of *PmPYL* by ABA treatment while downregulating the *PmPP2C* gene (Pm022723), which acted as a negative regulator of seed dormancy by inhibiting ABA signaling and subsequently activating GA signaling (Kim et al., 2013). This regulatory pathway might also be applicable to bud dormancy in *P. mume*. However, other two *PmPP2C* genes were upregulated after ABA treatment, which was consistent with the level of gene expression during bud dormancy. *PmSVP*, a *P. mume* MADS-box gene, was upregulated by ABA treatment, and half of its associated genes in the co-expression network were also upregulated. The two most prominent upregulated genes, *PmIF2γ* (Pm005029, a translation elongation factor EF1A/initiation factor IF2gamma family protein) and *PmNF-YA3* (Pm027028, a nuclear transcription factor Y subunit A-3), were co-expressed with *PmABF2/4* and *PmSVP*, respectively. In downregulated genes, *EMB2766* (Pm017771) was most significantly inhibited by ABA treatment, whose function has been described as the structural constituent of nuclear pore in *Arabidopsis* (Tamura et al., 2010). At the same time, the functional annotations of some downregulated genes were related to plant growth and development, such as *PIN6* (Pm022685), an auxin efflux carrier family protein gene (**Figure 6B**). These gene expressions confirmed and extended the dormancy process analysis above, furthering the description of a conserved ABA treatment response.

DISCUSSION

Bud dormancy is a common adaptive response of perennial plants to survive in adverse environmental conditions

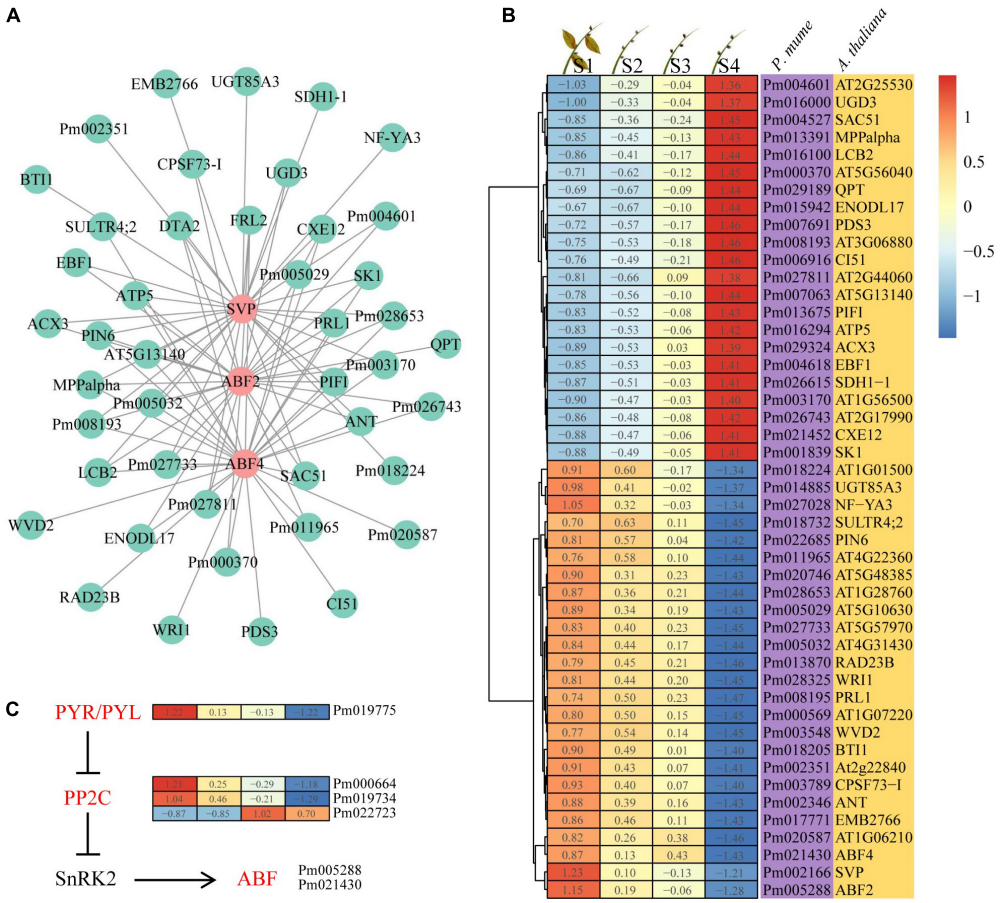


FIGURE 5 | Interaction network and expression of the hub genes. **(A)** Interaction of gene co-expression network of the hub genes. **(B)** Heat map showing the genes that were expressed at four different dormancy stages of *Prunus mume*. **(C)** Abscisic acid (ABA) signal transduction pathway. The ABA-related genes in yellow module were highlighted in red.

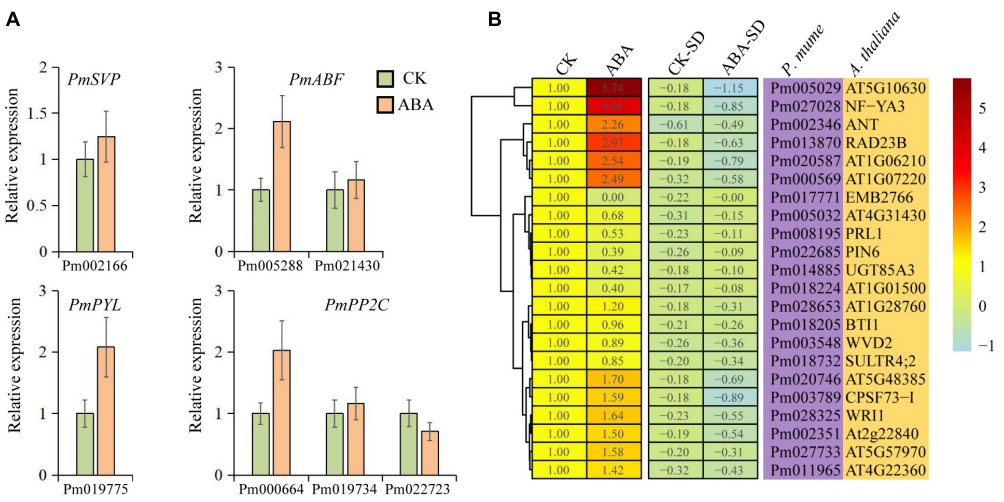


FIGURE 6 | Expression patterns of candidate genes under abscisic acid (ABA) treatment. **(A)** qRT-PCR of 7 selected genes under ABA treatment. **(B)** Heatmap of 22 genes under ABA treatment. CK represents the control group, ABA represents the ABA treatment group, and CK-SD and ABA-SD represent the standard deviations of relative expression levels in the groups, respectively.

(Tarancón et al., 2017; Liu and Sherif, 2019). Plant dormancy greatly preserves reproductive success, productivity, and survival through stalled growth and reduced metabolic activities (Tarancón et al., 2017). *P. mume* has the winter-flowering property, which also means that its bud dormancy period is relatively short. Although some cultivated species are able to tolerate temperatures as low as −19°C, it is very different from the freezing tolerance of apricot, a close relative species with a longer bud dormancy period (Zhang, 1987, 1989). Bud dormancy is likely one of the key factors in the poor freezing tolerance of *P. mume*.

Plant hormone-mediated dormancy has been extensively demonstrated, especially plant hormones ABA and GA (Liu and Sherif, 2019; Chen et al., 2020). The changes in ABA and GA content are inversely correlated during dormancy, with the ABA/GA ratio varying with the level of dormancy (Liu and Sherif, 2019). In *P. mume*, the content of ABA decreased gradually, while the content of GA increased gradually from bud dormancy through dormancy release (Wen et al., 2016; Zhang et al., 2018). The ABA signal transduction pathway has been confirmed by a large number of studies, mainly including *PYR/PYL*, *PP2C*, *SnRK2s*, and *AREB/ABF* genes (Raghavendra et al., 2010; Soon et al., 2012). The observed changes in the expression of genes involved in ABA signal transduction, including *PmPYL*, *PmPP2C*, and *PmABF*, suggest a central role for ABA during *P. mume* bud dormancy (**Figure 5C**). The regulation of bud dormancy by the ABA signal transduction pathway has been demonstrated in woody plants such as grape, pear, and poplar (Ruttink et al., 2007; Boneh et al., 2012; Li J. et al., 2018). ABA responses during bud dormancy may also be mediated by ABA-dependent pathways in *P. mume*. Interestingly, *PmCBF* genes, as well-known cold response conduits, are co-expressed with *PmABF* genes, belonging to profile 3 (**Figure 4**). CBF/DREB TFs, as well-known cold response genes, have been shown to bind directly to interact with dormancy regulator DAM TFs in an ABA-independent manner (Zhao et al., 2018b,c). These results suggest a multiple manner in the dormancy regulation of each other, as demonstrated by the TF interaction network. Based on this network, we speculate that these TFs and the regulated target genes may further regulate plant dormancy by responding to low temperature or photoperiod to cause changes in endogenous hormones. The expressions of ABA-related genes were upregulated by low temperature in grape (Boneh et al., 2012) and upregulated by short day in poplar (Singh et al., 2018; Tylewicz et al., 2018).

In addition, the molecular mechanism of ABA involvement in bud dormancy by regulating *SVP/SVL* genes has been widely reported. ABA content is increased by environmental signals, thereby inhibiting the expression of a chromodomain remodeling factor *PICKLE* (*PKL*) gene (Tylewicz et al., 2018). *PKL* negatively regulates the expression of *SVP/SVL* gene and further negatively regulates *flowering locus T* (*FT*) and GA-related gene (Singh et al., 2018, 2019). Meanwhile, SVP/SVL, acting as a MADS box, positively regulates *callose synthase 1* (*CALS1*), thereby promoting bud dormancy (Singh et al., 2019). In *P. mume*, the expression pattern of *PmSVP* gene is gradually decreasing during bud dormancy. The *PmSVP* gene is upregulated by ABA treatment, and half of its associated genes in the co-expression network are also upregulated. Taken together, the *PmSVP* gene is central to the regulation of *P. mume* bud dormancy. Short day can induce the expression of *SVL* gene, while low temperature inhibits gene expression in poplar (Singh et al., 2018, 2019).

The cell cycle also plays a key role in bud dormancy, in contrast to the dormancy mechanism of intercellular communication involving the *SVL* gene. The cells of the dormant bud stagnate in the G1 phase during the cell cycle (Cespedes et al., 1995; Gutierrez et al., 2002). ABA can maintain cells in the G1 phase by interfering with DNA replication (Swiatek et al., 2002; Vergara et al., 2017). In the yellow module, *PmCYCD2;1* (Pm002137) is enriched during *P. mume* bud dormancy, which is negatively correlated with ABA content. CYCD proteins interact with cyclin-dependent kinases, which is activated by cell division cycle protein (Riou-Khamlichi et al., 1999, 2000). In grape bud dormancy, ABA represses the expression of cell cycle genes, including two *VvCDKB*s, three *VvCYCA*s, *VvCYCB*, and *VvCYCD3.2a* (Vergara et al., 2017). According to current studies, the molecular mechanism of bud dormancy is mainly based on ABA, which has been studied from different directions.

CONCLUSION

In the present study, we performed WGCNA of the association between bud dormancy and plant hormones. A total of 23 distinct co-expression modules were identified by WGCNA, each consisting of 35 to 3,825 correlated genes, of which 12 modules were significantly associated with IAA, ABA, GA1, GA3, and GA4. Most of the gene co-expression modules showed an inverse relationship between ABA and GA. The enrichment analysis revealed that these genes were related to the dormancy process, seed dormancy process, and plant hormone signal transduction. In ABA signal transduction, *PmABF2* and *PmABF4* genes, belonging to profile 3 in the yellow module, were positively correlated with *P. mume* bud dormancy. Meanwhile, *PmSVP* gene, a MADS box gene, was upregulated by ABA treatment, and half of its associated genes in the co-expression network were also upregulated. Taken together, *PmABF2*, *PmABF4*, and *PmSVP* might be the hub genes involved in the regulation of *P. mume* bud dormancy through changes of ABA content.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

PL, TZ, and QZ conceived and designed the experiments. PL and ZZ performed the experiments. PL, WL, ZZ, JW, TC, and LQ analyzed the data. PL and TZ wrote the manuscript. All the authors approved the final manuscript.

FUNDING

This research was funded by the National Key R&D Program of China (2019YFD1001500) and Special Fund for Beijing Common Construction Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.698598/full#supplementary-material

Supplementary Figure 1 | Analysis of the scale-free fit index (left) and the mean connectivity (right) for various soft-thresholding powers.

Supplementary Figure 2 | The interactive relationship network of 400 genes.

Supplementary Table 1 | Primer sequences used for qRT-PCR.

Supplementary Table 2 | Cellular component Gene Ontology terms for genes in the yellow module.

Supplementary Table 3 | Molecular function Gene Ontology terms for genes in the yellow module.

Supplementary Table 4 | Biological process Gene Ontology terms for genes in the yellow module.

Supplementary Table 5 | Cellular component Gene Ontology terms for genes in the turquoise module.

Supplementary Table 6 | Molecular function Gene Ontology terms for genes in the turquoise module.

Supplementary Table 7 | Biological process Gene Ontology terms for genes in the turquoise module.

Supplementary Table 8 | Kyoto Encyclopedia of Genes and Genomes analysis for genes in the yellow module.

Supplementary Table 9 | Kyoto Encyclopedia of Genes and Genomes analysis for genes in the turquoise module.

REFERENCES

An, F., Zhao, Q., Ji, Y., Li, W., Jiang, Z., Yu, X., et al. (2010). Ethylene-induced stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 is mediated by proteasomal degradation of EIN3 binding F-box 1 and 2 that requires EIN2 in *Arabidopsis*. *Plant Cell* 22, 2384–2401. doi: 10.1105/tpc.110.076588

Barrero, J. M., Jacobsen, J. V., Talbot, M. J., White, R. G., Swain, S. M., Garvin, D. F., et al. (2012). Grain dormancy and light quality effects on germination in the model grass *Brachypodium distachyon*. *N. Phytol.* 193, 376–386. doi: 10.1111/j.1469-8137.2011.03938.x

Boneh, U., Biton, I., Zheng, C., Schwartz, A., and Ben-Ari, G. (2012). Characterization of potential ABA receptors in *Vitis vinifera*. *Plant Cell Rep.* 31, 311–321. doi: 10.1007/s00299-011-1166-z

Bossi, F., Cordoba, E., Dupré, P., Mendoza, M. S., Román, C. S., and León, P. (2009). The *Arabidopsis ABA-INSENSITIVE* (*ABI*) 4 factor acts as a central transcription activator of the expression of its own gene, and for the induction of *ABI5* and *SBE2.2* genes during sugar signaling. *Plant J.* 59, 359–374. doi: 10.1111/j.1365-313X.2009.03877.x

Cespedes, R. D., Peretsman, S. J., and Blatt, S. P. (1995). Significance of hematuria in patients infected with human immunodeficiency virus. *J. Urol.* 154, 1455–1456. doi: 10.1016/S0022-5347(01)66890-3

Chao, W. S., Doğramacı, M., Horvath, D. P., Anderson, J. V., and Foley, M. E. (2017). Comparison of phytohormone levels and transcript profiles during seasonal dormancy transitions in underground adventitious buds of leafy spurge. *Plant Mol. Biol.* 94, 281–302. doi: 10.1007/s11103-017-0607-7

Chen, C., Xia, R., Chen, H., and He, Y. (2018). TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* [preprint] doi: 10.1101/289660

Chen, K., Li, G. J., Bressan, R. A., Song, C.-P., Zhu, J. K., and Zhao, Y. (2020). Abscisic acid dynamics, signaling, and functions in plants. *J. Integr. Plant Biol.* 62, 25–54. doi: 10.1111/jipb.12899

Cheng, G., Zhang, L., Wang, H., Lu, J., Wei, H., and Yu, S. (2020). Transcriptomic profiling of young cotyledons response to chilling stress in two contrasting cotton (*Gossypium hirsutum* L.) genotypes at the seedling stage. *Int. J. Mol. Sci.* 21:5095. doi: 10.3390/ijms21145095

de Wit, M., Spoel, S. H., Sanchez-Perez, G. F., Gommers, C. M. M., Pieterse, C. M. J., Voesenek, L. A. C. J., et al. (2013). Perception of low red:far-red ratio compromises both salicylic acid- and jasmonic acid-dependent pathogen defences in *Arabidopsis*. *Plant J.* 75, 90–103. doi: 10.1111/tpj.12203

Farhadian, M., Rafat, S. A., Panahi, B., and Mayack, C. (2021). Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. *Sci. Rep.* 11:2367. doi: 10.1038/s41598-021-81888-z

Gene Ontology, C. (2021). The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi: 10.1093/nar/gkaa1113

Gutierrez, C., Ramirez-Parra, E., Castellano, M. M., and del Pozo, J. C. (2002). G1 to S transition: more than a cell cycle engine switch. *Curr. Opin. Plant Biol.* 5, 480–486. doi: 10.1016/S1369-5266(02)00301-1

Hao, Q., Yin, P., Li, W., Wang, L., Yan, C., Lin, Z., et al. (2011). The molecular basis of ABA-independent inhibition of PP2Cs by a subclass of PYL proteins. *Mol. Cell* 42, 662–672. doi: 10.1016/j.molcel.2011.05.011

Holdsworth, M. J., Bentsink, L., and Soppe, W. J. J. (2008). Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *N. Phytol.* 179, 33–54. doi: 10.1111/j.1469-8137.2008.02437.x

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

Kainer, K. A., Duryea, M. L., White, T. L., and Johnson, J. D. (1991). Slash pine bud dormancy as affected by lifting date and root wrenching in the nursery1. *Tree Physiol.* 9, 479–489. doi: 10.1093/treephys/9.4.479

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi: 10.1093/nar/gkaa970

Kegge, W., Ninkovic, V., Glinwood, R., Welschen, R. A. M., Voesenek, L. A. C. J., and Pierik, R. (2015). Red:far-red light conditions affect the emission of volatile organic compounds from barley (*Hordeum vulgare*), leading to altered biomass allocation in neighbouring plants. *Ann. Bot.* 115, 961–970. doi: 10.1093/aob/mcv036

Kim, W., Lee, Y., Park, J., Lee, N., and Choi, G. (2013). HONSU, a protein phosphatase 2C, regulates seed dormancy by inhibiting ABA signaling in *Arabidopsis*. *Plant Cell Physiol.* 54, 555–572. doi: 10.1093/pcp/pct017

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

Lau, O. S., and Deng, X. W. (2010). Plant hormone signaling lightens up: integrators of light and hormones. *Curr. Opin. Plant Biol.* 13, 571–577. doi: 10.1016/j.pbi.2010.07.001

Li, J., Xu, Y., Niu, Q., He, L., Teng, Y., and Bai, S. (2018). Abscisic acid (ABA) promotes the induction and maintenance of pear (*Pyrus pyrifolia* White Pear Group) flower bud endodormancy. *Int. J. Mol. Sci.* 19:310. doi: 10.3390/ijms19010310

Li, S., and Dami, I. (2015). Responses of *Vitis vinifera* 'Pinot gris' grapevines to exogenous abscisic acid (ABA): I. yield, fruit quality, dormancy, and freezing tolerance. *J. Plant Growth Regul.* 35, 245–255. doi: 10.1007/s00344-015-9529-2

Li, S., Tan, Q., Sun, M., Xu, G., Li, C., Fu, X., et al. (2018). Protein changes in response to photoperiod during dormancy induction in peach leaves and flower buds. *Sci. Hortic.* 239, 114–122. doi: 10.1016/j.scienta.2018.05.010

Liu, J., and Sherif, S. M. (2019). Hormonal orchestration of bud dormancy cycle in deciduous woody perennials. *Front. Plant Sci.* 10:1136. doi: 10.3389/fpls.2019.01136

Liu, W., and Stewart, C. N. (2016). Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* 37, 36–44. doi: 10.1016/j.copbio.2015.10.001

Liu, X., and Hou, X. (2018). Antagonistic regulation of ABA and GA in metabolism and signaling pathways. *Front. Plant Sci.* 9:251. doi: 10.3389/fpls.2018.00251

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

Maurya, J. P., and Bhalerao, R. P. (2017). Photoperiod- and temperature-mediated control of growth cessation and dormancy in trees: a molecular perspective. *Ann. Bot.* 120, 351–360. doi: 10.1093/aob/mcx061

Moenga, S. M., Gai, Y., Carrasquilla-Garcia, N., Perilla-Henao, L. M., and Cook, D. R. (2020). Gene co-expression analysis reveals transcriptome divergence between wild and cultivated chickpea under drought stress. *Plant J.* 104, 1195–1214. doi: 10.1111/tpj.14988

Panahi, B., and Hejazi, M. A. (2021). Weighted gene co-expression network analysis of the salt-responsive transcriptomes reveals novel hub genes in green halophytic microalgae *Dunaliella salina*. *Sci. Rep.* 11:1607. doi: 10.1038/s41598-020-80945-3

Panahi, B., Farhadian, M., and Hejazi, M. A. (2020). Systems biology approach identifies functional modules and regulatory hubs related to secondary metabolites accumulation after transition from autotrophic to heterotrophic growth condition in microalgae. *PLoS One* 15:e0225677. doi: 10.1371/journal.pone.0225677

Parada, F., Noriega, X., Dantas, D., Bressan-Smith, R., and Pérez, F. J. (2016). Differences in respiration between dormant and non-dormant buds suggest the involvement of ABA in the development of endodormancy in grapevines. *J. Plant Physiol.* 201, 71–78. doi: 10.1016/j.jplph.2016.07.007

Park, S. Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., et al. (2009). Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science (New York, N.Y.)* 324, 1068–1071. doi: 10.1126/science.1173041

Presson, A. P., Sobel, E. M., Papp, J. C., Suarez, C. J., Whistler, T., Rajeevan, M. S., et al. (2008). Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst. Biol.* 2:95. doi: 10.1186/1752-0509-2-95

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

Riou-Khamlichi, C., Huntley, R., Jacqmard, A., and James, A. H. M. (1999). Cytokinin activation of *Arabidopsis* cell division through a D-Type cyclin. *Science* 283, 1541–1544. doi: 10.1126/science.283.5407.1541

Riou-Khamlichi, C., Menges, M., Healy, J. M., and Murray, J. A. (2000). Sugar control of the plant cell cycle: differential regulation of *Arabidopsis* D-type cyclin gene expression. *Mol. Cell. Biol.* 20, 4513–4521. doi: 10.1128/mcb.20.13.4513-4521.2000

Ruttink, T., Arend, M., Morreel, K., Storme, V. R., Rombauts, S., Fromm, J. R., et al. (2007). A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* 19, 2370–2390. doi: 10.1105/tpc.107.052811

Shu, K., Liu, X. D., Xie, Q., and He, Z. H. (2016). Two faces of one seed: hormonal regulation of dormancy and germination. *Mol. Plant* 9, 34–45. doi: 10.1016/j.molp.2015.08.010

Singh, R. K., Maurya, J. P., Azeez, A., Miskolczi, P., Tylewicz, S., Stojkoviè, K., et al. (2018). A genetic network mediating the control of bud break in hybrid aspen. *Nat. Commun.* 9, 4173. doi: 10.1038/s41467-018-06696-y

Singh, R. K., Miskolczi, P., Maurya, J. P., and Bhalerao, R. P. (2019). A tree ortholog of SHORT VEGETATIVE PHASE floral repressor mediates photoperiodic control of bud dormancy. *Curr. Biol.* 29, 128–133.e122.

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics (Oxford, England)* 27, 431–432. doi: 10.1093/bioinformatics/btq675

Soon, F. F., Ng, L. M., Zhou, X. E., West, G. M., Kovach, A., Tan, M. H. E., et al. (2012). Molecular mimicry regulates ABA signaling by SnRK2 kinases and PP2C phosphatases. *Science (New York, N.Y.)* 335, 85–88. doi: 10.1126/science.1215106

Sun, S., Wang, D., Li, J., Lei, Y., Li, G., Cai, W., et al. (2021). Transcriptome analysis reveals photoperiod-associated genes expressed in rice anthers. *Front. Plant Sci.* 12:621561. doi: 10.3389/fpls.2021.621561

Swiatek, A., Lenjou, M., Van Bockstaele, D., Inzé, D., and Van Onckelen, H. (2002). Differential effect of jasmonic acid and abscisic acid on cell cycle progression in tobacco BY-2 cells. *Plant Physiol.* 128, 201–211. doi: 10.1104/pp.128.1.201

Swift, J., Alvarez, J. M., Araus, V., Gutiérrez, R. A., and Coruzzi, G. M. (2020). Nutrient dose-responsive transcriptome changes driven by Michaelis-Menten kinetics underlie plant growth rates. *Proc. Natl. Acad. Sci. U.S.A.* 117, 12531–12540. doi: 10.1073/pnas.1918619117

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937

Tamura, K., Fukao, Y., Iwamoto, M., Haraguchi, T., and Hara-Nishimura, I. (2010). Identification and characterization of nuclear pore complex components in *Arabidopsis thaliana*. *Plant Cell* 22, 4084–4097. doi: 10.1105/tpc.110.079947

Tarancón, C., González-Grandío, E., Oliveros, J. C., Nicolas, M., and Cubas, P. (2017). A conserved carbon starvation response underlies bud dormancy in woody and *Herbaceous* species. *Front. Plant Sci.* 8:788. doi: 10.3389/fpls.2017.00788

Tylewicz, S., Petterle, A., Marttila, S., Miskolczi, P., Azeez, A., Singh, R. K., et al. (2018). Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication. *Science* 360:212. doi: 10.1126/science.aan8576

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Vergara, R., Noriega, X., Aravena, K., Prieto, H., and Pérez, F. J. (2017). ABA represses the expression of cell cycle genes and may modulate the development of endodormancy in grapevine buds. *Front. Plant Sci.* 8:812. doi: 10.3389/fpls.2017.00812

Wen, L., Zhong, W., Huo, X., Zhuang, W., Ni, Z., and Gao, Z. (2016). Expression analysis of ABA- and GA-related genes during four stages of bud dormancy in Japanese apricot (*Prunus mume* Sieb. et Zucc). *J. Hortic. Sci. Biotechnol.* 91, 1–8. doi: 10.1080/14620316.2016.1160546

Wu, R., Wang, T., Warren, B. A. W., Allan, A. C., Macknight, R. C., and Varkonyi-Gasic, E. (2017). Kiwifruit *SVP2* gene prevents premature budbreak during dormancy. *J. Exp. Bot.* 68, 1071–1082. doi: 10.1093/jxb/erx014

Xu, Z., Zhang, Q., Sun, L., Du, D., Cheng, T., Pan, H., et al. (2014). Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in *Prunus mume*. *Mol. Genet. Genom.* 289, 903–920. doi: 10.1007/s00438-014-0863-z

Yuan, L., Qian, G., Chen, L., Wu, C. L., Dan, H. C., Xiao, Y., et al. (2018). Co-expression network analysis of biomarkers for adrenocortical carcinoma. *Front. Genet.* 9:328. doi: 10.3389/fgene.2018.00328

Zhang, Q. (1987). The interspecific crossing of mei flower and cold hardiness breeding. *J. Beijing For. Univ.* 9, 69–79.

Zhang, Q. (1989). The effect of growth regulators on fruit setting of interspecies hybridization between mei flower (*Prunus mume* S. Et Z.) and apricot (*P. armeniaca*) etc. *Acta Hortic.* 251, 403–406. doi: 10.17660/ActaHortic.1989.251.56

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., et al. (2012). The genome of *Prunus mume*. *Nat. Commun.* 3, 1318. doi: 10.1038/ncomms2290

Zhang, Z., Zhuo, X., Zhao, K., Zheng, T., Han, Y., Yuan, C., et al. (2018). transcriptome profiles reveal the crucial roles of hormone and sugar in the bud dormancy of *Prunus mume*. *Sci. Rep.* 8:5090. doi: 10.1038/s41598-018-23108-9

Zhao, K., Zhou, Y., Ahmad, S., Xu, Z., Li, Y., Yang, W., et al. (2018a). Comprehensive cloning of *Prunus mume* dormancy associated MADS-Box genes and their response in flower bud development and dormancy. *Front. Plant Sci.* 9:17. doi: 10.3389/fpls.2018.00017

Zhao, K., Zhou, Y., Ahmad, S., Yong, X., Xie, X., Han, Y., et al. (2018b). PmCBFs synthetically affect PmDAM6 by alternative promoter binding and protein complexes towards the dormancy of bud for *Prunus mume*. *Sci. Rep.* 8, 4527. doi: 10.1038/s41598-018-22537-w

Zhao, K., Zhou, Y., Li, Y., Zhuo, X., Ahmad, S., Han, Y., et al. (2018c). Crosstalk of PmCBFs and PmDAMs Based on the changes of phytohormones under seasonal cold stress in the stem of *Prunus mume*. *Int. J. Mol. Sci.* 19:15. doi: 10.3390/ijms19020015

Zhou, W., Chen, F., Luo, X., Dai, Y., Yang, Y., Zheng, C., et al. (2020). A matter of life and death: Molecular, physiological, and environmental regulation of seed longevity. *Plant Cell Environ.* 43, 293–302. doi: 10.1111/pce.13666

Zhuo, X., Zheng, T., Zhang, Z., Zhang, Y., Jiang, L., Ahmad, S., et al. (2018). Genome-Wide analysis of the NAC transcription factor gene family reveals differential expression patterns and cold-stress responses in the woody plant *Prunus mume*. *Genes* 9:494. doi: 10.3390/genes9100494

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.

Copyright © 2021 Li, Zheng, Zhang, Liu, Qiu, Wang, Cheng and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 July 2021
doi: 10.3389/fgene.2021.706930

Combined Analysis of MicroRNAs and Target Genes Revealed miR156-SPLs and miR172-AP2 Are Involved in a Delayed Flowering Phenomenon After Chromosome Doubling in Black Goji (*Lycium ruthencium*)

Shupei Rao[1,2], *Yue Li*[1,2] *and Jinhuan Chen*[1,2]*

[1] College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [2] National Engineering Laboratory for Tree Breeding, Beijing Forestry University, Beijing, China

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Hikmet Budak,
Montana Bioagriculture Inc., United States
Xiaozeng Yang,
Beijing Academy of Agricultural and Forestry Sciences, China

***Correspondence:**
Jinhuan Chen
chenjh@bjfu.edu.cn

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

Received: 08 May 2021
Accepted: 14 June 2021
Published: 15 July 2021

Citation:
Rao S, Li Y and Chen J (2021) Combined Analysis of MicroRNAs and Target Genes Revealed miR156-SPLs and miR172-AP2 Are Involved in a Delayed Flowering Phenomenon After Chromosome Doubling in Black Goji (Lycium ruthencium). Front. Genet. 12:706930. doi: 10.3389/fgene.2021.706930

Polyploidy, which is widely distributed in angiosperms, presents extremely valuable commercial applications in plant growth and reproduction. The flower development process of higher plants is essential for genetic improvement. Nevertheless, the reproduction difference between polyploidy and the polyploid florescence regulatory network from the perspective of microRNA (miRNA) remains to be elucidated. In this study, the autotetraploid of *Lycium ruthenicum* showed late-flowering traits compared with the progenitor. Combining the association of miRNA and next-generation transcriptome technology, the late-flowering characteristics triggered by chromosome duplication may be caused by the age pathway involved in miR156-SPLs and miR172-AP2, which inhibits the messenger RNA (mRNA) transcripts of *FT* in the leaves. Subsequently, FT was transferred to the shoot apical meristem (SAM) to inhibit the expression of the flowering integration factor *SOC1*, which can eventually result in delayed flowering time. Our exploration of the flowering regulation network and the control of the flowering time are vital to the *goji* producing in the late frost area, which provides a new perspective for exploring the intrinsic molecular mechanism of polyploid and the reproductive development of flowering plants.

Keywords: polyploid, florescence, miRNA, *Lycium ruthenicum*, molecular mechanism

INTRODUCTION

The integrated lifecycle of higher plants experiences a series of developmental processes subsuming seed germination, vegetative growth, flowering, fertilization, embryonic development, and seed formation. Among them, flowering is a momentous physiological behavior and the symbol from vegetative to reproductive growth in the growth course of higher plants

(Spanudakis and Jackson, 2014). Blooming in the punctual time point sets the stage for the successful sexual reproduction and subsequent seed and fruit development (Jung and Muller, 2009). Nevertheless, the regulation of flowering time, whose onset and progression are strictly controlled, is composed of an intricate gene network integrated with environmental and endogenous cues to control the expression of a group of crucial flowering genes in the shoot apical meristem (SAM) (Michaels, 2009; Lee et al., 2020). Florigen as a flowering hormone in plants is transmitted from the light-receiving organs to the SAM to form a floral response. Studies have proved that *CONSTANS* (*CO*) can induce *Flowering Locus T* (*FT*) expression in leaves. When both *CO* and *FT* were expressed in the SAM, only the *FT* can induce flowering in plants, which showed that the FT protein can transfer to the SAM to promote the flowering transformation and is the key component of florigen. *FT* is an important integration factor in various pathways of flowering, which promotes onset of the reproductive phase and flower formation and is further transmitted to downstream flowering development genes, thereby promoting flowering (Zhu et al., 2020). Insomuch as flowering time will noticeably affect plant health and crop yields, shifting the seasonal timing of reproduction is also the main goal of plant breeding. A meticulous grasp of the flowering time regulatory mechanisms, to produce new varieties that are better suitable for the change in the local climatic conditions, is essential for continuous improvement of agricultural practices.

Polyploidization has always been recognized as a crucial force and channel to facilitate plant evolution and speciation during the evolution process. The biological and genetic advantages of polyploid over diploid are tremendous. Most evidence have concluded that polyploid usually has novel characteristics that are unattainable by the diploid progenitor, such as increased organ biomass, resistance to adversity, and changes in flowering time, which may cause polyploidy to enter a new environmental niche and domesticate plants from a new perspective (Braynen et al., 2021). To better utilize the commercial worthlessness brought by polyploid, people often make use of the newly synthesized polyploid to simulate the polyploid generation process. A comparison of multiple biological genetics between the polyploid and the parent species in the existing research found that polyploid has undergone major changes in many aspects. At the moment, various scientific progresses in flowering transition has authenticated that many allotetraploid plants have a delayed flowering phenotype (Mayfield et al., 2011). Nevertheless, the pivotal factors that trigger polyploidy-induced late-flowering traits are ambiguous and complex. The influence not only exists in heritable variations, containing chromosome rearrangement and sequence elimination, but also in epigenetic variations, such as methylation, gene silencing, and small RNA changes.

Noncoding RNAs (ncRNAs) play an important role in regulating plant growth and development networks. With the emergence of deep sequencing technology, the research on sRNAs that cannot encode functional proteins has been accelerated. The structural categories of ncRNAs include ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snRNA) (Brant and Hikmet, 2018). Regulatory ncRNAs are generally divided into long ncRNAs (lncRNAs $>$ 200 nt) and small ncRNAs (sRNAs, 18–30 nt). sRNAs are mainly divided into two categories, namely, siRNA and microRNA (miRNA), which play an important role in gene regulation and plant development. siRNA is a nonendogenous double-stranded RNA, which can act on any part of mRNA and degrade target gene. Unlike siRNA, miRNA is a type of noncoding RNAs abundantly encoded by endogenous genes with a length of approximately 21–24 bases, which can precisely and effectively monitor gene expression at the posttranscriptional level through negative regulation methods comprising translation inhibition or degradation of the target gene mRNA (Song et al., 2019; Zhang et al., 2019a). Recently, the majority of previous research on miRNAs focuses on transcriptome analysis and function prediction, yet reports on the interaction of miRNAs in the reproductive biology of polyploid plants are relatively rare.

Lycium ruthencium is a traditional Chinese medicine of the Solanaceae and famous for its rich nutrients and medicinal value, which is mainly distributed in the late frost-prone areas in the northwestern region of China (Wang et al., 2018). *L. ruthenicum* contains water-soluble antioxidant proanthocyanidins and polysaccharides, which can reduce the production of reactive oxygen species (ROS) and inflammation and inhibit the growth of cancer cells (Zhang et al., 2019b). Geographically, the special climate of *L. ruthencium* growth zone is characterized by different degrees of frost every year, which is usually the most serious in spring. The transition from bud stage to flowering and fruiting stage of *L. ruthencium* occurs in spring. When the minimum temperature is lower than 0°C, the flower buds are likely to be frozen, dead, and withered, which directly leads to the decline of the yield and quality of the first fruit of *L. ruthencium*. However, the nutrient components of the first fruit are higher than those of the later summer and autumn fruits and occupy an important position in the wolfberry industry. Therefore, staggering the flowering and frost periods has a greater impact on the production and the development of the wolfberry industry. Here, we focused on the difference in the flowering time of *L. ruthencium* tetraploid and analyzed the molecular mechanism of tetraploid on the change in flowering period from the perspective of miRNA and transcriptome association analysis, in order to provide a theoretical basis for the creation of late-flowering.

MATERIALS AND METHODS

Plant Growth Condition and Phenotype Identification in the Floral Organs of *Lycium ruthencium*

Diploid seeds of *L. ruthencium* from Xinjiang Province were disinfected with sodium hypochlorite and planted in sterile Murashige and Skoog (MS) medium for 30 days. The 4-week-old leaves were used for colchicine chromosome doubling into autotetraploid. The diploid and autotetraploid seedlings that had

grown homogeneously were planted in a mixture of sterilized turfy soil and vermiculite at a ratio of 3:1 under a 16-h light/8-h dark cycle at a temperature of 24 ± 1°C in the greenhouse for 14 days. The flowering time of tetraploid and diploid flowers planted in Ningxia under consistent growth conditions was observed. About 0.2 g true leaves of different ploidy *L. ruthencium* was collected as sample for sequencing. We selected a typical individual to serve as a biological replicate. A total of three replicates were obtained. Six samples were immediately frozen in liquid nitrogen after harvest and stored at −80°C for further use.

RNA Isolation and miRNA Sequencing

Total RNA was extracted using RNAprep Pure Plant Kit (TIANGEN Biotech Co., Ltd., Beijing, China) following the manufacturer's recommendations. The RNA was monitored by 1% agarose gel electrophoresis for RNA quality, and the concentration was detected using NanoDrop 2000. After the ends of the RNA are ligated to the adaptor, the library was subjected to reverse transcription by reverse transcriptase using RT Primer. After PCR amplification on the library, polyacrylamide gel electrophoresis (PAGE) was used to recover a small RNA library of about 150 bp to generate a stable library for sequencing. The stable small RNA library was mixed, denatured, and added to the Illumina HiSeq X Ten sequencing platform (Illumina Inc., San Diego, CA, United States) for high-throughput single-read sequencing (Cao et al., 2018).

Data Filtering and Statistics of sRNA

The reads of low-quality or contaminants sequence were filtered, and the reads containing polyA/polyT and outside the miRNA length range (15–34 nt) from high-throughput raw data were removed using cutadapt (Martin, 2011) to prevent the data quality and subsequent analysis being affected. Repeated reads were combined, and the distribution of reads in each sample was calculated. Based on clean reads small RNA (sRNA) sequence types, number and length distribution. The small RNA sequences were searched and annotated by BLASTN[1], and the Rfam database[2], which integrates ncRNA family sequences, can remove possible rRNA, tRNA, and snRNA with Bowtie (Langmead, 2009) and classify the small RNA annotated or predicted as mature miRNAs to analyze subsequent experiments.

MiRNA Analysis

The remaining sRNA tags without rRNA, tRNA, snRNA, and snoRNA are compared with the mature miRNA sequences of other species through BLAST and miRbase ([3], version 22) databases considering maximum two mismatches or gaps to identify the known mature miRNA in the sequencing data (Cagirici et al., 2021). Sequences that have not been identified as known miRNAs are mapped to the reference genome of homologous species by Bowtie 2, and then, the MIREAP[4] was used to predict novel miRNAs, which was recognized by the defining characteristics such as the secondary structures (Langmead and Salzberg, 2012). Simultaneously, the miRNA count, length, and nucleotide bias at each position were calculated for each candidate novel miRNA.

[1]http://blast.ncbi.nlm.nih.gov/Blast.cgi

[2]http://www.sanger.ac.uk/software/Rfam

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

[4]http://soxircefoige.net/projects/mireap/

Analysis of Differentially Expressed miRNA

The reads number of each miRNA in all samples was analyzed statistically, and the expression was normalized by the reads per million (RPM) value to the same magnitude:

$$RPM = \frac{each\ miRNA\ reads}{all\ miRNA\ reads}$$

The DEGseq R package was used to identify differentially miRNAs expression between the two samples based on the significant principle that expression | fold change (FC)| > 2 and $p < 0.05$. Statistically tested values were adjusted by the false discovery fate (FDR). Differential miRNA expression was visually displayed *via* volcano plot and cluster heatmap by scatter plot and MEV software.

Target Prediction and Enrichment Analysis

miRanda and RNAhybrid as the classical algorithms for identifying sRNAs' target loci were used for all miRNAs detected in this study. miRanda considers sequence complementarity, free energy of miRNA targets double strand, and cross-species conservation of target sites. RNAhybrid can be used to analyze the sequence complementarity, target site abundance, and minimum free energy (MFE). MFE ≤ -22 kcal/mol and $p \leq 0.1$ were used for RNAhybrid analysis. The sequence complementarity score = 145 and energy = -10 kcal/mol were parameters for miRanda analysis (Ye et al., 2018). The target genes of whole differentially expressed miRNA were putatively predicted using miRanda and RNAhybrid, respectively, and the intersection was taken as the ultimate target prediction results. Gene Ontology (GO) is an internationally standardized gene function classification system that provides a dynamically updated standard vocabularies to comprehensively describe the attributes of genes and products in organisms. The GO term was taken as the unit and the hypergeometric test applied to find out the GO term with a $p < 0.05$ compared with all expression backgrounds and defined it as a GO term that is significantly enriched in the target gene.

Validation of miRNA and Target mRNA Expression Using qRT-PCR

Total RNA was extracted using a plant polysaccharide polyphenol RNA kit (TIANGEN Biotech Co.). The miRNA RT/qPCR Detection Kit (Aidlab Biotechnologies Co., Ltd., Beijing, China) was employed to carry out the poly(A) tailing and reverse transcription reaction on the 3′ end of the extracted total RNA. The quantitative real-time PCR (qRT-PCR) analysis of transcriptome was performed using the 2 × SYBR® Green qPCR Mix Kit (Aidlab Biotechnologies Co., Ltd., Beijing, China).

A three-step method was performed for fluorescence quantitative PCR detection using an ABI PRISM 7500 Real-Time PCR system (Applied Biosystems, Foster City, CA, United States). The relative expression of interesting miRNA and DEGs was normalized by actin. The real-time PCR data were analyzed by the $2^{-\Delta\Delta CT}$ method.

RESULTS

Autotetraploid Exhibits Delayed Flowering in *Lycium ruthencium*

To observe whether the flowering time of *L. ruthenicum* is different from that of diploid after chromosome duplication, we found a ubiquitous phenomenon existing in flowering plants; the floral transition-induced tetraploid was delayed about 7–10 days compared with diploids, which is the same as *L. ruthenicum*.

Summary of Small RNAs in Different Ploidy Black *goji*

To explore the role of chromosomal replication events in the miRNA regulatory network of *L. ruthenicum*, six libraries of small RNA derived from diploid and autotetraploid leaves were sequenced (PRJNA727809). In summary, more than 9.0×10^6 raw reads were drawn from six small RNA libraries performed by Illumina sequencing technology. After filtering low-quality sequences, more than 8.40×10^6 high-quality clean reads were obtained, ranging from 8.40×10^6 to 1.44×10^7 for each sample, accounting for 91.80–94.90% (**Table 1**). Generally, the peaks of the length distribution can help us to determine the types of small RNAs, so as to make statistics on the types and quantity distributions of unique reads between two samples. Statistics on the length of small RNAs according to the sequencing results displayed that the lengths ranged from 15 to 34 nt, with peaks at 21 and 24 nt, and the most abundant miRNA length was mainly concentrated at 21 nt (**Figures 1A,B**). The base analysis of different positions of miRNA found that the different positions showed obvious bases preference. When miRNA precursors were transformed into mature ones, the specificity of the Dicer restriction site makes the first base of the mature sequence have a strong preference for U, and the occurrence probability is > 90% (**Figure 1C**). Analyzing the first position of miRNA with a length of 18–26 nt showed that, except for the total number of miRNA sequences of 18 nt, 19 nt, 25 nt, and 26 nt was too few to be statistically significant, the initiation preference for other miRNAs of different lengths had certain differences. The 20–23 nt miRNA is prone to U, especially that the probability of U base appearing in the first position of 21 nt length is as high as 60%, and the first base of miRNA of 24 nt is prone to appear to be A base (**Figure 1D**). From the perspective of sRNA classification and annotation, the number of unannotated small RNAs in all samples far exceeds the annotated small RNAs, and the rRNAs, with a determination rate remaining at about 19%, were the most abundant noncoding RNAs followed by other sRNAs, tRNA, and snRNA. The abundance of known miRNAs was higher than that of putative novel miRNAs (**Supplementary Table 1** and **Figure 1E**).

Conserved and Novel miRNA Identification

To analyze miRNAs in subsequent experiments, we first compared the miRBase database and performed bioinformatics predictions to identify mature miRNAs and sorted out known and novel miRNAs. Considering that *L. ruthenicum* has no reference genome, whole plant miRNA sequences in miRbase have been extracted for comparison and quantitative analysis, and the results revealed that 646, 707, 807, 658, 655, and 659 known miRNAs were identified, respectively, totaling 1,309 miRNAs and 706 novel miRNAs (**Table 2**). There were 1,309 known miRNAs from 735 miRNA families, of which 604 families have only one member, accounting for 82.18%. The species distribution statistics of the identified known miRNAs illustrated that the types of the known miRNA distributed in *Solanum tuberosum* (Stu) are the most abundant, approximately 119 categories, indicating that the miRNA sequence of *L. ruthenicum* has the closest homology relationship with the *S. tuberosum* (**Figure 2A**). Novel miRNAs were expressed at a relatively low level compared with conserved miRNAs. A total of 248 and 290 novel miRNAs were specifically expressed only in *L. ruthenicum* tetraploid and diploid, respectively (**Figure 2B**).

Differentially Expressed miRNAs

For exploring the miRNA expression pattern after chromosome doubling in *Lycium*, the normalized miRNA expression levels of diploid and tetraploid plants were compared. A total of 100 differentially expressed miRNAs including 89 known and 11 novel miRNAs were screened to be significantly different among different ploidy plants according to the principle of $|\text{fold change (FC)}| > 2$ and $p < 0.05$ (**Figure 3A**). Thirty-two and 68 miRNAs play an up- and downregulated role in the growth and development of *L. ruthenicum*, indicating that the function of most miRNAs is inhibited after chromosome

TABLE 1 | Statistics of small RNA sequencing in different ploidy samples of *L. ruthenicum*.

Sample	Raw Reads (M)	Raw Bases (G)	Clean Reads (M)	Clean Bases (G)	Clean Q20 (G)	Clean Q30 (G)
Diploid_2	9.158	1.383	8.406 (91.8%)	0.194 (14.0%)	0.194 (99.9%)	0.193 (99.5%)
Diploid_5	11.357	1.715	10.729 (94.5%)	0.246 (14.3%)	0.246 (99.9%)	0.245 (99.7%)
Diploid_6	15.199	2.295	14.423 (94.9%)	0.346 (15.1%)	0.346 (99.9%)	0.345 (99.7%)
Tetraploid_2	11.185	1.689	10.172 (90.9%)	0.233 (13.8%)	0.233 (99.9%)	0.232 (99.7%)
Tetraploid_3	12.821	1.936	11.876 (92.6%)	0.275 (14.2%)	0.275 (99.9%)	0.274 (99.6%)
Tetraploid_5	10.373	1.566	9.534 (91.9%)	0.209 (13.3%)	0.209 (99.9%)	0.208 (99.8%)

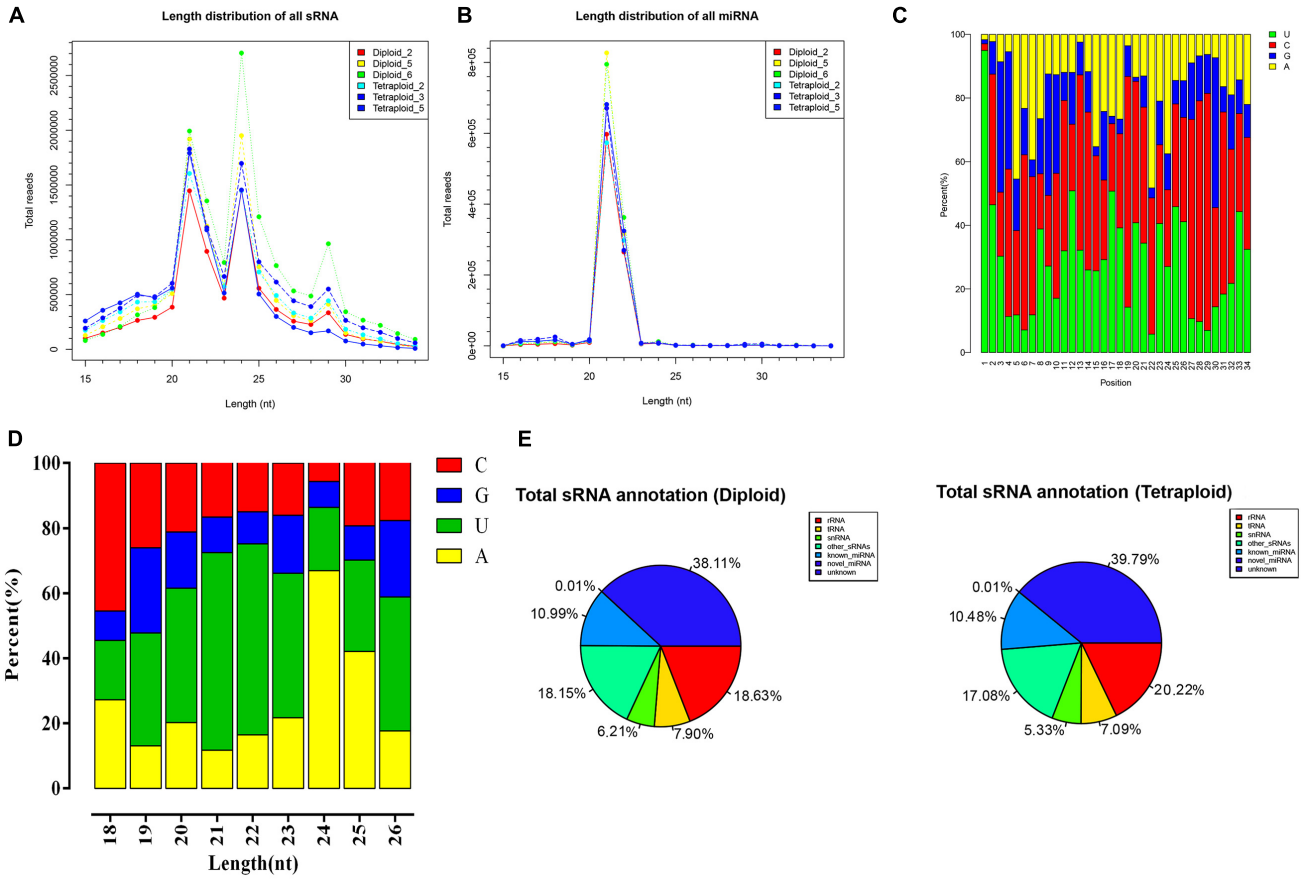


FIGURE 1 | The length distribution and base preference of small RNAs in different ploidy *L. ruthenicum*. **(A,B)** The length distribution of sRNA and mRNA of different ploidy *L. ruthenicum*. **(C)** The base preference analysis of different positions of all miRNA. **(D)** The base preference analysis of first position in miRNA with 18–26 nt length. **(E)** The percentage graph of sRNA annotation status of each sample.

doubling (**Figure 3B**). Only novel m0925-5p was upregulated in the novel miRNAs, others were downregulated, and three of them even were not expressed in tetraploid (**Supplementary Table 2**). The gene expression trend reflects the potential function, and the miRNAs expression levels in all samples were statistically quantified and normalized using the fragments per kilobase per million (FPKM). Differentially expressed miRNAs were analyzed by hierarchical cluster analysis, in which miRNAs with the similar expression patterns were clustered. By clustering 100 differentially expressed miRNAs, the distribution statistics found that the differentially expressed miRNAs showed a variety of expression trends. For example, the different members of miR408 family were clustered together, showing the same trend; the expression level of three diploid varieties was extremely high, while that of tetraploid varieties was significantly lower (**Figure 3C**).

Prediction and Annotation of miRNA Target Genes

Plants can exercise their regulatory function by the complementation of miRNA and mRNA. Based on the mechanism of plant miRNA, an efficient method to find the target gene of predicted miRNA and corresponding gene sequence is to use miRanda and RNAhybrid software. A total of 775 putative target genes were predicted *via* 100 differential miRNAs, of which 57 target genes corresponded to upregulated miRNAs, and 718 target genes were downregulated. Among the miRNAs mainly involved in the florescence regulation, miR156 and miR172 showed significant antagonistic expression, and the corresponding target genes were *SPLs* and *AP2*, respectively (**Table 3**).

TABLE 2 | Classification of known and novel miRNAs in different ploidy black goji.

Type	Known miRNA	Novel miRNA	Total miRNA
Diploid_2	646	148	794
Diploid_5	707	184	891
Diploid_6	807	288	1095
Tetraploid_2	658	185	843
Tetraploid_3	655	187	842
Tetraploid_5	659	162	821
Total	1309	706	2015

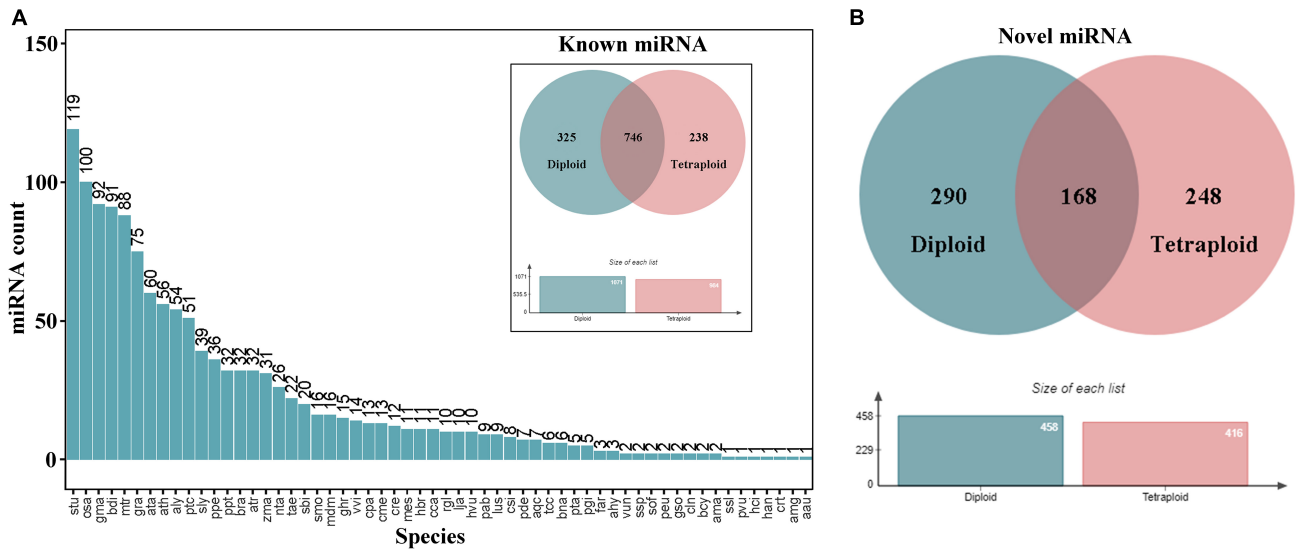


FIGURE 2 | Species distribution and expression of different ploidy miRNAs. **(A)** The species distribution of all miRNAs. **(B)** The expression of known and novel miRNA in different ploidy *L. ruthenicum*.

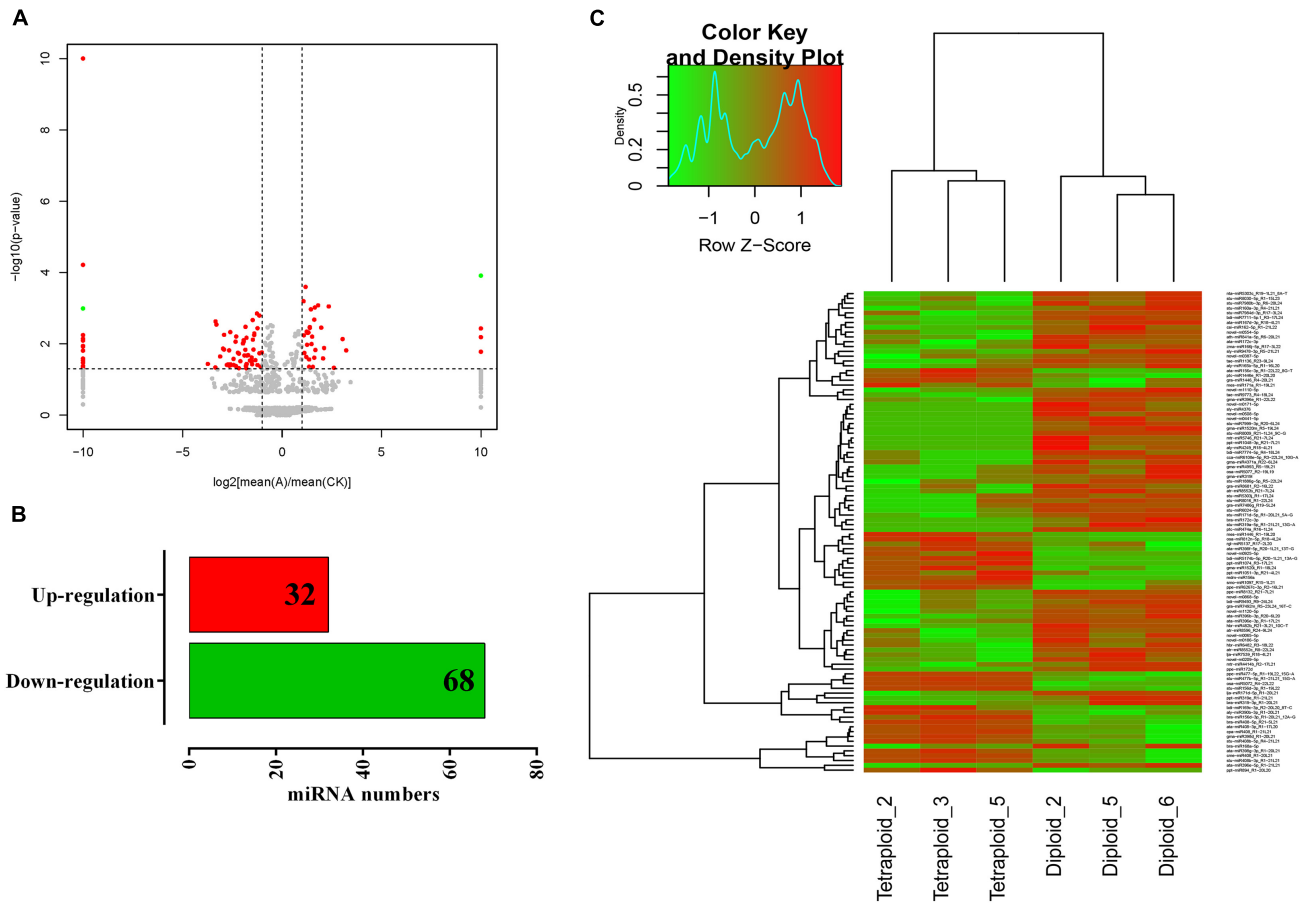


FIGURE 3 | Differential miRNA screening and expression cluster analysis. **(A)** The volcano map of differentially expressed genes in diploid and tetraploid *L. ruthenicum*. **(B)** The number of up- and downregulated genes of differentially expressed genes. **(C)** The cluster heat map of differentially expressed genes. Green indicates the downregulation, and red indicates the upregulation.

TABLE 3 | The differential miRNAs and target genes that affect the flowering time.

miRNA ID	miRNA family	Log_2 (fold change)	Target gene
ata-miR156c-3p	miR156	1.485	Squamosa promoter-binding-like protein
bra-miR156d-3p		2.119	
stu-miR156d-3p		1.188	
mdm-miR156s		Inf	
ata-miR172c-3p	miR172	−1.528	APETALA 2
bra-miR172c-3p		−Inf	TOE3
ppe-miR172d		−Inf	

TABLE 4 | The terms of GO enrichment by differential miRNA.

Category	Term	*p* value
GO:0006351	Transcription, DNA-templated	0.00004
GO:0003774	Motor activity	0.00006
GO:0006355	Regulation of transcription, DNA-templated	0.00008
GO:0051645	Golgi localization	0.00082
GO:0060151	Peroxisome localization	0.00082
GO:0014823	Response to activity	0.00093
GO:0032982	Myosin filament	0.00093
GO:0051646	Mitochondrion localization	0.00097
GO:0010321	Regulation of vegetative phase change	0.00117
GO:0006030	Chitin metabolic process	0.00118
GO:0080155	Regulation of double fertilization	0.00142
GO:0048653	Anther development	0.00152
GO:0009856	Pollination	0.0019
GO:0016459	Myosin complex	0.00199
GO:0042425	Choline biosynthetic process	0.00199
GO:0003677	DNA binding	0.00214
GO:0030048	Actin filament-based movement	0.00239
GO:0010468	Regulation of gene expression	0.00245
GO:0030016	Myofibril	0.00275
GO:0003700	Transcription factor activity…	0.00341

To determine the main biological functions performed by the target gene through the significant enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) function, GO/KEGG enrichment analysis was conducted on the predicted target genes, respectively. In GO enrichment analysis, there were 63 upregulated terms and 165 downregulated terms. A total of the most significant first 20 terms were intercepted, manifesting that the biological functions after chromosome doubling were gathered in the plant reproduction, including the development of floral organs and plant pollination (**Table 4**).

qRT-PCR Validation of miRNA and mRNA Expression

To prove the reliability and accuracy of miRNA sequencing, we selected the same samples as the sequencing to perform qRT-PCR verification on flowering-related miRNAs and several typical members of the corresponding target genes. The results of miRNA expression changes determined by qRT-PCR showed similar trends, indicating that the miRNA sequencing results are extremely accurate. In addition, the qRT-PCR identification of the target genes involved in the regulation of flowering time confirmed that the expression of these genes in tetraploid *L. ruthenicum* changed significantly (**Figure 4**).

DISCUSSION

Polyploid organisms generally display vigorous vitality and survivability, which is the reason why they easily produce in harsh environments such as glaciers and plateaus. Polyploid provides genomic deformability for the functional differences in duplicate gene, chromosome recombination, transcriptome changes, and gene dosage effects, which makes an important role in promoting species evolution (Van De Peer et al., 2017). As the outstanding abilities of polyploid are continuously excavated, breeders aim to attract the high-quality traits and artificially formulate polyploid germplasm, expecting to broaden the selectable population and promote the breeding industry development in China. Despite that the research location of polyploidy has been re-established, the molecular mechanism of isogeny polyploid traits with the same subgenome is still lacking.

Autotetraploids have been permanently created for many plant species in nature, including Salicaceae (Guo et al., 2017; Braynen et al., 2021), Cruciferae (Storme and Geelen, 2019), Zizyphus (Gu et al., 2005), and Actinidia (Wu et al., 2012). The majority of innovative traits formed by autotetraploid is mostly reflected in the gigantism caused by the genetic multiplication, the nutrients utilization, and the adversity adaptability. The research on the flowering time of autotetraploid is relatively scarce. In angiosperms, the regulation of floral organs and flowering period directly affects the fruits yield and the reproduction of offspring, which are vital to the life course of plants (Ionescu et al., 2017). In our study, we used the *L. ruthenicum* autotetraploid created in our laboratory, which exhibited a consistent phenotype the same as most tetraploids in vegetative growth, such as slower growth, larger and thicker leaves, and thicker stem (Rao et al., 2019). Compared with the first flowering period in spring with the diploid, it was found that flowering of tetraploid was universally delayed, indicating that chromosome doubling affects the vegetative growth and reproductive behavior of plants. The conclusion that chromosome doubling induced delayed flowering fully provides a reference for avoiding the decline of the first stubble *L. ruthenicum* fruit quality under the effect of late spring frost.

The continuous upgrading of high-throughput sequencing technology can achieve pivotal transcripts, which has become a broad-spectrum method for exploring plant biomolecules (Ma et al., 2015). The expression of miRNAs has been found to be altered in different plant growth and adverse circumstances,

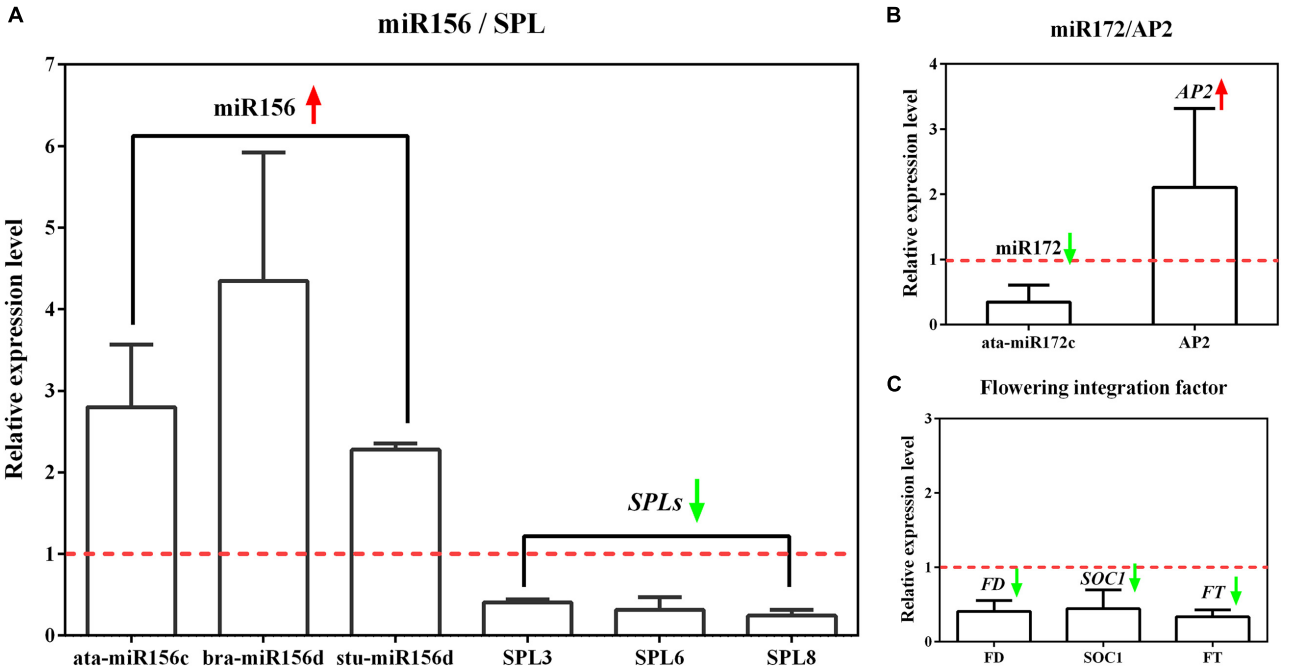


FIGURE 4 | The relative expression of miRNA related to flowering time control and its target genes by qRT-PCR analysis. **(A)** The expression analysis of differential miR156 members and target gene SPL. **(B)** The expression analysis of differential miR172 members and target gene AP2. **(C)** The expression analysis of flowering integration factor FT, FD, and SOC1. The relative expression level represents the tetraploid compared to diploid under the premise that the expression level of diploid is equal to 1.

which helps shed light on plant internal mechanism from a new perspective (Ferdous et al., 2015; Muslu et al., 2021). We applied high-throughput sequencing to consider how chromosome doubling influences the florescence of *L. ruthenicum* at the miRNA level and eventually got the transcript and miRNA information of the autotetraploid and diploid ancestor. A total of six sRNA complementary DNA (cDNA) libraries of different ploidy *L. ruthenicum* were obtained, and more than 8.40×10^6 high-quality clean reads were constructed. To study miRNA function, it is important to directly find the corresponding regulated target genes. Based on libraries screening, 100 differentially expressed miRNAs and 775 putative target genes of different ploidy were obtained according to the FDR standard, confirming that when the chromosome ploidy changes, the miRNA expression level in plants will have a tremendous change. The selected miRNAs were mainly concentrated in the plant reproduction by GO enrichment, including the development of floral organs and plant pollination, which further demonstrates that miRNAs play a vital role in chromosome doubling to regulate the delayed flowering phenotype of *L. ruthenicum*.

Some of the multitudinous screened differential miRNAs are significantly expressed and have momentous significance in the flowering regulation, such as miR156 and miR172. miR156 and miR172 belong to the miRNA family of extremely conserved and functional homeostasis. The expression of miR156 and miR172 in tetraploid compared with diploid showed a trend that the former was significantly upregulated and the latter was significantly downregulated. miR156 and miR172 manifested antagonistic regulation in *Arabidopsis* and some xylophyta plants. The former delayed flowering, and the latter promoted flowering, indicating that miR156 and miR172 are the fundamental miRNA families affecting late flowering of tetraploid. Researchers found that miR156 and target genes are key regulators to mediate the age pathway, and the content resolves the plants physiological age (Roussin-Leveillee et al., 2020). miR156 mainly regulates the physiological processes of plant growth, morphogenesis, flowering phase, secondary metabolism, and stress response through the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factor. SPL is a kind of plant-specific transcription factor that plays an important regulatory role in plant growth and development and also is a regulatory hub in the process of flower development (Wang et al., 2009). The expression of most SPLs is regulated by miR156/157 and has corresponding recognition sites. miRNA and transcriptome association analysis showed that miR156 and miR172 mainly targeted SPL3 and AP2 transcription factors. SPL3 can promote the mRNA expression of FT by combining with the GTAC element on the FT promoter and complete the transition from vegetative growth to reproductive growth (Jung et al., 2016). miR172 controls flowering time and organ formation by targeting AP2 transcription factor as FT repressor (Li et al., 2019). The protein product encoded by the FT gene is a flower-forming hormone that can be transported over long distances and is an important integration factor for the regulation of plant flowering.

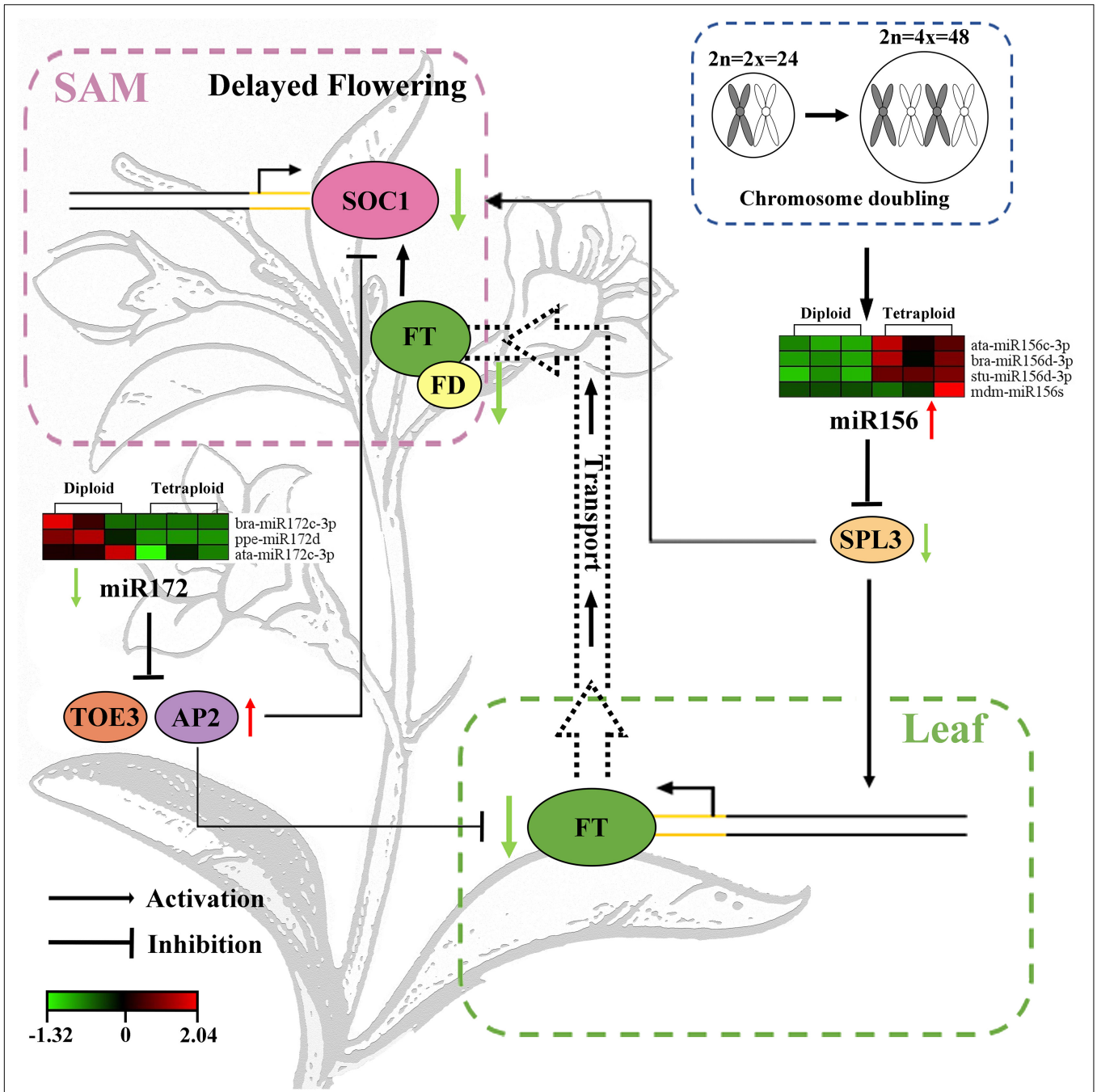


FIGURE 5 | Molecular mechanism model of flowering delay regulated by chromosome doubling in *L. ruthenicum*. The green and red colors indicate low and high expression level, respectively. The left and right columns indicate diploid and tetraploid, respectively.

As a mobile flower formation signal, FT can combine with FLOWERING LOCUS D (FD) in the shoot apex meristem to form an FT–FD complex. The FT–FD dimer upregulates the expression of SOC1 gene and completes flowering induction (Smith et al., 2011).

To summarize, the mechanism model of chromosome doubling induced late flowering of *L. ruthenicum* by participating in the age pathway involved in miR156-SPLs and miR172-AP2, which inhibited the mRNA expression of FT in the leaves. Subsequently, FT was transferred to the SAM to inhibit the expression of the flowering integration factor SOC1, which can eventually bring about the delay of flowering time (**Figure 5**). The exploration of mechanism provided the new available means for the regulation of the flowering period in polyploidy species,

deepened the basic research of the *goji* industry, and may directly affect the crops yield.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI (accession: PRJNA727809).

AUTHOR CONTRIBUTIONS

JC designed the experiments. SR performed the experiments and analyzed the data. JC and SR wrote the manuscript. JC and YL revised the manuscript. All authors read and approved the manuscript.

FUNDING

This research was funded by the National Key Research and Development Program of China (No. 2018YFD1000607) and the National Natural Science Foundation of China (31670610).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.706930/full#supplementary-material

Supplementary Table 1 | The relative abundance of each small RNA annotation in black goji.

Supplementary Table 2 | The fold change of all miRNAs between diploid and tetraploid.

REFERENCES

Brant, E. J., and Hikmet, B. (2018). Plant small non-coding RNAs and their roles in biotic stresses. *Front. Plant Sci.* 9:1038. doi: 10.3389/fpls.2018.01038

Braynen, J., Yang, Y., Yuan, J. C., Xie, Z. Q., Cao, G. Q., Wei, X. C., et al. (2021). Comparative transcriptome analysis revealed differential gene expression in multiple signaling pathways at flowering in polyploid Brassica rapa. *Cell Biosci.* 11:17.

Cagirici, H. B., Sen, T. Z., and Budak, H. (2021). Mirmachine: a one-stop shop for plant miRNA annotation. *J. Vis. Exp.* doi: 10.3791/62430

Cao, C., Long, R., Zhang, T., Kang, J., Wang, Z., Wang, P., et al. (2018). Genome-wide identification of micrornas in response to salt/alkali stress in *Medicago truncatula* through high-throughput sequencing. *Int. J. Mol. Sci.* 19:4076. doi: 10.3390/ijms19124076

Ferdous, J., Hussain, S. S., and Shi, B. J. (2015). Role of micrornas in plant drought tolerance. *Plant Biotechnol. J.* 13, 293–305. doi: 10.1111/pbi.12318

Gu, X. F., Yang, A. F., Meng, H., and Zhang, J. R. (2005). In vitro induction of tetraploid plants from diploid Zizyphus jujuba mill. cv. Zhanhua. *Plant Cell Rep.* 24, 671–676. doi: 10.1007/s00299-005-0017-1

Guo, L. Q., Xu, W. T., Zhang, Y., Zhang, J. F., and Wei, Z. Z. (2017). Inducing triploids and tetraploids with high temperatures in Populus sect. Tacamahaca. *Plant Cell Rep.* 36, 313–326. doi: 10.1007/s00299-016-2081-0

Ionescu, I. A., Mller, B. L., and Sánchez-Pérez, R. (2017). Chemical control of flowering time. *J. Exp. Bot.* 68:369.

Jung, C., and Muller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14, 563–573. doi: 10.1016/j.tplants.2009.07.005

Jung, J. H., Lee, H. J., Ryu, Y. J., and Park, C. M. (2016). SPL3/4/5 integrate developmental aging and photoperiodic signals into the FT-FD module in arabidopsis flowering. *Mol. Plant* 9, 1647–1659. doi: 10.1016/j.molp.2016.10.014

Langmead, B. (2009). Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol.* 10:R25.

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, J. E., Goretti, D., Neumann, M., Schmid, M., and You, Y. (2020). A gibberellin methyltransferase modulates the timing of floral transition at the *Arabidopsis* shoot meristem. *Physiol. Plant.* 170, 474–487. doi: 10.1111/ppl.13146

Li, X. Y., Guo, F., Ma, S. Y., Zhu, M. Y., Pan, W. H., and Bian, H. W. (2019). Regulation of flowering time via miR172-mediated APETALA2-like expression in ornamental gloxinia (*Sinningia speciosa*). *J. Zhejiang Univ. Sci. B* 20, 322–331. doi: 10.1631/jzus.b1800003

Ma, X. X., Tang, Z. H., Qin, J. P., and Meng, Y. J. (2015). The use of high-throughput sequencing methods for plant microRNA research. *RNA Biol.* 12, 709–719. doi: 10.1080/15476286.2015.1053686

Martin, M. (2011). Cut adapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Mayfield, D., Chen, Z. J., and Pires, J. C. (2011). Epigenetic regulation of flowering time in polyploids. *Curr. Opin. Plant Biol.* 14, 174–178. doi: 10.1016/j.pbi.2011.03.008

Michaels, S. D. (2009). Flowering time regulation produces much fruit. *Curr. Opin. Plant Biol.* 12, 75–80. doi: 10.1016/j.pbi.2008.09.005

Muslu, T., Biyiklioglu-Kaya, S., Akpinar, B. A., Yuce, M., and Budak, H. (2021). Pan-genome miRNomics in *Brachypodium*. *Plants* 10:991. doi: 10.3390/plants10050991

Rao, S. P., Kang, X. Y., Li, J., and Chen, J. H. (2019). Induction, identification and characterization of tetraploidy in *Lycium ruthenicum*. *Breed. Sci.* 69, 160–168. doi: 10.1270/jsbbs.18144

Roussin-Leveillee, C., Silva-Martins, G., and Moffett, P. (2020). ARGONAUTE5 represses age-dependent induction of flowering through physical and functional interaction with miR156 in *Arabidopsis*. *Plant Cell Physiol.* 61, 957–966. doi: 10.1093/pcp/pcaa022

Smith, H. M. S., Ung, N., Lal, S., and Courtier, J. (2011). Specification of reproductive meristems requires the combined function of shoot meristemless and floral integrators flowering locus T and FD during Arabidopsis inflorescence development. *J. Exp. Bot.* 62, 583–593. doi: 10.1093/jxb/erq296

Song, X. W., Li, Y., Cao, X. F., and Qi, Y. J. (2019). MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. *Ann. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334

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

Storme, N. D., and Geelen, D. (2019). Induction and characterization of diploid pollen grains in *Arabidopsis thaliana*. *Plant Meiosis* 2061, 331–346. doi: 10.1007/978-1-4939-9818-0_23

Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

Wang, H. Q., Li, J. N., Tao, W. W., Zhang, X., Gao, X. J., Yong, J. J., et al. (2018). *Lycium ruthenicum* studies: molecular biology, phytochemistry and pharmacology. *Food Chem.* 240, 759–766. doi: 10.1016/j.foodchem.2017.08.026

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

Wu, J. H., Ferguson, A. R., Murray, B. G., Jia, Y. L., Datson, P. M., and Zhang, J. L. (2012). Induced polyploidy dramatically increases the size and alters the shape of fruit in *Actinidia chinensis*. *Ann. Bot.* 109, 169–179. doi: 10.1093/aob/mcr256

Ye, R. S., Li, M., Chen, T., Wei, X. C., Qi, Q. E., Cheng, X., et al. (2018). miRNAome, mRNAome and degradome analysis of Tibetan minipigs anterior pituitary. *Gen. Comp. Endocrinol.* 259, 104–114. doi: 10.1016/j.ygcen.2017.11.009

Zhang, S. F., Yan, S. S., Zhao, J. L., Xiong, H. H., An, P. Q., Wang, J. H., et al. (2019a). Identification of miRNAs and their target genes in *Larix olgensis* and verified of differential expression miRNAs. *BMC Plant Biol.* 19:247. doi: 10.1186/s12870-019-1853-4

Zhang, S. H., He, F., Chen, X., and Ding, K. (2019b). Isolation and structural characterization of a pectin from *Lycium ruthenicum* Murr and its anti-pancreatic ductal adenocarcinoma cell activity. *Carbohydr. Polym.* 223:115104. doi: 10.1016/j.carbpol.2019.115104

Zhu, Y., Klasfeld, S., Jeong, C. W., Jin, R., Goto, K., Yamaguchi, N., et al. (2020). Terminal flower 1-FD complex target genes and competition with flowering locus T. *Nat. Commun.* 11:5118.

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.

Copyright © 2021 Rao, Li and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 July 2021
doi: 10.3389/fgene.2021.690264

Comparative Transcriptome and Weighted Gene Co-expression Network Analysis Identify Key Transcription Factors of *Rosa chinensis* 'Old Blush' After Exposure to a Gradual Drought Stress Followed by Recovery

Xin Jia[1], Hui Feng[2], Yanhua Bu[2], Naizhe Ji[2], Yingmin Lyu[1] and Shiwei Zhao[2]**

[1] *Beijing Key Laboratory of Ornamental Germplasm Innovation and Molecular Breeding, China National Engineering Research Center for Floriculture, College of Landscape Architecture, Beijing Forestry University, Beijing, China,* [2] *Beijing Key Laboratory of Greening Plant Breeding, Beijing Institute of Landscape Architecture, Beijing, China*

OPEN ACCESS

Edited by:
Shang-Qian Xie, Hainan University, China

Reviewed by:
Jingang Wang, Northeast Agricultural University, China
Waltram Ravelombola, Texas A&M University, United States

***Correspondence:**
Yingmin Lyu
luyingmin@bjfu.edu.cn
Shiwei Zhao
zhaoshiwei@beijingbg.com

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

Received: *02 April 2021*
Accepted: *24 May 2021*
Published: *15 July 2021*

Citation:
Jia X, Feng H, Bu Y, Ji N, Lyu Y and Zhao S (2021) Comparative Transcriptome and Weighted Gene Co-expression Network Analysis Identify Key Transcription Factors of Rosa chinensis 'Old Blush' After Exposure to a Gradual Drought Stress Followed by Recovery. Front. Genet. 12:690264. doi: 10.3389/fgene.2021.690264

Rose is one of the most fundamental ornamental crops, but its yield and quality are highly limited by drought. The key transcription factors (TFs) and co-expression networks during rose's response to drought stress and recovery after drought stress are still limited. In this study, the transcriptomes of leaves of 2-year-old cutting seedlings of *Rosa chinensis* 'Old Blush' from three continuous droughted stages (30, 60, 90 days after full watering) and rewatering were analyzed using RNA sequencing. Weighted gene co-expression network analysis (WGCNA) was used to construct a co-expression network, which was associated with the physiological traits of drought response to discovering the hub TFs involved in drought response. More than 45 million high-quality clean reads were generated from the sample and used for comparison with the rose reference genome. A total of 46433 differentially expressed genes (DEGs) were identified. Gene Ontology (GO) term enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis indicated that drought stress caused significant changes in signal transduction, plant hormones including ABA, auxin, brassinosteroid (BR), cytokinin, ethylene (ET), jasmonic acid (JA) and salicylic acid (SA), primary and secondary metabolism, and a certain degree of recovery after rewatering. Gene co-expression analysis identified 18 modules, in which four modules showed a high degree of correlation with physiological traits. In addition, 42 TFs including members of *NACs, WRKYs, MYBs, AP2/ERFs, ARFs,* and *bHLHs* with high connectivity in navajowhite1 and blue modules were screened. This study provides the transcriptome sequencing report of *R. chinensis* 'Old Blush' during drought stress and rewatering process. The study also identifies the response of candidate TFs to drought stress, providing guidelines for improving the drought tolerance of the rose through molecular breeding in the future.

Keywords: drought stress, comparative transcriptome analysis, weighted gene co-expression network analysis, rose, transcription factors

INTRODUCTION

Rose is one of the most fundamental ornamental crops. With an important economic, cultural, and symbolic value, the rose has a significant market share in the floriculture industry. Cut rose flowers account for 32% of the world's total cut flower trade. The rose occupies nearly a quarter of the Netherlands floriculture market (Mordor Intelligence, 2020). However, their growth conditions are closely related to changes in the external environment. Most roses suffer from various degrees of drought, which result in a sizable loss of productivity and quality around the world. As a kind of abiotic stress, drought is becoming an even more serious issue due to global warming. Changes in the earth's climate have an adverse impact on plant growth, crop production, and distribution (Zhu, 2002). It is therefore pressing to identify the response mechanism in rose under the drought stress.

To resistance drought stress, plants have evolved several adaptive mechanisms depending on the activation of the molecular networks involved in stress perception, signal transcription and transduction, expression of specific stress-related genes, and physiological reaction of tolerance (Chaves et al., 2003). Damage to plants under drought stress is manifested in the peroxidation caused by the accumulation of active oxygen in the plant (Niu et al., 2013). The role of various antioxidant protection enzymes in plants under drought stress has been extensively studied. In addition, malondialdehyde (MDA) is an important indicator of cell stability (Fang et al., 2018). Studies have shown that drought stress leads to different trends in plant antioxidant systems (Zandalinas et al., 2017; Niu et al., 2018). The activity of four antioxidant enzymes fluctuate while the MDA content continues to increase in Herbaceous peony with drought stress (Niu et al., 2018). Drought stress accumulates amounts of osmotic adjustment substances in plants such as proline and soluble sugars. In research on the drought stress of rose, it was found that as the drought stress deepened the content of proline and soluble sugar increased (Gadzinowska et al., 2019; Adamipour et al., 2020; Al-Yasi et al., 2020). It is crucial to identify the genes and biological pathways involved in drought stress tolerance and improve the understanding of the molecular mechanisms relating to drought resistance in plants (Valliyodan and Nguyen, 2006). A large number of signaling metabolic pathways and numerous genes have been determined in several model plants exposed to drought environments, such as Arabidopsis (Huang et al., 2018), rice (Liu et al., 2018), and so on. Transcription factors (TFs) control the regulating downstream stress response genes that play essential roles in various abiotic stress response processes (Joshi et al., 2016). During signal transduction, TFs directly regulate the expression of related genes by acting as a molecular switch. These TFs interact specifically with *cis*-elements located in the promoter region of the regulator (Franco-Zorrilla et al., 2014). As the plant drought tolerance network is extremely complex, this study started by exploring the transcriptional regulatory network of TFs for drought resistance. To date, *RcLEA* (Zhang et al., 2014), *RcMYBPA2* (Li et al., 2019c), *RhMYB96* (Jiang et al., 2018), *RcNAC3* (Jiang et al., 2014), and *RcNAC31* (Ding et al., 2019) genes have been reported to improve the abiotic stress resistance of the transgenic plant. However, information on the mechanism of rose response to drought stress is still limited.

As a deep sequencing technology, transcriptome technology has been successfully used to explore the gene network of non-model plant species in response to drought stress. Drought-responsive transcriptome studies have been conducted in ornamental plants such as chrysanthemum (Xu et al., 2013), tree peony (*Paeonia section* Moutan DC.) (Zhao et al., 2019a), and *Dianthus spiculifolius* (Zhou et al., 2017). Previously, only a suppression subtractive hybridization library was constructed for the dehydration stress of cut rose (*Rosa hybrida* cv. Samantha) petals, which contained 3513 unique expression sequence tags, and its expression profile during the dehydration cycle was analyzed (Dai et al., 2012). Although the transcriptome of *R. chinensis* 'Mutabilis' in the three stages of drought stress has been reported (Li et al., 2021a) and provided some basis, drought stress generally happens continuously, and it is more essential to discover how the rose recovers after drought stress. In other words, knowledge about the gene co-expression network involved in rose drought and recovery is very limited. WGCNA has become an important tool to identify gene co-expression related to associating with function, which has been widely used in the research of horticultural plants (Lu et al., 2019; Liu et al., 2021).

Rosa chinensis 'Old Blush' originated from China and has participated in modern rose hybrid breeding. *R. chinensis* 'Old Blush' is the diploid species in Rosa, which may represent a convenient genomic model for Rosa research and may be used as a model plant to study rose response to drought stress. Currently, the genome of *R. chinensis* 'Old Blush' was published, providing a valuable database for further function and genomic research in rose (Raymond et al., 2018). To obtain unique insights into the molecular mechanisms at the transcriptome level in response to drought and recovery, this study compared transcriptome profiles of *R. chinensis* 'Old Blush' at continuous drought stages (30, 60, 90 days after full watering) and rewatering, and constructed co-expression modules using WGCNA. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis in differentially expressed genes (DEGs) were performed on the modules with significant physiological data and the key TFs in the module were identified. The study aims to obtain an understanding of the drought responsive regulatory network at the molecular level and identify the key TFs. Furthermore, the results of this study can provide the basis for a better understanding of not only other *R. hybrida* cultivars but also other woody plants responding to drought stress, and odder candidate genes for further rose resistance research.

MATERIALS AND METHODS

Plant Materials

Two-year-old healthy cutting seedlings of the *R. chinensis* 'Old Blush' used in the experiments were collected from the Beijing Institute of Landscape Architecture (Beijing, China). The plant material was planted in the pot with an inner size of 26 cm in diameter, 22 cm in height, and 18 cm in diameter at the

bottom. It contained sterilized soil with a mixture of nutritive soil, garden soil, perlite, and vermiculite with the ratio of 5:3:1:1 (v/v). The size and height of plant materials used in all experiments were almost the same. Before drought stress, samples were cultured under the following growth conditions: 50% relative humidity, 25°C/18°C, day/night temperatures in the same artificial climate chamber in Beijing Forestry University (BJFU) (116.3°E, 40.0°N). Then, the samples were subjected to a continuously dry environment (no-watering) for 15, 30, 45, 60, 75, and 90 days. Then, after 7 days with rewatering treatment, the leaves were totally recovered. Fully watered plants were used as a control. Refer to experimental methods of drought treatment from David Gamboa-Tuz et al. (2018). At each time point, the fourth or fifth fully expanded leaves from the plants were collected, then immediately frozen in liquid nitrogen and stored at −80°C until the physiological measurement or RNA extraction. Each treatment group contained three replicates, and each sample from the three replicates was used to obtain biological replicates.

Measurement of Physiological Parameters

Before transcriptome sequencing using leaf tissue, which was collected and stored at −80°C from plants subjected to 0 to 90 days of drought stress, the relative water content of leaves, superoxide dismutase (SOD) activity, MDA, and relative soil water content (RSWC) was evaluated. RSWC was analyzed using direct drying method. The relative water content of leaves was determined by the initial fresh weight of the leaf samples, referencing from Zhou et al. (2019) with slight modifications, the samples were soaked in distilled water placed on the lab bench with dark cloth for 24 h and finally weighted, The thiobarbituric acid (TBA) reaction was used for evaluation of the MDA content as described by Heath and Packer (1968) with minor adjustments. The SOD activity was measured after the spectrophotometer by determining its ability to inhibit the photochemical reduction of nitro blue tetrazolium (NBT). Each experiment included three biological replicates. The correlation analysis for various physiological parameters used the Pearson correlation coefficient in SPSS software.

RNA Isolation, cDNA Library Preparation, and Transcriptome Sequencing

According to the manufacturer's instructions, total RNA was isolated with an Easy Spin Plus RNA Extraction Kit (RN53, Aidlab China). RNA integrity was verified using electrophoresis on 1% agarose gel. The 2100 Bioanalyzer (Agilent Technologies, United States) and ND-2000 (NanoDrop Technologies) were employed to test the quality and purity of RNA.

The RNA-Seq library was generated using the NEB#7530 RNA Library Prep Kit (#E7530, New England Biolabs). The total mixed RNA from *R. chinensis* leaves in all treatments was used for cDNA library construction. The library quality was subsequently detected through the High Sensitivity DNA assay Kit (Agilent Technologies). The three biological replicates from each experiment were sequenced using an Illumina HiSeq^{TM} 4000 platform by Genedenovo Biotechnology, Co., Ltd. (Guangzhou, China).

To investigate genes involved in the transcriptome, a cDNA sample was prepared from an equal mixture of t all RNA. The total RNA was isolated from the leaves of 2-year-old cutting seedlings *R. chinensis* from the control, drought-treated (30, 60, 90 days) and rehydration groups, designated as CK1, CK2, CK3, DT1-1, DT1-2, DT1-3, DT2-1, DT2-2, DT2-3, DT3-1, DT3-2, DT3-3, RW1, RW2, RW3. Three biological replicates from each group were analyzed on the Illumina HiSeq 4000 platform.

To get high quality clean reads, we were further filtered by fastp (version 0.18.0). The parameters were as follows: first removing reads containing adapters, and then removing reads containing more than 10% of unknown nucleotides (N), finally, removing low quality reads containing more than 50% low quality (Q-value $\leq$ 20) bases.

Mapping to the Rose Genome and Gene Expression Quantification

HISAT2.2.4 software (Kim et al., 2015) was used to carry out comparative analysis based on the rose genome (Raymond et al., 2018). The comparison results of HISAT2 comparison StringTie (Pertea et al., 2015, 2016) were used to reconstruct the transcript, and the expression of all genes in each sample was calculated.

Identification of DEGs and Genes Co-expression Network Analysis

The Fragments Per Kilobase of Exon Per Million Mapped Fragments (FPKM) method was used to analyze the expression level of each transcript. The reference-based approach was used to map the reads of each sample assembled by StringTie v1.3.1. For each transcription region, an FPKM value was calculated by StringTie software to quantify its expression abundance and variations. Differential expression analysis was performed by DESeq2 software (Love et al., 2014) between two different groups [also by edgeR (Robinson et al., 2010) between two samples]. The genes with the parameter of false discovery rate (FDR) below 0.05 and absolute fold change $\geq$ 2 were considered DEGs. GO enrichment analyses of DEGs (Ashburner et al., 2000) were able to recognize the main biological functions. Then, KEGG (Ogata et al., 1999) pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in DEGs.

Co-expression networks were established using WGCNA (v1.47) package in R (Langfelder and Horvath, 2008). Gene expression values were imported into WGCNA to construct co-expression modules using automatic network construction function blockwise Modules with default settings, except that the power was 10, TOMType unsigned, and minModuleSize was 50. The correlation coefficient between module eigengenes and physiological data was calculated to find out significant modules, and the hub TFs were selected from those modules (Huang et al., 2020). The networks were visualized using Cytoscape_3.3.0 (Shannon et al., 2003), GO and KEGG enrichment analyses were conducted for genes in each module.

The calculated p-value was subjected to FDR and correction, taking FDR ≤ 0.05 as a threshold.

Real-Time Quantitative PCR Verification

Total RNAs were isolated from the leaves of the treatment and controlled specimens as described above. According to the manufacturer's instructions, first-strand cDNA synthesis was performed using Prime Script II 1st strand cDNA Synthesis Kit (Takara, Shiga, Japan). Based on the manufacturer's protocol, the qRT-PCR was performed using a Bio-Rad/CFX Connect^{TM} Real-Time PCR Detection System (Bio-Rad, CA, United States) with SYBR® qPCR mix (Takara, Shiga, Japan). Relative mRNA content was calculated using the $2^{-\Delta\Delta Ct}$ method against the internal reference gene *RcPP2A* (Klie and Debener, 2011). The primers used in this study were designed with Primer Premier 5 and are listed in **Supplementary Table 12**. Three biological replicates were performed for all reactions.

RESULTS

Changes in Phenotype and Physiology at Drought Stress in *R. chinensis*

The experiment applied drought stress by stop watering. Plant leaf phenotypes were recorded and physiological indexes were measured under control condition (CK) at 15, 30, 45, 60, 75, 90 days after stress and rewatering. Leaves were dark green and shiny during normal conditions (**Figure 1A**). After 15 days of drought treatment, the leaves has no obvious morphological changes (**Figure 1B**). Similarly, the RSWC and leaf relative water content were not significantly different from the control (**Figures 2A,B**). At 30 days the leaves were slightly wrinkled (**Figure 1C**). When drought stress lasted for 45 days and 60 days, the leaves were all curled (**Figures 1D,E**). At this time the relative water content of the soil considerably decreased, reaching a moderate degree of stress (**Figure 2A**). At 60 days, the relative water content of the leaves decreased significantly compared with the control, which was consistent with the phenotype (**Figure 2B**). The curl degree of the leaf deepened and the leaf became thinner, and the relative water content of the leaf has a material decrease during 75 days of drought (**Figures 1F**, **2B**). Both leaf relative water content and soil relative water content reached the lowest value, and the leaves were severely curled and withered after 90 days of drought (**Figures 1G**, **2A,B**). After rewatering, the leaves were green and thin (**Figure 1H**). In order to determine the exact point for transcriptome sequencing, physiological indicators were also measured. The MDA content of leaves treated with drought and rewatering was determined. With the deepening degree of drought stress, the content of MDA rises in fluctuation and reached the highest point after 90 days of the drought treatment (**Figure 2D**). In addition, SOD increased with the increasing of drought treatment time and achieved

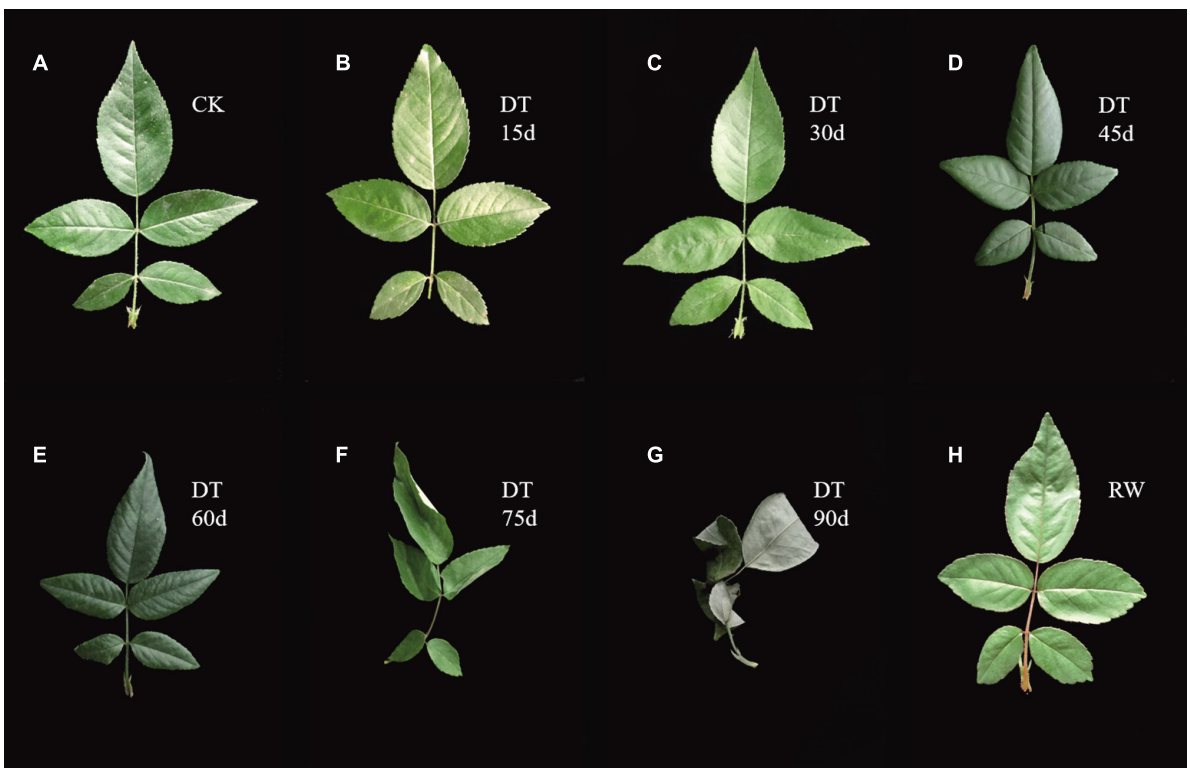


FIGURE 1 | Leaf phenotypes of *Rosa chinensis* 'Old Blush' leaves during drought and rewatering stages. **(A)** The control plants. **(B)** Drought treatment for 15 days. **(C)** Drought treatment for 30 days. **(D)** Drought treatment for 45 days. **(E)** Drought treatment for 60 days. **(F)** Drought treatment for 75 days. **(G)** Drought treatment for 90 days. **(H)** Rewatering treatment.

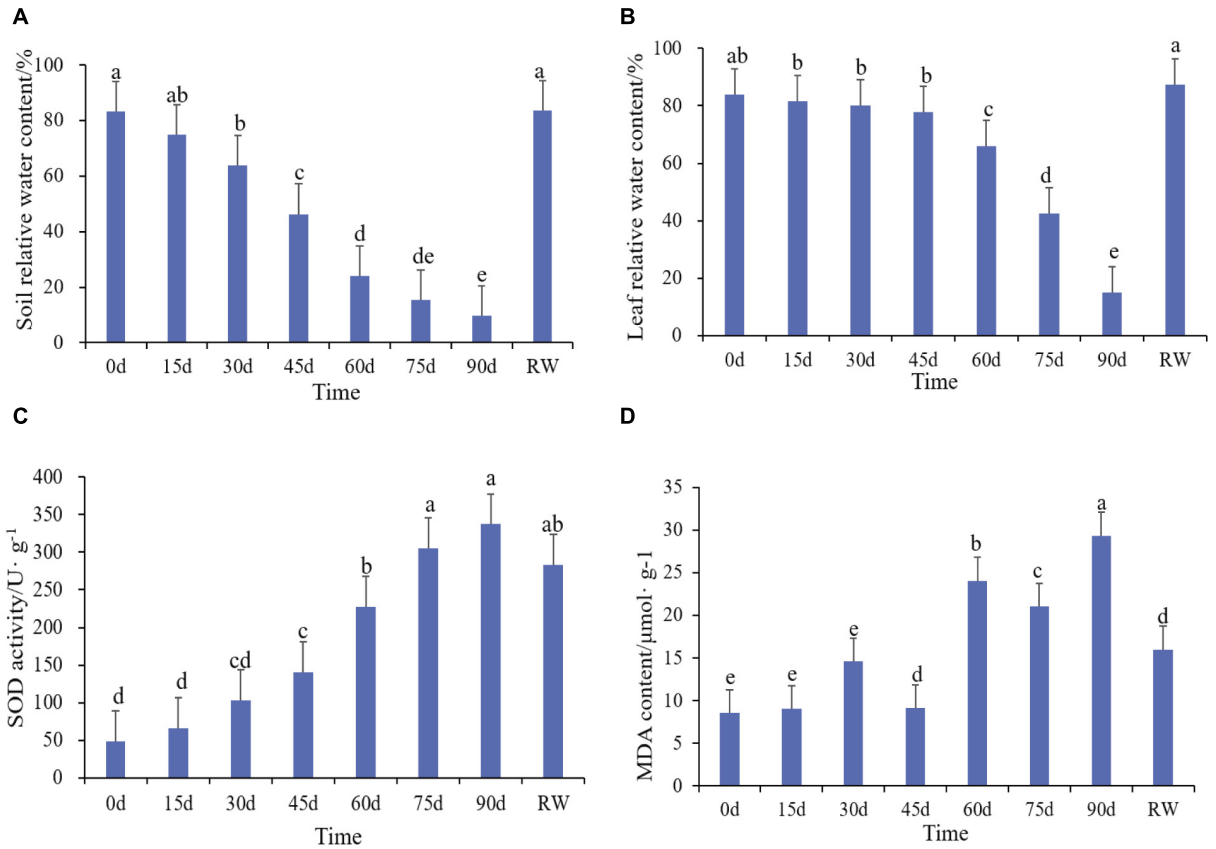


FIGURE 2 | Physiological and biochemical changes of *Rosa chinensis* 'Old Blush' under drought and rewatering treatment. **(A)** Soil relative water content. **(B)** Leaf relative water content. **(C)** MDA content. **(D)** SOD activity.

the maximum value after 90 days (**Figure 2C**). Both MDA and SOD content increased significantly during drought treatment for 30 days, 60 days, and 90 days. We conducted correlation analysis on physiological data. MDA content and SOD activity were positively correlated with different treatments and were significantly correlated at the 0.01 level (0.885, 0.979), while the relative leaf water content and soil relative water content were negatively correlated, at 0.01 significantly correlated horizontally (−0.891, −0.987). After rewatering, the MDA and SOD contents were all reduced. Based on these results, leaves were selected under normal conditions, drought treatment for 30 days, 60 days, 90 days, and rewatering treatment for transcriptome sequencing.

Transcriptome Analysis of *R. chinensis*

The clean reads were obtained by removing adaptor sequences, ambiguous nucleotides, and low-quality reads (Q-value < 20), which ranged from 41397356 to 54756474. The Q20 and Q30 values of all 15 libraries were more than 97% and 92%. The GC content was approximately ranged from 48% to 49%. Between 92.33% and 93.89% of the sequenced reads could be aligned to the rose reference genome (**Table 1**). In this study, the proportion of exon sequences ranged from 95.30% to 96.38%, and the proportion of intron sequences ranged from 1.45% to 2.66% (**Supplementary Figure 1**).

Comparisons of DEGs Under the Different Drought Stress Stages

To analyze the DEGs between leaves that suffered from drought stress at four different time duration (30, 60, 90 days and rewatering) and the control, comparisons were conducted between five groups. The DEGs were filtrated according to an expression level $|\log_2(FC)| > 1$ and FDR < 0.05 in each pairwise comparison. 6380 genes were identified exclusively at DT3, while 866 genes were specific in DT1. Especially 6380 (35.14%) genes were identified in DT3, but only 1680 (9.25%) and 1804 (9.94%) genes were identified in DT1 and DT2 (**Figure 3B**).

The number of DEGs is an increase from DT1 to DT3 and reached a peak at the CK vs. DT3 comparison set, indicating most of the drought regulated genes had a late response. The results showed DEG enrichment in CK vs. DT1, CK vs. DT2, and CK vs. DT3, illustrating the number of down-regulated DEGs was significantly higher than the number of upregulated DGEs. Meanwhile, a total of 11545 DEGs including 7886 downregulated and 3659 upregulated between CK and DT3 which contains the largest number of observations among all drought stress groups.

TABLE 1 | RNA sequencing data and corresponding quality control.

Sample name	Clean reads	Raw reads	Mapped reads	GC Content (%)	Q20 (%)	Q30 (%)
CK1	48126136	48181768	44510062 (93.27%)	48.28	97.51	92.92
CK2	49106032	49156710	45429755 (93.89%)	48.66	98.03	94.17
CK3	43793126	43853234	40291282 (93.01%)	48.29	97.48	92.88
DT1-1	45651270	45706946	42334367 (93.31%)	49.52	97.67	93.28
DT1-2	48888834	48972378	44986749 (92.70%)	48.11	97.38	92.70
DT1-3	53820314	53902378	49546849 (92.94%)	48.30	97.46	92.85
DT2-1	47463436	47521292	43862640 (93.07%)	48.19	98.12	94.31
DT2-2	54353944	54420772	49969726 (93.02%)	48.23	98.10	94.27
DT2-3	51937470	52002482	47795161 (92.33%)	48.09	98.13	94.32
DT3-1	50304612	50359618	40931525 (93.07%)	49.61	98.27	94.71
DT3-2	44365148	44421624	38029228 (93.35%)	49.58	98.10	94.28
DT3-3	49072308	49130228	41525793 (93.40%)	49.60	98.18	94.50
RW1	54756474	54838556	50746877 (93.11%)	48.65	97.99	94.00
RW2	41397356	41446080	38442300 (93.52%)	48.94	97.91	93.88
RW3	48571772	48639152	45164692 (93.24%)	48.56	98.16	94.46

Only 2503 upregulated and 2367 downregulated were observed between CK and DT1, suggesting that the differentiation of expressed genes between CK and DT3 is larger than that of CK and DT1. The result indicates that transcript abundance changed dramatically at key switched among the drought stages, which drought response genes could be induced and expressed largely. Compared with CK, there are slightly more genes upregulated than genes downregulated in RW (**Figure 3A**).

GO Functional Enrichment and KEGG Pathway Enrichment Analysis of DEGs

The drought resistance process in *R. chinensis* is complex According to GO functional enrichment analysis, as the drought stress deepened, the most enriched GO category among these DEGs was 'metabolic process,' followed by 'cellular process,' 'biological regulation,' 'regulation of biological process,' 'response to stimulus,' 'catalytic activity,' 'binding,' 'transport activity,' 'cell part,' 'cell,' 'organelle,' 'membrane,' and 'membrane part.' Meanwhile, based on the GO analysis of up-regulated DEGs, the GO terms of up-regulated DEGs in CK vs. DT1, CK vs. DT2, and CK vs. RW were mostly concentrated on membranes, including the cellular components category, such as 'membrane' (GO:0016020), 'membrane part' (GO:0044425), and 'intrinsic component of membrane' (GO:0031224) (**Supplementary Table 1**). CK vs. DT3 also involved more cells and metabolic processes including the biological process category (metabolic process, GO:0008152, cellular, GO:0009987, single-organism process, GO:0044699, cellular metabolic process, GO:0044237) (**Supplementary Table 1**). Moreover, the GO analysis of the down-regulated DEGs found that among the top 20 enriched, 13 of them belong to the category of biological processes, mainly focusing on the response to stimuli in CK vs. DT1 (**Supplementary Table 2**). It is worth noting that the GO analysis of DT3 vs. RW mainly focused on cells and cell membranes, indicating that the rose has been restored to a certain extent at the cell level (**Supplementary Tables 1, 2**).

The KEGG pathway analysis revealed that it is mainly enriched in the metabolism under the different drought degrees and rewatering treatment. Among the top 10 enriched

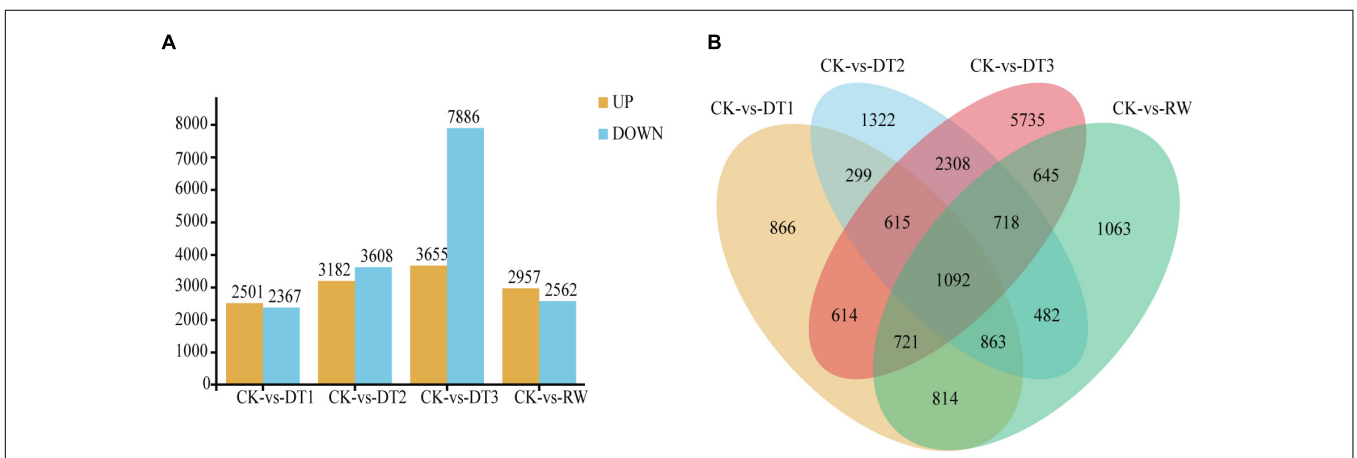


FIGURE 3 | RNA-seq data expression profiles during drought and rewatering treatment in *Rosa chinensis* 'Old Blush.' **(A)** The number of genes up-regulated and down-regulated in four drought and rewatering treatment stages compared with the control. Genes up-and down-regulated are shown as yellow and blue bars. **(B)** Venn diagram analysis of the DEGs from four drought and rewatering treatment stages compared with control plants.

pathways, 'plant–pathogen interaction' (ko04626), 'biosynthesis of secondary metabolites' (ko01110), 'plant hormone signal transduction' (ko04075), 'MAPK signaling pathway-plant' (ko04016), were regulated in response to DT1, indicating that drought stress was sensed and signaling transduction pathways were activated. However, more pathways related to metabolism were enriched at DT2 and DT3. The 'photosynthesis' (ko00195) and 'photosynthesis-antenna proteins' (ko00196) associated with photosynthesis were all enriched in CK vs. DT2 and CK vs. DT3 (**Supplementary Table 3**). Metabolite-related pathways such as 'biosynthesis of secondary metabolites' (ko01110), 'alpha-Linolenic acid metabolism' (ko00592), 'tryptophan metabolism' (ko00380), 'glycosphingolipid biosynthesis-lacto and neolacto series' (ko00601), 'alanine, aspartate and glutamate metabolism' (ko00250), 'galactose metabolism' (ko00052), and 'glyoxylate and dicarboxylate metabolism' (ko00630) were enriched in CK vs. DT2 and CK vs. DT3. These results indicated that the metabolism processes of *R. chinensis* may act as the predominant process in the middle and later stages of drought stress. This result was unsurprising, as drought causes changes in the metabolism of plants to improve their drought tolerance. Meanwhile, secondary metabolism, primary metabolism, and photosynthesis were focused in DT3 vs. RW, indicating that the metabolism and photosynthesis of the rose after rewatering recovered, and that drought stress did not cause serious damage to the rose (**Supplementary Table 3**). Apart from metabolism enrichment, 'plant–pathogen interaction' (ko04626) was also enriched in CK vs. RW, indicating that plants are vulnerable to pathogenic bacteria after recovering from drought stress (**Supplementary Table 3**).

Identification of Plant Hormone and Signal Transduction-Related DEGs

In this study, signal transduction in plants was influenced by drought stress which was verified by both the GO and KEGG enrichment analysis of the DEGs. More than 240 genes were predicted to encode protein kinases with varying expression levels in *R. chinensis*. Genes encoding receptor-like protein kinases (*RLKs*), in which LRR receptor kinases were the majority accounted for the largest proportion of these genes. Twenty-nine of them were up-regulated in DT1 and DT2. Among them, *MIK1* (RchiOBHm_Chr6g0286591) was up-regulated 8.02-fold, respectively. More than 66 *RLK* genes were included in the group of serine/threonine-protein kinases during the drought stress. Furthermore, one gene related to G-type lectin S-receptor-like serine/threonine-protein kinase showed the highest degree of upregulation in DT3. In the KEGG pathway analysis of DEGs, the 'MAPK signaling pathway' (ko04016) was induced during drought treatment. For DEGs detected in DT1 and DT2, two *MAPKKKs* including RchiOBHm_Chr2g0158821 and RchiOBHm_Chr5g0002581 and four *MAPKKs* including, *MPK9* (RchiOBHm_Chr7g0178621), *MKK10* (RchiOBHm_Chr7g0211301), *MEKK1* (RchiOBHm_Chr7 g0240941), *MPK4* (RchiOBHm_Chr7g0196781) were up-regulated. *MKK10* (RchiOBHm_Chr7g0211301) was particularly up-regulated 9.18-fold. However, three *MAPKKs* consist of *MPK3* (RchiOBHm_Chr5g0061451), *MKK4* (RchiOBHm_Chr2g0150341), and *MEKK1* (RchiOBHm_Chr1g0352161) were down-regulated. Also, *MPK4* (RchiOBHm_Chr7g0196781) and *MPK19* (RchiOBHm_Chr5g0028971) were substantially increased in DT3 (**Supplementary Table 4**).

Ca^{2+} signal transmission is achieved through the generation, decryption, and transmission of specific calcium signals and the corresponding physiological and biochemical reactions downstream. The study identified DEGs related to calcium ion, encoding calcium-binding proteins, Ca^{2+}-binding protein EF hand, calmodulin-like, calcium-dependent protein kinases, calcium-transporting ATPase, and calcium channel protein. One gene (RchiOBHm_Chr2g0158571) encoding the calcium channel protein sustained high expression under drought stress and was up-regulated 6.44-fold in DT1 vs. CK (**Supplementary Table 5**).

According to the GO and KEGG analyses of the DEGs, the GO term 'hydrogen peroxide metabolic' (GO:0042743) and 'response to reactive oxygen species' (GO:0000302) were enriched at DT1 and DT2, while peroxisome (ko04146) was induced at DT3. 69 DEGs encoded ROS production and scavenging. A total of 55 DEGs encoding enzymes related to ROS scavenging, including peroxidase (POD), ascorbate peroxidase (APX), glutathione *S*-transferase (GST), glutathione peroxidase (GPX), polyphenol oxidase (PPO), ferritin, glutaredoxin, thioredoxin, and peroxiredoxin. Among these antioxidant enzymes, *POD45* (RchiOBHm_Chr4g0400821) encoding POD was up-regulated by 4.79-fold at DT1, another gene encoding POD *PNC1* (RchiOBHm_Chr5g0072281) has no significant change from DT1 to the control, but 9.29-fold has a significant at DT2, reaching the peak of expression level at DT3 (**Supplementary Table 6**).

The expression of genes involved in hormone biosynthesis or signaling was substantially changed during the drought stress and rewatering stage. The changed genes included ABA, auxin, BR, cytokinin, ET, JA, and SA. Among these DEGs, most genes were involved in ABA response. Nine-*cis*-epoxycarotenoid dioxygenase (NCED) is a critical enzyme in ABA biosynthesis. For instance, one gene encoding *NCED3* (RchiOBHm_Chr5g0014331) was up-regulated by 2.70-fold at DT1, while another gene encoding NCED6 (RchiOBHm_Chr4g0397001) was significantly down-regulated 9.40-fold at DT3. The expression of a gene (RchiOBHm_Chr5g0049951) encoding *AAO* is also a vital enzyme in ABA biosynthesis, which was up-regulated by 2.78-fold at DT1 and down-regulated by 7.28-fold at DT3. Moreover, some genes involved in the ABA mediated signaling pathway were also induced by drought stress. Abscisic acid receptor *PYL4* (RchiOBHm_Chr1g0371101) was down-regulated. In contrast, one gene encoding *PP2C* (RchiOBHm_Chr5g0066401), a negative regulator of ABA, has an up-regulation (**Supplementary Table 7**).

A total of 24 genes involved in the auxin signaling pathway changed. Three genes encoding auxin response factors (*ARF3*, *ARF5*, *ARF9*) were drought induced. Particularly, *ARF3* (RchiOBHm_Chr5g0009381) expression continued to decline under drought stress, which was induced 4.41-fold

at DT3. At the same time, the expression level of *ARF5* (RchiOBHm_Chr6g0302551) continued to increase under drought stress. Two genes encoding *AUX/IAA* proteins were up-regulated at DT1 and down-regulated at DT3. *AUX22D* (RchiOBHm_Chr4g0389611) was lowered nearly 10-fold at DT3. Furthermore, some genes participated in auxin response and signal, for example of auxin-binding protein ABP19a-like (*ABP19A*), auxin transporter-like protein 2 (*LAX2*), and auxin efflux carrier (*PIN1C*). The gene encoding *ABP19A* (RchiOBHm_Chr2g0094781) was up-regulated 7.41-fold at DT1. Five genes encoding *SAUR* were identified, among these, *SAUR77* (RchiOBHm_Chr5g0026611) was up-regulated 9.78-fold at DT1 (**Supplementary Table 7**).

In this study, a large number of ethylene-responsive transcription factor (*ERF*) genes displayed a down-regulation trend under early drought stress. However, one gene encoding *ERF023* (RchiOBHm_Chr1g0360021) had an up-regulation 3.70-fold at DT1 and was then down-regulated 5.02-fold at DT3. Another gene encoding example, *ERF110* (RchiOBHm_Chr6g0274591) was up-regulated 7.34-fold at DT3. Meanwhile, ethylene receptor (*ERS1*), ethylene-overproduction protein 1 (*ETO1*), and ethylene insensitive3-like protein (*EIL3*) were identified (**Supplementary Table 7**).

Most of the genes associated with the biosynthesis or signaling of jasmonic acid (JA) were down-regulated. The gene encoding malate dehydrogenase (*MDHG*) was down-regulated 4.06-fold at DT3, while 12-oxophytodienoate reductase 3-like (*OPR3*) was up-regulated 4.13-fold at DT3. Additionally, the majority of genes were related to cytokinin, salicylic acid, and brassinosteroid pathways. One gene encoding cytokinin dehydrogenase 3 (RchiOBHm_Chr2g0093491) was down-regulated 7.40-fold at DT3 (**Supplementary Table 7**).

Expression of Genes Involved in Metabolism and Biosynthesis

The output of GO and KEGG enrichment analysis indicates that many DEGs were related to metabolism and biosynthesis. The 'polysaccharide catabolic process' (GO:0000272) and the 'Starch and sucrose metabolism' (ko00500) pathway were significantly induced by drought stress. Related to carbohydrate synthase and starch synthase were induced. The expression level of sucrose synthase increased, and one of the genes encoding *SUS7* (RchiOBHm_Chr3g0488651) up-regulated 4.03-fold at DT3. Only a small portion of starch synthase genes were down-regulated, others appeared to be up-regulated. Additionally, drought treatment significantly induced the expression of enzyme genes involved in starch and sugar metabolism, fructose, and mannose metabolism. Among these, genes encoding fructose-1,6-bisphosphatase, fructose-bisphosphate aldolase, mannose-1-phosphate guanyltransferase alpha, pyrophosphate–fructose 6-phosphate 1-phosphotransferase subunit beta-like and 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase were down-regulated at DT3 (**Supplementary Table 8**).

In total, 89 DEGs associated with lipid metabolism were enriched in the 'lipid metabolic process' (GO:0006629). Most of these were related to the biosynthesis and metabolic processes of lipoprotein, wax, keratin, lipid, and fatty acid. Moreover, most genes involved in wax biosynthesis were significantly up-regulated. For example, one gene encoding *CYP86B1* (RchiOBHm_Chr4g0432631) up-regulated 9.90-fold at DT3 (**Supplementary Table 9**). The KEGG functional annotation of these differentially expressed genes was related to metabolism and differentially expressed genes were also related to secondary metabolisms. A total of 95 DEGs involved in secondary metabolism were enriched by KEGG function, and the results are shown in **Figure 4**. Drought stress affected the transcription level of genes that related to phenylpropanoid biosynthesis pathways.

Photosynthesis-Related Genes Under Drought Stress

Many DEGs enriched in the category of 'photosynthesis' (ko00915) were overexpressed in DEGs down-regulated by DT2 and DT3. In the process of DT2 and DT3, many genes involved in photosynthesis were significantly down-regulated. A total of 70 DEGs involved in photosynthesis were identified. In the seam, three (*CAB40*, *LHCB5*, *psaA*) of 16 DEGs encoding the components of photosystem I (PSI) were slightly up-regulated in DT1, and other genes were significantly down-regulated during the drought treatment period. A total of 33 genes encoding PS II were identified, including 6 up-regulated genes (*psbE*, three genes encoding *psbC*, *psbD*, *PSB27-1*), and 27 down-regulated genes. Most of these genes decreased significantly in the middle and later periods of drought, indicating that PS I and PS II were suppressed under drought stress. Meanwhile, there were 15 genes encoding redox chains, all of which were significantly down-regulated in the middle and late periods of drought stress. Additionally, the expression of six gene encoding components of chloroplasts significantly decreased (**Supplementary Table 10**).

Transcription Factors Responding to Drought Stress

Transcription factors specifically binding to *cis*-acting regulatory elements in the promoter of target genes are crucial regulatory proteins, which can modulate numbers of genes up or down-regulation (Dhriti and Ashverya, 2015). In this study, *WRKY* (58), *NAC* (43), *MYB* (52), *bZIP* (19), *bHLH* (55) families contained large numbers of transcription factors. The expression level of the *WRKYs* family exhibited large changes under different treatments. The expression of some genes reached a maximum in the middle and later periods of drought. Genes encoding *WRKY75* and *WRKY71* all reached the highest expression level at DT3 with an up-regulation of more than 5-fold. *WRKY57* (RchiOBHm_Chr5g0083891) up-regulated regarding expression responses at different drought stress, especially increased 7.93-fold at DT3. However, more than half of the *WRKY* families have decreased expression levels at different drought levels. Twenty-four genes encoding *WRKYs* were down-regulated at DT1 by 1.02- to 3.71-fold (**Supplementary Table 11**).

More than half of the genes encoding *NACs* were up-regulated to varying degrees under drought stress. One gene encoding *NAC100* (RchiOBHm_Chr2g0167461) continued to increase in the middle and later periods of drought stress

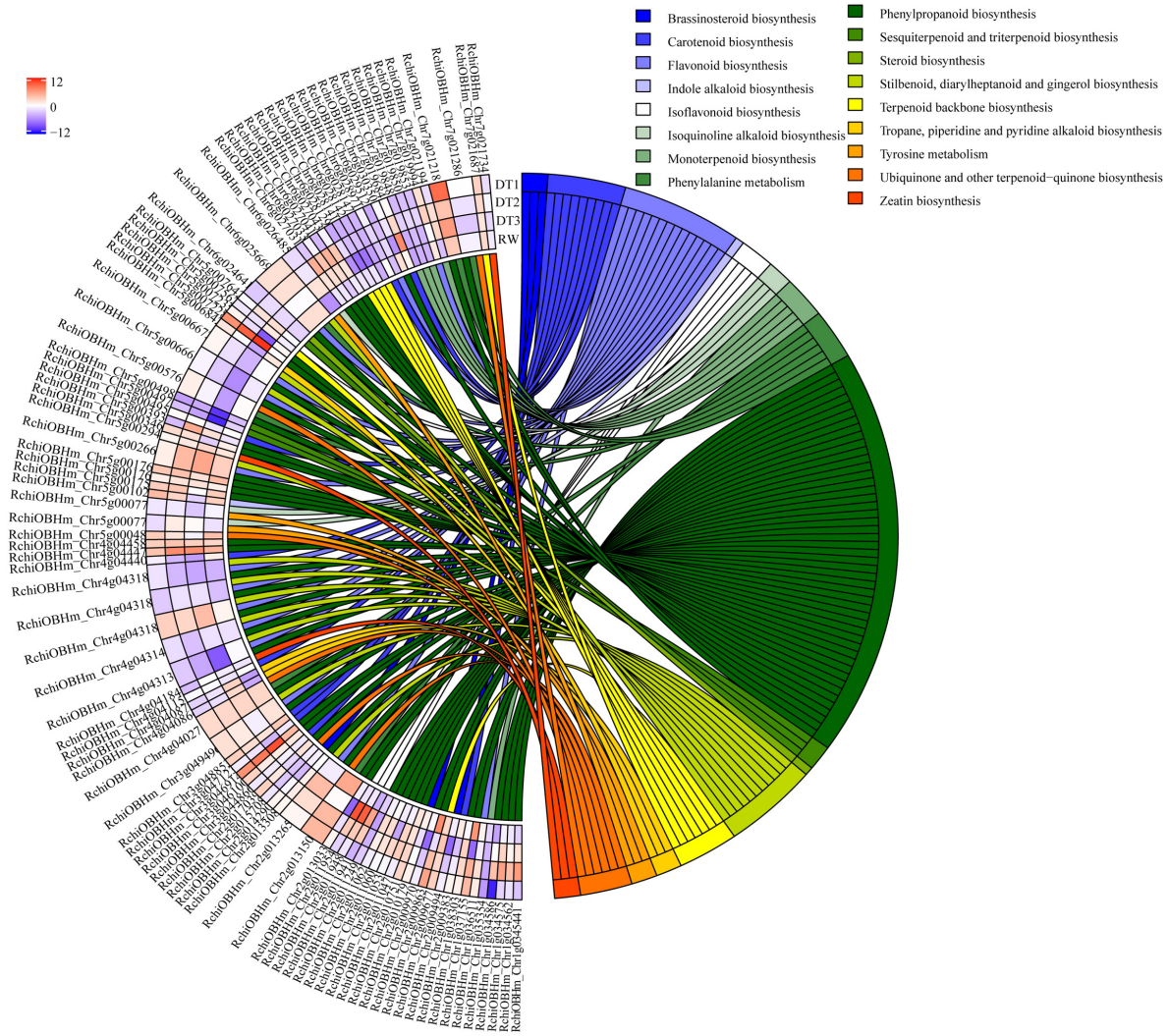


FIGURE 4 | KEGG enrichment of differentially expressed genes involved in secondary metabolism.

by 7.20-fold. A total of 11 genes upregulated during DT1. Some genes increased and then decreased under drought stress, but finally increased after rewatering. For example, *NAC073* (RchiOBHm_Chr0c19g0500091) was upregulated 5.78-fold at DT1, decreasing slightly at DT2, then down-regulated 2.64-fold at DT3, but had an increase of 6.96-fold after rewatering. *NAC037* (RchiOBHm_Chr2g0099671) had a similar expression trend, but the up-regulation decreased a small amount. On the contrary, one gene encoding *NAC072* (RchiOBHm_Chr5g0034761) was down-regulated at DT1 and continued to increase afterward (**Supplementary Table 11**).

Under different drought treatments, more than three-quarters of the *MYB* genes' expression levels showed different degrees of up-regulation. Among them, a relatively large number of up-regulated genes appeared during the DT1 period. For instance, *MYB46* (RchiOBHm_Chr1g0315931) and *MYB61* (RchiOBHm_Chr1g0315931) showed the highest up-regulation of expression with 7.09 and 7.15-fold. Moreover, some genes up-regulated significantly at DT3. *MYB75* (RchiOBHm_Chr2g0116041) displayed the most amount of up-regulation among all up-regulated genes at DT3, with an increase of 7.65-fold. Some genes showed a tendency to increase and then decrease during drought treatment, for example, gene encoding *MYB61* (RchiOBHm_Chr3g0458721) slightly increased at DT1 and then decreased by 6.10-fold. After rewatering, most gene expression did not change significantly compared with CK, only a small amount of gene expression up-regulated (**Supplementary Table 11**).

Among the three drought treatments of the *bHLH* families, the numbers of genes up-regulated in one period were more than the numbers of genes down-regulated. The number of down-regulated genes in two periods was more than that. Most genes in the *bHLH* family up-regulated during early drought treatment of DT1. While a large number of genes significantly down-regulated during the middle and later drought periods (DT2 and DT3). Among these, there was a slight upward adjustment,

up-regulated by 1.02- to 4.66-fold at DT1. Nevertheless, there were 37 genes down-regulated at DT2 or DT3. One gene encoding *bHLH72* (RchiOBHm_Chr7g0182341) down-regulated 6.94-fold. A gene *bHLH121* (RchiOBHm_Chr6g0245181) that belongs to the *bHLH* family has no significant change in gene expression during drought treatment but appeared to be up-regulated after rewatering (**Supplementary Table 11**).

Among *AP2/DREB* and *bZIP* transcription factors, more genes exhibited down-regulation under drought treatment. Nineteen members of *bZIPs* were found to be responsive to drought stress, while 12 down-regulated from 1.24- to 4.80-fold (**Supplementary Table 11**). Five genes encoding *DREBs* showed down-regulation that exceeded 6.60-fold at DT1. The most significant down-regulation occurred by *DREB1D* (RchiOBHm_Chr7g0199351), which down-regulated 9.64-fold (**Supplementary Table 11**).

Co-expression Network Construction and Identification of WGCNA Modules

Gene co-expression network gene clustering and module cutting combined genes with similar expression patterns on the same branch. Each branch represented a co-expression module with different colors representing different modules. The unexpressed genes in more than half of the samples were filtered, the expression patterns of 30,012 genes obtained from transcriptome sequencing were performed by WGCNA. A total of 18 modules were identified according to the similarity of expression patterns (**Figure 5A**). Cluster analysis was performed to evaluate genes in the modules and were observed through heat maps (**Figure 5B**). Furthermore, to identify modules that were significantly associated with different drought stress, the module-trait correlation relationships were constructed (**Figure 5C**). There were four modules including blue, navajowhite1, salmon4, and coral2 were closely associated, with traits related to four treatments (0.99, 0.99, 0.87, and 0.76, respectively). The expression pattern of the eigengenes represents the gene expression profile of the entire module, and the expression patterns of the four selected modules were analyzed (**Figure 6**).

Four modules of blue, navajowhite1, salmon4, and coral2, were annotated through GO and KEGG analysis. The blue module of DT1 was most significant in GO analysis, which was mainly enriched in signal transduction and stress response (**Supplementary Table 13**). The results of KEGG analysis were

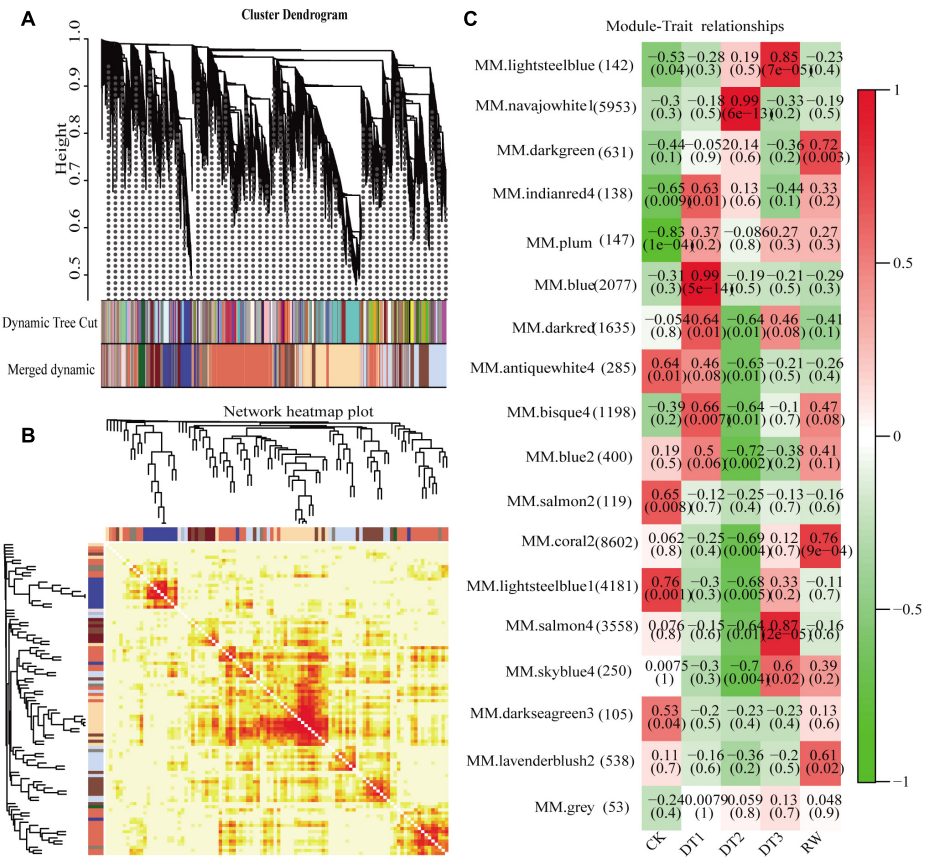


FIGURE 5 | Identification of DEGs by WGCNA. **(A)** Module level clustering diagram. **(B)** Network heatmap plot. **(C)** Module-trait associations. (Each column corresponds to different processing conditions, and each row corresponds to the characteristic gene of the module. The correlation between two is indicated in cell by Pearson correlation coefficient and p value in parentheses. Cell color ranges from red (high positive correlation) to green (high negative correlation), and the number of genes contained in each module is in left parentheses).

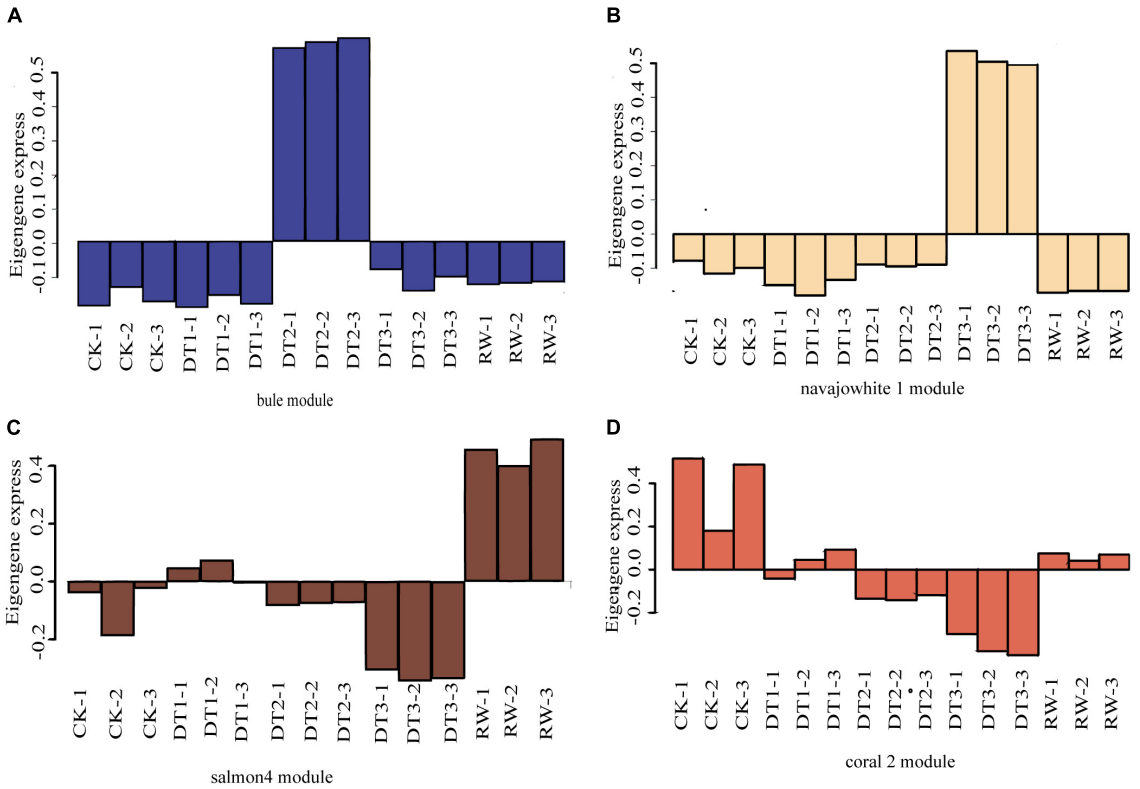


FIGURE 6 | The expression pattern of the co-expressed genes in the representative module. **(A)** Blue module. **(B)** Navajowhite1 module. **(C)** Salmon4 module. **(D)** Coral2 module.

similar, mainly focusing on signal transduction and synthesis of metabolites (**Supplementary Table 13**). The navajowhite1 module showed significant rich GO terms, indicating many genes were enriched in various stress responses. Those that were mainly enriched in 'response to stimulus' (GO:0050896), 'response to chemical' (GO:0042221), 'response to water' (GO:0009415), and 'response to water deprivation' (GO:0009414) were also significantly enriched (**Figure 7A**). KEGG enrichment analysis displayed the main enrichment in metabolic pathways (**Figure 7B**). In the salmon4 module, GO terms and KEGG pathways were enriched in metabolic progress, which indicates that the rose produces metabolites in the later drought stage to resist drought stress (**Supplementary Tables 13, 14**). However, the process of compound metabolisms was focused on the coral2 module (**Supplementary Tables 13, 14**). The GO terms about stress response are mainly concentrated in navajowhite1 and blue module. As a result, navajowhite1 and blue module genes may play a critical role in the process of rose drought stress and rewatering.

Identification of Hub TFs and Network Construction

Transcription factors are an essential class of regulatory proteins in biological processes. This study analyzed the TFs in each module. Although there was a difference in the distribution of TFs in different modules, it is mainly concentrated in *WRKY*, *NAC*, *MYB*, *ERF*, *bHLH*, *bZIP*, *C2H2,* and other transcription factor families. According to previous reports, these TFs were involved in the process of plant stress regulation (Erpen et al., 2018). TFs of navajowhite1 and the blue module were analyzed, the highly connected examples that were used as core genes were screened out. Many DEGs Identified in the navajowhite1 module were annotated as TFs, including several *WRKY*, *MYB*, *NAC*, *ERF*, *ARF,* and *bHLH* TFs (**Figure 8A**). Based on the hub gene correlation network and their high connectivity, 22 of these TFs were selected: two genes (RchiOBHm_Chr5g0034761 and RchiOBHm_Chr6g0308271) encoding *NAC072*, two genes (RchiOBHm_Chr2g0167461 and RchiOBHm_Chr7g0181811) encoding *NAC100*, *NAC087* (RchiOBHm_Chr3g0497471), *MYB33* RchiOBHm_Chr7g0228621), *MYB75* (RchiOBHm_Chr2g0116041), *MYB102* (RchiOBHm_Chr5 g0006031), *MYB78* (RchiOBHm_Chr4g0426181), *WRKY3* (RchiOBHm_Chr4g0439041), *WRKY4* (RchiOBHm_Chr2 g0166991), *WRKY6* (RchiOBHm_Chr5g0040801), *WRKY27* (RchiOBHm_Chr1g0378621), *WRKY28* (RchiOBHm_ Chr2g0151681), *WRKY71* (RchiOBHm_Chr5g0013131), two genes (RchiOBHm_Chr4g0429851 and RchiOBHm_ Chr2g0175911) encoding *WRKY75*, *ERF113* (RchiOBHm_Chr4g0428891), *ARF1* (RchiOBHm_Chr7g021 9771), *ARF5* (RchiOBHm_Chr6g0302551), *ARF8* (RchiOBHm_Chr3g0487771), and the *bHLH3*

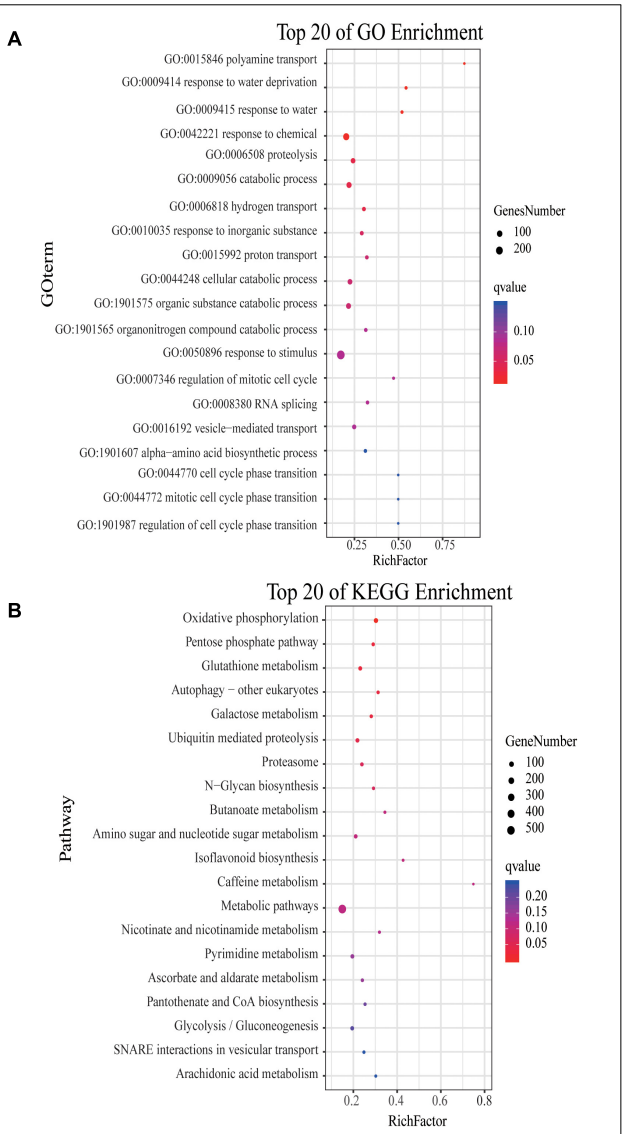


FIGURE 7 | GO and KEGG analysis of DEGs in the navajowhite1 module. **(A)** The top 20 of GO enrichment. **(B)** The top 20 of KEGG enrichment. (The *q*-value ranges from 0 to 0.05. The closer the *q*-value is to 0, the more significant the enrichment is. Genes number is the number of genes enriched in pathways).

(RchiOBHm_Chr6g0288981) (**Figure 8B**). The experiment showed the expression of these genes was significantly upregulated at DT3 except for *WRKY6*, *bHLH3*, *ARF1,* and *ARF8* (**Figure 8B**). In the blue module, 20 TFs were identified as hub genes including *ERFs*, *bHLHs*, *WRKYs*, *MYBs,* and *NACs* (**Figure 8C**). *ERF109* and *bHLH162* have higher connectivity among them (**Figure 8C**). The expression of all 20 TFs increased in DT2 (**Figure 8C**).

Validation of the DEGs by qRT-PCR

To verify the accuracy and reproducibility of the transcriptome analysis, 13 DEGs were selected to analyze the transcript abundance using qRT-PCR, including seven randomly selected transcription factors from the navajowhite1 module, four TFs, and two ABA synthesis-related DEGs. The results showed that the expression profiles detected by qRT-PCR were positively correlated with the RNA-Seq results (**Figure 9**). Therefore, the reliability of the RNA-seq data was confirmed by the consistency between the qRT-PCR results and RNA-seq analyses.

DISCUSSION

TFs Involved in the Drought Stress Response

Weighted gene co-expression network analysis is an analysis method for analyzing the gene expression patterns of multiple samples. Genes with similar expression patterns can be clustered, moreover, the relationship between modules and specific traits can be analyzed (Zhang and Horvath, 2005). WGCNA analysis can present the interaction relationship between genes, and identify the hub genes at the center of the regulatory network. It also can use the module eigengenes values to perform correlation analysis with specific traits, more accurately (Silva et al., 2019). More and more studies use WGCNA to analyze key genes in plants under abiotic stress (Shao et al., 2021). In this study, through the analysis of WGCNA and transcription regulation modules, the highly connected TFs in modules were selected. Numerous TFs were obtained in the transcriptome sequencing in this study including the *NACs*, *WRKYs*, *MYBs*, *AP2/ERFs*, *bHLHs,* and so on. Many studies have shown that *NACs* play positive roles in the regulation of plant stress resistance. In particular, Arabidopsis (Tran et al., 2004), wheat (Mao et al., 2014; Zhang et al., 2015), and rice (Hong et al., 2016; Sung et al., 2018) have received the most research attention on *NACs*, which found that overexpression of *NACs* transcription factor family genes can significantly enhance the tolerance of transgenic plants to multiple stresses. In addition, rose (*Rosa hybrida*) *RhNAC3* can improve rose petals dehydration tolerance, and can also improve the drought tolerance of transgenic Arabidopsis (Jiang et al., 2014). In the study of rose (*Rosa chinensis* 'Slater's crimson China') heat resistance, it was found that both *NAC27* and *NAC72* up-regulated (Li et al., 2019b). *NAC27* and *NAC72* were also involved in ABA signal transduction under heat stress of *Rhododendron hainanensis* (Zhao et al., 2018). Overexpression of *NAC72* in Arabidopsis thaliana can increase the drought resistance of plants, and also proved that it mediates the regulation of ABA-responsive genes (Li et al., 2016). Gene encoding *NAC072* with high connectivity in the navajowhite1 module was selected, indicating that it plays an important role in the regulation of rose drought stress. In agreement with previous studies, it can be concluded that *NAC072* may be involved in ABA signal transduction under drought stress in *R. chinensis*. Furthermore, the role of the core gene in the navajowhite1 module *NAC087* and *NAC100* in drought also requires further consideration.

The *WRKY* genes play pivotal roles in stress responses. Many researchers have demonstrated that *WRKYs* have crucial

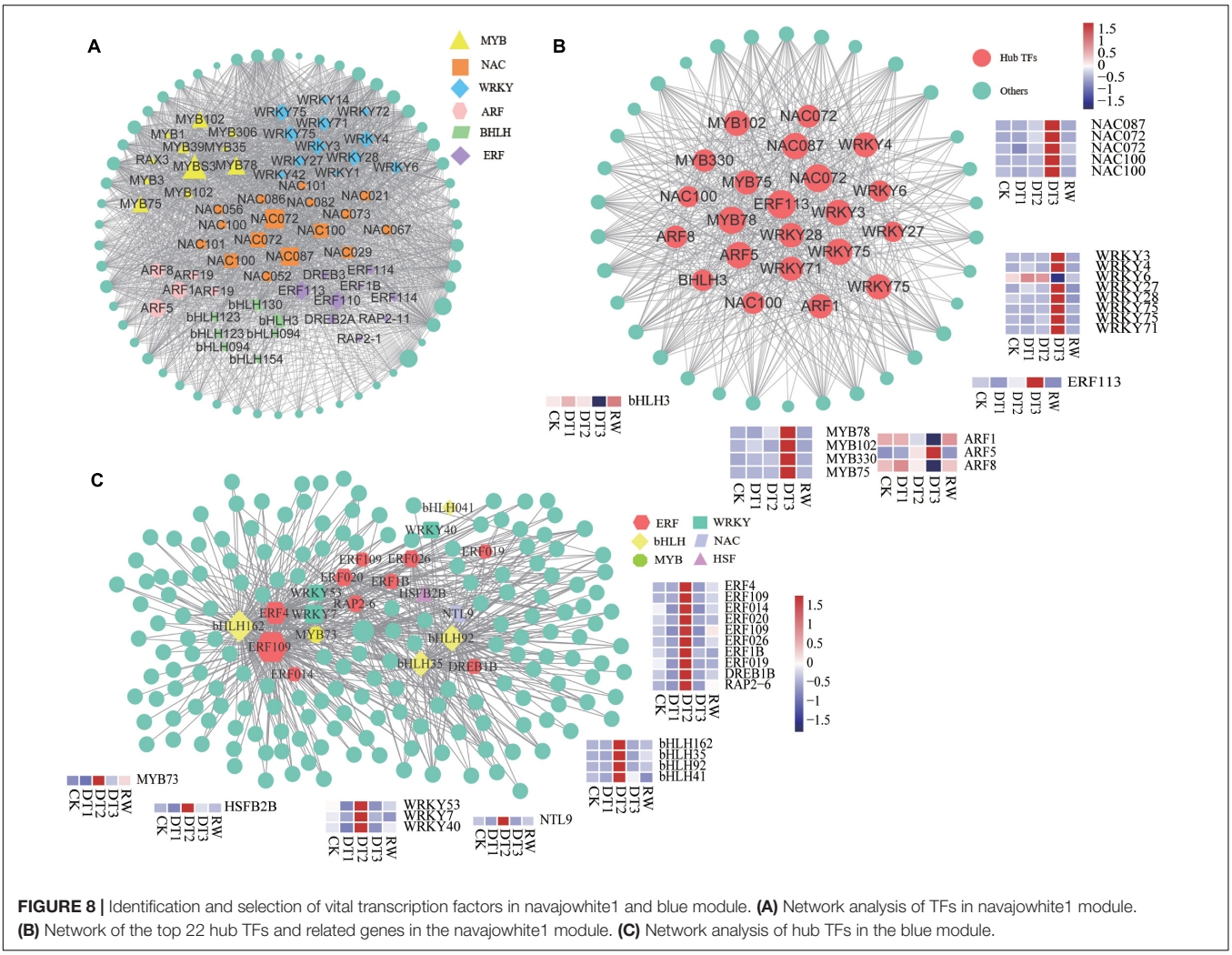


FIGURE 8 | Identification and selection of vital transcription factors in navajowhite1 and blue module. **(A)** Network analysis of TFs in navajowhite1 module. **(B)** Network of the top 22 hub TFs and related genes in the navajowhite1 module. **(C)** Network analysis of hub TFs in the blue module.

biological functions in the response of plants to different kinds of biotic and abiotic stresses (Jiang et al., 2017). Studies on Arabidopsis, rice, and soybean have shown the importance of the *WRKYs* in response to drought stress. For example, *AtWRKY57* (Jiang et al., 2012), *AtABO3* (Ren et al., 2010), *OsWRKY47* (Raineri et al., 2015), and *GmWRKY54* (Wei et al., 2019) can improve the drought tolerance of plants. Likewise, *VaWRKY14* responds to drought and cold stress, and hence the overexpression of *VaWRKY14* can enhance the drought tolerance of transgenic Arabidopsis (Zhang et al., 2018). However, some *WRKYs* can act as a negative regulator of abiotic stress in plants. For instance, *GhWRKY33* (Wang et al., 2019b) and *SlWRKY81* (Ahammed et al., 2020; Li et al., 2020) reduce the drought tolerance of transgenic plants. In the analysis, eight *WRKY* genes including *WRKY75*, *WRKY28*, and *WRKY27* were found to have high connectivity in the navajowhite1 module, indicating that they may be involved in drought response. Other research has found that *PagWRKY75* was down-regulated in the early stages of salt and osmotic stress, transgenic poplar lines overexpressing *PagWRKY75* were more sensitive to salt and osmotic (Zhao et al., 2019b). It can be assumed that *WRKY75* plays an essential regulatory role in abiotic stress, but the function of *WRKY75* in drought stress screened in this study requires further verification.

APETALA2/ethylene-responsive factor (*AP2/ERF*) TFs have been found to regulate plants and to enable them to be resistant to abiotic stress (Li et al., 2021b). In this study, AP2/ERF TFs were observed that may be involved in the regulation of rose drought stress. One gene encoding *ERF109* identified in a blue module appeared to play an important role in roses in resisting drought stress, the function of which requires further study. The study showed that *PtrERF109* is positively regulated with *Poncirus trifoliata* to resist the cold (Wang et al., 2019a). Moreover, *DREB2A*, *DREB3,* and *DREB1B* were found in this study. More and more studies have shown the function of *DREBs* in resisting abiotic stress in plants. Overexpression of *StDREB2* improved the drought resistance of cotton (El-Esawi and Alayafi, 2019). However, *RhDREB2B* is a negative regulator to resist abiotic stress (Li et al., 2021b).

The function of *bHLHs* in resisting abiotic stress in *Arabidopsis thaliana* has also been studied (Liu et al., 2014; Qiu et al., 2020). *bHLH162* and *bHLH35* have been identified in this

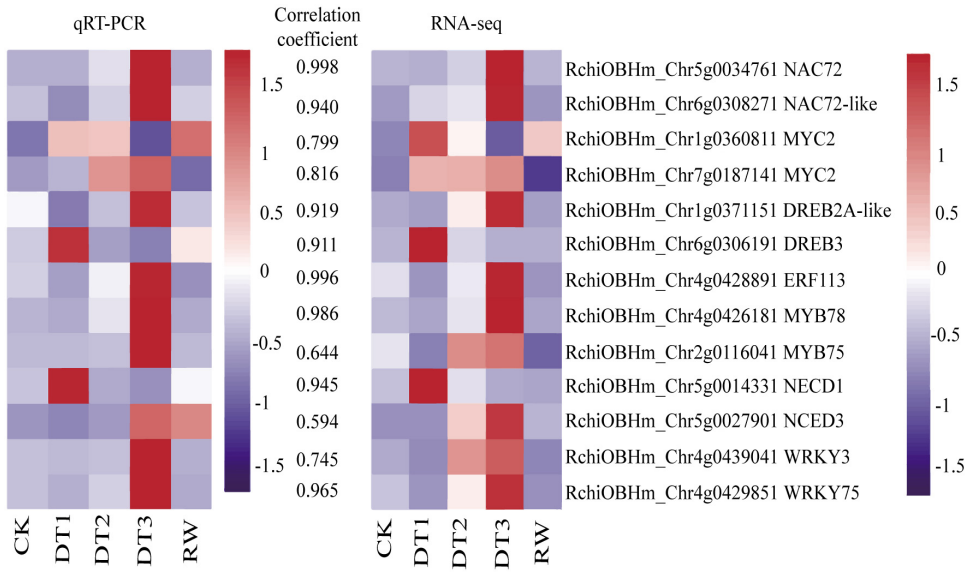


FIGURE 9 | Heatmaps of the validation and correlation analysis of 13 selected DEGs. The values between the two heatmaps represent the correlation coefficients of qRT-PCR and RNA-seq values from each gene.

study, providing new candidate genes for roses to resist drought stress. Meanwhile, *MYBs* including *MYBS3*, *MYB75,* and *MYB78* require further investigation. Particularly for rice, *OcMYBS3* is essential for resisting cold stress (Su et al., 2010).

Phytohormone Signals Under Drought Stress

Phytohormone plays a key role in abiotic stress responses and coordinates various signal transduction pathways. In this study, the gene expression involved in several plant hormone-related signal transduction pathways changed significantly after drought treatment, indicating that plant hormones may play a crucial role in the response of the rose to drought stress. ABA acts as an endogenous messenger in a plant's abiotic stress responses (Teng et al., 2014). Drought results in a substantial increase in plant ABA levels, accompanied by a major changes in gene expression and adaptive physiological responses. NCED is a key rate-limiting enzyme in the ABA synthesis pathway (Schwartz et al., 1997; Iuchi, 2002). Currently, some *NCED3* genes have been isolated and identified, corresponding gene function studies have also shown that *NCED3* genes play an important role in enhancing plant drought resistance (Li et al., 2019a). The study found that an *NCED3* gene was up-regulated by 2.70-fold at DT1. Interestingly, *NCED6* down-regulated by 9.40-fold in DT3. Lefebvre et al. (2006) state that *AtNCED6* is involved in ABA biosynthesis during seed development. The role of *NCED6* in drought stress still requires further study. Another ABA synthesis related to gene aldehyde oxidase (*AAO3*) and zeaxanthin cyclooxygenase (*ZEP*) were identified, both of which were significantly down-regulated in DT3. These results suggested that the synthesis of ABA may increase during the early drought stage and decrease during the later drought stage. Therefore, ABA-mediated signal transduction may be involved in the drought response of *R. chinensis*.

There is more and more evidence suggesting that the phytohormone ethylene is essential for regulating various developmental processes and stress responses of plants (Pei et al., 2017). Mcmichael et al. (1972) reported that the release of ethylene may increase under drought conditions. Studies on pineapples have shown that drought-stressed plants produce significantly less ethylene in leaf and stem tissues compared to control plants (Min and Bartholomew, 2005). These conflicting results may be due to the differences in distinct species. In general, the production of ethylene can be affected by changing *ACC* biosynthesis under drought stress (Larrainzar et al., 2014). In plants, ACC synthase (*ACS*) and ACC oxidase (*ACO*) are the two most essential enzymes in the ethylene biosynthetic pathway (Kende, 1989). In the study, the expression of three *ACS* genes was suppressed under drought stress, the expression of one *ACS* gene up-regulated, while the expression of two *ACO* genes was also up-regulated. It is not yet possible to speculate whether the ethylene content has increased or decreased under drought conditions. Ethylene response factors (*ERFs*) are downstream components of the ethylene signaling pathway (Guo and Ecker, 2004). Experiments showed that *ERFs* play important roles in plant abiotic stress response (Pan et al., 2012; Yu et al., 2017). A large number of *ERFs* exhibit different expressions under different drought levels. Their role under drought stress needs further study.

Jasmonates (JAs), a class of oxygenated lipid derivatives, are phytohormones necessary for plant growth and environmental adaptation (Wasternack and Hause, 2013). *MYCs* proteins are one of the JASMONATE-ZIM (JAZ) target proteins and a key transcription factor regulating the corresponding genes downstream of JA. According to the reports, *MYC2*, *MYC3,* and

MYC4 are the central nodes of the JA signal regulation (Chico et al., 2020). Lipoxygenase (*LOX*), allene oxide synthase (*AOS*), allene oxide cyclase (*AOC*), and oxophytodienoate reductase (*OPR*) are important enzyme genes in the JA synthesis pathway (Devoto and Turner, 2005). The study has observed that *RhHB1* binds to the *RhLOX4* promoter to suppress its expression in the cut rose (*Rosa hybrida*). As a result, dehydration tolerance decreases (Fan et al., 2020). These genes were identified, while most genes were up-regulated under drought stress. Other studies have shown that JA was observed to increase in *Pinus pinaster* (Pedranzani et al., 2007) plants and *Oryza sativa* (rice) (Du et al., 2013) leaves exposed to drought conditions. Based on these results, it can be conjectured that the JA pathway plays a role in the response to drought stress in *R. chinensis*.

Moreover, recent studies have indicated that the role of brassinosteroid (BR) in plant response to drought stress (Nolan et al., 2017a,b, 2020). In Fabregas et al.'s (2018) study, the overexpression of *BRL3*, a vascular-rich member of the BR family, can confer Arabidopsis drought resistance without compromising normal plant growth. In this study, the expression of two *BRL* genes increased under drought stress, and hence BRLs may play positive parts in the response to drought. In addition, BR and ABA regulated the drought resistance of plants. Indeed, the molecular basis of the antagonistic effect of the BR-ABA pathway has been defined (Wang et al., 2018; Nolan et al., 2020). The role of ABA in promoting the response to drought stress suggests that BRs will inhibit drought stress responses. Yet, only the role of hormone series resistance has been explored in model plants. This study also laid the foundation for the interaction of hormones in the drought resistance of rose.

Signaling Pathway Mediates Drought-Stress Responses

When plants are subjected to adversity stress due to changes in the external environment, they sense adversity stimuli from outside, recognize and transduce between cells and cells, and then transmit them to downstream genes, thereby causing downstream affect gene expression to resist adversity. Calcium ions are the second messengers of the cell, which mediates the calcium signal plays an essential role in the plant response to external stimuli (Kudla et al., 2010; Reddy et al., 2011). As a second messenger, calcium ion can not only maintain the stability of cell structure, but also play a vital role in signal transmission and participate in plant stress response (Xiong et al., 2002; Shinozaki et al., 2003). The three main families of calcium sensors in Ca^{2+} signaling in plants are calmodulin (*CaM*), calcineurin B-like (*CBL*), and calcium-dependent protein kinases (*CDPKs*) (Ranty et al., 2006; Boudsocq and Sheen, 2013). Analysis of the *CaM/CMLs* identified in this study revealed that five genes up-regulated, two up-regulated and then down-regulated, and fifteen down-regulated under drought stress. *AtCML8*, *AtCML13*, *AtCML18* and *AtCML25* down-regulated by salt damage and drought (McCormack et al., 2005). *ShCML44* up-regulated under cold, drought, and salt damage (Munir et al., 2016). In contrast, the gene encoding *CML44* down-regulated and the gene encoding *CML8* up-regulated under drought induction in this study. *CDPKs/CPKs* are involved in multiple stress signaling pathways. In Arabidopsis, Overexpressed *AtCPK6* confers drought tolerance (Xu et al., 2010), while *AtCPK21* is a negative regulator of osmotic response (Franz et al., 2011). In this study, six genes encoding putative Ca^{2+}-dependent protein kinases up-regulated during drought stress, and four of them showed high expression levels during the DT2 period. These results indicate that the CDPK pathway was significantly activated in DT2. CBLs communicate with downstream CIPKs for signal transmission. The CBL-CIPK pathways also participate in plant abiotic stress response (Xiang et al., 2007). In the current research, three genes encoding *CBLs-CIPKs* significantly down-regulated in DT3, which indicates that the CBL-CIPK pathway may contribute to regulating a later drought response in *R. chinensis*. In summary, the above results suggest that Ca^{2+} mediated signaling pathway has key functions in the response of *R. chinensis* to drought stress.

Plant's sense and mediate various intracellular and extracellular signals through cell surface receptor kinases after stimulating by the external environment. Studies have shown that receptor-like protein kinases (RLKs) play a vital role in plant growth and stress adaptation (Wang et al., 2007). Leucine-rich repeat receptor protein kinases (LRR-RLKs), a class of single transmembrane proteins, are the largest family of plant receptor protein kinases (Gou et al., 2010). The *LRR-RLKs* genes identified by the study were 1 up-regulated and 10 down-regulated under drought stress, which was similar to the expression of *LRR-RLKs* genes (3 up-regulated, 38 down-regulated) in chrysanthemum under dehydration stress (Xu et al., 2013). Additionally, multiple *MAPKs* were observed during the plant response to abiotic stresses such as salt, drought, cold, and heat, mainly rapid activation of *MPK3*, *MPK4,* and *MPK6* (Zhu, 2016). The results in this study are similar, as gene encoding *MPK3* significantly down-regulated on DT1, indicating that *MPK3* responded quickly when drought stress occurred.

MAPK cascade signals can also transmit reactive oxygen species (ROS) signals to downstream targets. When plants were subjected to environmental stress, ROS generated signals in the plant, triggering a series of changes. Plants also have an antioxidant system protecting against poisonous oxygen (Jaspers and KangasjcTrvi, 2010). Abiotic stress can cause plants to produce excessive ROS which can cause damage to plants. Plants have efficient enzymes including SOD, POD, APX, GPX, and so on. The antioxidant defense mechanisms can protect plants from oxidative stress (Gill and Tuteja, 2010). In this study, two genes encoding *APX* were up-regulated and four were down-regulated. Similarly, in the study of *Citrullus lanatus*, the changes of *APX* genes are different under drought stress (Goitseone et al., 2018). The post-harvest petals of cut rose (Rosa *hybrida* cv. Samantha) under water deficit stress reported that the Transcription level of *RhAPX* may be involved in the response of water deficit stress (Jin et al., 2006). Interestingly, three genes encoding *SOD* were identified, all of which were reduced under drought stress. When SOD activity was measured, the level of SOD activity increased as the degree of drought stress increased. The expression pattern of SOD genes in this study is similar to the expression pattern of the *SmFSD2* in drought. The *SmFSD2* did not change significantly

over 3 days, but decreased to the lowest value at 9 days (Shafi et al., 2019). These results indicate that the ROS signaling pathway may be involved in the regulation of *R. chinensis* response to drought stress.

Metabolism in Response to Drought Stress in *R. chinensis*

Primary and secondary metabolites act as signaling molecules or protective agents when plants respond to adversity stress. The primary metabolic processes were involved in the response to the drought stress of *R. chinensis*, producing a series of osmotic protective agents. Such as sugars, starches, amino acids, and lipids. Accumulation of sugars in various plants is related to high tolerance to drought stress (Ali et al., 2019), which is consistent with the observation of the experiment. Most genes related to carbohydrate synthesis including fructose, glucose, mannose, sucrose, and trehalose were induced and significantly up-regulated under drought stress. The gene encoding trehalose phosphate synthase (*TPP1*) down-regulated under drought stress. Among the 11 *OsTPS* genes in the rice (*Oryza sativa*) genome, only *OsTPS1* has TPS activity (Zang et al., 2011). Overexpression of *OsTPS1* can improve the tolerance of rice seedlings to drought stress. At the same time, it also causes the expression of some stress-related genes to be up-regulated and the phenotype has no obvious change (Li et al., 2011). Trehalose may involve in the process of rose's resistance to drought. The expression levels of the four genes encoding starch synthase significantly reduced at DT3, with no significant change in expression levels in the remaining periods compared with the control. It is speculated that the starch content may decrease in a later drought.

Lipids are essential components of cells and organelles. A series of genes that synthesize lipid biosynthesis have been identified in drought treated *R. chinensis* plants. Long-chain acyl-CoA synthetases (LACS) are indispensable in the pathway of lipids synthesis and degradation of higher plants. There are nine genes in the *LACSs* family in Arabidopsis, among which *LACS1* and *LACS2* are involved in the synthesis of wax and cutin together, with an overlap in function (Lü et al., 2009; Weng et al., 2010). Drought stress can cause rapid accumulation of wax on plant surface (Kosma et al., 2009). In this study, six genes encoding *LACS2* were induced under drought, while five of them significantly down-regulated at the DT3. Studies on bananas, *Agave sisalana* (Sarwar et al., 2019), and *Poa pratensis* (Ni et al., 2016) have reported that waxy synthetic transcripts significantly down-regulated under drought stress. These results confirm that wax may play key roles in regulating the drought resistance of *R. chinensis*. In brief, it can be speculated that the wax content may be reduced in the later period of drought stress. In addition, the

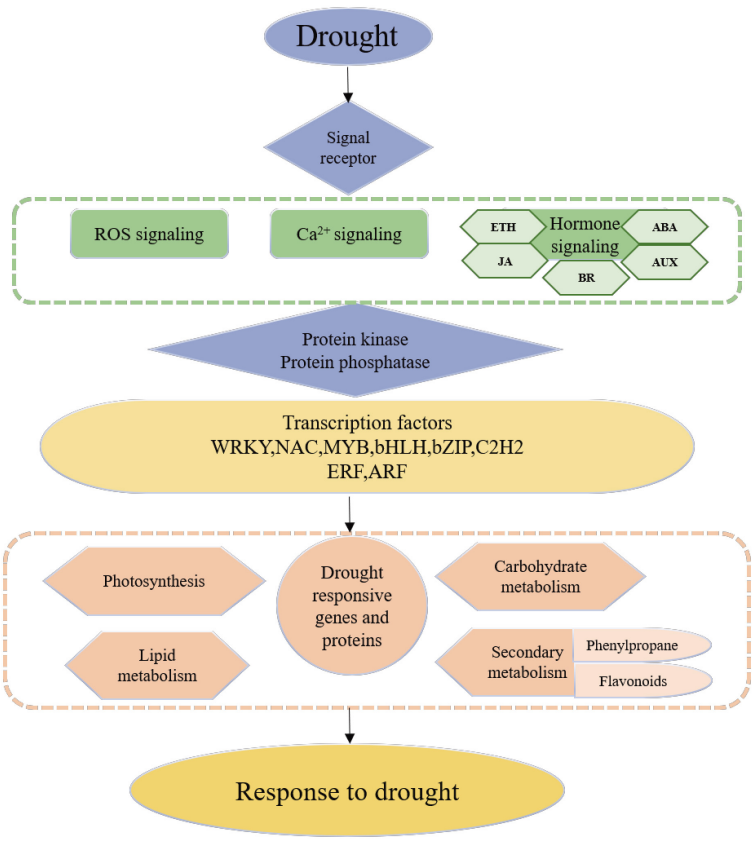


FIGURE 10 | The hypothetical model of *Rosa chinensis* 'Old Blush' in response to drought stress.

gene encoding GDSL esterase/lipase was significantly induced at the beginning of drought stress. Studies on the GDSL esterase/lipase gene of *Arabidopsis thaliana* revealed that this gene helps plants resist abiotic stress (Lai et al., 2017), and is induced in *M. wufengensis* by cold stress (Deng et al., 2019). In summary, lipid metabolism is crucial for *R. chinensis* adaptation to drought stress.

Furthermore, as part of the adaptation mechanism to the environment of plants, secondary metabolism is sensitive to both biological and abiotic stresses (Hartmann, 2007). It is known that the phenylpropane pathway has high responsiveness to different abiotic stresses such as injury, drought, and low or high temperature. White grapes (*Vitis vinifera* L.) cope with drought by stimulating the phenylpropane pathway, which reflects the result of this study. The experiment showed that most of the secondary metabolites are derived from phenylpropane biosynthesis and flavonoid biosynthesis pathways. The genes that encode phenylalanine ammonia lyase (*PAL*) and coumaric acid-CoA ligase (*4CL*) were identified. They are all related to enzyme genes in the phenylpropane metabolism pathway. They significantly up-regulated at DT1, indicating that the phenylpropane metabolism pathway was induced under drought stress. The transcription of *PAL* of *Caragana korshinskii* increased under field and laboratory drought conditions (Liu et al., 2019). These results also confirm the results of the present report. Another major derivative of the phenylpropane pathway is flavonoids, which play a key role in adversity stress. Flavonoids with free radical scavenging activity alleviate oxidative and drought stress in *Arabidopsis thaliana* (Nakabayashi et al., 2014). This study reported that genes encoding flavonoid synthase (*FLS*), chalcone isomerase (*CHS*), and anthocyanin synthase (*ANS*) were induced under drought. Flavonoid products may affect the regulation of *R. chinensis* by drought stress.

Model Construction of Rose Drought Response

This study aimed to enhance understanding of the mechanisms through which *R. chinensis* responds to drought and identify the TFs related to drought tolerance. **Figure 10** summarizes the model of *R. chinensis* drought response. According to the transcriptome data, signal receptors, such as ion channel proteins sense external drought stimuli and transmit through Ca^{2+}, ROS, and phytohormone signaling transduction. The signal transmits protein kinases and protein phosphorylases, which can activate the transcription factors. The activation of TFs triggers downstream drought-responsive gene transcription such as lipid metabolism, carbohydrate metabolism, and secondary metabolism to regulate cell homeostasis. The hub TFs involved in roses' drought response and rewatering were identified through WGCNA, providing resources for further research on the drought control network of roses. The new findings of this study, on the relationship between unreported TFs and drought resistance regulation, still require further experimental explore.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the NCBI repositories (BioProject ID: PRJNA722055).

AUTHOR CONTRIBUTIONS

YL, SZ, XJ, and HF conceived and designed the research. XJ performed the experiments, analyzed the data, and wrote the manuscript. YB and NJ cultivated and provided the rose plants. All authors have read and agreed to the published version of the manuscript.

FUNDING

Research was funded by Beijing Municipal Science and Technology Project (z191100008519005) and Chongming District, Shanghai Agricultural Science and Technology Innovation Project (2019CNKC-03).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.690264/full#supplementary-material

Supplementary Figure 1 | The distribution of cleaned RNA-Seq reads mapped to the rose reference genome. **(A)** Percent of mapped to genome regions of CK groups. **(B)** Percent of mapped to genome regions of DT1 groups. **(C)** Percent of mapped to genome regions of DT2 groups. **(D)** Percent of mapped to genome regions of DT3 groups. **(E)** Percent of mapped to genome regions of RW groups.

Supplementary Table 1 | Top 20 enriched GO terms of up-regulated DEGs.

Supplementary Table 2 | Top 20 enriched GO terms of down-regulated DEGs.

Supplementary Table 3 | Significantly enriched gene pathways involving DEGs under the drought stress treatment.

Supplementary Table 4 | Protein kinases-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 5 | Calcium-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 6 | ROS producing and scavenging system-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 7 | Differently expressed related to hormone metabolism and signaling in response to drought treatment.

Supplementary Table 8 | Starch and sucrose metabolism-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 9 | Lipid metabolism process-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 10 | Photosynthesis-related genes that were differentially expressed in response to drought treatment.

Supplementary Table 11 | Transcription factor families.

Supplementary Table 12 | The primers sequences of selected genes used in qRT-PCR.

Supplementary Table 13 | Top 20 enriched GO terms of modules.

Supplementary Table 14 | Top 20 enriched KEGG pathway of modules.

REFERENCES

Adamipour, N., Khosh-Khui, M., Salehi, H., Razi, H., Karami, A., and Moghadam, A. (2020). Metabolic and genes expression analyses involved in proline metabolism of two rose species under drought stress. *Plant Physiol. Biochem.* 155, 105–113. doi: 10.1016/j.plaphy.2020.07.028

Ahammed, G. J., Li, X., Yang, Y., Liu, C., Zhou, G., Wan, H., et al. (2020). Tomato WRKY81 acts as a negative regulator for drought tolerance by modulating guard cell H2O2-mediated stomatal closure. *Environ. Exp. Bot.* 171:103960. doi: 10.1016/j.envexpbot.2019.103960

Ali, Q., Shahid, S., Ali, S., Javed, M. T., and Haider, M. Z. (2019). *Trehalose Metabolism in Plants under Abiotic Stresses: Profiling and CounterAction*. Boca Raton, FL: CRC Press.

Al-Yasi, H., Attia, H., Alamer, K., Hassan, F., Ali, E., Elshazly, S., et al. (2020). Impact of drought on growth, photosynthesis, osmotic adjustment, and cell wall elasticity in Damask rose. *Plant Physiol. Biochem.* 150, 133–139. doi: 10.1016/j.plaphy.2020.02.038

Ashburner, A., Ball, C. A., Blake, J. A., and Botstein, D. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Boudsocq, M., and Sheen, J. (2013). CDPKs in immune and stress signaling. *Trends Plant Sci.* 18, 30–40. doi: 10.1016/j.tplants.2012.08.008

Chaves, M. M., Maroco, J. P., and Pereira, J. S. (2003). Understanding plant responses to drought - From genes to the whole plant. *Funct. Plant Biol.* 30, 239–264. doi: 10.1071/fp02076

Chico, J. M., Lechner, E., Fernandez-Barbero, G., Canibano, E., Garcia-Casado, G., Franco-Zorrilla, J. M., et al. (2020). CUL3(BPM) E3 ubiquitin ligases regulate MYC2, MYC3, and MYC4 stability and JA responses. *Proc. Natl. Acad. Sci. U S A.* 117, 6205–6215. doi: 10.1073/pnas.1912199117

Dai, F., Zhang, C., Jiang, X., Kang, M., Yin, X., Lu, P., et al. (2012). RhNAC2 and RhEXPA4 are involved in the regulation of dehydration tolerance during the expansion of rose petals. *Plant Physiol.* 160, 2064–2082. doi: 10.1104/pp.112.207720

David Gamboa-Tuz, S., Pereira-Santana, A., Alejandro Zamora-Briseno, J., Castano, E., Espadas-Gil, F., Tonatiuh Ayala-Sumuano, J., et al. (2018). Transcriptomics and co-expression networks reveal tissue-specific responses and regulatory hubs under mild and severe drought in papaya (*Carica papaya* L.). *Sci. Rep.* 8:14539. doi: 10.1038/s41598-018-32904-2

Deng, S., Ma, J., Zhang, L., Chen, F., Sang, Z., Jia, Z., et al. (2019). *De novo* transcriptome sequencing and gene expression profiling of *Magnolia wufengensis* in response to cold stress. *BMC Plant Biol.* 19:321. doi: 10.1186/s12870-019-1933-5

Devoto, A., and Turner, J. G. (2005). Jasmonate-regulated *Arabidopsis* stress signalling network. *Physiol. Plant.* 123, 161–172. doi: 10.1111/j.1399-3054.2004.00418.x

Dhriti, S., and Ashverya, L. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front. Plant Sci.* 6:895. doi: 10.3389/fpls.2015.00895

Ding, A., Li, S., Li, W., Hao, Q., Wan, X., Wang, K., et al. (2019). RhNAC31, a novel rose NAC transcription factor, enhances tolerance to multiple abiotic stresses in *Arabidopsis*. *Acta Physiol. Plantarum* 41:75. doi: 10.1007/s11738-019-2866-1

Du, H., Liu, H., and Xiong, L. (2013). Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. *Front. Plant Sci.* 4:397. doi: 10.3389/fpls.2013.00397

El-Esawi, M. A., and Alayafi, A. A. (2019). Overexpression of StDREB2 transcription factor enhances drought stress tolerance in cotton (*Gossypium barbadense* L.). *Genes* 10:142. doi: 10.3390/genes10020142

Erpen, L., Devi, H. S., Grosser, J. W., and Dutt, D. (2018). Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. *Plant Cell Tiss Org. Cult.* 132, 1–25. doi: 10.1007/s11240-017-1320-6

Fabregas, N., Lozano-Elena, F., Blasco-Escamez, D., Tohge, T., Martinez-Andujar, C., Albacete, A., et al. (2018). Overexpression of the vascular brassinosteroid receptor BRL3 confers drought resistance without penalizing plant growth. *Nat. Commun.* 9:4680. doi: 10.1038/s41467-018-06861-3

Fan, Y., Liu, J., Zou, J., Zhang, X., Jiang, L., Liu, K., et al. (2020). The RhHB1/RhLOX4 module affects the dehydration tolerance of rose flowers (Rosa hybrida) by fine-tuning jasmonic acid levels. *Hortic. Res.* 7:74. doi: 10.1038/s41438-020-0299-z

Fang, H., Wang, H. L., Li, H. G., Su, Y., and Shuang, L. (2018). PeCHYR1, a ubiquitin E3 ligase from *Populus euphratica*, enhances drought tolerance via ABA-induced stomatal closure by ROS production in Populus. *Plant Biotechnol. J.* 16, 1514–1528. doi: 10.1111/pbi.12893

Franco-Zorrilla, J. M., Lopez-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U S A.* 111, 2367–2372. doi: 10.1073/pnas.1316278111

Franz, S., Ehlert, B., Liese, A., Kurth, J., Cazale, A. C., and Romeis, T. (2011). Calcium-dependent protein kinase CPK21 functions in abiotic stress response in *Arabidopsis thaliana*. *Mol. Plant* 4, 83–96. doi: 10.1093/mp/ssq064

Gadzinowska, J., Ostrowska, A., Hura, K., Dziurka, M., Pawlowska, B., and Hura, T. (2019). Physiological traits determining high adaptation potential of sweet briar (*Rosa rubiginosa* L.) at early stage of growth to dry lands. *Sci. Rep.* 9:19390. doi: 10.1038/s41598-019-56060-3

Gill, S. S., and Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* 48, 909–930. doi: 10.1016/j.plaphy.2010.08.016

Goitseone, M., Hisashi, T., and Kinya, A. (2018). The cDNA structures and expression profile of the ascorbate peroxidase gene family during drought stress in wild watermelon. *J. Agricult. Sci.* 10:56. doi: 10.5539/jas.v10n8p56

Gou, X., He, K., Yang, H., Yuan, T., Lin, H., Clouse, S. D., et al. (2010). Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in *Arabidopsis thaliana*. *BMC Genomics* 11:19. doi: 10.1186/1471-2164-11-19

Guo, H., and Ecker, J. (2004). The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* 7, 40–49. doi: 10.1016/j.pbi.2003.11.011

Hartmann, T. (2007). From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* 68, 2831–2846. doi: 10.1016/j.phytochem.2007.09.017

Heath, R. L., and Packer, K. (1968). Leaf senescense: correlated with increased levels of membrane permeability and lipid peroxidation, and decreased levels of superoxide dismutase and catalase. *J. Exp. Bot.* 32, 93–101. doi: 10.1093/jxb/32.1.93

Hong, Y., Zhang, H., Lei, H., Li, D., and Song, F. (2016). Overexpression of a stress-responsive NAC transcription factor gene ONAC022 improves drought and salt tolerance in rice. *Front. Plant Sci.* 7:4. doi: 10.3389/fpls.2016.00004

Huang, X., Hou, L., Meng, J., You, H., Li, Z., Gong, Z., et al. (2018). The antagonistic action of abscisic acid and cytokinin signaling mediates drought stress response in *Arabidopsis*. *Mol. Plant* 11, 970–982. doi: 10.1016/j.molp.2018.05.001

Huang, Z. H., Ma, A. J., Yang, S. S., Liu, X. F., Zhao, T. T., Zhang, J. S., et al. (2020). Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot *Scophthalmus maximus*. *Comp. Biochem. Physiol. D-Genom. Proteom.* 33:100632. doi: 10.1016/j.cbd.2019.100632

Iuchi, S. (2002). Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis* (vol 27, pg 325, 2001). *Plant J.* 30, 611–611.

Jaspers, P., and KangasjcTrvi, J. (2010). Reactive oxygen species in abiotic stress signaling. *Physiol. Plant* 138, 405–413. doi: 10.1111/j.1399-3054.2009.01321.x

Jiang, J. J., Ma, S. H., Ye, N. H., Cao, J. S., and Zhang, J. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59, 86–101. doi: 10.1111/jipb.12513

Jiang, X., Li, S., Ding, A., Zhang, Z., Hao, Q., Wang, K., et al. (2018). The novel rose MYB transcription factor RhMYB96 enhances salt tolerance in transgenic *Arabidopsis*. *Plant Mol. Biol. Rep.* 36, 406–417. doi: 10.1007/s11105-018-1094-y

Jiang, X. Q., Zhang, C. Q., Lu, P. T., Jiang, G. M., Liu, X. W., Dai, F. W., et al. (2014). RhNAC3, a stress- associated NAC transcription factor, has a role in dehydration tolerance through regulating osmotic stress- related genes in rose petals. *Plant Biotechnol. J.* 12, 38–48. doi: 10.1111/pbi.12114

Jiang, Y., Liang, G., and Yu, D. (2012). Activated expression of WRKY57 confers drought tolerance in *Arabidopsis*. *Mol. Plant* 5, 1375–1388. doi: 10.1093/mp/sss080

Jin, J., Shan, N., Ma, N., Bai, J., and Gao, J. (2006). Regulation of ascorbate peroxidase at the transcript level is involved in tolerance to postharvest water deficit stress in the cut rose (*Rosa hybrida* L.) cv. Samantha. *Postharvest Biol. Technol.* 40, 236–243. doi: 10.1016/j.postharvbio.2006.01.014

Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., et al. (2016). Transcription factors and plants response to drought stress: current understanding and future directions. *Front. Plant Sci.* 7:1029. doi: 10.3389/fpls.2016.01029

Kende, H. (1989). Enzymes of ethylene biosynthesis. *Plant Physiol.* 91, 1–4. doi: 10.1104/pp.91.1.1

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

Klie, M., and Debener, T. (2011). Identification of superior reference genes for data normalisation of expression studies via quantitative PCR in hybrid roses (*Rosa hybrida*). *BMC Res. Notes* 4:518. doi: 10.1186/1756-0500-4-518

Kosma, D. K., Bourdenx, B., Bernard, A., Parsons, E. P., Lu, S., Joubes, J., et al. (2009). The impact of water deficiency on leaf cuticle lipids of *Arabidopsis*. *Plant Physiol.* 151, 1918–1929. doi: 10.1104/pp.109.141911

Kudla, J., Batistic, O., and Hashimoto, K. (2010). Calcium signals: the lead currency of plant information processing. *Plant Cell* 22, 541–563. doi: 10.1105/tpc.109.072686

Lai, C. P., Huang, L. M., Chen, L. F. O., Chan, M. T., and Shaw, J. F. (2017). Genome-wide analysis of GDSL-type esterases/lipases in *Arabidopsis*. *Plant Mol. Biol.* 95, 181–197. doi: 10.1007/s11103-017-0648-y

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

Larrainzar, E., Molenaar, J. A., Wienkoop, S., Gil-Quintana, E., Alibert, B., Limami, A. M., et al. (2014). Drought stress provokes the down-regulation of methionine and ethylene biosynthesis pathways in Medicago truncatula roots and nodules. *Plant Cell Environ.* 37, 2051–2063. doi: 10.1111/pce.12285

Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., et al. (2006). Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J.* 45, 309–319. doi: 10.1111/j.1365-313X.2005.02622.x

Li, H. W., Zang, B. S., Deng, X. W., and Wang, X. P. (2011). Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice. *Planta* 234, 1007–1018. doi: 10.1007/s00425-011-1458-0

Li, T., Sun, J., Li, C., Lu, Z., and Xia, J. (2019a). Cloning and expression analysis of the FvNCED3 gene and its promoter from ash (*Fraxinus velutina*). *J. Forestry Res.* 30, 471–482. doi: 10.1007/s11676-018-0632-7

Li, Z. Q., Xing, W., Luo, P., Zhang, F. J., Jin, X. L., and Zhang, M. H. (2019b). Comparative transcriptome analysis of *Rosa chinensis* 'Slaters' crimson China' provides insights into the crucial factors and signaling pathways in heat stress response. *Plant Physiol. Biochem.* 142, 312–331. doi: 10.1016/j.plaphy.2019.07.002

Li, Z. Y., Chen, W., Zhang, C., Du, C. X., Shao, G. Y., Cui, Y. Y., et al. (2019c). RcMYBPA2 of Rosa chinensis functions in proanthocyanidin biosynthesis and enhances abiotic stress tolerance in transgenic tobacco. *Plant Cell Tissue Organ. Culture* 137, 441–454. doi: 10.1007/s11240-019-01580-z

Li, W., Fu, L., Geng, Z., Zhao, X., Liu, Q., and Jiang, X. (2021a). Physiological characteristic changes and full-length transcriptome of rose (*Rosa chinensis*) roots and leaves in response to drought stress. *Plant Cell Physiol.* 61, 2153–2166. doi: 10.1093/pcp/pcaa137

Li, W., Geng, Z., Zhang, C., Wang, K., and Jiang, X. (2021b). Whole-genome characterization of *Rosa chinensis* AP2/ERF transcription factors and analysis of negative regulator RcDREB2B in *Arabidopsis*. *BMC Genomics* 22:90. doi: 10.1186/s12864-021-07396-6

Li, X., Li, X., Li, M., Yan, Y., Xu, L., and Ling, L. (2016). Dual function of NAC072 in ABF3-mediated ABA-responsive gene regulation in *Arabidopsis*. *Front. Plant Sci.* 7:1075. doi: 10.3389/fpls.2016.01075

Li, X., Wan, H., Zhou, G., Cheng, Y., and Ahammed, G. J. (2020). SlWRKY81 reduces drought tolerance by attenuating proline biosynthesis in tomato. *Sci. Horticult.* 270:109444. doi: 10.1016/j.scienta.2020.109444

Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., et al. (2018). Correction to: OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. *Plant Mol. Biol.* 97, 467–468. doi: 10.1007/s11103-018-0745-6

Liu, F., Xie, L., Yao, Z., Zhou, Y., and Gong, C. (2019). Caragana korshinskii phenylalanine ammonialyase is up-regulated in the phenylpropanoid biosynthesis pathway in response to drought stress. *Biotechnol. Biotechnol. Equipment* 33, 842–854. doi: 10.1080/13102818.2019.1623718

Liu, S., Zenda, T., Dong, A., Yang, Y., Wang, N., and Duan, H. (2021). Global transcriptome and weighted gene co-expression network analyses of growth-stage-specific drought stress responses in maize. *Front. Genet.* 12:645443. doi: 10.3389/fgene.2021.645443

Liu, W., Tai, H., Li, S., Gao, W., Zhao, M., Xie, C., et al. (2014). bHLH122 is important for drought and osmotic stress resistance in *Arabidopsis* and in the repression of ABA catabolism. *New Phytol.* 201, 1192–1204. doi: 10.1111/nph.12607

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

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

Lü, S., Song, T., Kosma, D. K., Parsons, E. P., Rowland, O., and Jenks, M. A. (2009). Arabidopsis CER8 encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. *Plant J.* 59, 553–564. doi: 10.1111/j.1365-313x.2009.03892.x

Mao, X., Chen, S., Li, A., Zhai, C., Jing, R., and Turgay, U. (2014). Novel NAC transcription factor TaNAC67 confers enhanced multi-abiotic stress tolerances in *Arabidopsis*. *PLoS One* 9:e84359. doi: 10.1371/journal.pone.0084359

McCormack, E., Tsai, Y. C., and Braam, J. (2005). Handling calcium signaling: *Arabidopsis* CaMs and CMLs. *Trends Plant Sci.* 10, 383–389. doi: 10.1016/j.tplants.2005.07.001

Mcmichael, B. L., Jordan, W. R., and Powell, R. D. (1972). An effect of water stress on ethylene production by intact cotton petioles. *Plant Physiol.* 49, 658–660. doi: 10.1104/pp.49.4.658

Min, X. J., and Bartholomew, D. P. (2005). Effects of flooding and drought on ethylene metabolism, titratable acidity and fruiting of pineapple. *Acta Horticulturae* 666, 135–148. doi: 10.17660/actahortic.2005.666.13

Mordor Intelligence (2020). *Netherlands Floriculture Market-Growth, Trends, COVID-19 Impact, and Forecasts (2021-2026)*. Hyderabad: Mordor Intelligence.

Munir, S., Liu, H., Xing, Y., Hussain, S., Ouyang, B., Zhang, Y., et al. (2016). Overexpression of calmodulin-like (ShCML44) stress-responsive gene from *Solanum habrochaites* enhances tolerance to multiple abiotic stresses. *Sci. Rep.* 6:31772.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., et al. (2014). Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. *Plant J.* 77, 367–379. doi: 10.1111/tpj.12388

Ni, Y., Guo, N., Zhao, Q., and Guo, Y. (2016). Identification of candidate genes involved in wax deposition in Poa pratensis by RNA-seq. *BMC Genom.* 17:314. doi: 10.1186/s12864-016-2641-2

Niu, J., Zhang, S., Liu, S., Ma, H., Chen, J., Shen, Q., et al. (2018). The compensation effects of physiology and yield in cotton after drought stress. *J. Plant Physiol.* 22, 30–48. doi: 10.1016/j.jplph.2018.03.001

Niu, Y., Wang, Y., Li, P., Zhang, F., Liu, H., and Zheng, G. (2013). Drought stress induces oxidative stress and the antioxidant defense system in ascorbate-deficient vtc1 mutants of *Arabidopsis thaliana*. *Acta Physiol. Plantarum* 35, 1189–1200. doi: 10.1007/s11738-012-1158-9

Nolan, T., Chen, J., and Yin, Y. (2017a). Cross-talk of brassinosteroid signaling in controlling growth and stress responses. *Biochem. J.* 474, 2641–2661. doi: 10.1042/BCJ20160633

Nolan, T. M., Brennan, Yang, M. R., Chen, J. N., and Zhang, M. C. (2017b). Selective autophagy of BES1 Mediated by DSK2 balances plant growth and survival. *Dev. Cell* 41, 33–46.e7.

Nolan, T. M., Vukasinovic, N., Liu, D., Russinova, E., and Yin, Y. (2020). Brassinosteroids: multidimensional regulators of plant growth, development, and stress responses. *Plant Cell* 32, 295–318. doi: 10.1105/tpc.19.00335

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34.

Pan, Y., Seymour, G. B., Lu, C., Hu, Z., Chen, X., and Chen, G. (2012). An ethylene response factor (ERF5) promoting adaptation to drought and salt tolerance in tomato. *Plant Cell Rep.* 31, 349–360. doi: 10.1007/s00299-011-1170-3

Pedranzani, H., Sierra-De-Grado, R., Vigliocco, A., Miersch, O., and Abdala, G. (2007). Cold and water stresses produce changes in endogenous jasmonates in

two populations of Pinus pinaster Ait. *Plant Growth Regulation* 52, 111–116. doi: 10.1007/s10725-007-9166-2

Pei, H., Wang, H., Wang, L., Zheng, F., and Dong, C. H. (2017). *Regulatory Function of Ethylene in Plant Responses to Drought, Cold, and Salt Stresses*. Hoboken, NJ: John Wiley & Sons Inc.

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, 1650–1667. doi: 10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Qiu, J. R., Huang, Z., Xiang, X. Y., Xu, W. X., Wang, J. T., Chen, J., et al. (2020). MfbHLH38, a *Myrothamnus flabellifolia* bHLH transcription factor, confers tolerance to drought and salinity stresses in *Arabidopsis*. *BMC Plant Biol.* 20:542. doi: 10.1186/s12870-020-02732-6

Raineri, J., Wang, S., Peleg, Z., Blumwald, E., and Chan, R. L. (2015). The rice transcription factor OsWRKY47 is a positive regulator of the response to water deficit stress. *Plant Mol. Biol.* 88, 401–413. doi: 10.1007/s11103-015-0329-7

Ranty, B., Aldon, D., and Galaud, J. P. (2006). Plant calmodulins and calmodulin-related proteins: multifaceted relays to decode calcium signals. *Plant Signal. Behav.* 1, 96–104. doi: 10.4161/psb.1.3.2998

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., et al. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nat. Genet.* 50, 772–777. doi: 10.1038/s41588-018-0110-3

Reddy, A. S. N., Ali, G. S., Celesnik, H., and Day, I. S. (2011). Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. *Plant Cell* 23, 2010–2032. doi: 10.1105/tpc.111.084988

Ren, X., Chen, Z., Liu, Y., Zhang, H., Zhang, M., Liu, Q., et al. (2010). ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in *Arabidopsis*. *Plant J. Cell Mol. Biol.* 63, 417–429. doi: 10.1111/j.1365-313x.2010.04248.x

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

Sarwar, M. B., Ahmad, Z., Rashid, B., Hassan, S., Gregersen, P. L., Leyva, et al. (2019). *De novo* assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci. Rep.* 9:396.

Schwartz, S. H., Tan, B. C., Gage, D. A., Zeevaart, J. A., and McCarty, D. R. (1997). Specific oxidative cleavage of carotenoids by VP14 of maize. *Science* 276, 1872–1874. doi: 10.1126/science.276.5320.1872

Shafi, A., Gill, T., Zahoor, I., Ahuja, P. S., Sreenivasulu, Y., Kumar, S., et al. (2019). Ectopic expression of SOD and APX genes in *Arabidopsis* alters metabolic pools and genes related to secondary cell wall cellulose biosynthesis and improve salt tolerance. *Mol. Biol. Rep.* 46, 1985–2002. doi: 10.1007/s11033-019-04648-3

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

Shao, A., Wang, W., Fan, S., Xu, X., Yin, Y., Erick, A., et al. (2021). Comprehensive transcriptional analysis reveals salt stress-regulated key pathways, hub genes and time-specific responsive gene categories in common bermudagrass (*Cynodon dactylon* (L.) Pers.) roots. *BMC Plant Biol.* 21:175. doi: 10.1186/s12870-021-02939-1

Shinozaki, K., Yamaguchi-Shinozaki, K., and Seki, M. (2003). Regulatory network of gene expression in the drought and cold stress responses. *Curr. Opin. Plant Biol.* 6:410. doi: 10.1016/s1369-5266(03)00092-x

Silva, K., Singh, J., Bednarek, R., Fei, Z., and Khan, A. (2019). Differential gene regulatory pathways and co-expression networks associated with fire blight infection in apple (Malus × domestica). *Horticulture Res.* 6:35.

Su, C.-F., Wang, Y.-C., Hsieh, T.-H., Lu, C.-A., Tseng, T.-H., and Yu, S.-M. (2010). A Novel MYBS3-Dependent Pathway Confers Cold Tolerance in Rice. *Plant Physiol.* 153, 145–158. doi: 10.1104/pp.110.153015

Sung, S. J., Nuri, O., Joong, C. P., Shic, K. Y., Do, C. Y., and Ju-Kon, K. (2018). Overexpression of OsNAC14 improves drought tolerance in rice. *Front. Plant Sci.* 9:310. doi: 10.3389/fpls.2018.00310

Teng, K., Li, J., Liu, L., Han, Y., Du, Y., Zhang, J., et al. (2014). Exogenous ABA induces drought tolerance in upland rice: the role of chloroplast and ABA biosynthesis-related gene expression on photosystem II during PEG stress. *Acta Physiologiae Plantarum* 36, 2219–2227. doi: 10.1007/s11738-014-1599-4

Tran, L. S., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16, 2481–2498. doi: 10.1105/tpc.104.022699

Valliyodan, B., and Nguyen, H. T. (2006). Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr. Opin. Plant Biol.* 9, 189–195. doi: 10.1016/j.pbi.2006.01.019

Wang, H., Chevalier, D., Larue, C., Ki, S., and John, C. (2007). The protein phosphatases and protein kinases of *Arabidopsis thaliana*. *Arabidopsis Book* 5:e0106.

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

Wang, M., Dai, W. S., Du, J., Ming, R. H., Dahro, B., and Liu, J. H. (2019a). ERF109 of trifoliate orange (*Poncirus trifoliata* (L.) Raf.) contributes to cold tolerance by directly regulating expression of Prx1 involved in antioxidative process. *Plant Biotechnol. J.* 17, 1316–1332. doi: 10.1111/pbi.13056

Wang, N.-N., Xu, S.-W., Sun, Y.-L., Liu, D., Zhou, L., Li, Y., et al. (2019b). The cotton WRKY transcription factor (GhWRKY33) reduces transgenic *Arabidopsis* resistance to drought stress. *Sci. Rep.* 9:724.

Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. an update to the 2007 review in annals of botany. *Ann. Bot.* 11, 1021–1058. doi: 10.1093/aob/mct067

Wei, W., Liang, D. W., Bian, X. H., Shen, M., Xiao, J. H., and Zhang, W. K. (2019). Gm WRKY 54 improves drought tolerance through activating genes in ABA and Ca 2+ signaling pathways in transgenic soybean. *Plant J.* 100, 384–398. doi: 10.1111/tpj.14449

Weng, H., Molina, I., Shockey, J., and Browse, J. (2010). Organ fusion and defective cuticle function in a lacs1lacs2 double mutant of *Arabidopsis*. *Planta* 231, 1089–1100. doi: 10.1007/s00425-010-1110-4

Xiang, Y., Huang, Y., and Xiong, L. (2007). Characterization of stress-responsive CIPK genes in rice for stress tolerance improvement. *Plant Physiol.* 144, 1416–1428. doi: 10.1104/pp.107.101295

Xiong, L., Schumaker, K. S., and Zhu, J. K. (2002). Cell signaling during cold. Drought, and Salt Stress. *Plant Cell* 14(Suppl.), S165–S183.

Xu, J., Tian, Y. S., Peng, R. H., Xiong, A. S., Zhu, B., Jin, X. F., et al. (2010). AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in *Arabidopsis*. *Planta* 231, 1251–1260. doi: 10.1007/s00425-010-1122-0

Xu, Y. J., Gao, S., Yang, Y. J., Huang, M. Y., Cheng, L. N., Wei, Q., et al. (2013). Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. *BMC Genom.* 14:662. doi: 10.1186/1471-2164-14-662

Yu, Y., Yang, D., Zhou, S., Gu, J., Wang, F., Dong, J., et al. (2017). The ethylene response factor OsERF109 negatively affects ethylene biosynthesis and drought tolerance in rice. *Protoplasma* 254, 401–408. doi: 10.1007/s00709-016-0960-4

Zandalinas, S. I., Damián, B., Vicent, A., and Aurelio, G. C. (2017). Modulation of antioxidant defense system is associated with combined drought and heat stress tolerance in citrus. *Front. Plant Sci.* 8:953. doi: 10.3389/fpls.2017.00953

Zang, B., Li, H., Li, W., Deng, X. W., and Wang, X. (2011). Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice. *Plant Mol. Biol.* 76, 507–522. doi: 10.1007/s11103-011-9781-1

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statist. Appl. Genet. Mol. Biol.* 4:Article17.

Zhang, L., Cheng, J., Sun, X., Zhao, T., Li, M., Wang, Q., et al. (2018). Overexpression of VaWRKY14 increases drought tolerance in *Arabidopsis* by modulating the expression of stress-related genes. *Plant Cell Rep.* 37, 1159–1172. doi: 10.1007/s00299-018-2302-9

Zhang, L., Zhang, L., Xia, C., Zhao, G., Jia, J., and Kong, X. (2015). The novel wheat transcription factor TaNAC47 enhances multiple abiotic stress tolerances in transgenic plants. *Front. Plant Sci.* 6:1174. doi: 10.3389/fpls.2015.01174

Zhang, X., Lu, S. C., Jiang, C. H., Wang, Y. F., Lv, B., Shen, J. B., et al. (2014). RcLEA, a late embryogenesis abundant protein gene isolated from Rosa chinensis, confers tolerance to *Escherichia coli* and *Arabidopsis thaliana* and stabilizes enzyme activity under diverse stresses. *Plant Mol. Biol.* 85, 333–347. doi: 10.1007/s11103-014-0192-y

Zhao, D. Q., Zhang, X. Y., Fang, Z. W., Wu, Y. Q., and Tao, J. (2019a). Physiological and transcriptomic analysis of tree peony (Paeonia section Moutan DC.) in response to drought stress. *Forests* 10:135. doi: 10.3390/f10020135

Zhao, K., Zhang, D., Lv, K., Zhang, X., and Jiang, T. (2019b). Functional characterization of poplar WRKY75 in salt and osmotic tolerance. *Plant Sci.* 289:110259. doi: 10.1016/j.plantsci.2019.110259

Zhao, Y., Yu, W., Hu, X., Shi, Y., Liu, Y., Zhong, Y., et al. (2018). Physiological and transcriptomic analysis revealed the involvement of crucial factors in heat stress response of *Rhododendron hainanense*. *Gene* 660, 109–119. doi: 10.1016/j.gene.2018.03.082

Zhou, A., Ma, H., Liu, E., Jiang, T., Shuang, F., Gong, S., et al. (2017). Transcriptome sequencing of dianthus spiculifolius and analysis of the genes involved in responses to combined cold and drought stress. *Int. J. Mol. Sci.* 18:849. doi: 10.3390/ijms18040849

Zhou, R., Yu, X., Zhao, T., Ottosen, C. O., Rosenqvist, E., and Wu, Z. (2019). Physiological analysis and transcriptome sequencing reveal the effects of combined cold and drought on tomato leaf. *BMC Plant Biol.* 19:377. doi: 10.1186/s12870-019-1982-9

Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

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

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.

Copyright © 2021 Jia, Feng, Bu, Ji, Lyu and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fgene.2021.700469

Identification of Candidate Genes Regulating the Seed Coat Color Trait in Sesame (*Sesamum indicum* L.) Using an Integrated Approach of QTL Mapping and Transcriptome Analysis

*Chun Li[1,2], Yinghui Duan[1,2], Hongmei Miao[1,2], Ming Ju[1,2], Libin Wei[1] and Haiyang Zhang[1,2]**

[1] Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, China, [2] Henan Key Laboratory of Specific Oilseed Crops Genomics, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, China

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Hong-Bin Zhang,
Texas A&M University, United States
Xingwei Zheng,
Shanxi Academy of Agricultural Sciences, China
Xiaobiao Zhu,
Anhui Agricultural University, China

***Correspondence:**
Haiyang Zhang
zhanghaiyang@zzu.edu.cn

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

Received: 26 April 2021
Accepted: 05 July 2021
Published: 04 August 2021

Citation:
Li C, Duan Y, Miao H, Ju M, Wei L and Zhang H (2021) Identification of Candidate Genes Regulating the Seed Coat Color Trait in Sesame (Sesamum indicum L.) Using an Integrated Approach of QTL Mapping and Transcriptome Analysis. Front. Genet. 12:700469. doi: 10.3389/fgene.2021.700469

Seed coat color is an important seed quality trait in sesame. However, the genetic mechanism of seed coat color variation remains elusive in sesame. We conducted a QTL mapping of the seed coat color trait in sesame using an F_2 mapping population. With the aid of the newly constructed superdense genetic linkage map comprised of 22,375 bins distributed in 13 linkage groups (LGs), 17 QTLs of the three indices (i.e., *L*, *a*, and *b* values) of seed coat color were detected in seven intervals on four LGs, with a phenotype variance explanation rate of 4.46–41.53%. A new QTL *qSCa6.1* on LG 6 and a QTL hotspot containing at least four QTLs on LG 9 were further identified. Variants screening of the target intervals showed that there were 84 genes which possessed the variants that were high-impact and co-segregating with the seed coat color trait. Meanwhile, we performed the transcriptome comparison of the developing seeds of a white- and a black-seeded variety, and found that the differentially expressed genes were significantly enriched in 37 pathways, including three pigment biosynthesis related pathways. Integration of variants screening and transcriptome comparison results suggested that 28 candidate genes probably participated in the regulation of the seed coat color in sesame; of which, 10 genes had been proved or suggested to be involved in pigments biosynthesis or accumulation during seed formation. The findings gave the basis for the mechanism of seed coat color regulation in sesame, and exhibited the effects of the integrated approach of genome resequencing and transcriptome analysis on the genetics analysis of the complex traits.

Keywords: seed coat color, sesame, genetic linkage map, genome re-sequencing, RNA-Seq

INTRODUCTION

Sesame (*Sesamum indicum* L.) is a specific and important oilseed crop with a long cultivation history (Bedigian, 2003). At present, sesame is mainly grown in Asia, Africa, and Latin America. Sesame seeds possess abundant nutritional substances, such as unsaturated fatty acids, proteins, digestible fiber, and beneficial antioxidants (such as lignans) (Namiki, 2007; Kumar et al., 2013;

Ha et al., 2017; Keskin, 2019). For sesame seeds, seed coat color varies from white, shallow yellow, yellow, golden, brown, gray, reddish brown, and other medium colors, to black. In particular, the white- and the black-seeded seeds are more popular for consumption in the world. Black sesame seeds are shown to possess abundant nutritional facts (Ganesan and Xu, 2017; Khoo et al., 2017). Some studies showed that the contents of some substances, such as phenolics, vitamins, and phloretin, in black sesame seeds are high (Ha et al., 2017; Wang et al., 2018). In China, black sesame seeds have also been used as a traditional Chinese medicine in the last 3,000 years (Wang et al., 2018; Zhang H. et al., 2019).

In sesame, seed coat color trait is considered as a quantitative trait (Zhang H. et al., 2013; Wang L. et al., 2016). Zhang H. et al. (2013) firstly performed the genetic background and QTL mapping analysis of sesame seed coat color with an F_2 hybridization population using simple sequence repeat (SSR), amplified fragment length polymorphism (AFLP), and random selective amplification of microsatellite polymorphic loci (RSAMPL) markers. Four QTLs were located on three LGs with the phenotypic variation from 9.6 to 39.95% (Zhang H. et al., 2013). Subsequently, Wang L. et al. (2016) investigated the seed coat color indices of the *L*, *a*, and *b* values of a recombinant inbred lines (RIL) population and determined four QTLs on a genetic map constructed by restriction-site associated DNA sequencing (RAD-Seq). The four QTLs were found to be on three LGs and with the phenotypic variation from 3 to 46%. Wei et al. (2015) performed a genome-wide association study (GWAS) for the seed coat color and some other traits using a panel of 705 sesame accessions and determined seven loci associated with the seed coat color. Meanwhile, Du et al. (2019) detected seven QTLs related with seed coat color on three LGs using specific length amplified fragment sequencing (SLAF-seq) and an F_2 population. Wang et al. (2020) identified 20 candidate genes associated with pigment biosynthesis using transcriptome data from a black and white sesame variety. However, only the *SiPPO* (polyphenol oxidase) gene was determined for the seed coat color in sesame until now (Wei et al., 2015).

For angiosperm plants, the seed coat color trait is predominantly determined by the content of various pigments, which mainly belong to the secondary flavonoid metabolites, such as flavonols, anthocyanins, and proanthocyanidins (PAs), and are synthesized through the flavonoid biosynthesis pathway (Winkel-Shirley, 2001). At present, dozens of genes regulating the seed coat color regulation have been detected in crops; of which, most genes encode the enzymes related to flavonoid biosynthesis (such as *C2* in maize, and *TT6* and *TT7* in *Arabidopsis*), transfer proteins (*ZmMRP3* in maize and *TT12* and *AHA10* in *Arabidopsis*), or regulatory factors (*TT1* and *TT10* in *Arabidopsis*) (Pourcel et al., 2005; Marinova et al., 2007; Owens et al., 2008; Badone et al., 2010; Han et al., 2010; Appelhagen et al., 2011, 2015; Eloy et al., 2017). Flavonoid biosynthesis pathway is highly conserved in different plant species (Dong et al., 2001; Tohge et al., 2017). To date, numerous transcription factors, such as MYBs, basic helix-loop-helix (bHLHs), Transparent Testa Glabra1 (*TTG1*), and MYB-bHLH-WDR (MBW) ternary complexes, have been found to be participating in the regulation of flavonoid biosynthesis pathway in a tissue-specific manner (Gonzalez et al., 2009; Xu et al., 2014). In addition, post-translation modification and microRNA have also been proven to be related to the flavonoid biosynthesis pathway (Wang H. et al., 2016).

In this study, in order to identify the major effect of genes underlying sesame seed coat color, we performed a QTL mapping for the sesame seed coat color trait based on a superdense genetic linkage map constructed using an F_2 population and genome resequencing technology. The transcriptome comparison analysis of a black- and white-seeded variety during seed development was also carried out. As a result, a total of 28 candidate genes were determined for the seed coat color trait in sesame according to the integration of QTL mapping and transcriptome analysis. The findings provide a deep understanding of the genetic mechanism underlying the seed coat color regulation in sesame.

MATERIALS AND METHODS

Materials and Data

Two varieties, Yuzhi DS899 (white-seeded, mutated from a sibling of Yuzhi 11) and JS012 (black-seeded), and their 120 F_2 individuals were used for the QTL mapping in this study (Zhang et al., 2016). The F_2 population and both parents were grown in Sanya (109°50′E and 18°25′N), Hainan, China in 2014 for seed coat color trait evaluation. For the transcriptome analysis, two varieties with a contrasting seed coat color, i.e., Yuzhi 11 and the black-seeded parent JS012, were cultivated in Yuanyang (113°97′E and 35°05′N), China in 2020. Twenty capsules from 10 plantlets of each variety at the time points 5, 10, 15, 20, and 25 DAF, respectively, were collected. The developing seeds were then peeled from the capsules, frozen in liquid nitrogen, and stored at −80°C for RNA extraction. Three biological replicates were set for RNA sequencing. All the materials above were chosen from the sesame germplasm reservoir of Henan Sesame Research Center, Henan Academy of Agricultural Sciences (HSRC, HAAS).

The genome resequencing data (reserved in NCBI Bioproject PRJNA315474) of Yuzhi DS899 and JS012 and their 120 F_2 individuals (Zhang et al., 2016) were used for the genetic map construction and QTL mapping in this study. The sesame genome of v*ar.* Yuzhi11 (white-seeded; version 2) (Zhang et al., 2016) was used as the reference genome for genotyping, variant digging, and candidate gene screening analysis.

Seed Coat Color Trait Evaluation

Seed coat color of the F_2 family was investigated based on a single plant. For each of the sequenced 120 F_2 individuals above, the seeds were harvested after the plant was totally matured. The seeds from each individual were randomly divided into three equal pieces (used as three replicates) after the unmatured seeds were discarded. As to evaluate the seed coat color, the seeds were put into a quartz cell, and then scanned under the Colorflex EZ spectrophotometer (Hunter Associates Laboratory, United States). Three indices (i.e., *L*, *a*, and *b* values) were measured and recorded. Of which, "*L*" defines the luminosity of

the seed coat, with a range from –100 for black to 100 for white; "*a*" and "*b*" indicate the shade of color pairs, with a range from –60 for green to +60 for red, and from –60 for blue to +60 for yellow, respectively. For each of the indices, the mean value of the three replicates was finally used.

Construction of a Superdense Genetic Linkage Map for Sesame

Sequencing reads of the parents Yuzhi DS899 and JS012 and their 120 F_2 individuals (Zhang et al., 2016) were firstly filtered using the Trimmomatic 0.33 software (Bolger et al., 2014), and then mapped to the reference sesame genome (*cv.* Yuzhi 11) (Zhang et al., 2016) using the Burrows Wheeler Aligner (BWA) 0.7.17 software with the default settings (Li and Durbin, 2009). Joint variant calling was performed using the GATK 4.0 program following its best practice (Van der Auwera et al., 2013). Genotype coding was performed using the GenosToABH plugin in TASSEL 5.2.43 (Glaubitz et al., 2014). The homozygous genotype of the parent Yuzhi DS899 was coded as "A," the parent JS012 as "B," and the heterozygous genotype of their hybrid as "H." Since several consecutive heterozygotes ("H") would be called as identical homozygotes ("A" or "B") within a very short distance due to the low reads coverage, the coded genotype was processed according to Miao et al. (2018) using the following procedures to improve the accuracy of linkage map construction: (1) combining the consecutive and redundant (with identical information for all individuals in the progeny) markers within a short region into a representative marker, and the minimum length between two continuous markers was set to 100 bp (base pair); (2) correcting the genotype of representative markers based on a sliding window of 15 markers; (3) based on the corrected markers, grouping the adjacent and redundant markers into bins; and (4) filtering out the bins of distorted segregation with the number of plants with "A" or "B" genotypes <20, and the number of plants with H genotypes <40. The threshold for distorted segregation loci was selected based on a pilot study of linkage map construction, in which we found that at least 80 individuals (with ideal segregation ratio 1:2:1 for genotype "A," "B," and "H") were needed to form 13 linkage groups (corresponding 13 chromosomes of sesame).

The genetic map was constructed based on the filtered genotype matrix using the Lep-MAP3 program with the default parameters and a logarithm of the odds (LOD) of 12.0 to separate chromosomes (Rastas, 2017). LGs with less than 100 markers were discarded. The Kosambi function was applied to transfer the recombinant rate into the genetic distance (Morgan).

QTL Mapping and Comparison

QTL mapping was performed using the QTL Cartographer 2.5 (Wang et al., 2012), QTL.gCIMapping.GUI v1.0 (Wen et al., 2019; Zhang Y. W. et al., 2019), and QTL IciMapping 4.2.53 (Meng et al., 2015), respectively. The Composite interval mapping (CIM) method in the QTL Cartographer, the Genome-wide composite interval mapping (GCIM) method in the QTL.gCIMapping.GUI, and the Inclusive composite interval mapping (ICIM) and Interval mapping (IM) methods in the QTL IciMapping were individually used to detect the QTLs and to estimate the effects. For the CIM method, the map was scanned in 2 cM (centimorgan) intervals with a window size of 10 cM, and the LOD was determined by running 1,000 permutation tests. For the GCIM method, random model, critical LOD score = 2.5, and walk step for genome wide scanning = 1 cM were used. For the ICIM and IM methods, LOD was determined by running 1,000 permutation tests at type I error equals 0.05, and walk step was set at 1 cM. The ratio of phenotypic variance explained by genotype was taken from the marker at the peak QTL position. QTL physical position was determined by the flanking markers.

To make a distinction between the new QTLs identified in this study and from the reported results, QTL position comparison was performed with flanking sequences using the MUMER 4 program (Marcais et al., 2018). The flanking sequences of the QTLs or the associated loci were applied as the queries for alignment against the sesame reference genome (*cv.* Yuzhi 11) (Zhang et al., 2016).

RNA-Seq and Analysis

Total RNA of each sample was extracted using the RNAiso Plus Reagents (TaKaRa, Dalian, China). Genomic DNA was removed by DNase I treatment. RNA-seq libraries were constructed and sequenced on the Illumina platform HiSeq 2500 (Illumina, Inc., San Diego, CA, United States). The clean reads were mapped to the reference sesame genome using the STAR 2.7.0 (Dobin et al., 2013). The reads mapped to each genomic feature were counted using the featureCounts 2.0.1 (Yang et al., 2014). Default parameters were used for both read mapping with STAR and read counting with featureCounts. Differentially expressed genes (DEG) were screened and analyzed using the DESeq 2 software with the default parameters (Love et al., 2014). False discovery rate (FDR) adjusted p-values < 0.05 and absolute value of $\log_2$ (fold change) > 1 were set as the cutoff for the DEGs.

Gene Function Annotation, Homologous Gene Screening, and Variant Effect Prediction

Gene function annotation was performed using the eggNOG-Mapper 5.0.1 (Huerta-Cepas et al., 2017). Gene ontology (GO) plotting was conducted in the WEGO 2.0 (Ye et al., 2018). Homologous gene screening in the target intervals was conducted using the BLAST + 2.2.31 with the E-value < 1e-10. The effect of gene variant was estimated using the SnpEff 4.1 (Cingolani et al., 2012), and the variants that caused frame shift, stop gain, stop lost, start lost, splice acceptor, splice donor, exon loss, frame-shift, or nonsense mutation were regarded as high-impact and screened according to Cingolani et al. (2012). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the R package clusterProfiler 3.12 with the default parameters (Yu et al., 2012), and a q-value < 0.05 was set as the cutoff.

RESULTS

Seed Coat Color Genetic Analysis of the Mapping Population

In order to clarify the genetic background of the seed coat color in sesame, we firstly investigated the phenotypic variation of the seed coat color trait of the F_2 population in 2016 (**Supplementary Table 1** and **Supplementary Figure 1**). *L*, *a*, and *b* values of the white-seeded parent Yuzhi DS899 were 62.81, 5.50, and 19.62, respectively, while those of the black seeded parent were 21.62, 6.73, and 7.27, respectively. All the F_1 seeds were black, whereas F_2 plants exhibited an obvious variation in seed coat color with various values of *L*, *a*, and *b* (**Supplementary Figure 1**). As to the 120 F_2 plants, "*L*" values ranged from 16.23 to 62.81, with a standard deviation of 8.40; the "*a*" values ranged from 1.20 to 9.34, with a standard deviation of 2.67; while "*b*" varied from 1.42 to 19.62, with a standard deviation of 4.27 (**Supplementary Table 1**). In particular, we found some F_2 plants that presented a darker seed coat color than the black parent (**Supplementary Figure 1**). However, the color of all the F_2 plantlets was darker than the white-seeded parent.

We further performed the pairwise correlation analysis of the three indicators of the seed coat color in the population (**Supplementary Table 1**). The results showed that the *L*, *a*, and *b* values were significantly correlated with each other. The *L*, *a*, and *b* values of the F_2 population were all out of the normal distribution (**Figure 1** and **Supplementary Table 1**), indicating the existences of a few major QTLs controlling the three indicators.

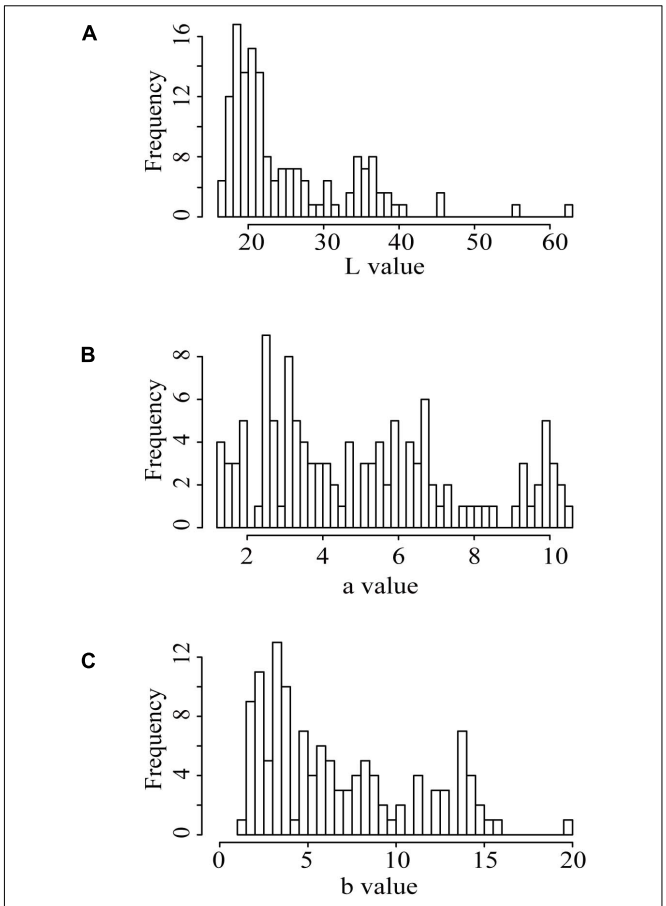


FIGURE 1 | Distribution of the seed coat color indices presented by the *L*, *a*, and *b* values in the F_2 population. **(A)** "*L*" value; **(B)** "*a*" value; **(C)** "*b*" value.

Construction of a Superdense Genetic Linkage Map

Before determining the QTLs for the seed coat color trait, we firstly constructed a superdense genetic map using the F_2 population above and the genome resequencing data. Based on the reference genome (*var.* Yuzhi 11), a total of 1,264,646 SNP/InDel variants for the two parents and their 120 F_2 plants were identified (**Supplementary Table 2**). The number of unique variants for the two parents Yuzhi DS899 and JS012 was 122,361 and 528,671, respectively. After filtered, a total of 425,011 variants were successfully genotyped into "A," "B," or "H" types in the population (**Supplementary Table 2**).

Then, we constructed the matrix consisting of the above 425,011 genotype codes for the 120 F_2 individuals and fed it to the linkage map construction pipeline (**Table 1**). As the bins with the distorted segregation (the numbers of plants having A/B genotypes < 20 and H genotypes < 40) were excluded, a total of 22,375 bins containing 380,544 SNP/InDel markers were obtained for the genetic map construction (**Table 1** and **Supplementary Table 3**). As a result, a genetic map with 13 LGs was successfully constructed (**Figure 2**). The total length of the genetic map was 1,576.14 cM, and the length per LG ranged from 96.96 to 157.96 cM. The average marker interval was 0.82 cM per bin marker. The marker density was 14.20 bins or 241.44 SNP/InDel markers per cM. In addition, 34 gaps with the length ≥ 5.0 cM remained in the genetic map, with the largest gap size of 18.27 cM (**Table 1**).

QTL Mapping of the Seed Coat Color Trait in the F_2 Population

We then performed a QTL mapping of the seed coat color in the crossing population using four methods (GCIM, ICIM, IM, and CIM) (**Table 2** and **Supplementary Table 4**). The results from the four methods were similar, and QTLs were found on LG3, 5, 6, and 9 for the three seed coat color indicators *L*, *a*, and *b* (**Table 2** and **Supplementary Table 4**). Of these, the QTL intervals observed using the CIM method were usually wider, while the intervals from the GCIM method were narrower. The overlapping relationship and boundaries between the QTLs were obvious (**Supplementary Table 4**). Therefore, based on the QTL interval boundaries, we summarized the QTL mapping resulting into 17 QTLs, including 5, 4, and 8 QTLs for *L*, *a*, and *b*, respectively (**Table 2**). The genetic effect (the explanation rate of phenotype variance or V_G/V_P value) of the QTLs on the seed coat color trait ranged from 4.66% (for *qSCb5.2* estimated by GCIM) to 41.53% (for *qSCL9.3* estimated by ICIM).

TABLE 1 | Characteristics of the superdense genetic linkage map for sesame.

LG	Length (cM)	Bin number	Variants number	Average interval (cM)	Gap number[1]	Largest gap (cM)
1	110.96	1,512	27,577	0.66	2	7.56
2	106.05	1,823	28,880	0.87	3	12.97
3	105.61	981	20,903	1.03	4	14.57
4	118.46	2,234	37,066	0.63	1	8.41
5	102.08	1,339	18,558	0.77	2	11.60
6	137.67	1,794	26,621	0.81	3	11.47
7	141.10	3,090	49,745	0.73	2	15.97
8	126.98	1,930	32,522	0.86	3	13.34
9	157.96	1,705	24,462	1.36	5	18.27
10	122.99	1,383	29,655	0.8	2	17.36
11	96.97	1,379	25,062	0.81	2	9.30
12	110.80	1,835	30,175	0.67	1	13.27
13	138.52	1,370	29,318	0.89	4	13.27
Total	1,576.15	22,375	380,544	0.82	34	18.27

[1]*Only the marker interval with the size of >5.0 cM was considered as a gap.*

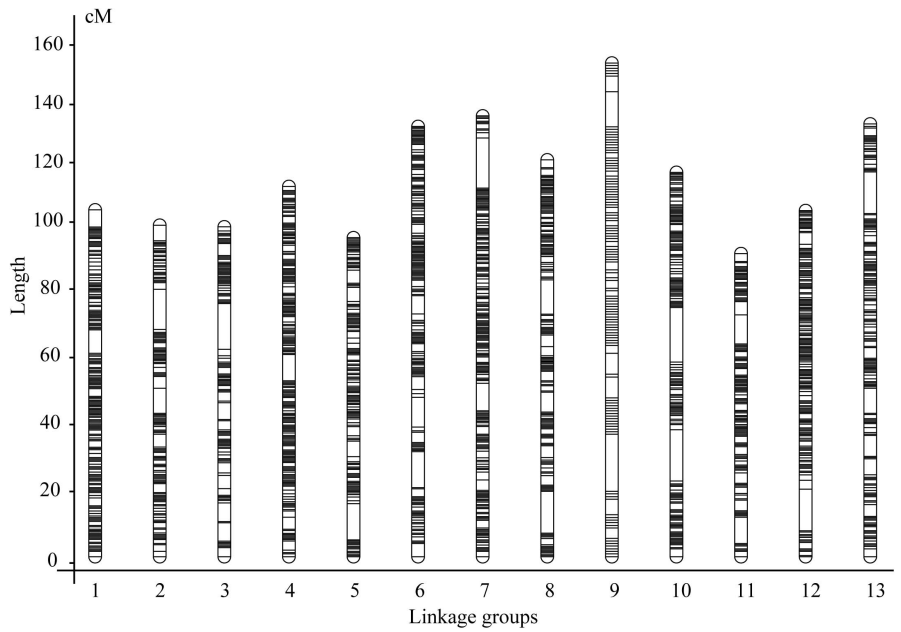


FIGURE 2 | The superdense genetic map constructed using an improved pipeline and the F_2 population derived from a cross between Yuzhi DS899 and JS012.

Due to the overlapping positions of the QTLs controlling *L*, *a*, and *b*, a total of seven non-overlapping intervals, with one on LG3 (43.75–47.09 cM), one on LG5 (20.44–22.11 cM), one on LG6 (78.80–84.23 cM), and four on LG9 (9.17–20.4, 20.4–21.23, 21.23–27.07, and 28.73–30.40 cM), were identified for the sesame seed coat color in this study (**Table 2**). As the QTLs intervals on LG9 were adjacent to each other, we thus named the QTL hotspot "*qSC* cluster" for the seed coat color trait. By further merging this *qSC* cluster into one interval, a total of four intervals (hereinafter referred to as "*qSC* intervals") on four LGs were thus harvested for the seed coat color trait in sesame, i.e., LG3: 43.75–47.09 cM, LG5: 20.44–22.11 cM, LG6: 78.80–84.23 cM, and LG9: 9.17–30.40 cM (**Table 2**).

Genes in the qSC Intervals

We further screened the genes in the four *qSC* intervals according to the fine sesame reference genome (**Table 3**). The results showed that the above *qSC* intervals were located into six contigs (i.e., C34, C_2_7, C4, U0167, C3, and C_2_3) in the reference genome. The physical size of the six contig segments varied from 26.80 Kb (kilobases) (in contig U0167) to 2,090.72 Kb (in contig C_2_3). A total of 625 genes were detected in the six contig segments, and the number of genes in each of the contig segments varied from one on contig C4 to 235 on C_2_3 (**Table 3** and **Supplementary Table 5**).

In order to determine the candidate genes, we firstly compared the two parents and determined that a total of 10,719 SNP/InDel

TABLE 2 | QTL distribution of the seed coat color using *L*, *a*, and *b* values in sesame.

Trait	LG	QTL	Left marker		Right marker		VG/VP (%)[1]	Peak LOD[2]	Detected method
			Name	Pos (cM)	Name	Pos (cM)			
L	3	*qSCL3.1*	C34-3876888	43.75	C34-3707103	47.09	6.58, 28.19	8.84, 15.00, 8.11	GCIM, ICIM, and CIM
	9	*qSCL9.1*	C3-4785376	9.17	C3-5518941	20.40	24.61,	13.82, 23.35	IM and CIM
		qSCL9.2	C3-5518941	20.40	C3-5518941	21.23	6.65	9.38	GCIM
		qSCL9.3	C3-5518941	21.23	C_2_3-7327811	27.07	39.42, 41.53, 20.57	38.07, 19.86, 24.11, 39.44	GCIM, ICIM, IM, and CIM
		qSCL9.4	C_2_3-7239571	28.73	C_2_3-6815285	30.40	19.39	19.36, 24.82	IM and CIM
a	5	*qSCa5.1*	C_2_7-2469423	20.44	C_2_7-2494612	22.11	7.00, 14.80	7.20, 7.49, 6.88	GCIM, ICIM, and CIM
	6	*qSCa6.1*	C4-7225533	78.80	U0167-32152	84.23	7.99, 11.92	6.44, 6.54, 7.01	GCIM, ICIM, and CIM
	9	*qSCa9.1*	C3-5590000	21.23	C_2_3-7327811	27.07	13.38, 31.97, 24.42	16.60, 15.19, 17.95, 18.45	GCIM, ICIM, IM, and CIM
		qSCa9.2	C_2_3-7239571	28.73	C_2_3-6815285	30.40	24.59	15.91, 16.08	IM and CIM
b	3	*qSCb3.1*	C34-3876888	43.75	C34-4228322	44.17	14.80	7.49	ICIM,
		qSCb3.2	C34-4201909	44.59	C34-4201909	44.59	4.73	5.16	GCIM
	5	*qSCb5.1*	C_2_7-2469423	20.44	C_2_7-2475883	21.27	11.96,	6.08	ICIM
		qSCb5.2	C_2_7-2476632	22.11	C_2_7-2476632	22.11	4.66	5.77	GCIM
	9	*qSCb9.1*	C3-4785376	9.17	C3-5518941	20.40	22.85,	10.47, 29.43	IM and CIM
		qSCb9.2	C3-5518941	20.40	C3-5518941	21.23	12.50	10.94	GCIM
		qSCb9.3	C3-5518941	21.23	C_2_3-7327811	27.07	31.58, 32.93, 19.33	29.97, 14.08, 22.60, 36.17	GCIM, ICIM, IM, and CIM
		qSCb9.4	C_2_3-7239571	28.73	C_2_3-6815285	30.40	18.00	17.76, 22.42	IM and CIM

[1]The values from the CIM method were not included because the QTLs mapped on LG9 are clustered, CIM seems unable to distinguish the effect of each QTL. [2]Peak LOD was calculated according to the GCIM, ICIM, IM, and CIM methods, respectively. The detail information is listed in ***Supplementary Table 4****.*

TABLE 3 | Genes and variants in the four merged *qSC* intervals based on the sesame genome data.

LG	*qSC* Interval (cM)	Physical position (Kb)[1]				Gene number	SNP/InDel number[2]	Unique in JS012[3]	
		Located contig	Start	End	Size			Low-impact	High-impact
3	43.75–47.09	C34	3,690.83	4,394.65	703.82	94	2,705	2,455	101
5	20.44–22.11	C_2_7	2,344.66	2,796.34	451.68	60	2,723	2,367	142
6	78.80–84.23	C4	7,225.53	7,261.46	35.93	1	7	7	0
		U0167	5.35	32.15	26.80	4	346	329	5
9	9.17–30.40	C3	4,285.99	6,348.88	2,062.88	231	695	163	3
		C_2_3	6,429.48	8,520.20	2,090.72	235	4,243	3,679	142
Total	–	–	–	–		625	10,719	9,000	393

[1]The physical location of the QTL intervals was calculated based on the genome Yuzhi 11 (Zhang et al., 2016).
[2]SNP and InDel number is calculated based on the white- and black-seeded parents joint calling; "–" Not detected.
[3]Unique in JS012 (the black-seeded parent), comparing with Yuzhi DS899 (the white-seeded parent) and Yuzhi 11 (the variety used for reference genome sequencing, white-seeded). See ***Supplementary Table 5*** *for detail.*

variants existed in the six target contig segments between the parents (**Table 3** and **Supplementary Table 6**). Of which, 9,393 variants were unique in JS012 (black-seeded) compared with Yuzhi DS899 (white-seeded) and Yuzhi 11 (the variety used for reference genome sequencing, white-seeded), and thus, co-segregated with the seed coat color trait. Annotation results of SnpEff indicated that 393 variants in 84 genes caused frame shift, stop gain, stop lost, start lost, splice acceptor, splice donor, exon loss, frame-shift, or nonsense mutation, and were named as high-impact variants (**Table 3** and **Supplementary Table 6**).

DEGs of White- and Black-Seeded Varieties

To further explore the characteristics of the genes related with the seed coat color trait, we compared the expression profiles of the developing seeds in a white variety (Yuzhi 11) and a black variety (JS012) (**Supplementary Figure 2**). A total of 1,210.93 million reads (181.64 Gbases), with an average of 43.25 million clean reads, for the samples were harvested, and the mapping rates to the reference were all above 96.03% (**Supplementary Table 7**). The number of DEGs at the stages 5, 10, 15, 20, and 25 DAF between the two varieties were 2,148 (838 up and 1,310 down expressed in JS012), 5,176 (3,019 up and 2,157 down), 3,725 (1,598 up and 2,127 down), 2,984 (1,092 up and 1,892 down), and 5,115 (2,115 up and 3,000 down), respectively (**Supplementary Table 8**). GO annotation showed that the DEGs from the five stages presented a similar functional profile (**Supplementary Figure 3**). Most DEGs were involved in metabolic and cellular biological processes and the molecular functions of catalytic activity and



FIGURE 3 | Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the differentially expressed genes (DEGs) identified at the stages 5, 10, 15, 20, and 25 DAF between JS012 and Yuzhi 11.

binding. We further performed the KEGG enrichment analysis and found that the DEGs were significantly enriched in 37 pathways (q_value < 0.05), including pigment biosynthesis related phenylpropanoid biosynthesis (KEGG: ko00940, at the stages 10, 15, 20, and 25 DAF), flavonoid biosynthesis (KEGG: ko00941, at the stages 10, 15, and 25 DAF), and anthocyanin Biosynthesis (KEGG: ko00942, at the stage 10 DAF) pathways (**Supplementary Table 9** and **Figure 3**). Notably, the three pigment biosynthesis related pathways above were enriched after the stage 5 DAF, and no genes in the three pathways were located in our *qSC* intervals (**Supplementary Table 9**).

Seed Coat Color Candidate Genes Identification in White- and Black-Seeded Varieties

In order to screen the candidate genes related with the seed coat color regulation in sesame, we compared the expression profiles of the 625 genes in the *qSC* intervals during the seed development in the two varieties (**Supplementary Table 10**). Of the 625 genes, 62, 138, 101, 83, and 135 were differentially expressed at the stages 5, 10, 15, 20, and 25 DAF, respectively (**Supplementary Table 10**). Considering that the pigment was accumulated in JS012 gradually from 8 DAF and that no enrichment of the seed coat color formation related pathways were observed at the stage 5 DAF, we took the stage 5 DAF as the control and the other stages as the case to search for the candidate genes, and a Venn diagram consisting of the genes that were DEGs at the stages 10, 15, 20, and 25 DAF and not-DEGs at the stage 5 DAF was created (**Figure 4A**). We found 223 genes (the sum of the red numbers in **Figure 4**) that were DEGs in at least at one of the stages 10, 15, 20, and 25 DAF but not at 5 DAF. By selecting the intersection of the 223 genes and the 84 genes containing high-impacts variants (and co-segregated with the seed coat color trait; **Supplementary Table 6**), 28 genes were identified (**Figure 4B**).

The 28 genes were selected as the highly-potential candidate genes for the sesame seed coat color. Of the 28 candidates, seven were located on LG3, 2 on LG5, and 19 on LG9, while no genes were found on LG6 (**Table 4**). The 28 genes had various

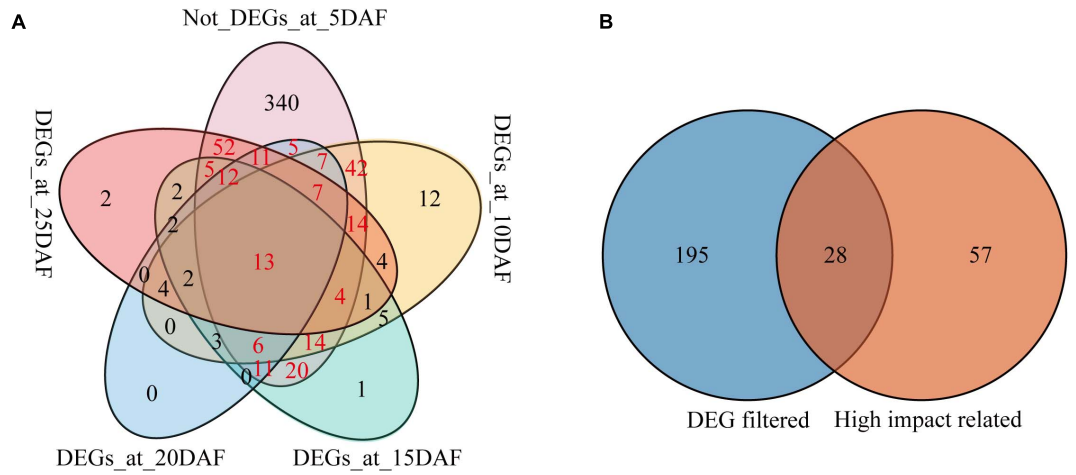


FIGURE 4 | Venn diagram of genes in the *qSC* intervals. **(A)** Venn diagram of the DEGs after 5 DAF and not-DEGs at the stage 5 DAF; **(B)** Venn diagram of the genes from DEG filtered (the genes that were DEGs in at least one of the stages 10, 15, 20 and 25 DAF but not at 5 DAF), and high-impact related (the genes containing high-impacts variants and co-segregated with the seed coat color trait).

functions, including E3 ubiquitin-protein ligase (*C_2_3.773* and *C_2_3.787*), oxidase (*C_2_7.207* and *C34.291*), transcription factor (*C_2_3.819* and *C34.285*), and so on (**Table 4**). Of the 28 genes, 14 were expressed predominantly in JS012 (**Supplementary Figure 4A**), 11 expressed predominantly in Yuzhi 11 (**Supplementary Figure 4B**), and the other three genes presented the hybridized expression pattern (**Supplementary Figure 4C**).

DISCUSSION

The Multiple-Fold Improvement of the Sesame Genetic Linkage Map

In this study, a superdense genetic linkage map was constructed for sesame by genome resequencing and an updated map construction pipeline. Compared with the published genetic maps (Zhang Y. et al., 2013; Zhang et al., 2018; Wu et al., 2014; Uncu et al., 2016; Wang L. et al., 2016; Mei et al., 2017; Du et al., 2019; Asekova et al., 2021; Liang et al., 2021), the new genetic linkage map contains the highest number of both SNP/InDel markers (380,544) and bins (22,375) for sesame to date.

During the genetic map construction, the genotype data which resulted from genome resequencing usually have many noisy genotypes and always lead to inaccurate co-segregating marker (bins) grouping. To overcome the problem above, we developed a genetic map construction pipeline in 2016 and updated the method in this study. The previous pipeline basically consists of five steps: (1) calling variants individually; (2) choosing the variant loci that were called in either parent and in at least 20 progenies; (3) filtering out the loci that were heterozygous in either parent or that have a read depth under five in any progeny; (4) genotype coding; and (5) genotype correction (Zhang et al., 2016). In this study, the pipeline was improved, and the genotype matrix for map construction was prepared as follows: (1) calling variants jointly; (2) performing genotype coding; (3) combining the consecutive same markers within a short region; (4) genotype correction; and (5) filtering out the loci of distorted segregation. In the updated pipeline, joint variant calling was applied with a higher efficiency than the previous single-sample variant calling and loci filtration strategy. Joint variant calling shared the information across all samples in a cohort and therefore had greater sensitivity, particularly for the site with a low sequence coverage or misalignment (Van der Auwera et al., 2013). As to the loci filtration strategy, in order to "rescue" the deleted but informative sites, we just filtered out the loci with distorted segregation and omitted the filtration step before genotype correction. Thus, more makers and bins could be applied for map construction. In the study, the located SNP/InDel makers increased from 30,193 to 380,544 in the genetic map, with more than an 11-fold increase. The number of bins in the new map increased from 3,041 to 22,375, with more than a six-fold increase. Furthermore, to improve the accuracy of the pipeline, we had two times of merging involved in the pipeline, i.e., combining the consecutive and redundant markers within a short region into a representative marker, and grouping the adjacent and redundant markers into bins. The involvement of merging twice lead to an average of 17 markers per bin in the new map, and, thus, the new map should be of a greater accuracy.

The Superdense Genetic Map Had More Robustly Dissected the Genetic Mechanism Underlying the Seed Coat Color Trait

Seed coat color is one of the earliest traits studied in crops for genetics analysis and Mendel's law discovery. In sesame, seed coat color seems to be correlated with the species evolution and the beneficial effects on seed nutrition (Namiki, 2007; Zhang et al., 2012). Based on our investigation, the sesame seed

TABLE 4 | The 28 candidate genes related to seed coat color in sesame.

LG	Gene	Annotated function	Related references[1]
3	*C34.269*	Cyclic nucleotide-gated ion channel 1-like	
	C34.277	TRAFAC class dynamin-related protein 5A-like	
	C34.281	Protein aspartic protease in guard cell 2-like	
	C34.285	Transcription factor bHLH13-like isoform X1	Nesi et al., 2000; Nesi et al., 2001
	C34.289	Uncharacterized protein At4g00950-like	
	C34.291	Pyridoxine/pyridoxamine 5′-phosphate oxidase 2	Sang et al., 2011; Frank et al., 1999
	C34.308	NAC domain-containing protein 43-like	
5	*C_2_7.207*	Polyphenol oxidase I	Wei et al., 2015
	C_2_7.220	Hyoscyamine 6-dioxygenase	Polturak and Aharoni, 2018
9	*C_2_3.640*	Uncharacterized protein C1F8.04c-like	
	C_2_3.644	Serine/threonine-protein kinase CDL1-like	
	C_2_3.683	Xyloglucan endotransglucosylase/hydrolase protein 5	Wang et al., 2019
	C_2_3.684	Uncharacterized protein LOC105168501	
	C_2_3.743	Late blight resistance protein homolog R1A-4	
	C_2_3.763	Short-chain dehydrogenase TIC 32	
	C_2_3.765	Cell division cycle protein 48 homolog	
	C_2_3.766	Phosphatidylinositol transfer protein CSR1-like isoform X2	
	C_2_3.767	Purine permease 3-like	
	C_2_3.768	Deoxycytidylate deaminase	
	C_2_3.771	Uncharacterized protein LOC105179396	
	C_2_3.773	E3 ubiquitin-protein ligase BOI-like isoform X1	Tang et al., 2015
	C_2_3.774	Triose phosphate/phosphate translocator	Hilgers et al., 2018
	C_2_3.780	Histidine biosynthesis bifunctional protein hisIE, chloroplastic	
	C_2_3.782	Uncharacterized protein LOC105179408	
	C_2_3.787	E3 ubiquitin-protein ligase SIS3-like	Tang et al., 2015
	C_2_3.819	Transcription factor MYB1R1-like	Nesi et al., 2000; Nesi et al., 2001
	C_2_3.850	Cytochrome P450 94A2-like	Tanaka and Brugliera, 2013
	C3.671	Receptor-like protein kinase	

[1]*In these references, the corresponding genes had been proved or suggested to be involved in pigments biosynthesis or accumulation during seed formation.*

coat color is stable under various environments, and is mainly influenced by the seed ripening level. With the aid of a molecular genetic map, Zhang H. et al. (2013) detected four QTLs for the seed coat color on three LGs using an F_2 population derived from a cross between a white-seeded parent and a black-seeded parent. Wei et al. (2015) identified seven loci on three LGs for the trait by GWAS using a natural population. Wang L. et al. (2016) identified four QTLs on three LGs for the trait using a population consisting of 430 RILs. Lately, Du et al. (2019) found seven seed coat color related QTLs on three LGs. The results suggest that the seed coat color trait is controlled by a few major QTLs in sesame, and our phenotype analysis also agrees with this. In the present study, 17 QTLs in seven non-overlapping intervals on four LGs were determined for sesame coat color. Comparison of the results indicated that our QTLs covered all the previous published QTLs, except two loci, i.e., *qSCL-8.1* identified by Wang L. et al. (2016) and *qsccZ12* identified by Du et al. (2019). High consistency reflects the reliability of the QTL mapping results in this study. Furthermore, we found that the specific QTL *qSCL-8.1* reported by Wang L. et al. (2016) was close to the *qSC* cluster in this study. We proposed that *qSCL-8.1* should be another member of the cluster. The QTL *qsccZ12* was identified based on the color space index Z, which is different from the indices used in this study, and had the lowest phenotypic variation (5.58%) among the seven QTLs reported by Du et al. (2019). This may be the reason why we missed it in this study. Comparing with these previous studies, we (1) identified a new QTL for the seed coat color, i.e., *qSCa6.1*, and (2) successfully distinguished more QTLs than the previous studies.

Candidate Genes for the Seed Coat Color in Sesame

In this study, in order to clarify the expression profiles of the candidate genes above, we also analyzed and compared the seed transcriptome during the seed development in the white- and black-seeded varieties. Compared to the stage 5 DAF, the DEGs at the following developing stages seem more important for the sesame seed coat color formation. Previous investigation indicated that pigment accumulation was usually observed approximately from 8 DAF in the black variety JS012 (data not shown). In this study, we determined that the pigment formation related pathways were not enriched at the time point 5 DAF. We thus used the DEGs at the stage 5 DAF to filter the other stages for candidate gene identification. Similarly, Wang et al. (2020) did a detailed research using RNA-seq data from a black- and white-sesame, and proposed that the genes controlling the sesame seed coat color were mainly expressed after 5 DAF, and thus took the 5 DAF as the initiated control and grouped 11–20 DAF to identify the candidate genes linked with pigment biosynthesis.

By integrating the QTL mapping and transcriptome analysis, we determined 28 genes as the candidate genes for the seed coat color in sesame (**Table 4**). Of the 28 candidate genes, the gene *SiPPO* (*C_2_7.207*) has been proved to determine the seed coat color in sesame (Wei et al., 2015), while the genes *C_2_3.819* and *C34.285* were annotated as transcription factor MYB-like and bHLH-like, respectively (**Table 4**). MYB and bHLH domain proteins act as a key regulator in PA accumulation in a developing seed (Nesi et al., 2000, 2001). Meanwhile, the gene *C34.291* was annotated to encode a pyridoxine/pyridoxamine 5′-phosphate oxidase (*PPOX*) gene, which was suggested to be related to pigment biosynthesis in both plant (Sang et al., 2011) and human (Frank et al., 1999). The gene *C_2_7.220* encodes a dioxygenase gene that was proven to be involved in plant pigments biosynthesis (Polturak and Aharoni, 2018). The gene *C_2_3.683* encodes a xyloglucan endotransglucosylase/hydrolase 5 (*XTH5*) genes, and *XTH5* has been reported to be involved in fruit pigment accumulation in the ripening processes in

tomato (Wang et al., 2019). In addition, we found that the genes *C_2_3.773* and *C_2_3.787* were annotated as E3 ubiquitin-protein ligases, which were suggested to play important roles in tomato pigment accumulation (Tang et al., 2015). The genes *C_2_3.774* and *C_2_3.850* encode triose phosphate/phosphate translocator and cytochrome P450 gene, respectively, and were also suggested to be related to pigments biosynthesis in many plants (Tanaka and Brugliera, 2013; Hilgers et al., 2018). To sum up, 10 out of the 28 candidate genes were proved or suggested to be involved in pigments biosynthesis or accumulation. The results exhibited the high confidence of the candidate gene detection. However, we should remember that some SNP and InDels existed in the upstream and downstream sequences of the genes that have been omitted in the study. In the future, the function of the candidate genes and other genes in the target intervals should be further validated to clarify the regulation for the seed coat color trait in sesame.

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/, PRJNA315474 and PRJNA742439.

AUTHOR CONTRIBUTIONS

CL performed the data analysis and drafted the manuscript. YD conducted the main experiments and participated in the data analysis. HM guided the experiments and drafted the manuscript. MJ conducted the cross-section. LW performed the genetic experiments. HZ conceived and designed the experiments, and guided manuscript correction for publishing. All authors have seen and approved the final version of the manuscript.

FUNDING

This work was supported by the Key Laboratory of Specific Oilseed Crops Genomics of Henan Province. This work was financially supported by the earmarked fund for China Agriculture Research System (CARS-14); the Key Project of Science and Technology of Henan province (201300110600); Zhongyuan Science and Technology Innovation Leading Talent Plan (214200510020), the Henan Province Specific Professor Position Program (SPPP2020), the Zhongyuan Scientist Workshop Construction (ZSWC2019 and 214400510026), and the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (ISTTCPHP2016).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.700469/full#supplementary-material

Supplementary Figure 1 | Phenotypes of seed coat color of the two parents, F_1 and some F_2 progenies. **(A)** Yuzhi DS899; **(B)** JS012; **(C)** F_1 seeds of the cross between Yuzhi DS899 and JS012; **(D)** seeds of some F_2 progenies.

Supplementary Figure 2 | The heatmap of the samples used for DEG screening between white- and black-seeded varieties.

Supplementary Figure 3 | GO annotation of the DEGs identified between JS012 and Yuzhi 11 at the stages 5, 10, 15, 20, and 25 DAF, respectively.

Supplementary Figure 4 | The expression pattern of the 28 candidate genes. The X-axis indicates the time points at 5, 10, 15, 20, and 25 DAF, respectively; the Y-axis indicates the expression level of the genes; *Indicates significantly differentially expressed; the orange line stands for the gene expression level in JS012, while the blue line stands for in Yuzhi 11. **(A)** the genes expressed predominantly in JS012; **(B)** the genes expressed predominantly in Yuzhi 11; **(C)** the genes with hybridized expression pattern.

Supplementary Table 1 | Statistics of the *L*, *a*, and *b* values of the seed coat color trait in the F_2 population derived from the cross between Yuzhi DS899 and JS012.

Supplementary Table 2 | The variant counts of the two parents and their F_2 population harvested from joint variant calling.

Supplementary Table 3 | Distribution of bin markers in the superdense genetic map.

Supplementary Table 4 | QTL mapping of seed coat color traits in sesame using the GCIM, ICIM, IM, and CIM methods.

Supplementary Table 5 | The genes located in the *qSC* intervals.

Supplementary Table 6 | The SNPs or InDels in the six contig segments between Yuzhi DS899 and JS012.

Supplementary Table 7 | Statistics of the RNA-seq datasets used in the study.

Supplementary Table 8 | The DEGs identified between JS012 and Yuzhi 11 at the stages 5, 10, 15, 20, and 25 DAF, respectively.

Supplementary Table 9 | The enriched KEGG pathways in the DEGs between JS012 and Yuzhi 11 at the stages 5, 10, 15, 20, and 25 DAF, respectively.

Supplementary Table 10 | The DEGs in the *qSC* intervals.

REFERENCES

Appelhagen, I., Lu, G. H., Huep, G., Schmelzer, E., Weisshaar, B., and Sagasser, M. (2011). TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant J.* 67, 406–419. doi: 10.1111/j.1365-313X.2011.04603.x

Appelhagen, I., Nordholt, N., Seidel, T., Spelt, K., and Koes, R. (2015). TRANSPARENT TESTA 13 is a tonoplast P3A -ATPase required for vacuolar deposition of proanthocyanidins in *Arabidopsis thaliana* seeds. *Plant J.* 82, 840–849. doi: 10.1111/tpj.12854

Asekova, S., Oh, E., Kulkarni, K. P., Siddique, M. I., Lee, M. H., Kim, J. I., et al. (2021). An integrated approach of QTL mapping and genome-wide association analysis identifies candidate genes for phytophthora blight resistance in sesame (*Sesamum indicum* L.). *Front. Plant Sci.* 12:604709. doi: 10.3389/fpls.2021.604709

Badone, F. C., Cassani, E., Landoni, M., Doria, E., Panzeri, D., Lago, C., et al. (2010). The low phyticacid1-241 (lpa1-241) maize mutation alters the accumulation

of anthocyanin pigment in the kernel. *Planta* 231, 1189–1199. doi: 10.1007/s00425-010-1123-z

Bedigian, D. (2003). Evolution of sesame revisited: domestication, diversity and prospects. *Genet. Resour. Crop Evol.* 50, 779–787. doi: 10.1023/A:1025029903549

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

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118, iso-2, iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dong, X., Braun, E. L., and Grotewold, E. (2001). Functional conservation of plant secondary metabolic enzymes revealed by complementation of *Arabidopsis* flavonoid mutants with maize genes. *Plant Physiol.* 127, 46–57. doi: 10.1104/pp.127.1.46

Du, H., Zhang, H., Wei, L., Li, C., Duan, Y., and Wuang, H. (2019). A high-density genetic map constructed using specific length amplified fragment (SLAF) sequencing and QTL mapping of seed-related traits in sesame (*Sesamum indicum* L.). *BMC Plant Biol.* 19:588. doi: 10.1186/s12870-019-2172-5

Eloy, N. B., Voorend, W., Lan, W., Saleme, M. L., Cesarino, I., Vanholme, R., et al. (2017). Silencing CHALCONE SYNTHASE in maize impedes the incorporation of tricin into lignin and increases lignin content. *Plant Physiol.* 173, 998–1016. doi: 10.1104/pp.16.01108

Frank, J., McGrath, J. A., Poh-Fitzpatrick, M. B., Hawk, J. L., and Christiano, A. M. (1999). Mutations in the translation initiation codon of the protoporphyrinogen oxidase gene underlie variegate porphyria. *Clin. Exp. Dermatol.* 24, 296–301. doi: 10.1046/j.1365-2230.1999.00484.x

Ganesan, K., and Xu, B. (2017). A critical review on polyphenols and health benefits of black soybeans. *Nutrients* 9:455. doi: 10.3390/nu9050455

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346

Gonzalez, A., Mendenhall, J., Huo, Y., and Lloyd, A. (2009). *TTG1* complex MYBs, *MYB5* and *TT2*, control outer seed coat differentiation. *Dev. Biol.* 325, 412–421. doi: 10.1016/j.ydbio.2008.10.005

Ha, T. J., Lee, M. H., Seo, W. D., Baek, I. Y., Kang, J. E., and Lee, J. H. (2017). Changes occurring in nutritional components (phytochemicals and free amino acid) of raw and sprouted seeds of white and black sesame (*Sesamum indicum* L.) and screening of their antioxidant activities. *Food Sci. Biotechnol.* 26, 71–78. doi: 10.1007/s10068-017-0010-9

Han, Y., Vimolmangkang, S., Soria-Guerra, R. E., Rosales-Mendoza, S., Zheng, D., Lygin, A. V., et al. (2010). Ectopic expression of apple *F3′H* genes contributes to anthocyanin accumulation in the *Arabidopsis tt7* mutant grown under nitrogen stress. *Plant Physiol.* 153, 806–820. doi: 10.1104/pp.109.152801

Hilgers, E. J. A., Schöttler, M. A., Mettler-Altmann, T., Krueger, S., Dörmann, P., Eicks, M., et al. (2018). The combined loss of triose phosphate and xylulose 5-phosphate/phosphate translocators leads to severe growth retardation and impaired photosynthesis in *Arabidopsis thaliana* tpt/xpt double mutants. *Front. Plant Sci.* 9:1331. doi: 10.3389/fpls.2018.01331

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Mering, C. V., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Keskin, H. (2019). "Active Compounds, Health Effects, and Extraction of Unconventional Plant Seed Oils," in *Plant and Human Health*, eds M. Ozturk and K. Hakeem (Cham: Springer), 245–285.

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* 61:1361779. doi: 10.1080/16546628.2017.1361779

Kumar, N., Mudgal, J., Parihar, V. K., Nayak, P. G., Kutty, N. G., and Rao, C. M. (2013). Sesamol treatment reduces plasma cholesterol and triacylglycerol levels in mouse models of acute and chronic hyperlipidemia. *Lipids* 48, 633–638. doi: 10.1007/s11745-013-3778-2

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Liang, J., Sun, J., Ye, Y., Yan, X., Yan, T., Rao, Y., et al. (2021). QTL mapping of PEG-induced drought tolerance at the early seedling stage in sesame using whole genome re-sequencing. *PLoS One* 16:e0247681. doi: 10.1371/journal.pone.0247681

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

Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* 26:14. doi: 10.1371/journal.pcbi.1005944

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., and Barron, D. (2007). The *Arabidopsis* MATE transporter *TT12* acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* 19, 2023–2038. doi: 10.1105/tpc.106.046029

Mei, H., Liu, Y., Du, Z., Wu, K., Cui, C., Jiang, X., et al. (2017). High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. *Front. Plant Sci.* 8:636. doi: 10.3389/fpls.2017.00636

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Miao, C., Fang, J., Li, D., Liang, P., and Zhang, X. (2018). Genotype-Corrector: improved genotype calls for genetic mapping in F_2 and RIL populations. *Sci. Rep.* 8:10088. doi: 10.1038/s41598-018-28294-0

Namiki, M. (2007). Nutraceutical functions of sesame: a review. *Crit. Rev. Food Sci. Nutr.* 47, 651–673. doi: 10.1080/10408390600919114

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* 12, 1863–1878. doi: 10.1105/tpc.12.10.1863

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114. doi: 10.1105/tpc.010098

Owens, D. K., Crosby, K. C., Runac, J., Howard, B. A., and Winkel, B. S. (2008). Biochemical and genetic characterization of *Arabidopsis* flavanone 3beta-hydroxylase. *Plant Physiol. Biochem.* 46, 833–843. doi: 10.1016/j.plaphy.2008.06.004

Polturak, G., and Aharoni, A. (2018). "La Vie en Rose": biosynthesis, sources, and applications of betalain pigments. *Mol. Plant* 11, 7–22. doi: 10.1016/j.molp.2017.10.008

Pourcel, L., Routaboul, J. M., Kerhoas, L., Caboche, M., and Lepiniec, L. (2005). TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. *Plant Cell* 17, 2966–2980. doi: 10.1105/tpc.105.035154

Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. doi: 10.1093/bioinformatics/btx494

Sang, Y., Locy, R. D., Goertzen, L. R., Rahotte, A. M., and Si, Y. (2011). Expression, in vivo localization and phylogenetic analysis of a pyridoxine 5′-phosphate oxidase in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 49, 88–95. doi: 10.1016/j.plaphy.2010.10.003

Tanaka, Y., and Brugliera, F. (2013). Flower colour and cytochromes P450. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20120432. doi: 10.1098/rstb.2012.0432

Tang, X., Miao, M., Niu, X., Zhang, D., Cao, X., Jin, X., et al. (2015). Ubiquitin-conjugated degradation of golden 2-like transcription factor is mediated by CUL4-DDB1-based E3 ligase complex in tomato. *New Phytol.* 209, 1028–1039. doi: 10.1111/nph.13635

Tohge, T., de Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi: 10.1093/jxb/erx177

Uncu, A., Fray, A., Karlovsky, P., and Doganlar, S. (2016). High-throughput single nucleotide polymorphism (SNP) identification and mapping in the sesame

(*Sesamum indicum* L.) genome with genotyping by sequencing (GBS) analysis. *Mol. Breeding* 36:173. doi: 10.1007/s11032-016-0604-6

Van der Auwera, G. A., Carneiro, M. O., Hart, C., Poplin, R., and Del Angel, G. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

Wang, A., Chen, D., Ma, Q., Rose, J. K., Fei, Z., Liu, Y., et al. (2019). The tomato HIGH PIGMENT1/DAMAGED DNA BINDING PROTEIN 1 gene contributes to regulation of fruit ripening. *Hortic Res.* 6:15. doi: 10.1038/s41438-018-0093-3

Wang, D., Zhang, L., Huang, X., Wang, X., Yang, R., Mao, J., et al. (2018). Identification of nutritional components in black sesame determined by widely targeted metabolomics and traditional Chinese medicines. *Molecules* 23:1180. doi: 10.3390/molecules23051180

Wang, H., Ma, S., Xue, L., Li, Y., Wang, J., He, X., et al. (2016). miR-488 determines coat pigmentation by down-regulating the pigment-producing gene pro-opiomelanocortin. *Cell Mol. Biol.* 62, 37–43. doi: 10.14715/cmb/2016.62.12.7

Wang, L., Dossou, S., Wei, X., Zhang, Y., Li, D., Yu, J., et al. (2020). Transcriptome dynamics during black and white sesame (*Sesamum indicum* L.) seed development and identification of candidate genes associated with black pigmentation. *Gene*. 11:1399. doi: 10.3390/genes11121399

Wang, L., Xia, Q., Zhang, Y., Zhu, X., and Zhu, X. (2016). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genomics* 17:31. doi: 10.1186/s12864-015-2316-4

Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL Cartographer 2.5*. Raleigh: North Carolina State University.

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6:8609. doi: 10.1038/ncomms9609

Wen, Y. J., Zhang, Y. W., Zhang, J., Feng, J. Y., Dunwell, J. M., and Zang, Y. M. (2019). An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F_2. *Brief. Bioinform.* 20, 1913–1924. doi: 10.1093/bib/bby058

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485

Wu, K., Liu, H., Yang, M., Tao, Y., and Ma, H. (2014). High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (*Sesamum indicum* L.) based on RAD-Seq technology. *BMC Plant Biol.* 14:274. doi: 10.1186/s12870-014-0274-7

Xu, W., Grain, D., Bobet, S., Le, G. J., and Thévenin, J. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620

Yang, L., Gordon, K. S., and Wei, S. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 46, W71–W75. doi: 10.1093/nar/gky400

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 6, 284–287. doi: 10.1089/omi.2011.0118

Zhang, H., Miao, H., and Ju, M. (2019). "Potential for adaptation to climate change through genomic breeding in sesame," in *Genomic Designing of Climate-Smart Oilseed Crops*, ed. C. Kole (Cham: Springer), 371–440.

Zhang, H., Miao, H., Li, C., Wei, L., Duan, Y., Ma, Q., et al. (2016). Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. *Sci. Rep.* 6:31556. doi: 10.1038/srep31556

Zhang, H., Miao, H., Li, C., Wei, L., and Ma, Q. (2012). Analysis of sesame karyotype and resemblance-near coefficient. *Chin. Plant Bull.* 47, 602–614. doi: 10.3724/SP.J.1259.2012.00602

Zhang, H., Miao, H., Wei, L., Li, C., Zhao, R., and Wang, C. (2013). Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). *PLoS One* 8:e63898. doi: 10.1371/journal.pone.0063898

Zhang, Y., Wang, L., Gao, Y., Li, D., and Yu, J. (2018). Genetic dissection and fine mapping of a novel *dt* gene associated with determinate growth habit in sesame. *BMC Genet.* 19:38. doi: 10.1186/s12863-018-0614-y

Zhang, Y., Wang, L., Xin, H., Li, D., and Ma, C. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13:141. doi: 10.1186/1471-2229-13-141

Zhang, Y. W., Wen, Y. J., Dunwell, J. M., and Zhang, Y. M. (2019). QTL.gCIMapping.GUI v2.0: an R software for detecting small-effect and linked QTLs for quantitative traits in bi-parental segregation populations. *Comput. Struct. Biotechnol. J.* 18, 59–65. doi: 10.1016/j.csbj.2019.11.005

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.

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.

Copyright © 2021 Li, Duan, Miao, Ju, Wei and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 August 2021
doi: 10.3389/fgene.2021.664814

Dissection of Allelic Variation Underlying Floral and Fruit Traits in Flare Tree Peony (*Paeonia rockii*) Using Association Mapping

Xin Guo[1,2,3,4,5], *Chunyan He*[1,2,3,4,5], *Fangyun Cheng*[1,2,3,4,5]*, *Yuan Zhong*[1,2,3,4,5], *Xinyun Cheng*[6] *and Xiwen Tao*[6]

[1]*Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China,* [2]*Peony International Institute, School of Landscape Architecture, Beijing Forestry University, Beijing, China,* [3]*Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, Beijing Forestry University, Beijing, China,* [4]*Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, Beijing Forestry University, Beijing, China,* [5]*National Engineering Research Center for Floriculture, Beijing Forestry University, Beijing, China,* [6]*Beijing Guose Peony Technology Co. Ltd., Beijing, China*

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Guoan Shi,
Henan University of Science and Technology, China
Rakesh Chahota,
Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, India

***Correspondence:**
Fangyun Cheng
chengfy8@263.net
orcid.org/0000-0002-3928-5731

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

Received: *06 February 2021*
Accepted: *29 June 2021*
Published: *11 August 2021*

Citation:
Guo X, He C, Cheng F, Zhong Y, Cheng X and Tao X (2021) Dissection of Allelic Variation Underlying Floral and Fruit Traits in Flare Tree Peony (Paeonia rockii) Using Association Mapping. Front. Genet. 12:664814. doi: 10.3389/fgene.2021.664814

Allelic variation in floral quantitative traits, including the elements of flowers and fruits, is caused by extremely complex regulatory processes. In the genetic improvement of flare tree peony (*Paeonia rockii*), a unique ornamental and edible oil woody species in the genus *Paeonia*, a better understanding of the genetic composition of these complex traits related to flowers and fruits is needed. Therefore, we investigated the genetic diversity and population structure of 160 *P. rockii* accessions and conducted single-marker association analysis for 19 quantitative flower and fruit traits using 81 EST-SSR markers. The results showed that the population had a high phenotypic diversity (coefficients of variation, 11.87–110.64%) and a high level of genetic diversity (mean number of alleles, $N_A = 6.09$). These accessions were divided into three subgroups by STRUCTURE analysis and a neighbor-joining tree. Furthermore, we also found a low level of linkage disequilibrium between these EST-SSRs and, by single-marker association analysis, identified 134 significant associations, including four flower traits with 11 EST-SSRs and 10 fruit traits with 32 EST-SSRs. Finally, based on the sequence alignment of the associated markers, P280, PS2, PS12, PS27, PS118, PS131, and PS145 may be considered potential loci to increase the yield of flare tree peony. These results laid the foundation for further analysis of the genetic structure of some key traits in *P. rockii* and had an obvious potential application value in marker-assisted selection breeding.

Keywords: flare tree peony (*P. rockii*), SSR markers, floral and fruit traits, association mapping, linkage disequilibrium

INTRODUCTION

Flare tree peony (*Paeonia rockii*; FTP) is one of the most representative species in the *Paeonia* section *Moutan* that is native to China (Hong and Pan, 1999). Approximately 300 cultivars have been derived from this species, mainly distributed in northwest China, consisting of a large cultivar group that is distinct from the common tree peony (*Paeonia suffruticosa*) in China (Cheng et al., 2005; Cheng and Yu, 2008; Yuan et al., 2012). In addition to being cultivated as an ornamental species since the Tang Dynasty (618–906 AD), FTP has rapidly developed into a new woody edible oil plant in recent years as its seeds contain a high unsaturated fatty acid content (Li et al., 2015a,b; Zhang et al., 2017a,b; Wang et al., 2020). Therefore, improving ornamental quality and increasing the potential yield have become important objectives for FTP breeding. Among them, ornamental quality mainly includes floral traits, while yield is a complex character affected by multiple factors, in which the most important contributing factor is the fruit trait (Liu and Cheng, 2020b). Quantitative flower and fruit traits are controlled by multiple genes and have moderate to high heritability. As conventional breeding methods take more than 10 years to develop new FTP cultivars with stable comprehensive characteristics (Cheng, 2012), molecular breeding approaches have become inevitable selection methods to breed new tree peony cultivars (Gao et al., 2013; Zhou et al., 2015).

The development of DNA markers has provided broad prospects for accelerating the selection of intricate quantitative traits in trees, especially in tree peonies (Du et al., 2013; Yu et al., 2013; Wang et al., 2014; Resende et al., 2017). Quantitative trait locus (QTL) mapping studies have been carried out on parental populations and attempts have been made to use QTLs for breeding (Cai et al., 2015; Guo et al., 2017; Zhang et al., 2019). When used in selective breeding, substantial deficiencies in QTL mapping have been widely discussed (Bernardo, 2008; Grattapaglia and Resende, 2011). Due to the method's low mapping ability, only a few QTLs hidden under target traits were found, and the variances explained by these QTLs were overestimated (Beavis, 1998). Association mapping based on linkage disequilibrium (LD), proposed as a way to deconvolute QTL mapping, can identify natural alleles for specific phenotypes, providing a valuable opportunity to shorten breeding years and improve breeding efficiency (Neale and Savolainen, 2004; Thumma et al., 2005). In molecular marker-assisted selection (MAS) breeding, expressed sequence tag (EST)-simple repeat sequence (SSR) markers are ideal choices because they are highly variable, codominant, and highly informative (Varshney et al., 2005). In FTP, a few associations between polymorphisms in EST-SSR and certain traits have been reported (Wu et al., 2017; Cui et al., 2018; Liu and Cheng, 2020a). Wu et al. (2017) employed single-marker association analysis on a sample of 462 individuals to identify 46 significant associations, including 37 EST-SSRs involving 11 flower traits, which explained 2.68–23.97% (mean 5.50%) of the phenotypic variation. Cui et al. (2018) found that 15 EST-SSRs were significantly associated with five oil-related traits based on association analysis of 205 individuals with phenotypic traits over three consecutive years. In an association analysis of 420 individuals, Liu et al. identified 141 significant associations involving 17 yield-related traits and 41 EST-SSRs, and the phenotypic variation was relatively small (mean 11.34%; Liu and Cheng, 2020a). To reduce environmental disturbances and measurement errors, asexual reproduction is often necessary to increase the accuracy of phenotypic measurements (Du et al., 2013). Although there have been previous studies on the association analysis of important traits in FTP, the samples used were all individuals without asexual reproduction; the core germplasm resources of FTP were not used.

In this study, 160 accessions (each containing three clones) from core germplasm resources of FTP and 81 EST-SSRs were used for single-marker association analysis to explore the allelic effects of floral and fruit trait variation. The results of this study laid a foundation for further identifying the key trait linkage loci of FTP, which is of great value to the genetic improvement of related traits.

MATERIALS AND METHODS

Plant Materials

A total of 160 accessions (each containing three clones), representing the core germplasm resources of FTP that had been demonstrated and described by Guo et al. (2020), were used in this study. All the samples were originally introduced from Gansu Province in Northwest China, which was the main cultivation area, and had been randomly cultivated in the same nursery field of Beijing Guose Peony Garden in Beijing, China (40°45'N, 115°97'E) for more than 10 years. All the measured plants, approximately 15 years old, can produce stable flowers and fruits annually. The complete list of accessions used in this study is provided in **Supplementary Table S1**.

Phenotyping

Seven flower quantitative traits, namely, flower diameter (FD), petal length (PL), petal width (PW), flare length (FL), flare width (FW), petal number (PN), and carpel number (CN), in the 160 accessions were measured at full bloom using digital calipers (YB5001B, Kraftwelle Industrial Co. Ltd., China) in May 2019. In addition, a total of 12 fruit quantitative traits, namely, the number of carpels with seeds (NCWS), multiple fruit fresh weight (MFFW), single fruit length (SFL), single fruit width (SFW), single fruit height (SFH), single fruit pericarp thickness (SFPT), multiple fruit seed number (MFSN), multiple fruit seed fresh weight (MFSFW), individual fruit number (IFN), individual seed number (ISN), individual fruit fresh weight (IFFW), and individual seed fresh weight (ISFW), were measured using digital calipers and electronic balances in September 2019. All measurements were carried out, as described in **Supplementary Table S2**. A total of nine flowers and nine fruits of each accession were collected, with three flowers and three fruits from one plant per replicate. The average value for each trait from three replicates for each accession was used for statistical and association analysis.

DNA Extraction and SSR Marker Genotyping

The total genomic DNA of 160 accessions was extracted from fresh young leaves using a DNAsecure plant kit (Tiangen Biotech, Beijing, China), following the manufacturer's instructions. The quality of the extracted DNA was determined by electrophoresis on 2% agarose gels and visualization using a UnicoUV-visible Spectrophotometer (Agilent, Palo Alto, CA, United States). Then, the DNA were diluted with deionized water to 20–30 ng/μl and stored in a freezer at −20°C.

The polymorphism of 140 previously developed SSR markers (Homolka et al., 2010; Hou et al., 2011; Gai et al., 2012; Zhang et al., 2012; Yu et al., 2013; Wu et al., 2014; Liu and Cheng, 2020a) was evaluated using a random sample of 12 accessions. After screening, a total of 81 EST-SSRs (**Supplementary Table S3**) were used to reveal the polymorphism of these 160 accessions. The SSR-PCR amplification reaction was conducted in a 10-μl solution, including 5 μl of 2×Power Taq PCR Master MIX (Aidlab Biotechnologies, Beijing, China), 3 μl of ddH_2O, 1 μl of 20–25 ng/μl genomic DNA, and 0.5 μl of 10 μmol/L each of forward and reverse primers, and the procedure was performed as described by Wu et al. (2014). The PCR products were differentiated by capillary electrophoresis using an ABI3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA, United States). Micro-Checker 2.2.3 (Van Oosterhout et al., 2004) was applied to identify and correct genotyping errors.

Data Analysis

The statistics software SPSS 18.0 (IBM Inc., Chicago, IL, United States; Davis, 2008) was used to analyze the average values of the investigated traits per accession. The coefficient of variation (CV) of each trait was calculated as follows: (standard deviation/mean) × 100%. The variation in 19 traits was estimated by one-way analysis of variance (ANOVA), and Pearson's correlations between traits were calculated. Prior to ANOVA and Pearson's correlation analysis, all SM data were tested for normality with the Shapiro-Wilk W test and for homogeneity of variance with Levene's test, and the nonnormal data were logarithmically transformed. In addition, Benjamini-Hochberg (BH) FDR correction was used to correct the values of p for Pearson's correlation analysis.

The capillary electrophoresis data were analyzed using GeneMarker 2.2.0. GenAIEx 6.5 (Peakall and Smouse, 2012) was used to calculate the following statistics: number of different alleles (N_A), number of effective alleles (N_E), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficients (F_{IS}), and Nei's genetic diversity (GD). The polymorphism information content (PIC) of each locus was calculated using the Microsatellite Toolkit. GENEPOP 4.2 (Rousset, 2008) was used to detect microsatellite loci deviating from the Hardy-Weinberg equilibrium (HWE), and the results were applied to multiple tests with Bonferroni correction. Additionally, MEGA-X was used to construct a neighbor-joining (NJ) phylogenetic tree based on Nei's unbiased genetic distance (Kumar et al., 2018).

The number of subpopulations (K) was detected by STRUCTURE 2.3.4 through an admixture model (Pritchard et al., 2000). First, the K value ranges from 1 to 10, and 10 independent operations were carried out for each K value, with a burn-in period of 100,000 times and 200,000 replications. Then, the results were uploaded to Structure Harvester (Earl and VonHoldt, 2012) to determine the final optimal K value. The optimum K value was inferred from $\mathrm{Ln}P(D)$ and ΔK (Evanno et al., 2005). For 10 repetitions of each K value, CLUMPP 1.1 was used to analyze the results from replicate analyses (Jakobsson and Rosenberg, 2007). Then, the outputs of each K value were visualized using CLUMPP and DISTRUCT (Rosenberg, 2004). The matrix corresponding to the K value of the optimal population structure was used for association analysis to correct false positives.

The degree of LD between loci was evaluated by the square of the allelic frequency (r^2), which was calculated by using TASSEL 2.0.1. $r^2 = 0.1$ was taken as the critical value to determine whether two loci had LD ($r^2 > 0.1$: LD). Then, a mixed linear model (MLM) of TASSEL 2.0.1 was used to incorporate SSR data, phenotypic traits, Q matrix, and kinship matrix for association analysis. Q was the matrix of the optimal K value obtained through Structure Harvester. The kinship matrix was calculated by SPAGeDi 1.2 (Hardy and Vekemans, 2002). An adjusted value of p was employed for association analysis using false discovery rate (FDR) correction under QVALUE in R (Storey and Tibshirani, 2003). The ratio of dominance (d) to additive (a) effects was used to assess the gene effects of the significant loci obtained by association analysis. The boundaries of additive effect, partial to full dominance and overdominance, were $|d/a| \leq 0.50$, $0.50 < |d/a| < 1.25$, and $|d/a| > 1.25$, respectively. The calculation formulas of additive (a) effects and dominance (d) were as follows: $2a = |G_{BB} - G_{bb}|$; $d = G_{Bb} - 0.5\ (G_{BB} + G_{bb}$; G represents the average of phenotypes corresponding to genotype, BB and bb: homozygous genotypes, and Bb: heterozygous genotypes; Eckert et al., 2009).

Open Reading Frame Finder (ORF Finder) was used to find the complete ORFs of the associated marker sequences and translate the ORF sequences into protein sequences (Rombel et al., 2002). Then, the results were compared in the Arabidopsis Information Resource (TAIR) to find protein sequences with higher homology (Berardini et al., 2015). Finally, DNAMAN was used to construct a phylogenetic tree and protein sequence alignment map.

RESULTS

Statistical and Correlation Analyses of Phenotypic Traits

As the tested samples can represent the current situation of FTP germplasm resources in China, ANOVA showed that the phenotypic variation range of all measured traits were wide (**Table 1**). The CVs ranged from 11.87% (FD) to 110.64% (ISFW), with an average of 48.01%. The CVs of fruit traits (mean 58.44%) were higher than those of flower traits (mean 30.13%), among them the traits ISFW (110.64%), PN (109.70%),

TABLE 1 | Descriptive statistics of 19 quantitative traits in 160 flare tree peony (FTP) accessions.

Trait	Minimum	Maximum	Mean ± SD	Standard deviation	Coefficient of variation (%)	*F*
FD	9.67	19.67	15.40 ± 0.14	1.83	11.87	3.340**
PL	5.43	10.07	7.54 ± 0.07	0.90	11.93	0.810**
PW	4.68	9.93	7.42 ± 0.09	1.11	15.00	1.237**
FL	1.43	4.30	2.69 ± 0.05	0.57	21.26	0.327**
FW	0.67	3.72	1.98 ± 0.04	0.53	27.03	0.286**
PN	7.00	257.00	39.00 ± 3.38	42.78	109.70	1,829.992**
CN	4.00	10.33	5.33 ± 0.06	0.75	14.13	0.567**
NCWS	3.56	11.33	5.52 ± 0.08	1.05	19.02	1.100**
MFFW	5.93	83.83	21.48 ± 0.80	10.15	47.23	102.935**
SFL	14.93	67.32	40.49 ± 0.47	5.98	14.77	35.761**
SFW	9.89	22.26	15.33 ± 0.19	2.44	15.91	5.951**
SFH	3.34	98.85	15.15 ± 0.56	7.07	46.69	50.017**
SFPT	1.20	17.57	2.59 ± 0.10	1.29	49.78	1.667**
MFSN	0.11	52.00	12.28 ± 0.77	9.79	79.67	95.761**
MFSFW	0.05	28.26	5.56 ± 0.37	4.74	85.29	22.486**
IFN	2.50	74.00	29.12 ± 1.24	15.69	53.87	246.051**
ISN	0.93	2,119.79	368.61 ± 29.95	378.83	102.77	143,513.737**
IFFW	43.18	2,294.00	509.98 ± 30.50	385.85	75.66	148,883.795**
ISFW	0.43	960.84	170.24 ± 14.89	188.35	110.64	35,474.893**

**$p < 0.01$.

and ISN (102.77%) were relatively higher. Correlation analysis of different traits showed 81 significant correlations ($p < 0.05$), of which 65 were very significant correlations ($p < 0.01$; **Table 2**). Considering the flower traits, excluding FL and PN, FW and CN, and PN and CN, all other traits were significantly correlated. Among the 66 correlation factors in fruit traits, 39 were significantly correlated. Especially when we considered ISFW as the yield index, MFFW, SFL, SFW, MFSN, MFSFW, IFN, ISN, and IFFW were very significantly positively correlated with ISFW, and the correlation coefficients were 0.726, 0.474, 0.599, 0.730, 0.820, 0.515, 0.960, and 0.814, respectively.

Genetic Diversity

In total, 81 polymorphic EST-SSRs were used to evaluate the diversity of 160 accessions (**Table 3**). Then, 493 alleles were identified, ranging from 2 to 20 (N_A). The N_E ranged from 1.006 to 8.516, with an average of 2.603. The I varied from 0.021 to 2.357 (mean 1.026). The mean values of H_O and H_E were 0.501 and 0.524, respectively. The mean F_{IS} of 81 SSRs was −0.439, among which 63 pairs were negative. In addition, the PIC values ranged from 0.006 to 0.871, with an average of 0.476. After Bonferroni multiple comparisons, 43 SSR sites deviated significantly from HWE. Therefore, these deviated sites were removed in the subsequent analysis and 38 SSR sites were ultimately used for subsequent population structure analysis and association mapping.

Population Structure

Through STRUCTURE analysis of 160 accessions with 38 EST-SSRs, as the K value increased, Ln$P(D)$ progressively increased overall (**Figure 1**). When K reached 3, the rising trend of Ln$P(D)$ slowed down. Moreover, the ΔK value corresponding to the maximum K value was the population structure; therefore, the 160 accessions could be divided into three subgroups. The outputs of $K = 3$ were visualized through CLUMPP and DISTRUCT, and all samples showed a wide range of mixed lineages. Then, we visualized the outputs from $K = 2$–5 and found that each sample was highly heterozygous. The Q matrix output of the three subgroups can be used for structure-based association analysis. The phylogenetic tree divided these samples into three major clades (**Figure 2**), revealing similar clustering results to those obtained using STRUCTURE. In summary, the two methods classified these samples into three subgroups.

LD Level

The LD levels of 38 EST-SSRs in the 160 accessions were evaluated (**Figure 3**). Based on the r^2 estimates, only 0.71% ($r^2 \geq 0.1$) of the loci sites had significant LD. In all 698 locus pairs ($r^2 < 0.1$), 17.62, 30.23, and 50.14% of the locus pairs displayed linkage equilibrium at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. Therefore, the overall LD level of 38 EST-SSRs in the 160 accessions was low, and most loci were in linkage equilibrium with each other. However, there were also several loci with significant LD levels between them, such as marker P235-PS50 ($r^2 > 0.3$; $p < 0.001$).

Single-Marker Associations of Floral Traits

For flower traits, 266 (38 EST-SSRs × 7 traits) single-marker associations were detected by the MLM model, and a total of 21 (7.89%) significant associations between four floral traits and 14 EST-SSRs were detected under the condition of $p < 0.01$. Then, FDR test was carried out on these 21 significantly associated combinations ($Q < 0.05$), and 17 significantly associated combinations of four flower traits and 11 EST-SSRs were ultimately identified, with an explainable rate of 2.23–26.34% (mean 11.80%; **Table 4**). The statistical results of the gene effects indicated that 2 (11.76%) associated combinations

TABLE 2 | The correlation analysis of 19 quantitative traits in 160 FTP accessions.

Trait	FD	PL	PW	FL	FW	PN	CN	NCWS	MFFW	SFL	SFW	SFH	SFPT	MFSN	MFSFW	IFN	ISN	IFFW	ISFW
FD	1.000																		
PL	0.680**	1.000																	
PW	0.634**	0.733**	1.000																
FL	0.339**	0.546**	0.405**	1.000															
FW	0.281**	0.342**	0.500**	0.659**	1.000														
PN	−0.158*	−0.233**	−0.370**	−0.103	−0.167*	1.000													
CN	0.172*	0.181*	0.230**	0.175*	0.154	0.103	1.000												
NCWS	0.096	0.064	0.085	0.097	0.053	0.234**	0.654**	1.000											
MFFW	0.235**	0.177*	0.217**	−0.073	0.021	−0.235**	−0.012	−0.035	1.000										
SFL	0.271**	0.220**	0.227**	0.044	0.114	−0.098	−0.168*	−0.182*	0.789**	1.000									
SFW	0.257**	0.185*	0.262**	−0.014	0.138	−0.322**	−0.067	−0.106	0.867**	0.703**	1.000								
SFH	0.077	0.010	0.068	−0.030	0.058	−0.119	0.061	−0.056	0.231**	0.263**	0.209**	1.000							
SFPT	−0.005	0.049	0.130	0.040	0.180*	−0.032	0.013	0.124	0.270**	0.216**	0.298**	0.044	1.000						
MFSN	0.137	0.030	0.076	−0.084	−0.067	−0.299**	−0.048	−0.097	0.677**	0.368**	0.588**	0.107	−0.007	1.000					
MFSFW	0.150	0.065	0.102	−0.127	−0.087	−0.273**	−0.051	−0.088	0.841**	0.533**	0.694**	0.160*	0.061	0.938**	1.000				
IFN	0.042	−0.078	−0.097	−0.175*	−0.119	0.034	−0.114	−0.002	0.140	0.117	0.119	−0.032	−0.056	0.072	0.113	1.000			
ISN	0.095	−0.014	0.021	−0.158*	−0.101	−0.233**	−0.066	−0.058	0.605**	0.363**	0.529**	0.093	−0.021	0.775**	0.772**	0.523**	1.000		
IFFW	0.176*	0.096	0.097	−0.144	−0.075	−0.192*	−0.097	−0.036	0.726**	0.581**	0.634**	0.140	0.148	0.460**	0.591**	0.681**	0.739**	1.000	
ISFW	0.100	0.011	0.035	−0.191*	−0.122	−0.220**	−0.070	−0.059	0.726**	0.474**	0.599**	0.131	0.029	0.730**	0.820**	0.515**	0.960**	0.814**	1.000

**$p < 0.01$; *$p < 0.05$.

presented additive effects, 2 (11.76%) presented partial to full dominance, and 13 (76.47%) presented overdominance.

The number of significantly associated combinations for each trait ranged from two (FD) to eight (FL). There were two markers that were significantly related to FD and CN, and the highest interpretation rates were found for locus PS19, which were 26.34 and 25.38%, respectively. There were eight EST-SSRs that were significantly related to FL, with interpretation rates of 5.03–12.53% (mean 10.15%), and the highest explanatory rate was locus PS91. Among them, there were one and seven combinations with additive effects and overdominance, respectively. There were five EST-SSRs that were significantly related to FW, with explanatory rates ranging from 6.72 to 14.68% (mean 9.43%), and the highest interpretation rate was locus PS43. No significant associations with PL, PW, and PN were detected in this study. Additionally, five EST-SSRs were markedly associated with more than one trait. For instance, P26 was significantly related to FD, FL, and CN, while P235, P281, and PS24 were significantly associated with FL and FW.

Single-Marker Associations of Fruit Traits

For fruit traits, 456 (38 EST-SSRs × 12 traits) single-marker associations were detected by the MLM model, and a total of 118 (25.88%) significant associations between 10 fruit traits and 32 EST-SSRs were detected under the condition of $p < 0.01$. Then, an FDR test was carried out on these 118 significantly associated combinations ($Q < 0.05$), and 117 significantly associated combinations of 10 fruit traits and 32 EST-SSRs were ultimately identified, with explanatory rates of 0.18–43.40% (average 17.68%; **Table 5**). The statistical results of gene effects showed that 13 (11.11%) associated combinations appeared to have additive effects, 28 (23.93%) appeared to have partial to full dominance, and 76 (64.96%) appeared to have overdominance.

The number of significantly associated combinations for each trait ranged from 1 (IFN) to 25 (MFSN). Of these, IFN had one significant association; NCWS had two significant associations; SFW had three significant associations; ISN had eight significant associations; SFL had 12 significant associations; IFFW had 13 significant associations; MFFW and IFSW had 15 significant associations each; MFSFW had 23 significant associations; and MFSN had 25 significant associations. We also found that in ISN, ISFW, and MFSFW, the interpretation rates of P280 were the highest, while in MFFW and SFL, the interpretation rates of P318 were the highest. No locus significantly associated with SFPT and SFH was detected in this study. The phenomenon in which one marker was significantly associated with multiple traits was more common in this part of the study. In the correlation analysis of fruit traits, ISN, IFFW, and ISFW were all significantly correlated with each other. In the association analysis results, six markers were significantly associated with these three traits at the same time, namely, loci P150, P242, P280, PS12, PS59, and PS91. Similarly, loci P242, PS12, PS131, PS145, PS24, PS27, PS50, PS59, PS85, and PS91 were significantly associated with MFSN, MFFW, and MFSFW.

There were 10 marker-trait combinations with explanatory rates exceeding 30%, including the following: MFFW associated with PS145, PS85, P318, and PS12; SFL associated with P318;

TABLE 3 | Diversity information parameter at 81 SSRs in 160 FTP accessions.

Locus	N_A	N_E	I	H_o	H_E	F_{IS}	PIC	p
PS2	4	1.268	0.403	0.225	0.211	–0.853	0.193	0.986^{ns}
PS7	8	2.446	1.279	0.388	0.591	–0.064	0.558	0.000
PS8	6	2.884	1.172	0.738	0.653	–0.476	0.584	0.046^{ns}
PS10	6	2.681	1.217	0.163	0.627	0.368	0.571	0.000
PS12	7	2.924	1.367	0.688	0.658	–0.387	0.624	0.303^{ns}
PS17	4	2.118	0.829	0.513	0.528	–0.443	0.418	0.000
PS19	8	3.453	1.483	0.794	0.710	–0.407	0.673	0.090^{ns}
PS21	20	8.516	2.357	0.925	0.883	–0.165	0.871	0.000
PS24	3	1.013	0.042	0.013	0.012	–0.992	0.012	1.000^{ns}
PS25	10	3.234	1.480	0.406	0.691	0.103	0.648	0.000
PS27	6	1.422	0.589	0.275	0.297	–0.631	0.272	0.693^{ns}
PS30	5	2.262	0.930	1.000	0.558	–1.235	0.460	0.000
PS31	7	2.390	1.053	0.519	0.582	–0.310	0.493	0.000
PS33	7	3.620	1.446	0.650	0.724	–0.174	0.678	0.000
PS36	3	1.730	0.628	0.513	0.422	–0.793	0.335	0.049^{ns}
PS43	9	1.563	0.808	0.388	0.360	–0.716	0.342	1.000^{ns}
PS46	4	1.284	0.416	0.200	0.221	–0.684	0.200	0.000
PS47	4	2.100	0.903	0.219	0.524	0.106	0.447	0.000
PS49	11	5.214	1.927	0.425	0.808	0.282	0.788	0.000
PS50	6	1.660	0.745	0.413	0.397	–0.641	0.352	0.326^{ns}
PS53	4	2.661	1.050	1.000	0.624	–0.978	0.546	0.000
PS55	4	2.620	1.055	0.550	0.618	–0.271	0.550	0.000
PS56	6	3.402	1.321	1.000	0.706	–0.710	0.652	0.000
PS57	3	1.759	0.639	0.463	0.432	–0.640	0.341	0.741^{ns}
PS59	4	1.322	0.436	0.281	0.243	–0.912	0.217	0.638^{ns}
PS62	11	2.764	1.318	0.288	0.638	0.188	0.588	0.000
PS64	4	1.801	0.802	0.413	0.445	–0.483	0.392	0.226^{ns}
PS66	6	2.132	1.045	0.113	0.531	0.319	0.485	0.000
PS73	12	7.077	2.101	0.819	0.859	–0.095	0.843	0.000
PS75	12	5.269	1.913	0.600	0.810	0.070	0.786	0.000
PS85	6	2.438	1.107	0.650	0.590	–0.512	0.538	0.643^{ns}
PS90	9	4.010	1.729	0.563	0.751	0.001	0.726	0.000
PS91	4	2.247	1.043	0.544	0.555	–0.425	0.511	0.764^{ns}
PS93	4	2.606	1.040	1.000	0.616	–1.007	0.540	0.000
PS94	6	2.762	1.162	0.463	0.638	–0.087	0.568	0.235^{ns}
PS96	13	3.993	1.714	0.800	0.750	–0.318	0.723	0.000
PS97	5	1.181	0.371	0.163	0.153	–0.906	0.149	1.000^{ns}
PS98	4	2.076	0.782	0.494	0.518	–0.434	0.402	0.948^{ns}
PS102	5	3.381	1.306	1.000	0.704	–0.716	0.651	0.000
PS103	8	4.111	1.612	0.700	0.757	–0.168	0.721	0.000
PS105	2	1.098	0.189	0.094	0.089	–0.960	0.085	0.534^{ns}
PS113	4	2.050	0.760	1.000	0.512	–1.440	0.393	0.000
PS114	10	4.047	1.557	0.494	0.753	0.097	0.711	0.000
PS116	5	2.376	1.108	0.263	0.579	0.126	0.538	0.000
PS117	5	2.791	1.213	0.288	0.642	0.194	0.597	0.000
PS118	3	2.061	0.860	0.563	0.515	–0.578	0.444	0.327^{ns}
PS122	4	1.494	0.634	0.138	0.331	–0.085	0.305	0.000
PS123	3	1.918	0.687	0.781	0.478	–1.154	0.367	0.000
PS129	5	2.223	0.953	0.113	0.550	0.346	0.473	0.000
PS131	4	2.517	1.020	0.588	0.603	–0.372	0.527	0.077^{ns}
PS142	6	1.515	0.720	0.056	0.340	0.174	0.320	0.000
PS145	6	2.535	1.127	0.663	0.606	–0.488	0.547	0.892^{ns}
PS147	11	4.610	1.787	0.838	0.783	–0.286	0.755	0.000
PS151	12	6.593	2.047	0.456	0.848	0.310	0.831	0.000
PS159	12	2.639	1.527	0.350	0.621	0.058	0.602	0.000
PS160	9	3.615	1.607	0.656	0.723	–0.184	0.691	0.000
PS163	11	4.922	1.753	0.838	0.797	–0.254	0.767	0.000
P002	7	1.536	0.794	0.131	0.349	–0.027	0.335	0.000
P026	4	1.712	0.637	0.431	0.416	–0.621	0.334	0.986^{ns}
P061	4	2.024	0.731	0.525	0.506	–0.532	0.384	0.895^{ns}
P067	6	2.570	1.224	0.494	0.611	–0.197	0.573	0.000
P068	5	3.046	1.281	0.750	0.672	–0.445	0.620	0.056^{ns}
P138	3	2.023	0.726	0.650	0.506	–0.780	0.384	0.152^{ns}

(Continued)

TABLE 3 | Continued

Locus	N_A	N_E	I	H_o	H_E	F_{IS}	PIC	p
P150	3	2.540	0.999	0.594	0.606	−0.373	0.527	0.785^{ns}
P162	3	1.253	0.384	0.188	0.202	−0.728	0.185	0.672^{ns}
P180	3	1.246	0.384	0.206	0.197	−0.848	0.182	0.871^{ns}
P221	5	2.538	1.035	0.656	0.606	−0.477	0.524	0.000
P235	3	1.589	0.631	0.375	0.370	−0.642	0.320	0.080^{ns}
P242	2	1.013	0.038	0.013	0.012	−0.994	0.012	0.937^{ns}
P260	2	1.479	0.505	0.356	0.324	−0.777	0.271	0.204^{ns}
P265	3	1.733	0.650	0.481	0.423	−0.715	0.341	0.234^{ns}
P280	7	2.844	1.368	0.669	0.648	−0.383	0.614	0.422^{ns}
P281	2	1.006	0.021	0.006	0.006	−0.997	0.006	0.968^{ns}
P290	4	1.703	0.689	0.513	0.413	−0.829	0.348	0.000
P296	5	1.963	0.891	0.481	0.491	−0.491	0.431	0.964^{ns}
P318	8	3.820	1.448	0.844	0.738	−0.405	0.691	0.430^{ns}
P333	5	1.429	0.580	0.281	0.300	−0.637	0.274	0.400^{ns}
Seq6	7	1.683	0.777	0.450	0.406	−0.703	0.360	0.000
50F, R	12	3.960	1.713	0.775	0.748	−0.289	0.721	0.000
5F, R	5	1.187	0.365	0.169	0.158	−0.911	0.152	0.999^{ns}
PCA1	5	2.509	1.069	1.000	0.601	−1.061	0.524	0.000

Significant level for deviations from Hardy-Weinberg equilibrium (significance is $p < 0.01$). ns, not significant.

MFSN associated with PS145; MFSFW associated with P280, PS145, and PS12; and IFFW associated with PS85. Through sequence alignment *via* the National Center for Biotechnology Information (NCBI), the gene sequences of P280, PS12, and PS145 were related to fruit and seed development in other species. The explanatory rate of P280 with MFSFW was the highest, at 43.4%, which was predicted to be *WRINKLED Like 1* (*WRIL1*) belonging to the basal *ANT* subgroup of the *APETALA2/ETHYLENE-RESPONSIVE ELEMENT-BINDING FACTOR* (*AP2/ERF*) family *AINTEGUMENTA* (*ANT*) group. This gene, which has a typical AP2 domain, has a similarity of 44.79% with AtWRIL1 in *Arabidopsis*, and its main function is now known to regulate lipid biosynthesis (**Supplementary Figure S1**; Kong et al., 2020; Zhai et al., 2021). After comparison, PS145 had no similarity with the genes studied in the database, but the protein sequences that were translated had WD40 domains, which can control seed weight and volume in *Arabidopsis* (You et al., 2011). The predicted results of PS12 showed that it has the highest similarity with *MYB5* of the MYB gene family in *Arabidopsis* and is mainly involved in seed coat development and oil biosynthesis (**Supplementary Figure S2**; Li et al., 2009, 2020; Cheng et al., 2021). In addition, PS2 and PS131 were also significantly associated with fruit traits in Liu's research (Liu and Cheng, 2020a). PS2 was associated with MFFW, SFL, SFW, and MFSFW, and the explanation rate of MFFW was the highest, at 26.8%. This site has a MADS-box functional domain, and the study of the similar *AtSEP3* sequence in *Arabidopsis* showed that it controlled carpel and ovule development (**Supplementary Figure S3**; Liljegren et al., 2000; Favaro et al., 2003; Renard et al., 2020). The predicted result of PS131 was *WRKY*22 of the WRKY gene family, which can mediate dark-induced leaf senescence in *Arabidopsis* (**Supplementary Figure S4**; Zhou et al., 2011; Hsu et al., 2013; Gaudinier et al., 2018). Furthermore, *WRKY*22 (PS131), *GATA*8 (PS118; **Supplementary Figure S5**), and *ERF*3 (PS27; **Supplementary Figure S6**) jointly participate in nitrogen metabolism to regulate growth and development in *Arabidopsis* (Dubois et al., 2015; Taylor-Teeples et al., 2015; Gaudinier et al., 2018); whether these genes affect fruit and seed development in FTP through nitrogen metabolism needs further research. Therefore, the abovementioned markers provide important references for the molecular breeding of FTP.

DISCUSSION

Genetic Diversity of the Associated Population

The selection of populations with high genetic diversity and molecular markers with high polymorphism is crucial for association analysis (Ingvarsson, 2008). Here, 19 quantitative traits and 81 SSRs were utilized to evaluate the diversity of 160 accessions. In previous studies of phenotypic variation of FTP, Pang et al. (2012) reported that the CVs of 32 traits in 150 individuals ranged from 10 to 30%, while Wu et al. (2017) found that the CVs of 29 quantitative traits in 462 individuals ranged from 9.52 to 112.1%. In addition, Liu et al. studied 24 quantitative traits of 420 individuals and found that the CVs ranged from 12.03 to 106.63% (Liu and Cheng, 2020a). This study revealed CVs ranging from 11.87 to 110.64% for 19 quantitative traits in 160 accessions, which was consistent with the results of Wu et al. and Liu et al. Because the sampling strategies of these three studies were essentially the same, all the associated populations were constructed through the screening of a large number of FTP genotypes, reflecting the genetic background and structure of cultivated FTP to a certain extent. Additionally, the phenotypic variation of the 160 accessions in this study was similar to the CVs of other studies, indicating that the associated population composed of 160 accessions has high phenotypic diversity.

Phenotypic variation analysis allowed us to obtain a basic understanding of the associated population, but phenotypic

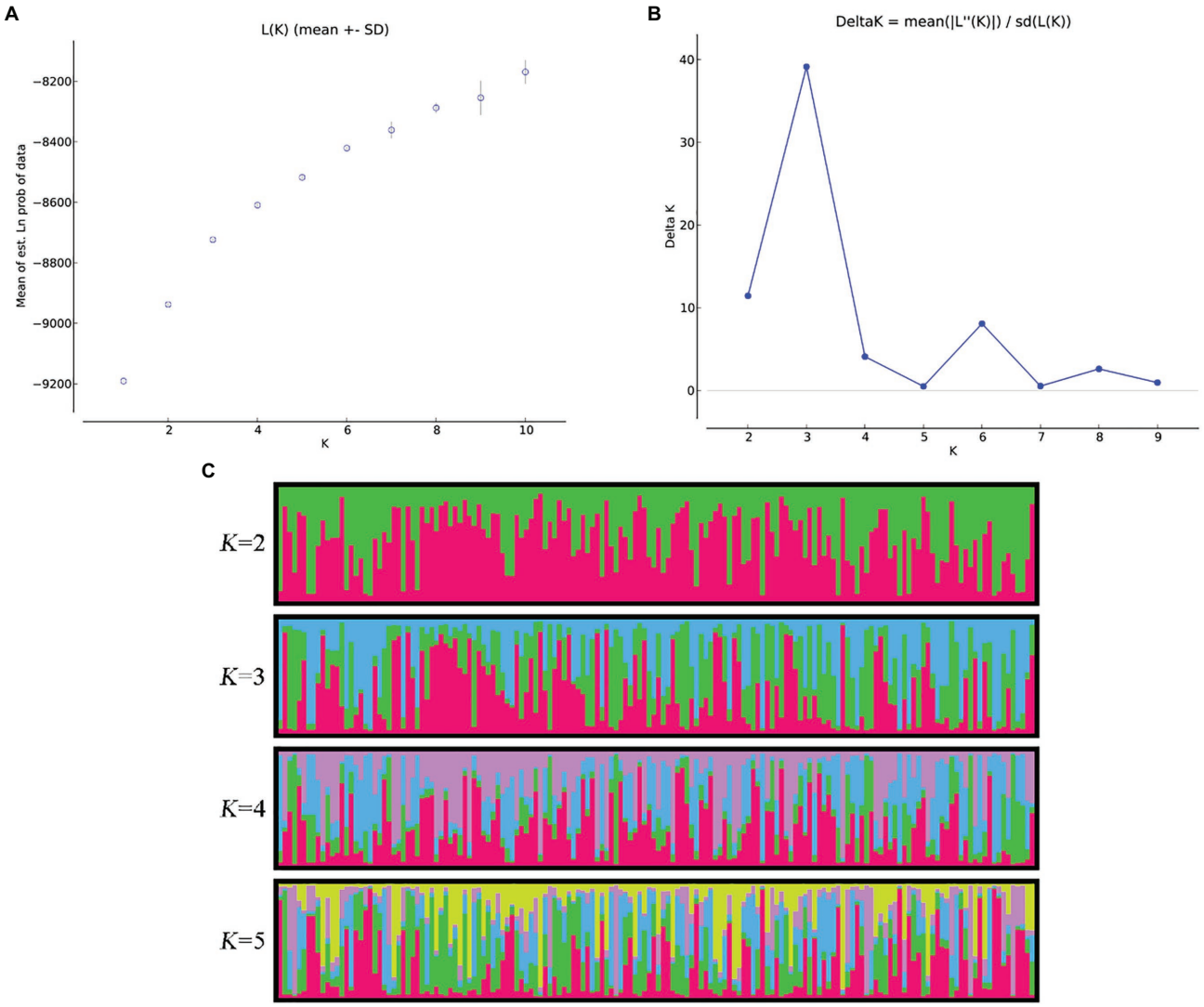


FIGURE 1 | Estimation of genetic structure of 160 accessions using 38 simple repeat sequences (SSRs) based on the STRUCTURE. **(A)** Log probability data [Ln*P*(*D*)] for each *K* value (10 replicates). **(B)** Δ*K* estimates of the posterior probability distribution of the data for a given *K*. **(C)** Estimated population structure and clustering of 160 accessions with *K* = 2–5. Accessions are shown by thin vertical lines.

variation was easily affected by the environment, and some individuals exhibited small differences in phenotypic variation, which was difficult to distinguish only by phenotypic traits. Therefore, it was very important to use molecular markers to analyze genetic diversity and population structures. In this study, 81 SSRs detected a total of 493 alleles in 160 accessions, and the N_A was 6.09, which was larger than 40 SSRs in 462 individuals (N_A = 4.5; Wu et al., 2017) and 34 SSRs in 282 cultivars (N_A = 5.441; Guo et al., 2020), but smaller than 12 SSRs in 335 individuals of wild *P. rockii* (N_A = 9.15; Yuan et al., 2012). This suggested that, on the one hand, the SSR primers and plant materials used in this study had comparatively higher allele variation, and on the other hand, the diverse primers and materials used in the studies gave rise to discrepancies. Additionally, the mean F_{IS} of 81 SSRs was −0.439, among which 63 pairs were negative, demonstrating the existence of a heterozygote surplus in the 160 accessions. We speculated that this was correlated with the hybridization origin and self-incompatibility of tree peony, which was in accordance with the reported results (Yuan et al., 2014; Zhou et al., 2014). Simultaneously, compared with other outcrossing woody plants, the genetic diversity of this species was at a moderate level (PIC = 0.476; Botstein et al., 1980), which was higher than that of *Populus tomentosa* (Du et al., 2012), but inferior to that of *Prunus avium* (Ganopoulos et al., 2011).

LD and Population Structure

The level of LD of the association population was a prerequisite for association analysis. There was a low level of LD between different SSR markers in this study. In general, the LD levels of outcrossing woody plants were low (Kulheim et al., 2009; Chhetri et al., 2019), and tree peony was also an outcrossing

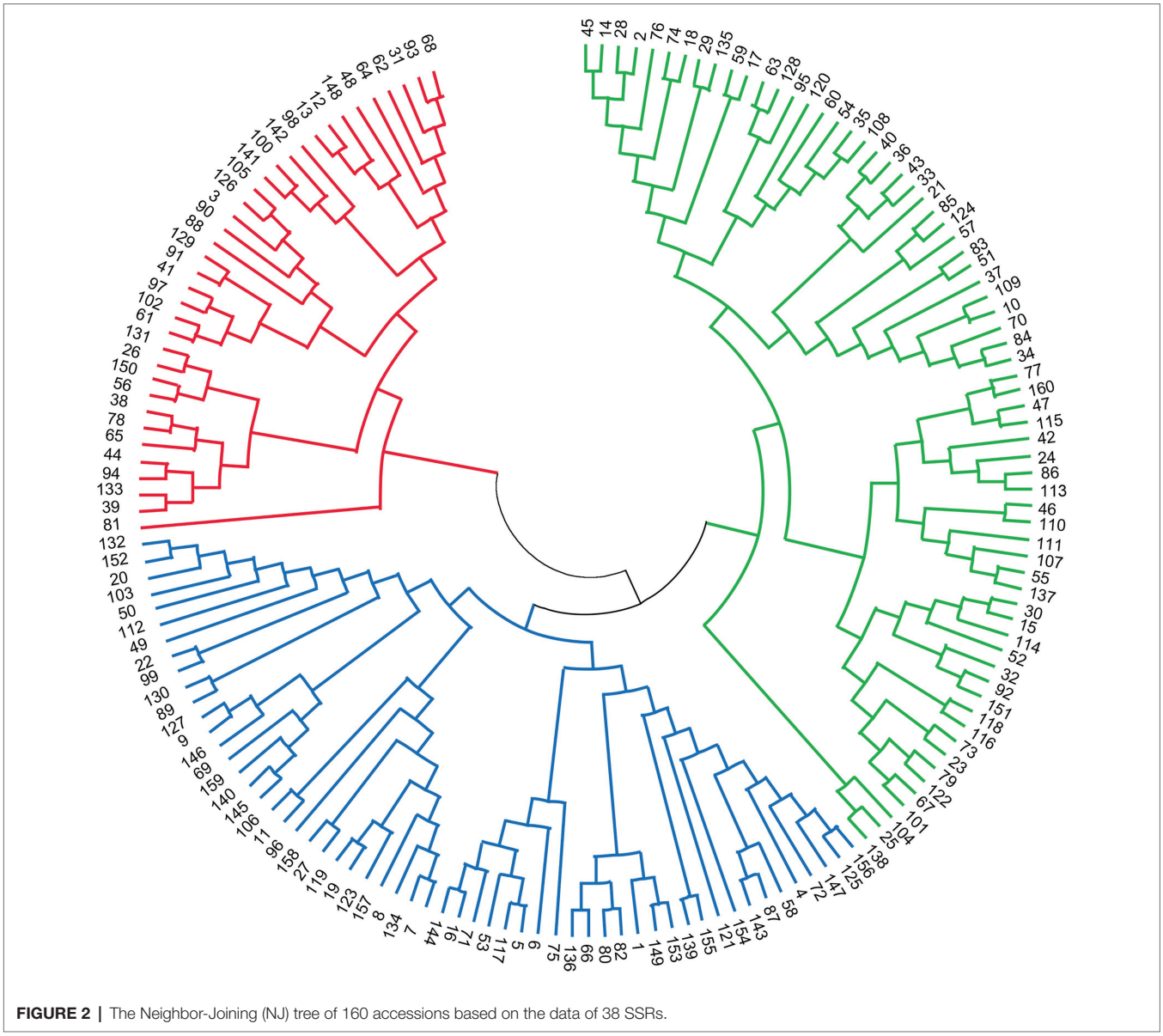


FIGURE 2 | The Neighbor-Joining (NJ) tree of 160 accessions based on the data of 38 SSRs.

species so the LD levels of FTP in other studies were also low (Wu et al., 2017; Cui et al., 2018; Liu and Cheng, 2020a). Moreover, we speculated that many human interventions, such as selective breeding, were also important contributors to the low LD levels. However, the mechanism of the LD level was still not clear for this associated population because the distances of these loci on chromosomes were unclear. Additionally, low LD levels at a small number of SSR sites did not represent the level of the entire genome or the intergenomic region. Therefore, it is necessary to use multiple markers distributed throughout the whole genome to determine the LD level in one population, because different types of markers may provide different types of insight depending on their characteristics (Neale and Savolainen, 2004).

In addition to considering the LD level of the associated population, we should also pay attention to the genetic structure of the population, because in practical research, for various reasons, it is impossible to have a population without population structure. The effect of sample structure in populations used for genetic association studies has been well documented and identified as the cause of some false associations (Newman et al., 2001; Kang et al., 2010; Sul et al., 2018). Predicting the genetic structure of the population was the premise of association analysis, which can improve the accuracy and avoid the appearance of false positives as much as possible (King et al., 2010). SSR loci that deviate from HWE may indicate genotyping error, inbreeding, population subdivision, or selection (Balding, 2006). In this study, after the Bonferroni correction for multiple testing, 43 loci were found to deviate significantly from HWE, and they were excluded. Then, the two methods were utilized to evaluate population structure and produced coincident consequences. Through STRUCTURE analysis, the 160 accessions

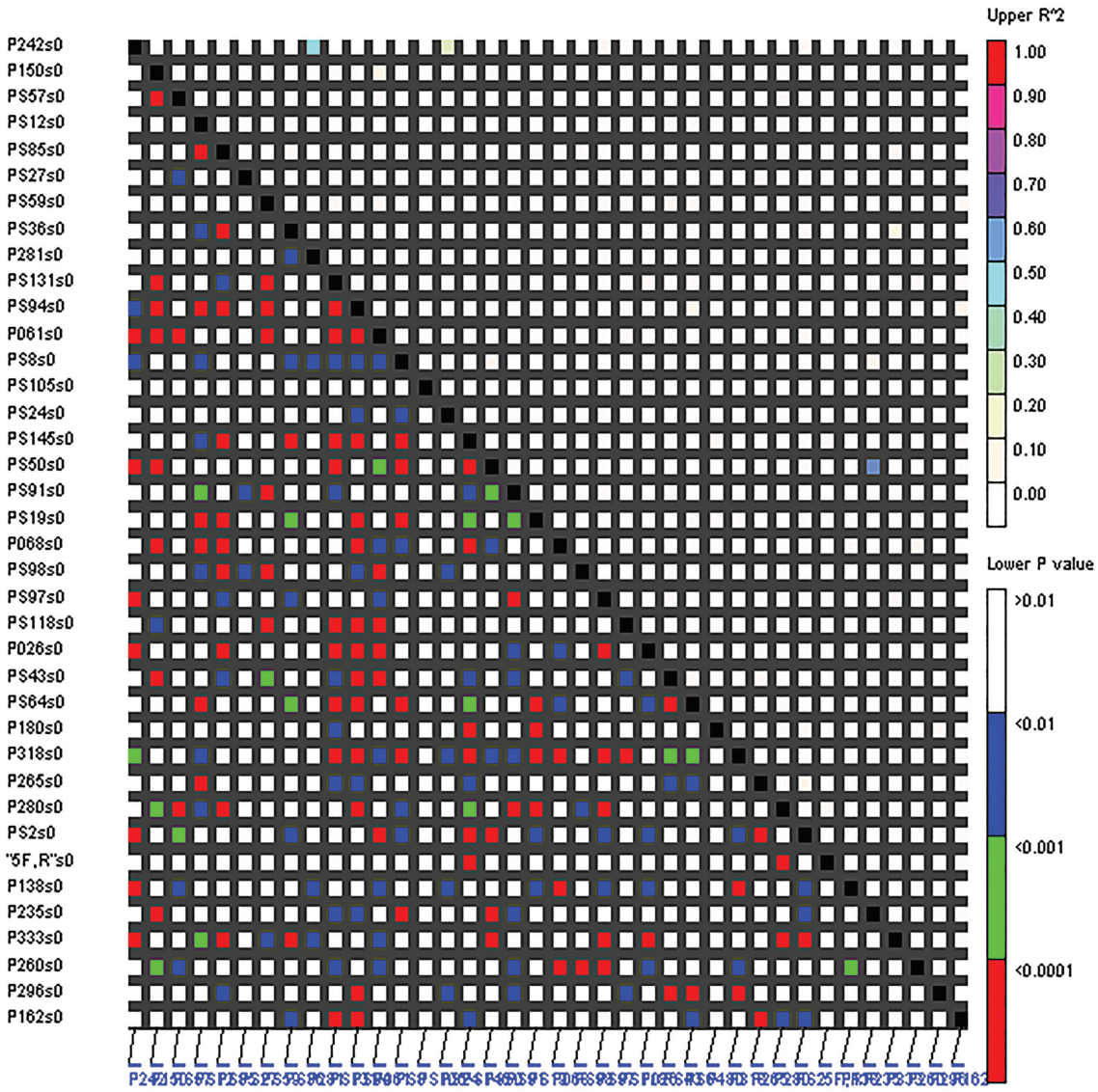


FIGURE 3 | Pairwise linkage equilibrium (LD; r^2) between 38 SSRs. The *x*- and *y*-axes represent the 38 SSRs, $r^2 < 0.1$ represents linkage equilibrium, and the different colors correspond to the thresholds of r^2 and *p*.

were divided into three subgroups, and to a large extent, this was supported by the NJ tree analysis, which was similar to previous reports of three subgroups in FTP (Wu et al., 2017; Guo et al., 2020). Furthermore, Yuan et al. (2012) revealed that 335 wild *P. rockii* individuals were mainly divided into three subpopulations, which were strongly linked to the geographical distribution pattern of wild *P. rockii*. We hypothesized that the three subgroups prescribed in this research correspond to the same three gene banks that were previously reported and also reflect their geographic origins.

Even if the subgroup structure of the associated population was considered, false positives could not be completely controlled by the general linear model (GLM; Sun et al., 2015). MLM has been demonstrated to be effective in controlling false positives and can effectively control the error rate due to its consideration of the population structure matrix (*Q*) and the kinship matrix (*K*; Yu et al., 2006). To further improve the accuracy of the association analysis results, FDR correction was carried out for all values of *p* of associations, greatly reducing the expansion of values of *p*. In this study, we found 139 significant associations, but this number dropped to 134 after FDR correction, akin to previous studies (Wu et al., 2017; Liu and Cheng, 2020a).

Associations With Floral and Fruit Traits in Flare Tree Peony

In this study, 11 SSRs were demonstrated to be significantly associated with four floral traits, four of which were derived from the transcriptome sequences of tree peony flower buds (Wu et al., 2014), and a total of 117 significant associations of 32 SSR markers related to 10 fruit traits were identified.

TABLE 4 | Significant SSR marker-trait pairs from the association test results of flower traits.

Trait	Locus	*p*	*Q*-value	R^2 (%)	2*a*	*d*	*d*/*a*	$2a/s_p$
FD	P26	2.00E−06	7.60E−05	2.23	0.0405	0.2332	11.5136	0.0221
	PS19	1.15E−03	2.19E−02	26.34	1.2277	0.5450	0.8878	0.6714
PL	P26	1.13E−03	4.28E−02	11.02	0.0096	0.1294	26.9583	0.0168
	PS24	2.93E−03	3.71E−02	7.16	1.0545	−2.1573	−4.0915	1.8448
	P235	3.53E−03	3.35E−02	10.69	0.3987	0.3841	1.9265	0.6975
	PS8	3.96E−03	3.01E−02	14.6	0.3015	−0.0402	−0.2663	0.5275
	P281	4.34E−03	2.75E−02	5.03	1.6221	−3.4890	−4.3018	2.8378
	PS97	5.05E−03	2.74E−02	9.08	0.1067	−2.7200	−50.9831	0.1867
	P333	5.59E−03	2.65E−02	11.12	0.1372	−0.1193	−1.7391	0.2400
	PS91	8.27E−03	3.49E−02	12.53	0.0055	0.0513	18.6364	0.0096
PW	P281	9.29E−04	3.53E−02	6.72	1.7530	−2.8435	−3.2442	3.2801
	PS24	1.70E−03	3.24E−02	7.8	0.7014	−1.6207	−4.6213	1.3124
	P235	5.26E−03	5.00E−02	10.14	0.4548	0.2835	1.2467	0.8510
	PS57	5.34E−03	4.06E−02	7.79	0.1996	0.0357	0.3577	0.3735
	PS43	6.90E−03	4.37E−02	14.68	0.0872	0.2411	5.5287	0.1632
CN	P26	2.00E−06	7.60E−05	18.26	0.0615	0.1042	3.3870	0.0817
	PS19	1.15E−03	2.19E−02	25.38	0.0320	0.4916	30.7250	0.0425

p, significance level for association (significance is $p < 0.01$); R^2, percentage of the phenotypic variance explained; Q-value, a correction for multiple testing [false discovery rate FDR (Q) < 0.05]; s_p, standard deviation for the phenotypic trait under consideration.

More than two loci were significantly associated with each trait, indicating that quantitative traits were controlled by microeffect polygenes. Complex quantitative traits of plants for association analysis can be significantly associated with many sites (Sun et al., 2015); similar conclusions have been reported in other studies of trees (Dillon et al., 2012; Porth et al., 2013). Simultaneously, we also found that a marker was significantly associated with multiple traits, such as P26, which was associated with FD, FL, and CN, which may be due to the significant correlation between these phenotypic traits, and may reflect the characteristics of pleiotropism (Thudi et al., 2014).

Phenotyping is an important part of tree association analysis. Typical associated populations are usually composed of different unrelated individuals grown under the same ambient conditions and augment what is known about the measurement of phenotypes, which must usually be asexually reproduced to reduce environmental interference and measurement errors (Du et al., 2013). Although there have been previous studies on the association analysis of important traits in FTP, the samples used were all individuals and had no asexual reproduction (Wu et al., 2017; Cui et al., 2018; Liu and Cheng, 2020a). Hence, in this study, we used 480 phenotypes (160 genotypes × 3 clones) to compensate for the limited number of EST-SSR markers, and repeated data from each accession could be integrated to generate a phenotypic mean for analysis, which reduced the impact of measurement errors. In this study, the average explanatory rates of floral and fruit traits were 11.80 and 17.68%, respectively, higher than those of Wu et al. (2017; flower traits: 5.50%) and Liu and Cheng (2020a; fruit traits: 6.53%), which also reflected the improvement of the effectiveness of association mapping.

Additionally, replication of genotype-phenotype associations is crucial in association mapping to distinguish false-positive associations. Therefore, of the 38 SSRs used in this study, 17 markers were in conformity with Wu et al. (2017), but the association analysis results of flower traits in the two studies were inconsistent. This may be due to differences in sample size, gene-environment interactions, genetic background, gene-gene interactions, or other factors; therefore, some real associations may not be repeated in unrelated datasets (Greene et al., 2009; Beaulieu et al., 2011; Du et al., 2013). In addition, there were 21 markers identical to those described in Liu et al. Most of the associations were different, but associations between the same markers and traits were still found in the two studies. In both studies, PS2 was significantly associated with MFFW, PS12 was significantly associated with MFFW and MFSFW, and PS131 was significantly associated with SFL (Liu and Cheng, 2020a). We hypothesized that these repeated associations might help identify important genomic regions. These findings were also of great value in the use of marker-assisted selective breeding for trait improvement. Future studies will require multiple germplasm populations to combine multiyear and multiplot phenotypic data to validate the developed markers.

Flare tree peony has developed into an emerging woody oil crop in China, and its output mainly refers to ISFW. PS12, PS27, PS131, PS118, and ps280 were significantly associated with ISFW, and may be the key sites affecting the yield. Studies have shown that the *MYB5* gene of PS12 is involved in seed coat development to control seed size and affect yield (Li et al., 2008; Su et al., 2011; Dong et al., 2017). The oil content in tree peony seeds can reach 27–33%, which can directly affect seed weight (Cui et al., 2016). The *WRIL1* gene corresponding to P280 may be an important gene regulating oil synthesis in tree peony. Nitrogen metabolism is closely related to plant growth and development, thereby affecting yield (Tilman et al., 2002). In this study, *GATA8* (PS118), *ERF3* (PS27), and *WRKY22* (PS131) may form a network to participate in nitrogen metabolism. In addition, *WRKY22* was also involved in leaf senescence. Relevant studies have shown that the combination of related genes can delay the senescence process of plants and significantly increase yield (Distelfeld et al., 2014; Lira et al., 2017). This showed that these loci were potential genes to increase the yield of FTP, which is worthy of further study.

TABLE 5 | Significant SSR marker-trait pairs from the association test results of fruit traits.

Trait	Locus	*p*	*Q*-value	R^2 (%)	2*a*	*d*	*d*/*a*	$2a/s_p$
NCWS	P26	7.17E−07	2.73E−05	19.63	0.0240	−0.0310	−2.5833	0.0229
	PS19	1.12E−04	2.13E−03	28.87	0.3633	0.6151	3.3859	0.3465
MFFW	P242	3.89E−09	1.48E−07	19.76	40.4683	−41.2109	−2.0367	3.9887
	PS24	3.99E−10	7.59E−09	24.09	62.7752	−52.4424	−1.6708	6.1873
	PS59	4.07E−10	5.16E−09	25.79	2.9592	−21.8023	−14.7353	0.2917
	PS2	6.90E−10	6.55E−09	26.8	62.3351	−33.4870	−1.0744	6.1439
	PS145	1.61E−08	1.23E−07	31.98	4.5925	−0.2254	−0.0982	0.4527
	PS85	1.70E−08	1.08E−07	31.93	0.7140	1.8037	5.0524	0.0704
	P318	3.68E−08	2.00E−07	34.1	2.5125	−1.7301	−1.3772	0.2476
	PS131	4.25E−08	2.02E−07	26.53	1.2683	−0.9481	−1.4950	0.1250
	PS27	4.74E−08	2.00E−07	25.19	3.2202	−0.2004	−0.1245	0.3174
	PS91	8.90E−08	3.38E−07	26.91	6.1555	−1.2124	−0.3939	0.6067
	PS12	2.80E−05	9.67E−05	30.77	2.2168	−1.4778	−1.3333	0.2185
	P280	6.70E−05	2.12E−04	28.73	2.9589	1.5406	1.0413	0.2916
	P162	2.70E−04	7.89E−04	11.45	23.4680	−13.0793	−1.1146	2.3131
	PS50	6.85E−04	1.86E−03	16.13	23.1626	−14.0836	−1.2161	2.2830
	P260	2.00E−03	5.06E−03	7.61	15.0406	−10.1388	−1.3482	1.4824
SFL	P318	5.54E−08	2.11E−06	33.66	0.4855	−0.2852	−1.1747	0.0812
	PS2	1.10E−05	2.09E−04	16.53	26.6660	−14.2622	−1.0697	4.4588
	PS24	1.20E−05	1.52E−04	13.44	27.0421	−53.7990	−3.9789	4.5217
	PS131	3.50E−05	3.33E−04	18.88	1.8602	−0.3113	−0.3347	0.3110
	PS85	6.90E−05	5.24E−04	22.24	5.8173	2.4138	0.8299	−0.9727
	PS59	7.70E−05	4.88E−04	12.92	0.4488	−40.4321	−180.1787	0.0750
	PS145	8.70E−05	4.72E−04	21.93	1.8258	−0.1060	−0.1161	0.3053
	PS27	1.20E−04	5.70E−04	16.23	4.9738	2.0351	0.8183	0.8317
	PS43	1.21E−04	5.11E−04	20.5	25.3766	12.7560	1.0053	4.2432
	P61	3.36E−04	1.28E−03	12.52	0.6538	−0.5210	−1.5938	0.1093
	PS12	1.17E−03	4.05E−03	25.35	3.6484	−2.0494	−1.1235	0.6100
	P242	4.03E−03	1.28E−02	5.11	12.1354	−46.4023	−7.6474	2.0291
SFW	P242	2.69E−04	1.02E−02	8.08	6.2226	−18.3637	−5.9023	2.5511
	PS91	1.47E−03	2.79E−02	15.07	1.8603	0.1002	0.1078	0.7627
	PS2	2.95E−03	3.73E−02	9.78	6.8639	−4.1569	−1.2112	2.8140
MFSN	P242	1.26E−11	4.78E−10	5.53	20.6461	−22.3470	−2.1648	2.1098
	P150	1.45E−08	2.75E−07	17.77	2.1017	−6.8179	−6.4879	0.2148
	PS57	2.00E−06	2.53E−05	9.43	0.7912	−6.2323	−15.7540	0.0809
	PS12	3.00E−06	2.85E−05	29.76	4.4831	−2.3336	−1.0410	0.4581
	PS85	3.00E−06	2.28E−05	17.62	9.3424	2.4571	0.5260	0.9547
	PS27	3.00E−06	1.90E−05	4.67	2.5904	−2.9028	−2.2412	0.2647
	PS59	9.00E−06	4.89E−05	12.82	6.5719	−13.6332	−4.1489	0.6716
	PS36	1.10E−05	5.23E−05	11.36	2.6301	−7.5893	−5.7711	0.2688
	P281	1.20E−05	5.07E−05	0.18	5.2508	−14.8746	−5.6657	0.5366
	PS131	2.80E−05	1.06E−04	18.9	2.6919	−1.9727	−1.4656	0.2751
	PS94	3.50E−05	1.21E−04	15.52	5.5730	−3.9375	−1.4131	0.5695
	P61	5.90E−05	1.87E−04	11.22	3.5347	−5.9028	−3.3399	0.3612
	PS8	8.20E−05	2.40E−04	19.99	6.2574	−8.0737	−2.5805	0.6394
	PS105	8.40E−05	2.28E−04	11.81	11.4994	−16.9536	−2.9486	1.1751
	PS24	2.91E−04	7.37E−04	4.12	24.5754	−24.3823	−1.9843	2.5113
	PS145	5.12E−04	1.22E−03	32.1	0.2785	−5.1720	−37.1415	0.0285
	PS50	8.13E−04	1.82E−03	23.24	15.8575	−8.3573	−1.0540	1.6204
	PS91	9.54E−04	2.01E−03	22.89	3.7400	−6.8927	−3.6859	0.3822
	PS19	1.44E−03	2.88E−03	21.49	6.8567	7.4117	2.1619	0.7007
	P68	1.88E−03	3.56E−03	19.51	18.8487	−8.2082	−0.8710	1.9261
	PS98	2.49E−03	4.51E−03	12.63	3.6896	−5.9812	−3.2422	0.3770
	PS97	2.76E−03	4.77E−03	16.59	13.7810	−17.7528	−2.5764	1.4083
	PS118	4.61E−03	7.61E−03	17.99	1.2030	−8.3587	−13.8964	0.1229
	P26	7.19E−03	1.14E−02	9.2	2.6599	−6.4400	−4.8422	0.2718
	PS43	8.84E−03	1.34E−02	14.04	7.9873	3.5050	0.8776	0.8162
MFSFW	P242	4.01E−07	1.52E−05	15.05	16.5058	−13.6071	−1.6488	3.4808
	P280	1.36E−10	2.59E−09	43.4	3.6135	0.0060	0.0033	0.7620
	PS145	3.68E−09	4.66E−08	33.51	0.5759	−1.6869	−5.8581	0.1214
	PS91	2.59E−08	2.46E−07	28.22	1.8375	−2.6293	−2.8618	0.3875
	PS59	8.72E−08	6.62E−07	20.39	2.5827	−6.0147	−4.6576	0.5447
	PS131	8.96E−07	5.67E−06	23.2	0.4428	−0.7498	−3.3866	0.0934

(Continued)

TABLE 5 | Continued

Trait	Locus	*p*	*Q*-value	R^2 (%)	2*a*	*d*	*d*/*a*	$2a/s_p$
	PS24	4.00E−06	2.17E−05	14.67	22.8577	−16.8312	−1.4727	4.8204
	P318	4.00E−06	1.90E−05	28.81	1.8230	−0.8516	−0.9343	0.3844
	PS2	5.00E−06	2.11E−05	17.55	22.8615	−11.7372	−1.0268	4.8212
	PS50	7.00E−06	2.66E−05	21.96	10.9331	−6.1514	−1.1253	2.3056
	PS85	1.00E−05	3.45E−05	24.67	3.0100	1.1058	0.7348	0.6348
	PS12	1.40E−05	4.43E−05	31.7	0.9607	−1.0020	−2.0859	0.2026
	PS118	4.10E−05	1.20E−04	15.06	1.3036	−3.9239	−6.0201	0.2749
	PS64	4.30E−05	1.17E−04	16.27	4.2022	0.7440	0.3541	0.8862
	PS27	2.69E−04	6.81E−04	15.21	0.4414	−1.1026	−4.9959	0.0931
	PS8	8.01E−04	1.90E−03	16.93	2.8796	−2.7622	−1.9185	0.6073
	PS36	8.43E−04	1.88E−03	10.08	2.3534	−3.8258	−3.2513	0.4963
	P162	9.78E−04	2.06E−03	9.9	10.7435	−5.3716	−1.0000	2.2657
	P150	1.22E−03	2.45E−03	12.1	1.1925	−3.0146	−5.0558	0.2515
	PS105	1.53E−03	2.91E−03	6.17	4.0299	−7.1978	−3.5722	0.8499
	PS97	1.67E−03	3.02E−03	10.51	4.9218	−7.4541	−3.0290	1.0379
	P260	1.75E−03	3.02E−03	7.77	6.3885	−4.7610	−1.4905	1.3472
	PS57	6.33E−03	1.05E−02	7.58	0.8793	−2.8582	−6.5010	0.1854
IFN	P150	2.56E−04	9.73E−03	14.1	6.9075	−3.8298	−1.1089	0.4404
ISN	P280	3.67E−04	1.39E−02	26.29	195.6839	96.1550	0.9828	0.5165
	P150	7.50E−04	1.43E−02	12.73	130.2499	−238.2099	−3.6577	0.3438
	PS118	9.46E−04	1.20E−02	11.23	20.7663	−235.6122	−22.6918	0.0548
	PS91	1.09E−03	1.04E−02	15.48	80.4876	−227.2271	−5.6463	0.2125
	PS64	2.65E−03	2.02E−02	11.07	175.7995	155.3601	1.7675	0.4641
	PS59	5.70E−03	3.61E−02	7.71	173.9827	−403.9357	−4.6434	0.4593
	PS12	6.83E−03	3.71E−02	22.36	208.7671	−32.2896	−0.3093	0.5511
	P242	9.67E−03	4.59E−02	4.16	693.3875	−706.6413	−2.0382	1.8303
IFFW	PS85	1.77E−09	6.71E−08	34.25	102.9559	−122.7455	−2.3844	0.2668
	PS24	3.40E−05	6.46E−04	12.28	1,672.3147	−1,333.3727	−1.5946	4.3341
	PS59	6.30E−05	7.98E−04	13.16	85.4445	−521.1641	−12.1989	0.2214
	PS27	2.73E−04	2.59E−03	15.19	210.2070	10.6415	0.1012	0.5448
	P242	3.11E−04	2.36E−03	7.93	974.6330	−985.1185	−2.0215	2.5259
	PS2	4.02E−04	2.55E−03	12.3	1,666.0573	−867.6269	−1.0415	4.3178
	P318	4.72E−04	2.56E−03	22.41	252.5740	−40.4510	−0.3203	0.6546
	PS131	4.79E−04	2.28E−03	15.56	12.9643	−78.5108	−12.1119	0.0336
	PS91	7.68E−04	3.24E−03	15.97	132.2274	−110.9912	−1.6788	0.3427
	PS145	1.58E−03	6.02E−03	17.87	98.3069	−44.8050	−0.9115	0.2548
	PS12	2.69E−03	9.30E−03	23.98	21.5857	−14.9735	−1.3873	0.0559
	P150	4.62E−03	1.46E−02	10.32	177.2492	−174.8099	−1.9725	0.4594
	P280	9.06E−03	2.65E−02	21.04	318.1074	2.3626	0.0149	0.8244
ISFW	P242	2.80E−05	1.06E−03	10.55	548.9454	−437.8473	−1.5952	2.9145
	PS59	3.40E−05	6.46E−04	13.87	67.6135	−181.3082	−5.3631	0.3590
	PS91	4.60E−05	5.83E−04	19.67	43.3480	−93.1077	−4.2958	0.2301
	P280	8.60E−05	8.17E−04	28.37	69.8530	51.2509	1.4674	0.3709
	PS24	9.30E−05	7.07E−04	11.16	795.7008	−562.9896	−1.4151	4.2246
	PS85	1.56E−04	9.88E−04	21.15	22.8575	10.7868	0.9438	0.1214
	PS2	2.76E−04	1.50E−03	12.77	796.2987	−404.4477	−1.0158	4.2278
	PS12	3.13E−04	1.49E−03	27.38	73.4092	−6.9552	−0.1895	0.3898
	PS131	6.79E−04	2.87E−03	15.09	7.6998	−29.8505	−7.7536	0.0409
	PS64	2.16E−03	8.22E−03	11.35	42.4772	85.9071	4.0449	0.2255
	PS145	2.26E−03	7.80E−03	17.34	20.6608	−64.2305	−6.2176	0.1097
	P318	2.39E−03	7.57E−03	19.95	120.9243	−30.5276	−0.5049	0.6420
	P150	2.55E−03	7.45E−03	11.13	68.8411	−111.4819	−3.2388	0.3655
	PS118	3.01E−03	8.16E−03	9.75	23.1819	−120.6926	−10.4127	0.1231
	PS27	3.04E−03	7.71E−03	11.98	25.3010	−24.1708	−1.9107	0.1343

p, significance level for association (significance is p < 0.01); R^2, percentage of the phenotypic variance explained; Q-value, a correction for multiple testing [false discovery rate FDR(Q) < 0.05]; s_p, standard deviation for the phenotypic trait under consideration.

The analysis of the genetic regulatory relationships between different significant association sites of the same quantitative trait was helpful for breeding with significant association combinations. For example, in floral traits, FD was an important ornamental character that determined its ornamental value to a large extent. P26 was significantly associated with FD and showed an overdominant effect (d/a = 11.5136), indicating that individuals with P26 heterozygosity sites may produce flowers of larger diameter. For fruit traits, ISFW was considered a yield indicator, with 15 markers significantly associated with it.

Among them, 11 markers, namely, P242, PS59, PS91, P280, PS24, PS131, PS64, PS145, P150, PS118, and PS27, showed the same pattern of gene action (overdominance), while the effects of P280 (d/a = 1.4674) and PS64 (d/a = 4.0449) were positive, showing that individuals with P280 and PS64 heterozygous loci might generate heavier seeds. In summary, marker combinations based on the gene effect value of associated loci can be used for the early selection of target traits.

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

XG and FC: conceptualization and validation. XG: formal analysis, data curation, writing – original draft preparation, and visualization. XG, XC, and XT: investigation. XC and XT: resources. XG, CH, FC, YZ, XC, and XT: writing – review and editing. FC: supervision and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This research was mainly supported by the Special Foundation for National Science and Technology Basic Research Program of China (Grant No. 2019FY100802), the National Key Research and Development Program of China (Grant No. 2019YFD1001502), and the Science and Technology Project of Beijing (Grant No. Z181100002518001).

ACKNOWLEDGMENTS

We would like to thank Lidan Sun (Beijing Forestry University) for her suggestions on this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.664814/full#supplementary-material

REFERENCES

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791. doi: 10.1038/nrg1916

Beaulieu, J., Doerksen, T., Boyle, B., Clément, S., Deslauriers, M., Beauseigle, S., et al. (2011). Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics* 188, 197–214. doi: 10.1534/genetics.110.125781

Beavis, W. D. (1998). "QTL analyses: power, precision, and accuracy," in *Molecular Dissection of Complex Traits*. ed. A. H. Patterson (Boca Raton, FL: CRC Press), 145–162.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331. doi: 10.1016/0165-1161(81)90274-0

Cai, C. F., Cheng, F. Y., Wu, J., Zhong, Y., and Liu, G. X. (2015). The first high-density genetic map construction in tree peony (*Paeonia* sect. *Moutan*) using genotyping by specific-locus amplified fragment sequencing. *PLoS One* 10:e128584. doi: 10.1371/journal.pone.0128584

Cheng, F. Y. (2012). Advances in the breeding of tree peonies and a cultivar system for the cultivar group. *Int. J. Plant Breed.* 1, 89–104.

Cheng, F. Y., Li, J. J., Chen, D. Z., and Zhang, Z. S. (2005). *Chinese Paeonia rockii*. Beijing: Chinese Forestry Publishing House.

Cheng, F. Y., and Yu, X. N. (2008). Flare tree peonies (*Paeonia rockii* hybrids) and the origin of the cultivar group. *Acta Hortic.* 766, 375–382. doi: 10.17660/ActaHortic.2008.766.49

Cheng, T. H., Zhao, P., Ren, Y. R., Zou, J., and Sun, M. X. (2021). AtMIF1 increases seed oil content by attenuating GL2 inhibition. *New Phytol.* 229, 2152–2162. doi: 10.1111/nph.17016

Chhetri, H. B., Macaya-Sanz, D., Kainer, D., Biswal, A. K., Evans, L. M., Chen, J. G., et al. (2019). Multi-trait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. *New Phytol.* 223, 293–309. doi: 10.1111/nph.15777

Cui, H. L., Chen, C. R., Huang, N. Z., and Cheng, F. Y. (2018). Association analysis of yield, oil and fatty acid content, and main phenotypic traits in *Paeonia rockii* as an oil crop. *J. Hortic. Sci. Biotechnol.* 93, 425–432. doi: 10.1080/14620316.2017.1381045

Cui, H. L., Cheng, F. Y., and Peng, L. P. (2016). Determination of the fatty acid composition in tree peony seeds using near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* 93, 943–952. doi: 10.1007/s11746-016-2839-0

Davis, J. W. (2008). Handbook of univariate and multivariate data analysis and interpretation with SPSS. *Am. Stat.* 62:268. doi: 10.1198/000313008X332287

Dillon, S. K., Brawner, J. T., Meder, R., Lee, D. J., and Southerton, S. G. (2012). Association genetics in *Corymbia citriodora subsp. variegata* identifies single nucleotide polymorphisms affecting wood growth and cellulosic pulp yield. *New Phytol.* 195, 596–608. doi: 10.1111/j.1469-8137.2012.04200.x

Distelfeld, A., Avni, R., and Fischer, A. M. (2014). Senescence, nutrient remobilization, and yield in wheat and barley. *J. Exp. Bot.* 65, 3783–3798. doi: 10.1093/jxb/ert477

Dong, H., Dumenil, J., Lu, F. H., Na, L., Vanhaeren, H., Naumann, C., et al. (2017). Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in *Arabidopsis*. *Genes Dev.* 31, 197–208. doi: 10.1101/gad.292235.116

Du, Q. Z., Pan, W., Xu, B. H., Li, B. L., and Zhang, D. Q. (2013). Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (PtoCesA) genes are associated with growth and wood properties in *Populus tomentosa*. *New Phytol.* 197, 763–776. doi: 10.1111/nph.12072

Du, Q. Z., Wang, B. W., Wei, Z. Z., Zhang, D. Q., and Li, B. L. (2012). Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered.* 103, 853–862. doi: 10.1093/jhered/ess061

Dubois, M., Van den Broeck, L., Claeys, H., Van Vlierberghe, K., Matsui, M., and Inzé, D. (2015). The ETHYLENE RESPONSE FACTORs ERF6 and ERF11 antagonistically regulate mannitol-induced growth inhibition in *Arabidopsis*. *Plant Physiol.* 169, 166–179. doi: 10.1104/pp.15.00335

Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Eckert, A. J., Bower, A. D., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Krutovsky, K. V., et al. (2009). Association genetics of coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* 182, 1289–1302. doi: 10.1534/genetics.109.102350

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., et al. (2003). MADS-Box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611. doi: 10.1105/tpc.015123

Gai, S. P., Zhang, Y. X., Mu, P., Liu, C. Y., Liu, S., Dong, L., et al. (2012). Transcriptome analysis of tree peony during chilling requirement fulfillment: assembling, annotation and markers discovering. *Gene* 497, 256–262. doi: 10.1016/j.gene.2011.12.013

Ganopoulos, I. V., Kazantzis, K., Chatzicharisis, I., Karayiannis, I., and Tsaftaris, A. S. (2011). Genetic diversity, structure and fruit trait associations in Greek sweet cherry cultivars using microsatellite based (SSR/ISSR) and morpho-physiological markers. *Euphytica* 181, 237–251. doi: 10.1007/s10681-011-0416-z

Gao, Z. M., Wu, J., Liu, Z. A., Wang, L. S., Ren, H. X., and Shu, Q. Y. (2013). Rapid microsatellite development for tree peony and its implications. *BMC Genomics* 14:886. doi: 10.1186/1471-2164-14-886

Gaudinier, A., Rodriguez-Medina, J., Zhang, L. F., Olson, A., Liseron-Monfils, C., Bagman, A., et al. (2018). Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* 563, 259–264. doi: 10.1038/s41586-018-0656-3

Grattapaglia, D., and Resende, M. D. V. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genomes* 7, 241–255. doi: 10.1007/s11295-010-0328-4

Greene, C. S., Penrod, N. M., Williams, S. M., and Moore, J. H. (2009). Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One* 4:e5639. doi: 10.1371/journal.pone.0005639

Guo, X., Cheng, F. Y., and Zhong, Y. (2020). Genetic diversity of *Paeonia rockii* (flare tree peony) germplasm accessions revealed by phenotypic traits, EST-SSR markers and chloroplast DNA sequences. *Forests* 11:672. doi: 10.3390/f11060672

Guo, Q., Guo, L. L., Zhang, L., Zhang, L. X., Ma, H. L., Guo, D. L., et al. (2017). Construction of a genetic linkage map in tree peony (*Paeonia* sect. *Moutan*) using simple sequence repeat (SSR) markers. *Sci. Hortic.* 219, 294–301. doi: 10.1016/j.scienta.2017.03.017

Hardy, O. J., and Vekemans, X. (2002). Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x

Homolka, A., Berenyi, M., Burg, K., Kopecky, D., and Fluch, S. (2010). Microsatellite markers in the tree peony, *Paeonia suffruticosa* (Paeoniaceae). *Am. J. Bot.* 97, 42–44. doi: 10.3732/ajb.1000127

Hong, D. Y., and Pan, K. Y. (1999). Taxonomical history and revision of *Paeonia* sect. *Moutan* (Paeoniaceae). *Acta Phytotaxon. Sin.* 37, 48–65.

Hou, X. G., Guo, D. L., Cheng, S. P., and Zhang, J. Y. (2011). Development of thirty new polymorphic microsatellite primers for *Paeonia suffruticosa*. *Biol. Plant.* 55, 708–710. doi: 10.1007/s10535-011-0172-x

Hsu, F. C., Chou, M. Y., Chou, S. J., Li, Y. R., Peng, H. P., and Shih, M. C. (2013). Submergence confers immunity mediated by the WRKY22 transcription factor in *Arabidopsis*. *Plant Cell* 25, 2699–2713. doi: 10.1105/tpc.113.114447

Ingvarsson, P. K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* 180, 329–340. doi: 10.1534/genetics.108.090431

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354. doi: 10.1038/ng.548

King, R. A., Harris, S. L., Karp, A., and Barker, J. H. A. (2010). Characterisation and inheritance of nuclear microsatellite loci for use in population studies of the allotetraploid *Salix alba-Salix fragilis* complex. *Tree Genet. Genomes* 6, 247–258. doi: 10.1007/s11295-009-0245-6

Kong, Q., Yang, Y. Z., Guo, L., Yuan, L., and Ma, W. (2020). Molecular basis of plant oil biosynthesis: insights gained from studying the WRINKLED1 transcription factor. *Front. Plant Sci.* 11:24. doi: 10.3389/fpls.2020.00024

Kulheim, C., Yeoh, S. H., Maintz, J., Foley, W. J., and Moran, G. F. (2009). Comparative SNP diversity among four *eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* 10:452. doi: 10.1186/1471-2164-10-452

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Li, S. F., Allen, P. J., Napoli, R. S., Browne, R. G., Pham, H., and Parish, R. W. (2020). MYB-bHLH-TTG1 regulates *Arabidopsis* seed coat biosynthesis pathways directly and indirectly via multiple tiers of transcription factors. *Plant Cell Physiol.* 61, 1005–1018. doi: 10.1093/pcp/pcaa027

Li, S. F., Milliken, O. N., Pham, H., Seyit, R., Napoli, R., Preston, J., et al. (2009). The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* 21, 72–89. doi: 10.1105/tpc.108.063503

Li, S. S., Wang, L. S., Shu, Q. Y., Wu, J., Chen, L. G., Shao, S., et al. (2015a). Fatty acid composition of developing tree peony (*Paeonia* section *Moutan* DC.) seeds and transcriptome analysis during seed development. *BMC Genomics* 16:208. doi: 10.1186/s12864-015-1429-0

Li, S. S., Yuan, R. Y., Chen, L. G., Wang, L. S., Hao, X. H., Wang, L. J., et al. (2015b). Systematic qualitative and quantitative assessment of fatty acids in the seeds of 60 tree peony (*Paeonia* section *Moutan* DC.) cultivars by GC-MS. *Food Chem.* 173, 133–140. doi: 10.1016/j.foodchem.2014.10.017

Li, Y. H., Zheng, L. Y., Corke, F., Smith, C., and Bevan, M. W. (2008). Control of final seed and organ size by the *DA1* gene family in *Arabidopsis thaliana*. *Genes Dev.* 22, 1331–1336. doi: 10.1101/gad.463608

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089

Lira, B. S., Gramegna, G., Trench, B. A., Alves, F. R. R., Silva, E. M., Silva, G. F. F., et al. (2017). Manipulation of a senescence-associated gene improves fleshy fruit yield. *Plant Physiol.* 175, 77–91. doi: 10.1104/pp.17.00452

Liu, N., and Cheng, F. Y. (2020a). Association mapping for yield traits in *Paeonia rockii* based on SSR markers within transcription factors of comparative transcriptome. *BMC Plant Biol.* 20:245. doi: 10.1186/s12870-020-02449-6

Liu, N., and Cheng, F. Y. (2020b). Variation and yield grading standards of fruit traits in *Paeonia rockii* (S. G. Haw et L. A. Lauener) T. Hong et J. J. Li. *Plant Sci. J.* 38, 378–389. doi: 10.11913/PSJ.2095-0837.2020.30378

Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi: 10.1016/j.tplants.2004.05.006

Newman, D. L., Abney, M., McPeak, M. S., Ober, C., and Cox, N. J. (2001). The importance of genealogy in determining genetic associations with complex traits. *Am. J. Hum. Genet.* 69, 1146–1148. doi: 10.1086/323659

Pang, L. Z., Cheng, F. Y., Zhong, Y., Cai, C. F., and Cui, H. L. (2012). Phenotypic analysis of association population for flare tree peony. *J. Beijing For. Univer.* 34, 115–120.

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Porth, I., Klapste, J., Skyba, O., Hannemann, J., McKown, A. D., Guy, R. D., et al. (2013). Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytol.* 200, 710–726. doi: 10.1111/nph.12422

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Renard, J., Niñoles, R., Martínez Almonacid, I., Gayubas, B., Mateos Fernández, R., Bissoli, G., et al. (2020). Identification of novel seed longevity genes related to oxidative stress and seed coat by genome-wide association studies and reverse genetics. *Plant Cell Environ.* 43, 2523–2539. doi: 10.1111/pce.13822

Resende, R. T., Resende, M. D., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., et al. (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in eucalyptus. *New Phytol.* 213, 1287–1300. doi: 10.1111/nph.14266

Rombel, I. T., Sykes, K. F., Rayner, S., and Johnston, S. A. (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282, 33–41. doi: 10.1016/S0378-1119(01)00819-8

Rosenberg, N. A. (2004). Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4, 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rousset, F. (2008). Genepop' 007: a complete re-implementation of the Genepop software for windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genome-wide experiments. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445. doi: 10.1073/pnas.1530509100

Su, Z. Q., Hao, C. Y., Wang, L. F., Dong, Y. C., and Zhang, X. Y. (2011). Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Sul, J. H., Martin, L. S., and Eskin, E. (2018). Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* 14:e1007309. doi: 10.1371/journal.pgen.1007309

Sun, X. Y., Du, Z. M., Ren, J., Amombo, E., Hu, T., and Fu, J. M. (2015). Association of SSR markers with functional traits from heat stress in diverse tall fescue accessions. *BMC Plant Biol.* 15:116. doi: 10.1186/s12870-015-0494-5

Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T. W., Gaudinier, A., et al. (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575. doi: 10.1038/nature14099

Thudi, M., Upadhyaya, H. D., Rathore, A., Mal Gaur, P., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS One* 9:e96758. doi: 10.1371/journal.pone.0096758

Thumma, B. R., Nolan, M. F., Evans, R., and Moran, G. F. (2005). Polymorphisms in *Cinnamoyl CoA reductase* (*CCR*) are associated with variation in microfibril angle in *eucalyptus* spp. *Genetics* 171, 1257–1265. doi: 10.1534/genetics.105.042028

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

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4, 535–538. doi: 10.1111/j.1471-8286.2004.00684.x

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

Wang, X. W., Fan, H. M., Li, Y. Y., Sun, X., Sun, X. Z., Wang, W. L., et al. (2014). Analysis of genetic relationships in tree peony of different colors using conserved DNA-derived polymorphism markers. *Sci. Hortic.* 175, 68–73. doi: 10.1016/j.scienta.2014.05.026

Wang, X. Q., Li, C. H., Contreras, M. D. M., Verardo, V., Gómez-Caravaca, A. M., and Xing, C. (2020). Integrated profiling of fatty acids, sterols and phenolic compounds in tree and herbaceous peony seed oils: marker screening for new resources of vegetable oil. *Foods* 9:770. doi: 10.3390/foods9060770

Wu, J., Cai, C. F., Cheng, F. Y., Cui, H. L., and Zhou, H. (2014). Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol. Breed.* 34, 1853–1866. doi: 10.1007/s11032-014-0144-x

Wu, J., Cheng, F. Y., Cai, C. F., Zhong, Y., and Jie, X. (2017). Association mapping for floral traits in cultivated *Paeonia rockii* based on SSR markers. *Mol. Gen. Genomics.* 292, 187–200. doi: 10.1007/s00438-016-1266-0

You, X. H., Li, W., Tao, Q. C., Sun, X. F., and Tang, K. X. (2011). *At1g65030*, a WD40-repeat protein gene, regulates seed mass and size in *Arabidopsis*. *Plant Physiol. J.* 47, 715–725. doi: 10.13592/j.cnki.ppj.2011.07.013

Yu, H. P., Cheng, F. Y., Zhong, Y., Cai, C. F., Wu, J., and Cui, H. L. (2013). Development of simple sequence repeat (SSR) markers from *Paeonia ostii* to study the genetic relationships among tree peonies (Paeoniaceae). *Sci. Hortic.* 164, 58–64. doi: 10.1016/j.scienta.2013.06.043

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

Yuan, J. H., Cheng, F. Y., and Zhou, S. L. (2012). Genetic structure of the tree peony (*Paeonia rockii*) and the Qinling Mountains as a geographic barrier driving the fragmentation of a large population. *PLoS One* 7:e34955. doi: 10.1371/journal.pone.0034955

Yuan, J. H., Cornille, A., Giraud, T., Cheng, F. Y., and Hu, Y. H. (2014). Independent domestications of cultivated tree peonies from different wild peony species. *Mol. Ecol.* 23, 82–95. doi: 10.1111/mec.12567

Zhai, Z. Y., Keereetaweep, J., Liu, H., Feil, R., Lunn, J. E., and Shanklin, J. (2021). Expression of a bacterial trehalose-6-phosphate synthase otsA increases oil accumulation in plant seeds and vegetative tissues. *Front. Plant Sci.* 12:656962. doi: 10.3389/fpls.2021.656962

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

Zhang, Q. Y., Niu, L. X., Yu, R., Zhang, X. X., Bai, Z. Z., Duan, K., et al. (2017a). Cloning, characterization, and expression analysis of a gene encoding a putative lysophosphatidic acid acyltransferase from seeds of *Paeonia rockii*. *Appl. Biochem. Biotechnol.* 182, 721–741. doi: 10.1007/s12010-016-2357-4

Zhang, J. J., Shu, Q. Y., Liu, Z. A., Ren, H. X., Wang, L. S., and De Keyser, E. (2012). Two EST-derived marker systems for cultivar identification in tree peony. *Plant Cell Rep.* 31, 299–310. doi: 10.1007/s00299-011-1164-1

Zhang, Q. Y., Yu, R., Sun, D. Y., Bai, Z. Z., Li, H., Xue, L., et al. (2017b). PrLPAAT4, a putative lysophosphatidic acid acyltransferase from *Paeonia rockii*, plays an important role in seed fatty acid biosynthesis. *Molecules* 22:1694. doi: 10.3390/molecules22101694

Zhou, H., Cheng, F. Y., Wu, J., and He, C. Y. (2015). Isolation and functional analysis of flowering locus T in tree peonies (PsFT). *J. Am. Soc. Hortic. Sci.* 140, 265–271. doi: 10.21273/JASHS.140.3.265

Zhou, X., Jiang, Y. J., and Yu, D. Q. (2011). WRKY22 transcription factor mediates dark-induced leaf senescence in *Arabidopsis*. *Mol. Cells* 31, 303–313. doi: 10.1007/s10059-011-0047-1

Zhou, S. L., Zou, X. H., Zhou, Z. Q., Liu, J., Xu, C., Yu, J., et al. (2014). Multiple species of wild tree peonies gave rise to the 'king of flowers', *Paeonia suffruticosa* Andrews. *Proc. R. Soc. B* 281:20141687. doi: 10.1098/rspb.2014.1687

Conflict of Interest: XC and XT are employees of Beijing Guose Peony Technology Co. Ltd., Beijing, China.

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

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

Copyright © 2021 Guo, He, Cheng, Zhong, Cheng and Tao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fgene.2021.732432

Comparative Analysis of Genomic and Transcriptome Sequences Reveals Divergent Patterns of Codon Bias in Wheat and Its Ancestor Species

Chenkang Yang[1†], Qi Zhao[1†], Ying Wang[1], Jiajia Zhao[2], Ling Qiao[2], Bangbang Wu[2], Suxian Yan[2], Jun Zheng[1,2] and Xingwei Zheng[1,2*]*

[1]School of Life Science, Shanxi University, Taiyuan, China, [2]State Key Laboratory of Sustainable Dryland Agriculture, Institute of Wheat Research, Shanxi Agricultural University, Linfen, China

The synonymous codons usage shows a characteristic pattern of preference in each organism. This codon usage bias is thought to have evolved for efficient protein synthesis. Synonymous codon usage was studied in genes of the hexaploid wheat *Triticum aestivum* (AABBDD) and its progenitor species, *Triticum urartu* (AA), *Aegilops tauschii* (DD), and *Triticum turgidum* (AABB). *Triticum aestivum* exhibited stronger usage bias for G/C-ending codons than did the three progenitor species, and this bias was especially higher compared to *T. turgidum* and *Ae. tauschii*. High GC content is a primary factor influencing codon usage in *T. aestivum*. Neutrality analysis showed a significant positive correlation ($p < 0.001$) between GC12 and GC3 in the four species with regression line slopes near zero (0.16–0.20), suggesting that the effect of mutation on codon usage was only 16–20%. The GC3s values of genes were associated with gene length and distribution density within chromosomes. tRNA abundance data indicated that codon preference corresponded to the relative abundance of isoaccepting tRNAs in the four species. Both mutation and selection have affected synonymous codon usage in hexaploid wheat and its progenitor species. GO enrichment showed that GC biased genes were commonly enriched in physiological processes such as photosynthesis and response to acid chemical. In some certain gene families with important functions, the codon usage of small parts of genes has changed during the evolution process of *T. aestivum*.

Keywords: codon usage, *Triticum aestivum*, GC content, tRNA, gene function

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Yinping Li,
Institute of Pomology, Chinese Academy of Agricultural Sciences (CAAS), China
Panfeng Guan,
Zhengzhou University, China
Wang Yuquan,
Henan Institute of Science and Technology, China

***Correspondence:**
Xingwei Zheng
smilezxw@126.com
Jun Zheng
sxnkyzj@126.com

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

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

Received: *29 June 2021*
Accepted: *29 July 2021*
Published: *20 August 2021*

Citation:
Yang CK, Zhao Q, Wang Y, Zhao JJ, Qiao L, Wu BB, Yan SX, Zheng J and Zheng XW (2021) Comparative Analysis of Genomic and Transcriptome Sequences Reveals Divergent Patterns of Codon Bias in Wheat and Its Ancestor Species. Front. Genet. 12:732432. doi: 10.3389/fgene.2021.732432

INTRODUCTION

In protein synthesis, triplet codons in the mRNA are translated into amino acids to form polypeptide chains. From 2 to 6 synonymous codons can be assigned to a specific amino acid, except for methionine and tryptophan, which are encoded only by AUG and UGG, respectively. Such synonymous codons are not used equally but are instead biased to some optimal codon (Sirihongthong et al., 2019). Codon usage bias, also known as codon bias, can

vary among species and even within genes of the same organism (Mukhopadhyay et al., 2007; Bali and Bebok, 2015). Codon usage bias reflects a mutation-selection balance, which can be affected by mutation, translational selection, and genetic drift in a population (Shah and Gilchrist, 2011; Duan et al., 2021). Therefore, understanding the codon usage bias can reveal the effects of long-term evolution on plant genomes.

Long-term studies account of the non-random variation in codon usage have shown that it is not simply a neutral process driven by mutational bias and genetic drift. There is substantial evidence that codon usage bias in some organisms is under selection pressure for translational efficiency (Zalucki et al., 2007; Szövényi et al., 2017; LaBella et al., 2019). Theories suggested that the neutralist model and selective model are not mutually exclusive and that codon usage might reflect a balance between mutational and selection effects (Duret and Mouchiroud, 1999). The "selection-mutation-drift" model was proposed to describe this balance. This model proposes that codon usage bias in genes with high expression is primarily the result of selection at the translational level and that in genes with low expression the bias is mainly caused by either mutation or genetic drift with the selective effect being relatively weak (Sharp and Li, 1987; Bulmer, 1991; Mukhopadhyay et al., 2007). Codon usage bias in unicellular organisms, such as *Escherichia coli* and *Saccharyomyces cerevisiae*, is adequately described by the selection-mutation-drift model. In contrast, codon usage bias is more complex in multicellular eukaryotic organisms. Lavner and Kotlar (2005) found codon bias tends to be high in some weakly expressed genes in mammalian eukaryotes. In such genes, non-optimal codons may have been selected to reduce the expression by reducing the elongation rate during gene translation. In mammals and plants, GC-biased gene conversion (gBGC) has also been proposed as a main driving force in the evolution of base composition, especially in grasses such as the genus *Oryza* and in some non-grass monocot species (Muyle et al., 2011; Clément et al., 2014; Mazumdar et al., 2017).

Codon usage bias in evolution research has been reported on model organisms, including *E. coli* (Krisko et al., 2014), *Dengue virus* (Manokaran et al., 2019), *S. cerevisiae*, *Drosophila melanogaster* (Kliman et al., 2003), and humans (Sauna and Kimchi-Sarfaty, 2011), as well as studies assessing codon usage in plant evolution research (Rensing et al., 2005; Liu et al., 2010; Sahoo et al., 2019; Chi et al., 2020). Plant species are broadly diverse in their gene expression, physiology, and stress response in varied environments. Therefore, the knowledge of codon usage and codon-pair context patterns in plants and the underlying evolutionary forces will advance understanding of the molecular mechanisms of environmental adaptation and biological diversity. Shen et al. (2020) analyzed codon usage patterns in eight citrus species based on coding sequence data, drawing the conclusion that few differences in codon features among citrus species and the genomes of citrus species were conserved. Researches showed that *Arabidopsis* genes show relatively little variation in codon usage, whereas an extreme degree of heterogeneity in codon usage patterns within the rice genome, which is highly correlated with differences in GC content between the genes (Wang and Hickey, 2007). Paralogs with high and low GC contents in rice and other cereals were analyzed and provided evidence for selectively driven codon usage, which represented a potential evolutionary process for the origin of genes with a high GC content in *Gramineae* (Guo et al., 2007).

With three closely related A, B, and D genomes, hexaploid wheat (*Triticum aestivum* L.) is the most widely cultivated crop on earth and is a model for allopolyploidy in plants (Zhang et al., 2020). The progenitor species of the A genome is diploid wild einkorn wheat, *Triticum urartu* (AA; Ling et al., 2018). The ancestor of the D genome is the wild diploid grass *Aegilops tauschii* (DD; Jia et al., 2013), which spontaneously hybridized with cultivated tetraploid wheat *Triticum turgidum* (AABB; Avni et al., 2017) resulting in hexaploid wheat (AABBDD; Appels et al., 2018). Genome sequencing and assembly of wheat and its genome donor species have been finished. The aim of the present study is to assess which evolutionary forces have a significant impact on codon usage bias in hexaploid wheat and its genomic donors.

MATERIALS AND METHODS

Sequence Data

The CDS genomic sequences of *T. urartu* accession G1812, wild tetraploid wheat (*T. turgidum*) accession "Zavitan," *Ae. tauschii* accession AL8/78, and common wheat cultivar "Chinese Spring" were downloaded from the Ensembl Plants.[1] The CDS sequences were filtered using Python and the following criteria: (1) the gene is a complete CDS, (2) the length of the CDS sequence is more than 300 bp, and (3) the CDS begins with ATG and ends with a termination codon (Liu et al., 2020). The leaf transcriptome data in common wheat and its ancestors were collected from the NCBI[2] under accessions SRR3274670 (*T. urartu*), SRR11292316 (*T. turgidum*), SRR11292312 (*Ae. tauschii*), and SRR9647023 (*T. aestivum*). In addition, the sequences of genes for grain size (GS; Li and Yang, 2017), the CCT gene family (Zheng et al., 2017), and the prolamin genes (Ling et al., 2018) identified from previous studies were used. The collected fasta sequences of genes from different families were aligned to the local protein database of other species using BALSTX, which was downloaded from URGI database[3] with the expected value (*E*-value) of the lowest. Meantime, the conserved domains of the gene family were identified using HMMER 3.0.

Codon Bias Analysis

The codon usage of the *T. aestivum* genome and its ancestor species were analyzed using the software program CodonW 1.4.2[4] (Peden, 2000). The parameters for the codon usage calculation were listed as following: Relative Synonymous Codon

[1]http://plants.ensembl.org/index.html
[2]https://www.ncbi.nlm.nih.gov/
[3]https://urgi.versailles.inra.fr/download/iwgsc/
[4]http://codonw.sourceforge.net/

Usage (RSCU), Effective Number of Codon (ENC), GC content (GC), codon first and second and the third position for GC content (GC1, GC2, and GC3), the average GC content of the first and second codons (GC12) and the content of each base in the third position of the synonymous codon (A3s, T3s, G3s, and C3s).

(1) RSCU:

Among the synonymous codons, the codon with the larger RSCU value has a higher usage probability (Wright, 1990). RSCU values were estimated using the following formula:

$$RSCU = \frac{x_{ij}}{\sum_j^{n_i} x_{ij}}$$

(2) ENC:

A value below 35 indicates high codon usage bias and values above 50 indicates low bias (Sharp and Li, 1987). ENC were calculated by the following formula:

$$ENC_{\exp} = 2 + s + 29 \Big/ \left(s^2 + (1-s)^2\right)$$

Effective Number of Codon-plot and neutrality plot analysis were used to analyze factors affecting the codon preference of *T. aestivum* and its ancestor species. In ENC-plot analysis, when ENC value is within the allowable range of the standard curve, the codon preference is caused by base mutation, otherwise the codon preference is caused by natural selection (Cai et al., 2006).

Neutral plot analysis was used to investigate the effect of mutational and selection pressure on the codon usage pattern. The neutrality plot was mapped with the scatterplot to analyze the correlation between GC12 and GC3. When the regression coefficient was close to 0, the correlation was low, indicating that the GC composition for the three positions of the codon differed, the GC content is highly conserved, and the codon preference for the species is mainly affected by natural selection (Su et al., 2009; Nasrullah et al., 2015). Otherwise, the codon bias of species is more affected by mutations.

PR2-plot was used to analyze whether the third base of a codon is biased according to the bias rule (Sueoka, 2001). The bias rule states that if A = T and G = C between the two complementary strands of DNA then there is no mutation and selection bias. If otherwise, then the codon is biased.

Correspondence Analysis and Determination of Codon

Correspondence analysis (COA) was generated to analyze codon usage changes in different genes. Correlation analysis is used to determine the main factors affecting codon bias (Jia et al., 2015). In COA, all the genes were distributed in multi-dimensional vector spaces using the RSCU COA function of codonW. The first axis represented the most different codon usage changes with sequential decreases represented in the second, third, and fourth axes.

To determine optimal codons of *T. aestivum* genome and its ancestor species, genes with the 5% highest and lowest ENC values were screened to establish a low preference group and a high preference group. Among genes from two groups, the RSUC difference was greater than 0.08, and the codon with the RSUC value greater than 1 was determined to be the optimal codon (Liu et al., 2020).

Major Factors of Variations in Synonymous Codon Usages Analysis

To determine the factors affecting the codon bias of the *T. aestivum* genome and its ancestor species, we analyzed the number of tRNA anti-codons and the length of genes. Firstly, CDS sequences of common wheat and its ancestor species were divided into four length categories: <1,000 bp, 1,000–2,000 bp, 2,000–3,000 bp, and >3,000 bp. Secondly, codon bias analysis for genes of different length was done using CodonW 1.4.2 and the numbers of genes and values of GC3s were calculated. Anti-codons of common wheat and its ancestor species were identified using tRNAscan-SE 2.0 (Chan and Lowe, 2019).

Gene Function

The rawdata of common wheat and its ancestors were used for quality control using fastp (Chen et al., 2018). To analyze the relationship between codon preference and function of expressed genes, the TPM of CDS sequences from common wheat and its ancestors were calculated using Salmon v1.3.0 (Patro et al., 2017). To determine differences in the function of genes with different codon bias, gene ontogeny (GO) enrichment were conducted separately using genes with different codon bias. For gene annotation, coding sequence regions of each gene in wheat and its ancestor species reference genomes were compared against the default database using eggNOG-mapper (Huerta-Cepas et al., 2017, 2019). The CDS sequences of the TPM > 1 of all species were divided into high GC bias genes and high AT bias genes according to GC content of each species. The CDS sequences with different codon bias were calculated *via* GO enrichment analysis using the clusterProfiler R package (Yu et al., 2012).

RESULTS

Effective Number of Codons Value and GC Contents

The codon usage characteristics of the *T. aestivum* genome and those of its ancestor species, *T. urartu*, *T. turgidum*, and *Ae. tauschii*, were investigated using CodonW1.4.2. *Triticum aestivum* showed a stronger codon usage bias (ENC 48.6) compared to its ancestor species *T. turgidum* (ENC 51.6) and *Ae. tauschii* (ENC 51.2; **Supplementary Table 1**). ENC values of the coding genes in the A and B genomes were compared. Among the 57,979 coding genes from the A genome of *T. turgidum*, 3,874 sequences (6.68%) showed high codon usage bias with ENC values less than 35. In genes from the A genome of *T. aestivum*, there were 5,606 sequences (13.66%) with ENC values less than 35. Sequences with ENC values less than 35 accounted for 6.10% of the *T. turgidum* B genome and 12.53% of the *T. aestivum* B genome. Thus, both the A and B genome from *T. aestivum*

showed more codon usage bias than the A and B genome from *T. turgidum*.

GC3s frequencies in *T. aestivum* (63.7%) and *T. urartu* (62.0%) were higher than those in *T. turgidum* (52.6%) and *Ae. tauschii* (54.6%; **Figure 1**; **Supplementary Table 1**). Codon usage bias has been shown to correlate with GC composition (Wan et al., 2004). The mean GC contents were examined for the whole coding sequences of the four species and the higher GC content partly explained the bias for G/C-ending codons seen in *T. aestivum* and *T. urartu*. PR2-bias plot analysis showed a non-proportional use of A/T and C/G in the four genomes, with T3>A3 and C3>G3 (**Supplementary Figure 1**). Neutrality plot results showed that the correlation between GC12 and GC3 was statistically significant ($p < 0.001$) in *T. aestivum* ($r = 0.650$), *Ae. tauschii* ($r = 0.596$), *T. turgidum* ($r = 0.551$), and *T. urartu* ($r = 0.517$). The slopes of the regression line were close to zero (0.16–0.20, **Supplementary Figure 2**), indicating that selective pressure should be the dominant factor shaping codon usage pattern in *T. aestivum* and its ancestor species.

Correspondence Analysis

In the present study, COA based on RSCU was expected to further identify the major factors affecting codon usage frequencies and synonymous codon preferences observed in the coding sequences of the four genomes. Genes from the four species were widely distributed along the two explanatory axes, suggesting a strong codon usage bias in these coding sequences (**Figure 2**). The first and second explanatory axis accounted for 40.2/4.3, 39.1/3.6, 47.9/4.1, and 40.6%/3.7% in *T. urartu*, *T. turgidum*, *T. aestivum*, and *Ae. tauschii*, respectively. Thus, the first axis reflects the primary factor that explains the differences in codon usage among all genes and genes with high GC and high AT were separated along this axis. The corresponding distribution of synonymous codons shows the separation of C/G-ending codons (except AGG that encodes arginine) and A/U-ending codons along this same axis. High G/C genes from *T. urartu* and *T. turgidum* were distributed on the right side of axis 1. The separation of genes on the second axis appears to be largely due to frequency differences in C-ending and G-ending codons among the GC rich genes; however, genes with more A-ending and T-ending codons tended to cluster at the center of axis 2. Correlation analysis showed that the third base preference of coding genes positioned on axis 1 of the COA were significantly related (**Supplementary Table 2**), indicating that base composition is the main factor affecting codon usage and the COA could be further used in optimal codon identification.

The optimal codons of common wheat and its ancestor species were identified based on the RSCU value (**Supplementary Table 3**) and the number of optimal codons in *T. urartu*, *T. turgidum*, *T. aestivum*, and *Ae. tauschii* were 21, 14, 22, and 16, respectively. All the optimal codons showed G/C at the third codon position, indicating that wheat and its ancestor species have similar optimal codon preferences for codons with G/C-endings instead of A/U-endings.

Variations in Synonymous Codon Usages in *Triticum* Genomes

To investigate the effect of coding sequences length on codon usage, coding sequences were classified into four groups: <1,000 bp, 1,000–2,000 bp, 2,000–3,000 bp, and >3,000 bp, and the average GC3s values computed for genes from each species (**Figure 3**). Bias for G/C-ending synonymous codon usage was much stronger in genes coding for short vs. long proteins (**Supplementary Table 4**). Among coding sequences longer than 1,000 bp, more GC-ending synonymous codons were observed in *T. aestivum* than in its ancestor species. However, among sequences <1,000 bp, *T. urartu* sequences showed the

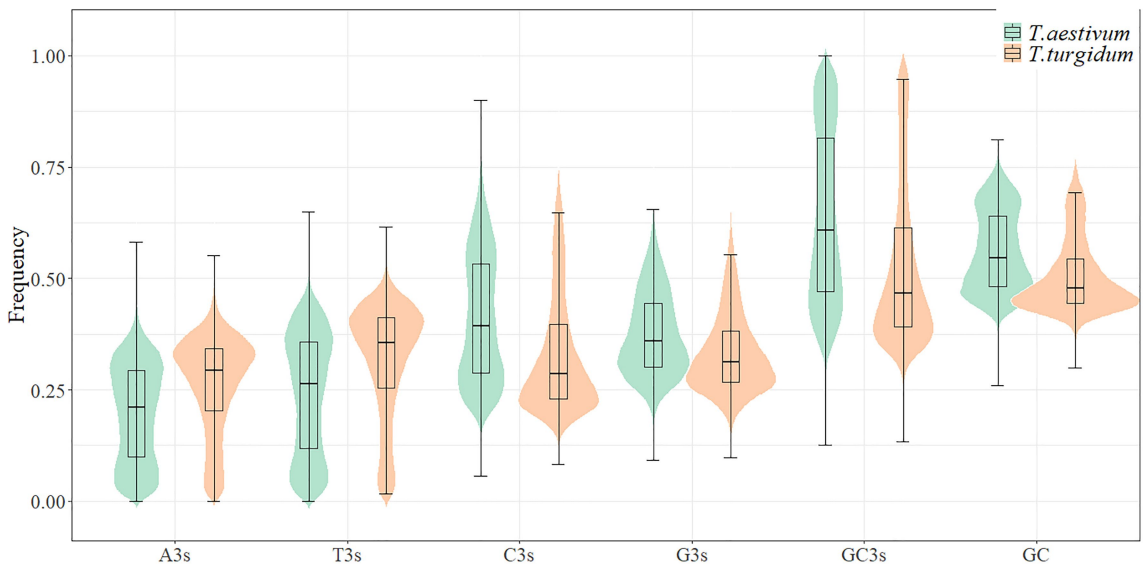


FIGURE 1 | Comparison of base composition between *Triticum aestivum* and *Triticum turgidum*.

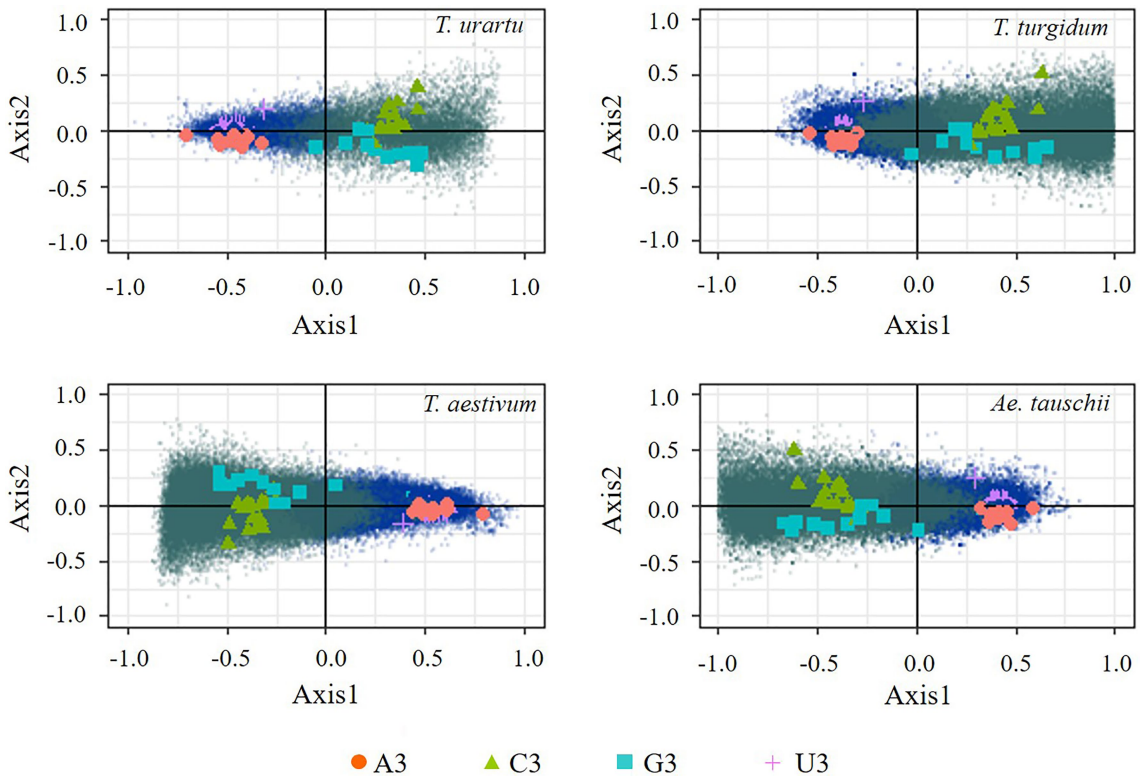


FIGURE 2 | Correspondence analysis based on Relative Synonymous Codon Usage (RSCU) of genes and codons. Blue dots represent genes of AT preference and green dots represent genes of GC preference.

greatest bias for GC-ending codons. Thus, in longer genes *T. aestivum* showed the greatest tendency to use G/C at the third codon position.

In humans, GC distribution and codon preference have been proved to be related to gene density (Versteeg et al., 2003). In the present study, substantially higher gene density occurred in the distal regions of chromosome arms and genes distributed in these regions exhibited higher GC3s values. In contrast, the centromere regions showed lower gene density and the GC3s values of genes in these regions were low (**Figure 4**).

Corresponding tRNA Abundance

Biased usage of synonymous codons can also be constrained by the amount of isoacceptor tRNA to increase translation efficiency (Duret and Mouchiroud, 1999). The frequency of each synonymous codon encoding 21 amino acids was estimated, and we further investigated the amount of corresponding tRNA in the four species (**Supplementary Table 5**). Significant differences were found among the species for codon frequency and corresponding tRNA abundance. For example, *T. urartu* and *T. aestivum* showed high correlations ($r = 0.667$ and 0.724, $p < 0.01$) for the amino acids with two synonymous codons (Tyr, Phe, Lys, His, Glu, Gln, Cys, Asp, and Asn; **Figures 5A,B**). In contrast, amino acid with more than four synonymous codons (Val, Thr, Ser, Pro, Leu, Gly, Arg, and Ala), *T. turgidum* and *Ae. tauschii* showed high correlations ($r = 0.511$ and 0.488, $p < 0.01$; **Figures 5C,D**).

Gene Ontogeny Enrichment of GC/AT Biased Genes

Based on GO categories, the functional enrichment of codon usage biased genes from the four species was analyzed, and the top 10 GO terms with values of $p < 0.05$ for genes from each species were identified (**Figure 6**). G/C-ending genes performed different functions compared to A/T-ending genes. The GC3s biased genes from *T. aestivum* and its ancestor species were commonly enriched in genes related to photosynthesis (**Figure 6A**). Also enriched were transmembrane transport (72/625), and response to acid chemical (71/625) in *T. urartu*, photosynthesis (46/802) and photosynthesis light reaction (43/802) in *T. turgidum*, response to acid chemical (158/1,214) and response to inorganic substance (113/1,214) in *T. aestivum*, and photosynthesis (46/751) and photosynthesis light reaction (40/751) in *Ae. tauschii* (**Supplementary Table 6**). AT3s biased genes likely perform similar functions in *T. aestivum* and *T. urartu* (**Figure 6B**) as these genes were enriched in GO terms such as cellular location, macromolecule localization, and protein location. A/T ending genes from *Ae. tauschii* were mainly enriched in response to extracellular stimulus (50/836) and response to nutrient levels (47/836; **Supplementary Table 7**). The GO enrichment (**Supplementary Tables 8** and **9**) demonstrated that the evolutionary features reliably reflect the relative function of different codon biased genes.



FIGURE 3 | GC3s values in coding sequences from *Triticum urartu*, *Triticum turgidum*, *Triticum aestivum*, and *Aegilops tauschii*. *N* is the number of coding sequences (CDS).

Codon Preference in Important Functional Gene Families

The present study shows that synonymous codons had bias at the whole genome level in *T. aestivum* and its ancestor species. We next sought to investigate if codon usage preference changed during evolution for gene families with important functions, including grain size (GS) genes, the CCT gene family regulating flowering, and the quality related prolamin gene family. Based on 36 grain size related genes identified in previous studies in *Triticum* (Li and Yang, 2017), similar GS genes from the A genome of *T. aestivum*, *T. turgidum*, and *T. urartu* were isolated and the GC3s variation of GS genes during polyploidization was calculated (**Supplementary Table 10**). A difference in the GC3s value between the homologous genes of >0.1 was classified as being codon biased. *Triticum urartu* is the progenitor of the A subgenome of tetraploid *T. turgidum* and hexaploid *T. aestivum* (Ling et al., 2018). From *T. urartu* to *T. aestivum*, seven GS genes were identified with different preferences in codon usage patterns (**Table 1**). Five of the seven genes, *TRIUR3_05970-T1*, *TRIUR3_33526-T1*, *TRIUR3_34310-T1*, *TRIUR3_08952-T1*, and *TRIUR3_09477-T1*, showed codon usage bias from the time *T. urartu* evolved into *T. turgidum* (**Table 2**).

The RSCU of each codon was measured for CCT and prolamin genes to investigate the variation in codon usage bias of these genes across the four genomes. Consistent with the results at the genome level, the G/C-ending codons in most of the CCT family genes were overrepresented (RSCU > 1), except *CCT24* which showed obvious A/U bias in the third position of synonymous codons. Among the species, the codon usage in *T. aestivum* CCT genes was most like that for genes in the D genome ancestor *Ae. tauschii* (**Figure 7**). Euclidean clustering of RSCU for codons in each gene showed that among the 26 CCT genes in *T. aestivum*, 15 were like *Ae. tauschii*, eight were like *T. turgidum*, and the last three were like *T. urartu* (**Figure 8**). For the prolamin genes, some optimal codons identified at the genome level (**Supplementary Table 5**) were underrepresented (RSCU < 1) including GUG encoding Val, UAC encoding Try, CCG encoding Pro, UUC encoding Phe, CUC encoding Leu, CAG encoding Glu, GAC encoding Asp, and CGC encoding Arg (**Figure 9**).

DISCUSSION

Triticum aestivum has a close phylogenetic relationship with its ancestor species. However, due to the long-term evolutionary processes and the various environments, where the plants evolved, the codon usage patterns among the species differ. In the present study, we analyzed the codon bias patterns in *T. aestivum* and its progenitor species and found that *T. aestivum* has similar preference in codon usage with *T. urartu* and that it differs most from *T. turgidum*. This phenomenon is mainly explained by the

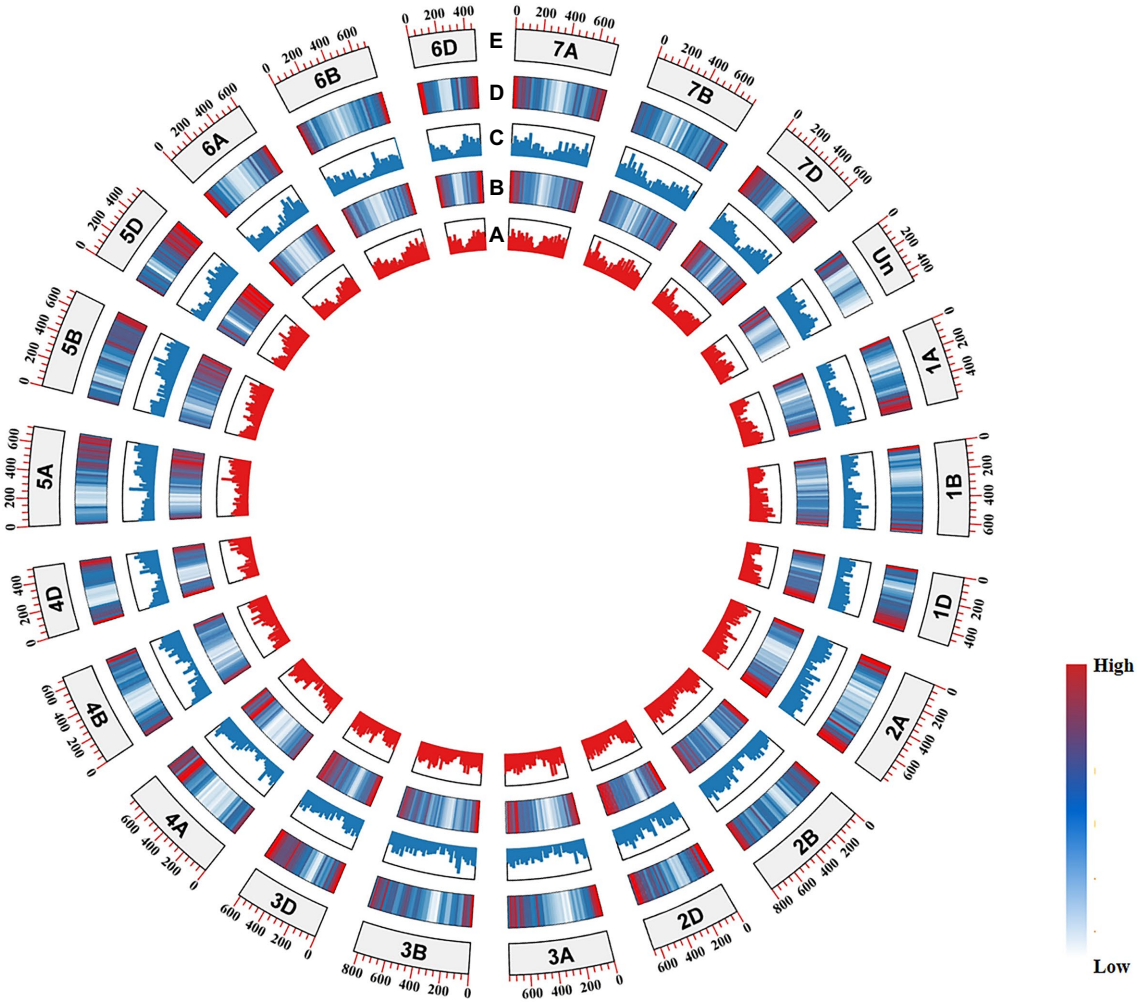


FIGURE 4 | CIRCOS visualization of various *Triticum aestivum* genome data. **(A)** Median GC3 values for genes from the (+) strand. **(B)** Gene density of the (+) strand. **(C)** Median CG3 values for genes from the (–) strand. **(D)** Gene density of the (–) strand. **(E)** Chromosome name and size.

similar genomic GC content in *T. aestivum* and *T. urartu* (**Supplementary Table 1**), which considered to be the strongest determinant of codon usage variation caused by mutational processes across species (Muyle et al., 2011; Plotkin and Kudla, 2011).

tRNA Abundance Supports the Role of Selection for Translational Efficiency

Codon bias is determined by a balance between mutation, drift, and selection for optimal translational efficiency and/or accuracy (De La Torre et al., 2015). In many cases, codon usage correlates with the cognate tRNA abundance with highly used codons generally having higher intracellular tRNA concentrations than low-use codons (Zalucki et al., 2009). This fact indicates that tRNA abundance may be the main selection pressure for synonymous codon usage (Mohanta et al., 2020). In the present study of *Triticum*, more codon-anticodon interaction occurred in amino acids with two synonymous codons in *T. aestivum* and *T. urartu* and in amino acids with four or six synonymous codons in *T. turgidum* and *Ae. tauschii*. This result also provides evidence for the similar codon usage bias in *T. aestivum* and *T. urartu*. In contrast, for codons encoding glutamine from *T. aestivum* and *T. urartu*, corresponding tRNA copy numbers for the optimal codon (CAG) were much less than for the non-optimal codon (CAA), suggesting that the primary factor influencing codon-usage for these amino acids is not translation efficiency (Kanaya et al., 2001). Simultaneously, we considered the influence of codon swinging on tRNA abundance and codon usage preference. Crick (1966) proposed the wobble hypothesis based on stereochemistry to explain this phenomenon: when the first anticodon is A or C, only one codon can be recognized; when the first anticodon is G or U, two codons can be recognized, where G can identify C and U, where U can identify A and G. Due to the wobble phenomenon, a tRNA anticodon can be combined with more than one mRNA codon. We conjecture that UUG oscillates when recognizing CAA, so that more UUG can recognize the optimal codon CAG. We also found the same phenomenon in Ser. Wobble rules say that tRNAs with a GCU anticodon can pair with AGU, which explains the apparent

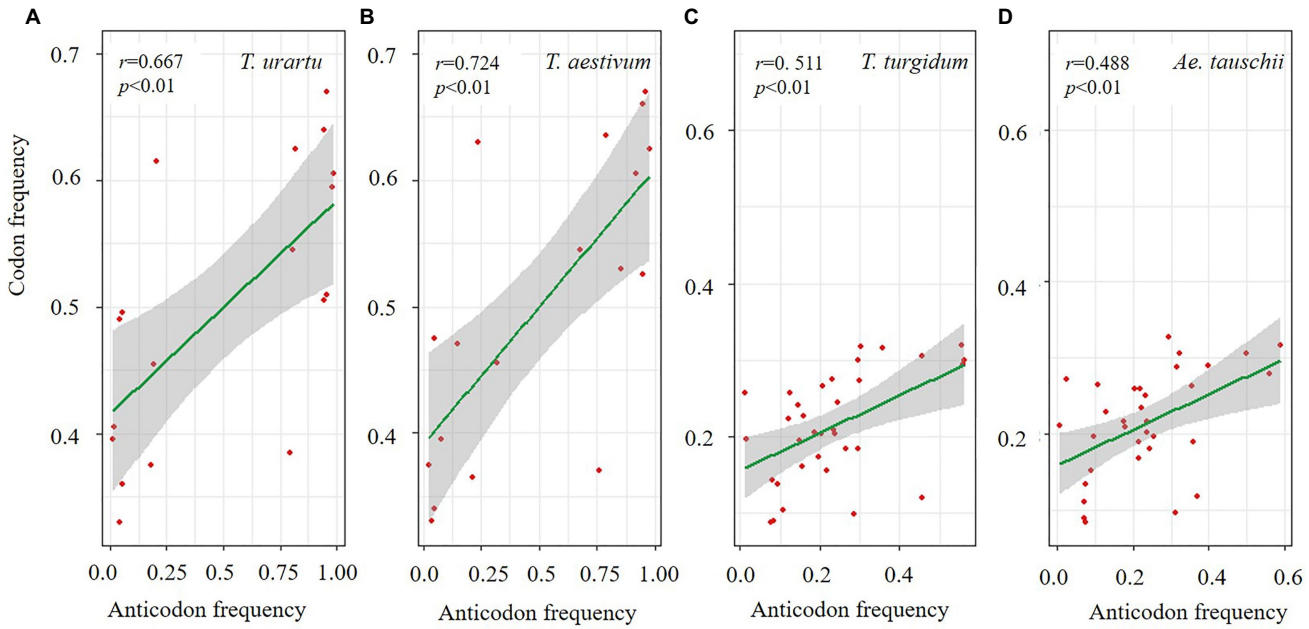


FIGURE 5 | Correlation for common wheat and its three progenitor species between codon and anticodon frequency for amino acids with two codons **(A,B)** and more than four codons **(C,D)**.

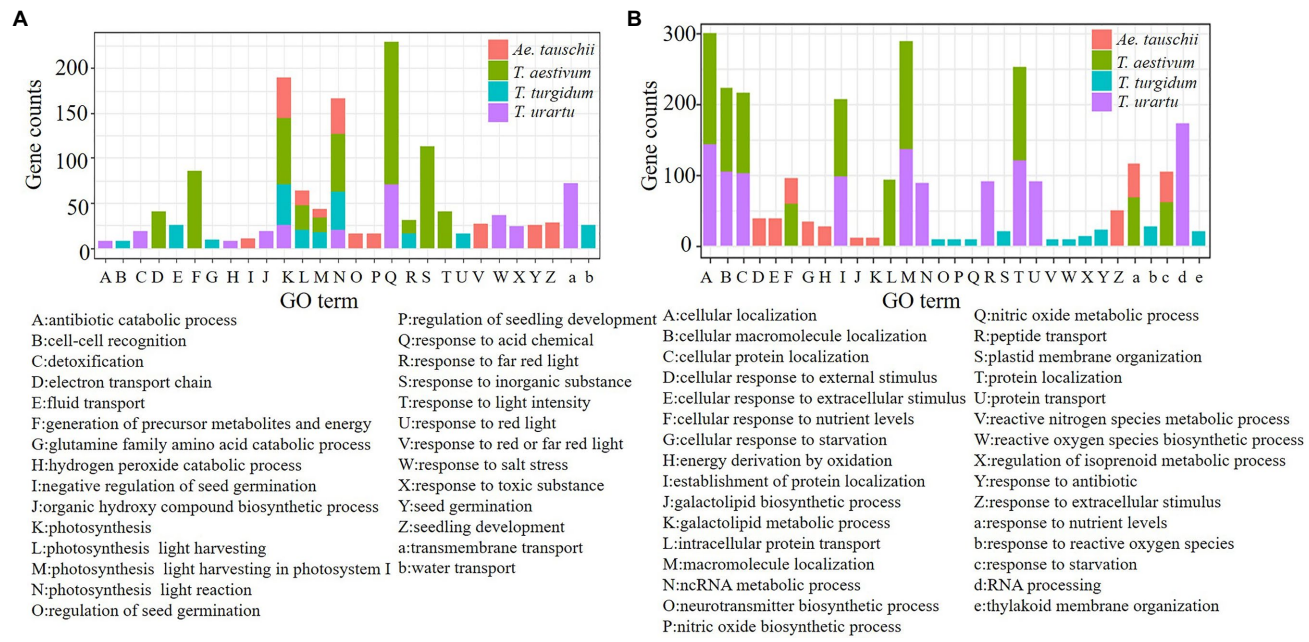


FIGURE 6 | GO enrichment of genes with GC3s **(A)** and AT3s **(B)** biased codons from four species.

discrepancy in the ability to translate AGU codons to Ser with extremely few tRNAs with ACU anticodons (**Supplementary Table 5**). In a recent research of 128 species of the plants (Mohanta et al., 2020), high and low frequencies of wobble base-pairing were occurred at the G:U base pairing to meet translational demand.

Triticum aestivum Showed More GC3 Codon Bias in Genes With Optimal Length

Duret and Mouchiroud (1999) found that codon usage bias is negatively correlated with protein length in *Caenorhabditis elegans*, *D. melanogaster*, and *Arabidopsis thaliana*. Gene length showed a negative correlation with GC3 content in non-grass monocot species

TABLE 1 | The GC3s variation of grain size genes during polyploidization with *Triticum aestivum* and *Triticum urartu*.

Triticum aestivum GS genes	*Triticum urartu* GS genes	GC3s of *Triticum aestivum*	GC3s of *Triticum urartu*	Difference of GC3s
TraesCS1A02G083000.1	*TRIUR3_14996-T1*	0.841	0.524	0.317
TraesCS4A02G012100.1	*TRIUR3_05970-T1*	0.975	0.793	0.182
TraesCS2A02G013200.1	*TRIUR3_33526-T1*	0.756	0.654	0.102
TraesCS5A02G215100.1	*TRIUR3_34310-T1*	0.766	0.953	−0.187
TraesCS7A02G120000.1	*TRIUR3_08952-T1*	0.603	0.340	0.263
TraesCS7A02G246500.1	*TRIUR3_09477-T1*	0.779	0.536	0.243
TraesCS4A02G430600.1	*TRIUR3_35245-T1*	0.963	0.566	0.397

TABLE 2 | The GC3s variation of grain size genes during polyploidization with *Triticum turgidum* and *Triticum urartu*.

Triticum turgidum GS genes	*Triticum urartu* GS genes	GC3s of *Triticum turgidum*	GC3s of *Triticum urartu*	Difference of GC3s
TRIDC4AG041610.1	*TRIUR3_05970-T1*	0.911	0.793	0.118
TRIDC6AG039380.2	*TRIUR3_33526-T1*	0.784	0.654	0.130
TRIDC5AG033880.1	*TRIUR3_34310-T1*	0.763	0.953	−0.190
TRIDC7BG002860.6	*TRIUR3_08952-T1*	0.512	0.340	0.172
TRIDC7AG069150.1	*TRIUR3_09477-T1*	0.911	0.536	0.375
TRIDC7AG067090.1	*TRIUR3_33769-T1*	0.910	0.441	0.469
TRIDC7AG065170.7	*TRIUR3_09129-T1*	0.830	0.655	0.175

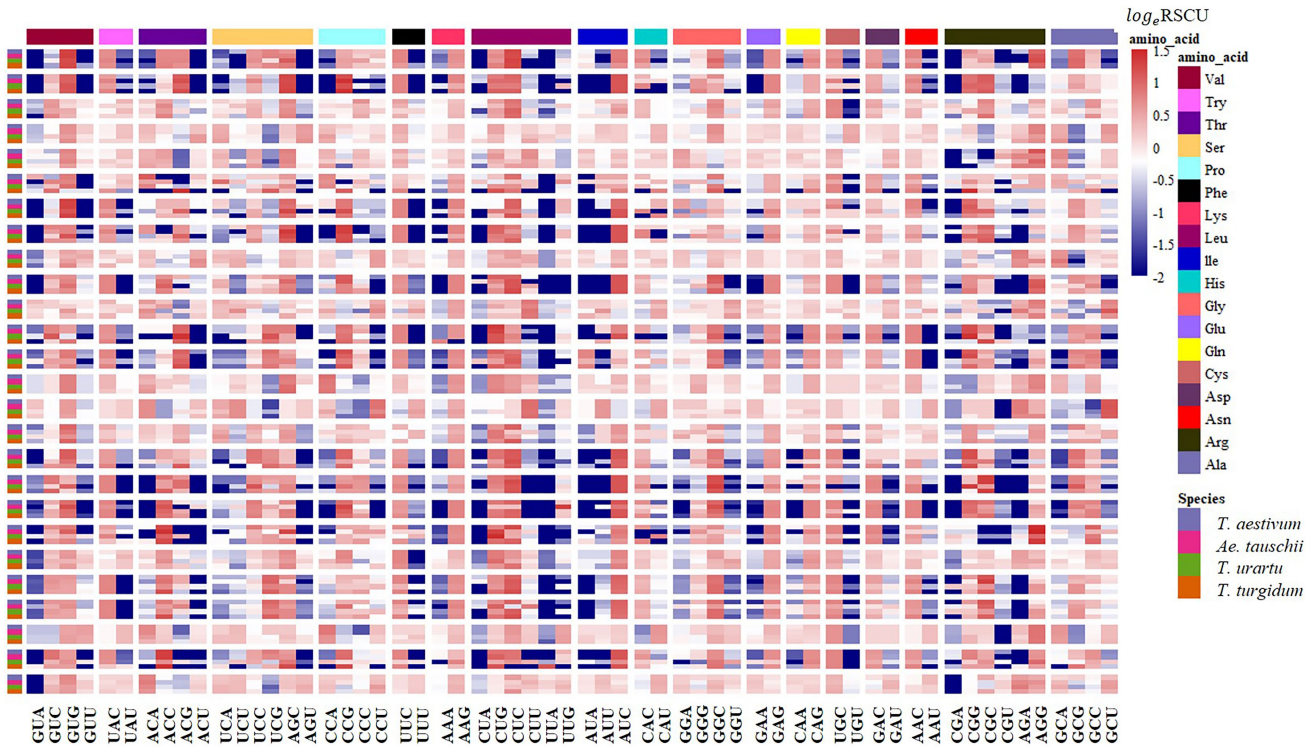


FIGURE 7 | RSCU analysis showing synonymous codon usage preference of the CCT gene families in common wheat and its ancestor species. Columns correspond to the 59 nondegenerate, non-stop codons. Rows correspond to special family genes in the four species. Blue cells indicate low bias codons and red cells indicate high bias codons.

(Mazumdar et al., 2017). The same phenomenon was found in the three progenitor species, as coding sequences shorter than 1,000 bp had the highest GC3 codon bias (**Figure 3**). In *T. aestivum*, though, synonymous codon bias increased among coding sequences of 1,000–2,000 bp compared to short sequences (<1,000 bp). In the *T. aestivum* genome, we found that most coding sequences were

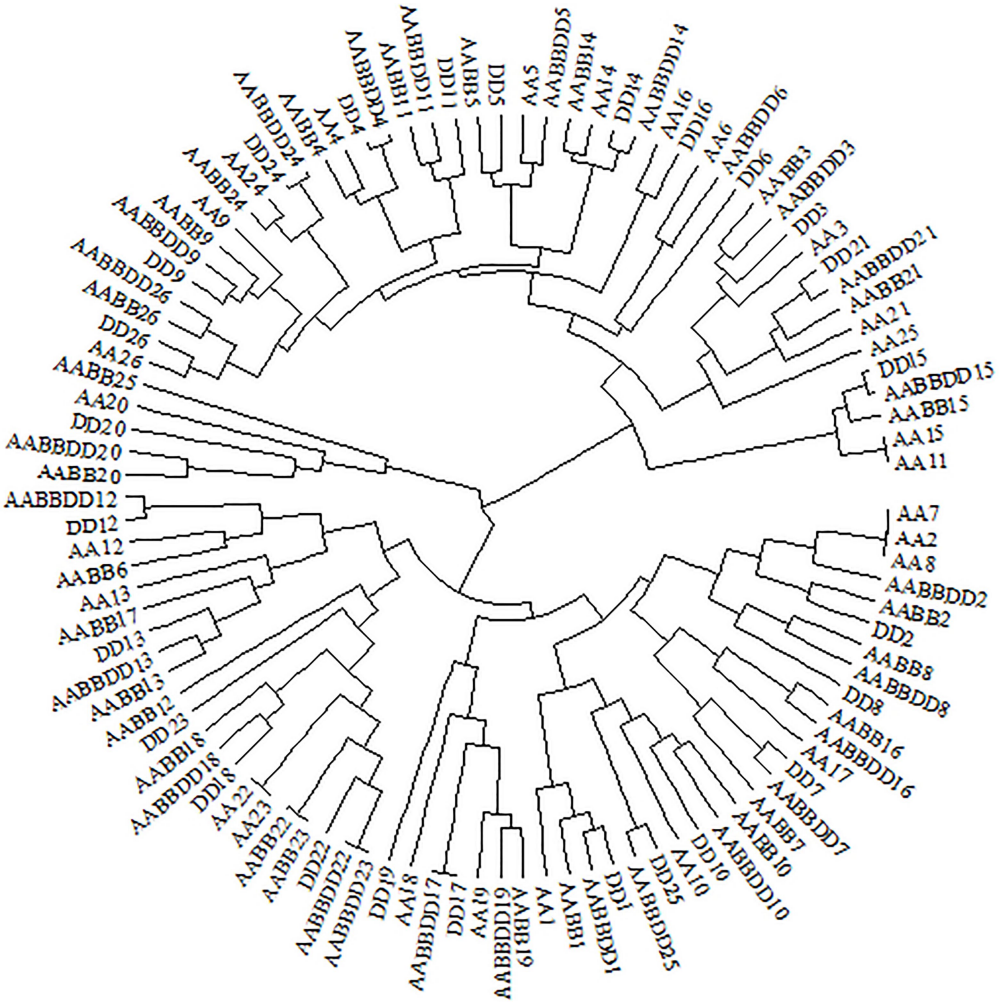


FIGURE 8 | Clustering tree of CCT genes from *Triticum aestivum* (AABBDD), *Aegilops tauschii* (DD), *Triticum turgidum* (AABB), and *Triticum urartu* (AA) according to RSUC values for 59 synonymous codons.

between 1,000 and 2,000 bp (N = 53, 051), which is consistent with the result for *T. aestivum* transcript sequences reported by Wang et al. (2019). All the optimal codons identified in the *Triticeae* in the present study preferred G/C-endings at the genome wide level, and the GC3 content of coding sequences shorter than 2,000 bp was higher than that for longer genes. This result supports the idea that in *T. aestivum* sequences of less than 2,000 bp are the optimal length for translation, especially for important functional genes, and that the use of optimal codons increases their translational efficiency and accuracy. The optimal gene length of *T. aestivum* may be longer than that of its ancestral species, and perhaps the process of evolving into a complex hexaploid species enabled *T. aestivum* to translate longer proteins with minimal mismatches.

Particular Gene Families Showed Different Codon Usage Preference Compared With the Whole Genome

Genes for particular physiological processes, such as photosynthesis, response to salt stress/acid chemical/cold stress, and seed germination, were biased to GC-ending codons. In contrast, A/U-ending biased genes tended to be associated with the function of basic biological processes, like RNA processing, protein localization, protein transport, and membrane organization (**Supplementary Tables 8** and **9**). Some important genes always maintained conserved codon usage bias. For example, we identified 1,755 disease resistance genes in *Ae. tauschii* (**Supplementary Table 11**; Jia et al., 2013) and compared the GC3s contents with the corresponding homoeologous genes in D genome of *T. aestivum*. Only 19 genes had a GC3 difference of more than 0.1, an exceptionally low level (**Supplementary Table 12**).

CCT family genes are important for the flowering process, which enables the adaptation for reproductive success in various geographical environments (Yang et al., 2013). RSCU values of 59 synonymous codons in the CCT gene family from the four species were investigated and most CCT genes from *T. aestivum* (15/26) were clustered together with genes from *Ae. tauschii*, indicating that CCT genes from *T. aestivum* and *Ae. tauschii* were similar in codon usage frequency. We further measured the GC3

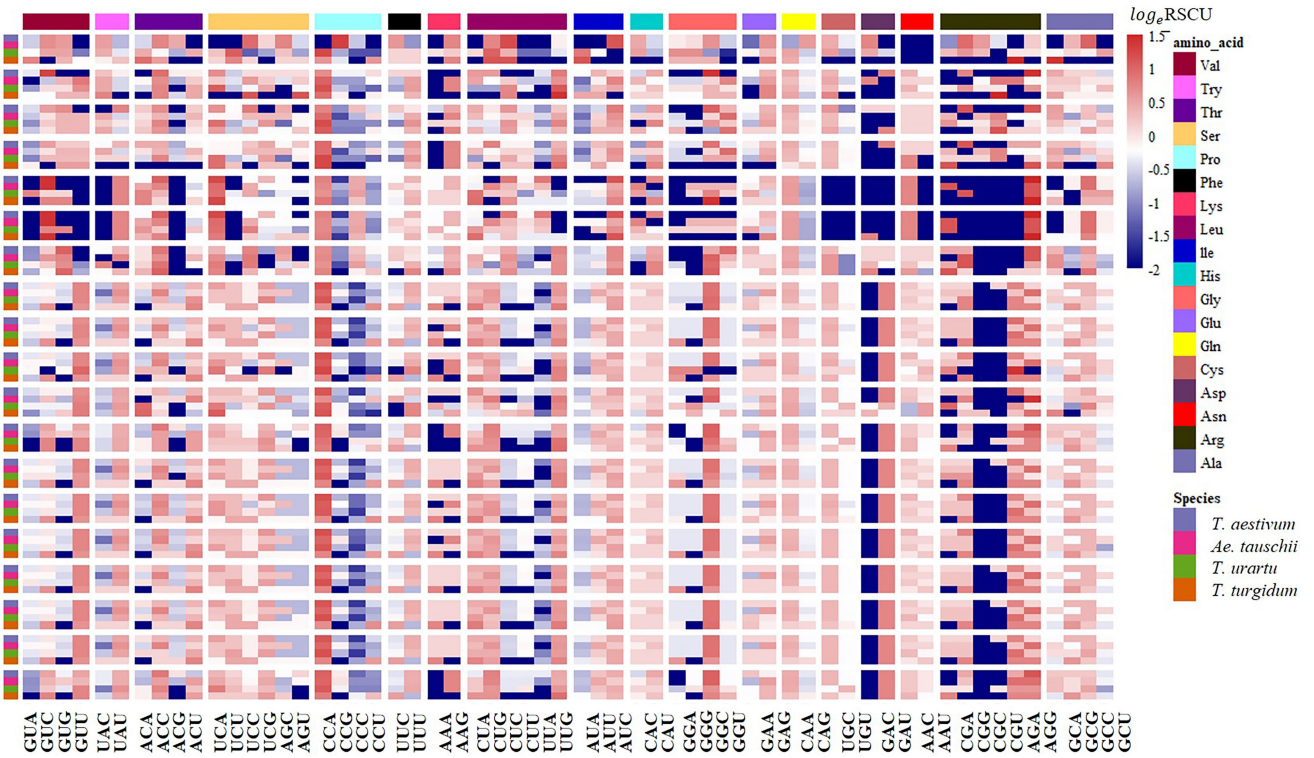


FIGURE 9 | RSCU analysis showing synonymous codon usage preference of the Prolamin gene families in common wheat and its ancestor species. Columns correspond to the 59 nondegenerate, non-stop codons. Rows correspond to special family genes in the four species. Blue cells indicate low bias codons and red cells indicate high bias codons.

values of CCT genes and found that *T. aestivum* and *Ae. tauschii* CCTs showed similar G/C preference at the third position of codons (**Supplementary Figure 3A**; **Supplementary Table 13**), which was consistent with the RSCU results. Previous studies showed that the rapid evolution of the CCT gene family promotes the adaptability of *Ae. tauschii* (Zheng et al., 2017), the similar codon usage bias implies that CCT genes from *T. aestivum* maintained the functions from its D progenitor, *Ae. tauschii*, presumably to enhance environment adaptability. The prolamin sequences of the four species are highly similar, but there are obvious differences in the usage of codons, especially the optimal codons (**Supplementary Figure 3B**; **Supplementary Table 14**). As a donor of the D genome of wheat, *Ae. tauschii* provided wheat with better adaptability and quality traits (Jia et al., 2013).

For codon optimization, not only must the codon preference and tRNA abundance be considered, but also other factors including local mRNA folding, codon coordination, and codon correlation. By both natural evolution and genetic engineering, these properties can be altered to gain additional transcriptional regulatory ability for appropriate gene expression under specific cell conditions.

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

CY collected datasets and performed the bioinformatics analysis. CY, QZ, and YW wrote original draft. XZ and JuZ conducted funding acquisition, project administration, and manuscript review and editing. JiZ and LQ plotted the figures. BW and SY prepared the supplementary data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported financially by the Research Program Sponsored by State Key Laboratory of Sustainable Dryland Agriculture (in preparation), Shanxi Agricultural University (nos. 202002-1 and 202105D121008-2-1), General Program of China Postdoctoral Science Foundation (2020M670701), and Agricultural Science Research of Shanxi Academy of Agricultural Sciences (YCX2020BH2 and YBSJJ2006).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.732432/full#supplementary-material

REFERENCES

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032

Bali, V., and Bebok, Z. (2015). Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *Int. J. Biochem. Cell B* 64, 58–74. doi: 10.1016/j.biocel.2015.03.011

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907. doi: 10.1093/genetics/129.3.897

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

Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14. doi: 10.1007/978-1-4939-9173-01

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Chi, X., Zhang, F., Dong, Q., and Chen, S. (2020). Insights into comparative genomics, codon usage bias, and phylogenetic relationship of species from *Biebersteiniaceae* and *Nitrariaceae* based on complete chloroplast genomes. *Plants* 9:1605. doi: 10.3390/plants9111605

Clément, Y., Fustier, M. A., Nabholz, B., and Glémin, S. (2014). The bimodal distribution of genic GC content is ancestral to monocot species. *Genome Biol. Evol.* 7, 336–348. doi: 10.1093/gbe/evu278

Crick, F. H. (1966). Codon--anticodon pairing: the wobble hypothesis. *J. Biochem. Mol. Biol.* 19, 548–555. doi: 10.1016/s0022-2836(66)80022-0

De La Torre, A. R., Lin, Y. C., Van de Peer, Y., and Ingvarsson, P. K. (2015). Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in picea gene families. *Genome Biol. Evol.* 7, 1002–1015. doi: 10.1093/gbe/evv044

Duan, H., Zhang, Q., Wang, C., Li, F., Tian, F., Lu, Y., et al. (2021). Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ* 9:e10787. doi: 10.7717/peerj.10787

Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4482–4487. doi: 10.1073/pnas.96.8.4482

Guo, X. Y., Bao, J. D., and Fan, L. J. (2007). Evidence of selectively driven codon usage in rice: implications for GC content evolution of *Gramineae* genes. *FEBS Lett.* 581, 1015–1021. doi: 10.1016/j.febslet.2007.01.088

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Jia, X., Liu, S. Y., Zheng, H., Li, B., Qi, Q., Wei, L., et al. (2015). Non-uniqueness of factors constraint on the codon usage in *Bombyx mori*. *BMC Genomics* 16:356. doi: 10.1186/s12864-015-1596-z

Jia, J. Z., Zhao, S. C., Kong, X. Y., Li, Y. R., Zhao, G. Y., He, W. M., et al. (2013). *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496, 91–95. doi: 10.1038/nature12028

Kanaya, S., Yamada, Y., Kinouchi, M., Kudo, Y., and Ikemura, T. (2001). Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J. Mol. Evol.* 53, 290–298. doi: 10.1007/s002390010219

Kliman, R. M., Irving, N., and Santiago, M. (2003). Selection conflicts, gene expression, and codon usage trends in yeast. *J. Mol. Evol.* 57, 98–109. doi: 10.1007/s00239-003-2459-9

Krisko, A., Copic, T., Gabaldón, T., Lehner, B., and Supek, F. (2014). Inferring gene function from evolutionary change in signatures of translation efficiency. *Genome Biol.* 15:R44. doi: 10.1186/gb-2014-15-3-r44

LaBella, A. L., Opulente, D. A., Steenwyk, J. L., Hittinger, C. T., and Rokas, A. (2019). Variation and selection on codon usage bias across an entire subphylum. *PLoS Genet.* 15:e1008304. doi: 10.1371/journal.pgen.1008304

Lavner, Y., and Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 345, 127–138. doi: 10.1016/j.gene.2004.11.035

Li, W. L., and Yang, B. (2017). Translational genomics of grain size regulation in wheat. *Theor. Appl. Genet.* 130, 1765–1771. doi: 10.1007/s00122-017-2953-x

Ling, H. Q., Ma, B., Shi, X. L., Liu, H., Dong, L. L., Sun, H., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* 557, 424–428. doi: 10.1038/s41586-018-0108-0

Liu, H., He, R., Zhang, H., Huang, Y., Tian, M., and Zhang, J. (2010). Analysis of synonymous codon usage in Zea mays. *Mol. Biol. Rep.* 37, 677–684. doi: 10.1007/s11033-009-9521-7

Liu, X. Y., Li, Y., Ji, K. K., Zhu, J., Ling, P., Zhou, T., et al. (2020). Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in *Cuscuta australis*. *Genomics* 112, 2695–2702. doi: 10.1016/j.ygeno.2020.03.002

Manokaran, G., Sujatmoko, X., McPherson, K. G., and Simmons, C. P. (2019). Attenuation of a dengue virus replicon by codon deoptimization of nonstructural genes. *Vaccine* 37, 2857–2863. doi: 10.1016/j.vaccine.2019.03.062

Mazumdar, P., Binti, O. R., Mebus, K., Ramakrishnan, N., and Ann, H. J. (2017). Codon usage and codon pair patterns in non-grass monocot genomes. *Ann. Bot.* 120, 893–909. doi: 10.1093/aob/mcx112

Mohanta, T. K., Mishra, A. K., Hashem, A., Abd Allah, E. F., Khan, A. L., and Al-Harrasi, A. (2020). Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine ($tRNA^{Sec}$). *BMC Genomics* 21:804. doi: 10.1186/s12864-020-07216-3

Mukhopadhyay, P., Basak, S., and Ghosh, T. C. (2007). Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes. *Gene* 400, 71–81. doi: 10.1016/j.gene.2007.05.027

Muyle, A., Serres-Giardi, L., Ressayre, A., Escobar, J., and Glémin, S. (2011). GC-biased gene conversion and selection affect GC content in the *Oryza* genus (rice). *Mol. Biol. Evol.* 28, 2695–2706. doi: 10.1093/molbev/msr104

Nasrullah, I., Butt, A. M., Tahir, S., Idrees, M., and Tong, Y. (2015). Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol. Biol.* 15:174. doi: 10.1186/s12862-015-0456-4

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

Peden, J. F. (2000). Analysis of codon usage. dissertation/master's thesis. Nottingham: University of Nottingham.

Plotkin, J. B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42. doi: 10.1038/nrg2899

Rensing, S. A., Fritzowsky, D., Lang, D., and Reski, R. (2005). Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics* 6:43. doi: 10.1186/1471-2164-6-43

Sahoo, S., Das, S. S., and Rakshit, R. (2019). Codon usage pattern and predicted gene expression in *Arabidopsis thaliana*. *Gene X* 2:100012. doi: 10.1016/j.gene.2019.100012

Sauna, Z. E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* 12, 683–691. doi: 10.1038/nrg3051

Shah, P., and Gilchrist, M. A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10231–10236. doi: 10.1073/pnas.1016719108

Sharp, P. M., and Li, W. H. (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol. Biol. Evol.* 4, 222–230. doi: 10.1093/oxfordjournals.molbev.a040443

Shen, Z., Gan, Z., Zhang, F., Yi, X., Zhang, J., and Wan, X. (2020). Analysis of codon usage patterns in citrus based on coding sequence data. *BMC Genomics* 21(Suppl. 5):234. doi: 10.1186/s12864-020-6641-x

Sirihongthong, T., Jitobaom, K., Phakaratsakul, S., Boonarkart, C., Suptawiwat, O., and Auewarakul, P. (2019). The relationship of codon usage to the replication strategy of parvoviruses. *Arch. Virol.* 164, 2479–2491. doi: 10.1007/s00705-019-04343-5

Su, M. W., Lin, H. M., Yuan, H. S., and Chu, W. C. (2009). Categorizing host-dependent RNA viruses by principal component analysis of their codon usage preferences. *J. Comput. Biol.* 16, 1539–1547. doi: 10.1089/cmb.2009.0046

Sueoka, N. (2001). Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses. *J. Mol. Evol.* 53, 469–476. doi: 10.1007/s002390010237

Szövényi, P., Ullrich, K. K., Rensing, S. A., Lang, D., van Gessel, N., Stenøien, H. K., et al. (2017). Selfing in haploid plants and efficacy of selection: codon usage bias in the model moss *Physcomitrella patens*. *Genome Biol. Evol.* 9, 1528–1546. doi: 10.1093/gbe/evx098

Versteeg, R., van Schaik, B. D., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., et al. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* 13, 1998–2004. doi: 10.1101/gr.1649303

Wan, X. F., Xu, D., Kleinhofs, A., and Zhou, J. Z. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol. Biol.* 4:19. doi: 10.1186/1471-2148-4-19

Wang, X. M., Chen, S. Y., Shi, X., Liu, D. N., Zhao, P., Lu, Y. Z., et al. (2019). Hybrid sequencing reveals insight into heat sensing and signaling of bread wheat. *Plant J.* 98, 1015–1032. doi: 10.1111/tpj.14299

Wang, H. C., and Hickey, D. A. (2007). Rapid divergence of codon usage patterns within the rice genome. *BMC Evol. Biol.* 7(Suppl. 1):S6. doi: 10.1186/1471-2148-7-S1-S6

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Yang, Q., Li, Z., Li, W. Q., Ku, L. X., Wang, C., Ye, J. R., et al. (2013). CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16969–16974. doi: 10.1073/pnas.1310949110

Yu, G. C., Wang, L. G., Han, Y. Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zalucki, Y. M., Beacham, I. R., and Jennings, M. P. (2009). Biased codon usage in signal peptides: a role in protein export. *Trends Microbiol.* 17:19307122, 146–150. doi: 10.1016/j.tim.2009.01.005

Zalucki, Y. M., Power, P. M., and Jennings, M. P. (2007). Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. *Nucleic Acids Res.* 35:5748. doi: 10.1093/nar/gkm577

Zhang, R. Z., Huang, S. Y., Li, S. M., Song, G. Q., Li, Y. L., Li, W., et al. (2020). Evolution of PHAS loci in the young spike of allohexaploid wheat. *BMC Genomics* 21:200. doi: 10.1186/s12864-020-6582-4

Zheng, X. W., Li, X. H., Ge, C., Chang, J. Z., Shi, M. M., Chen, J. L., et al. (2017). Characterization of the CCT family and analysis of gene expression in *Aegilops tauschii*. *PLoS One* 12:e0189333. doi: 10.1371/journal.pone.0189333

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.

Copyright © 2021 Yang, Zhao, Wang, Zhao, Qiao, Wu, Yan, Zheng and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fgene.2021.685788

Integrated Analysis of mRNA and Non-coding RNA Transcriptome in Pepper (*Capsicum chinense*) Hybrid at Seedling and Flowering Stages

*Huang-ying Shu, He Zhou, Hai-ling Mu, Shu-hua Wu, Yi-li Jiang, Zhuang Yang, Yuan-yuan Hao, Jie Zhu, Wen-long Bao, Shan-han Cheng, Guo-peng Zhu and Zhi-wei Wang**

Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, College of Horticulture, Hainan University, Haikou, China

Pepper is an important vegetable in the world. In this work, mRNA and ncRNA transcriptome profiles were applied to understand the heterosis effect on the alteration in the gene expression at the seedling and flowering stages between the hybrid and its parents in *Capsicum chinense*. Our phenotypic data indicated that the hybrid has dominance in leaf area, plant scope, plant height, and fruit-related traits. Kyoto Encyclopedia of Genes and Genomes analysis showed that nine members of the plant hormone signal transduction pathway were upregulated in the seedling and flowering stages of the hybrid, which was supported by weighted gene coexpression network analysis and that *BC332_23046* (auxin response factor 8), *BC332_18317* (auxin-responsive protein IAA20), *BC332_13398* (ethylene-responsive transcription factor), and *BC332_27606* (ethylene-responsive transcription factor WIN1) were candidate hub genes, suggesting the important potential role of the plant hormone signal transduction in pepper heterosis. Furthermore, some transcription factor families, including bHLH, MYB, and HSF were greatly over-dominant. We also identified 2,525 long ncRNAs (lncRNAs), 47 micro RNAs (miRNAs), and 71 circle RNAs (circRNAs) in the hybrid. In particular, downregulation of miR156, miR169, and miR369 in the hybrid suggested their relationship with pepper growth vigor. Moreover, we constructed some lncRNA–miRNA–mRNA regulatory networks that showed a multi-dimension to understand the ncRNA relationship with heterosis. These results will provide guidance for a better understanding of the molecular mechanism involved in pepper heterosis.

Keywords: pepper, heterosis, RNA-seq, lncRNA, miRNA, transcription factor

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Cheng Qin,
Zunyi Vocational and Technical College, China
Jianbin Hu,
Henan Agricultural University, China
Guangwu Zhao,
Zhejiang Agriculture and Forestry University, China

****Correspondence:***
Zhi-wei Wang
zwwang22@163.com

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

Received: *25 March 2021*
Accepted: *26 July 2021*
Published: *20 August 2021*

Citation:
Shu H-Y, Zhou H, Mu H-L, Wu S-H, Jiang Y-L, Yang Z, Hao Y-Y, Zhu J, Bao W-L, Cheng S-H, Zhu G-P and Wang Z-W (2021) Integrated Analysis of mRNA and Non-coding RNA Transcriptome in Pepper (Capsicum chinense) Hybrid at Seedling and Flowering Stages. Front. Genet. 12:685788. doi: 10.3389/fgene.2021.685788

INTRODUCTION

Heterosis is a fundamental biological phenomenon characterized by a superior agricultural performance that enhances plant biomass, growth rates, seed yield, and stress tolerance, compared with their inbred parents. Hybrids have been comprehensively applied to agriculture, to increase yield and quality. Three classical hypotheses have been proposed to explain the genetic control of heterosis: dominance (Xiao et al., 1995; Garcia et al., 2008; Li et al., 2014), overdominance (East, 1936; Larièpe et al., 2012; Zhou et al., 2012), and epistasis (Powers, 1944). Subsequently, comprehensive research suggested that quantities of superior alleles were related to heterosis (Yu et al., 1997; Melchinger et al., 2007). Tremendous progress has been made in characterizing the

genetic mechanisms of plant heterosis, including *Arabidopsis* (Groszmann et al., 2014; Yang et al., 2017), rice (Li et al., 2016; Shao et al., 2019), maize (Ding et al., 2014; Hu et al., 2016), oilseed (Shen et al., 2017), and tomato (Krieger et al., 2010). For example, studies suggest that DNA methylation (Kawanabe et al., 2016; Lauss et al., 2017) and non-coding RNA (ncRNA) are linked to heterosis (Ng et al., 2012; Li et al., 2014). These data provide important hints for studying the heterosis of other plants. However, the molecular mechanism of heterosis still remains to be further deciphered (Govindaraju, 2019).

Pepper (*Capsicum* L.) is a genus of the *Solanaceae* family, which is an important vegetable and ornamental plant. Pepper spread all over the world because of its superior adaptability to diverse agroclimatic regions (Qin et al., 2014). The global pepper planting area was about ∼1.99 million hectares with a production of ∼36.77 million tons (FAO, 2018). Moreover, pepper plays an important role in pharmaceuticals, cosmetics, and pigmentation (Kothari et al., 2010). Pepper hybrids show strong heterosis and have been broadly used in commercial production. However, the molecular mechanism of heterosis in pepper is poorly understood, which seriously impedes the efficient breeding of pepper hybrids.

The application of new biotechnologies, including next-generation sequencing on the *Capsicum chinense* (one of the five cultivated species of the *Capsicum* genus), helped to explore the heterosis of pepper at the genomic level (Kim et al., 2014, 2017; Qin et al., 2014). The weighted gene coexpression network analysis (WGCNA) is a promising tool to reveal coexpressed gene networks or modules that are based on transcriptome data, which combines phenotypic features with these modules to detect the key genes in the networks by WGCNA (Langfelder and Horvath, 2008). This procedure has been extensively used to identify the module key genes of plant horticulture traits, but fewer studies have applied WGCNA to illustrate the gene networks underlying pepper heterosis.

In the present study, we aimed to provide comprehensive information to study heterosis and hybrid breeding by using non-directional sequencing, strand-specific RNA sequencing, and small RNA analysis. We implemented transcriptomic analysis at the pepper seedling stage (S-stage) and flowering stage (F-stage) to explore the candidate genes related to heterosis. Through WGCNA, we identified some possible genes that play a role in the networks. Moreover, we detected the differential transcriptome containing lncRNA, circRNA, and miRNA among two parents and their F_1 hybrid at the S-stage of *C. chinense*. Overall, these data will be helpful for further illustration of the molecular mechanism of heterosis and open new ways to study heterosis in pepper.

MATERIALS AND METHODS

Plant Materials and Morphological Observation

Pepper plants (*C. chinense*) of HNCc16, HNCc22, and their hybrid HNCy01 with HNCc22 as the maternal parent were planted in the growth room with 16 h light (3,500 Lx) at 26°C followed by 8 h dark at 20°C. The plant height (PH), stem diameter, and leaf area (LA) were measured (30 days at the S-stage and 70 days at the F-stage). The LA was calculated using ImageJ (v1.51). Leaves were frozen immediately in liquid nitrogen after harvest and stored at −80°C for RNA extraction (triplicate biological duplications for each sample).

RNA Extraction and RNA-Seq

The total RNA of each sample was isolated with Trizol reagents under the manufacturer's instruction (Thermo Fisher Scientific, Shanghai, China). Eighteen non-directional libraries were produced using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) and sequenced on Illumina Novaseq platform. Nine strand-specific libraries were produced by rRNA-depleted RNA by NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® with the dUTP second-strand marking (NEB, Ipswich, MA, United States) and used for generating mRNA, lncRNA, and circRNA data. Moreover, nine small RNA libraries were constructed using NEBNext® Multiplex Small RNA Library Prep Kit for Illumina® (NEB, Ipswich, MA, United States) and used for generating miRNA data. The cleaned Illumina reads were aligned to the pepper reference genome[1] using HISAT2 (v.2.0.4) (Kim et al., 2015). All raw data were deposited in the National Center for Biotechnology Information Sequence Read Archive with accession numbers: PRJNA694379 for non-directional sequencing library and PRJNA655561 for strand-specific RNA sequencing library and small RNA library. The samples MRS30, RS30, and YS30 were from the S-stage, and MRF60, RF60, and YF60 were from the F-stage. The sample names MRs, Rs, and Ys in the database are equivalent to HNCy01, HNCc22, and HNCc16 in this paper[2].

Differential Expression Analysis

Fragments per kilo-base of exon per million fragments mapped (FPKM) of lncRNAs and mRNAs in each sample was calculated by StringTie (v2.1.1) (Pertea et al., 2016). Expression levels of miRNAs and circRNAs were estimated by transcript per million (TPM) (Zhou et al., 2010). Differentially expressed genes (DEGs), DE lncRNAs, DE miRNAs, and DE circRNAs were analyzed by the DESeq2 (v1.16.1) (Love et al., 2014). $|\text{Log2(foldchange)}| > 1$ and $p\text{-adjust} < 0.05$ were set as the threshold for significantly differential expression by default. The DEGs were classified into five major expression patterns based on the gene expression level: additive, high-parental dominance, low-parental dominance, under-dominance, and over-dominance (Li et al., 2014). The coexpression network visualization of differential ncRNAs was conducted by the Cytoscape (v3.8.0) software[3]. All of the Venn diagrams were drawn by Jvenn[4] (Bardou et al., 2014), and heat maps were drawn by R package (pheatmap).

[1]https://www.ncbi.nlm.nih.gov/genome/?term=Capsicum+chinense

[2]https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?

[3]https://cytoscape.org/

[4]http://jvenn.toulouse.inra.fr/app/usermanual.html

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis

Gene Ontology (GO) enrichment of DEGs was analyzed by GOseq (Young et al., 2010). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a public database resource for annotating the biological system (Kanehisa et al., 2008). Using KOBAS software, we identified the enrichment of the target DE candidate genes, DE lncRNAs, and DE miRNAs in KEGG pathways (Mao et al., 2005).

Transcription Factors Analysis

Transcription factors (TFs) were analyzed by iTALK (v1.2) software[5](Pérez-Rodríguez et al., 2010).

Analysis of WGCNA Gene Coexpression Network

The weighted gene coexpression network analysis was used to identify genes for network construction, gene cluster, and visualization (Langfelder and Horvath, 2008). The correlation of pepper phenotypic traits was analyzed by R package corPvalueStudent function. The gene correlation and soft thresholding power analysis were based on Pearson correlation matrix. In WGCNA networks, a power of 10 was chosen (model fitting index R^2 = 0.8). The gene dendrogram was used for module detection, and dynamic tree cutting method was used (minModuleSize and mergeCutHeight value were set as 100 and 0.25, respectively). The phenotype data were imported into WGCNA package, and the correlation between phenotype and gene module was calculated using default settings.

RESULTS

Morphological Comparison of F_1 Hybrid and Its Parents

We collected the pepper hybrid and its parents at S-stage and F-stage for RNA sequencing analysis (RNA-seq). These two periods play a vital role in the growth and heterosis interpretation of pepper. Moreover, both F_1 hybrid HNCy01 and maternal line HNCc22 were more vigorous at these two phases, and the hybrid showed higher PH than the paternal line HNCc16. At S-stage, the hybrid LA was larger than the parents, and the diameter of the stem of the hybrid was similar to that of the thick stem parent (HNCc16). The hybrid crown diameter and PH have a greater advantage than its parents at F-stage. Moreover, fruit diameter, fruit length (FL), and fruit weight (FW) also showed strong hybrid vigor (**Figure 1**).

Sequencing and Mapping Reads to the Pepper Reference Genome

We performed RNA sequencing of the three genotypes at the S-stage and F-stage and constructed 18 libraries. A total of 83.72 million raw reads were generated from the three genotypes at S-stage and F-stage through RNA-seq on Illumina HiSeq 2500 platform. The paired-end sequences with low-quality reads were filtered out. Finally, 80.92 million clean reads were obtained. On average, GC content was 42.42%, and Q30 was 93.5%. Interestingly, about 91.02% of reads were mapped to the *C. chinense* reference genome, and 86.71% were aligned to unique positions. Among these reads, 51.35–67.77% of reads were distributed to exonic regions, 27.25–43.84% to intergenic, and 4.49–7.94% to intronic regions, respectively (**Supplementary Table 1**).

To comprehensively investigate how the hybrid influences the transcript profile, we also constructed nine strand-specific RNA libraries (used for acquiring mRNAs, lncRNAs, and circRNAs) and nine small RNA libraries. For strand-specific RNA sequencing, the raw data size of each sample ranges from 87.73 to 158.25 million reads, obtaining 83.56–150.49 million clean reads. For small RNA sequencing, the numbers of raw reads were distributed from 13.02 to 20.36 million, including 11.59–19.72 million clean reads. Thus, at least 84.9% of strand-specific clean data in each sample were aligned with the reference genome, and over 76.07% of small RNAs from each sample were mapped to the reference genome (**Supplementary Table 2**).

Differential Gene Expression Analysis

At the S-stage, 28,566 DEGs were identified between hybrid HNCy01 and the parental line, and 4,562 DEGs were discovered between the parent HNCc22 and HNCc16 with a similar number of upregulated DEGs and downregulated DEGs. At the F-stage, there were 10,727 among HNCy01 and its parents and 6,759 between inbred, respectively (**Supplementary Figure 1**). Through a Venn diagram, we identified 2,672 and 3,315 DEGs between the hybrid HNCy01vs HNCc22 and HNCc22vs HNCc16 at S-stage and F-stage, respectively, with 2,892 and 2,585 DEGs between HNCy01vs HNCc16 and HNCc22vs HNCc16 at both stages. In addition, 1,491 and 956 DEGs were identified among three genotype combinations at the two stages (**Figures 2A,B**). Moreover, hierarchy analysis showed that different samples at the same development stages clustered together (**Figure 2C**).

By studying differential gene expression among the hybrid and its parents in two development stages, we found that a total of 633 and 800 DEGs showed additive effects in the S-stage and F-stage among hybrid HNCy01 and its parents, respectively. Whereas 12,719 and 5,042 DEGs at S-stage and F-stage displayed non-additivity effects, the non-additive genes were further divided into four patterns: high-parent dominance, low-parent dominance, over-dominance, and under-dominance (**Table 1**). Moreover, research on the comparison of gene expression patterns displayed that 3,008 non-additive genes were retained in both stages. In the F-stage, 384 genes showed an additive pattern, but in the S-stage, they showed a non-additive pattern. Besides, 334 additive genes were maintained in these two stages, and a total of 227 genes displayed non-additive expression patterns at the F-stage but additive patterns at the S-stage (**Figure 2D**). These results indicated that the non-additive

[5]http://planttfdb.gao-lab.org/

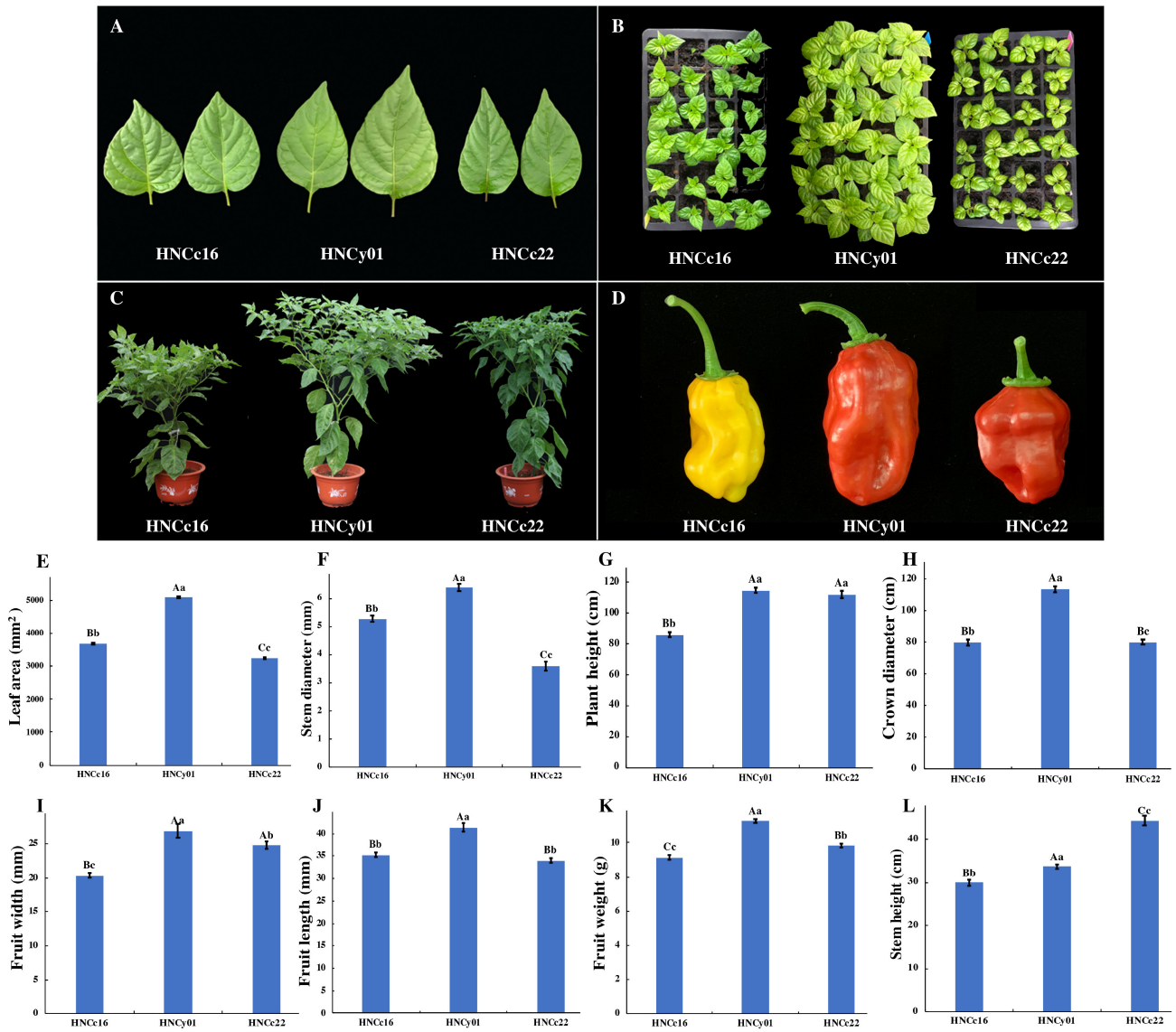


FIGURE 1 | The growth of the F_1 hybrid and its pepper parents at the S-stage and F-stage. **(A)** Leaves from F_1, HNCc22, and HNCc16 at the S-stage. **(B)** Plants of the F_1 and its parents at 35 days after germination. **(C)** Plants of the hybrid and its parents at F-stage of 70 days after germination. **(D)** The fruit of the hybrid and its parents. **(E,F)** Plant LA and stem diameter of the hybrid and parents at the S-stage. **(G–L)** Plant height, crown diameter, fruit width, fruit length, fruit weight, and stem height of hybrid and its parents at the F-stage. Different letters above the columns of bars indicate significant differences at *p*-value less than 0.05 (lowercase letters) or less than 0.01 (capital letters).

expression genes played an important role in the different development stages.

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

To delve into an overall insight in the molecular and biological functions of DEGs between the hybrid and its parents, we identified 12,719 and 5,042 non-additively expressed genes at S-stage and F-stage, respectively. Among HNCy01 and its parents at these two stages, the biology process, cellular process, organic substance metabolic process, and primary metabolic process were enriched most significantly in the biological process category. In the cellular component category, the cellular component, cell, cell part, and photosystem were enriched. In the molecular function category, the transferase activity, transferring phosphorus-containing groups, phosphotransferase activity, alcohol group as acceptor, and kinase activity were significantly abundant (**Figure 3**). Furthermore, we inspected that serval terms exhibited over-dominance expression patterns at both development stages (**Table 2**).

Kyoto Encyclopedia of Genes and Genomes analysis of DEGs at the two development stages displayed non-additional enrichment pathways, including ribosome biogenesis in eukaryotes, proteasomes, plant hormone signal transduction,

TABLE 1 | Differentially expressed genes between the hybrid and its parents at the seedling and flowering stages.

Expression patterns		S-stage	F-stage
Total		13,352	5,842
Additive		633 (4.74%)	800 (13.69%)
Non-additive		12,719 (95.26%)	5,042 (86.31)
High-parental dominance	ELD-F	721	1,286
	ELD-M	818	987
Low-parental dominance	ELD-F	448	1,047
	ELD-M	565	627
Over-dominance		4,869	550
Under-dominance		5,298	545

ribosomes, photosynthesis-antenna proteins, and photosynthesis at the S-stage. However, at the F-stage, protein processing in the endoplasmic reticulum, photosynthesis, photosynthesis-antenna proteins, plant hormone signal transduction, and carotenoid biosynthesis were significantly enriched (**Supplementary Figure 2**). Interestingly, we found that plant hormone signal transduction, photosynthesis, and photosynthesis-antenna proteins were high-parental dominant at the S-stage and over-dominant at the F-stage. Particularly, the plant hormone signal transduction pathways were extremely enriched, where *BC332_21619*, *BC332_18317*, *BC332_12434*, *BC332_25121*, and *novel.2052* were over-dominant in both developmental stages. These results strongly suggest the contribution to the pathway for the growth vigor in the F_1 hybrid (**Figure 4** and **Supplementary Tables 3**, **4**).

Analysis of Coexpressed Gene Networks and Candidate Genes

To reveal the transcriptional regulation of pepper heterosis based on RNA-seq data among the hybrid and inbred at two development stages, WGCNA was used to identify 42 modules of coexpressed genes to reveal the correlation between

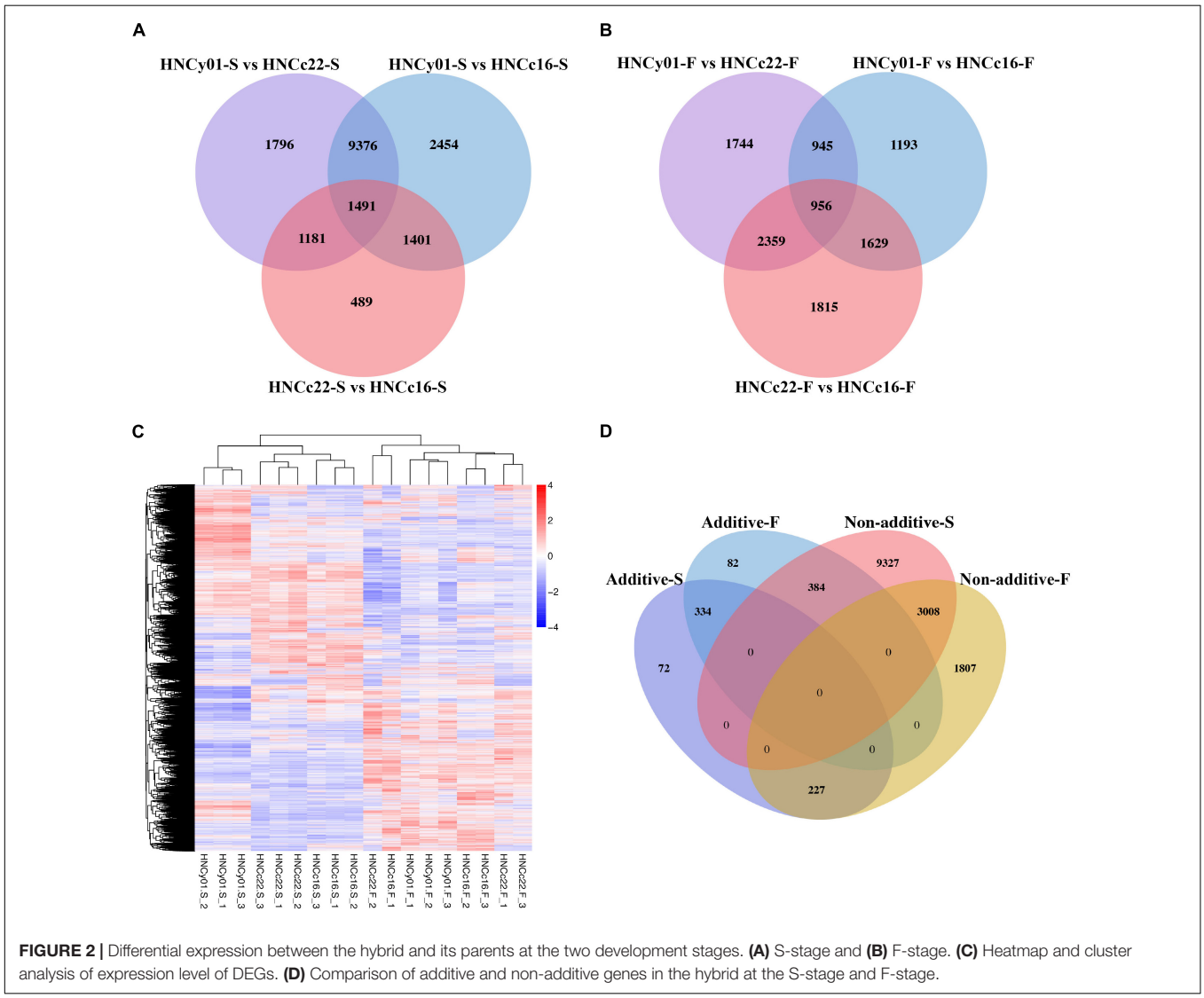


FIGURE 2 | Differential expression between the hybrid and its parents at the two development stages. **(A)** S-stage and **(B)** F-stage. **(C)** Heatmap and cluster analysis of expression level of DEGs. **(D)** Comparison of additive and non-additive genes in the hybrid at the S-stage and F-stage.

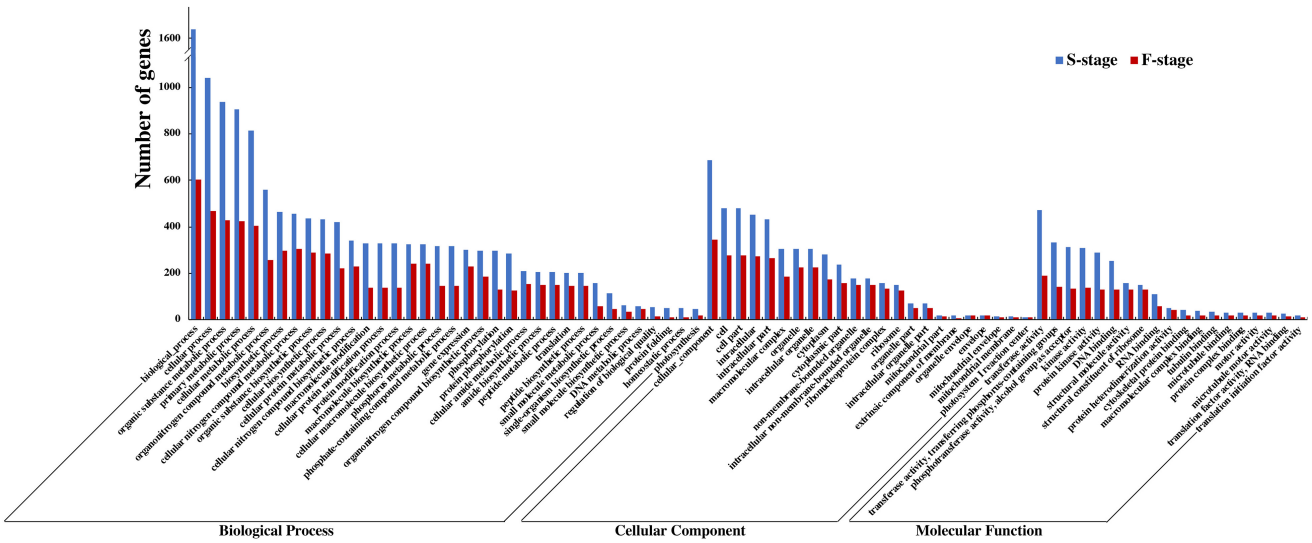


FIGURE 3 | Comparison of Gene Ontology (GO) classification of differentially expressed genes (DEGs) at the seedling and flowering stages.

genes and traits indexes. These modules were represented by different colors and displayed with a heatmap (**Figure 5A** and **Supplementary Figure 3**).

Among the 42 coexpressed gene modules, each expression cluster was displayed in a heatmap, which can directly visualize the relationship between the clusters of pepper at two development stages (**Figure 5B**). Subsequently, eight horticulture phenotypic data of module–trait correlations were analyzed at the two developmental stages. We found that physiological indexes were mainly concentrated in purple at LA and plant stem diameter (PSD) with a moderate correlation coefficient of 0.45–0.56 ($p < 0.05$). Besides, the ligthsteelblue1 module displayed a positive correlation with plant scope (PS), FL, and main stem height (MSH), while the black and white modules were significantly positively correlated with PH and FW at F-stage. Furthermore, the genes with a higher weight in each module were chosen for network constructing and analysis (**Figure 6**). By searching the hub gene of the major gene network, *BC332_ 11557*, *novel.744*, *BC332_18317*, and *BC332_23046* were related to β-glucosidase and auxin responsive (**Table 3**). These results exhibited that these genes may have a relationship with the heterosis of pepper.

TABLE 2 | Significant Gene Ontology (GO) terms of over-dominance at the seedling and flowering stages.

GO accession	Description	Term type	*Q*-value
GO:0016310	Phosphorylation	biological_process	1.9034E-09
GO:0006468	Protein phosphorylation	biological_process	2.7427E-09
GO:0006793	Phosphorus metabolic process	biological_process	5.7423E-07
GO:0006796	Phosphate-containing compound metabolic process	biological_process	5.7423E-07
GO:0016773	Phosphotransferase activity, alcohol group as acceptor	molecular_function	5.6832E-10
GO:0016301	Kinase activity	molecular_function	8.1118E-10
GO:0004672	Protein kinase activity	molecular_function	1.5464E-09
GO:0016772	Transferase activity, transferring phosphorus-containing groups	molecular_function	1.9132E-07
GO:0016740	Transferase activity	molecular_function	0.0001406

Analysis of DE lncRNAs, DE miRNAs and DE circ RNAs

The heat map illustrated that the F_1 was more similar to its parental line HNCc16 **(Figure 7A)**. Venn diagram showed that 1,201 and 96 genes had a unique expression in HNCy01vs HNCc16 and HNCy01vs HNCc22. Moreover, 93 DE lncRNAs were commonly expressed among all three genotypes (**Figure 7B**). In addition, we identified 1,932 DE lncRNAs (985 upregulated and 947 downregulated) and 593 DE lncRNAs (405 upregulated and 188 downregulated) in HNCy01 compared with HNCc16 and HNCc22 at S-stage, respectively (**Figure 7C** and **Supplementary Table 5**).

A total of 17 DE miRNAs (seven upregulated and 10 downregulated) were identified in HNCy01and HNCc16, and 30 DE miRNAs (15 upregulated and 15 downregulated) were identified in HNCy01 and HNCc22. These DE miRNAs include 13 known miRNA and 34 novel miRNAs, of which conserved miRNAs belong to seven families (**Supplementary Figure 4** and **Supplementary Table 6**). There were 54 DE circRNAs (27 upregulated and 27 downregulated) in HNCy01 compared to HNCc16. In comparison with HNCc22, we identified 17 DE circRNAs (two upregulated and 15 downregulated) in HNCy01 (**Supplementary Figure 5** and **Supplementary Table 7**).

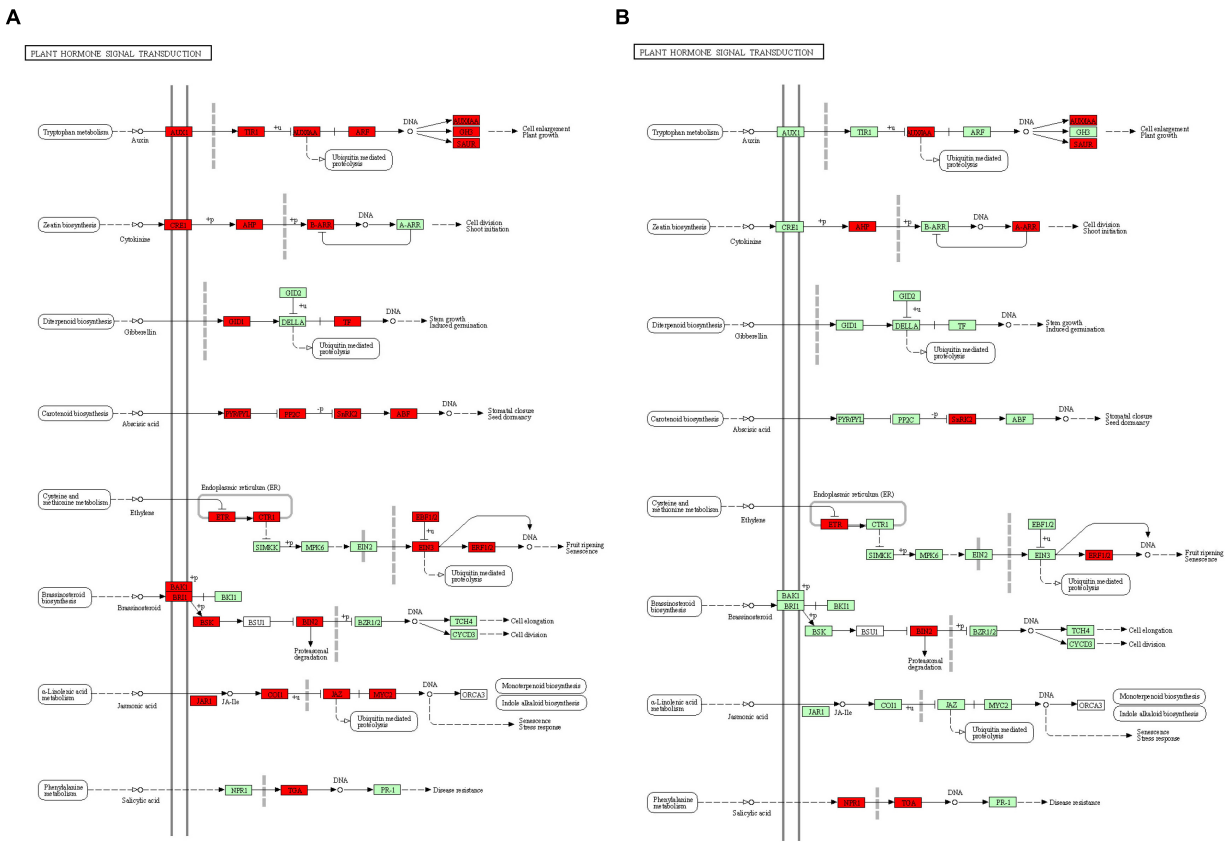


FIGURE 4 | The "plant hormone signal transduction" pathway enriched by Kyoto Encyclopaedia of Genes and Genomes (KEGG) analysis. **(A)** At S-stage. **(B)** At F-stage. Upregulated KEGG Orthology (KO) nodes are marked with red.

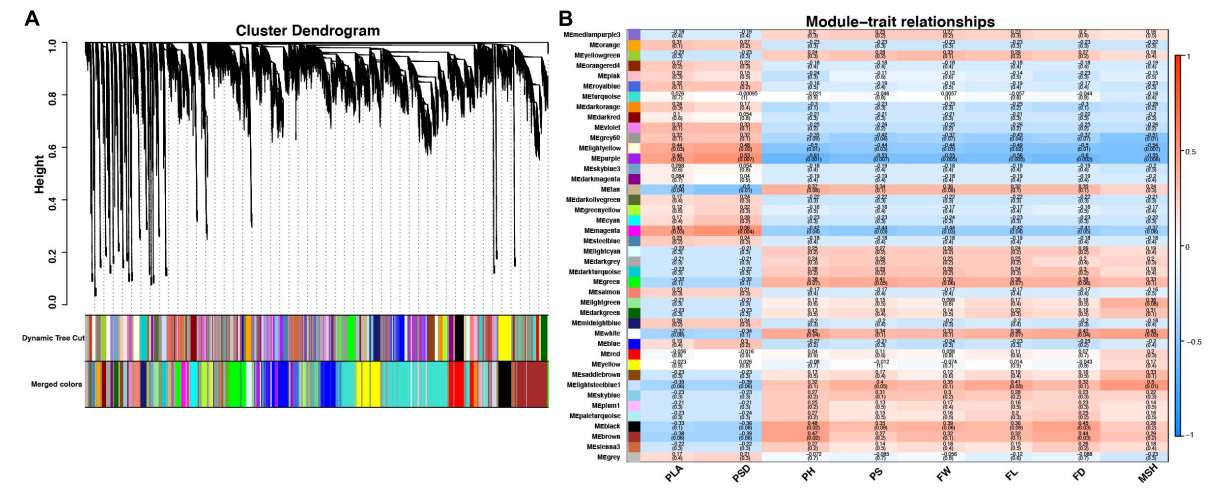


FIGURE 5 | Cluster dendrogram and network heatmap of coexpressed genes. **(A)** Hierarchical cluster of 42 modules coexpressed genes. **(B)** Module–trait relationship. Each row represents a module eigengene, and each column presents a trait. Each module includes the corresponding correlation and *p*-value. PLA, plant leaf area; PSD, plant stem diameter; PH, plant height; PS, plant scope; FL, fruit length; FD, fruit diameter; FW, fruit weight; MSH, main stem height.

To better reveal the DE lncRNA and DE miRNA functions, through analysis of targeted mRNA of DE lncRNA and DE miRNA, we identified 777 (725) upregulated targeted mRNA of upregulated lncRNA and 176 (339) downregulated targeted mRNA of downregulated lncRNAs in the F_1 hybrid in comparison with its parent HNCc22 (HNCc16). Similarity,

one (11) upregulated targeted mRNA of downregulated lncRNAs and 41 (99) downregulated targeted mRNAs of upregulated lncRNAs in the F_1 hybrid in comparison with its parent HNCc22 (HNCc16) were identified (**Figures 8A,B** and **Supplementary Table 8**). Furthermore, by intersection analysis of DE miRNAs, we found 39 (19) upregulated targeted mRNAs of downregulated miRNAs and 22 (11) downregulated targeted mRNAs of upregulated miRNAs in the F_1 hybrid compared with its parent HNCc22 (HNCc16) (**Figures 8C,D** and **Supplementary Table 9**). We also identified one upregulated targeted mRNA from one upregulated circRNA and two downregulated targeted mRNAs from two downregulated circRNAs (**Table 4**).

By GO enrichment analysis of the targeted mRNA of DE lncRNAs, GO terms binding (GO:0005488) was remarkably enriched in molecular function categories. Among the biological process categories, the signal–organism process (GO:0044699) was highly represented. Membrane (GO:0016020) and membrane part (GO:0044425) were dominated in cellular components. GO enrichment of the targeted mRNA of DE miRNAs showed that binding (GO:0005488) was enriched in molecular function categories. In addition, phosphorylation (GO:0016310) and protein phosphorylation (GO:0006468) were significantly enriched in biological process categories. KEGG enrichment analysis of DE lncRNAs suggested that photosynthesis and metabolic pathways were enriched (**Supplementary Figure 6**).

We constructed the lncRNA–miRNA–mRNA coexpression network. In HNCy01 and HNCc22, 249 lncRNA–miRNA–mRNA combinations were identified, including 26 lncRNAs as decoys, nine miRNAs as centers, and 70 mRNAs as targets. HNCy01 vs. HNCc16 showed 293 lncRNA–miRNA–mRNA combinations with 55 lncRNAs, 14 miRNAs, and 60 mRNAs (**Figure 9** and **Supplementary Table 10**).

DEGs Encoding Transcription Factor

A total of 2,145 differentially expressed TFs were identified at the two developmental stages and belonged to 57 TF families, including 213 bHLH, 176 MYB-related, 128 ERF, 118 WRKY, 116 NAC, 110 B3, 89 C3H, 81 MYB, 76 M-type MADS, and 39 HSF. Among these TFs, the bHLH family accounted for the largest proportion, followed by MYB-related family ERF family protein and WRKY family at two developmental stages. Furthermore, 32.17% (690/2,145) TFs showed high-parental dominance or over-dominance, which indicated that these TFs might play an important function in pepper heterosis (**Supplementary Table 11**).

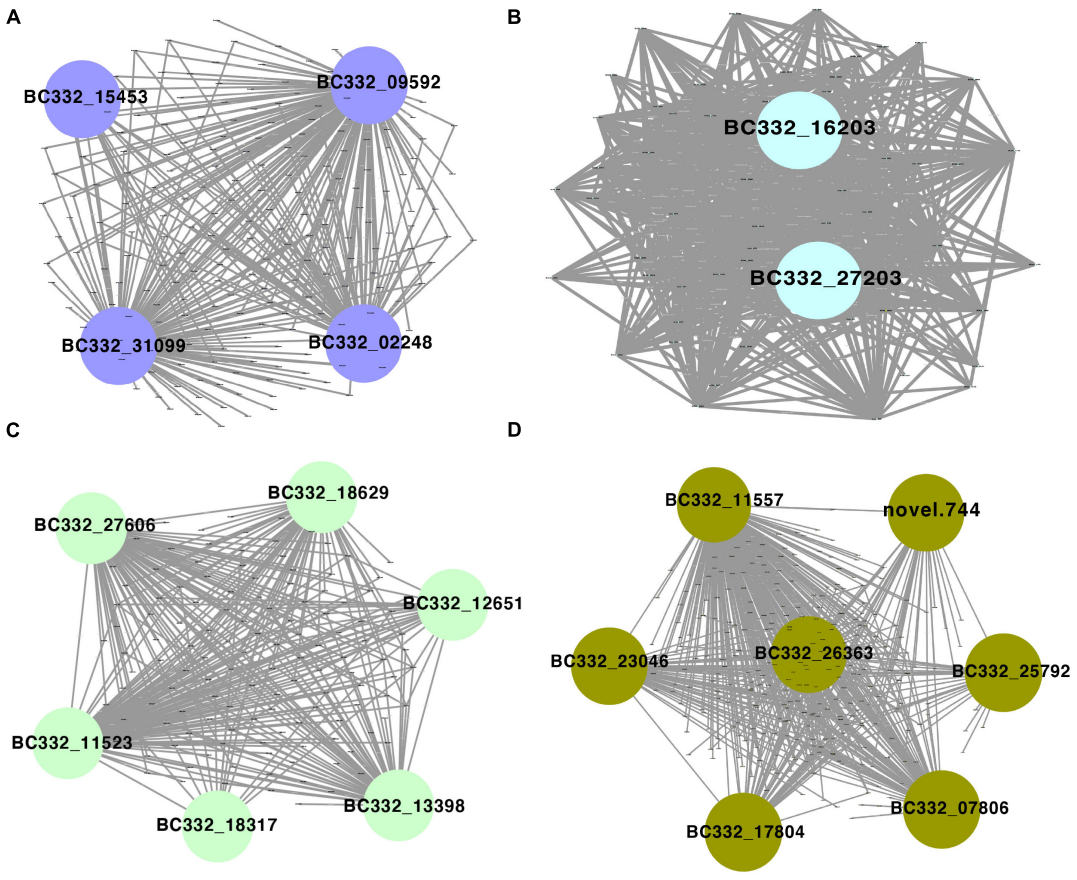


FIGURE 6 | Gene networks and hub genes involved in heterosis regulation during pepper development by WGCNA analysis. **(A)** Gene network of the purple module. **(B)** Gene network of the lightsheetblue1 module. **(C)** Gene network of the white module. **(D)** Gene network of the black module.

DISCUSSION

Heterosis utilization plays a crucial role in crop production. Our current knowledge of heterosis molecular mechanism is mainly based on rice and *Arabidopsis*. However, the understanding of the pepper molecular mechanism heterosis is still limited. Here, we used non-directional sequencing, strand-specific RNA sequencing, and miRNA analysis to understand the heterosis of pepper hybrid at two development stages, providing important preliminary data and clues for the further study of heterosis in pepper.

Phenotype Change in the F_1 Hybrid

The leaf photosynthesis is crucial for plant development (Yang et al., 2018). A recent study has proved that the *Arabidopsis* F_1 hybrid germinated earlier and has a larger leaf size than its parents. Furthermore, the larger size of the leaf was associated with increased cell numbers with a greater number of chloroplasts, which increased photosynthate production contributing to higher biomass (Fujimoto et al., 2012; Groszmann et al., 2014; Liu et al., 2020a). In our study, the pepper F_1 hybrid has a larger LA and grows quicker than its parents. Besides, compared with its parents, the hybrid HNCy01 has the over-dominance advantage in crown diameter. However, the physiological basis of these advantages needs further analysis.

TABLE 3 | The detailed information of hub genes.

Gene ID	Module	Description	Pattern
BC332_31099	Purple	Putative protein phosphatase 2C 53	Over-dominance
BC332_09592	Purple	Protein phosphatase 2C 16	Over-dominance
BC332_02248	Purple	Transcription factor bHLH80	Over-dominance
BC332_15453	Purple	Heat stress transcription factor A-2e	Over-dominance
BC332_27606	White	Ethylene-responsive transcription factor WIN1	Over-dominance
BC332_13398	White	Ethylene-responsive transcription factor	Over-dominance
BC332_11523	White	Serine/threonine-protein kinase SAPK2	Over-dominance
BC332_18629	White	Transcription factor PIF1	Over-dominance
BC332_12651	White	Transcription factor TGA7	Over-dominance
BC332_18317	White	Auxin-responsive protein IAA20	Over-dominance
BC332_07806	Black	Flowering locus K -like proteiny domain	Over-dominance
BC332_17804	Black	Transcription factor bHLH66	Over-dominance
BC332_23046	Black	Auxin response factor 8	Over-dominance
BC332_25792	Black	Transcription factor RF2b	High-parental dominance
BC332_26363	Black	Serine/threonine-protein kinase CTR1	Over-dominance
BC332_11557	Black	Beta-1%2C3-galactosyltransferase 7	Over-dominance
novel.744	Black	Beta-glucosidase 9	Over-dominance
BC332_27203	Lightsheetblue1	Putative receptor-like protein kinase	Over-dominance
BC332_16203	Lightsheetblue1	ATP-dependent zinc metalloprotease FtsH 3	N/A

Non-additive Effect Role in Heterosis

Using RNA-seq analysis, our comparative transcriptome analysis revealed a subset of differential expression transcripts between the F_1 hybrid and its parents at the S-stage and F-stage. Several reports suggested that a great number of genes showed additive expression patterns in hybrids, with only a small proportion of genes having non-additive expression patterns (Hu et al., 2016; Shen et al., 2017), while it has also been reported that non-additive genes play a vital role in gene expression and heterosis (Zhang et al., 2008; Wei et al., 2009; Schnable and Springer, 2013; Yang et al., 2018). In this study, we have investigated that a majority proportion of differential expressed genes possessed a non-additive pattern in the F_1 hybrid from HNCc22 and HNCc16, suggesting that dominance models could be associated with growth vigor (Hedgecock et al., 2007).

Coexpression Facilitates the Identification of Traits-Related Key Genes

The weighted gene coexpression network analysis revealed the module hub genes related to the eigengene traits of pepper at two developmental stages. Here, we found that several modules have a significantly positive correlation between LA and seedling stem diameter. In the purple module, we found that β-glucosidase was highly expressed, supported by a recent report that β-glucosidase plays a crucial role in stomatal traits and photosynthesis (Zhou et al., 2021). We found that *BC332_18317* and *BC332_23046* were over-dominance expressed in the white and black modules, encoding auxin-responsive protein. KEGG pathway analysis also identified these genes in plant hormone regulation, indicating their possible roles in plant morphology construction and pepper growth at a specific developmental stage.

Plant Hormone Signal Transduction Contributing to Heterosis in Hybrid

In multicellular organisms, hormones have been shown to coordinate with cell division, expansion, differentiation, and stress (Malamy et al., 1990; Tan et al., 2007; Santiago et al., 2009; Kyndt et al., 2012; Davière and Achard, 2013; Zhai et al., 2013; Lavy and Estelle, 2016; Huang et al., 2017; Shani et al., 2017; Ding et al., 2018; Kunkel and Harper, 2018; Luo et al., 2018; Wang et al., 2020). This study found that many DEGs were over-dominant and related to plant hormone signal transduction including auxin, gibberellin, abscisic acid (ABA), ethylene, jasmonic acid, and salicylic acid. For example, members from ABA signal transduction pathway, *PYR/PYL* (ABA receptor group), were upregulated at S-stage and downregulated at F-stage, and PP2C were upregulated in the F_1 hybrid (Santiago et al., 2009), which showed the negative-feedback regulatory mechanism of ABA signal transduction, similar to the previous study on rice root

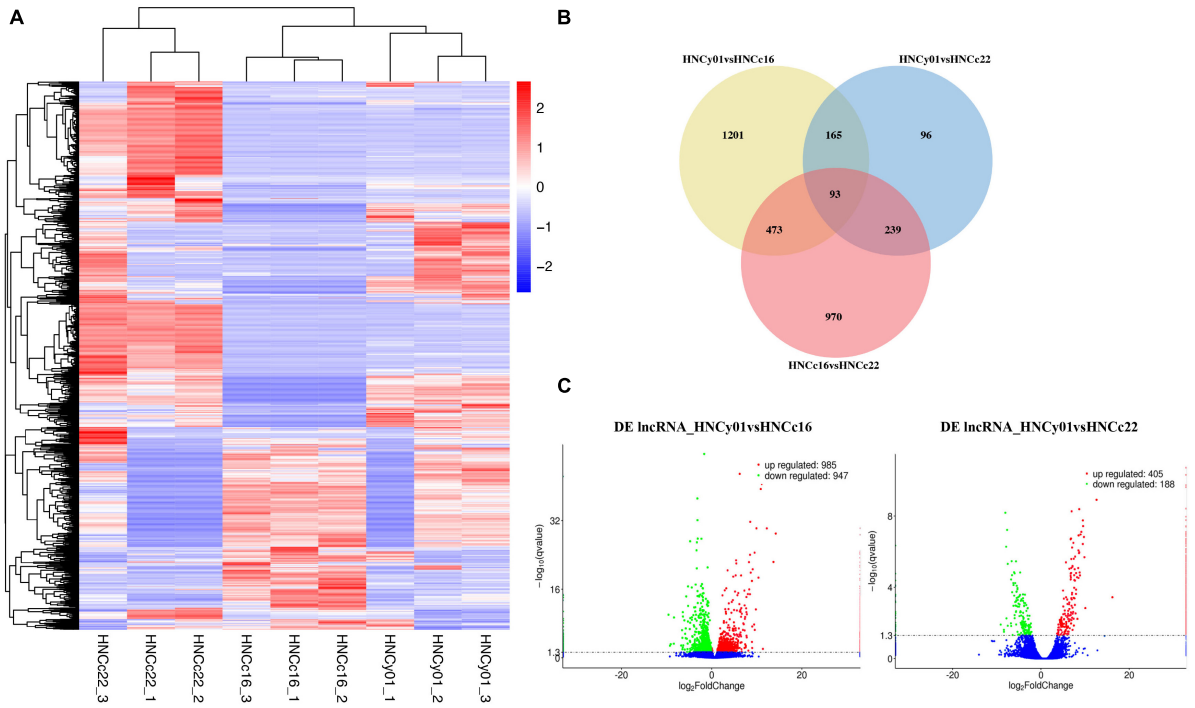


FIGURE 7 | Significantly differentially expressed lncRNAs between the hybrid and its parents at the seedling stage. **(A)** Hierarchical cluster analysis of DE lncRNAs. **(B)** Venn diagram of DE lncRNAs in HNCy01vsHNCc16, HNCy01vs HNCc22, and HNCc16vs HNCc22. **(C)** Volcano plot of DE lncRNAs in HNCy01vsHNCc16 and HNCy0vsHNCc22. Upregulated, downregulated, and non-differentially expressed genes are represented by red, green, and blue dots, respectively.

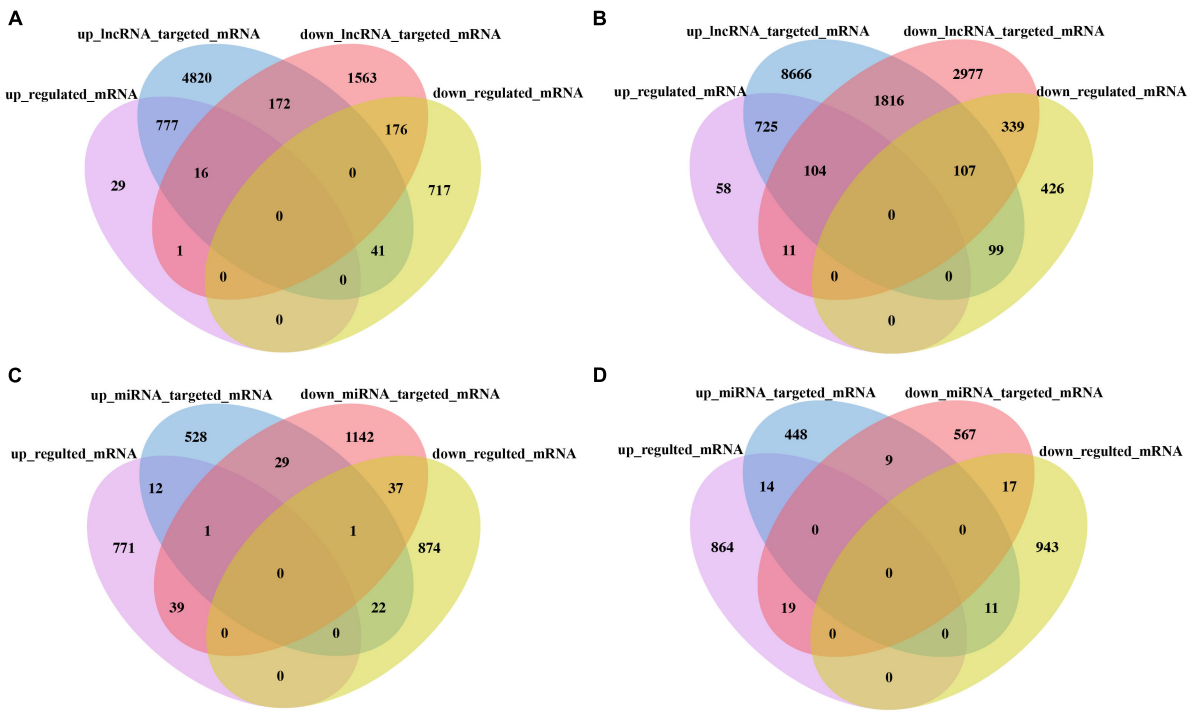


FIGURE 8 | Analysis of the targeted mRNA of lncRNA and miRNA between the hybrid and its parents. **(A)** Venn diagram of DE lncRNA targeted mRNA between HNCy01 and HNCc22. **(B)** Venn diagram of DE lncRNA targeted mRNA between HNCy01 and HNCc16. **(C)** Venn diagram of DE miRNA targeted mRNA between HNCy01 and HNCc22. **(D)** Venn diagram of DE miRNA targeted mRNA between HNCy01 and HNCc16.

TABLE 4 | The detailed information of DE circRNA and its targeted mRNA.

circRNA ID	circRNA regulation	DE targeted mRNA ID	mRNA regulation	Description
novel_circ_0001361	Up	BC332_22932	Up	Probable aminotransferase
novel_circ_0000872	Down	BC332_16958	Down	Chloroplast stem-loop binding protein
novel_circ_0001206	Down	BC332_22540	Down	Kinesin-like protein

heterosis (Zhai et al., 2013). TGAs, positive regulators of SA-induced PR genes (Ding et al., 2018), were upregulated in the F_1 hybrid in the salicylic acid pathway. And NPR protein as an SA receptor was upregulated at F-stage, suggesting that the hybrid may regulate plant immunity based on heterosis (Wang et al., 2020). In auxin signaling pathway, AUX/IAA and SAUR controlling plant growth by cell enlargement were upregulated in the F_1 hybrid. Moreover, previous research reported that the gibberellin (GA) signaling pathway gives impetus to the transformation of vegetative into reproduction growth and stress tolerance (Colebrook et al., 2014). DELLA proteins restrain plant growth and GID1 as a GA receptor. In the present study, GID1 was upregulated at the S-stage but downregulated at the F-stage, which indicates that GA signal promotes growth by overcoming DELLA-mediated growth restraint (Davière and Achard, 2013). We also identified that *BC332_21619* and *BC332_25121* belong to the bHLH family, which were over-dominant in both the S-stage and the F-stage. These results will be helpful for future studies on molecular mechanisms of hormones underlying pepper heterosis.

miRNA Roles in Heterosis

The miRNA is a kind of small RNA, which plays an important role in gene expression, defense responses, and cell function regulation in plants and animals (Zhang et al., 2013). Recent studies showed that several miRNAs showed non-additive expression and led to the non-additive expression of target genes affecting growth vigor and adaptability (Ng et al., 2012; Chen, 2013). Here, we also found plenty of DE miRNAs and their target DE mRNAs in the F_1 hybrid. It has been reported that miR156 may participate in flowering and abiotic stress through targeted gene regulation (Wang et al., 2009; Frazier et al., 2011; Khraiwesh et al., 2012). For example, the upregulation of miR156 may be beneficial for higher anthocyanin synthesis under drought stress (González-Villagra et al., 2017). Besides, it was also found that *GmmiR156b* might improve shoot architecture and yield in soybean (Sun et al., 2019; Liu et al., 2020b). We found that *miR156a* and *miR156d-5p* expressions were high-parental dominant in the pepper F_1 hybrid HNCy01. The miR169 family has been reported to be associated with ABA-responsive TFs (Ding et al., 2012; Song et al., 2018). In the allopolyploid wheat, *miR169.2* and *miR169.6* showed an expression pattern with low-parental ELD-ab. Similarly, in our pepper hybrid, miR169c was downregulated compared with both parents. The miR171 family was known to regulate chlorophyll biosynthesis and leaf growth by target TFs DELLA (Ma et al., 2014). The miR369 was reported to be involved in growth-regulating factors (GRF) silence, which may negatively regulate in disease resistance (Chandran et al., 2018). The other research reported that the overexpression of miR369 has a negative influence on PH, which may be associated with biomass yield (Liu et al., 2021), while in our study, miR369a-3p and miR369a-5p were downregulated in HNCy01 and HNCc22, which may have advantageous impacts on pepper growth. Based on these results, it is suggested that miRNA could play an important role in pepper heterosis.

LncRNA Function in Heterosis

Plant lncRNAs have strong relationships with abiotic and biotic responses, photomorphogenesis, flowering time regulation, etc.

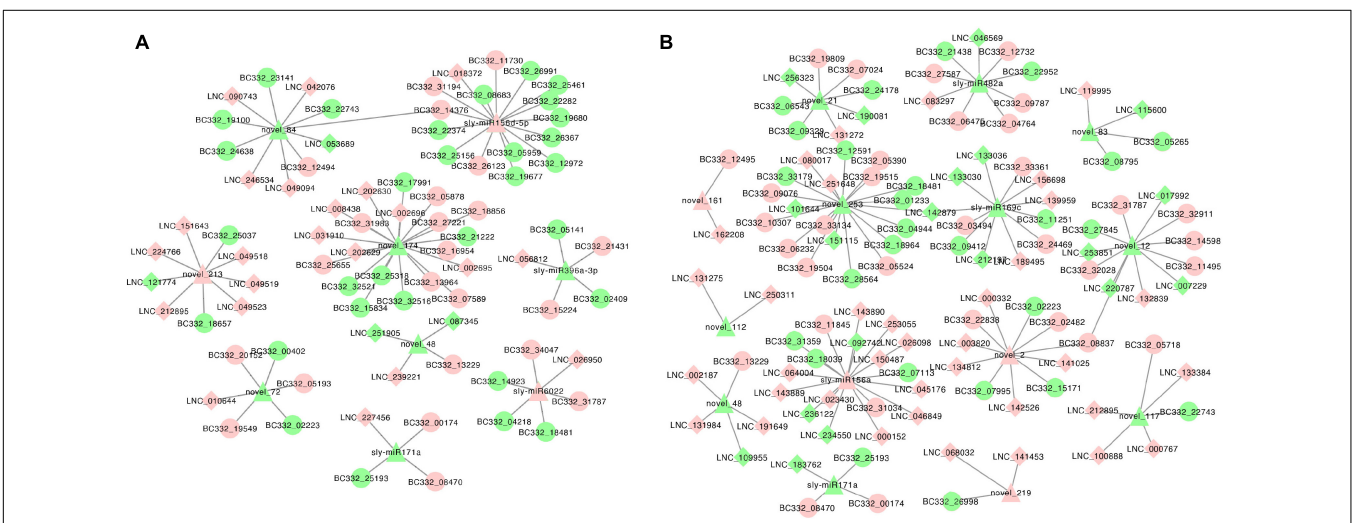


FIGURE 9 | The visualization of coexpression network in the hybrid and its parents. **(A)** lncRNA–miRNA–mRNA coexpression network between HNCy01 and HNCc22. **(B)** lncRNA–miRNA–mRNA coexpression network in HNCy01 vs. HNCc16. Red indicates upregulated, green indicates downregulated, ellipse indicates mRNA, diamond indicates lncRNA, and triangle indicates miRNA, respectively.

(Wang H. et al., 2014, Wang Y. et al., 2014, Wang Z.W. et al., 2014, Wang et al., 2015). In this study, we identified a great number of lncRNAs by strand-specific RNA-seq. In the F_1 hybrid HNCy01, more upregulated lncRNAs were enriched compared to downregulated lncRNAs. Meanwhile, the target mRNAs of many DE lncRNAs were also differentially expressed, suggesting their roles in pepper heterosis. Furthermore, we identified 74 upregulated lncRNAs in the plant–pathogen interaction pathway from HNCy01 vs. HNCc16. However, the functions of some plant lncRNAs, such as COOLAIR mediating epigenetic silence of the floral repressor FLOWERING LOCUS C, were described in detail (Zhao et al., 2018). Although large-scale lncRNAs were identified, it is still a challenge to reveal the function and mechanism of lncRNA, and the roles of lncRNAs in plant heterosis remain to be further clarified.

Transcription Factors Regulate the Heterosis of Pepper

Transcription factors, a group of DNA-binding proteins controlling gene transcription, play crucial roles in heterosis (Zhang et al., 2008; Guo et al., 2017; Chen et al., 2018). Numerous TFs have been reported to be correlated with plant development and resistance ability. For example, MYB TFs play a crucial role in various biological processes. Previous research demonstrates that the MYB family functioned in plant growth and development and even abiotic stress (Wang et al., 2011). bZIP TFs are involved in PH and flower development (Jakoby et al., 2002).

Furthermore, among these TF families, 690 TF genes were displaying high-parental dominance or over-dominance. ARF family genes, which are responsive to plant hormone transduction, were highly expressed in shoots (Liu et al., 2014). We found that ARF family genes had over-dominance expression at the S-stage and high-parental dominance expression at the F-stage. Besides, we identified 39 HSF TFs implementing important functions in response to high-temperature stress and plant development (Guo et al., 2015). In our study, 22 TFs displayed over-dominance or high-parental dominance, suggesting that these TFs may play a crucial factor in resistance ability.

The bHLH family plays a significant role in phytohormone signal and abiotic stress (Zhou et al., 2021). MYB TFs are key factors in cell proliferation and differentiation (Dubos et al., 2010). For the network analysis, we identified that bHLH and MYB TFs were overdominant in the white and black module, suggesting that these genes may have a relationship with pepper development, indicating their potential roles in heterosis (Ni et al., 2009).

CONCLUSION

This study used RNA-seq analysis to investigate the transcriptome of pepper hybrid and their parents at the S-stage and F-stage. We identified plenty of DEGs and novel ncRNAs in three genotypes. Comparing the significantly enriched gene between the F_1 hybrid and its parents, we have identified some candidate transcripts that may correlate with heterosis. This study will provide molecular resources for further interpreting the pepper heterosis mechanism.

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

Z-WW and H-YS: conceptualization and methodology. H-YS: software and writing—original draft preparation. H-YS, HZ, and H-LM: validation. H-YS, HZ, H-LM, Y-LJ, ZY, Y-YH, JZ, W-LB, S-HC, and G-PZ: investigation. Z-WW, H-YS, H-LM, Y-LJ, ZY, Y-YH, JZ, W-LB, S-HC, and G-PZ: writing—review and editing. Z-WW: supervision. All authors read and agreed to the published version of the manuscript.

FUNDING

This work was supported by a grant from the National Key Research and Development Program of China (2018YFD1000800) and Innovation Research Projects for Postgraduate in Hainan Province (Hys2019-05).

ACKNOWLEDGMENTS

We greatly thank the reviewers for their valuable comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.685788/full#supplementary-material

REFERENCES

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinform.* 15:293.

Chandran, V., Wang, H., Gao, F., Cao, X. L., Chen, Y. P., Li, G. B., et al. (2018). miR396-OsGRFs module balances growth and rice blast disease-resistance. *Front. Plant Sci.* 9:1999. doi: 10.3389/fpls.2018.01999

Chen, L., Bian, J., Shi, S., Yu, J., Khanzada, H., Wassan, G. M., et al. (2018). Genetic analysis for the grain number heterosis of a super-hybrid rice WFYT025 combination using RNA-Seq. *Rice (N Y)* 11:37. doi: 10.1186/s12284-018-0229-y

Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482. doi: 10.1038/nrg3503

Colebrook, E. H., Thomas, S. G., Phillips, A. L., and Hedden, P. (2014). The role of gibberellin signalling in plant responses to abiotic stress. *J. Exp. Biol.* 217(Pt 1), 67–75. doi: 10.1242/jeb.089938

Davière, J. M., and Achard, P. (2013). Gibberellin signaling in plants. *Development* 140, 1147–1151. doi: 10.1242/dev.087650

Ding, D., Wang, Y., Han, M., Fu, Z., Li, W., Liu, Z., et al. (2012). MicroRNA transcriptomic analysis of heterosis during maize seed germination. *PLoS One* 7:e39578. doi: 10.1371/journal.pone.0039578

Ding, H., Qin, C., Luo, X., Li, L., Chen, Z., Liu, H., et al. (2014). Heterosis in early maize ear inflorescence development: a genome-wide transcription analysis for two maize inbred lines and their hybrid. *Int. J. Mol. Sci.* 15, 13892–13915. doi: 10.3390/ijms150813892

Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y., Li, X., et al. (2018). Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity. *Cell* 173, 1454–1467.e15. doi: 10.1016/j.cell.2018.03.044

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

East, E. M. (1936). Heterosis. *Genetics* 21, 375–397.

FAO. (2018). Available online at: http://www.fao.org/statistics/en/

Frazier, T. P., Sun, G., Burklew, C. E., and Zhang, B. (2011). Salt and drought stresses induce the aberrant expression of microRNA genes in tobacco. *Mol. Biotechnol.* 49, 159–165. doi: 10.1007/s12033-011-9387-9385

Fujimoto, R., Taylor, J. M., Shirasawa, S., Peacock, W. J., and Dennis, E. S. (2012). Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci. U S A.* 109, 7109–7114. doi: 10.1073/pnas.1204464109

Garcia, A. A., Wang, S., Melchinger, A. E., and Zeng, Z. B. (2008). Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics* 180, 1707–1724. doi: 10.1534/genetics.107.082867

González-Villagra, J., Kurepin, L. V., and Reyes-Díaz, M. M. (2017). Evaluating the involvement and interaction of abscisic acid and miRNA156 in the induction of anthocyanin biosynthesis in drought-stressed plants. *Planta* 246, 299–312. doi: 10.1007/s00425-017-2711-y

Govindaraju, D. R. (2019). An elucidation of over a century old enigma in genetics-Heterosis. *PLoS Biol.* 17:e3000215. doi: 10.1371/journal.pbio.3000215

Groszmann, M., Gonzalez-Bayon, R., Greaves, I. K., Wang, L., Huen, A. K., Peacock, W. J., et al. (2014). Intraspecific *Arabidopsis* hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant Physiol.* 166, 265–280. doi: 10.1104/pp.114.243998

Guo, H., Mendrikahy, J. N., Xie, L., Deng, J., Lu, Z., Wu, J., et al. (2017). Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci. Rep.* 7:40139. doi: 10.1038/srep40139

Guo, M., Lu, J. P., Zhai, Y. F., Chai, W. G., Gong, Z. H., and Lu, M. H. (2015). Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum* L.). *BMC Plant Biol.* 15:151. doi: 10.1186/s12870-015-0512-517

Hedgecock, D., Lin, J. Z., DeCola, S., Haudenschild, C. D., Meyer, E., Manahan, D. T., et al. (2007). Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*). *Proc. Natl. Acad. Sci. U S A.* 104, 2313–2318. doi: 10.1073/pnas.0610880104

Hu, X., Wang, H., Diao, X., Liu, Z., Li, K., Wu, Y., et al. (2016). Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. *BMC Genomics* 17:959. doi: 10.1186/s12864-016-3296-3298

Huang, H., Liu, B., Liu, L., and Song, S. (2017). Jasmonate action in plant growth and development. *J. Exp. Bot.* 68, 1349–1359. doi: 10.1093/jxb/erw495

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* 7, 106–111. doi: 10.1016/s1360-1385(01)02223-2223

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

Kawanabe, T., Ishikura, S., Miyaji, N., Sasaki, T., Wu, L. M., Itabashi, E., et al. (2016). Role of DNA methylation in hybrid vigor in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U S A.* 113, E6704–E6711. doi: 10.1073/pnas.1613372113

Khraiwesh, B., Zhu, J. K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta* 1819, 137–148. doi: 10.1016/j.bbagrm.2011.05.001

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18:210. doi: 10.1186/s13059-017-1341-1349

Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877

Kothari, S. L., Joshi, A., Kachhwaha, S., and Ochoa-Alejo, N. (2010). Chilli peppers–a review on tissue culture and transgenesis. *Biotechnol. Adv.* 28, 35–48. doi: 10.1016/j.biotechadv.2009.08.005

Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene single flower truss drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463. doi: 10.1038/ng.550

Kunkel, B. N., and Harper, C. P. (2018). The roles of auxin during interactions between bacterial plant pathogens and their hosts. *J. Exp. Bot.* 69, 245–254. doi: 10.1093/jxb/erx447

Kyndt, T., Denil, S., Haegeman, A., Trooskens, G., De Meyer, T., Van Criekinge, W., et al. (2012). Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. *J. Exp. Bot.* 63, 2141–2157. doi: 10.1093/jxb/err435

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

Larièpe, A., Mangin, B., Jasson, S., Combes, V., Dumas, F., Jamin, P., et al. (2012). The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). *Genetics* 190, 795–811. doi: 10.1534/genetics.111.133447

Lauss, K., Wardenaar, R., Oka, R., Hulten, M. H. A. V., and Johannes, F. (2017). Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiol.* 176, 1627–1645. doi: 10.1104/pp.17.01054

Lavy, M., and Estelle, M. (2016). Mechanisms of auxin signaling. *Development* 143, 3226–3229. doi: 10.1242/dev.131870

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., et al. (2014). mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26, 1878–1900. doi: 10.1105/tpc.114.124388

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. U S A.* 113, E6026–E6035. doi: 10.1073/pnas.1610115113

Liu, M. J., Zhao, J., Cai, Q. L., Liu, G. C., Wang, J. R., Zhao, Z. H., et al. (2014). The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* 5:5315. doi: 10.1038/ncomms6315

Liu, P. C., Peacock, W. J., Wang, L., Furbank, R., Larkum, A., and Dennis, E. S. (2020a). Leaf growth in early development is key to biomass heterosis in *Arabidopsis*. *J. Exp. Bot.* 71, 2439–2450. doi: 10.1093/jxb/eraa006

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

Liu, Y., Yan, J., Wang, K., Li, D., Yang, R., Luo, H., et al. (2021). MiR396-GRF module associates with switchgrass biomass yield and feedstock quality. *Plant Biotechnol J.* doi: 10.1111/pbi.13567 Online ahead of print.

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/s13059-014-0550-558

Luo, J., Zhou, J. J., and Zhang, J. Z. (2018). Aux/IAA gene family in plants: molecular structure, regulation, and function. *Int. J. Mol. Sci.* 19:259. doi: 10.3390/ijms19010259

Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., et al. (2014). Arabidopsis miR171-targeted scarecrow-like proteins bind to GT cis-elements and mediate gibberellin-regulated chlorophyll biosynthesis under light conditions. *PLoS Genet* 10:e1004519. doi: 10.1371/journal.pgen.1004519

Malamy, J., Carr, J. P., Klessig, D. F., and Raskin, I. (1990). Salicylic Acid: a likely endogenous signal in the resistance response of tobacco to viral infection. *Science* 250, 1002–1004. doi: 10.1126/science.250.4983.1002

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

Melchinger, A. E., Piepho, H. P., Utz, H. F., Muminovic, J., Wegenast, T., Törjék, O., et al. (2007). Genetic basis of heterosis for growth-related traits in

Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics* 177, 1827–1837. doi: 10.1534/genetics.107.080564

Ng, D. W., Lu, J., and Chen, Z. J. (2012). Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* 15, 154–161. doi: 10.1016/j.pbi.2012.01.007

Ni, Z., Kim, E. D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457, 327–331. doi: 10.1038/nature07523

Pérez-Rodríguez, P., Riaño-Pachón, D. M., Corrêa, L. G., Rensing, S. A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38, D822–D827. doi: 10.1093/nar/gkp805

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Powers, L. (1944). An expansion of Jones's theory for the explanation of heterosis. *Am. Nat.* 78, 275–280. doi: 10.1086/281199

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl. Acad. Sci. U S A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S. Y., Jamin, M., et al. (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* 462, 665–668. doi: 10.1038/nature08591

Schnable, P. S., and Springer, N. M. (2013). Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64, 71–88. doi: 10.1146/annurev-arplant-042110-103827

Shani, E., Salehin, M., Zhang, Y., Sanchez, S. E., Doherty, C., Wang, R., et al. (2017). Plant stress tolerance requires auxin-sensitive Aux/IAA transcriptional repressors. *Curr. Biol.* 27, 437–444. doi: 10.1016/j.cub.2016.12.016

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. U S A.* 116, 5653–5658. doi: 10.1073/pnas.1820513116

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C. Y., Cai, D., et al. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *Plant J.* 91, 874–893. doi: 10.1111/tpj.13605

Song, S., Xu, Y., Huang, D., Ashraf, M. A., Li, J., Hu, W., et al. (2018). Identification and characterization of miRNA169 family members in banana (*Musa acuminata* L.) that respond to *Fusarium oxysporum* f. sp. cubense infection in banana cultivars. *PeerJ* 6:e6209. doi: 10.7717/peerj.6209

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

Tan, X., Calderon-Villalobos, L. I., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., et al. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* 446, 640–645. doi: 10.1038/nature05731

Wang, D., Pan, Y., Zhao, X., Zhu, L., Fu, B., and Li, Z. (2011). Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *BMC Genomics* 12:149. doi: 10.1186/1471-2164-12-149

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., et al. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24, 444–453. doi: 10.1101/gr.165555.113

Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., et al. (2014). Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci. U S A.* 111, 10359–10364. doi: 10.1073/pnas.1409457111

Wang, Z. W., Wu, Z., Raitskin, O., Sun, Q., and Dean, C. (2014). Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proc. Natl. Acad. Sci. U S A.* 111, 7468–7473. doi: 10.1073/pnas.1406635111

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, T. Z., Liu, M., Zhao, M. G., Chen, R., and Zhang, W. H. (2015). Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. *BMC Plant Biol.* 15:131. doi: 10.1186/s12870-015-0530-535

Wang, W., Withers, J., Li, H., Zwack, P. J., Rusnac, D. V., Shi, H., et al. (2020). Structural basis of salicylic acid perception by *Arabidopsis* NPR proteins. *Nature* 586, 311–316. doi: 10.1038/s41586-020-2596-y

Wei, G., Tao, Y., Liu, G., Chen, C., Luo, R., Xia, H., et al. (2009). A transcriptomic analysis of superhybrid rice LYP9 and its parents. *Proc. Natl. Acad. Sci. U S A.* 106, 7695–7701. doi: 10.1073/pnas.0902340106

Xiao, J., Li, J., Yuan, L., and Tanksley, S. D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* 140, 745–754. doi: 10.1093/genetics/140.2.745

Yang, H., Wang, X., Wei, Y., Deng, Z., Liu, H., Chen, J., et al. (2018). Transcriptomic analyses reveal molecular mechanisms underlying growth heterosis and weakness of rubber tree seedlings. *BMC Plant Biol.* 18:10. doi: 10.1186/s12870-017-1203-1203

Yang, M., Wang, X., Ren, D., Huang, H., Xu, M., He, G., et al. (2017). Genomic architecture of biomass heterosis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U S A.* 114, 8101–8106. doi: 10.1073/pnas.1705423114

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, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., et al. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U S A.* 94, 9226–9231. doi: 10.1073/pnas.94.17.9226

Zhai, R., Feng, Y., Wang, H., Zhan, X., Shen, X., Wu, W., et al. (2013). Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics* 14:19. doi: 10.1186/1471-2164-14-19

Zhang, H. Y., He, H., Chen, L. B., Li, L., Liang, M. Z., Wang, X. F., et al. (2008). A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol. Plant* 1, 720–731. doi: 10.1093/mp/ssn022

Zhang, Y. C., Yu, Y., Wang, C. Y., Li, Z. Y., Liu, Q., Xu, J., et al. (2013). Overexpression of microRNA OsmiR397 improves rice yield by increasing grain size and promoting panicle branching. *Nat. Biotechnol.* 31, 848–852. doi: 10.1038/nbt.2646

Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y. (2018). Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* 9:5056. doi: 10.1038/s41467-018-07500-7507

Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., et al. (2012). Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U S A.* 109, 15847–15852. doi: 10.1073/pnas.1214141109

Zhou, J., Chen, S., Shi, W., David-Schwartz, R., Li, S., Yang, F., et al. (2021). Transcriptome profiling reveals the effects of drought tolerance in Giant Juncao. *BMC Plant Biol.* 21:2. doi: 10.1186/s12870-020-02785-2787

Zhou, L., Chen, J., Li, Z., Li, X., Hu, X., Huang, Y., et al. (2010). Integrated profiling of microRNAs and mRNAs: microRNAs located on Xq27.3 associate with clear cell renal cell carcinoma. *PLoS One* 5:e15224. doi: 10.1371/journal.pone.0015224

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.

Copyright © 2021 Shu, Zhou, Mu, Wu, Jiang, Yang, Hao, Zhu, Bao, Cheng, Zhu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 August 2021
doi: 10.3389/fgene.2021.716163

Transcriptome and Metabolome Analyses of the Flowers and Leaves of *Chrysanthemum dichrum*

Hua Liu[1†], Xiaoxi Chen[1,2†], Haixia Chen[1,3], Jie Lu[4], Dongliang Chen[1], Chang Luo[1], Xi Cheng[1], Yin Jia[2] and Conglin Huang[1*]*

[1] Beijing Key Laboratory of Agricultural Genetic Resources and Biotechnology, Beijing Agro-Biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing Engineering Research Center of Functional Floriculture, Beijing, China, [2] Sichuan Agricultural University, College of Landscape Architecture, Chengdu, China, [3] Beijing University of Agriculture, College of Landscape Architecture, Beijing, China, [4] Shandong Forestry Protection and Development Service Center, Jinan, China

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Cheng Qin,
Zunyi Vocational and Technical College, China
Youlu Yuan,
Cotton Research Institute, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Conglin Huang
conglinh@126.com
Yin Jia
jiayin_cn@163.com

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

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

Received: 28 May 2021
Accepted: 27 July 2021
Published: 31 August 2021

Citation:
Liu H, Chen X, Chen H, Lu J, Chen D, Luo C, Cheng X, Jia Y and Huang C (2021) Transcriptome and Metabolome Analyses of the Flowers and Leaves of Chrysanthemum dichrum. Front. Genet. 12:716163. doi: 10.3389/fgene.2021.716163

Chrysanthemum dichrum is an important wild species in the family Asteraceae. However, because of a lack of genetic information, there has been relatively little research conducted on the molecular mechanisms in *C. dichrum*. There is no report describing the transcriptome and metabolome of *C. dichrum* flowers and leaves at different developmental stages. In this study, high-throughput sequencing and RNA-seq analyses were used to investigate the transcriptome of *C. dichrum* leaves, flower buds, and blooming flowers. Additionally, these three tissues also underwent a metabolomics analysis. A total of 447,313,764 clean reads were assembled into 77,683 unigenes, with an average length of 839 bp. Of the 44,204 annotated unigenes, 42,189, 28,531, 23,420, and 17,599 were annotated using the Nr, Swiss-Prot, KOG, and KEGG databases, respectively. Furthermore, 31,848 differentially expressed genes (DEGs) were detected between the leaves and flower buds, whereas 23,197 DEGs were detected between the leaves and blooming flowers, and 11,240 DEGs were detected between the flower buds and blooming flowers. Finally, a quantitative real-time Polymerase Chain Reaction (qRT-PCR) assay was conducted to validate the identified DEGs. The metabolome data revealed several abundant metabolites in *C. dichrum* leaves, flower buds, and blooming flowers, including raffinose, 1-kestose, asparagine, glutamine, and other medicinal compounds. The expression patterns of significant DEGs revealed by the transcriptome analysis as well as the data for the differentially abundant metabolites in three *C. dichrum* tissues provide important genetic and metabolic information relevant for future investigations of the molecular mechanisms in *C. dichrum*. Moreover, the results of this study may be useful for the molecular breeding, development, and application of *C. dichrum* resources.

Keywords: *Chrysanthemum dichrum*, transcriptome, RNA-Seq, metabolites, metabolomics

INTRODUCTION

Chrysanthemum dichrum, which is a perennial herb belonging to the genus *Chrysanthemum* in the family Asteraceae, is a very important germplasm resource. It is a wild relative of the famous ornamental plant *Chrysanthemum morifolium*. A *C. dichrum* plant can be more than 30 cm tall. Its main stem, which lies prone or is slanted, is bare, brown, and has multiple branches on the upper part, which contains densely appressed pubescent leaves. It is a highly edible plant with medicinal value and is used for landscaping as a garden greening material. *Chrysanthemum dichrum* is distributed in southwestern Hebei province in China (Hsianshiu, 1993). However, it can be introduced to other regions relatively easily, making it a wild flower species potentially useful as a functional flower crop.

Current research on *C. dichrum* mainly focuses on cross breeding and stress resistance, including tolerance to cold and drought. Previous studies proved that *C. dichrum*, *Chrysanthemum indicum* "Nankingense," and other interspecific hybrid progeny are highly tolerant to drought conditions (Li et al., 2020). Chi et al. in order to solve the economic benefit and ornamental quality decrease caused by low temperature in northern China. The genetic control mechanism of interspecific cold tolerance between *C. dichrum* and *C. nankingense* was clarified by analyzing the cold-resistant traits such as Low Semi-Lethal Temperature (LT50), basal shoots height and number in the basal shoots stage of the hybrids (Chi et al., 2018). A research team found a 1,474 bp stress-induced CdDREBa promoter in *C. dichrum*. Identified a stress-inducible CdDREBa promoter from *C. dichrum*, Clothes several candidate stress-related *cis*-acting elements (MYC-box, MYB site, GT-1, And W-box) within it. CdDREBa promoter features a strong low temperature- and right-inducible promoter. This provides an attractive target for engineering inducible promoters in transgenic crops (Chen Y. et al., 2012). Additionally, the *CdICE1* gene in *C. dichrum* was cloned and an analysis of its cold tolerance-related function in *Arabidopsis thaliana* revealed that the miR398-CSD pathway is involved in the induction of freezing resistance (Chen Y. et al., 2013). Earlier research also confirmed that the structural expression of *ICE1* gene in *C. dichrum* in large Chrysanthemum improves the tolerance to low temperature, salt and drought. *CdICE1* represents a promising candidate for a biotechnological approach to improve the level of crop abiotic stress tolerance (Chen L. et al., 2012). Other studies applied RNA-seq technology to investigate chrysanthemum and related species (Chen et al., 2009; Wang et al., 2013; Xu et al., 2013; Ren et al., 2014; Hong et al., 2015; Liu et al., 2015), but the *C. dichrum* transcriptome and metabolome have been rarely examined. Accordingly, research regarding the *C. dichrum* transcriptome may increase our understanding of the gene regulatory mechanism and biological pathways in this wild species, while also identifying the downstream target genes of key transcription factors to clarify specific processes (Liu et al., 2013). Therefore, in this study, we analyzed the transcriptome and metabolome of *C. dichrum* leaves, flower buds, and blooming flowers. Moreover, we identified the genes differentially expressed in the examined plant tissues. We also determined the metabolite composition and content in the leaves, flower buds, and blooming flowers and identified some highly abundant economically valuable metabolites with substantial developmental potential. The findings of this study lay the foundation for future investigations regarding chrysanthemum resource development as well as molecular research and breeding related to *C. dichrum*.

RESULTS

Illumina Sequencing and Assembly

This study involved analyses of *C. dichrum* leaves, flower buds, and blooming flowers (**Figure 1**). The transcriptomes of these plant tissues were sequenced using the Illumina HiSeq^{TM} 4000 platform. A total of 445,396,894 clean reads were obtained and 77,683 unigene sequences were assembled using *de novo* assembly technology. The N50 was 1,413 bp and the maximum and minimum lengths were 14,459 and 201 bp, respectively, with an average length of 839 bp (**Table 1**).

Gene Annotation and Functional Classification

Of the 77,683 unigenes, 44,204 were annotated. More specifically, 42,189 unigenes were annotated using the Nr database, 28,531 unigenes were annotated using the Swiss-Prot database, 23,420 unigenes were annotated using the KOG database, and 17,599 unigenes were annotated using the KEGG database. A total of 33,479 unigenes were not been annotated (**Figure 2**).

We further analyzed the annotated genes regarding their KOG classifications. As indicated in **Figure 3**, of the 25 KOG database functional categories, "General function prediction only" (5,862 unigenes), "Signal transduction mechanisms" (4,961 unigenes), and "Posttranslational modification, protein turnover, chaperones" (3,970 unigenes) had the most unigenes. In contrast, "Extracellular structures" (119 unigenes), "Nuclear structure" (105 unigenes), and "Cell motility" (16 unigenes) had the fewest unigenes.

Regarding the GO analysis, 55,748 unigenes were classified into 50 groups in the three main categories. The biological process category had the most unigenes (27,682). "Metabolic process" and "Cellular process" were the main GO terms in this category, with 6,476 and 5,866 unigenes, respectively. A total of 16,466 unigenes were classified in the cell component category, with most annotated with the "Cells" (3,756 unigenes) and "Cell parts" (3,754 unigenes) GO terms. The molecular function category included 11,599 annotated unigenes, with "Catalytic activity" (6,303 unigenes) and "Binding" (4,296 unigenes) representing the main terms. Additionally, of the GO terms in the biological process, cell component, and molecular function categories, "Locomotion," "Nucleoid," and "Translation regulator activity" were assigned to the fewest genes, respectively (**Figure 4**).

The enriched KEGG pathways among the annotated unigenes were also determined. A total of 7,696 unigenes were associated with 134 pathways (**Table 2**), with "Metabolic pathways" assigned the most unigenes (3,480, 38.66%), followed

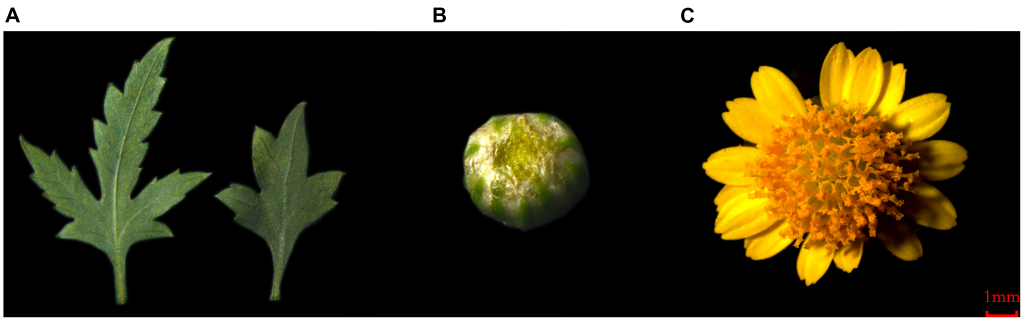


FIGURE 1 | The plant material of *C. dichrum*. **(A)** *C. dichrum* leaves. **(B)** *C. dichrum* flower bud. **(C)** *C. dichrum* blooming flower.

TABLE 1 | *De novo* sequence assembly results.

Genes Num	GC%	N50 (bp)	Max length (bp)	Min length (bp)	Average length (bp)	Total assembled bases
77683	39.4979	1,413	14,459	201	839	65,184,963

by "Biosynthesis of secondary metabolites" (1,914, 21.26%). Notably, "Glycosphingolipid biosynthesis-lacto and neolacto series" (2, 0.02%) and "Anthocyanin biosynthesis" (1, 0.01%) were assigned the fewest unigenes.

Analysis of Differentially Expressed Genes in *Chrysanthemum dichrum*

We used a false discovery rate (FDR) < 0.05 and a $|\log_2(\text{fold-change})| > 1$ as the thresholds for identifying differentially expressed genes (DEGs). A total of 31,848 genes were differentially expressed between the *C. dichrum* leaves and flower buds. Compared with the leaf expression levels, 18,718 and 13,130 DEGs had significantly up-regulated and down-regulated expression levels in the flower buds. A total of 23,197 DEGs were identified between the leaves and the blooming flowers. Compared with the corresponding expression during the blooming period, 13,640 and 9,557 DEGs had significantly up-regulated and down-regulated expression levels in the leaves. A comparison between the flower buds and blooming flowers revealed 11,240 DEGs, of which 4,436 and 6,804 DEGs had significantly up-regulated and down-regulated expression levels in the flower buds (**Figure 5**).

The DEGs underwent GO and KEGG pathway enrichment analyses to determine the differences in biological processes and pathways among the leaves, flower buds, and blooming flowers. The DEGs between the leaves and flower buds were annotated with 47 GO terms. In the biological process category, the main GO terms were "Metabolic process" (GO:0008152), "Cellular process" (GO:0009987), and "Single-organism process" (GO:0044699), which were assigned to 3,146, 2,729, and 940 DEGs, respectively. In the cellular component category, the most represented GO terms were "Cell" (GO:0005623), "Cell part" (GO:0044464), and "Organelle" (GO:0043226), which were assigned to 1,706, 1,706, and 1,056 DEGs, respectively. The most common molecular function GO terms among the DEGs were "Catalytic activity" (GO:0003824), "Binding" (GO:0005488), and "Transporter activity" (GO:0005215), which were assigned to 3,142, 2,008, and 284 DEGs, respectively. A total of 134 KEGG pathways were enriched among 4,253 DEGs. The main enriched KEGG pathways were "Metabolic pathways" (ko01100), "Biosynthesis of secondary metabolites" (ko01110), and "Ribosome" (ko03010), which were associated with 1,853, 1,063, and 364 DEGs, respectively (**Figure 6A**).

The DEGs between the leaves and blooming flowers were annotated with 46 GO terms. In the biological process category, the main GO terms were "Metabolic process" (GO:0008152), "Cellular process" (GO:0009987), and "Single-organism process" (GO:0044699), which were assigned to 2,511, 2,187, and 1,716 DEGs, respectively. In the cellular component category, the most represented GO terms were "Cell" (GO:0005623), "Cell part" (GO:0044464), and "Membrane" (GO:0016020), which were assigned to 1,333, 1,333, and 867 DEGs, respectively. The most common GO terms in the molecular function category were "Catalytic activity" (GO:0003824), "Binding" (GO:0005488), and "Transporter activity" (GO:0005215), which were assigned to 2,555, 1,631, and 221 DEGs, respectively. A total of 132 KEGG pathways were enriched among 3,373 DEGs. The main pathways were "Metabolic pathways" (ko01100), "Biosynthesis of secondary metabolites" (ko01110), and "Ribosome" (ko03010), which were associated with 1,477, 843, and 270 DEGs, respectively (**Figure 6B**).

The DEGs between the flower buds and blooming flowers were annotated with 45 GO terms. In the biological process category, the main GO terms were "Metabolic process" (GO:0008152), "Cellular process" (GO:0009987), and "Single-organism process" (GO:0044699), which were assigned to 1,008, 959, and 810 DEGs, respectively. In the cellular component category, the most represented GO terms were "Cell" (GO:0005623), "Cell part" (GO:0044464), and "Membrane" (GO:0016020), which were assigned to 433, 433, and 412 DEGs, respectively. The most common molecular function GO terms were "Catalytic activity" (GO:0003824), "Binding" (GO:0005488), and "Transporter activity" (GO:0005215), which were assigned

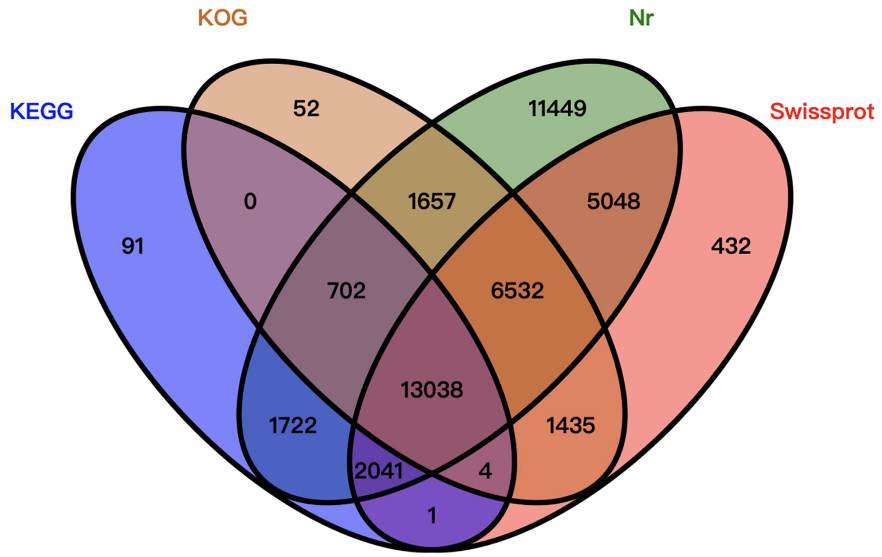


FIGURE 2 | Venn diagram summarizing the unigene annotations based on four databases.

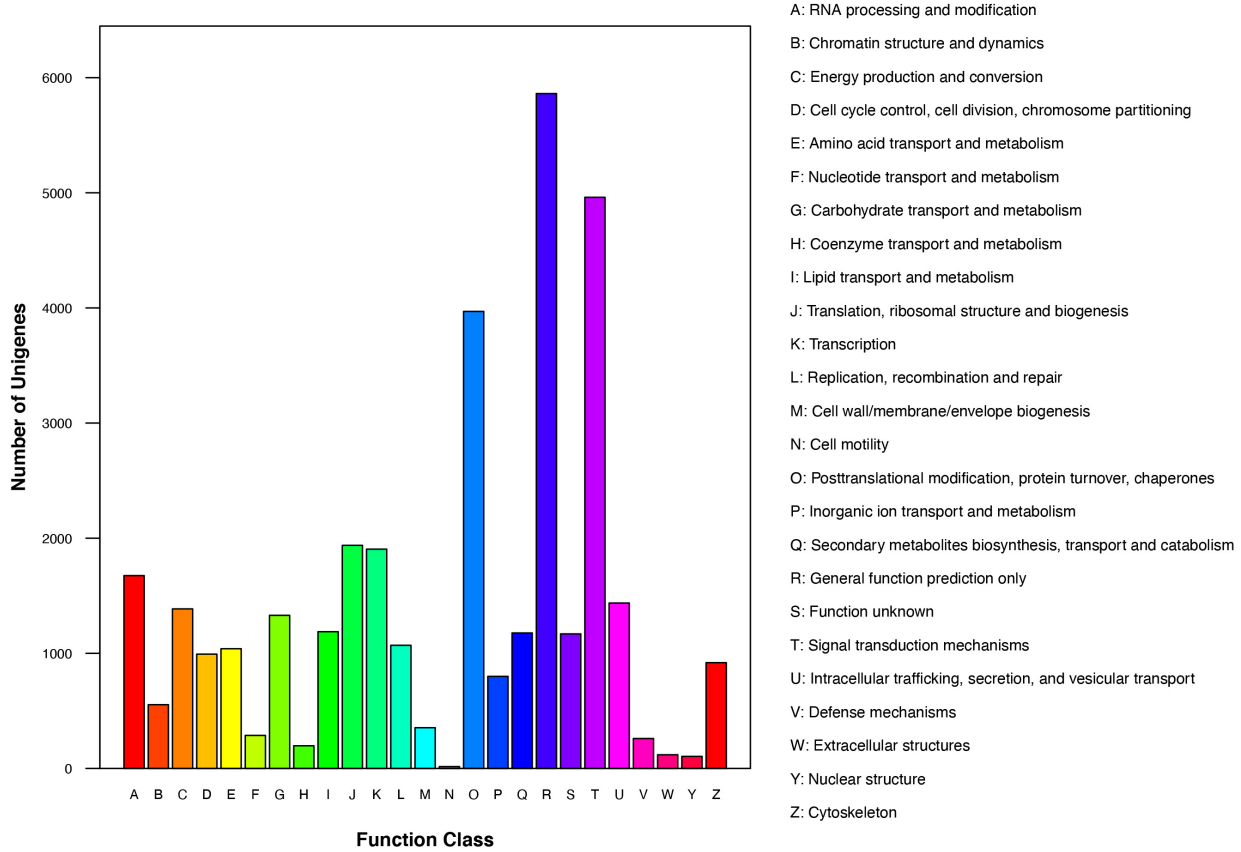


FIGURE 3 | KOG functional classification of *Chrysanthemum dichrum* unigenes.

to 1,144, 672, and 136 DEGs, respectively. A total of 128 KEGG pathways were enriched among 1,457 DEGs. The main pathways were "Metabolic pathways" (ko01100), "Biosynthesis of secondary metabolites" (ko01110), and "Plant–pathogen interaction" (ko04626), which were associated with 649, 374, and 118 DEGs, respectively (**Figure 6C**).

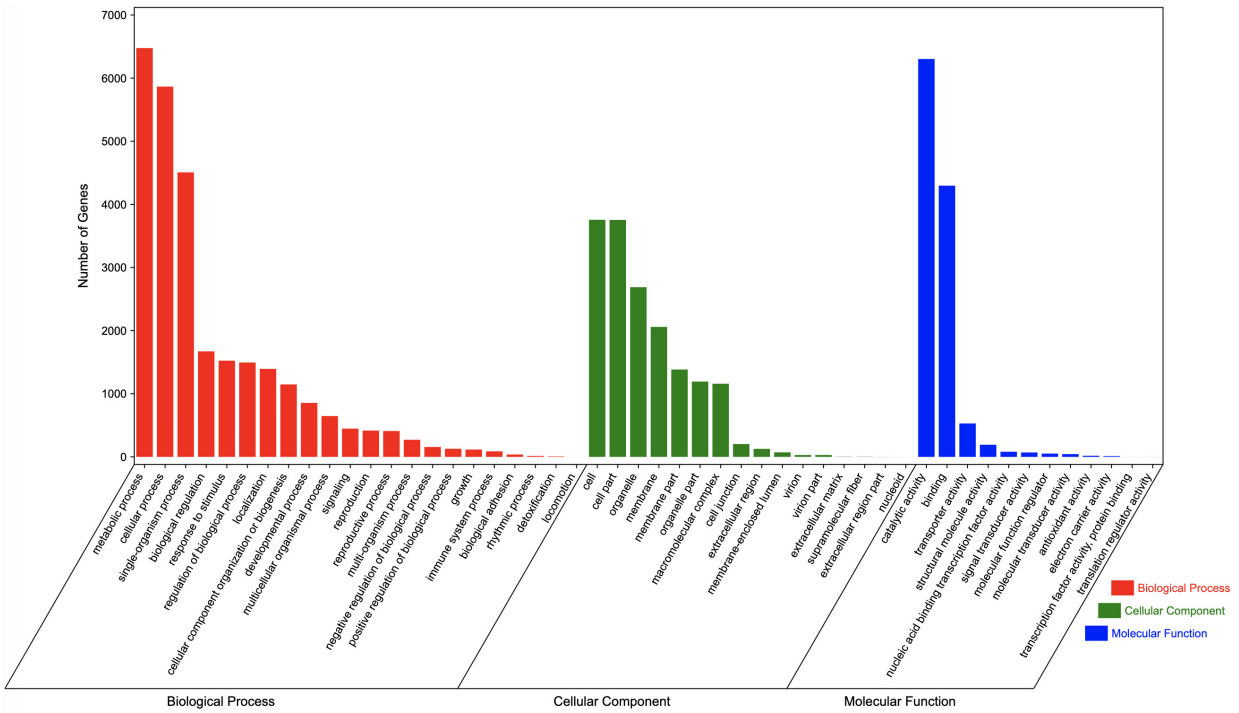


FIGURE 4 | GO terms assigned to *Chrysanthemum dichrum* unigenes.

Metabolomic Analysis

Multivariate Statistical Analysis of the Metabolome

A principal component analysis (PCA) can generally reflect the overall metabolic differences between samples in each group and the degree of variability between samples within a group. The first and second principal components (PC1 and PC2, respectively) of the sample group data for the *C. dichrum* leaves, flower buds, and blooming flowers accounted for 54.8 and 28.4% of the total variability, respectively (**Figure 7**). These results reflect relatively large overall metabolic differences, but low diversity between samples.

Qualitative and quantitative data were obtained for 568 metabolites in *C. dichrum* leaves, flower buds, and blooming flowers. Additionally, 216 annotated substances were highly abundant in the three plant tissues, including m-cresol, phytosphingosine, tartaric acid, abietic acid, neohesperidin, oxalic acid, salicylic acid, xylitol, glutamine, asparagine, 1-kestose, and other heterochromia-related metabolites. These compounds have diverse uses in medicines, skin care products, and food additives, indicative of the importance and utility of *C. dichrum*.

We screened 48 metabolites with significant differences between the *C. dichrum* leaves and flower buds. Of these metabolites, 22 and 26 were, respectively, more and less abundant in flower buds than in leaves. A total of 49 metabolites differed significantly between the leaves and blooming flowers, of which 28 and 21 metabolites were respectively more and less abundant in blooming flowers than in leaves. A comparison between the flower buds and blooming flowers revealed 22 significantly different metabolites, with all but one more abundant in the blooming flowers than in the flower buds (**Figure 8**).

Clustering Analysis of Differentially Abundant Metabolites

We compared the three tissue samples in groups, normalized the data for the differentially abundant metabolites, and conducted a cluster analysis. The following metabolites were more abundant in leaves than in flower buds: aminooxyacetic acid, alpha-ketoglutaric acid, melezitose, cellobiose, DL-dihydrosphingosine, galactinol, lactulose, conduritol, D-glyceric acid, glycerol, 2-hydroxypyridine, threonic acid, 1-kestose, proline, alanine, and palmitic acid. The opposite pattern was detected for the following metabolites: raffinose, isopropyl-beta-D-thiogalactopyranoside, methylmalonic acid, 3-cyanoalanine, glucose, methyl, phosphate, L-malic acid, glutamine, threonine, fumaric acid, oxoproline, asparagine, xylose, isoleucine, and aspartic acid (**Figure 9A**).

The following metabolites were more abundant in blooming flowers than in leaves: raffinose, isopropyl-beta-D-thiogalactopyranoside, leucine, lysine, methylmalonic acid, 3-cyanoalanine, phenylalanine, glucose, ribose, phosphate, L-malic acid, threonine, asparagine, oxoproline, xylose, serine, isoleucine, ethanolamine, aspartic acid, and valine. The opposite pattern was detected for the following metabolites: aminooxyacetic, alpha-ketoglutaric, melezitose, cellobiose, DL-dihydrosphingosine, galactinol, lactulose, D-glyceric, glycerol, and proline (**Figure 9B**).

Only one metabolite was more abundant in flower buds than in blooming flowers, but it was not identified. However, the

TABLE 2 | Enriched KEGG pathways among *Chrysanthemum dichrum* unigenes.

KEGG Categories	Unigene number	Pathway ID
Metabolic pathways	3480 (38.66%)	ko01100
Biosynthesis of secondary metabolites	1914 (21.26%)	ko01110
Ribosome	778 (8.64%)	ko03010
Carbon metabolism	570 (6.33%)	ko01200
Plant-pathogen interaction	503 (5.59%)	ko04626
Protein processing in endoplasmic reticulum	459 (5.1%)	ko04141
Biosynthesis of amino acids	444 (4.93%)	ko01230
Plant hormone signal transduction	411 (4.57%)	ko04075
Spliceosome	397 (4.41%)	ko03040
Endocytosis	360 (4%)	ko04144
RNA transport	320 (3.55%)	ko03013
Oxidative phosphorylation	285 (3.17%)	ko00190
Purine metabolism	285 (3.17%)	ko00230
Glycolysis / Gluconeogenesis	275 (3.05%)	ko00010
Ubiquitin mediated proteolysis	267 (2.97%)	ko04120
MAPK signaling pathway–plant	227 (2.52%)	ko04016
Starch and sucrose metabolism	224 (2.49%)	ko00500
mRNA surveillance pathway	224 (2.49%)	ko03015
Pyrimidine metabolism	213 (2.37%)	ko00240
Aminoacyl-tRNA biosynthesis	213 (2.37%)	ko00970
Phenylpropanoid biosynthesis	211 (2.34%)	ko00940
RNA degradation	202 (2.24%)	ko03018
Pyruvate metabolism	200 (2.22%)	ko00620
Amino sugar and nucleotide sugar metabolism	187 (2.08%)	ko00520
Phagosome	185 (2.06%)	ko04145
Carbon fixation in photosynthetic organisms	181 (2.01%)	ko00710
Cysteine and methionine metabolism	168 (1.87%)	ko00270
Ribosome biogenesis in eukaryotes	168 (1.87%)	ko03008
Glutathione metabolism	166 (1.84%)	ko00480
Glycerophospholipid metabolism	164 (1.82%)	ko00564
Glyoxylate and dicarboxylate metabolism	164 (1.82%)	ko00630
Fatty acid metabolism	163 (1.81%)	ko01212
Nucleotide excision repair	163 (1.81%)	ko03420
Homologous recombination	162 (1.8%)	ko03440
Peroxisome	161 (1.79%)	ko04146

following metabolites were determined to be more abundant in blooming flowers than in flower buds: 3,6-anhydro-D-galactose, isopropyl-beta-D-thiogalactopyranoside, leucine, lysine, gluconic acid, phenylalanine, glucose, threonine, glutamic acid, oxoproline, lactic acid, 1-kestose, xylose, serine, isoleucine, aspartic acid, valine, and palmitic acid (**Figure 9C**).

KEGG Annotation and KO Enrichment Analysis

On the basis of the KEGG pathway enrichment analysis, 296 metabolites were assigned to 17 metabolic pathways. The most represented pathway was the "Global and overview " (51 metabolites), followed by "Carbohydrate metabolism" (32 metabolites) and "Amino acid metabolism" (18 metabolites). The least annotated metabolic pathway was "Metabolism of terpenoids and polyketides" (1 metabolite) (**Figure 10**).

The four KO pathways with the most differentially abundant metabolites and the smallest *P/Q* value between the leaves and flower buds were arginine biosynthesis, glycerolipid metabolism, citrate cycle (TCA cycle), and carbon metabolism (**Figure 11A**). Additionally, the four KO pathways with the most differentially abundant metabolites and the smallest *P/Q* value between the leaves and blooming flowers were glycerolipid metabolism, glyoxylate and dicarboxylate metabolism, ABC transporters, and carbon metabolism (**Figure 11B**). Furthermore, the four KO pathways with the most differentially abundant metabolites and the smallest *P/Q* value between the flower buds and blooming flowers were fatty acid elongation; cutin, suberine, and wax biosynthesis; glutathione metabolism; and fatty acid metabolism (**Figure 11C**).

Verification of Gene Expression Profiles Using qRT-PCR

To further verify the expression profiles of genes in the Illumina sequencing analyses, 15 ungenes with high expression levels and large differential multiple were selected for qRT-PCR, and leaf (YS1), flower bud (YS2) and blooming flower (YS3) were selected for RNA-seq. According to transcriptome sequencing data, the expression level of unigenes in the former group was significantly higher than that in the latter group in the three control groups. The RT-PCR results showed that the expression patterns of these 15 genes were consistent with the sequencing data (Additional file 10, **Figure 12**).

Conjoint Analyses of Transcriptome and Metabolome Sequencing

Common Metabolic Pathway Analysis

Due to the few types of differential metabolites among sample groups, and the absence of target metabolites in some differential metabolites, the common pathway analysis of all genes and all metabolites was carried out.

Combined analysis of transcriptome and metabolome sequencing showed that common pathway analysis was conducted for all genes and all metabolites. Due to the small number of differential metabolites among sample groups and the lack of target metabolites in some differential metabolites, we conducted common pathway analysis for all genes and metabolites.

There were 64 metabolic pathways shared by all genes and metabolites, 9016 candidate genes with pathway annotations, and 64 metabolites with pathway annotations. A total of 3,485 candidate genes (38.65%) and 42 metabolites (65.63%) were annotated for Metabolic pathways. A total of 1,917 candidate genes were annotated by Biosynthesis of secondary metabolites pathway, accounting for 21.26% of the total, and 18 metabolites were annotated, accounting for 28.13% of the total. A total of 570 candidate genes were annotated by Carbon Metabolism Pathway, accounting for 6.32% of the total, and 10 metabolites were annotated, accounting for 15.63% of the total. These three paths are the most annotated (Additional File 14).

Correlation Coefficient Model Analysis of Transcriptome and Metabolome

The correlation between genes and metabolites was evaluated according to Pearson's correlation coefficient. The top 250

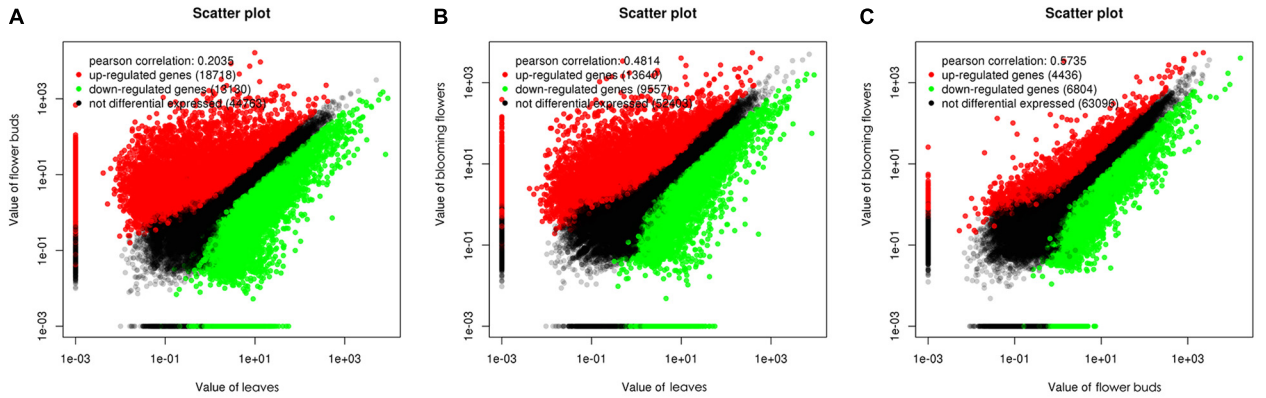


FIGURE 5 | Scatter plots of differentially expressed genes among three *Chrysanthemum dichrum* plant tissues. The values in the vertical and vertical coordinates are gene expression levels. **(A)** The x-coordinate is the gene expression level of leaves, and the y-coordinate is the gene expression level of flower buds. **(B)** The x-coordinate is the expression level of leaves, and the y-coordinate is the expression level of blooming flowers. **(C)** The abscissa is the gene expression level of flower buds, and the ordinate is the gene expression level of blooming flowers.

differential genes and metabolites with the absolute correlation coefficient greater than 0.5 were screened out, and the correlation network diagram of gene expression and metabolite abundance was drawn (**Figure 13**). The results showed that there were more positive correlations between different genes and different metabolites, and only 12 groups showed negative correlations.

DISCUSSION

On the basis of a metabolomic analysis of *C. dichrum* tissue samples from three different developmental periods, we identified several differentially abundant functional metabolites among the 568 total metabolites. Two metabolites associated with food additives, raffinose and 1-kestose, were detected in leaves.

Raffinose, which is also called melitriose, is a trisaccharide comprising galactose, fructose, and glucose. Melitriose is a functional oligosaccharide that can induce substantial bifidobacterial proliferation. It can simultaneously promote the reproduction and growth of beneficial bacteria, such as bifidobacteria and lactobacilli, while also inhibiting the reproduction of harmful bacteria in the intestines, ultimately helping to establish a healthy intestinal microbial flora (Cox, 1966; Mussatto and Mancilha, 2007). Previous studies confirmed that melitriose can also prevent constipation and inhibit diarrhea; function as a two-way regulator to detoxify and protect the liver; inhibit the production of toxins in the body; minimize the burden on the liver; regulate the human immune system; enhance immunity; decrease allergic reactions following an oral administration (Zhang et al., 2004, 2008; Zhang R. et al., 2013). It also can improve lipid metabolism, lower blood lipid and cholesterol levels, and prevent dental caries; however, it is not effective against oral cariogenic bacteria. Additionally, even in the presence of sucrose, melitriose can decrease tartar formation and protect regions where oral microorganisms are deposited and produce acid, thereby preventing corrosion. It can also whiten and strengthen teeth. Because melitriose is a low-calorie trisaccharide, it does not affect the blood sugar level in the human body and can be consumed by people with diabetes (Yazawa and Tamura, 1982; Muthukumaran et al., 2018).

As the smallest oligofructose (i.e., fructooligosaccharide) compound, 1-kestose is an active ingredient naturally found in diverse foods, including fruits, vegetables, and honey, and it is an excellent water-soluble dietary fiber. Fructooligosaccharides, which comprise 1-kestose (GF2), nystose (GF3), and 1F-fructofruranosy 1 nystose (GF4) (Jayalakshmi et al., 2021), can activate bifidobacteria, while also regulating the intestinal microbial flora balance, ensuring the bowel is appropriately hydrated, regulating blood lipid levels, lowering cholesterol and blood sugar levels, and enhancing immunity (Li et al., 2004; Probert et al., 2004; Mabel et al., 2008; Hidaka et al., 2010; Delgado et al., 2012). Owing to its superior physiological functions, oligofructose has become a widely popular functional food component in the international food market during the past 10 years. It has been used in more than 500 food and health products and medicines. Moreover, it has been described as a healthy sugar for the 21st century (Yun, 1996).

Metabolites (e.g., glutamine) in edible medicinal plants have been detected in *C. dichrum* flower buds. Basic and clinical experiments demonstrated that glutamine can decrease catabolic activities, promote protein synthesis, improve immune functions, protect the intestinal mucosal barrier, and accelerate wound healing. It can also be used to improve gastrointestinal health by preventing gastric and duodenal ulcers, gastritis, and hyperacidity (Taxuke et al., 2003). Asparagine is a medically relevant metabolite that can lower blood pressure, expand bronchial tubes (useful for treating asthmatics), and prevent peptic ulcers and gastric dysfunction. It can also be used for cultivating microorganisms and treating acrylonitrile wastewater.

In this study, raffinose and asparagine contents were higher in *C. dichrum* blooming flowers than in leaves. Additionally, another metabolite, xylose, was also highly abundant in blooming flowers. This sugar has been used as a food additive. As a component of xylan, it is widely distributed in plants. Xylose is

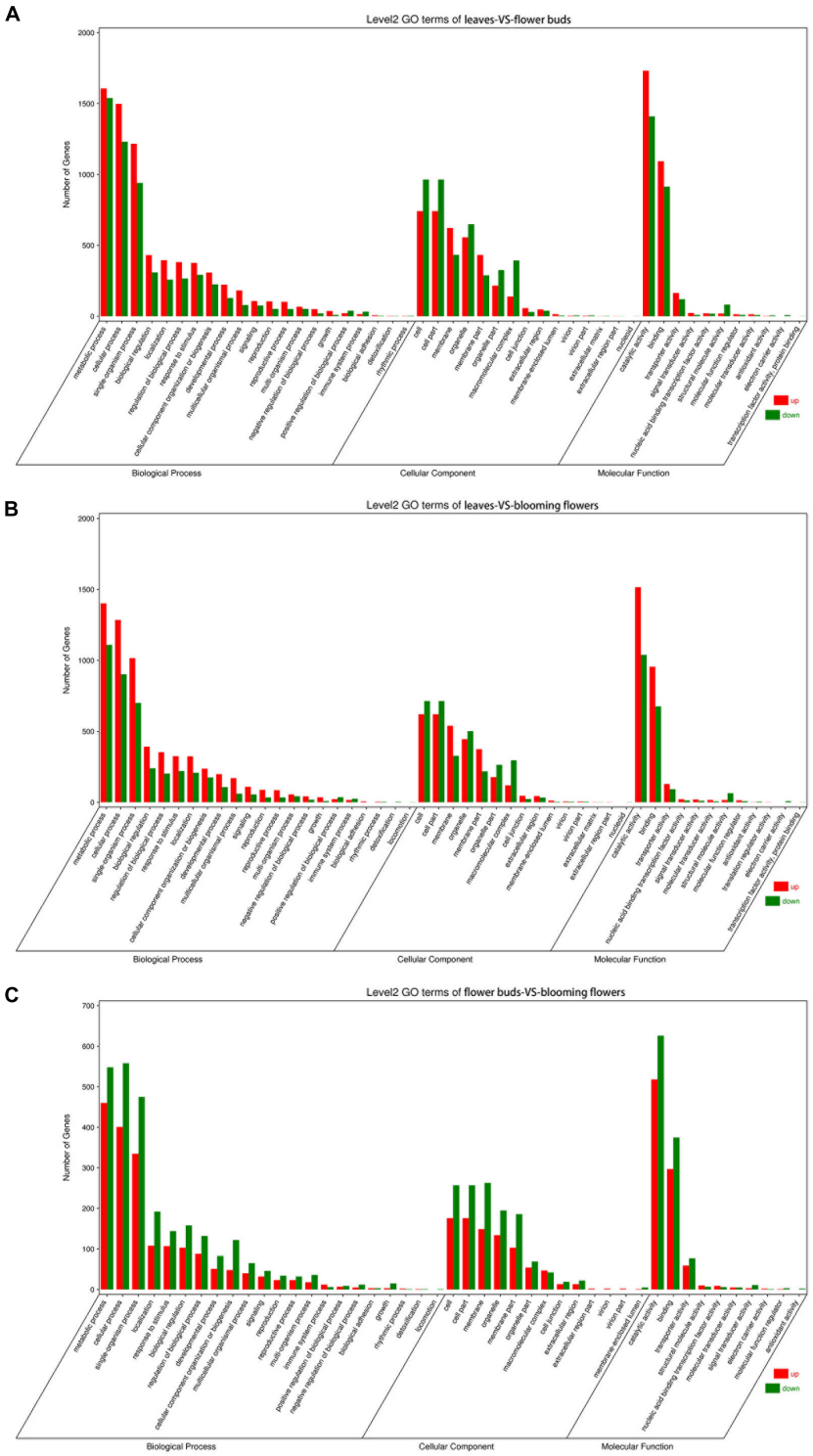


FIGURE 6 | GO term classification of differentially expressed genes among *Chrysanthemum dichrum* plant tissues. The GO classification of all genes was divided into three parts: Biological Process, Cellular Component and Molecular Function. **(A)** Shows the GO classification of the differentially expressed genes in the leaves and flower buds comparison group, and **(B)** shows the GO classification of the differentially expressed genes in the leaves and blooming flowers comparison group. **(C)** Shows the GO classification of differentially expressed genes between the flower buds and blooming flowers comparison group.

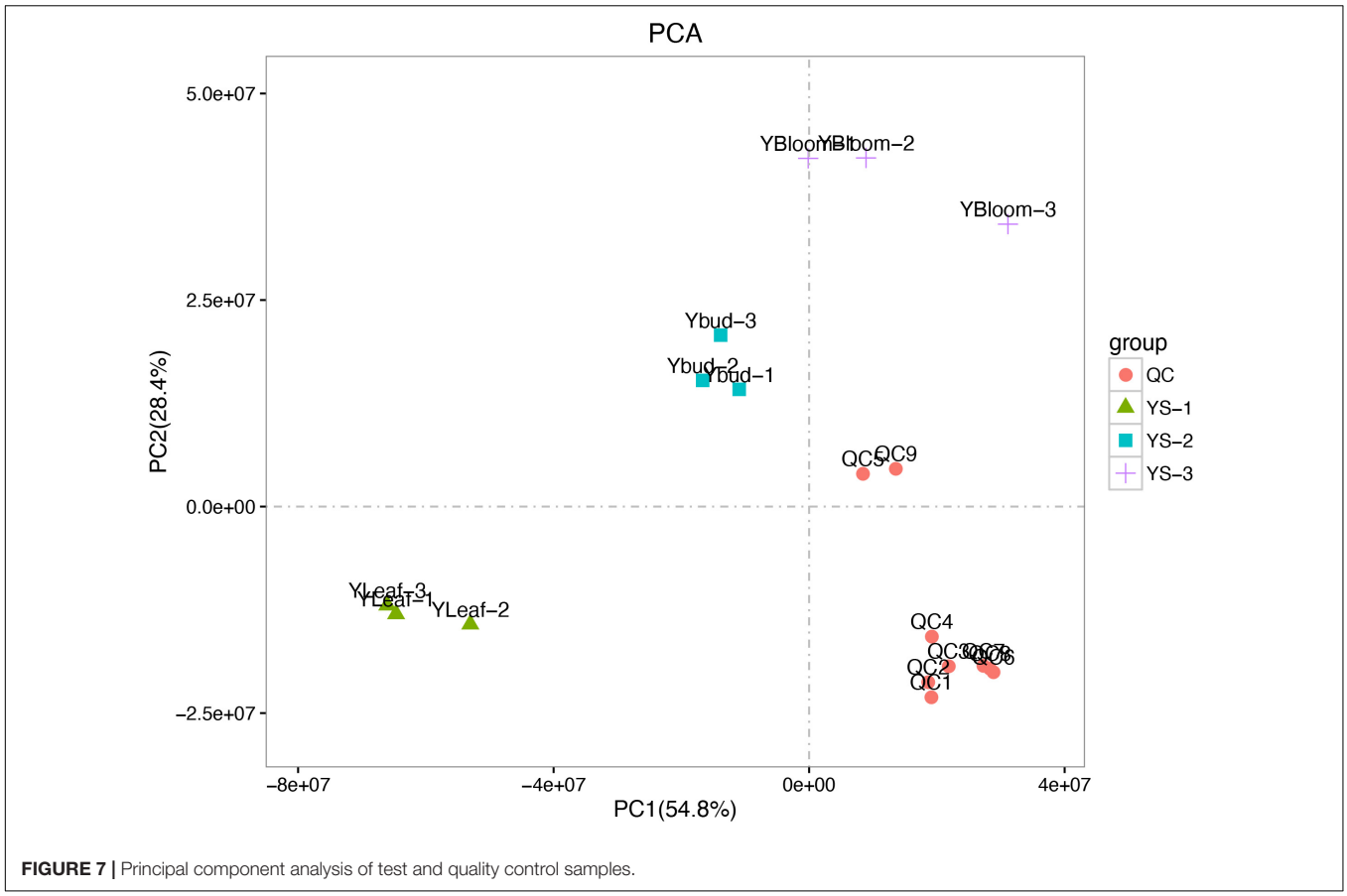


FIGURE 7 | Principal component analysis of test and quality control samples.

not digested and absorbed, making it useful as a sweetener that will not lead to weight gain. It can also activate and promote the growth of bifidobacteria in the human intestine, with beneficial effects on health. Ingesting foods containing xylose can improve the microbial environment of the human body and enhance immunity. It cannot be used by microorganisms in the oral cavity. Additionally, it has some physiological functions of dietary fiber and it can decrease serum cholesterol levels and prevent colon cancer. Hence, adding a small amount of xylose to food can lead to increased positive health effects (Liang and Zhao, 2011). Xylose is widely present in plant hemicellulose as macromolecular xylan and can be obtained by degrading xylan with acids or enzymes (Harner et al., 2015). Xylooligosaccharides are an important component of functional foods. Because they are stable and non-toxic sugars that can be efficiently purified, xylooligosaccharides are increasingly being used in food, feed, and medicine, while also being applied in other fields (Zhang Y.H. et al., 2013; Baker et al., 2021).

In the current study, the 1-kestose and xylose contents were high in blooming flowers. Similarly, many other functional metabolites were more abundant in blooming flowers than in flower buds and leaves, including tartaric acid and salicylic acid. Tartaric acid is an antioxidant that has been added to foods. It is most commonly used as a beverage additive and a raw material for the pharmaceutical industry (Kroschwitz and Seidel, 2004). Earlier research proved that salicylic acid is useful for treating acne and for lightening post-acne pigmentation, minimizing pore size, removing fine wrinkles, and preventing sun-induced skin aging. The renewal of skin cells in response to salicylic acid will result in smoother skin (Kligman and Kligman, 1998). In small amounts, salicylic acid is also applicable as a food preservative (Chrubasik et al., 2000). It is also an important raw material for the pharmaceutical industry.

The data presented herein indicate that *C. dichrum* flowers and leaves are rich in metabolites with important functions. Consequently, *C. dichrum* is an important plant source of useful compounds. In terms of the extraction of functional metabolites from *C. dichrum*, the flowers at full bloom are likely the most appropriate plant tissues because they contain large amounts of diverse metabolites.

MATERIALS AND METHODS

Plant Materials and RNA Extraction

The *C. dichrum* plants used in this study were cultivated at 23°C in a greenhouse at the Beijing Academy of Agriculture and Forestry Sciences (116.3°E, 39.9°N) with an 8-h light/16-h dark cycle. Leaves, flower buds, and blooming flowers were collected from *C. dichrum* plants, with three biological replicates per sample. This plant samples was used for both transcriptome analysis and Metabolomic

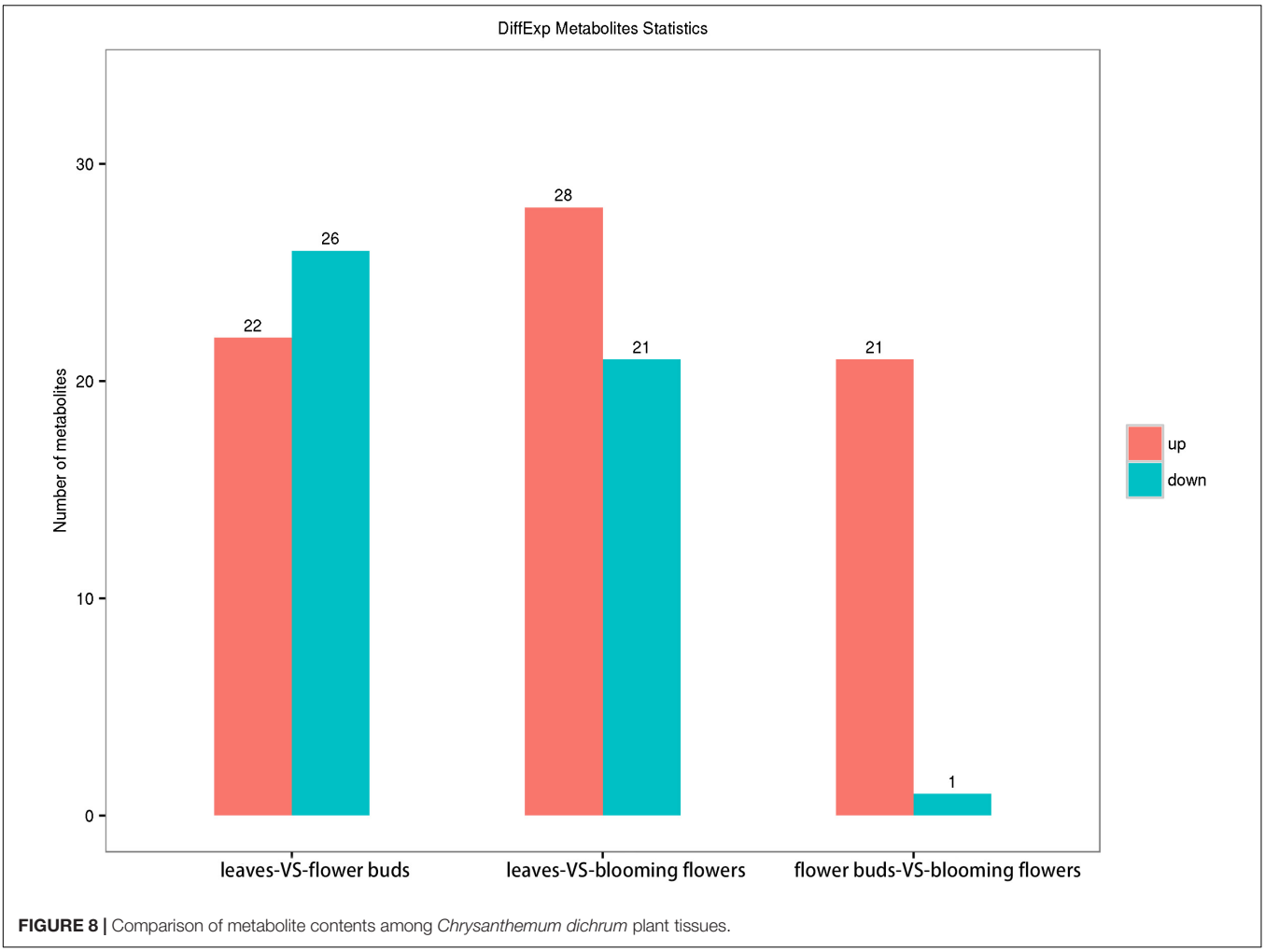


FIGURE 8 | Comparison of metabolite contents among *Chrysanthemum dichrum* plant tissues.

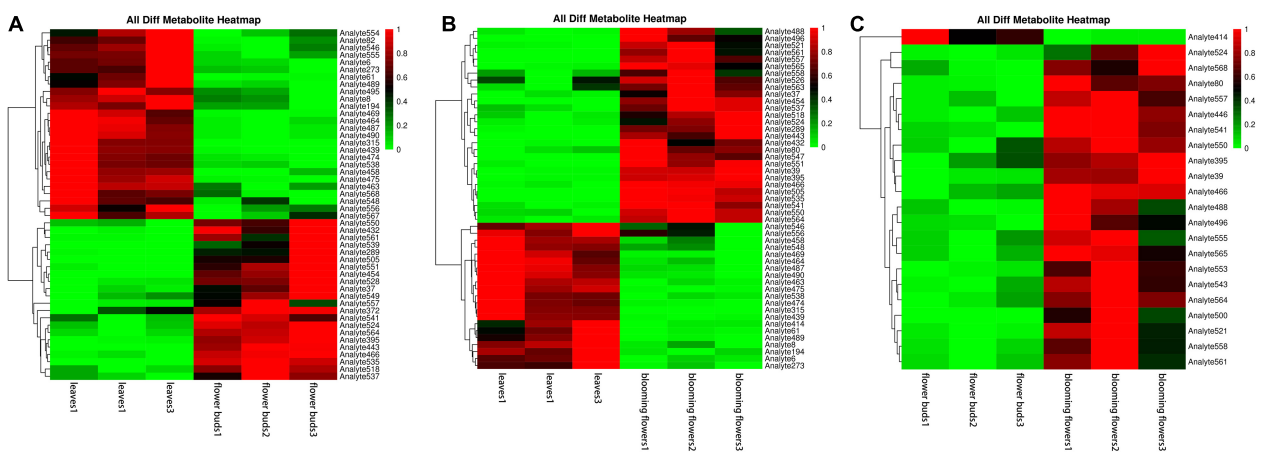


FIGURE 9 | Clustering heat maps of differentially abundant metabolites among *Chrysanthemum dichrum* plant tissues. The abscissa is the sample of each period, and the ordinate on the right is the name of the metabolite. The redder part of the heat map indicates the higher the metabolite content. **(A)** Is the differential metabolite clustering heat map of leaves and flower buds, and **(B)** is the differential metabolite clustering heat map of leaves and blooming flowers comparison group. **(C)** Is the heat map of different metabolites clustering between flower buds and blooming flowers comparison group. The abscissa is the sample of each period, and the ordinate on the right is the name of the metabolite. The redder part of the heat map indicates the higher the metabolite content.

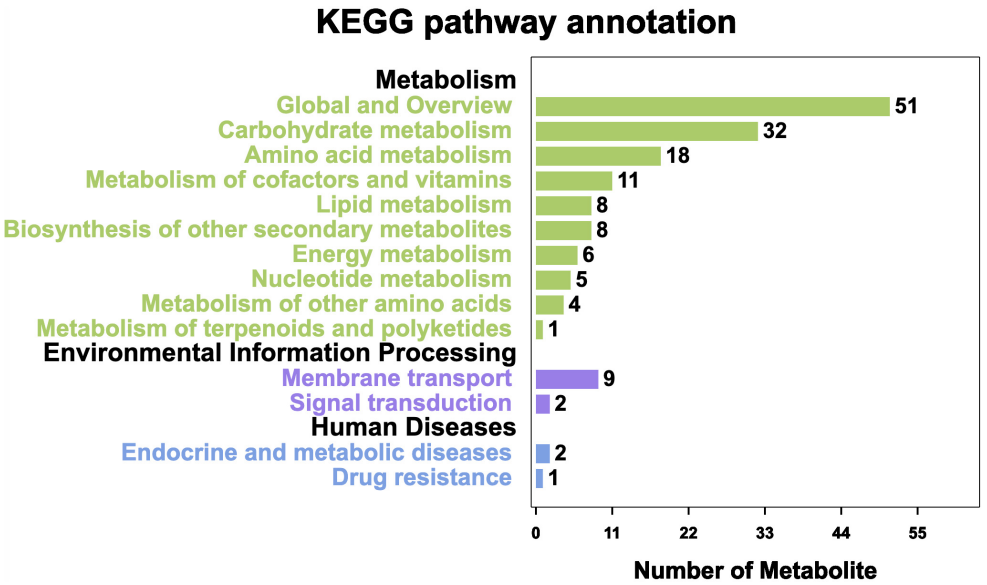


FIGURE 10 | KEGG pathway annotation of metabolites.

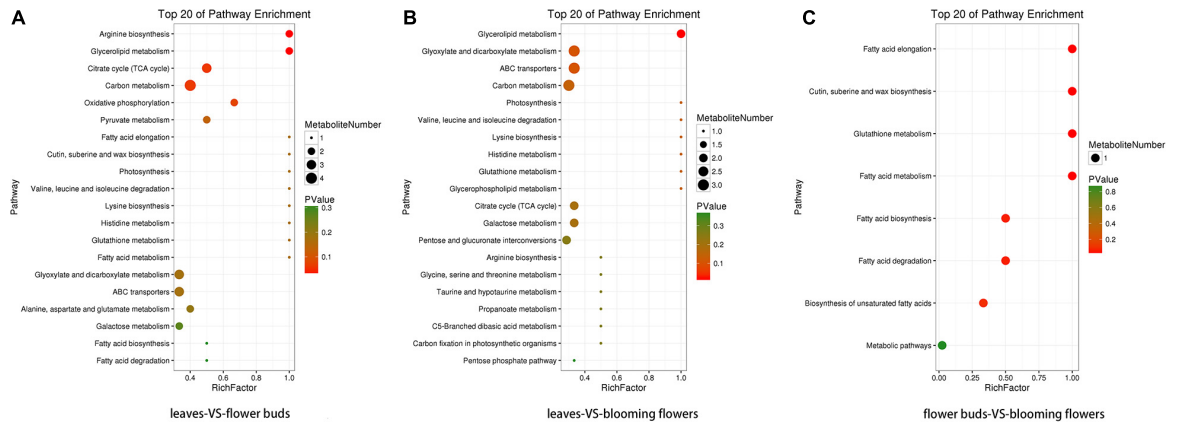


FIGURE 11 | Enriched KO pathway bubble diagrams. **(A)** Shows the enrichment of KO pathways in leaves and flower buds, and **(B)** shows the main KO pathways in which metabolites in leaves and blooming flowers are enriched. **(C)** Shows the enrichment of KO pathway in flower buds and blooming flowers. The top 20 Pathways with the smallest *P* value (or *Q* value) were used for mapping. The ordinate was Pathway, and the abscissa was enrichment factor (the number of differences in this pathway divided by all numbers). The size represented the number, and the redder the color, the smaller the *P/Q* value.

analysis. The collected samples were quickly frozen in liquid nitrogen and stored at −80°C. The RNeasy Plant Mini Kit (Qiagen, China) was used to extract total RNA from the frozen samples. The concentration of the RNA was determined using the NanoDrop ND2000 spectrophotometer (Thermo Scientific).

Library Construction and Sequencing

To construct a sequencing library, rRNA was eliminated from the total RNA extracted from the leaves, flower buds, and blooming flowers using the Ribo-Zero^{TM} Magnetic Kit (Epicenter), after which the mRNA was enriched using oligo-(dT) beads. Fragmentation buffer was used to produce short mRNA fragments, which were then reverse transcribed into cDNA. The second cDNA strand was synthesized in buffer containing DNA polymerase I, RNase H, and dNTP. The cDNA fragments were purified using the QiaQuick PCR extraction kit. After repairing the ends and adding a poly(A) tag, an Illumina sequencing adapter was ligated. The size of the ligation products was determined by Gene *de novo* using the Illumina HiSeq^{TM} 4000 system (Illumina, San Diego, CA, United States) for the subsequent agarose gel electrophoresis, PCR amplification, and sequencing. The PacBio Sequel system (PacBio, CA, United States) was used for sequencing. To improve the accuracy of the PacBio data, the reads were filtered by deleting reads with adapters, reads with more than 10% unknown nucleotides (N), and reads with more than 40% low-quality nucleotides (*Q* value $\leq$ 20). The CD-HIT (version 4.6.7) software

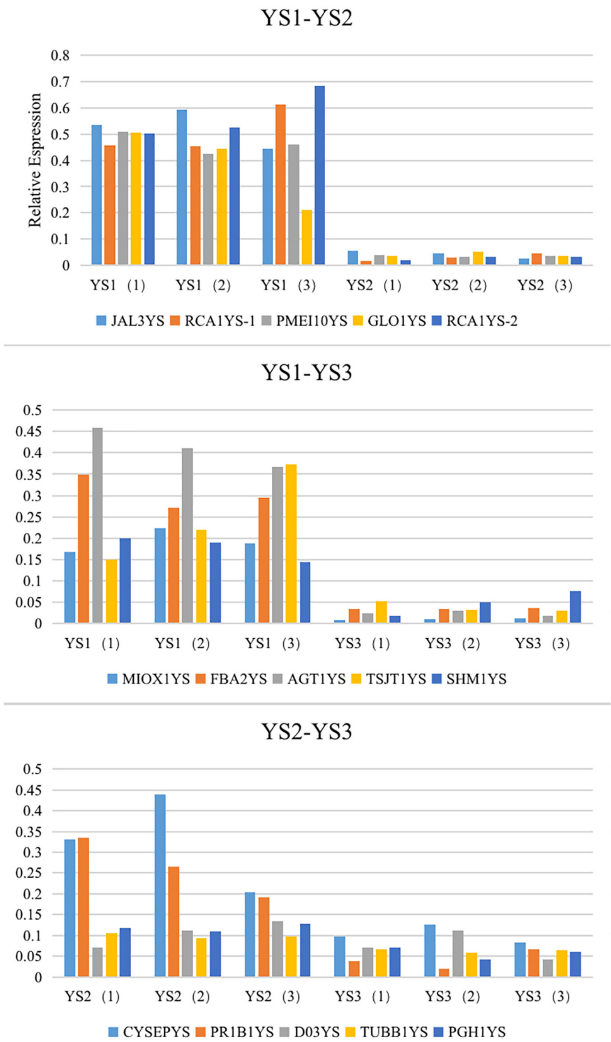


FIGURE 12 | The expression profiles of 15 transcripts in *C. dichrum* by qRT-PCR.

(with a sequence identity threshold of 0.99) was used to remove redundant sequences to obtain the final unigene sequences.

Basic Annotation of Unigenes

Unigenes were annotated using the BLASTX algorithm and the following four databases (E value of 1.00E-5): Nr (NCBI), Swiss-Prot[1], KEGG[2], and KOG[3]. The sequence direction of each gene was determined according to the best alignments. Unigenes were annotated with GO terms using the Blast2GO program (Conesa et al., 2005). On the basis of the Blast2GO analysis, we selected high-quality unigenes with a high hit rate. These unigenes were functionally classified using the WEGO software (Ye, 2006).

Analysis of *Chrysanthemum dichrum* Transcriptome Sequencing Results

The RPKM values derived from the RNA-seq data were used to represent unigene expression levels (Fourquin et al., 2005). The DEGs in the chrysanthemum transcriptome were screened and analyzed as previously described (Samarskiĭ and Claverie, 1997). To verify the *P*-value threshold, the FDR was used in multiple tests and analyses (Endress, 2006). The threshold for determining a significant difference in gene expression was set as follows: absolute value of the $\log_2$(ratio) ≥ 2 and FDR < 0.05 (Mortazavi et al., 2008). Differential gene expression was analyzed for the DEGs with expression levels that differed by at least 2-times between samples.

Alternative Splicing Detection

To analyze alternative splicing events in a unigene, transcripts were divided into gene families according to k-mer similarity. This process uses the coding genome reconstruction tool (Cogent) and then recombines each family into a coding reference genome. This process uses the De Bruijn diagram (Li et al., 2017). The SUPPA tool was used to analyze the alternative splicing events of a unigene (Alamancos et al., 2015).

Gene Expression Analysis Based on qRT-PCR

To conduct qRT-PCR assays, the total RNA extracted from the leaves, flower buds, and blooming flowers were treated with DNase (Promega, United States) to eliminate any residual genomic DNA. The purified RNA served as the template for synthesizing cDNA using a commercial reverse transcription kit (Tsingke, China). The qRT-PCR analysis was completed using the PikoReal system (Thermo Fisher Scientific, Germany). The 20-μL reaction solution comprised 1 μL reverse transcribed cDNA as the template. The PCR program was as follows: 95°C for 30 s; 40 cycles of 95°C for 5 s and 60°C for 30 s. Details regarding the qRT-PCR primers used to determine the relative expression levels of specific genes are listed in Additional file 13. Each sample was analyzed in triplicate, with three biological replicates. The $2^{-\Delta\Delta Ct}$ method was used to calculate relative gene expression levels. The reference control was the *Aspergillus* gene encoding protein phosphatase 2A (PP2Ac) (Xue et al., 2014).

Chemicals and Reagents

The chemicals and reagents used in this study, such as methanol, ethanol, and acetonitrile, were purchased from Merck (Germany;[4]). The Milli-Q system (Millipore, Bedford, MA, United States) was used to produce ultrapure water. Authentic standards were purchased from BioBioPha Co., Ltd.[5] and Sigma-Aldrich[6]. All chemicals and reagents were of analytical grade.

Sample Preparation and Extraction

Freeze-dried samples were ground to a powder using zirconia beads and a mixing mill (MM 400, Retsch) set at 30 Hz for

[1]http://www.expasy.ch/sprot

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

[3]http://www.ncbi.nlm.nih.gov/KOG

[4]www.merckchemicals.com

[5]www.biobiopha.com/

[6]www.sigmaaldrich.com/unitedStates.html

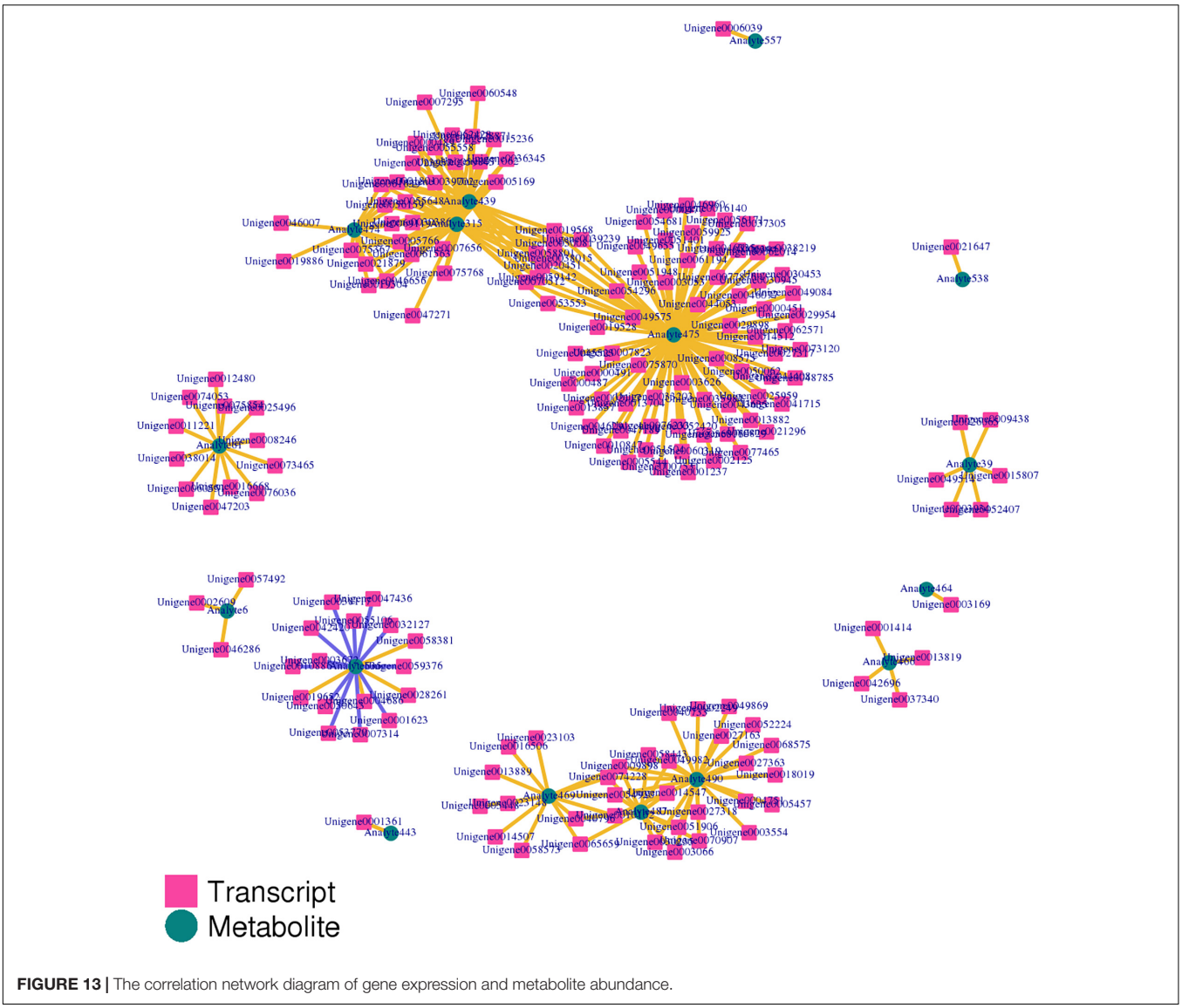


FIGURE 13 | The correlation network diagram of gene expression and metabolite abundance.

1.5 min. The powdered material was weighed, after which 100 mg was mixed with 1.0 mL 70% methanol aqueous solution (containing 0.1 mg/L lidocaine as an internal standard) overnight at 4°C. After centrifuging the mixture at 10,000 *g* for 10 min, the supernatant was collected and filtered (SCAA-104; pore size 0.22 μm; ANPEL, Shanghai, China;[7]) for the subsequent LC-MS/MS analysis. To assess the repeatability of the entire experiment, quality control samples were mixed with all samples.

AB Sciex QTRAP6500 (UPLC) Analysis

The extracted compounds were analyzed using an LC-ESI-MS/MS system [UPLC: Shim-pack UFLC SHIMADZU CBM30A[8]; MS/MS: Applied Biosystems 6500 QTRAP[9]] (Chen W. et al., 2013; Wishart et al., 2013). For each sample, a 2-μL aliquot was injected into the Waters ACQUITY UPLC HSS T3 C18 column (2.1 mm × 100 mm, 1.8 μm) operating at 40°C with a flow rate of 0.4 mL/min. The mobile phases used were acidified water (0.04 % acetic acid) (Phase A) and acidified acetonitrile (0.04 % acetic acid) (Phase B). Compounds were separated using the following gradient: 95:5 Phase A/Phase B at 0 min; 5:95 Phase A/Phase B at 11.0 min; 5:95 Phase A/Phase B at 12.0 min; 95:5 Phase A/Phase B at 12.1 min; and 95:5 Phase A/Phase B at 15.0 min. The eluent was analyzed by an ESI-triple quadrupole-linear ion trap (QTRAP) mass spectrometer. The LIT and triple quadrupole (QQQ) scans were acquired using a triple quadrupole-linear ion trap mass spectrometer (QTRAP; AB Sciex QTRAP6500 System) equipped with an ESI-Turbo Ion-Spray interface. The system was operated in the positive ion mode and controlled with the Analyst 1.6.1 software (AB Sciex). The operating parameters were as follows: ESI source temperature, 500°C; ion spray voltage, 5,500 V; curtain gas,

[7]www.anpel.com.cn/

[8]http://www.shimadzu.com.cn/

[9]http://www.appliedbiosystems.com.cn/

25 psi; and collision-activated dissociation, highest setting. The QQQ scans were acquired as MRM experiments with optimized declustering potential and collision energy for each MRM transition. The m/z range was set between 50 and 1,000.

Data Preprocessing and Metabolite Identification

To generate a matrix containing relatively little offset and redundant data, the peak signal/noise was manually checked (>10) and an internal Perl software was used to remove redundant signals resulting from different isotopes and to analyze intra-source fragmentation and K^+, Na^+, and NH_4^+ adducts and dimers. The accurate m/z of each Q1 was obtained, after which the area of each chromatographic peak was calculated. The peaks for the different samples were aligned according to the spectrum and retention time. Metabolites were identified by searching internal and public databases [MassBank, KNApSAcK, HMDB (Zhu et al., 2013), MoTo DB, and METLIN (Worley and Powers, 2013)]. The m/z values, RT, and lysis patterns were compared with standards. The data filtering as well as the peak detection, comparison, and calculation were performed using the Analyst 1.6.1 software.

Multivariate Statistical Analysis

For all samples, the R package model [10]was used to perform the unsupervised dimensionality reduction PCA to initially visualize the differences between different sample groups (Kanehisa et al., 2008).

Differentially Abundant Metabolite Analysis

The *T* test and the variable importance in the projection (VIP) score of the (O)PLS model were used to rank the metabolites with significant differences in abundance between two groups. The *P*-value threshold for the *T* test was set at < 0.05 and the VIP score threshold was set at ≥ 1.

KEGG Pathway Analysis

The enriched KEGG metabolic pathways among the metabolites were determined. The metabolic and signal transduction pathways significantly enriched among the differentially abundant metabolites were identified using the following formula:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of all metabolites annotated using the KEGG database, n is the number of different metabolites in N, M is the number of all metabolites assigned to a specific pathway, and m is the number of different metabolites in M. The *P*-value was adjusted using an FDR ≤ 0.05. The pathways satisfying this condition were identified as significantly enriched among the differentially abundant metabolites.

[10]http://www.r-project.org/

Transcriptome Data

The Submission ID: SUB9976541.
The BioProject ID: PRJNA744998.
Access link: https://www.ncbi.nlm.nih.gov/sra/PRJNA744998.

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 NCBI, accession number PRJNA744998.

AUTHOR CONTRIBUTIONS

HL, XiaC, HC, and JL performed the research. XiaC analyzed the data and prepared the manuscript. CH and YJ guided the research. CL, DC, and XiC provided assistance for the research. All authors read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (31901354) and the Innovation Foundation of the Beijing Academy of Agriculture and Forestry Sciences (KJCX20200112).

ACKNOWLEDGMENTS

We would like to thank Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.716163/full#supplementary-material

Supplementary Table 1 | Genes differentially expressed between *Chrysanthemum dichrum* leaves and flower buds.

Supplementary Table 2 | Genes differentially expressed between *Chrysanthemum dichrum* leaves and blooming flowers.

Supplementary Table 3 | Genes differentially expressed between *Chrysanthemum dichrum* flower buds and blooming flowers.

Supplementary Table 4 | GO functional categorization of significantly up-regulated and down-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and flower buds.

Supplementary Table 5 | Enriched KEGG pathways among the significantly up-regulated and down-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and flower buds.

Supplementary Table 6 | GO functional categorization of significantly down-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and flower buds.

Supplementary Table 7 | Enriched KEGG pathways among the significantly up-regulated and down-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and blooming flowers.

Supplementary Table 8 | GO functional categorization of significantly up-regulated and down-regulated differentially expressed genes between *Chrysanthemum dichrum* flower buds and blooming flowers.

Supplementary Table 9 | Enriched KEGG pathways among the significantly up-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and blooming flowers.

Supplementary Table 10 | Details regarding the 15 primers used for the quantitative real-time PCR analysis of *Chrysanthemum dichrum* genes.

Supplementary Table 11 | Enriched KEGG pathways among the significantly down-regulated differentially expressed genes between *Chrysanthemum dichrum* leaves and blooming flowers.

Supplementary Table 12 | Significantly different metabolites between *Chrysanthemum dichrum* leaves and blooming flowers.

Supplementary Table 13 | Significantly different metabolites between *Chrysanthemum dichrum* flower buds and blooming flowers.

Supplementary Table 14 | Annotation of metabolic pathways common to all genes and metabolites.

REFERENCES

Alamancos, G. P., Pagès, A., Trincado, J. L., Bellora, N., and Eyras, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* 21, 1521–1531. doi: 10.1261/rna.051557.115

Baker, J. T., Duarte, M. E., Holanda, D. M., and Kim, S. W. (2021). Friend or foe impacts of dietary xylans, xylooligosaccharides, and xylanases on intestinal health and growth performance of monogastric animals. *Animals* 11:609. doi: 10.3390/ani11030609

Chen, L., Chen, Y., Jiang, J., Chen, S., Chen, F., Guan, Z., et al. (2012). The constitutive expression of *Chrysanthemum dichrum* ICE1 in *Chrysanthemum grandiflorum* improves the level of low temperature, salinity and drought tolerance. *Plant Cell Rep.* 31, 1747–1758. doi: 10.1007/s00299-012-1288-y

Chen, S., Miao, H., Chen, F., Jiang, B., Lu, J., and Fang, W. (2009). Analysis of expressed sequence tags (ESTs) collected from the inflorescence of *Chrysanthemum*. *Plant Mol. Biol. Rep.* 27, 503–510. doi: 10.1007/s11105-009-0103-6

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

Chen, Y., Chen, S., Chen, F., Li, P., Chen, L., Guan, Z., et al. (2012). Functional characterization of a *Chrysanthemum dichrum* stress-related promoter. *Mol. Biotechnol.* 52, 161–169. doi: 10.1007/s12033-011-9483-6

Chen, Y., Jiang, J., Song, A., Chen, S., Shan, H., Luo, H., et al. (2013). Ambient temperature enhanced freezing tolerance of *Chrysanthemum dichrum* CdICE1 *Arabidopsis* via miR398. *BMC Biol.* 11:121. doi: 10.1186/1741-7007-11-121

Chi, T., Xu, T., Liu, Y., Ma, J., Guan, Z., Fang, W., et al. (2018). Genetic variation for cold tolerance in an interspecific *C. dichrum* × *C. nankingense* population. *J. Nuclear Agric. Sci.* 32, 2298–2304.

Chrubasik, S., Eisenberg, E., Balan, E., Weinberger, T., Luzzati, R., and Conradt, C. (2000). Treatment of low back pain exacerbations with willow bark extract:a randomized double-blind study. *Am. J. Chin. Med.* 109, 9–14. doi: 10.1016/s0002-9343(00)00442-3

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Cox, C. S. (1966). The survival of *Escherichia coli* sprayed into air and into nitrogen from distilled water and from solutions of protecting agents, as a function of relative humidity. *J. Gen. Microbiol.* 43, 383–399. doi: 10.1099/00221287-43-3-383

Delgado, G. T., Thomé, R., Gabriel, D. L., Tamashiro, W. M., and Pastore, G. M. (2012). Yacon(*Smallanthus sonchifolius*)-derived fructooligosaccharides improves the immune parameters in the mouse. *Nutr. Res.* 32, 884–892. doi: 10.1016/j.nutres.2012.09.012

Endress, P. K. (2006). Angiosperm floral evolution: morphological developmental framework. *Adv. Bot. Res.* 44, 1–61. doi: 10.1016/s0065-2296(06)44001-5

Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C., and Scutt, C. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4649–4654. doi: 10.1073/pnas.0409577102

Harner, N. K., Wen, X., Bajwa, P. K., Austin, G. D., Ho, C. Y., Habash, M. B., et al. (2015). Genetic improvement of native xylose-fermenting yeasts for ethanol production. *J. Ind. Microbiol. Biotechnol.* 42, 1–20. doi: 10.1007/s10295-014-1535-z

Hidaka, H., Eida, T., Takizawa, T., Tokunaga, T., and Tashiro, Y. (2010). Effects of fructooligosaccharides on intestinal flora and human health. *Bifidobacteria Microflora* 5, 37–50. doi: 10.12938/bifidus1982.5.1_37

Hong, Y., Tang, X., Huang, H., Zhang, Y., and Dai, S. (2015). Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in *Chrysanthemum*. *BMC Genomics* 16:202. doi: 10.1186/s12864-015-1428-1

Hsianshiu, K. (1993). *Flora Republicae Popularis Sinicae (FRPS)*, Vol. 76. Beijing: Science Press, 47.

Jayalakshmi, J., Mohamed Sadiqa, A., and Sivakumarb, V. (2021). Microbial enzymatic production of fructooligosaccharides from sucrose in agricultural harvest. *Asian J. Microbiol. Biotechnol. Environ. Sci.* 23, 84–88.

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.

Kligman, D., and Kligman, A. M. (1998). Salicylic acid peels for the treatment of photoaging. *Dermatol. Surg.* 24, 325–328. doi: 10.1111/j.1524-4725.1998.tb04162.x

Kroschwitz, J., and Seidel, A. (2004). *Kirk-Othmer Encyclopedia of Chemical Technology. CRC Handbook*, 5 Edn. Hoboken, NJ: Wiley-Interscience.

Li, J., Harata-Lee, Y., Denton, M. D., Feng, Q., Rathjen, J. R., Qu, Z., et al. (2017). Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov.* 3:17031.

Li, J., Wang, J., Kaneko, T., Qin, L. Q., and Sato, A. (2004). Effects of fiber intake on the blood pressure,lipids,and heart rate in Goto Kakizaki rats. *Nutrition* 20, 1003–1007. doi: 10.1016/j.nut.2004.08.010

Li, P., Chi, T., Liu, Y., Fan, H., Wang, H., Guan, Z., et al. (2020). Genetic polymorphism analysis,evaluation of drought tolerance and association analysis using SSR markers in the interspecific *Chrysanthemum dichrum*×*C.nankingense* F1hybrids. *J. Nanjing Agric. Univ.* 43, 238–246.

Liang, C. S., and Zhao, G. Z. (2011). Terahertz spectroscopic inspection and analysis of Xylitol and D-Xylose. *Spectrosc. Spectr. Anal.* 31, 323–327.

Liu, H., Sun, M., Du, D., Pan, H., Cheng, T., Wang, J., et al. (2015). Whole-transcriptome analysis of differentially expressed genes in the vegetative buds, floral buds and buds of *Chrysanthemum morifolium*. *PLoS One* 10:e0128009. doi: 10.1371/journal.pone.0128009

Liu, Z., Ma, L., Nan, Z., and Wang, Y. (2013). Comparative transcriptional profiling provides insights into the evolution and development of the zygomorphic flower of *Vicia sativa* (Papilionoideae). *PLoS One* 8:e57338. doi: 10.1371/journal.pone.0057338

Mabel, M. J., Sangeetha, P. T., Kalpana, P., Srinivasan, K., and Prapulla, S. G. (2008). Physicochemical characterization of fructooligosaccharides and evaluation of their suitability as a potential sweetener for diabetics. *Carbohydr. Res.* 343, 56–66. doi: 10.1016/j.carres.2007.10.012

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Mussatto, S. I., and Mancilha, I. M. (2007). Non-digestible oligosaccharides: a review. *Carbohydr. Polym.* 68, 587–597. doi: 10.1016/j.carbpol.2006.12.011

Muthukumaran, P., Thiyagarajan, G., Babu, R. A., and Lakshmi, B. S. (2018). Raffinose from *Costus speciosus* attenuates lipid synthesis through modulation of PPARs/SREBP1c and improves insulin sensitivity through PI3K/AKT. *Chem. Biol. Interact.* 284, 80–89. doi: 10.1016/j.cbi.2018.02.011

Probert, H. M., Apajalahti, J. H., Rautonen, N., Stowell, J., and Gibson, G. R. (2004). Polydeztrose,lactitol and fructo-ligosaccharide fermentation by colonic bacteria in a three stage continuous culture system. *Appl. Environ. Microbiol.* 70, 4505–4516. doi: 10.1128/aem.70.8.4505-4511.2004

Ren, L., Sun, J., Chen, S., Gao, J., Dong, B., Liu, Y., et al. (2014). A transcriptomic analysis of *Chrysanthemum nankingense* provides insights into the basis of low temperature tolerance. *BMC Genomics* 15:844. doi: 10.1186/1471-2164-15-844

Samarskiĭ, A., and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7, 986–995. doi: 10.1101/gr.7.10.986

Taxuke, Y., Wasa, M., Shimizu, Y., Wang, H. S., and Okada, A. (2003). Alanyl-glulamine-supplement edparenteral nut nition prevents intesti nal islemia-reperfusion injury in rats. *J. Parenter. Enteral Nutr.* 27, 110–115. doi: 10.1177/014860710302700211O

Wang, H., Jiang, J., Chen, S., Qi, X., Peng, H., Li, P., et al. (2013). Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. *PLoS One* 8:e62293. doi: 10.1371/journal.pone.0062293

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., et al. (2013). HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* 41, D801–D807.

Worley, B., and Powers, R. (2013). Multivariate analysis in metabolomics. *Curr. Metabolomics* 1, 92–107. doi: 10.2174/2213235x130108

Xu, Y., Gao, S., Yang, Y., Huang, M., Cheng, L., Wei, Q., et al. (2013). Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. *BMC Genomics* 14:662. doi: 10.1186/1471-2164-14-662

Xue, C. M., Xie, T. N., Ye, S. D., and Chen, C. (2014). Reference gene selection for quantitative real-time PCR in *Pandora neoaphidis*. *J. Agric. Biotechnol.* 22, 1575–1583.

Yazawa, K., and Tamura, Z. (1982). Search for;sugar sources for selective increase of bifido-bacteria. *Bifidobacteria Microfora* 1, 39–44.

Ye, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297.

Yun, J. W. (1996). Fructooligosaccharides—occurrence, preparation and application. *Enzyme Microb. Technol.* 19, 107–117. doi: 10.1016/0141-0229(95)00188-3

Zhang, R., Zhao, Y., Sun, Y., Lu, X., and Yang, X. (2013). Isolation, characterization, and he-patoprotective effects of the raffinose family oligosaccharides from *Rehmannia glu-tinosa* Libosch. *J. Agric. Food Chem.* 61, 7786–7793. doi: 10.1021/jf4018492

Zhang, R., Zhou, J., Jia, Z., Zhang, Y., and Gu, G. (2004). Hypoglycemic effect of *Rehmannia glutinosa* oligosaccharide in hyperglycemic and alloxan-induced diabetic rats and its mechanism. *J. Ethnopharmacol.* 90, 39–43. doi: 10.1016/j.jep.2003.09.018

Zhang, R. X., Li, M. X., and Jia, Z. P. (2008). *Rehmannia glutinosa*: review of botany, chemistry and pharmacology. *J. Ethnopharmacol.* 117, 199–214. doi: 10.1016/j.jep.2008.02.018

Zhang, Y. H., Zhou, W., and Li, B. (2013). Determination of sugar alcohols sweeteners in sugar-free food by derivatization capillary gas chromatography. *Chin. J. Anal. Chem.* 41, 911–916. doi: 10.3724/sp.j.1096.2013.20897

Zhu, Z. J., Schultz, A. W., Wang, J., Johnson, C. H., Yannone, S. M., Patti, G. J., et al. (2013). Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat. Protoc.* 8, 451–460. doi: 10.1038/nprot.2013.004

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.

Copyright © 2021 Liu, Chen, Chen, Lu, Chen, Luo, Cheng, Jia and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fgene.2021.715731

Insights Into Walnut Lipid Metabolism From Metabolome and Transcriptome Analysis

Suxian Yan[1†], Xingsu Wang[2†], Chenkang Yang[1], Junyou Wang[1], Ying Wang[1], Bangbang Wu[1], Ling Qiao[1], Jiajia Zhao[1], Pourkheirandish Mohammad[3], Xingwei Zheng[1], Jianguo Xu[2], Huming Zhi[1]* and Jun Zheng[1]**

[1] State Key Laboratory of Sustainable Dryland Agriculture, Institute of Wheat Research, Shanxi Agricultural University, Linfen, China, [2] College of Food Science, Shanxi Normal University, Linfen, China, [3] Plant Molecular Biology and Biotechnology Laboratory, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC, Australia

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Jieqin Li,
Anhui University of Science and Technology, China
Caiguo Tang,
Hefei Institutes of Physical Science (CAS), China
Bin Zhang,
Temasek Life Sciences Laboratory, Singapore

***Correspondence:**
Jun Zheng
sxnkyzj@126.com
Jianguo Xu
xjg71@163.com
Huming Zhi
sxnkysx@126.com

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

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

Received: 27 May 2021
Accepted: 30 June 2021
Published: 03 September 2021

Citation:
Yan S, Wang X, Yang C, Wang J, Wang Y, Wu B, Qiao L, Zhao J, Mohammad P, Zheng X, Xu J, Zhi H and Zheng J (2021) Insights Into Walnut Lipid Metabolism From Metabolome and Transcriptome Analysis. Front. Genet. 12:715731. doi: 10.3389/fgene.2021.715731

Walnut oil is an excellent source of essential fatty acids. Systematic evaluation of walnut lipids has significance for the development of the nutritional and functional value of walnut. Ultra-performance liquid chromatography/Orbitrap high-resolution mass spectrometry (UHPLC-Orbitrap HRMS) was used to characterize the lipids of walnut. A total of 525 lipids were detected and triacylglycerols (TG) (18:2/18:2/18:3) and diacylglycerols (DG) (18:2/18:2) were the main glycerolipids present. Essential fatty acids, such as linoleic acid and linolenic acid, were the main DG and TG fatty acid chains. Many types of phospholipids were observed with phosphatidic acid being present in the highest concentration (5.58%). Using a combination of metabolome and transcriptome analysis, the present study mapped the main lipid metabolism pathway in walnut. These results may provide a theoretical basis for further study and specific gene targets to enable the development of walnut with increased oil content and modified fatty acid composition.

Keywords: walnut, lipidomic, UHPLC-Orbitrap HRMS, metabolism, transcriptome

INTRODUCTION

As one of the four major nut crop species in the world, walnut (*Juglans regia* L.) is widely distributed in Asia, Europe, North America, and Africa. In 2019, the world output of walnut was about 3.66 million tons (Zhou et al., 2018). The cultivated area and yield of walnut rank first among all types of dried fruits, and the crop has high economic value (Martínez et al., 2010). The oil content of walnut kernels is 52–70% and walnut oil is an excellent source of essential fatty acids with high nutritional value. Walnut kernels can be eaten fresh or dried. Dried walnuts are currently the most important walnut product. Less than 10% of dried walnuts are highly processed into walnut food. Walnut is also a good source of vegetable oil, which can be used for cooking and as an ingredient in paint and cosmetics (Zambón et al., 2000). The major constituents of walnut oil are triacylglycerols (TG) and diacylglycerols (DG). TGs are the major storage lipids and are an important energy reserve for the seed for germination and development (Shiu-Cheung and Randall, 2006). TG composition indicates the quality and purity of vegetable oils and is increasingly being used by the food industry to confirm oil authenticity. TG and DG constitute a good source of essential fatty acids of which

linoleic acid and linolenic acid are the most common (Bouabdallah et al., 2014). In walnut oil, the ratio of n-3 and n-6 unsaturated fatty acids is 4∼6:1, which is in line with healthy dietary standards for humans (Croitoru et al., 2019). Therefore, it is vital to analyze and compare the lipid composition of walnut comprehensively from the perspective of the lipidome.

Lipids are currently classified into eight accepted categories by "The International Lipid Classification and Nomenclature Committee" as follows: fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), saccharolipids (SL), sterol lipids, prenol lipids, and polyketides (Fahy et al., 2009). For lipidomic separation and investigation, thin-layer chromatography (TLC) was first used, and it has been gradually replaced by gas chromatography (GC) and liquid chromatography (LC) for lower resolution and sensitivity. The combination of GC/LC and mass spectrometry can efficiently separate and accurately detect lipid molecules. However, GC can only analyze small lipid molecules (e.g., fatty acids and tocopherols) that are thermally stable and sufficiently volatile, and long-chain unsaturated fatty acids are easily destroyed (Hamide et al., 2015). High-performance liquid chromatography, including high-performance liquid chromatography (HPLC), ultra-high-performance liquid chromatography (UHPLC), and two-dimensional HPLC (2D HPLC), can quickly achieve high efficiency separation (Li et al., 2011). Modern mass spectrometry (MS) mass analyzers offer very high mass resolution and mass accuracy, such as Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap and time of flight (TOF) (Lee and Yokomizo, 2018). At present, ultra-performance liquid chromatography-high-resolution mass spectrometry (UPLC-MS) is the most used analytical platform for the analysis of plant lipid metabolism. For instance, LC-ESI-MS was used to extend our understanding of the dynamic changes in lipid molecules in high oleic acid peanut at different development stages (Liu et al., 2019). Three hundred phospholipid molecules were detected by liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-Q-TOF) in the seeds of *Eryngium maritimum* and *Cakile maritima* (Zitouni et al., 2016). A total of 165 phospholipids were separated by hydrophilic action chromatography-electrospray atomization-ion trap-time-of-flight mass spectrometry (HILIC-ESI-IT-TOF-MS) in six nut species (Song et al., 2018).

Transcriptomics can reflect the gene expression of cells, tissues, and organisms at a specific time and location (Qi et al., 2011). Many candidate genes related to lipid metabolism can be found with transcriptomics. For example, 4,817 differentially expressed genes were found from the dynamic changes of the transcriptome associated with oil accumulation at different developmental stages in walnut embryos. Among them, *ACCase*, *LACS*, and *FAD7* were identified as key genes for fatty acid synthesis (Zhao et al., 2018). Huang et al. (2021) found 108 genes related to lipid synthesis, including 60 genes for the fatty acid synthesis pathway, 33 for the triglyceride synthesis pathway, seven genes for the formation of oil bodies, and eight transcription factors. By analyzing the miRNA and mRNA transcriptome data of walnut kernels at different developmental stages, 104 miRNAs related to oil accumulation were found (Zhao et al., 2020). Lipid synthesis is the result of the interaction of a multilayer network. Single omics data cannot fully reflect the metabolic activity of cells. Multi-omics analysis is more robust. For example, Rothenberg et al. (2019) analyzed the molecular mechanisms driving anthocyanin accumulation in the development of mutant pink tea flowers (*Camellia sinensis* L.) by combining transcriptomics and metabolomics. Since multi-omics analysis can more clearly identify the genes regulating walnut oil metabolism, it was used in the present study.

UHPLC-Orbitrap HRMS was used to systematically compare the kernel lipid composition of different walnut varieties. The results provide a reference for studying walnut functional lipid components and improving the nutritional quality of walnuts.

MATERIALS AND METHODS

Plant Materials

Xin 2, a precocious walnut variety, produces fruit early with high yield. However, Xin 2 has an astringent taste. The variety Qingxiang combines the advantages of precocious walnut and late walnut with long storage life and good quality (Wang et al., 2012). Walnut samples were collected in the XI county test station in China (110°57′E, 36°42′N, elevation 1,100 m, annual average precipitation 570 mm, annual average temperature 9.5°C). Normal plants were selected from orchards with stable yield. After harvest in mid-October, the green fruit husks were removed and washed. The fruits were then dried at 32°C for 10 h, at 37°C for 24 h, and at 35°C for 15 h before further analysis.

Instruments and Reagents

The following were used in the experiments: UHPLC Nexera LC-30A ultra-high-performance liquid chromatograph (Shimadzu Co. Ltd., Tokyo, Japan), Q-Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States), low-temperature high-speed centrifuge (Eppendorf 5430R, Framingham, MA, United States), Acquity UPLC CSH C18 column (1.7 μm, 2.1 mm × 100 mm, Waters Corporation, Milford, MA, United States). Acetonitrile, isopropanol, methanol, methyl tert-butyl ether and 13 isotopic internal standards: Cer, LPC, PC, LPE, PE, PI, PS, PA, PG, SM, Chol Ester, DG, and TG (Thermo Fisher Scientific, Beijing, China).

Sample Processing

Ten smooth, plump, uniform kernels each of Qingxiang and Xin 2 were selected. The embryos were frozen in liquid nitrogen and ground into a homogenized powder. Thirty milligrams of the powder was thoroughly mixed with 200 μl distilled water and 20 μl internal standard solution. Next, 800 μl of methyl tert-butyl ether and 240 μl of precooled methanol were added.

Abbreviations: DG, diacylglycerol; TG, triacylglycerol; PA, phosphatidic acid; PG, phosphatidylglycerol; PS, phosphatidylserine; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; PIP, phosphatidylinositol diphosphate; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPG, lysophosphatidylglycerol; LPI, lysophosphatidylinositol; CL, cardiolipin; Cer, ceramide; SM, sphingomyelin; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol.

A vortex mixer was used to agitate the mixture throughout the process. The samples were subjected to ultrasound mixing in cold water for 20 min and then allowed to stand at room temperature for 30 min. Samples were centrifuged at 14,000 × *g* at 10°C for 15 min and the upper organic phase was removed and blown dry with nitrogen. Before analysis, 200 μl of 90% isopropanol/acetonitrile solution was added to redissolve the samples, and 90 μl of the sample solution was centrifuged for 15 min at 14,000 × *g* and 10°C. Three microliters of the supernatant was used for analysis. All reagents used were chromatographically pure. Each sample was tested four times in succession.

Chromatographic Conditions

The separation was performed on a UHPLC Nexera LC-30A. The chromatography column was at 45°C. The flow rate was 300 μl/min. Mobile phase A was acetonitrile–water solution (acetonitrile:water = 6:4, v/v) and phase B was acetonitrile–isopropanol solution (acetonitrile:isopropanol = 1:9, v/v). The gradient elution was programmed as follows: 0–2 min with 30% B, 2–25 min with 30–100% B, and 25–35 min with 30% B. The sample was placed in a 10°C automatic sampler for analysis.

Mass Spectrometry Conditions

The samples were separated by UHPLC and analyzed by mass spectrometry with a Q Exactive mass spectrometer. Electrospray ionization (ESI) was performed in positive and negative ion modes. ESI source conditions were as follows: sheath gas flow rate 45 arb, auxiliary gas flow rate 15 arb, collision gas flow rate 1 arb, spray voltage 3.0 kV, capillary temperature 350°C, atomization temperature 300°C, S-Lens RF level 50%, and MS_1 scanning range *m*/*z* 200–1,800. The mass charge ratio of lipid molecules and lipid fragments was obtained by collecting 10 fragment maps (MS_2 scan, HCD) after each full scan. MS_1 had a resolution of 70,000 at *m*/*z* 200 and MS_2 had a resolution of 17,500 at *m*/*z* 200. The above experiments were completed by Applied Protein Technology Company.

Data Analysis

The internal standard method was used for absolute quantification. The absolute content of the analyte was calculated by the response abundance ratio (peak area ratio) of the analyte and the internal standard when the concentration of the internal standard was known. Lipid data were obtained using Analyst^{R} TF 1.6 and Multi Quant^{TM} software (Taguchi and Ishikawa, 2010), and the peaks of lipid molecules and the internal standard lipid molecules were identified by LipidSearch. The main parameters were precursor tolerance 5 ppm, product tolerance 5 ppm, and product ion threshold 5%. The identification of lipid molecular species was mainly based on retention time, accurate *m*/*z*, and fragmentation ion patterns. Quantitative statistics and lipid composition analysis were performed with Microsoft Excel 2007 and Origin 8.5.

The original transcriptome sequencing data were obtained from Huang et al. (2021). The BioProject accession number of the data was PRJNA643637. Quality control of the downloaded transcriptome raw data was performed with FASTP V0.20.1 (Chen et al., 2018). After quality control, the clean data were aligned to a reference genome using HISAT2 (v2.0.5).[1] FeatureCounts (Yang et al., 2014) was used for gene quantitative analysis. EggNOG V5.0 (Huerta-Cepas et al., 2017) was used for gene annotation.

RESULTS

Data Quality Assessment

For each sample, good repeatability of the experiment was evidenced by both the response strength of the chromatographic peak and retention time being nearly identical between runs (**Figure 1**).

Lipid Separation

The elution order of the same type of lipid molecules is determined by the number of carbon atoms and double bonds in the fatty acid chains. Retention time increased and elution slowed as the number of carbon atoms increased and vice versa. Nearly all sample peaks were detected within about 25 min. The peak shape, resolution, and response values were good. In the positive ion mode, glycerides (TG, DG) and some phospholipids (PC, PE, and LPC) had better mass spectrometry response intensity (**Figure 2A**). TG and DG generated primarily $[M + NH_4]^+$, PC, and PE, and LPC generated primarily $[M + H]^+$. PI, PA, PS, PG, PIP, CL, LPE, LPG, LPI, some PE, PC, and saccharolipids had better responses under the negative ion mode (**Figure 2B**). PI, PA, PS, PG, PIP, PE, CL, LPE, LPG, and LPI produced primarily $[M + H]^-$ and PC and glycolipid produced $[M + HCOO]^-$.

For lipid identification, the databases LIPID MAPS[2] and Lipid Bank[3] were searched. In addition, walnut lipids can be distinguished by retention time in either positive or negative ion modes and by MS_1 (primary mass spectrum) and MS_2 (secondary mass spectrum) data. For example, the mass spectrum behavior of DG (18:2/18:2) can be explained as follows. The main mass spectral peak of DG (18:2/18:2) in positive ion mode was $[M + NH_4]^+$ (*m*/*z* 634.5405). The secondary mass spectrometry of DG generated fragmentation ions *m*/*z* 599.5035 and *m*/*z* 337.2734, corresponding to M-OH and NL $[FA (18:2)-H + NH_4]^+$, and generated characteristic fragmentation ion *m*/*z* 263.2367 after dissociation (**Supplementary Figure 2A**) corresponding to fatty acid C18:2. The molecule was identified as DG (18:2/18:2).

Variance Analysis

Using partial least squares discrimination analysis (PLS-DA) (**Supplementary Figure 3A**), the model evaluation parameters (R^2Y, Q^2) obtained are listed in **Supplementary Table 1**. Generally, if Q^2 is greater than 0.5, the model is the most stable

[1] http://daehwankimlab.github.io/hisat2/

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

[3] http://www.lipidbank.jp/

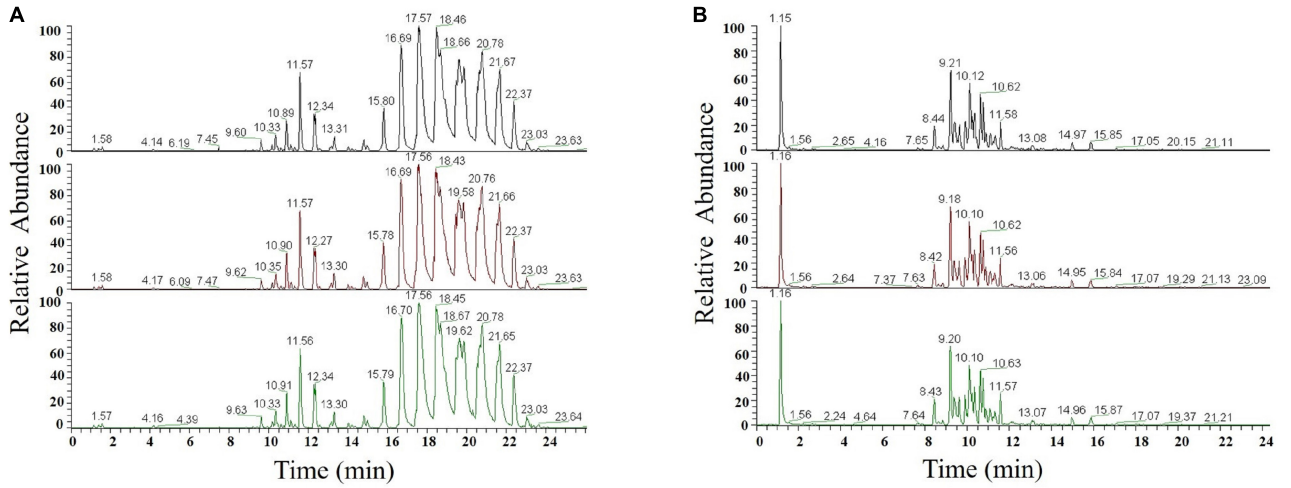


FIGURE 1 | Chromatogram in ESI-positive **(A)** and ESI-negative ion modes **(B)**.

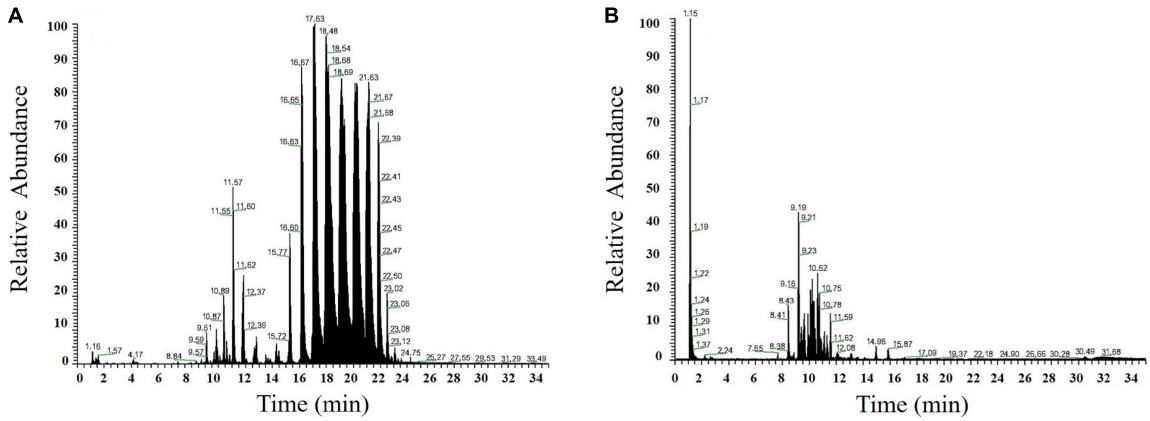


FIGURE 2 | Chromatogram in ESI-positive **(A)** and ESI-negative ion modes **(B)** of XE.

and reliable; if $0.3 < Q^2 \leq 0.5$, the model is stable; and if $Q^2 < 0.3$, the reliability of the model is low.

Difference analysis of all detected lipid molecules was performed and the results are expressed in a volcano graph (**Figure 3A**).

Of 525 lipid molecules in the two walnut samples, 12 molecules with significant differences (OPLS-DA variable importance for the projection > 1 and $p < 0.05$) were identified using the OPLS-DA model (**Supplementary Figure 2B**). Among these, there were seven TG molecules—TG (14:0/18:2/18:3), TG (18:1/14:0/18:3), TG (15:0/18:1/18:3), TG (16:1/18:3/18:3), TG (16:0/18:3/18:3), TG (18:3/17:1/18:3), and TG (17:0/18:2/18:3); one DG molecule—DG (16:0/18:3); three PC molecules—PC (34:1), PC (16:0/18: 2), and PC (18:0/18:1); and one LPC molecule—LPC (18:1) (**Table 1**).

Hierarchical clustering of the differential lipid molecules based on the expression levels showed that the contents of PC (34:1), PC (18:0/18:1), and LPC (18:1) were higher in Qingxiang, and the contents of TG (14:0/18:2/18:3), TG (18:1/14:0/18:3), TG (15:0/18:1/18:3), TG (16:1/18:3/18:3), TG (16:0/18:3/18:3), TG (18:3/17:1/18:3), TG (17:0/18:2/18:3), DG (16:0/18:3), and PC (16:0/18:2) were higher in Xin 2 (**Figure 3B**).

Lipid Species in Walnut Kernels

A total of 525 lipid molecules representing 20 lipid subclasses were identified. The number of lipid molecules in different subtypes varied greatly (**Figure 4**). A total of 250 types of GLs were detected, of which TG was most frequent, with 207 types of TGs, and DGs were the next most common with 43 types. There were 221 types of GPs consisting of 50 PCs, 31 PAs, 36 PEs, 35 PSs, 19 PIs, 14 PGs, 12 LPCs, 5 LPEs, 2 LPGs, 2 LPIs, 5 PIPs, and 10 CLs. There were 36 kinds of SPs, consisting of 28 Cers, 7 CerG1s, and 1 SM. There were 18 kinds of SLs, including 3 MGDGs, 3 DGDGs, and 6 SQDGs.

In addition, there was a rich array of fatty acids present, including 21 saturated fatty acids, namely, C4:0, C8:0, C10:0,

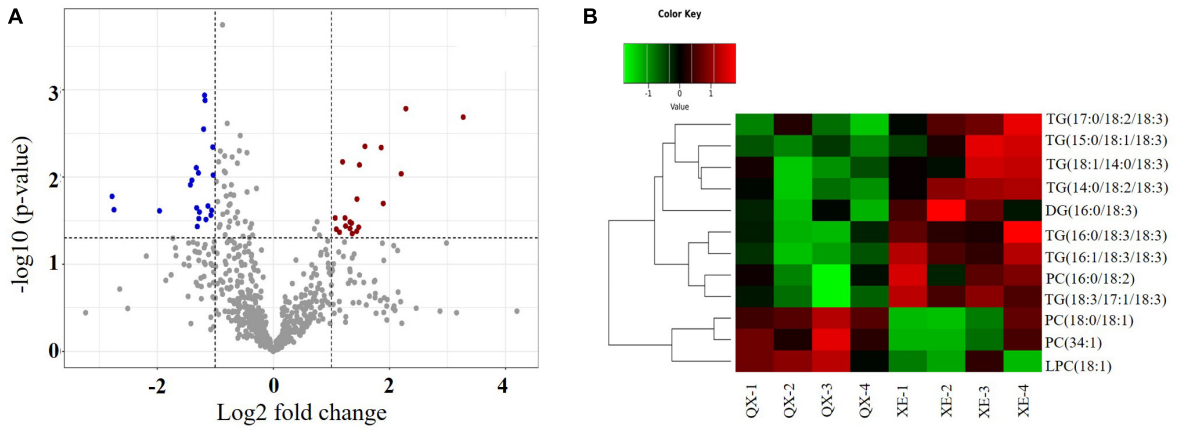


FIGURE 3 | Analysis of differential lipid molecules in walnut seed kernels. **(A)** Volcano plot. The dots represent lipid molecules, among which the blue and red dots are differential lipid molecules that satisfies FC < 0.5, FC > 2, and $p < 0.05$. **(B)** Hierarchical and clustering analysis.

TABLE 1 | Lipid molecules showing significant differences between two walnut varieties ($p < 0.05$).

Lipid group	Class	Fatty acid	Ion formula	Cal Mz	QX (μg/g) ± SD	XE (μg/g) ± SD	*p*-value
PC (34:1) + H	PC	(34:1)	C_{42} H_{83} O_8 N_1 P_1	760.58	241.96 ± 75.77	119.26 ± 56.20	0.0337
TG (50:5) + NH_4	TG	(14:0/18:2/18:3)	C_{53} H_{96} O_6 N_1	842.72	119.48 ± 16.05	181.00 ± 15.00	0.0052
TG (50:4) + NH_4	TG	(18:1/14:0/18:3)	C_{53} H_{98} O_6 N_1	844.73	259.03 ± 42.34	381.81 ± 51.73	0.0382
TG (51:4) + NH_4	TG	(15:0/18:1/18:3)	C_{54} H_{100} O_6 N_1	858.75	167.36 ± 9.22	235.70 ± 36.14	0.0344
TG (52:7) + NH_4	TG	(16:1/18:3/18:3)	C_{55} H_{96} O_6 N_1	866.72	106.63 ± 22.35	200.55 ± 23.52	0.0024
TG (52:6) + NH_4	TG	(16:0/18:3/18:3)	C_{55} H_{98} O_6 N_1	868.74	416.58 ± 125.62	751.36 ± 165.34	0.0204
TG (53:7) + NH_4	TG	(17:1/18:3/18:3)	C_{56} H_{98} O_6 N_1	880.74	124.57 ± 23.16	274.53 ± 12.97	0.0050
TG (53:5) + NH_4	TG	(17:0/18:2/18:3)	C_{56} H_{102} O_6 N_1	884.77	207.39 ± 58.96	382.70 ± 60.74	0.0252
DG (34:3) + NH_4	DG	(16:0/18:3)	C_{37} H_{70} O_5 N_1	608.52	945.60 ± 310.82	1,443.94 ± 394.57	0.0352
LPC (18:1) + HCOO	LPC	(18:1)	C_{27} H_{53} O_9 N_1 P_1	566.35	69.83 ± 14.88	16.00 ± 7.11	0.0178
PC (34:2) + HCOO	PC	(16:0/18:2)	C_{43} H_{81} O_{10} N_1 P_1	802.56	442.17 ± 83.06	535.81 ± 71.90	0.0494
PC (36:1) + HCOO	PC	(18:0/18:1)	C_{45} H_{87} O_{10} N_1 P_1	832.61	43.45 ± 7.17	9.50 ± 5.11	0.0327

Cal Mz means mass-to-charge ratio. QX (μg/g) ± SD and XE (μg/g) ± SD are the means of four replicates ± standard deviation.

C12:0, C13:0, C15:0, C14:0, C16: 0, C17:0, C18:0, C19:0, C20:0, C21:0, C22:0, C23:0, C24:0, C25:0, C26:0, C27:0, C29:0, and C30:0, and 27 unsaturated fatty acids, namely, C10:1, C10:2, C12:1, C14:1, C14:2, C14:3, C16:1, C17:1, C18:1, C18:2 C18:3, C18:4, C19:1, C20:1, C21:1, C20:2, C20:4, C20:5, C22:4, C22:5, C22:6, C24:1, C24: 2. C26:1, C28:1, C29:1, and C30:1. Some rare medium-chain fatty acids (C4:0, C8:0, C10:0, C10:1, C10:2) and ultra-long-chain fatty acids (C24:2, C25:0, C26:0, C27:0, C26:1, C28:1, C29:1, C30:1) were present.

Lipid Content in Walnut Kernels

Comparing the lipid content of the two walnut varieties, Qingxiang had 140,711 μg/g and Xin 2 had 155,801 μg/g. The content trends of the lipid components of Qingxiang and Xin 2 were nearly identical, with both having the highest content of glycerides (including DG and TG) accounting for 88.37 and 86.18% of the total lipid content, respectively. Phospholipids were the second most common type (including PA, PG, PS, PC, PE, LPC, PI, LPG, LPI, LPE, PIP, and CL) accounting for, respectively, 10.9 and 13.2% of the total lipids.

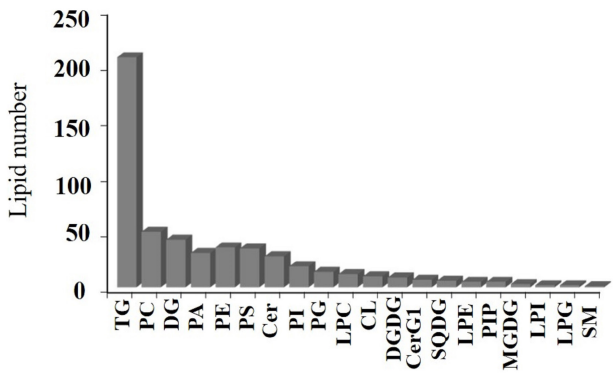


FIGURE 4 | Walnut kernel lipid composition.

Glycolipids (including DGDG, MGDG and SQDG) accounted for 0.7 and 0.61% and sphingolipids (including Cer and SM) accounted for 0.03 and 0.01%, respectively. In addition, comparing the lipid subtypes (**Figure 5**), the content of DG,

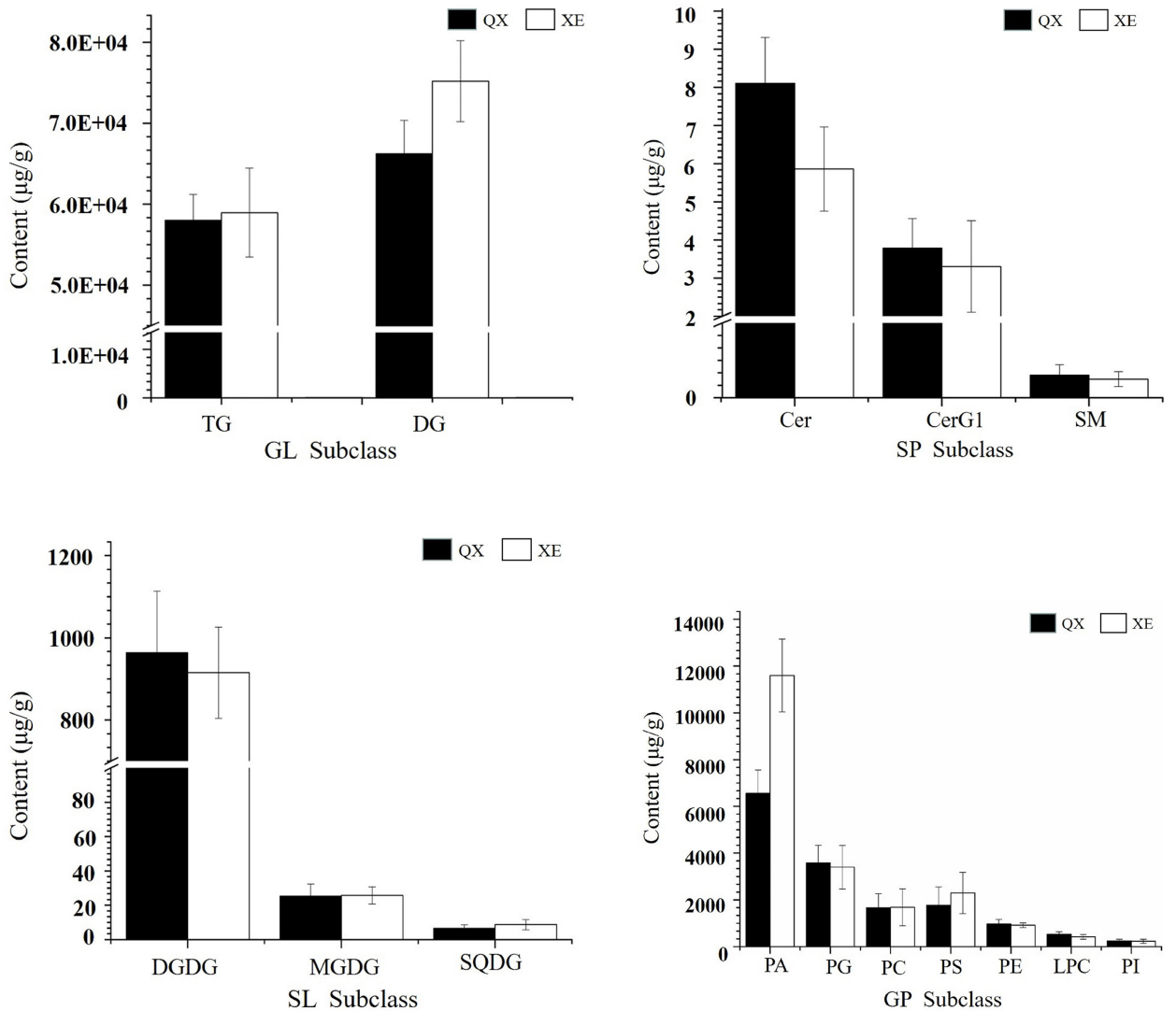


FIGURE 5 | The content of lipid subtypes in two walnut varieties. The ordinate represents the sum of the lipid molecules with the same lipid subtypes.

TG, PA, and PS in Xin 2 was relatively high compared with that in Qingxiang.

In both walnut varieties, the main molecules among the TGs were TG 54:7 (18:2/18:2/18:3, LLLn), TG 54:6 (including TG 18:1/18:2/18:3, OLLn; and TG 18:2/18:2/18:2, LLL), TG 54:3 (18:1/18:1/18:1, OOO), TG 52:4 (16:0/18:1/18:3, POLn), and TG 52:5 (16:0/18:2/18:3, PLLn), with mainly four kinds of fatty acids: C16:0, C18:1, C18:2, and C18:3 (**Figure 6A**). TG (14:0/18:2/18:3), TG (18:1/14:0/18:3), TG (15:0/18:1/18:3), TG (16:1/18:3/18:3), TG (16:0/18:3/18:3), TG (18:3/17:1/18:3), and TG (17:0/18:2/18:3) ($p < 0.05$) were significantly higher in Xin 2. Linoleic acid (18:2) and linolenic acid (18:3) contents were higher in Qingxiang, whereas the palmitic acid content was greater in Xin 2 (P: palmitic acid; S: stearic acid; O: oleic acid; L: linoleic acid; Ln: linolenic acid).

Analysis of the degree of unsaturation of the TG molecules indicated that seven double bonds were the most common, thus showing a higher degree of unsaturation. Eight TG molecules had saturated carbon chains, but their content was lower ($\leq$ 10 μg/g). The remainder of the TG molecules were unsaturated. This result shows that both Qingxiang and Xin 2 contain large amounts of unsaturated fatty acids.

In the TG molecules, 50–56 carbon atoms were present. The main fatty acids connected with TG were medium-chain C16, C17, and C18 and the content of long-chain fatty acids was low ($\leq$ 10 μg/g). There were more C54 TG molecules in Qingxiang, whereas C52 TG molecules were slightly higher in Xin 2 (**Supplementary Figure 3B**).

Phospholipids were rich in walnut kernels, including PA, PG, PS, PC, PE, LPC, PI, LPE, CL, PIP, LPI, and LPG. The phospholipids in Qingxiang and Xin 2 accounted for

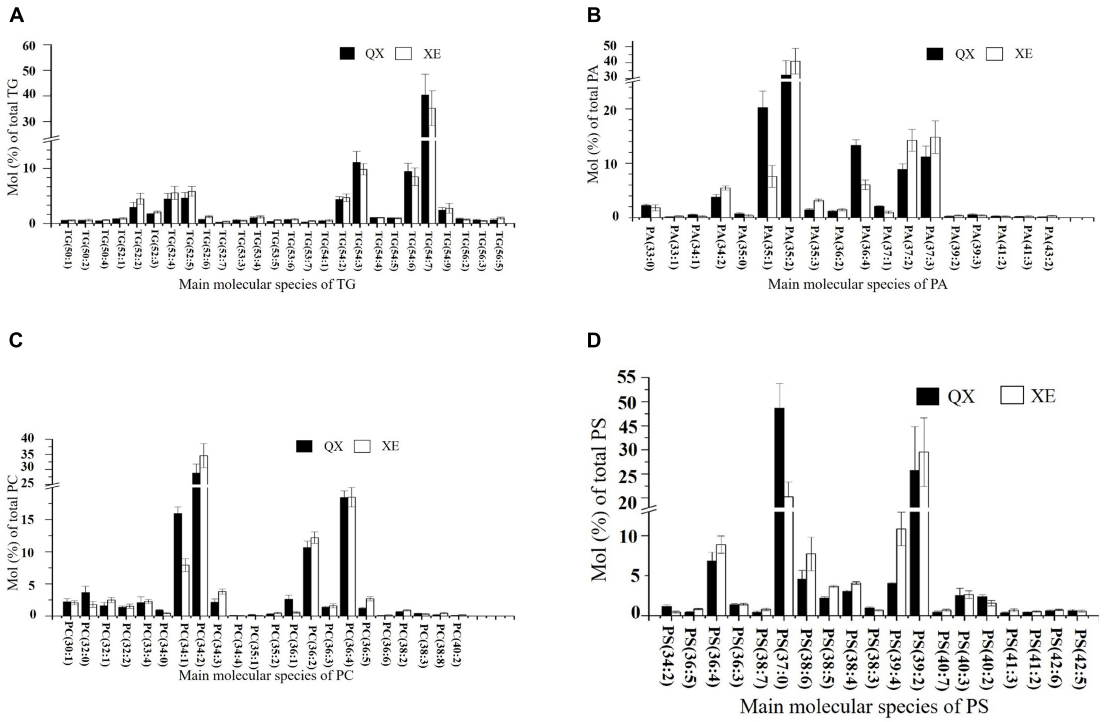


FIGURE 6 | The content of the main lipid molecules in kernels of the two walnut varieties. Major molecular species composition of TG **(A)**, PA **(B)**, PC **(C)**, and PS **(D)** determined by UHPLC-Orbitrap HRMS in the walnut varieties Qingxiang and Xin 2.

approximately 10.9 and 13.2% of the total lipids, respectively. The phospholipid content in Xin 2 was relatively high. PA was the main phospholipid subtype, which accounted for 4.65% of the total lipids in Qingxiang and 7.45% in Xin 2. The main molecular species detected in Qingxiang were PA 35:2 (17:0/18:2) (32.23%), PA 35:1 (17:0/18:1) (20.24%), and PA 36:4 (18:2/18:2) (13.29%), while PA 35:2 (17:0/18:2) (40.91%), PA 37:2 (14.21%), and PA 37:2 (14.79%) were main PAs in Xin 2 (**Figure 6B**). In both walnut varieties, PA containing C16:0, C18:1, and C18:2 accounted for more than 50%, and a small content of long-chain fatty acids, such as 40:2, 40:3, 41:2, 41:3, 42:2, 43:2, and 44:3, was also detected.

There was a little difference in PC content between Qingxiang and Xin 2, accounting for 1.17 and 1.07% of the total lipids, respectively. The main molecules in PCs were PC 34:1 (16:0/18:1), PC 34:2 (16:0/18:2), PC 36:2 (18:0/18:2), and PC 36:4 (18:2/18:2). PC mainly contained fatty acids C16:0, C18:0, and C18:1 (**Figure 6C**). The content of PC 34:1 (16:0/18:1) (p = 0.0337) and PC (18:0/18:1) (p = 0.0494) was higher in Qingxiang and the content of PC 34:2 (16:0/18:2) (p = 0.0327) was higher in Xin 2.

PS contains amino groups, which have antioxidant effects. Of the total lipids present, PS accounted for about 1.25% in Qingxiang and for 1.47% in Xin 2. The main PS molecules were PS 37:0, PS 39:2, PS 36:4, and PS 39:4 (**Figure 6D**).

SL (DGDG, MGDG, SQDG) in Qingxiang and Xin 2 accounted for 0.7 and 0.61% of the total lipids, respectively. DGDG was the main component of SL, accounting for 76.32 and 65.23% in Qingxiang and Xin 2, respectively. The more abundant molecules were DGDG (18:2/18:2), DGDG (18:2/18:3), and SQDG (39:12). Saccharolipids are the main components of the membrane lipid in walnut, although the content is relatively small. Saccharolipids have a variety of pharmacological functions, such as antiviral, antioxidant, antitumor, and anti-atherosclerosis activities (Schinitz and Ruebsaamen, 2010).

The SP in Qingxiang and Xin 2 walnut kernels accounted for only 0.03 and 0.01% of total lipids, respectively. The contents of Cer (d32:0) and Cer (d34:0) were higher in SP (including Cer and SM). As a secondary signal molecule of cells, SP promotes cell proliferation, apoptosis, and growth arrest; inhibits the occurrence and metastasis of tumors; and increases the sensitivity of tumors to chemotherapeutic drugs (Goldkorn et al., 2013). A small amount of sphingomyelin SM (d22:1) was detected in the two walnut lipids.

Lipid Metabolism of Walnut Analysis

Using the functional annotations of the expressed genes, the lipid metabolism-related genes in the transcriptome were identified. The proposed walnut lipid metabolism pathway map was generated corresponding to the main lipid molecules in the lipidome (**Figure 7**). A higher content of lipid molecules was detected in the lipid group, such as TG (36:4/18:3), TG (34:2/18:3), and TG (34:3/18:3). DGAT and PDAT related to TG synthesis were expressed in the transcriptome. In addition, there were many oleic acids (18:1) and the expression of PDH and ACCase related to oleic acid synthesis was also high. The content of lysophospholipids, saccharolipids, and sphingolipids

was low, and the expression of the corresponding synthetase genes was either low (CERS and MGD1) or undetected (LPGAT, LPGAT, and DGD) (**Supplementary Table 2**). Genes with high expression levels in the transcriptome corresponded to high levels of lipid metabolism molecules. Some genes related to lipid metabolism molecules with lower content were not detected in the transcriptome. Perhaps, low abundance RNA was below the detection sensitivity of our methods, or perhaps, an undescribed gene was present which would require further study.

Comparing the lipid data of the walnut varieties, the content of glyceride TG (18:2/18:2/18:3) was the highest in two kinds of walnuts, while the contents of TG (15:0/18:1/18:3), TG (16:1/18:3/18:3), TG (16:0/18:3/18:3), TG (17:0/18:2/18:3), TG (16:0/18:3/18:3) were higher in Xin 2. The fatty acid composition differed significantly between the oils of Qingxiang and Xin 2. TG synthesis in walnut has two pathways. The Kennedy pathway relies on acyl-CoA. The three acyltransferases (GPAT, LPAT, and DGAT) transfer the fatty acids of acyl-CoA to glycerol. Another pathway uses direct transfer of FA from PC to DG producing TG and LPC by the PDAT (Bates et al., 2013). Both DGAT and PDAT cooperated to produce TG. DGAT and PDAT can be used as target genes to regulate the oil content of walnut through molecular technology.

DG (36:4) had relatively higher content in the two kinds of walnuts, whereas DG (34:2) and DG (34:3) contents were relatively higher in Xin 2. The phospholipids were mostly C36 molecules. PA and DG are important intermediate products in lipid metabolism. Their synthesis starts with G-3-P and fatty acids as initial substrates and includes the endoplasmic reticulum pathway (eukaryotic pathway) and the plastid pathway (prokaryotic pathway), which occur in different subcellular locations (Hong et al., 2017). The sn-2 position of glycerolipid molecules synthesized by the prokaryotic pathway generally prefers C16:0, while the sn-2 position of the lipids derived from the eukaryotic pathway is C18:1 (Schmid-Siegert et al., 2016). There were many C18 fatty acids in the lipid molecules of Qingxiang and Xin 2, thus showing that the eukaryotic pathway is the primary pathway of glycerolipid synthesis in walnut. The analysis also found that MGDG and DGDG in the walnut lipid were mainly 36:5, indicating that the intermediate products DG and PA produced by the ER pathway were likely the main substrate sources of MGDG and DGDG.

DISCUSSION

Comparative Analysis of Lipid Composition

As an important oil tree species, walnut has high economic and nutritional value. Compared with other main nut crops, such as pistachios, cashews, peanuts, pecans, and almonds, walnuts have the most abundant phospholipids, accounting for 96 species (Song et al., 2018). Triglycerides were detected in the oils of walnut, sesame, water chestnut, hazelnut, and beechnut, with walnut oil mainly composed of highly unsaturated TG (54:6–8) (Bail et al., 2009). Research on the lipid composition of walnuts has been limited to the identification of the composition and content of single lipids such as fatty acids, phospholipids, and glycerides, and a systematic comparison of the total lipid composition of walnuts has not been done previously. The present study systematically analyzed and compared the lipid composition of the walnut varieties Qingxiang and Xin 2 and found a total of 525 lipid molecules in 20 subtypes. The lipid molecule contains 21 species of saturated fatty acids and 27 species of unsaturated fatty acids, including a low content of rare ultra-long-chain unsaturated fatty acids. The presence of these fatty acids indicates that special fatty acids dehydrogenase and elongase enzymes were likely responsible for the unsaturation and elongation of the glyceride chain (Ivanova et al., 2016).

There were more C54 TG molecules in Qingxiang than in Xin 2, and the C52 TG molecules in Xin 2 were slightly higher than in Qingxiang, both of which connected medium-chain fatty acids. Compared with long-chain fatty acid glycerides, medium-chain fatty acid triglycerides (MCT) in oils are easier to hydrolyze to produce unsaturated fatty acids. These fatty acids, which are absorbed easily by the body, can effectively reduce the levels of triglycerides and apolipoproteins and improve lipid metabolism (Fink et al., 2014). The unsaturation of TG molecules in Qingxiang was higher with greater linoleic acid and linolenic acid content. The quality of walnut oil mainly lies in the fact that it contains a large amount of unsaturated fatty acids, which can effectively reduce and prevent the occurrence of cholesterol, atherosclerosis, and heart disease (Ibáñez et al., 2017). Moreover, the oxidized linoleic acid will produce n-butyraldehyde and other volatile components that determine the flavor and taste of walnuts (Zhou et al., 2017). Qingxiang has good taste and high nutritional value, but its unsaturated fatty acids are oxidized easily, thus could reduce the shelf life of kernels and oil (Emilio and Mataix, 2006).

Twelve kinds of phospholipids were found in walnut oil encompassing 221 phospholipid molecules, which were the most abundant species. PA is a lipid signaling molecule that participates in various physiological processes, including signal transmission and response to environmental stress. Yu et al. (2010) found that under drought stress, PA 34:2, 34:3, 34:6, 36:3, and 36:6 increase significantly. The PA accounted for 7.45% in Xin 2, a relatively high content, which shows that Xin 2 is more resistant and adaptable. PG is rich in membrane lipids and is a biologically active lipid with antioxidant effects (Yang et al., 2007). PG (44:0) in Qingxiang and Xin 2 accounted for 98.37 and 97.44% of the total PG, respectively. These phospholipid molecules containing long-chain saturated fatty acids have antioxidant effects. The PS content in the Qingxiang and Xin 2 is low. However, PS contains amino groups, which can have synergistic antioxidant effects with vitamin E (Shen et al., 2013).

Lipidome-Combined Transcriptome Analysis

By combining the analysis of expressed genes and lipid metabolism molecules, preliminary metabolic pathways of the main lipid molecules in walnuts were constructed. Qingxiang

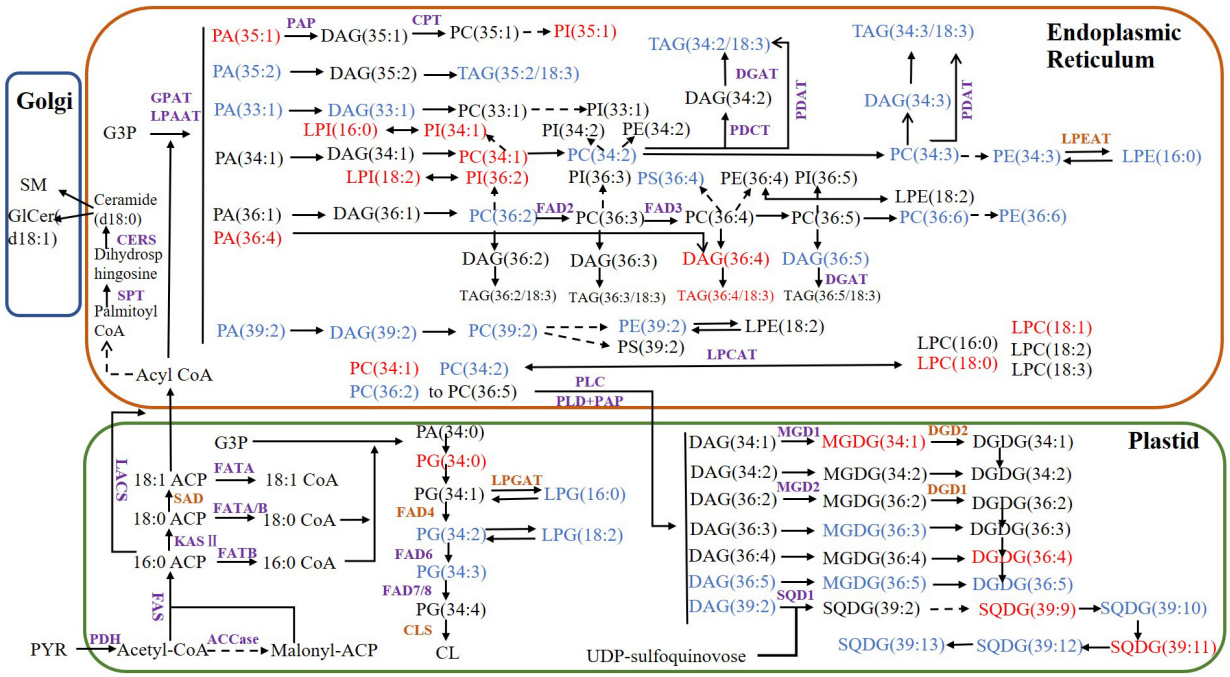


FIGURE 7 | Proposed lipid metabolism pathway map for walnut. The purple enzymes represent genes that have been detected in the transcriptome and the brown genes represent the undetected genes. The lipid molecules in blue have higher content of Xin 2 and those in red have higher content of Qingxiang. PYR, pyruvate; ACP, acyl carrier protein; PDH, pyruvate dehydrogenase; ACCase, acetyl-CoA carboxylase; acyl CoA, acetyl-coenzyme A; FAS, fatty acid synthase; SAD, stearoyl-ACP desaturase; LACS, long-chain acyl-CoA synthetase; CPT, CDP-choline:diacylglycerol cholinephosphotransferase; KASII: 3-ketoacyl-ACP synthase; DGAT, diacylglycerol acyltransferase; FAD2, oleoyl desaturase; FAD3, linoleoyl desaturase; FAD4, FAD6, FAD7/8, fatty acid desaturase; GAP, glyceraldehyde-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; PEP, phosphoenolpyruvate; MGD, monogalactosyldiacylglycerol synthase; DGD, digalactosyldiacylglycerol synthase; SQD, sulfoquinovosyldiacylglycerol synthase; CLS, cardiolipin synthase; SPT, serine palmitoyltransferase; CERS, ceramide synthase; SMS, sphingomyelin synthase; LPEAT, lysophosphatidylethanolamine acyltransferase; LPCAT, lysophosphatidylcholine acyltransferase; PLC, phospholipase C; PLD, phospholipase D; PAP, phosphatidic acid phosphatase.

contained more linoleic acid (18:2), while Xin 2 contained more palmitic acid (16:0) and long-chain fatty acids such as behenic acid. FAD2 and FAD3 control, respectively, the conversion of oleic acid to linoleic acid and linoleic acid to linolenic acid (Liu et al., 2020). The expression levels of FAD2 and FAD3 in the transcriptome were higher than those of FAD6 and FAD7/8. Primarily, linoleic and linolenic acids were found in the present study and their generation may have been catalyzed by FAD2 and FAD3 in the ER. ACCase is the key rate-limiting enzyme for the assembly of fatty acids. Analysis of the transcriptome showed that ACC-1 and ACC-2 expressed higher levels, which is likely related to the high oil content in walnut kernels. The genes of FATA and FATB were also expressed, with that of FATA being the higher of the two. These relative expression levels are likely the reason why many unsaturated C18 fatty acids were present.

The molecular composition of TG differed greatly between the two walnut varieties. Both DGAT1 and DGAT2 in the transcriptome were expressed in varieties. A previous work has shown that they may play different roles during plant development and produce TG with different fatty acid components (Oakes et al., 2011). It is likely that these two enzymes are related to the varietal differences in TG species observed in the preset study. Furthermore, PDAT has different expression levels during walnut kernel development, which may be related to the accumulation of walnut oil (Bernard et al., 2018).

CONCLUSION

In the present study, the UHPLC-Orbitrap HRMS system was used to compare the lipid content and composition in the kernels of the walnut varieties Qingxiang and Xin 2. Combined with transcriptome data, we constructed a preliminary molecular regulatory network of the main lipid metabolism in walnut. A total of 525 lipid molecules were identified in Qingxiang and Xin 2, consisting of 250 glycerides (including DG and TG), 221 phospholipids (including PA, PG, PS, PC, PE, LPC, PI, LPG, LPI, LPE, PIP, and CL), 18 types of glycolipids (including DGDG, MGDG, and SQDG), and 36 types of sphingolipids (including Cer and SM). The fatty acid chains in DG and TG are mainly composed of essential fats such as oleic acid, linoleic acid, and linolenic acid. The walnut lipid profile and the lipid metabolism pathway constructed here have important theoretical

and practical value for further study of walnut lipid metabolism and functional development.

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.ebi.ac.uk/metabolights/index, MTBLS2888.

AUTHOR CONTRIBUTIONS

JZhe and JX designed the experiment. CY, JW, YW, HZ, BW, XW, LQ, and JZha performed the analysis and interpretation of the data. SY and XZ drafted the manuscript. JZhe and PM revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the State Key Laboratory of Sustainable Dryland Agriculture, Shanxi Agricultural University (No. 202105D121008-2-1), the Key Research and Development Program of Shanxi province (No. 201803D221008-11), and Research Program Sponsored by the Agricultural Science Research of Shanxi Academy of Agricultural Sciences (Nos. 2020xshf45 and YCX2020BH2).

ACKNOWLEDGMENTS

We gratefully acknowledge help from Professor Janli Chen and Mike Bonman, University of Idahoin in United States, for improving and editing our English language.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.715731/full#supplementary-material

Supplementary Figure 1 | Chromatogram in ESI positive **(A)** and negative ion modes **(B)** of QX.

Supplementary Figure 2 | (A) Secondary mass spectrometry of DG (18:2/18:2) (m/z 634.5405) generated fragmentation ion; **(B)** bubble chart, bubble in figure present significant difference of lipid molecules, different colors present different lipid subclasses.

Supplementary Figure 3 | (A) Score and loading plots generated from PLS-DA classifying the lipid of QX and XE. **(B)** The content of different numbers of carbon in TGs of two walnut varieties.

Supplementary Table 1 | Model evaluation parameters: R^2X means the explanatory rate for the model of variable X; R^2Y means the explanatory rate for the model of variable Y; Q^2 evaluates the predictive power of PLS-DA model.

Supplementary Table 2 | The lipid molecules in lipid metabolism pathway of walnut kernel. QX (μg/g) and XE (μg/g) are the means four replicates ± standard deviation.

Supplementary Table 3 | The enzymes involved in main lipid metabolism pathway of walnut kernel. FPKM means the genes of these enzymes expression level. ND means the genes were not detected.

REFERENCES

Bail, S., Stuebiger, G., Unterweger, H., Buchbauer, G., and Krist, S. (2009). Characterization of volatile compounds and triacylglycerol profiles of nut oils using SPME-GC-MS and MALDI-TOF-MS. *Eur. J. Lipid Sci. Technol.* 111, 170–182. doi: 10.1002/ejlt.200800007

Bates, P. D., Stymne, S., and Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol.* 16, 1–7.

Bernard, A., Lheureux, F., and Dirlewanger, E. (2018). Walnut: past and future of genetic improvement. *Tree Genet. Genomes* 14, 1–28. doi: 10.1007/s11295-017-1214-0

Bouabdallah, I., Bouali, I., Martinez-Force, E., Albouchi, A., Perez Camino, M. C., Boukhchina, S., et al. (2014). Composition of fatty acids, triacylglycerols and polar compounds of different walnut varieties (*Juglans regia L.*) from Tunisia. *Nat. Prod. Res.* 28, 1826–1833. doi: 10.1080/14786419.2014.950573

Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.

Croitoru, A., Ficai, D., Craciun, L., Ficai, A., and Andronescu, E. (2019). Evaluation and exploitation of bioactive compounds of walnut, *Juglans regia*. *Curr. Pharm. Des.* 25, 119–131. doi: 10.2174/1381612825666190329 150825

Emilio, R., and Mataix, J. (2006). Fatty acid composition of nuts-implications for cardiovascular health. *Br. J. Nutr.* 96, S29–S35.

Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R., Shimizu, T., et al. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* 50:S9.

Fink, A., Rüfer, C. E., Le Grandois, J., Roth, A., Aoude-Werner, D., Marchioni, E., et al. (2014). Dietary walnut oil modulates liver steatosis in the obese Zucker rat. *Eur. J. Nutr.* 53, 645–660. doi: 10.1007/s00394-013-0573-z

Goldkorn, T., Chung, S., and Filosto, S. (2013). Lung cancer and lung injury: the dual role of ceramide. *Handb. Exp. Pharmacol.* 216, 93–113. doi: 10.1007/978-3-7091-1511-45

Hamide, Z. S., Vural, G., and Ebru, A. S. (2015). Future perspectives in Orbitrap^{TM} high resolution mass spectrometry in food analysis -a review. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* 32, 1568–606. doi: 10.1080/19440049.2015.1057240

Hong, C., Fang, W., and Xu, Y. D. (2017). Lipidomics in food science. *Curr. Opin. Food Sci.* 16, 80–87. doi: 10.1016/j.cofs.2017.08.003

Huang, R. M., Zhou, Y., Zhang, J. P., Ji, F., Jin, F., Fan, W., et al. (2021). Transcriptome analysis of walnut (*Juglans regia L.*) embryos reveals key developmental stages and genes involved in lipid biosynthesis and polyunsaturated fatty acid metabolism. *J. Agric. Food Chem.* 69, 337–396.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Ibáñez, C., Mouhid, L., Reglero, G., and Ramírez de Molina, A. (2017). Lipidomics insights in health and nutritional intervention studies. *J. Agricult. Food Chem.* 10, 1–45. doi: 10.1002/9781118683682.part1

Ivanova, S., Marinova, G., and Batchvarov, V. (2016). Comparison of fatty acid composition of various types of edible oils. *Bulgar. J. Agric. Sci.* 22, 849–856.

Lee, H. C., and Yokomizo, T. (2018). Applications of mass spectrometry-based targeted and non-targeted lipidomics. *Biochem. Biophys. Res. Commun.* 504, 576–581. doi: 10.1016/j.bbrc.2018.03.081

Li, M., Zhou, Z., Nie, H., Bai, Y., and Liu, H. (2011). Recent advances of chromatography and mass spectrometry in lipidomics. *Anal. Bioanal. Chem.* 399, 243–249. doi: 10.1007/s00216-010-4327-y

Liu, H., Hong, Y., Lu, Q., Li, H., Gu, J., Ren, L., et al. (2019). Integrated analysis of comparative lipidomics and proteomics reveals the dynamic changes of lipid molecular species in high oleic acid peanut seed. *J. Agric. Food Chem.* 68, 426–438. doi: 10.1021/acs.jafc.9b04179

Liu, K., Zhao, S., Wang, S., Wang, H., and Zhang, Z. (2020). Identification and analysis of the FAD gene family in walnuts (*Juglans regia L.*) based on transcriptome data. *BMC Genom.* 21:299. doi: 10.1186/s12864-020-6692-z

Martínez, M. L., Labuckas, D. O., Lamarque, A. L., and Maestri, D. M. (2010). Walnut (*Juglans regia L.*): genetic resources,chemistry, by-products. *J. Sci. Food Agric.* 90, 1959–1967. doi: 10.1002/jsfa.4059

Oakes, J., Brackenridge, D., Colletti, R., Daley, M., Hawkins, D. J., Xiong, H., et al. (2011). Expression of fungal diacylglycerol acyltransferase 2 genes to increase kernel oil in maize. *Plant Physiol.* 155, 1146–1157. doi: 10.1104/pp.110.167676

Qi, Y. X., Liu, Y. B., and Rong, W. H. (2011). RNA-Seq and its applications: a new technology for transcriptomics. *Hereditas* 33, 1191–1202. doi: 10.3724/sp.j.1005.2011.01191

Rothenberg, D. O. N., Yang, H. J., Chen, M., Zhang, W., and Zhang, L. (2019). Metabolome and transcriptome sequencing analysis reveals anthocyanin metabolism in pink flowers of anthocyanin-rich tea (Camellia sinensis). *Molecules* 24:1064. doi: 10.3390/molecules24061064

Schinitz, G., and Ruebsaamen, K. (2010). Metabolism and atherogenic disease association of Lysophosphatidycholine. *Atherosclerosis* 208, 10–18. doi: 10.1016/j.atherosclerosis.2009.05.029

Schmid-Siegert, E., Stepushchenko, O., Glauser, G., and Farmer, E. E. (2016). Membranes as structural antioxidants: recycling of malondialdehyde to its source in oxidation-sensitive chloroplast fatty acids. *J. Biol. Chem.* 291, 13005–13013.

Shen, Q., Dong, W., Yang, M., Li, L., Wang, Y., Cheung, H.-Y., et al. (2013). Lipidomic fingerprint of almonds (Prunus dulcis L. cv Nonpareil) using TiO2 nanoparticle based matrix solid-phase dispersion and MALDI-TOF/MS and its potential in geographical origin verifification. *Agric. Food Chem.* 61, 7739–7748. doi: 10.1021/jf4016448

Shiu-Cheung, L., and Randall, J. W. (2006). Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. *Lipids* 41, 1073–1087. doi: 10.1007/s11745-006-5057-y

Song, S., Cheong, L. Z., Wang, H., Man, Q. Q., Pang, S. J., Li, Y. Q., et al. (2018). Characterization of phospholipid profiles in six kinds of nut using HILIC-ESIIT-TOF-MS system. *Food Chem.* 240, 1171–1178. doi: 10.1016/j.foodchem.2017.08.021

Taguchi, R., and Ishikawa, M. (2010). Precise and global identification of phospholipid molecular species by an Orbitrap mass spectrometer and automated search engine Lipid Search. *J. Chromatogr. A.* 1217, 4229–4239. doi: 10.1016/j.chroma.2010.04.034

Wang, S. L., Nie, C. X., Fu, X., and Li, W. B. (2012). Introduction and cultivation techniques of walnut with fresh fragrance. *Northern Fruit* 5, 27–29.

Yang, L., Gordon, K. S., and Wei, S. (2014). FeatureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Yang, S., Qiao, B., Lu, S. H., and Yuan, Y. J. (2007). Comparative lipidomics analysis of cellular development and apoptosis in two Taxus cell lines. *Biochim. Biophys. Acta* 1771, 600–612. doi: 10.1016/j.bbalip.2007.02.011

Yu, L., Nie, J., Cao, C., Jin, Y., Yan, M., Wang, F., et al. (2010). Phosphatidic acid mediates salt stress response by regulation of MPK6 in Arabidopsis thaliana. *N. Phytol.* 188, 762–773. doi: 10.1111/j.1469-8137.2010.03422.x

Zambón, D., Sabaté, J., Muñoz, S., Campero, B., Casals, E., Merlos, M., et al. (2000). Substituting walnuts for monounsaturated fat improves the serum lipid profile of hypercholesterolemic men and women. A randomized crossover trial. *Ann. Inter. Med.* 132, 538–546. doi: 10.7326/0003-4819-132-7-200004040-00005

Zhao, S., Zhang, X., Su, Y., Chen, Y., Liu, Y., Sun, M., et al. (2018). Transcriptome analysis reveals dynamic fat accumulation in the walnut kernel. *Int. J. Genomics* 2018:8931651.

Zhao, X., Yang, G., Liu, X., Yu, Z., and Peng, S. (2020). Integrated analysis of seed microRNA and mRNA transcriptome reveals important functional genes and microRNA-targets in the process of walnut (Juglans regia) seed oil accumulation. *Int. J. Mol. Sci.* 21, 90–93.

Zhou, Y., Fan, W., and Chu, F. X. (2018). Identification of Volatile Oxidation Compounds as Potential Markers of Walnut Oil Quality. *J. Food Sci.* 83, 2745–2752. doi: 10.1111/1750-3841.14342

Zhou, Y., Fan, W., Zhang, J. P., and Dong, P. (2017). Comparison of volatile components between pressed and refined walnut oil and identification with electronic nose. *China Oils Fats* 42, 130–134.

Zitouni, M., Wewer, V., Dörmann, P., Abdelly, C., and Youssef, N. B. (2016). Quadrupole time-of -flight mass spectrometry analysis of glycerophospholipid molecular species in the two halophyte seed oils: eryngium maritimum and Cakile maritima. *Food Chem.* 213, 319–328. doi: 10.1016/j.foodchem.2016.06.083

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.

Copyright © 2021 Yan, Wang, Yang, Wang, Wang, Wu, Qiao, Zhao, Mohammad, Zheng, Xu, Zhi and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 September 2021
doi: 10.3389/fgene.2021.738649

Transcriptomic Time-Series Analyses of Gene Expression Profile During Zygotic Embryo Development in *Picea mongolica*

Jia Yan[1,2†], *Ha buer*[1†], *Ya ping Wang*[1], *Gegen zhula*[1] *and Yu´e Bai*[1*]

[1]*Institute of Forest Tree Genetic Breeding, Forestry College, Inner Mongolia Agricultural University, Hohhot, China,* [2]*Life Science of College, Inner Mongolia University, Hohhot, China*

Zygotic embryogenesis is a critical process during seed development in gymnosperms. However, knowledge on the genome-wide transcriptional activation that guides this process in conifers is limited, especially in *Picea mongolica*. This tree species is endemic to semiarid habitats of Inner Mongolia in China. To extend what is known about the molecular events underpinning its zygotic embryogenesis, comparative transcriptomic analyses of gene expression in zygotic embryos were performed by RNA sequencing in *P. mongolica*. Our results showed that most changes in transcript levels occurred in the early embryonic pattering determination and formation of mature embryos. Transcripts related to embryogenic competence, cell division pattern, hormones, and stress response genes were identified during embryogenesis. Auxin is essential for early embryo patterning and pre-cotyledon embryonic formation. However, ABA is a major regulator of embryo maturation. Moreover, we found that methylation-related gene expression is associated with activation of early-stage embryos, late embryogenesis abundant proteins, and storage/energy-related genes with late and mature embryos. Furthermore, network analysis revealed stage-specific and multistage gene expression clusters during embryogenesis. WOX, MYB, AP2, and HLH transcription factors seem more relevant to embryogenesis in different stages. Our results provide large-scale and comprehensive transcriptome data for embryo development in *P. mongolica*. These data will lay a foundation for the protection and utilization of *P. mongolica* resources.

Keywords: ***Picea mongolica*, embryogenesis, zygotic embryo, transcriptomics, WGCNA**

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Guanzheng Qu,
Northeast Forestry University, China
Chunjie Fan,
Chinese Academy of Forestry, China

***Correspondence:**
Yu´e Bai
baiyue@imau.edu.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: *09 July 2021*
Accepted: *01 September 2021*
Published: *29 September 2021*

Citation:
Yan J, buer H, Wang Yp, zhula G and Bai Y (2021) Transcriptomic Time-Series Analyses of Gene Expression Profile During Zygotic Embryo Development in Picea mongolica. *Front. Genet. 12:738649. doi: 10.3389/fgene.2021.738649*

INTRODUCTION

Embryonic development is a critical reproductive phase of sexually reproducing plant species. Embryogenesis takes place from the one-cell zygote stage to a mature embryo in plant seeds. The key events and regulation pathways of reproductive development and embryogenesis have been widely studied in model angiosperms, such as *Arabidopsis thaliana* (Hofmann et al., 2019; Dresselhaus and Jürgens, 2021; Hao et al., 2021). However, the gene regulation mechanism associated with embryogenesis progression in gymnosperms is still limited (Trontin et al., 2016). Moreover, striking differences have been observed during conifer embryo development, which have a brief period of free nuclear divisions before cellularization of the embryo (Singh, 1978). Therefore, the molecular regulation mechanism must be different in conifers.

Recently, with the establishment of plant regeneration *via* the somatic embryogenesis pathway, molecular research on embryogenesis has received extensive attention in conifers (Llebrés et al., 2018; Rodrigues et al., 2018; Liu et al., 2019; Rodrigues et al., 2019). It has been reported that douglas-fir LEAFY COTYLEDON1 (PmLEC1) is an active transcription factor during zygotic and somatic embryogenesis (Vetrici et al., 2021). Thermospermine synthase (ACL5) and diamine oxidase (DAO) expression is needed for zygotic embryogenesis in Scots pine (Vuosku et al., 2019). PaWOX2 and PaWOX8/9 are expressed in both zygotic and somatic embryos during early embryo development in *P. abies* (Palovaara et al., 2010). Although these results provide additional basic knowledge for further improvement of somatic embryogenesis in conifers, knowledge concerning the regulation of embryo development is still limited.

Picea mongolica is a species native to Inner Mongolia of China. It has adapted to extreme and highly stressful environments and has great commercial and ecological value for the construction of urban green spaces in arid areas (Lorenz et al., 2012). The conventional propagation methods for *P. mongolica* are mainly through seeds and cuttings. However, the bad growth environment affects the development of seeds, while propagation through cuttings is hindered by rooting problems. In a previous study, we established an effective somatic embryo (SE) reproduction system, the best pathway to enhance *P. mongolica* yield (Yan et al., 2021). We found that embryos of pre-cotyledon embryos stage are the best explants for EC induction. Therefore, a better understanding of the regulation mechanism and key genes involved in embryogenesis is necessary to improve the efficiency of our SE protocols.

It is known that large genomes and large amounts of repetitive DNA in conifers have severely restricted efforts to produce a conifer reference genome and characterize expressed sequences. To this end, RNA-Seq is an essential tool and advanced approach for transcriptome-wide analysis of differential gene expression. Recently, reference transcriptomes of many conifers during the period of zygotic embryogenesis have been reported, including maritime pine, Scots pine (*Pinus sylvestris*), and *Pinus pinaster* (de Vega-Bartol et al., 2013a; Vuosku et al., 2015; Merino et al., 2016). Based on the results of these studies, the transcripts associated with carbohydrate metabolism, monosaccharide transport, and epigenetic regulation may serve as reliable molecular markers for early embryogenesis (Vuosku et al., 2015; Merino et al., 2016).

In the present study, to gain insight into the regulatory pathways associated with zygotic embryo development and reveal the transcriptomic activation mechanism during embryogenesis, we sequenced the transcriptome of zygotic embryo (ZE) over several developmental stages covering most embryogenesis stages in *P. mongolica*. Then, differentially expressed genes (DEGs) were identified during seed development. Many genes related to transcriptional regulation, signal transduction, and metabolic pathways were identified. Additionally, weighted gene co-expression network analysis (WGCNA) was performed to reveal key genes for specific time points in seed development. Our results provide data on molecular and biochemical events associated with ZE development, a foundation for improving the vegetative propagation of conifers *via* somatic embryogenesis in *P. mongolica*. Moreover, this work expands our understanding of the development of ZEs in conifers.

MATERIALS AND METHODS

Plant Material

Immature seeds of *P. mongolica* were obtained every 10 days after pollination between June and September 2019, from adult trees in the Baiyinaobao Nature Reserve of Inner Mongolia, China. Cones were periodically collected every week for all desired developmental stages of ZEs. The megagametophytes containing different developmental stages of ZEs were divided for RNA sequencing based on stereomicroscope observations (von Arnold et al., 2002), as follows: SDYS1 collected after 1 week of fertilization, which likely included early embryos in the proembryogenic stage, SDYS2 including the cleavage polyembryony in the early embryogenic stage, SDYS3 including the dominant embryos, SDYS4 including the pre-cotyledonary embryos, SDYS5 including the cotyledonary embryo, and SDYS6 including the mature embryo (**Figure 1**).

RNA Sample Preparation

For transcriptome sequencing, total RNA was obtained using the Plant RNA kit (Sigma-Aldrich, St. Louis, MO, United States) following the manufacturer's instructions. RNA degradation and contamination were detected in 1% agarose gels. RNA quantity and integrity were assessed using the NanoPhotometer spectrophotometer (NanoPhotometer purchased from Implen, Munich, Germany) and RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, United States). Total RNA (1 μg) was used to generate sequencing libraries using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, United States) sequencing.

DEG Statistics

Differential expression analysis was performed using the DESeq2 R package based on all obtained reads per kilo base per million mapped reads (RPKM) (1.16.1) (Love et al., 2014). Pairwise differential expression analysis was conducted among six group samples in sequence. Each set of samples contains three replicates. Transcripts that have log2FC $\geq$ 1 or log2FC $\leq$ 1 and p-value $<$ 0.05 were used to determine DEGs. Further analysis will be based on the fold-change value of EEGs using Expander 7 software (Ulitsky et al., 2010) with the K-means algorithm (Shamir et al., 2005).

Functional Annotation and GO Analysis

The agriGO (version 2.0) (http://systemsbiology.cau.edu.cn/agriGOv2/) was used in Gene Ontology (GO) enrichment analysis of obtained DEGs, which was corrected by a p value less than 0.05. The cluster Profiler R package was used to test the enrichment of DEGs in KEGG pathways. Venn diagrams were constructed using online software (http://bioinformatics.psb.ugent.be/webtools/Venn/). Hierarchical clustering heat maps were built, and cluster analyses were conducted using MultiExperiment Viewer (http://www2.heatmapper.ca/expression/).

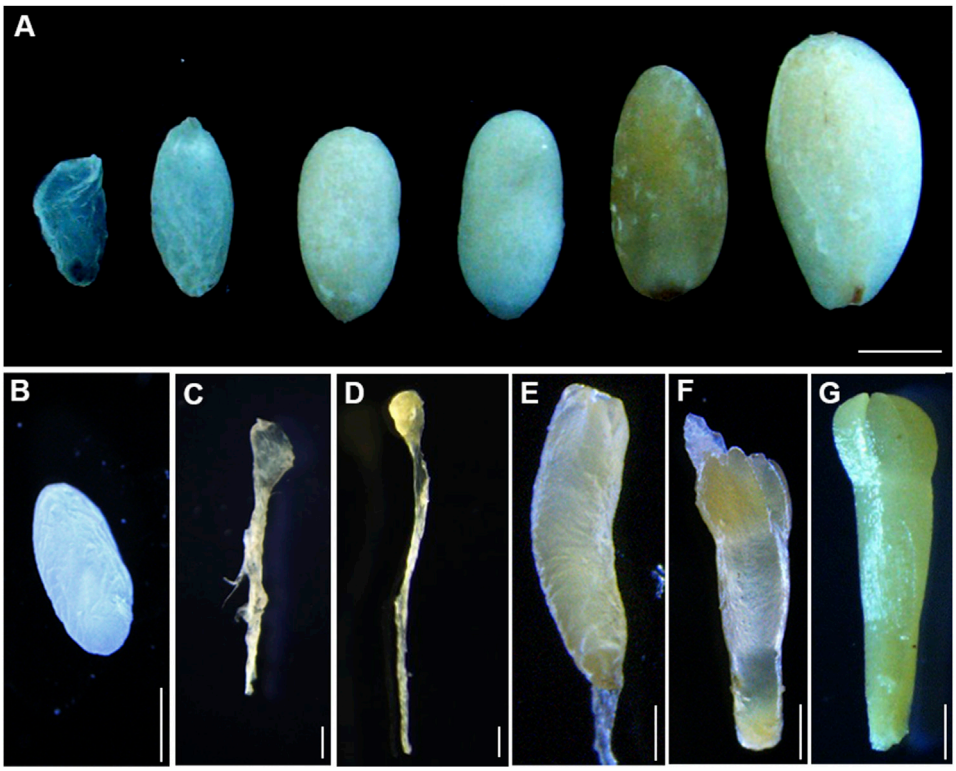


FIGURE 1 | Development stages of immature zygotic embryos. **(A)** Developing seeds collected from open-pollinated trees on different dates after fertilization. Bars 1 mm. **(B–D)** Immature zygotic embryo after fertilization. **(B)** Megagametophy, **(C)** cleavage polyembryony, **(D)** dominant embryos, **(E)** pre-cotyledon embryos, **(F)** cotyledon embryos, and **(G)** mature embryos. Bars 0.1 mm.

Gene Co-Expression Network

The R package WGCNA V1.41-1 was used for WGCNA analysis (Langfelder and Horvath, 2008). After filtering, the automatic network construction function blockwiseModules was used to construct co-expression modules based on gene expression values. Cluster correlation was used to construct cluster families within the network. The network image was created using Cytoscape software.

Quantitative Real-Time PCR

Total RNA was obtained from each sample using an RNA extraction kit (Takara, Japan), and cDNA was obtained using the PrimeScript™ RT Reagent Kit (Takara, Japan). PCRs were performed using AceQ qPCR SYBR Green Master Mix with the following thermal cycling conditions: 95°C for 10 min, 40 cycles of 95°C for 10 s, and 60°C for 60 s (de Vega-Bartol et al., 2013b). Relative gene expression was quantified using the 2−ΔΔCt method (Cao et al., 2016).

RESULTS

Dynamics of Gene Expression During the Development of ZE

To evaluate the molecular mechanisms governing the embryogenesis of *P. mongolica*, we isolated ZEs from June to September after pollination (**Figure 1**). Gene expression profiles were developed using RNA-seq as described in the methods section. After trimming adapter sequences and removing low-quality and multi-matched reads, more than 136.7 million unique clean reads and 20.5- to 30.4-G clean bases remained. A total of 76–80% clean reads were mapped to the reference genome of the *Norway spruce*. All samples were highly correlated and therefore clustered together (**Figure 2A**). Moreover, a principal component analysis (PCA) result demonstrated that the three replicates of each developmental stages can be clearly distinguished (**Figure 2B**).

We further analyzed DEGs through pairwise comparisons between consecutive stages of development. A two-fold difference (log2FC ≥ 1 or ≤ −1) and false discovery rate of 0.05 or less were used as thresholds. Each pairwise comparison identified a large proportion of DEGs, including 12,473 in proembryogenic versus cleavage polyembryony stages, 8,868 in cleavage polyembryony versus dominant embryo stages, 5,389 in dominant embryos versus pre-cotyledonary embryos, 9,077 in pre-cotyledonary embryos versus cotyledonary embryos, 15,878 in cotyledonary embryos versus mature embryos, and 16,987 in mature embryos versus proembryo embryos (**Figure 2C**). Next, we selected eight transcripts as candidate genes for further investigation of their expression patterns using RT-qPCR to evaluate their transcriptome data. The expression of these genes was consistent with the transcriptome data (**Figure 3**).

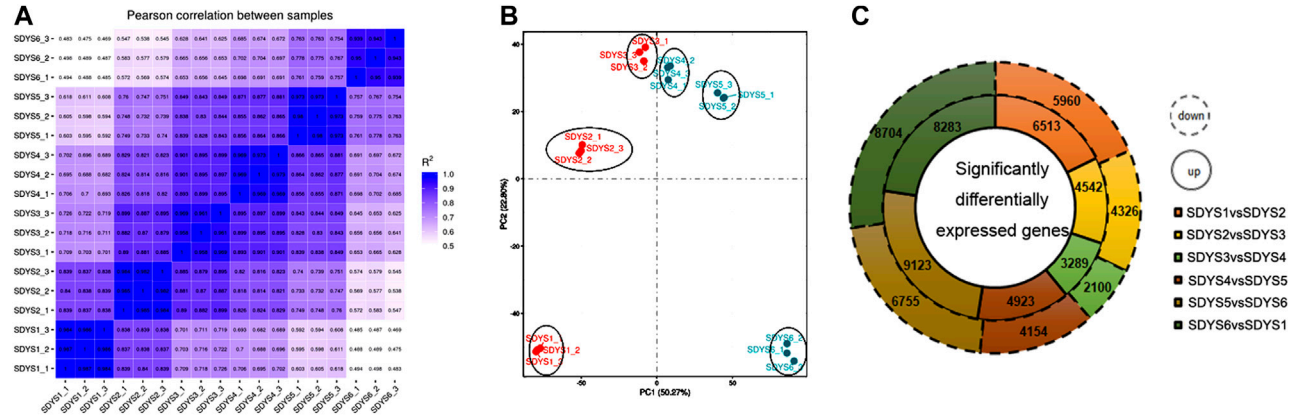


FIGURE 2 | Transcription during embryonic development of *P. mongolica*. **(A)** Correlation analysis among all samples. **(B)** Principal component analysis of the transcriptomes of samples from six stages. **(C)** Differentially expressed genes in each stage.

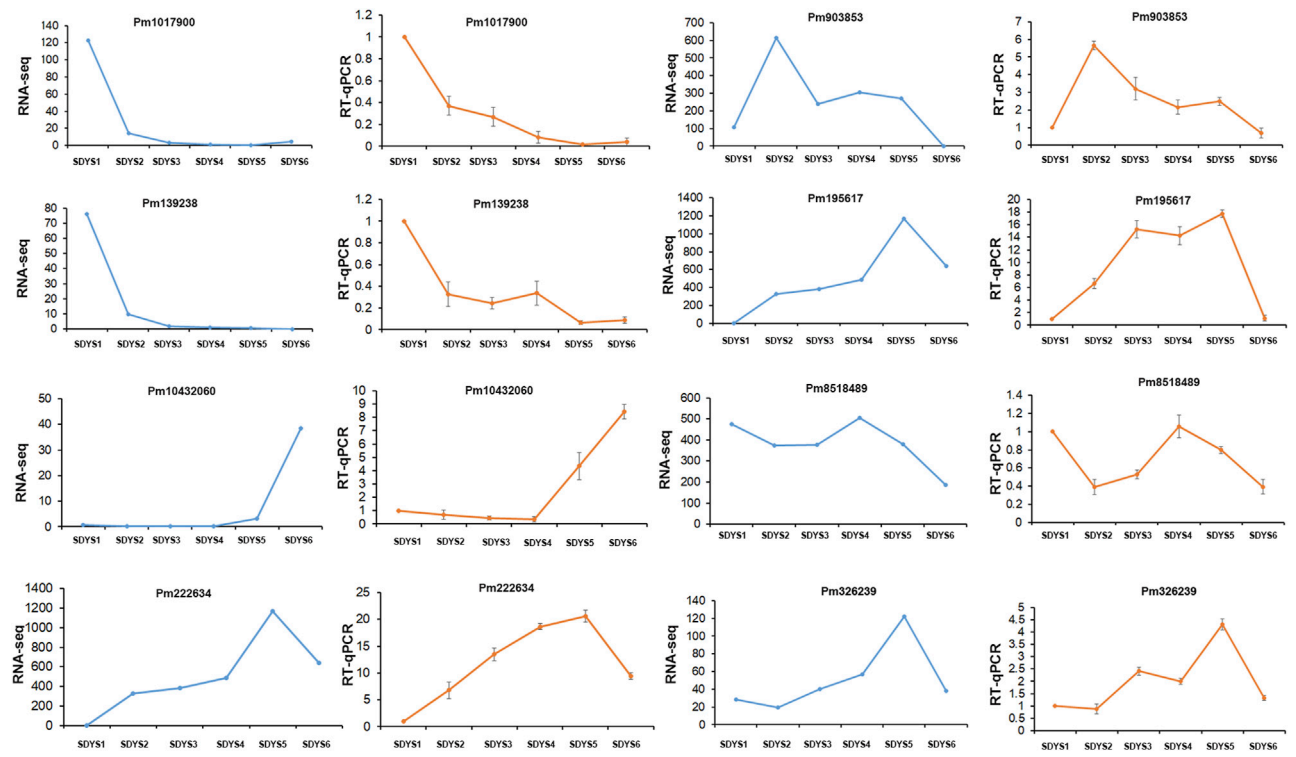


FIGURE 3 | Validation of RNA-seq results *via* RT-qPCR Fold changes for selected transcripts obtained by RNA-seq and RT-qPCR are shown for each developmental time point. Transcript levels are means ± SD of three biological replicates.

Based on the significantly expressed genes, a heat map was generated after background subtraction and normalization to represent the global dynamics of gene expression during zygotic embryogenesis in *P. mongolica* (**Figure 4A**). Furthermore, Venn diagram analysis showed that 5,039 transcripts were differentially expressed from stages 1 to 2, 1,606 from stages 2 to 3, 639 DEGs from stages 3 to 4, 1,646 from stages 4 to 5, and 4,425 from stages 5 to 6. Additionally, 1,279 genes appeared in five comparisons across all six stages of embryo development (**Figure 4B**). In all, the greatest difference in the number of DEGs was in the early embryogenic stage and later to mature embryonic stage. In contrary, little difference was observed between stages 3 and 4. Based on these results, embryonic developmental stages can be divided into the following: stage 1 is the proembryogenic phase; stage 2 is the early embryogenic phase; stages 3–4 are

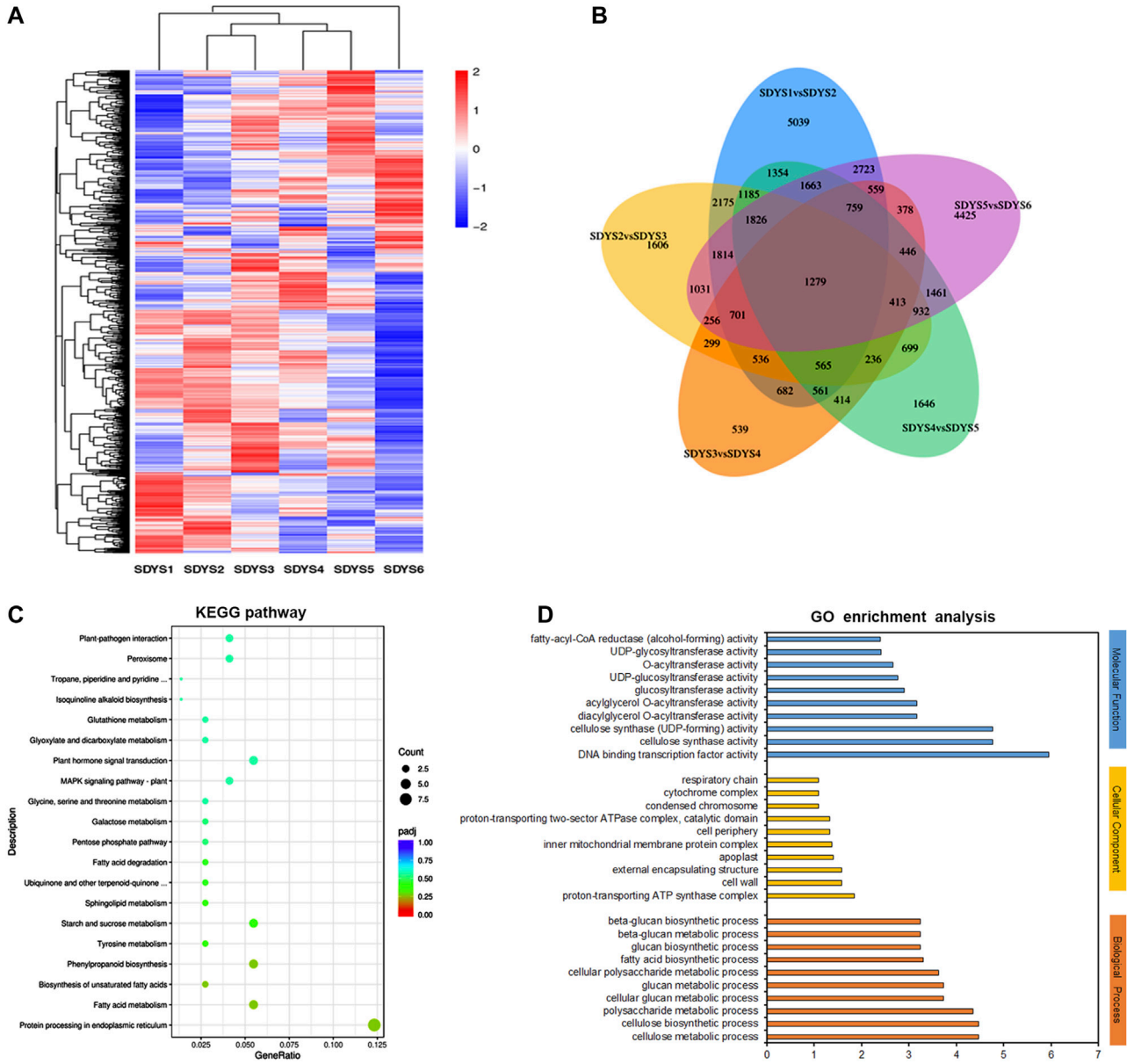


FIGURE 4 | Identification and analysis of differentially expressed genes during embryonic development stages. **(A)** Heat map, **(B)** Venn diagram, **(C)** KEGG annotation, and **(D)** GO enrichment analysis of differentially expressed genes in six embryonic development stages.

the mid-embryogenic phase; stage 5 is the latter embryogenic phase, and stage 6 is the mature embryo in *P. mongolica*.

Functional Regulation of DEGs During the Development of ZE

Furthermore, we classified these DEGs based on KEGG pathways, which were mainly enriched in protein processing of the endoplasmic reticulum, fatty acid metabolism, starch and sucrose metabolism, phenylpropanoid biosynthesis, plant hormone signal transduction, and MAPK signaling pathway (**Figure 4C**). Subsequently, GO enrichment analysis was performed to further characterize their biological roles. Most biological process (BP) terms were enriched in cellulose metabolic and biosynthetic processes, polysaccharide metabolic processes, and glucan metabolic processes. In the molecular function (MF) category, DNA-binding transcription factor activity and cellulose synthase activity were significantly enriched. The significantly enriched cellular component (CC) terms were cell wall and apoplast (**Figure 4D**). In brief, many DEGs were involved in the regulation of embryonic development, and these pathways have great significance for further revealing the molecular events involved in zygotic embryogenesis in *P. mongolica*.

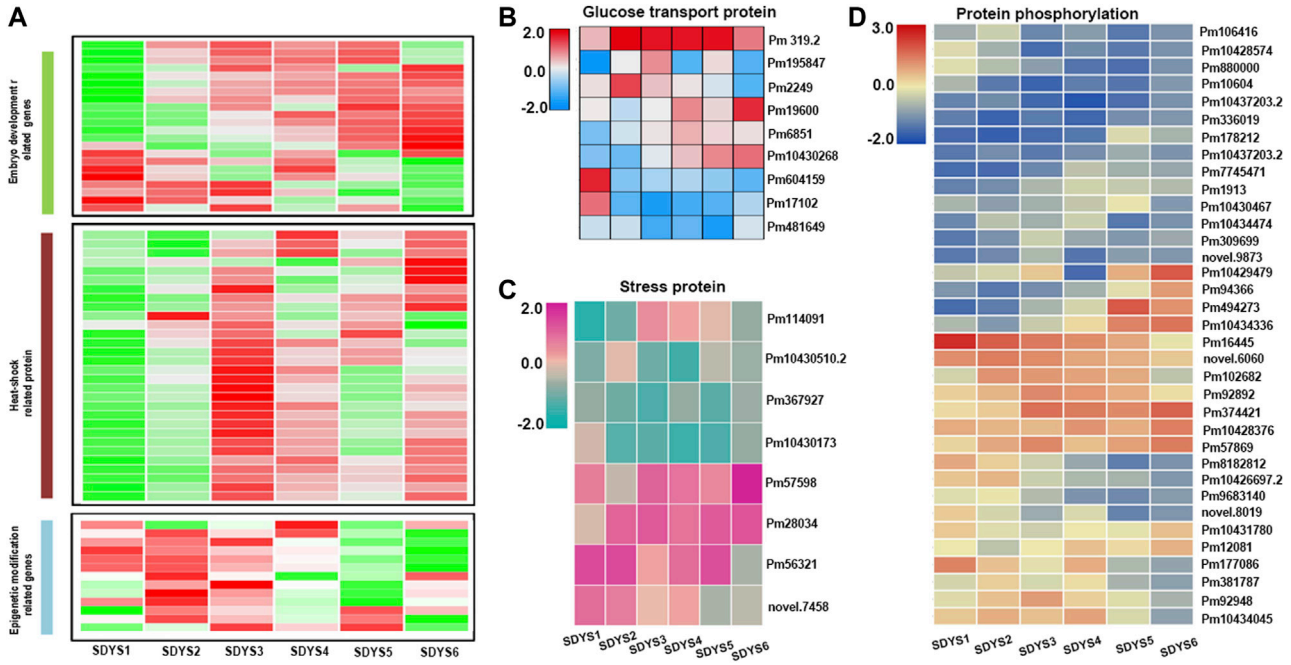


FIGURE 5 | Clustering of differentially expressed genes identified in five comparisons across all six stages of embryo development. **(A)** Heat map of differentially expressed genes. **(B)** Glucose transport genes, **(C)** stress response genes, and **(D)** protein kinases and phosphatases.

DEG Identification Across Whole-Embryo Developmental Stages

Among the DEGs identified, 1,279 genes appeared in five comparisons across all six stages of embryo development. Based on functional annotations and accumulated data on model plant systems, statistical results showed that the expressions of all six LEA genes, seven seed storage proteins, three seed maturation protein-encoding genes, and one seed dormancy gene were different in every developmental stage (**Figure 5A**). These associated genes were enriched in post-embryonic development, seed development, and fruit development (**Supplementary Figure S1**). In addition, many epigenetic regulation components were found during embryogenesis. Several putative histone subunit homologs and methyltransferases had different expression profiles across developmental stages. Additionally, genes associated with glucose transport and stress protein were also differentially expressed (**Figures 5B–D**). Five sugar efflux transporters and three sugar transporter genes involved in glucose metabolism and the transport pathway were expressed during ZE maturation. Moreover, stress response genes and HSP family genes were obviously differentially expressed in all six developmental stages during ZE maturation (**Supplementary Figure S1**). Importantly, several tyrosine phosphatases and PP2C were downregulated in later embryogenic stages, suggesting that the ABA signal transduction cascade plays an important role in later embryogenesis.

Specific Expression of DEGs During Each Embryogenesis Phase

Because multiple phytohormone-related genes are involved in the acquisition of embryogenic competence, we further analyzed the transcriptome data to gain insights into the function of plant hormones in zygotic embryogenesis (**Figure 6A**). In the early embryogenic stage, auxin responsive genes, auxin transport genes AUX/IAA, and gibberellin response genes were upregulated in early embryogenesis. Conversely, ethylene-responsive protein kinase Le-CTR1 and ABA-related protein PP2C phosphatase were downregulated in this stage. In addition, several PP2C was still upregulated in early to middle embryogenesis. EIN3, which could initiate downstream transcriptional cascades for ethylene responses, was upregulated in the pre-cotyledon embryo stage. Auxin transporter including AUX/IAA and PIN genes and auxin responsive genes were upregulated in the cotyledon embryo stage. Subsequent seed maturation, PP2C phosphatase, and AUX/IAA family genes were significantly increased. Taken together, these results suggested that auxin, GA, and ethylene may participate in early embryonic initiation and cotyledon embryonic formation. Auxin and ABA synergistically regulate late embryonic development and seed maturation in *P. mongolica*.

Furthermore, transcripts related to epigenetic regulation were also found in different developmental phases (**Figure 6B**). SAM-dependent carboxyl methyltransferase was found during early embryogenesis, suggesting that DNA methylation has an important role in embryonic development of *P. mongolica*. In addition, chromatin modification-related genes, such as Jumonji

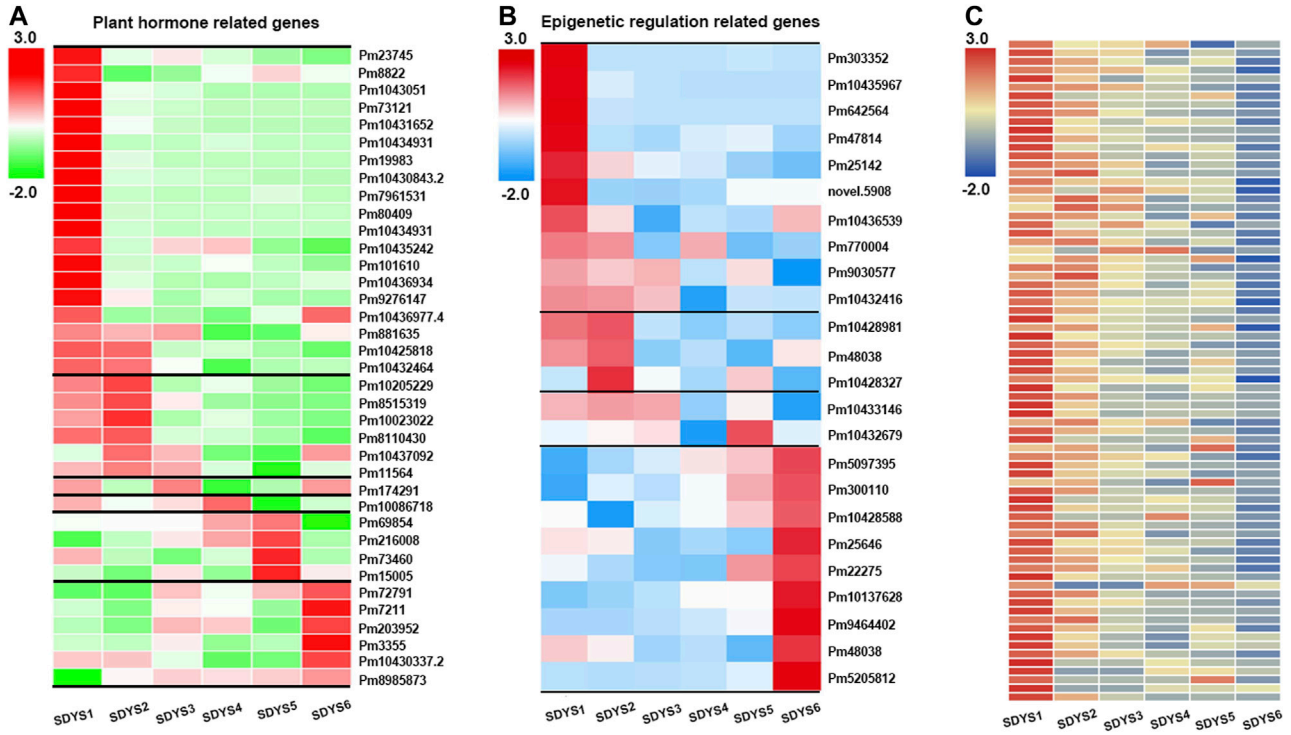


FIGURE 6 | Heat map of specific expression of differentially expressed genes identified during each embryogenesis phase. **(A)** Plant hormone-related gene, **(B)** epigenetic regulation-related gene, and **(C)** NB-ARC-related gene expression in six different developmental stages.

C-domain histone demethylase-encoding gene, showed maximum expression during early embryogenesis. The expression of three transcripts related to histone deacetylase, which is usually associated with transposon silencing, were also increased in early embryogenesis. Therefore, histone methylation and acetylation modification play a relevant role in the early embryogenic phase of *P. mongolica*.

Most changes in transcript levels occurred in the early embryonic pattering determination and mature embryos formation. DEGs between stage 6 and stage 1 were identified to further investigate transcriptional regulation in ZE development. In the megagametophytes including proembryos and auxin polar transport (AUX/IAA and PIN3), response TFs (MYB, AP2, bZIP, and B3) were highly abundant. More upregulated transcripts were highly enriched in response to stimulus, such as salt, drought, cold, and wound stress. Moreover, histone H2A, lysine-specific demethylase JMJ706 and JMJ25, and histone RNA stem-loop-binding two genes were up-expressed in early proembryos. On the contrary, histone acetylation-related genes (HAC1 and histone deacetylase 6) and chromatin formation or remodeling genes (MBD9, RRC1, and SNF2-related transcripts) were highly expressed at the mature phases. ABA response, dehydration, photosynthesis initiation, and storage/energy-related genes were associated with mature stages of embryo development. In addition, many DEGs were annotated as containing "terminal inverse repeats (TIR)," "leucine rich repeats," or "NB-ARC domains," which are molecular switches implicated in immune signal transduction mechanisms (**Figure 6C**). These genes were highly abundant at stage 1 but declined successively during later developmental stages. These results are consistent with the previous report of transcriptome in maritime pine (Gonçalves et al., 2005; Rodrigues et al., 2018).

Stage-Specific Gene Co-Expression Networks *via* WGCNA

Because all DEGs represent different functions, a WGCNA assay was used to define clusters and obtain gene modules with specific patterns of expression during the six-embryo development stages (**Figure 7**). Based on expression pattern and tissue specificity, 20 gene modules were identified. Six stage-specific modules were significantly correlated with one specific stage of embryonic development ($r > 0.8$ and p value < 0.001). Next, we analyzed the functions of these genes in each model and found that the translation initiation factor, CCAAT-box binding factor, and mRNA splicing factor were significantly associated with early embryogenic stage, suggesting that this period may be the transcriptional activation stage. Heat shock protein was significantly correlated with dominant embryonic developmental stage. Auxin-responsive protein was mainly enriched in the pre-cotyledon stage. Several late-embryogenesis abundant-like proteins were found during late embryogenesis and embryonic maturation stages. Therefore, LEA are likely key genes for embryo maturation in *P. mongolica*. In addition, the top 50 genes associated with MEdarkred were of unknown function.

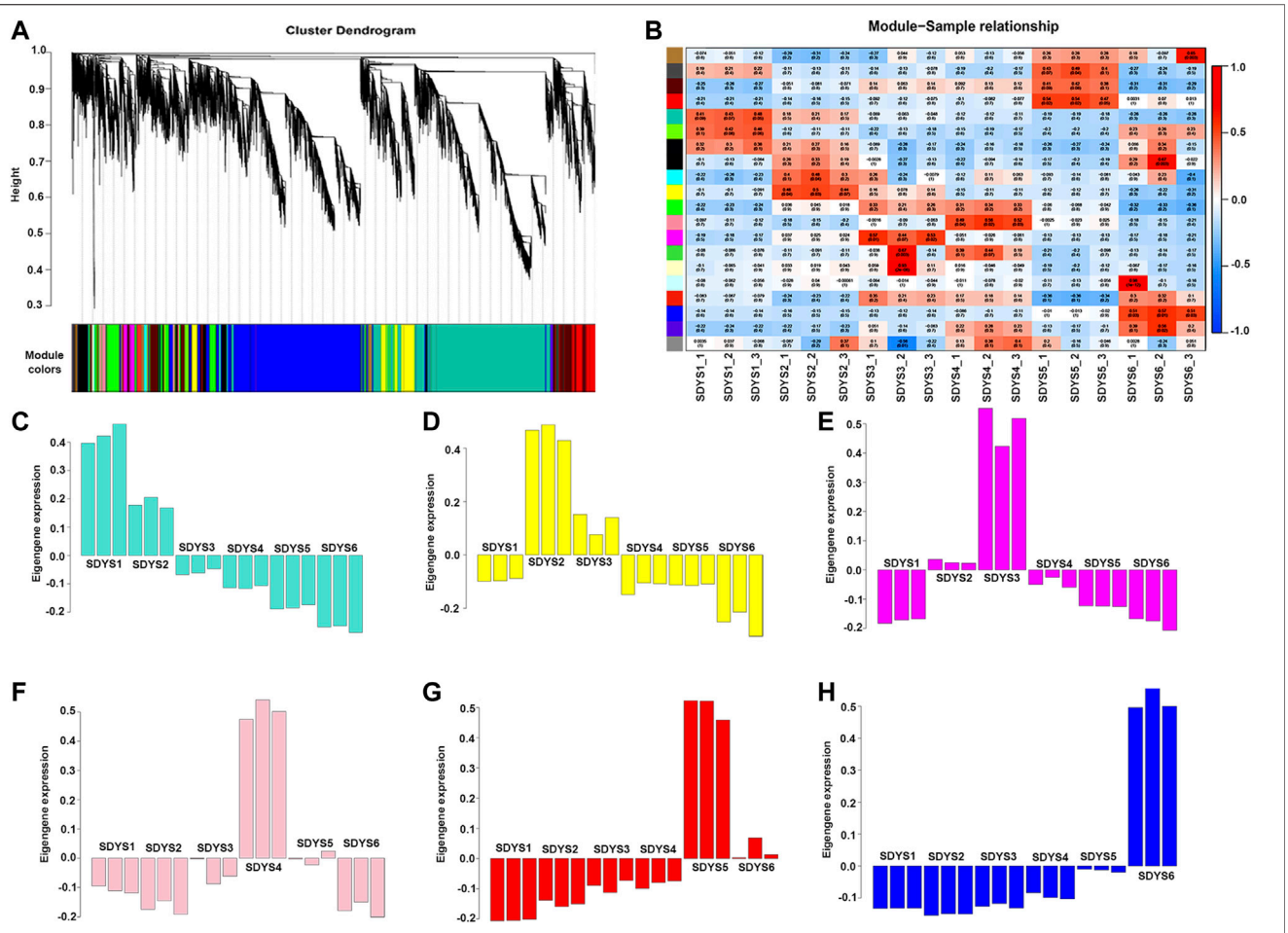


FIGURE 7 | Zygotic embryo development related module performed by WGCNA. **(A)** Cluster dendrogram of co-expressed genes. **(B)** Heat map of correlation between co-expression modules. **(C–H)** Correlation between co-expression modules and traits. A negative value represents a negative correlation, and a positive value represents a positive correlation.

This may be due to incomplete genome annotation in *P. mongolica*.

Identification of Transcripts for Embryo-Specific Markers

To further identify embryonic markers in *P. mongolica*, the MGFR tool (El Amrani et al., 2015) was used to analyze the DEGs of all samples. Fragments per kilobase of transcript per million mapped reads > 30 in at least one embryonic stage were used as the standard for DEG evaluation. We identified 184 transcripts in pro-embryos, 20 in early embryos, 7 in dominant embryos, 18 in pre-cotyledon embryo, 29 in cotyledon embryos, and 164 in mature seed embryos. These transcripts seem to be candidate genes for embryo-specific markers (**Figure 8A**). Considering that TFs are major determinants of embryogenesis, they are best for use as embryonic markers. We investigated transcripts likely TFs, whose levels were at least fourfold higher than other transcripts in a set of phase-enriched markers. There were MYB, GATA, TCP, and AP2 TFs associated with pro-embryo initiation. MYB was a key TF family in early embryo formation; HLH and AP2 TFs were highly expressed in the embryonic maturation stage (**Figure 8B**).

In addition, we also identified the known transcription factor marker gene WOX during embryonic morphogenesis (**Figure 8C**). Pm1910, Pm116716, and Pm10344604, likely putative orthologs of WOX8, were found to be downregulated in early embryogenesis. Pm10344604, a putative ortholog of WOX11, was upregulated in the dominant embryonic stage. Pm658686 was likely WOX12 and showed high expression in this stage. Pm81351 (a putative ortholog of WOX8) was strongly induced in the pre-cotyledonary stage. Pm74653 (WOX5) was upregulated in the cotyledonary stage.

Expression Profiles of Transcription Factor Genes in Different Developmental Stages

TFs play a significant role in cell fate determination during embryogenesis. Hence, we further identified TFs involved in embryonic development in *P. mongolica* based on model plant

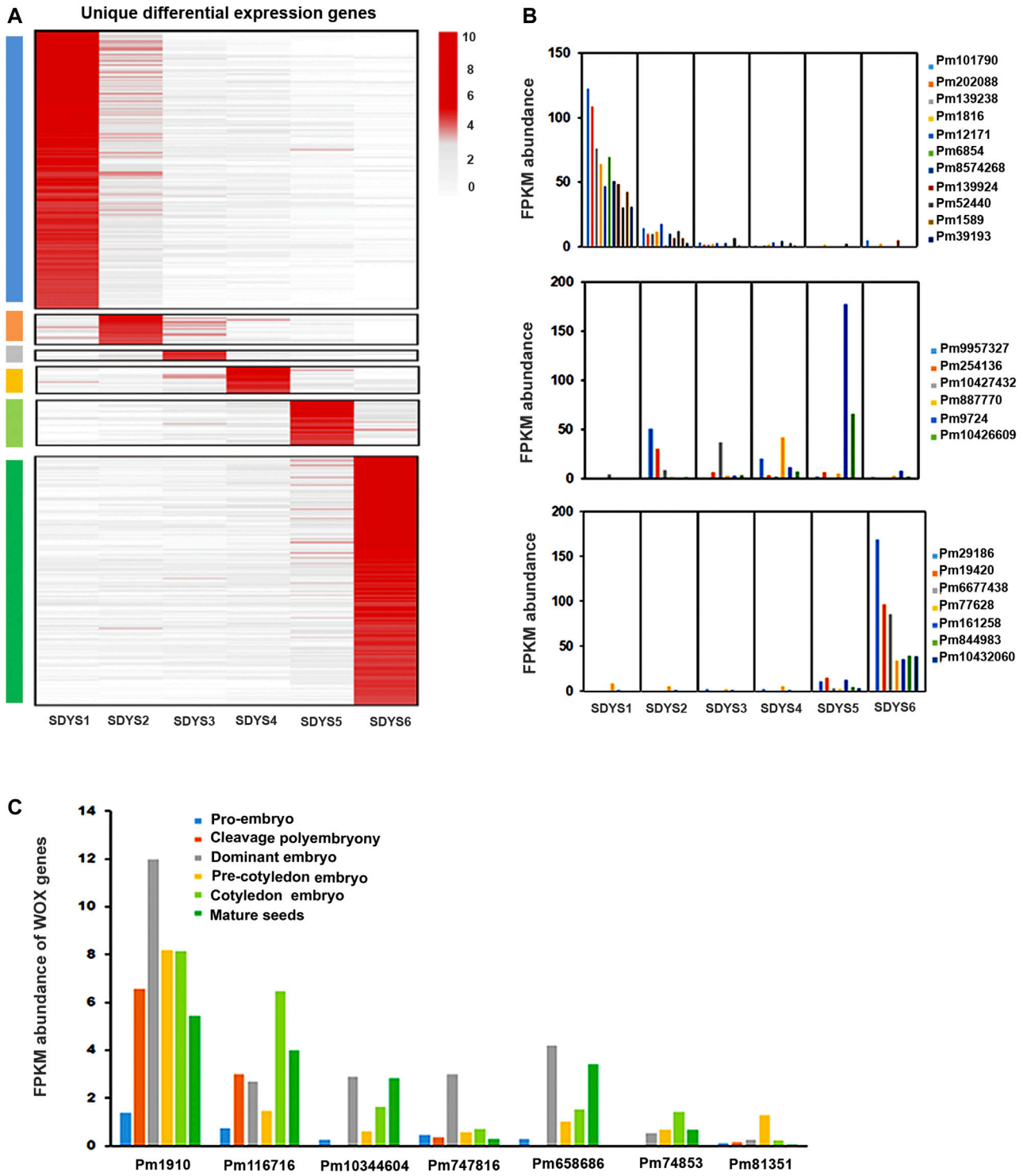


FIGURE 8 | Functions of embryonic development-specific differentially expressed genes. **(A)** Hierarchical clustering of fold changes in transcript abundance for all specific DEGs in each embryonic developmental stage. **(B,C)** Expression of TFs and WOX genes in different embryonic development stages.

systems (**Figure 9**). In all TFs, 144 were specifically upregulated and 90 were downregulated from pro-embryo to early embryo formation; 175 TFs were specifically upregulated and 71 were downregulated from early embryo to dominant embryo generation; 94 TFs (62 upregulated and 32 downregulated) in dominant embryo stages compared with pre-cotyledon

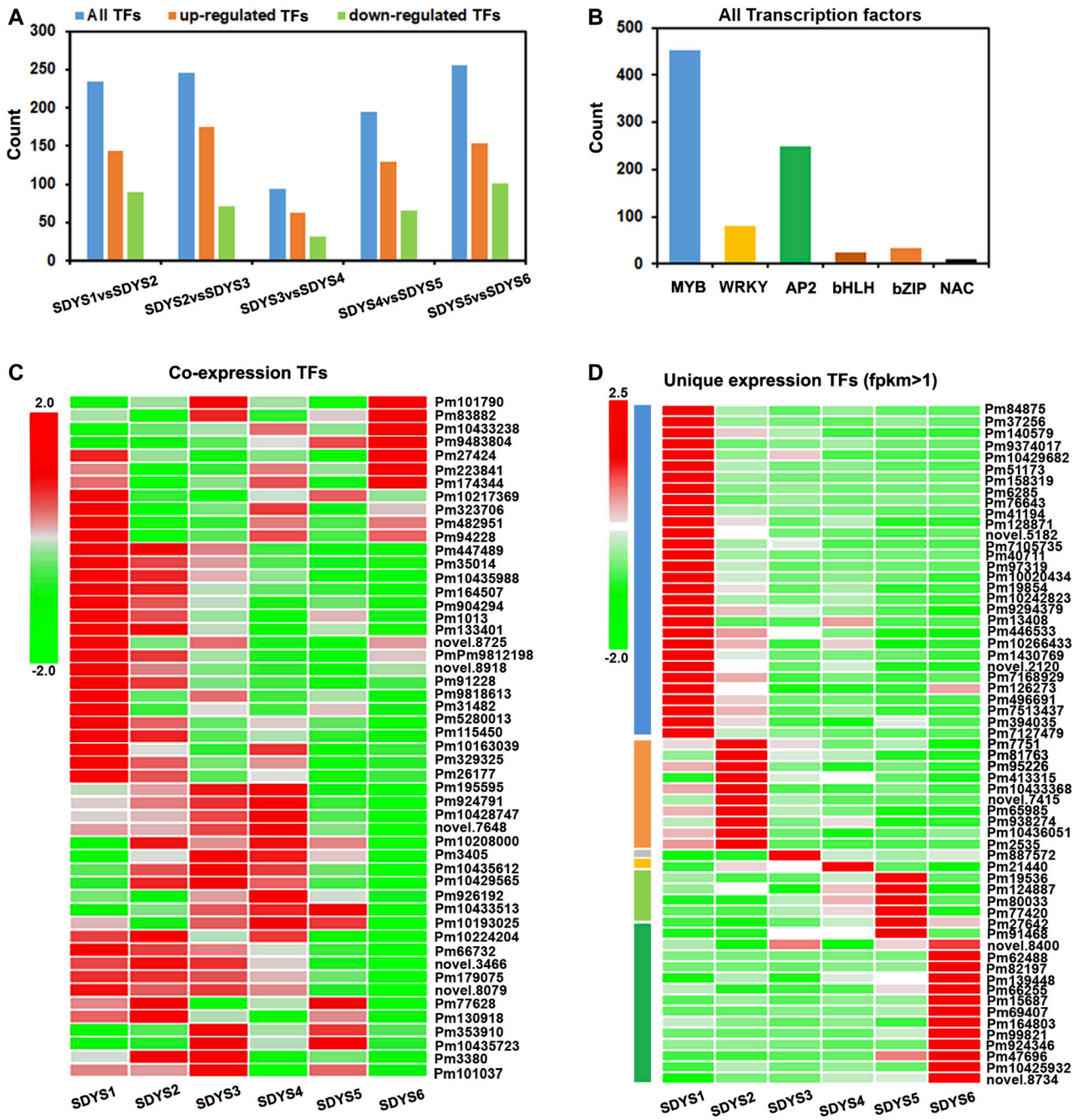


FIGURE 9 | Differentially expressed TFs related to embryo development of *P. mongolica*. **(A,B)** All TFs identified in pairwise comparisons. **(C)** Heat map of TFs from 1,279 DEGs shared in all six groups. **(D)** Heat map of TFs from each embryo development-specific DEG.

embryonic stages; and 129 TFs were upregulated and 66 downregulated in the cotyledon embryonic period. One hundred fifty-four were upregulated and 101 downregulated from the cotyledon embryonic stage to the mature seeds (**Figures 9A,C**). In addition, we further analyzed the types of transcription factors in all DEGs (**Figure 9B**). They are mainly enriched in MYB, AP2, WRKY, HLH, bZIP, and NAC TF families. Furthermore, we analyzed the TFs in differentially expressed transcripts shared by the six groups (**Figure 9D**). We found that MYB TFs were upregulated in early embryogenesis. A putative WRKY transcript was strongly upregulated in middle embryogenesis. The AP2 and MYB transcripts were crucial in the transition from early embryogenesis to the pre-cotyledonary stage. Additionally, GATA, TCP, BES1/BZR1, and WRKY TFs were upregulated in mature embryos.

DISCUSSION

Many genes are required for embryonic development. For example, 289 genes have been reported to be embryo-specific genes in *A. thaliana*. In this study, RNA-seq technology was used to provide a comprehensive overview of the transcriptome of developing embryo in *P. mongolica*. We focused on the regulation of gene expression over the entire developmental process. In here, plant embryogenesis related the plant hormone pathways, stress response, epigenetic control, protein phosphorylation, and transcription factor regulation patterns which were fully interpreted at the transcription level in *P. mongolica*.

Plant Hormone Pathways in ZE Development

Auxin plays a considerable role in cellular patterning during embryogenesis. Its biosynthesis and polar transport-related localization are crucial in plant embryogenesis (Pérez-Pastrana et al., 2019). A previous study found that auxin-induced genes, such as ARF16, significantly increase in early cotyledonary embryonic stages in *P. pinaster* (de Vega-Bartol et al., 2013a). ARF16 represses WOX5 transcription, which is required to maintain pluripotent *columella* stem cells (Radoeva et al., 2019). In our study, we found that the auxin transport carrier genes AUX and PIN3 were upregulated in early embryonic stages. Moreover, its expression was also increased from the early cotyledonary to later embryonic stages. Thus, these results suggest that auxin biosynthesis, auxin transport, and auxin response are more prominent in early embryogenesis and affect cotyledonary embryonic pattern formation. This result is consistent with a previous report that auxin influx (AUX1, LAX1, and LAX2) is needed for cellular patterning in early embryogenesis in *Arabidopsis* (Robert et al., 2015). In addition, we also found that WOX8 and several auxin response TFs were specifically expressed in early embryogenesis. These findings provide a framework for processes that the transportation and accumulation of auxin likely activate the expression of these TFs during early embryogenesis in *P. mongolica.*

In addition, abscisic acid (ABA) plays a critical role in maintaining embryogenesis by regulating tissue differentiation and determining the physiological characteristics of late embryos (Pérez et al., 2015). Moreover, protein accumulation is generally under the control of ABA in conifer (Högberg et al., 2001). In this study, we statistically analyzed the expression of PP2C, an ABA signal transducer and key factor protein phosphatase in different ZE developmental stages. Six transcripts annotated to PP2C phosphatase were downregulated in the cotyledonary embryonic stage. Moreover, genes associated with ABA response and dehydration are more highly expressed at the embryonic mature phases. These results suggest that ABA play a critical role in the embryonic maturation processes of *P. mongolica*.

Epigenetic Regulation in ZE Development

Epigenetic reprogramming, such as histone posttranslational modifications, DNA methylation, and chromatin-remodeling maintenance, has co-regulated functions during early embryogenesis in pine (de Vega-Bartol et al., 2013b; Trontin et al., 2016). HAD2, an acetylation regulation gene, is particularly increased in the early cotyledonary stage in *P. pinaster* (Rodrigues et al., 2018). We found that histone H2A histone subunits and lysine-specific demethylases show a higher abundance in early embryogenesis. Thus, histone posttranslational modifications might drive the expression of the transcriptome of early embryos in *P. mongolica*. Furthermore, histone acetylation-related genes and chromatin formation or remodeling genes were found in early embryogenesis. Similar results have been reported during maritime pine and *Norway spruce* embryogenesis (Uddenberg et al., 2011; Rodrigues et al., 2018). In addition, our results showed that several SAM-dependent carboxyl methyltransferase and other methyltransferases show different expression profiles. In general, DNA methylation is important for embryogenesis in *P. mongolica*.

Transcriptional Regulation During the Development of the ZE

A large number of TFs have been reported to regulate cell fate differentiation events, which determine the first zygotic division during embryogenesis. For example, the MYB TF family is expressed in early seed development (Avilés-Viñas et al., 2021). Moreover, MYB interacts with bHLH to regulate seed maturation. Network analysis showed that these TFs also associated with FUS3 and WOX12 in seed maturation (Tian et al., 2020). In here, we focused on TFs associated with plant development, phytohormones, and stress responses. The plant development-related MYB, AP2, and bZIP TF families were upregulated during early ZE development. The expressions of LEA, HSP, and HLH genes were increased during late ZE development. Additionally, homologues of the PLATZ and SRF family of plant-specific TFs, whose roles are poorly known, were also observed.

WOX genes (WOX2, WOX8, WOX1, and WOX3) play an important role during the ZE development. WOX2 is expressed in the apical daughter cell, and WOX8 is specifically expressed in the basal daughter cells of the zygote (Sakakibara et al., 2014; Armenta-Medina et al., 2021; Avilés-Viñas et al., 2021). The YDA-MAP signaling cascade phosphorylates and activates WRKY2 to promote WOX8 transcription during zygote development (Ueda et al., 2011; Horstman et al., 2017). We also found that WOX8 was specifically expressed in the early embryogenic stage, WOX11/12 were specifically expressed in the dominant embryonic stage, and WOX5 was upregulated in the cotyledonary stage. Additionally, a homolog of the WRKY family was also upregulated during this period. This may be consistent with regulation patterns in *Arabidopsis* that auxin response leads to activation of WRKY2 and expression WOX during embryogenesis.

Genes That Are Sufficient for Embryogenesis in *P. mongolica*

Several genes are sufficient for embryogenesis when ectopically expressed. For example, when LEC1 and LEC2 are expressed in seedlings, SE is induced (Osorio-Montalvo et al., 2020). SERK1 encodes a leucine-rich repeat (LRR) transmembrane RLK, which induces SE when ectopically expressed. In addition, several TFs, such as AGAMOUS-Like15 (AGL15) and AGAMOUS-Like18 (AGL18), may also induce SE development (Zheng et al., 2016; Wójcik et al., 2020). The APETALA2 (AP2) domain TFs BABY BOOM/PLETHORA4 (BBM/PLT4), PLETHORA2 (PLT2), and WOUND INDUCED DEDIFFERENTIATION (WIND) promote SE in seedlings (DuBuc et al., 2020). Our transcriptome profiling data represent a valuable foundation for identifying genes, which are specifically expressed in particular stages of embryogenesis. We established a set of stringently defined temporal markers for embryogenesis. Transcription factors, including MYB, AP2, GATA, and TCP, could be markers during morphogenesis in early embryonic stages. AP2 and bHLH transcription factors are markers in the later embryonic stages to seed maturation. Heat shock proteins are specifically expressed in the domain embryo stage and serval auxin response genes in the pre-cotyledon embryo stage; late-embryogenesis abundant proteins are abundant during later embryogenesis. Therefore, these genes may be sufficient for embryogenesis and SE formation when ectopically expressed in *P. mongolica*.

In summary, our results establish a set of stringently defined temporal markers for identifying the key players involved in embryogenesis of *P. mongolica*. Our datasets and approach are expected to facilitate the discovery of molecular mechanisms underlying patterns of embryonic development. Moreover, functional understanding of these shared and distinct expression patterns of signaling, transcriptional, and epigenetic factors will help to address how embryonic development shapes the divergence of seed development in *P. mongolica*. Furthermore, our data lay a foundation for the protection and utilization of *P. mongolica* resources.

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: SRA: SRR12338997, PRJNA649217.

AUTHOR CONTRIBUTIONS

JY wrote the paper and major bioinformatic analysis. HB collected all samples and performed the preprocessing work. YW and GZ conducted the transcriptomic experiments. YB conceived of and managed the project.

FUNDING

This research was supported by the National Natural Science Foundation of China (No. 316600212).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.738649/full#supplementary-material

REFERENCES

Armenta-Medina, A., Gillmor, C. S., Gao, P., Mora-Macias, J., Kochian, L. V., Xiang, D., et al. (2021). Developmental and Genomic Architecture of Plant Embryogenesis: from Model Plant to Crops. *Plant Commun.* 2 (1), 100136. doi:10.1016/j.xplc.2020.100136

Cao, X., Ma, F., Xu, T., Wang, J., Liu, S., Li, G., et al. (2016). Transcriptomic Analysis Reveals Key Early Events of Narciclasine Signaling in Arabidopsis Root apex. *Plant Cel. Rep.* 35 (11), 2381–2401. doi:10.1007/s00299-016-2042-7

de Vega-Bartol, J. J., Santos, R. R., Simões, M., and Miguel, C. M. (2013b). Normalizing Gene Expression by Quantitative PCR during Somatic Embryogenesis in Two Representative conifer Species: *Pinus pinaster* and *Picea Abies*. *Plant Cel. Rep.* 32, 715–729. doi:10.1007/s00299-013-1407-4

de Vega-Bartol, J. J., Simões, M., Lorenz, W. W., Rodrigues, A. S., Alba, R., Dean, J. F. D., et al. (2013a). Transcriptomic Analysis Highlights Epigenetic and Transcriptional Regulation during Zygotic Embryo Development of *Pinus pinaster*. *BMC Plant Biol.* 13, 123. doi:10.1186/1471-2229-13-123

Dresselhaus, T., and Jürgens, G. (2021). Comparative Embryogenesis in Angiosperms: Activation and Patterning of Embryonic Cell Lineages. *Annu. Rev. Plant Biol.* 72 (72), 641–676. doi:10.1146/annurev-arplant-082520-094112

DuBuc, T. Q., Schnitzler, C. E., Chrysostomou, E., McMahon, E. T., Febrimarsa, Gahan, J. M., et al. (2020). Transcription Factor AP2 Controls Cnidarian Germ Cell Induction. *Science* 367 (6479), 757–762. doi:10.1126/science.aay6782

El Amrani, K., Stachelscheid, H., Lekschas, F., Kurtz, A., and Andrade-Navarro, M. A. (2015). MGFM: a Novel Tool for Detection of Tissue and Cell Specific Marker Genes from Microarray Gene Expression Data. *BMC Genomics* 16 (1), 645. doi:10.1186/s12864-015-1785-9

Gonçalves, S., Cairney, J., Maroco, J., Oliveira, M. M., and Miguel, C. (2005). Evaluation of Control Transcripts in Real-Time RT-PCR Expression Analysis during Maritime pine Embryogenesis. *Planta* 222, 556–563. doi:10.1007/s00425-005-1562-0

Hao, Z., Zhang, Z., Xiang, D., Venglat, P., Chen, J., Gao, P., et al. (2021). Conserved, Divergent and Heterochronic Gene Expression during Brachypodium and Arabidopsis Embryo Development. *Plant Reprod.* 34, 207–224. doi:10.1007/s00497-021-00413-4

Hofmann, F., Schon, M. A., and Nodine, M. D. (2019). The Embryonic Transcriptome of *Arabidopsis thaliana*. *Plant Reprod.* 32 (1), 77–91. doi:10.1007/s00497-018-00357-2

Högberg, K.-A., Bozhkov, P. V., Grönroos, R., and Arnold, S. V. (2001). Critical Factors Affecting *Ex Vitro* Performance of Somatic Embryo Plants of *Picea Abies*. *Scand. J. For. Res.* 16, 295–304. doi:10.1080/02827580116772

Horstman, A., Li, M., Heidmann, I., Weemen, M., Chen, B., Muino, J. M., et al. (2017). The BABY BOOM Transcription Factor Activates the LEC1-ABI3-FUS3-LEC2 Network to Induce Somatic Embryogenesis. *Plant Physiol.* 175 (2), 848–857. doi:10.1104/pp.17.00232

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9 (9), 559. doi:10.1186/1471-2105-9-559

Liu, W., Wang, C., Shen, X., Liang, H., Wang, Y., He, Z., et al. (2019). Comparative Transcriptome Analysis Highlights the Hormone Effects on Somatic Embryogenesis in Catalpa Bungei. *Plant Reprod.* 32 (2), 141–151. doi:10.1007/s00497-018-0349-y

Llebrés, M.-T., Pascual, M.-B., Debille, S., Trontin, J.-F., Harvengt, L., Avila, C., et al. (2018). The Role of Arginine Metabolic Pathway during Embryogenesis and Germination in Maritime pine (*Pinus pinaster* Ait.). *Tree Physiol.* 38 (3), 471–484. doi:10.1093/treephys/tpx133

Lorenz, W. W., Ayyampalayam, S., Howe, J. M. G. T., Jermstad, K. D., Neale, D. B., Rogers, D. L., et al. (2012). Conifer DBMagic: a Database Housing Multiple De Novo Transcriptome Assemblies for 12 Diverse conifer Species. *Tree Genet. Genomes* 8, 1477–1485. doi:10.1007/s11295-012-0547-y

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

Merino, I., Abrahamsson, M., Sterck, L., Craven-Bartle, B., Canovas, F., and von Arnold, S. (2016). Transcript Profiling for Early Stages during Embryo Development in Scots pine. *BMC Plant Biol.* 16 (1), 255. doi:10.1186/s12870-016-0939-5

Osorio-Montalvo, P., De-la-Peña, C., Oropeza, C., Nic-Can, G., Córdova-Lara, I., Castillo-Castro, E., et al. (2020). A Peak in Global DNA Methylation Is a Key Step to Initiate the Somatic Embryogenesis of Coconut palm (Cocos Nucifera L). *Plant Cel. Rep.* 39 (10), 1345–1357. doi:10.1007/s00299-020-02568-2

Palovaara, J., Hallberg, H., Stasolla, C., and Hakman, I. (2010). Comparative Expression Pattern Analysis of WUSCHEL-Related Homeobox 2 (WOX2) and WOX8/9 in Developing Seeds and Somatic Embryos of the Gymnosperm Picea Abies. *New Phytol.* 188 (1), 122–135. doi:10.1111/j.1469-8137.2010.03336.x

Pérez, M., Viejo, M., LaCuesta, M., Toorop, P., and Cañal, M. J. (2015). Epigenetic and Hormonal Profile during Maturation of Quercus Suber L. Somatic Embryos. *J. Plant Physiol.* 173 (173), 51–61. doi:10.1016/j.jplph.2014.07.028

Pérez-Pastrana, J., Testillano, P. S., Barany, I., Canto-Flick, A., Álvarez-López, D., Pijeira-Fernández, G., et al. (2021). Endogenous Auxin Accumulation/localization during Zygotic and Somatic Embryogenesis of Capsicum Chinense Jacq. *J. Plant Physiol.* 258-259, 153333. doi:10.1016/j.jplph.2020.153333

Radoeva, T., Lokerse, A. S., Llavata-Peris, C. I., Wendrich, J. R., Xiang, D., Liao, C.-Y., et al. (2019). A Robust Auxin Response Network Controls Embryo and Suspensor Development through a Basic Helix Loop Helix Transcriptional Module. *Plant Cell* 31 (1), 52–67. doi:10.1105/tpc.18.00518

Robert, H. S., Grunewald, W., Sauer, M., Cannoot, B., Soriano, M., Swarup, R., et al. (2015). Plant Embryogenesis Requires Aux/lax-Mediated Auxin Influx. *Development* 142, 702–711. doi:10.1242/dev.115832

Rodrigues, A. S., Chaves, I., Costa, B. V., Lin, Y.-C., Lopes, S., Milhinhos, A., et al. (2019). Small RNA Profiling in *Pinus pinaster* Reveals the Transcriptome of Developing Seeds and Highlights Differences between Zygotic and Somatic Embryos. *Sci. Rep.* 9 (1), 11327. doi:10.1038/s41598-019-47789-y

Rodrigues, A. S., De Vega, J. J., and Miguel, C. M. (2018). Comprehensive Assembly and Analysis of the Transcriptome of Maritime pine Developing Embryos. *BMC Plant Biol.* 18 (1), 379. doi:10.1186/s12870-018-1564-2

Sakakibara, K., Reisewitz, P., Aoyama, T., Friedrich, T., Ando, S., Sato, Y., et al. (2014). WOX13-like Genes Are Required for Reprogramming of Leaf and Protoplast Cells into Stem Cells in the moss Physcomitrella Patens. *Development* 141 (8), 1660–1670. doi:10.1242/dev.097444

Shamir, R., Maron-Katz, A., Tanay, A., Linhart, C., Steinfeld, I., Sharan, R., et al. (2005). EXPANDER - an Integrative Program Suite for Microarray Data Analysis. *BMC Bioinformatics* 6, 232. doi:10.1186/1471-2105-6-232

Singh, H. (1978). *Embryology of Gymnosperms (Encyclopedia of Plant Anatomy)*. Berlin, Stuttgart, Germany: Gebrüder Borntraeger.

Tian, R., Paul, P., Joshi, S., and Perry, S. E. (2020). Genetic Activity during Early Plant Embryogenesis. *Biochem. J.* 477, 3743–3767. doi:10.1042/BCJ20190161

Trontin, J.-F., Klimaszewska, K., Morel, A., Hargreaves, C., and Lelu-Walter, M.-A. (2016). Molecular Aspects of Conifer Zygotic and Somatic Embryo Development: A Review of Genome-wide Approaches and Recent Insights. *Methods Mol. Biol.* 1359, 167–207. doi:10.1007/978-1-4939-3061-6_8

Uddenberg, D., Valladares, S., Abrahamsson, M., Sundström, J. F., Sundås-Larsson, A., and von Arnold, S. (2011). Transcriptomic Time-Series Analyses of Gene Expression Profile During Zygotic Embryo Development in *Picea mongolica*. *Planta* 234, 527–539. doi:10.1007/s00425-011-1418-8

Ueda, M., Zhang, Z., and Laux, T. (2011). Transcriptional Activation of Arabidopsis axis Patterning Genes WOX8/9 Links Zygote Polarity to Embryo Development. *Dev. Cel.* 20 (2), 264–270. doi:10.1016/j.devcel.2011.01.009

Ulitsky, I., Laurent, L. C., and Shamir, R. (2010). Towards Computational Prediction of microRNA Function and Activity. *Nucleic Acids Res.* 38 (15), e160, 2010 . Epub 2010 Jun 24. doi:10.1093/nar/gkq570

Vetrici, M. A., Yevtushenko, D. P., and Misra, S. (2021). Douglas-fir LEAFY COTYLEDON1 (PmLEC1) Is an Active Transcription Factor during Zygotic and Somatic Embryogenesis. *Plant Direct* 5 (7), e00333. doi:10.1002/pld3.333

von Arnold, S., Sabala, I., Bozhkov, P., Dyachok, J., and Filonova, L. (2002). Developmental Pathways of Somatic Embryogenesis. *Plant Cell, Tissue Organ. Cult.* 69, 233–249. doi:10.1023/A:1015673200621

Vuosku, J., Muilu-Mäkelä, R., Avia, K., Suokas, M., Kestilä, J., Läärä, E., et al. (2019). Thermospermine Synthase (ACL5) and Diamine Oxidase (DAO) Expression Is Needed for Zygotic Embryogenesis and Vascular Development in Scots Pine. *Front. Plant Sci.* 10, 1600. doi:10.3389/fpls.2019.01600.eCollection.2019

Vuosku, J., Sutela, S., Kestilä, J., Jokela, A., Sarjala, T., and Häggman, H. (2015). Expression of Catalase and Retinoblastoma-Related Protein Genes Associates with Cell Death Processes in Scots pine Zygotic Embryogenesis. *BMC Plant Biol.* 15, 88. doi:10.1186/s12870-015-0462-0

Wójcik, A. M., Wójcikowska, B., and Gaj, M. D. (2020). Current Perspectives on the Auxin-Mediated Genetic Network that Controls the Induction of Somatic Embryogenesis in Plants. *Ijms* 21 (4), 1333. doi:10.3390/ijms21041333

Yan, J., Peng, P., Duan, G., Lin, T., and Bai, Y. e. (2021). Multiple Analyses of Various Factors Affecting the Plantlet Regeneration of Picea Mongolica (H. Q. Wu) W.D. Xu from Somatic Embryos. *Sci. Rep.* 11 (1), 6694. doi:10.1038/s41598-021-83948-w

Zheng, Q., Zheng, Y., Ji, H., Burnie, W., and Perry, S. E. (2016). Gene Regulation by the AGL15 Transcription Factor Reveals Hormone Interactions in Somatic Embryogenesis. *Plant Physiol.* 172 (4), 2374–2387. doi:10.1104/pp.16.00564

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.

Copyright © 2021 Yan, buer, Wang, zhula and Bai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fgene.2021.714575

Genome-Wide Association Mapping of Late Blight Tolerance Trait in Potato (*Solanum tuberosum* L.)

Fang Wang[1,2,3,4†], *Meiling Zou*[5,6†], *Long Zhao*[1,2,3,4], *Zhiqiang Xia*[5,6*] and *Jian Wang*[1,2,3,4*]

[1] Academy of Agriculture and Forestry Sciences, Qinghai University, Xining, China, [2] National Key Laboratory of Sanjiangyuan Ecology and Plateau Agriculture and Animal Husbandry, Qinghai University, Xining, China, [3] Key Laboratory of Qinghai-Tibet Plateau Biotechnology Ministry of Education, Xining, China, [4] Qinghai Provincial Key Laboratory of Potato Breeding, Xining, China, [5] College of Tropical Crops, Hainan University, Haikou, China, [6] Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agriculture Sciences, Haikou, China

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Lihui Li,
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), China
Zewei Luo,
University of Birmingham, United Kingdom

***Correspondence:**
Zhiqiang Xia
zqxia@hainanu.edu.cn
Jian Wang
jianwang2197@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: 25 May 2021
Accepted: 23 August 2021
Published: 01 October 2021

Citation:
Wang F, Zou M, Zhao L, Xia Z and Wang J (2021) Genome-Wide Association Mapping of Late Blight Tolerance Trait in Potato (Solanum tuberosum L.). Front. Genet. 12:714575. doi: 10.3389/fgene.2021.714575

Uncovering the genetic basis and optimizing the late blight tolerance trait in potatoes (*Solanum tuberosum* L.) are crucial for potato breeding. Late blight disease is one of the most significant diseases hindering potato production. The traits of late blight tolerance were evaluated for 284 potato cultivars to identify loci significantly associated with the late blight tolerance trait. Of all, 37 and 15 were the most tolerant to disease, and 107 and 30 were the most susceptible. A total of 22,489 high-quality single-nucleotide polymorphisms and indels were identified in 284 potato cultivars. All the potato cultivars were clustered into eight subgroups using population structure analysis and principal component analysis, which were consistent with the results of the phylogenetic tree analysis. The average genetic diversity for all 284 potato cultivars was 0.216, and the differentiation index of each subgroup was 0.025–0.149. Genome-wide linkage disequilibrium (LD) analysis demonstrated that the average LD was about 0.9 kb. A genome-wide association study using a mixed linear model identified 964 loci significantly associated with the late blight tolerance trait. Fourteen candidate genes for late blight tolerance traits were identified, including genes encoding late blight tolerance protein, chitinase 1, cytosolic nucleotide-binding site–leucine-rich repeat tolerance protein, protein kinase, ethylene-responsive transcription factor, and other potential plant tolerance-related proteins. This study provides novel insights into the genetic architecture of late blight tolerance traits and will be helpful for late blight tolerance in potato breeding.

Keywords: *Solanum tuberosum* L., single-nucleotide polymorphism (SNP), genome-wide association analysis (GWAS), population structure, genetic diversity, late blight tolerance

INTRODUCTION

Potato (*Solanum tuberosum* L.) is a major food crop produced worldwide (Paul et al., 2012). Potato productivity is largely limited by late blight disease, a typical disease that often occurs at low temperatures and high humidity, caused by *Phytophthora infestans* (Fry et al., 2002). *Phytophthora infestans* propagate asexually or sexually and invade the potato cells by hyphae,

rapidly causing dead leaves and rotten tubers (McDonald and Linde, 2002; Zwankhuizen and Zadoks, 2002; Enciso-Rodriguez et al., 2008; Wu et al., 2016). In traditional agriculture, late blight is primarily controlled using protective agents or systemic fungicides. However, pesticides pollute the environment and accelerate the development of tolerance and the mutation of *Phytophthora infestans* (Maurice et al., 2009).

Risch and Merikangas (1996) and Hansen et al. (2001) first reported genome-wide association analysis (GWAS). With the development of high-throughput genotyping, particularly the genotyping sequencing (GBS) technique and amplified-fragment single-nucleotide polymorphism (SNP) and methylation (AFSM) technology, GWAS has become a powerful approach for genetic dissection of complex quantitative traits by locating quantitative trait loci affecting phenotype, using sufficient markers and linkage disequilibrium (LD) between alleles (Nordborg and Weigel, 2008; Zhu et al., 2008; Korte and Farlow, 2013; Pradhan et al., 2016; Zhang et al., 2018). Nowadays, GWAS is an efficient and reliable tool for deciphering the molecular basis of complex quantitative traits, such as pathogen tolerance in *Arabidopsis* (Aranzana et al., 2005), preharvest sprouting resistance in wheat (Zhou et al., 2017), and head smut tolerance in maize (Wang et al., 2012). In this study, we conducted GWAS for late blight tolerance traits using AFSM on 284 potato cultivars from China, Australia, Belarus, Canada, Britain, the International Potato Center (CIP), Israel, Netherlands, Russia, and the United States. Our objective was to dissect their genetic architecture, evaluate genetic diversity, identify loci, and molecular markers associated with the late blight tolerance trait, and identify candidate genes for potato-breeding improvement. Two randomly selected candidate genes were verified by qRT-PCR.

MATERIALS AND METHODS

Sample Collection

A total of 284 potato accessions, collected from the Qinghai Plateau Potato Experimental Station, were used for GWAS to dissect the genetic basis of late blight tolerance. Among them, 97 accessions originated from China, 1 from Australia, 3 from Belarus, 5 from Canada, 5 from Britain, 138 from CIP, 2 from Israel, 1 from the Netherlands, 5 from Russia, 1 from the United States, and 26 from unknown regions (**Supplementary Table 1**).

All 284 potato accessions used in this study were tetraploid and were grown in the Qinghai Plateau Potato Experimental Station from the Academy of Agriculture and Forestry Sciences, Qinghai University (36°68′N, 101°26′E) in 2018 and 2019 with 10 replicates of each accession in one row.

Phenotype Evaluation and Statistics

Preparation of Pathogen

The *Phytophthora infestans* used in this study were collected from different potato planting areas in Qinghai Province with physiologically dominant species, mainly 3, 4, and 10. The susceptible potatoes were cut into slices, approximately 0.5-cm thick. These slices were inoculated with the pathogen, placed in culture dishes in an 18°C incubator, incubated in the dark, and covered with moist filter papers. After 5 days, the collected sporangia on these slices were filtered using nylon and steel mesh. The concentration of *Phytophthora infestans* was adjusted to 50,000 per milliliter.

Indoor Inoculation and Identification

Ten leaves for each accession were picked for *in vitro* inoculation with *Phytophthora infestans* after the seedlings had 10 leaves. Each picked leaf of the 284 potato accessions was inoculated with 15 μL of *Phytophthora infestans* and placed in an 18°C humidified incubator with 8 h darkness and 16 h light every day. Then, the indoor investigation for late blight tolerance in potatoes was evaluated (**Table 1**) after inoculation for 5 days in 2018 and 2019. Each trait was scored on 10 leaves in each accession, and the mean of two replicates was used for subsequent statistical analysis and GWAS.

DNA Preparation and Sequencing

The improved cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson, 1980) was used to extract genomic DNA from potato leaves. After detecting and quantifying the concentration by 1% agarose gel electrophoresis, the working DNA solution was diluted to 100 ng/μL and stored at −20°C. The AFSM approach (Xia et al., 2014) was then used to construct EcoRI-MspI and EcoRI-HpaII libraries of 284 potato DNA samples. After the monoclonal detection met the requirements, the EcoRI-MspI and EcoRI-HpaII libraries were mixed into one library at a ratio of 1:1, and HiSeq 2500 was used to perform paired-end 150-bp sequencing on the constructed sequencing library. Out of 355 Gb of total sequencing data, 328 Gb of clean data were obtained with more than 1 Gb of data per sample.

Single-Nucleotide Polymorphism Calling and Annotation

We used a Perl script[1] to filter the original sequencing data, count the total number of reads obtained from sequencing, assign the reads to each individual based on the barcodes designed using the AFSM technology, and count the number of reads in each individual. Bowtie2 software (Langmead and Salzberg, 2012) aligned the optimized sequencing reads to the potato DM reference genome,[2] and SAMtools (Li et al., 2009) and VCFtools[3] were used to detect SNP and indel loci. Based on the potato DM reference genome v.4.03, the snpEff software (Cingolani et al., 2012) was used to identify the mutation locations (intergenic region, untranslated region/UTR, upstream gene region, or downstream gene region), mutation types (synonymous, missense, frameshift, and non-frameshift), and annotate them simultaneously.

[1]http://afsmseq.sourceforge.net/

[2]http://solanaceae.plantbiology.msu.edu/

[3]http://vcftools.sourceforge.net/

TABLE 1 | Identification criteria for late blight tolerance of potato leaves *in vitro*.

Grade of disease	Type	Degree of morbidity
1	Most tolerance	The lesion area is less than 3% or asymptomatic
2	More tolerance	The area of the lesion was 3–10%, forming an anaphylactic necrotic spot without chlorotic halo
3	Moderately tolerance	The area of the lesion was 10–30%. The lesion was water-soaked and surrounded by chlorotic halo. White mycelium could be seen on the surface of the lesion
4	Moderately susceptible	The lesion area ranged from 30 to 60%, there was an obvious chlorotic ring and white mold layer around the lesion
5	More susceptible	The lesion area was more than 60%, and a large number of white molds appeared on the lesion surface

Analysis of Population Structure and Genetic Diversity

We first used PHYLIP[4] to calculate the genetic distance matrix of the sample. We then used the Notepad++ software to save the genetic distance matrix file in a suitable format. A phylogenetic tree was constructed using the neighbor-joining method. After generating the tree file, iTOL[5] was used to draw the phylogenetic tree diagram. GCTA software was used to conduct principal component analysis (PCA) of the potato population materials with the detected SNPs as inputs (Yang et al., 2011). R software was then used to calculate the vector of each principal component and draw the PCA scatter plot. Additionally, ADMIXTURE software (Alexander et al., 2009) analyzed the population structure and estimated the optimal number of population subgroups. PLINK software (Purcell et al., 2007) was used to adjust the input file format for ADMIXTURE software, and then we input the file. The subgroups' *K*-value range was set to 1–12. The appropriate value of K for the number of subgroups was determined according to the obtained cross-validation error value. The genetic composition coefficient (Q) of each material in each subgroup was used to construct the population-genetic structure matrix. VCFtools software[6] was used to calculate the genetic diversity (π) and population pairwise *F*-statistics (F_{ST}) (Danecek et al., 2011). According to Wright, when F_{ST} is equal to zero or one, it indicates no differentiation or complete differentiation between subgroups, respectively. If $0 < F_{ST} < 0.05$, $0.05 \leq F_{ST} < 0.15$, $0.15 \leq F_{ST} < 0.25$, or $0.25 \leq F_{ST} < 1$, this indicates that the subgroups have weak, medium, strong, or very strong genetic differentiation, respectively (Dunia et al., 2011). In the entire group and each subgroup (determined by the population structure), the r^2-value was used to determine the genome-wide LD through pairwise comparisons between 22,489 SNP markers.

Linkage Disequilibrium Analysis

In the entire population and each subgroup (inferred using ADMIXTURE), the value of R^2 was used to evaluate the LD relationship between each pair of polymorphic sites throughout the genome, and the value of R^2 was calculated using PopLDdecay software (Zhang et al., 2019) for high-quality SNPs after filtering. The genetic distance was sorted from small to large, and then the average value of LD R^2 in the segment was calculated to draw a scatterplot with a smooth curve. The genetic distance interval in which the curve intersects with the straight line representing a non-collinear R^2 of 99% is the LD-decay distance.

Association Analysis

In this study, data from 22,489 high-quality SNPs and indels were typed to perform GWAS on this population's severity (lesion diameter) and tolerance grade. We used a compressed mixed linear model of TASSEL 5.0 software (Bradbury et al., 2003) for correlation analysis.

The threshold for the significance of tolerance grade was set at 0.05. Inputting the physical location of the SNP in the potato genome and its *P*-value, the qqman package of R software was used to draw the Manhattan plot and QQ plot. SAMtools was used to manually verify regions significantly correlated with the reordered read results of the potato reference genome PGSC_DM_v4.03.[7]

Candidate Gene Screening

Based on potato SNP annotation and LD decay and according to the functional annotation of the loci, the genes on which the loci were located were used as candidate genes. If a locus was located upstream and downstream of other genes simultaneously, the upstream and downstream genes were also used as candidate genes. If a locus was located in the intergenic region, the upstream and downstream genes closest to the locus were used as candidate genes. A mixed linear model was used to perform association analysis on the traits of late blight.

Real-Time Fluorescence Quantitative PCR Verification

Primers were designed based on the coding region sequence (CDS) of the candidate genes, and the actin gene was the internal reference (**Table 2**). A fluorescence reverse transcription kit (TaKaRa, Beijing) was used to generate cDNA using 500-ng RNA as the template. A fluorescence quantitative RT-PCR kit (TaKaRa, Beijing) was used to perform quantitative real-time polymerase chain reaction (qRT-PCR) with the thermal cycling program of

[4]http://evolution.genetics.washington.edu/phylip.html

[5]https://itol.embl.de/

[6]https://vcftools.github.io/index.html

[7]http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml/

95°C for 30 s followed by 40 cycles of 95°C for 5 s, and 60°C for 30 s. Excel 2016 was used to sort and analyze the gene expression fluorescence qRT-PCR (quantitative real-time polymerase chain reaction) data. The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression levels.

RESULTS

Genotype Analysis of Potato Population

We obtained a total of 4,786,675 SNPs and indels in this study. All SNPs and indels were filtered for minor allele frequency (MAF) > 0.05 and Hardy–Weinberg equilibrium *P*-value > 0.001, and 20,382 high-quality SNPs and 2,107 indels were obtained. Annotation of the high-quality SNPs and indels showed that 18,683 (83.08%) were in intergenic regions; 3,806 (16.92%) were in the gene regions of the genome of which 951 were in the untranscribed areas, 2,796 were in introns, and only 1,682 SNPs were in the coding areas. In the coding regions, 771 SNPs produced silent mutations, and 911 SNPs produced missense mutations at a ratio of 1.18 (**Table 3**).

Identification and Analysis of the Late Blight Tolerance of Potato Leaves *in vitro*

Among the 284 materials tested, 37 resources with disease area less than 3% or asymptomatic and disease grade 1 were selected. The area of the disease spot was 3–10%, forming an anaphylactic dead spot; there was no chlorotic halo around it; and there were 15 resources with disease grade 2. There were 30 resources with disease grade 3 with white mycelium visible on the disease spot surface, and the area of the disease spot was 10–30%. The disease spot was water-soaked and surrounded by a chlorotic halo. There were 107 resources with disease grade 4, and the area of the disease spot was 30–60%. There were obvious chlorotic circles and white mold layers around the disease spot. There were 95 resources with disease grade 5; the lesion area was more than 60%, and a large amount of white mildew appeared on the lesion's surface (**Supplementary Table 2**).

Analysis of Population Structure

The ADMIXTURE software was used to analyze 22,489 high-quality SNPs and indels; the largest cluster subgroup value (K) was assumed to be each integer from 1 to 12, and the cross-validation (CV) error of each *K*-value was calculated (**Figure 1A**). When K was 1–4, the CV error gradually increased. When K was greater than 4, the CV error dropped rapidly to a nadir at $K = 8$, and for $K > 8$, it gradually increased. Therefore, $K = 8$ was optimal; that is, the entire potato population was divided into 8 subgroups.

PCA was conducted using all high-quality SNPs and indels. The calculation and analysis process was conducted by R software (**Figure 1B**). After the analysis was completed, plots were generated by R. For plotting, the eight subgroups inferred by the ADMIXTURE software were used for grouping. The results showed that the eight subgroups could be distinguished on the PC1 axis, and the clustering results were consistent with the population structure division.

According to the *Q*-value of each material in these eight subgroups, each material was classified into the subgroup with the largest *Q*-value (**Figure 1C**). Subgroups 1–8 had 11, 33, 25, 16, 12, 30, 51, and 86 germplasm resources, respectively. The distribution of the eight subgroups showed differences on the PC1 axis, and the clustering results were consistent with the population structure division. The eight subgroups of potatoes could not all be clustered together on the phylogenetic tree.

The neighbor-joining method was used to construct a phylogenetic tree, and the tree diagram was drawn with iTOL software to explore the genetic relationships between the 284 potato germplasms. Overall, the clustering results were consistent with the division of the population structure: subgroups 1, 2, and 6 clustered together well, and samples of other subgroups could be clustered together, and there was a certain crossover between samples (**Figure 2A**). The results showed no significant relationship between the genetic relationship of potato germplasm and geographical origin (**Figure 2B**). Potatoes are native to the Andes of South America, and the history of artificial cultivation can be traced back to southern Peru from 8,000 to 5,000 BC.

Genetic Diversity Revealed by Single-Nucleotide Polymorphism Markers

According to the 22,489 high-quality SNPs and indel data, the genetic diversity (π) of all 284 potato germplasm resources was 0.216, and the genetic diversity index of the eight subgroups was between 0.164 and 0.250. Among them, subgroup 8 had the lowest genetic diversity index (0.164), and 6 had the highest (0.250) (**Table 4**). These data show that there is rich genetic diversity in the 284 potato germplasm resources.

The pairwise population *F*-statistics (F_{ST}), a measure of population differentiation, was used to evaluate the degree of difference between subgroups of the 284 potato germplasm resources (**Table 5**). It was found that the F_{ST} among the subgroups was between 0.025 and 0.149, and subgroups 1 and 8 had the highest F_{ST} (0.149) and subgroups 3 and 7 had the lowest F_{ST} (0.025). Subgroups 2 and 3, subgroups 2 and 7, subgroups 3 and 7, subgroups 3 and 8, and subgroups 7 and 8 were relatively weakly differentiated, and their genetic relationships were relatively close. In contrast, there was a moderate degree of differentiation between other subgroups.

Genome-Wide Association Analysis of Late Blight Tolerance of Potato Germplasms

The mixed linear model analysis performed correlation analysis on the disease tolerance grade (**Figure 3**). A total of 964 loci associated with late blight tolerance traits were identified (**Supplementary Table 3**). The results of the QQ map showed that the SNP sites associated with significant correlation were reliable (**Figure 4**). $P < 0.05$ was set as the threshold to determine the significant loci for the disease tolerance grade, and 14 candidate genes were obtained after annotating genes located at or near the significant loci.

TABLE 2 | qRT-PCR primers for the selected genes.

Gene ID	Forward primer (5′–3′)	Reverse primer (5′–3′)
Action	AGATGCTTACGCTGGATGGAATGC	TTCCGGTGTGGTTGGATTCTGTTC
PGSC0003DMG400000043	AGCAGCTCAAGCACAGAACTCTTC	CACGCTCCACCTCCAATTCATCTC
PGSC0003DMG400028682	TGGGCATCAACATTAACCCCAAA	CCAACCAGACAGGCTAGCCA

TABLE 3 | Summary of single-nucleotide polymorphisms and indels.

Totals SNPs and indels	Intergenic	Untranslated region		Intron	Coding sequence			Non-syn/synratio
		3′ UTR	5′ UTR		Total	Missense	Synonymous	
22,489	18,683	538	413	2,796	1,682	771	911	1.18

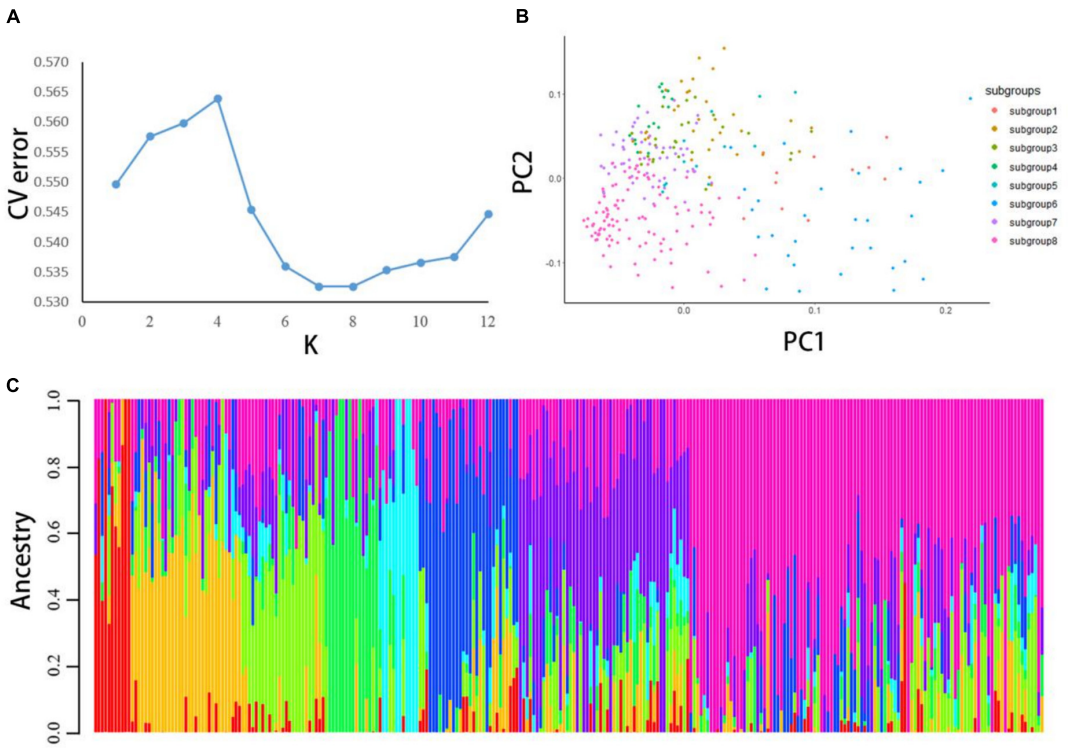


FIGURE 1 | (A) The population structure of 284 potato materials was analyzed by using ADMIXTURE software. CV error was calculated when K = 1–12. **(B)** PCA was performed on all 284 potato samples with high-quality polymorphic loci. Each dot represents a sample. **(C)** When K = 8. In this population structure, each individual is represented by a line with eight different colors. According to the proportion of colors, which subgroup the variety belongs to can be inferred.

Potato Genome-Wide Linkage Disequilibrium

The LD-decay curve was obtained by analyzing all 284 potato germplasm resources using the 22,489 SNP markers obtained from the whole genome (**Figure 5**). The results show that LD decreased as the physical distance between SNPs increased. When taking the coefficient of determination R^2 = 0.2 as the decay threshold, the results show that LD decreases with increased physical distance between SNPs. When the coefficient of determination R^2 = 0.2 is taken as the attenuation threshold. The attenuation distance from subgroups 1–8 was about 13.4, 1.3, 0.1, 0.7, 0.3, 1.9, 1.3, and 0.6 kb, respectively, and the attenuation distance of the whole population was about 0.9 kb. It was much lower than that of cultivated rice (123 kb), cultivated soybean (133 kb), and cultivated maize (30 kb), and slightly lower than that of cultivated cassava (8 kb) and a maize-inbred line population (1.5 kb) (Traut, 1994).

Candidate Genes

According to the LD, we found 14 candidate genes. Most of the candidate genes were within 10 kb of the mutation site, and the largest was no more than 20 kb. **Table 6** shows the detailed information. Among these genes, a gene

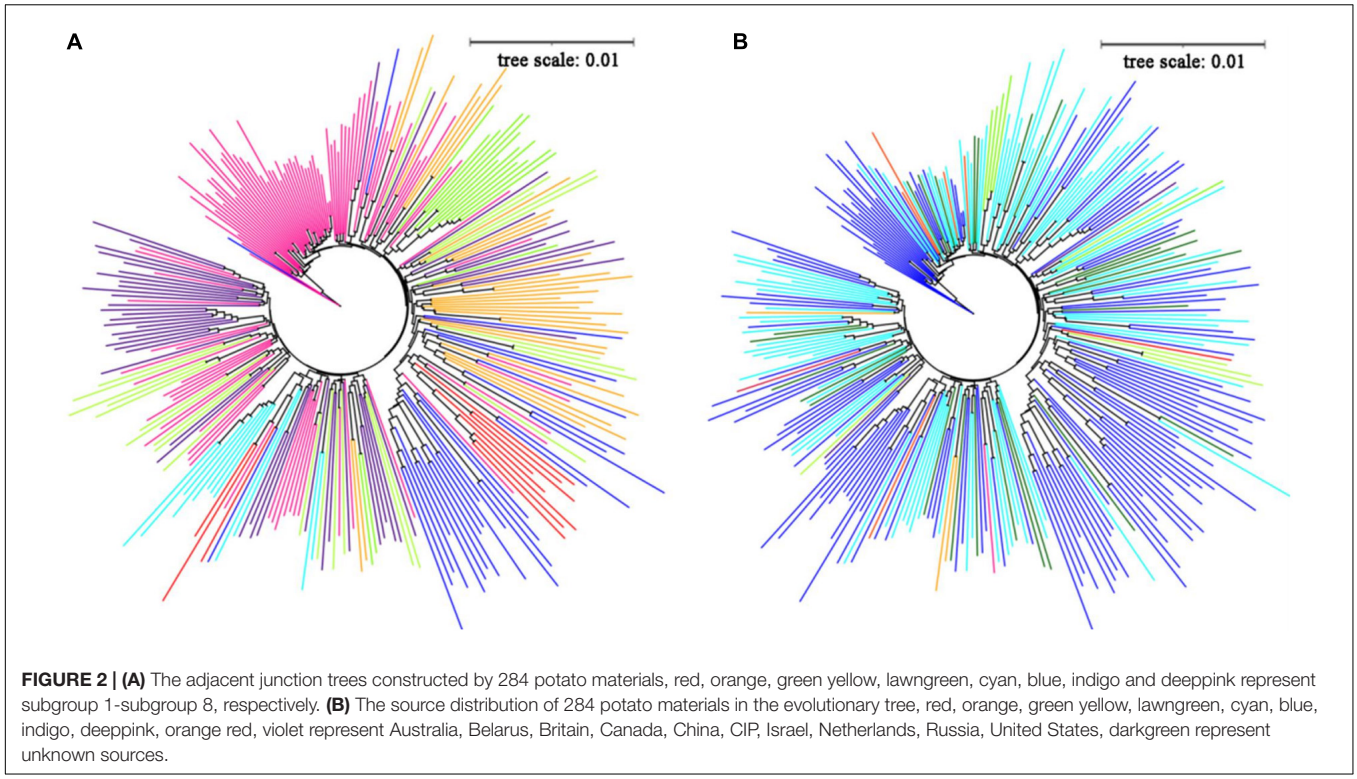


FIGURE 2 | (A) The adjacent junction trees constructed by 284 potato materials, red, orange, green yellow, lawngreen, cyan, blue, indigo and deeppink represent subgroup 1-subgroup 8, respectively. **(B)** The source distribution of 284 potato materials in the evolutionary tree, red, orange, green yellow, lawngreen, cyan, blue, indigo, deeppink, orange red, violet represent Australia, Belarus, Britain, Canada, China, CIP, Israel, Netherlands, Russia, United States, darkgreen represent unknown sources.

TABLE 4 | Summary of genetic diversity (π).

Subgroup	1	2	3	4	5	6	7	8	Total
π	0.238	0.239	0.235	0.201	0.208	0.250	0.213	0.164	0.216

TABLE 5 | Summary of population pairwise *F*-statistics (*F*st).

Subgroup	1	2	3	4	5	6	7	8
1								
2	0.092							
3	0.088	0.029						
4	0.144	0.060	0.071					
5	0.131	0.080	0.072	0.123				
6	0.074	0.064	0.066	0.117	0.094			
7	0.108	0.033	0.025	0.061	0.073	0.087		
8	0.149	0.060	0.048	0.097	0.110	0.105	0.030	

encoding chitinase 1 was on chromosome 11 related to the immune response; six protein kinases, serine-threonine protein phosphatase, and mitogen-activated protein kinase genes were found around the significant sites on chromosomes 1, 4, 5, and 6. These protein kinases can catalyze protein phosphorylation and mediate signal transduction to external stimuli; on chromosomes 3 and 7, genes encoding ERF transcription factor, ethylene-responsive transcription factor 4, and ethylene receptor 2 were associated with ethylene response. Ethylene can coregulate with jasmonic acid pathogen invasions; on chromosomes 5, 6, and 8, the plant resistance protein, NBS-LRR tolerance protein, and late blight tolerance protein may be related to the immune response of late blight and the specific identification of pathogen effectors. NBS-LRR protein is the most important tolerance protein of late blight, which is directly related to the tolerance of potatoes to late blight. Additionally, one gene with unknown functions was on chromosome 6, requiring further verification.

Expression Patterns of Candidate Genes

Two candidate genes were randomly selected to verify their gene expression patterns in late blight tolerant varieties A1, CIP10-1, and 0422-19 and susceptible varieties D8, UK7, and FAVORITA by qRT-PCR. After inoculation with *Phytophthora infestans in vitro* for 5 days, the green leaves around the plaque and leaves without inoculation were

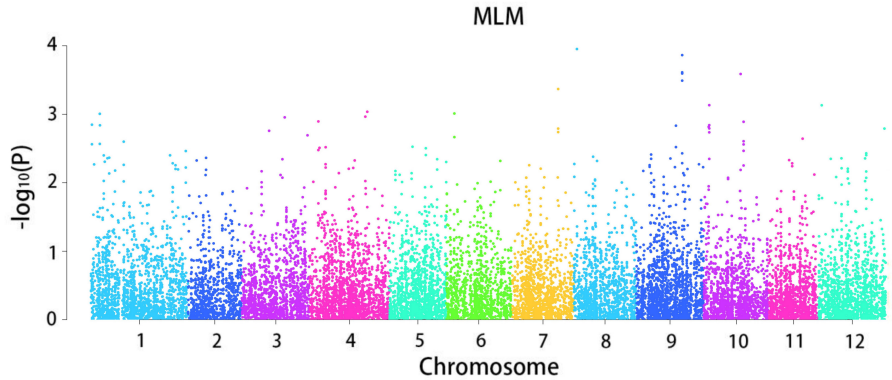


FIGURE 3 | Manhattan plot of the whole-genome association analysis of disease grade.

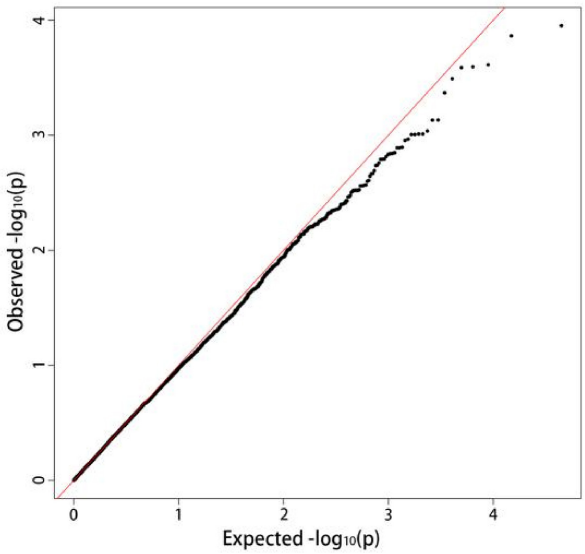


FIGURE 4 | QQ plot of the whole-genome association analysis of disease grade.

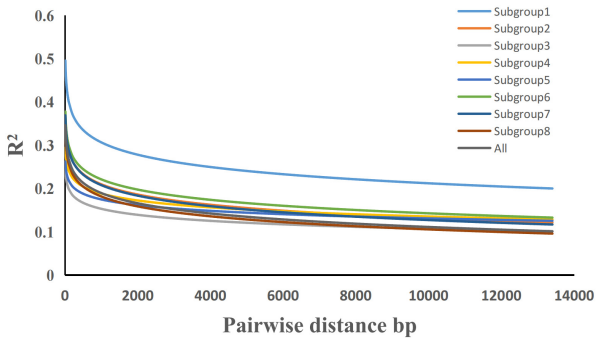


FIGURE 5 | Genome-wide average linkage disequilibrium decay estimated from all potato samples.

used. Then, their relative expression levels were calculated (**Figure 6**). The results showed that most of the candidate genes were upregulated after inoculation. The expression levels of PGSC0003DMG400028682 in varieties that were tolerant were higher than those in susceptible varieties, and the expression levels of PGSC0003DMG400000043 in susceptible variety UK7 were higher than those in the tolerant variety 0422-19. The results showed that the tolerance to late blight was a quantitative trait controlled by multiple genes.

DISCUSSION

In our study, we evaluated the tolerance to *Phytophthora infestans* in 284 potato germplasms. Of these, 37 potato germplasms showed high tolerance. Population structure analysis assigned 284 potato germplasms into eight subpopulations. The high-quality SNP markers revealed a moderate level in the main differentiation among the eight subpopulations. Simultaneously, GWAS was conducted to identify the significant loci for late blight tolerance traits in germplasm resources and screen candidate genes.

Identification and Analysis of Late Blight Tolerance of Germplasm Resources

In recent years, many reports present the identification of potato tolerance to late blight; however, there are some differences in the identification results. The mixed strains selected in this study have yet to be determined at the physiologic race level. With our studies for many years in the laboratory, we discovered that the tolerance of potato leaves to *Phytophthora infestans in vitro* was consistent with that at the adult stage. Therefore, the identification results of detached leaves could represent the tolerance at the adult stage. Favorita is a highly susceptible late blight variety, which is again consistent with our results. Identifying late blight-tolerant germplasm resources and discovering durable tolerant materials could be an important issue in

TABLE 6 | Candidate genes of significant association markers.

CHR	Position	Candidate gene	Start	End	Description
1	72,247,021	PGSC0003DMG400000043	72,233,966	72,243,677	Protein kinase
3	44,871,940	PGSC0003DMG400036493	44,865,282	44,866,451	ERF transcription factor
4	8,095,786	PGSC0003DMG400023584	8,093,725	8,096,358	Serine-threonine protein kinase
4	9,276,044	PGSC0003DMG400019737	9,274,796	9,280,465	Serine-threonine protein kinase
5	31,701,650	PGSC0003DMG400033661	31,687,765	31,689,237	Mitogen-activated protein kinase
5	3,253,2025	PGSC0003DMG400019926	32,513,930	32,514,640	Plant resistance protein
5	51,527,765	PGSC0003DMG400023346	51,518,931	51,524,033	Serine-threonine protein phosphatase
6	7,094,249	PGSC0003DMG400033671	7,082,039	7,082,791	Gene of unknown function
6	935787	PGSC0003DMG400031878	920,425	927,075	NBS-LRR resistance protein
6	39,855,777	PGSC0003DMG400016323	39,852,449	39854311	Serine-threonine protein kinase
7	49532127	PGSC0003DMG400026821	49,531,657	49,532,565	Ethylene-responsive transcription factor 4
7	51,816,228	PGSC0003DMG400027651	51,816,086	51,822,785	Ethylene receptor 2
8	11,363,730	PGSC0003DMG400047159	11,343,551	11,343,883	Late blight resistance protein
11	6,786,565	PGSC0003DMG400028682	6,788,959	6,789,991	Chitinase 1

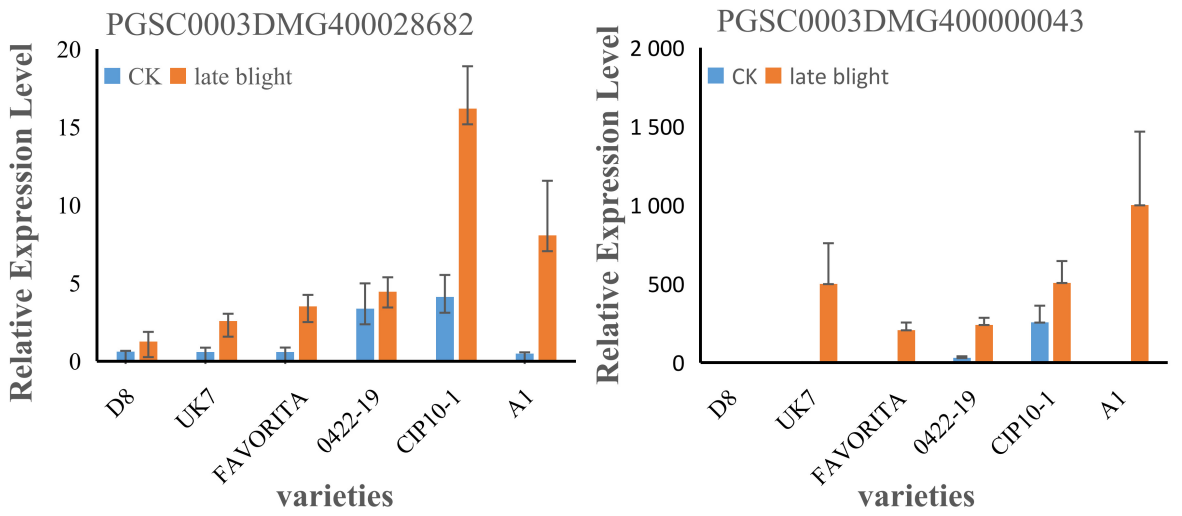


FIGURE 6 | Experimental verification of gene expression levels by qRT-PCR.

breeding high-quality, late blight–tolerant potato varieties in the future. In this study, 37 materials showed immunity, and 15 germplasms conferred high tolerance to late blight disease, providing valuable materials for late blight-tolerant potato breeding.

Compared with identifying natural diseases in the field, *in vitro* leaf inoculation identification has certain advantages and accuracy. First, the identification of *in vitro* leaf inoculation is indoor; the temperature and humidity conditions are controllable, which can create the most suitable conditions for the occurrence of late blight. The identification results are more accurate, avoiding the error caused by human factors in identifying natural diseases in the field. Additionally, the physiological races in the Qinghai Province of Northwest China were used to identify the tolerance to late blight for 2 years. Breeding tolerant varieties is the most economical and effective means to prevent and control late blight. The screening and utilization of tolerant resources is the basis and core of disease-tolerance breeding. In this study, *in vitro* leaf inoculation identification can identify its tolerance to late blight earlier than potato tuber inoculation identification, and the incidence can be clearly seen. Additionally, *in vitro* leaf inoculation identification can avoid releasing the late blight pathogen into the environment, causing a large area of late blight or becoming a potential risk.

Analysis of Potato Population Structure and Genetic Diversity

In this study, 284 potato germplasm resources were used for the association analysis of the tolerance to late blight, 138 of which were from the CIP. By testing genetic markers for Hardy–Weinberg equilibrium and suballelic frequency filtering, 22,489 high-quality polymorphic loci were obtained. The stratification of population structure and the uneven distribution

of alleles are important reasons for false associations between genotypes and traits (Flint-Garcia et al., 2003). The population structure analysis of 284 potato germplasm resources in this study found that, when K = 8, the CV error value was the smallest. Therefore, 284 potato germplasm resources were divided into eight subgroups. The genetic diversity index of the potato population was 0.216, indicating that there was abundant genetic diversity. The F_{ST} among subgroups was mostly between 0.05 and 0.15, and the differentiation between subgroups was mostly moderate, indicating a certain degree of differentiation in germplasm resources. Still, the degree of differentiation was not high.

Analysis of Candidate Genes

The reference genome is a diploid potato reference genome in this study, and no tetraploid genome existed. Through GWAS, significant variation sites are mined to reasonably analyze and mine candidate genes. The SNPs in annotated loci were obtained by combining the SNP genomic positions (Maria et al., 2017). In the future, if there is a better four ontology reference genome, we will fully consider LD, and then more detailed and accurate mining of candidate genes.

Late blight is a serious disease worldwide, threatening the potato industry and food security. Late blight-tolerance potato breeding is very important in China and abroad, and breeding to have multiple disease-tolerance genes in the same variety is important for preventing late blight. So far, 11 broad-spectrum tolerance genes (R1–R11) have been discovered, and these 11 major R genes have been successfully located on the potato genetic map (Li et al., 1998; Ballvora et al., 2002). Combinations of different disease-tolerance genes can provide ideal late blight tolerance (Rietman et al., 2012). Therefore, optimizing the known disease-tolerance gene combinations, making full use of R disease-tolerance genes with broad-spectrum tolerance characteristics, and discovering new broad-spectrum and longer-lasting disease-tolerance genes from the abundant potato resources are effective means for cultivating tolerant potato varieties in the future (Haverkort and Hillier, 2011). The abundant wild resources of potatoes are the source of R genes. Potato contains many genes encoding cytosolic NBS-LRR tolerance proteins (Jupe et al., 2012). The R gene can be introduced into cultivated varieties by crossing conventional varieties with wild species containing this gene, thus helping cultivate varieties to achieve durable tolerance to late blight (Halterman et al., 2008). In the abundant potato resources, many unknown late blight-tolerance genes are waiting to be discovered.

In this study, 19 candidate genes, respectively, were found by the GWAS of the disease severity grade identified from isolated potato leaves affected by late blight *in vitro*. Among them, the candidate gene encoding the chitinase 1 gene may be involved in the immune response, and chitinase can improve the tolerance of plants to fungi (Hiroshi et al., 1997; Tabei et al., 1998). The GWAS of disease tolerance grade found that three candidate genes were associated with ethylene response. Ethylene could coordinate with jasmonic acid to regulate plant immunity during pathogen invasion (Lieselotte et al., 2014). Two resistance proteins and one NBS-LRR tolerance protein may be directly related to potato late blight tolerance by the correlation analysis of the disease-tolerance grade. NBS-LRR proteins are the most critical late blight tolerance proteins, characterized by the same conserved structure containing an N-terminal leucine zipper or coiled-coil, a nucleotide-binding site, and leucine-rich repeats. The cloned anti–late blight genes have several highly conserved structures, such as a phosphate-binding domain (Ploop), a kinase-2 group, or a GLPL group in the NBS region (Traut, 1994; Meyers et al., 1999). This study provides genetic resources for follow-up research and lays the foundation for genetic improvement of potato tolerance to late blight.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://ngdc.cncb.ac.cn/search/?dbId=bioproject&q=PRJCA005945&page=1, PRJCA005945.

AUTHOR CONTRIBUTIONS

FW, ZX, and JW conceived and designed the experiments. FW and LZ conducted the experiments work. ZX, MZ, and LZ analyzed the experimental results. FW and MZ wrote the manuscript. ZX and JW reviewed and contributed to improve it and revised the last version of the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 31660417), the Modern Agricultural Industry Technology System (No. CARS-9), the Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (No. 1630052019022), and the Science and Technology Department of Qinghai Province (No. 2019-NK-A1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.714575/full#supplementary-material

Supplementary Table 1 | The 284 potato germplasm resources used for GWAS.

Supplementary Table 2 | Identification and analysis of the late blight tolerance of potato leaves *in vitro*.

Supplementary Table 3 | Significant variation sites.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1161/01.ATV.0000137190.63214.c5

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., et al. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* 1:e60. doi: 10.1371/journal.pgen.0010060

Ballvora, A., Ercolano, M. R., Weiss, J., Meksem, K., Bormann, C. A., Oberhagemann, P., et al. (2002). The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.* 30, 361–371. doi: 10.1046/j.1365-313X.2001.01292.x

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2003). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Dunia, P., Ram, K. B., Ric, C. H. D. V., Chris, M., Richard, V., and Guusje, B. (2011). The patterns of population differentiation in a *Brassica rapa* core collection. *Theor. Appl. Genet.* 122, 1105–1118. doi: 10.1007/s00122-010-1516-1

Enciso-Rodriguez, F., Douches, D., and Lopez-Cruz, M. (2008). Genomic selection for late blight and common scab resistance in tetraploid Potato (*Solanum tuberosum*). *G3 Genes Genomes Genet.* 8, 2471–2481. doi: 10.1534/g3.118.200273

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374.

Fry, W. E., Birch, P. R. J., Judelson, H. S., Grunwald, N. J., Danies, G., Everts, K. L., et al. (2002). Five reasons to consider *Phytophthora infestans* a reemerging pathogen. *Phytopathology* 105, 966–981. doi: 10.1094/PHYTO-01-15-0005-FI

Halterman, D. A., Kramer, L. C., Wielgus, S., and Jiming, J. (2008). Performance of transgenic potato containing the late blight resistance gene RB. *Plant Dis.* 92, 339–343. doi: 10.1094/PDIS-92-3-0339

Hansen, M., Kraft, T., Ganestam, S., Säll, T., and Nilsson, N. O. (2001). Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genetical Res.* 77, 61–66. doi: 10.1017/S0016672300004857

Haverkort, A. J., and Hillier, J. G. (2011). Cool farm tool - potato: model description and performance of four production systems. *Potato Res.* 54, 355–369. doi: 10.1007/s11540-011-9194-1

Hiroshi, A., Yoko, N., and Shigeru, A. (1997). Enhanced resistance against a fungal pathogen sphaerotheca humuli in transgenic strawberry expressing a rice chitinase gene. *Plant Biotechnol.* 14, 145–149. doi: 10.5511/plantbiotechnology.14.145

Jupe, F., Pritchard, L., Etherington, G. J., Mackenzie, K., Cock, P. J., Wright, F., et al. (2012). Identification and localisation of the NBS-LRR gene family within the potato genome. *BMC Genomics* 13:75. doi: 10.1186/1471-2164-13-75

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods.* 9:9. doi: 10.1186/1746-4811-9-29

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 9, 357–359. doi: 10.1038/nmeth.1923

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

Li, X., Eck, H. J. V., Voort, J. N. A. M. R., Huigen, D. J., Stam, P., and Jacobsen, E. (1998). Autotetraploids and genetic mapping using common AFLP markers: the R2 allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome 4. *Theor. Appl. Genet.* 96, 1121–1128. doi: 10.1007/s001220050847

Lieselotte, D. B., Monica, H., and David, D. V. (2014). Connecting growth and defense: the emerging roles of brassinosteroids and gibberellins in plant innate immunity. *Mol. Plant* 7, 943–959. doi: 10.1093/mp/ssu050

Maria, S. E., Ding, J., Enrique, R., Maria, J. P., Nicolás, C., Ekhard, T., et al. (2017). Physical mapping of QTL for tuber yield, starch content and starch yield in tetraploid potato (*Solanum tuberosum* L.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array. *BMC Genomics* 18:642. doi: 10.1186/s12864-017-3979-9

Maurice, S., Montes, M. S., Nielsen, B. J., Bdker, L., Martin, M. D., Jnck, C. G., et al. (2009). Population genomics of an outbreak of the potato late blight pathogen. *Phytophthora infestans*, reveals both clonality and high genotypic diversity. *Mol. Plant Pathol.* 20, 1134–1146. doi: 10.1111/mpp.12819

McDonald, B. A., and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40, 349–379.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-313X.1999.00606.x

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. *Nature* 456, 720–723. doi: 10.1038/nature07629

Paul, R. J. B., Glenn, B., Brian, F., Eleanor, M. G., Ingo, H., John, T. J., et al. (2012). Crops that feed the world 8: potato: are the trends of increased global production sustainable? *Food Secur.* 4, 477–508. doi: 10.1007/s12571-012-0220-1

Pradhan, S. K., Barik, S. R., Sahoo, A., Mohapatra, S., Nayak, D. K., Mahender, A., et al. (2016). Population structure, genetic diversity and molecular marker trait association analysis for high temperature stress tolerance in rice. *PLoS One* 11:e0160027. doi: 10.1371/journal.pone.0160027

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for Whole-Genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Rietman, H., Bijsterbosch, G., Cano, L. M., Lee, H. R., Vossen, J. H., Jacobsen, E., et al. (2012). Qualitative and quantitative late blight resistance in the potato cultivar sarpo mira Is determined by the perception of five distinct RXLR effectors. *Mol. Plant Microbe Interactions* 25, 910–919. doi: 10.1094/MPMI-01-12-0010-R

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516–1517. doi: 10.1097/00001648-199805000-00023

Tabei, Y., Kitade, S., Nishizawa, Y., Kikuchi, N., Kayano, T., Hibi, T., et al. (1998). Transgenic cucumber plants harboring a rice chitinase gene exhibit enhanced resistance to gray mold (*Botrytis cinerea*). *Plant Cell Rep.* 17, 159–164. doi: 10.1007/s002990050371

Traut, T. W. (1994). The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* 222, 9–19. doi: 10.1111/j.1432-1033.1994.tb18835.x

Wang, M., Yan, J. B., and Zhao, J. R. (2012). Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* 196, 125–131. doi: 10.1016/j.plantsci.2012.08.004

Wu, E. J., Yang, L. N., Zhu, W., Chen, X. M., Sheng, L. P., and Zhen, J. S. (2016). Diverse mechanisms shape the evolution of virulence factors in the potato late blight pathogen *Phytophthora infestans* sampled from China. *Sci. Rep.* 6:26182. doi: 10.1038/srep26182

Xia, Z. Q., Zou, M. L., Zhang, S. K., Feng, B. X., and Wang, W. Q. (2014). AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. *Sci. Rep.* 4:7300. doi: 10.1038/srep07300

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant

call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, S. K., Chen, X., Lu, C., Ye, J. Q., Zou, M. L., and Lu, K. D. (2018). Genome-Wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* Crantz). *Front. Plant Sci.* 9:503. doi: 10.3389/fpls.2018.00503

Zhou, Y., Tang, H., Cheng, M. P., Dankwa, K. O., Chen, Z. X., Li, Z. Y., et al. (2017). Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces. *Front. Plant Sci.* 8:401. doi: 10.3389/fpls.2017.00401

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. M. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi: 10.3835/plantgenome2008.02.0089

Zwankhuizen, M. J., and Zadoks, J. C. (2002). *Phytophthora infestans*'s 10-year truce with Holland: a long-term analysis of potato late-blight epidemics in the Netherlands. *Plant Pathol.* 51, 413–423. doi: 10.1046/j.1365-3059.2002.00738.x

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.

Copyright © 2021 Wang, Zou, Zhao, Xia and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

METHODS
published: 26 October 2021
doi: 10.3389/fgene.2021.758326

Transcriptomic and Lipidomic Analysis of Lipids in *Forsythia suspensa*

Bei Wu[1]†, Yinping Li[2]†, Wenjia Zhao[3]†, Zhiqiang Meng[1], Wen Ji[1] and Chen Wang[4]*

[1]Department of Pharmacy, Second Hospital of Shanxi Medical University, Taiyuan, China, [2]Institute of Pomology, Chinese Academy of Agricultural, Sciences (CAAS), Xingcheng, China, [3]College of Agriculture, Shanxi Agricultural University, Taigu, China, [4]Department of Pharmacy, Shanxi Eye Hospital, Taiyuan, China

Forsythiae Fructus (Lianqiao in Chinese) is widely used in traditional Chinese medicine. The lipid components in Forsythiae Fructus are the basis of plant growth and active metabolism. Samples were collected at two growth stages for a comprehensive study. Transcriptome and lipidomics were performed by using the RNA-seq and UPLC-Q-TOF-MS techniques separately. For the first time, it was reported that there were 5802 lipid components in Lianqiao comprised of 31.7% glycerolipids, 16.57% phospholipids, 13.18% sphingolipids, and 10.54% fatty acids. Lipid components such as terpenes and flavonoids have pharmacological activity, but their content was low. Among these lipids which were isolated from Forsythiae Fructus, 139 showed significant differences from the May and July harvest periods. The lipids of natural products are mainly concentrated in pregnenolones and polyvinyl lipids. RNA-Seq analysis revealed 92,294 unigenes, and 1533 of these were differentially expressed. There were 551 differential genes enriched in 119 KEGG pathways. The *de novo* synthesis pathways of terpenoids and flavonoids were explored. Combined with the results of lipidomics and transcriptomics, it is hypothesized that in the synthesis of abscisic acid, a terpenoid, may be under the dynamic regulation of genes EC: 1.1.1.288, EC: 1.14.14.137 and EC: 1.13.11.51 in balanced state. In the synthesis of gibberellin, GA20-oxidase (GA20ox, EC: 1.14.11.12), and GA3-oxidase (GA3ox, EC: 1.14.11.15) catalyze the production of active GAs, and EC: 1.14.11.13 is the metabolic enzymes of active GAs. In the synthesis of flavonoids, MF (multifunctional), PAL (phenylalanine ammonia-lyase), CHS (chalcone synthase), ANS (anthocyanidin synthase), FLS (flavonol synthase) are all key enzymes. The results of the present study provide valuable reference information for further research on the metabolic pathways of the secondary metabolites of *Forsythia suspensa.*

Keywords: Forsythiae Fructus, RNA sequencing, lipid, metabolism, transcriptome, candidate genes, terpenoids, falvonoids

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Linlin Jing,
940th hospital, China
Jinhui Li,
Boston University, United States

***Correspondence:**
Bei Wu
wubei8005@163.com

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

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

Received: *13 August 2021*
Accepted: *28 September 2021*
Published: *26 October 2021*

Citation:
Wu B, Li Y, Zhao W, Meng Z, Ji W and Wang C (2021) Transcriptomic and Lipidomic Analysis of Lipids in Forsythia suspensa. *Front. Genet. 12:758326. doi: 10.3389/fgene.2021.758326*

INTRODUCTION

Forsythia suspensa (Thunb.) Vahl, a plant in the family Oleaceae, is widely planted in East Asia and is a traditional medicinal tree species. It is one of the 40 commonly used medicinal species in China (Sun et al., 2018). Forsythiae Fructus (FF, called Lianqiao in Chinese) is the dried fruit of *F. suspensa,* and in traditional Chinese medicine it is thought to have "heat clearing" and "detoxifying" effects.

Currently, FF is used as a medical treatment for pyrexia, inflammation, gonorrhea, carbuncles and erysipelas. Due to different harvest times and processing methods, FF can be divided into "Qingqiao" and "Laoqiao". Clinically, "Qingqiao" is used more frequently (Guo et al., 2007; Wu et al., 2019). At present, 321 chemical components have been isolated from FF, including phenylethanoid glycosides, lignans, flavonoids, terpenes and volatile oils (Dong et al., 2017). Modern Chinese Medicine studies have found that most of these substances have pharmacological activity (Wang et al., 2018). Forsythiaside A is a phenylethanoid glycoside and it shows remarkable anti-inflammation, antivirus, neuroprotection, antioxidant, hepatoprotection, and antibacterial activity (Gong et al., 2021). Forsythin, a lignan substance, mitigates apoptosis and oxidative stress (Du et al., 2020), has anti-viral and anti-inflammatory activity (Ma et al., 2020), and improves insulin resistance (Xu et al., 2019). Rutin is a flavonoid compound with many pharmacological effects, including antioxidant, anti-inflammatory, anti-diabetic, neuroprotective (Chua, 2013), vascular protection (Stanely and Priya, 2010), and anti-cancer (Ben et al., 2016). The volatile oil of FF is a primary activeone of the main pharmacological components and has the effects of antiviral, antibacterial, antioxidant (Jiao et al., 2012), anti-inflammatory, anti-tumor, acaricidal (Lee and Lee, 2015). The basis of the heat-clearing effect of FF in Chinese medicine prescriptions is the anti-inflammatory and antioxidant properties. The detoxification effect is attributed to its antibacterial, antiviral and anticancer activities (Wang et al., 2018). Based on this, FF is the main active ingredient in many widely used classic Chinese patent medicine prescriptions, such as Shuanghuanglian injections (Zhou et al., 2011) and Lianhua Qingwen granules (Hu et al., 2021). Lianhua Qingwen is also recommended for the treatment of COVID 19 (Hu et al., 2021; Liu et al., 2021). Because FF is widely used, it is necessary to conduct research on this product and its pharmacologically active substances.

The pharmacologically active substances in FF are secondary metabolites. Plant growth, development and reproduction are achieved through material metabolism. Secondary metabolism produces compounds that are often involved in ecological interactions and include pigments, scents, and antibiotics. Secondary metabolites are used in medicine, food, dyes, and agriculture (Srivastava and Srivastava, 2007). Plant-derived medicinal secondary metabolites are mainly obtained by direct extraction from plants. However, to produce such compounds on a large scale would require medicinal crops to be grown extensively and most are currently not well suited to production agriculture. Also, the content of medicinal secondary metabolites in plants is low and extraction and separation are difficult. New production methods are being adopted, including extraction from plant tissue cell culture and the use of microbial cell factories for production (Jiang et al., 2021). The use of artificial microbial cell factories is considered feasible, and this approach has been used for the antimalarial drug artemisinin (Ro et al., 2006), the analgesic drug morphine (Thodey et al., 2014), the antitumor drug paclitaxel (Chen et al., 2019), the cardiovascular disease drug tanshinone (Yadav et al., 2020), and other aromatic secondary products. Efficient synthesis of metabolites is an important prerequisite for constructing a microbial cell factory for the secondary metabolites of FF and realizing efficient utilization of FF. The biosynthetic pathway analysis of secondary metabolites is an important prerequisite to improved production methods. Jiao et al. (2013) found that the oil yield of FF's seeds is about 30% and showed that FF seeds have great potential as a biodiesel raw material resource. FF's contain volatile oils, flavonols, anthocyanins, β-pinene and other compounds. The synthesis and metabolism of secondary metabolites are mostly related to lipids. Abscisic acid (ABA) is insoluble in water. As a plant hormone, it has physiological functions that include causing bud dormancy, leaf shedding, and inhibiting cell growth, and ABA can promote flavonoid metabolism. Lipid metabolism is a key process in energy homeostasis, carbon storage, membrane structure, cell signaling, transcription and translation regulation, and cell and protein interactions (Chapman and Feussner, 2016). Therefore, lipids play an important role in plant growth, development, and reproduction (Chapman and Feussner, 2016). Lipids are unique and functionally specific compared to other metabolites. Therefore, lipid composition and abundance can be used to monitor changes in plants over time and their response to specific stimuli. The lipid components of FF have not been fully studied. Because of the importance of these compounds, a comprehensive analysis and comparison of them in FF is needed. Clarifying the biosynthetic pathways and regulation methods of secondary metabolites in FF will be useful for stabilizing and improving this important product. The content of secondary metabolites, the improvement of the quality of medicinal materials, and the biological preparation technology of secondary metabolites are of great significance. Lipidomics deals with lipids and lipid interactions in biological systems. Lipidomic analysis methods include chromatography, nuclear magnetic resonance spectroscopy, and mass spectrometry-based analysis methods. Liquid chromatography-mass spectrometry is currently the most widely used method for lipidomics research. Chromatography mass spectrometry uses the separation ability of chromatography and the high sensitivity of mass spectrometry to achieve in-depth analysis of lipids (Xu et al., 2020). To identify and quantify all lipid components in FF and systematically analyze the lipid composition and expression changes in the growth process of FF, the present study applied high-resolution liquid phase tandem mass spectrometry technology for lipidomic analysis.

The syntheses of secondary metabolites requires the participation of various metabolic pathways, and is regulated by gene expression. High-throughput transcriptome sequencing analysis can reflect the gene expression of cells, tissues, or organisms over time and can identify candidate genes related to the metabolism of active substances. This analysis method provides gene regulation clues for the biosynthetic analysis of FF secondary metabolites. This study makes use of transcriptomics and lipidomics methods to 1) determine the lipid composition of FF over time during normal growth, and 2) screen candidate genes related to the synthesis of lipid secondary metabolites. Results can provide a reference for increasing the content of secondary metabolites in FF, breeding high-quality FF varieties, and realizing large-scale production of secondary metabolites.



FIGURE 1 | (A) *Forsythia suspensa* (Thunb.) Vahl; **(B)** *Forsythia* fruit growing on the tree **(C)** Forsythiae Fructus as a medicinal material.

MATERIALS AND METHODS

Plant Material and Treatments

FF was harvested in May and July 2021 in Taiyuan, Shanxi (112°33′E, 37°54′N, elevation 800 m, **Figure 1**). Normal plants were selected. Three to five fruits were taken from the same plant, frozen in liquid nitrogen for 120 min immediately after harvesting, and stored in an ultra-low temperature refrigerator at −80°C. The samples were similar in size and were identified by Jinping Luo, Chief Pharmacist of the Institute of Pharmaceutical Inspection Technology, Shanxi Provincial Inspection and Testing Center. The specimens were stored in the ultra-low temperature refrigerator at −80°C in the laboratory of the Department of Pharmacy, Second Hospital of Shanxi Medical University.

Chemicals and Reagents

Methanol: 67-56-1, LC-MS, Merck; Acetonitrile: 75-05-8, LC-MS, Merck; Ammonium acetate: 631-61-8; Formic acid: 64-18-6, LC-MS, TCI; Isopropanol: 67-63-0, LC-MS, Merck; Methyl tert-butyl ether, TCI LC-MS, TCI; Methyl tert-butyl ether (MTBE): 1634-04-4, LC-MS, Merck; Dichloromethane: 75-09-2, LC-MS, Merck.

Sample Preparation and Extraction

FF picked in May were designated as group A and those picked in July were designated as group B. Three samples of each group of the same size were selected. The sample is ground into powder with liquid nitrogen. A sample of 50 mg each were weighed, 200 μL of water was added, and the samples vortexed for 30 s. Steel beads were added, and the samples were ground at 1000 rpm for 12 min, then sonicated for 15 min in an ice water bath. Next, 480 μL of extraction solution (methyl tert-butyl ether: methanol = 5:1) was added, the samples vortexed for 30 s, sonicated for 10min in an ice-water bath, and allowed to stand for 1 hour at minus 20°C. The samples then were centrifuged at 10,000 rpm for 15 min at 4°C. A supernatant alloquot of 380 μL was removed, dried under vacuum, and 200 μL of 1:1methylene chloride: methanol was added for re-solubilization, the solution vortexed for 30 s, and sonicated for 10 min in an ice-water bath. The samples were centrifuged at 13,000 rpm for 15 min at 4°C. From each sample, 180 μL of supernatant wasremoved and mixed into a quality control sample by taking 10 μL of each sample.

Chromatographic Conditions

The sample extracts were analyzed using a UPLC Acquity I-Class PLUS column (Waters). The UPLC column analytical conditions were as follows. Waters Acquity UPLC CSH C18 (1.7 μm, 2.1 mm, 100 mm); injection volume 5 μL; temperature 55°C; flow rate 400 μL/min; mobile phase A 60% acetonitrile aqueous solution, 10 mM ammonium acetate, 0.1% formic acid; mobile phase B 90% isopropanol acetonitrile solution, 10 mM ammonium acetate, 0.1% formic acid, 10 mM ammonium acetate, 0.1% formic acid. Gradient elution parameters: 0.0–2.0 min, 60–57% A, 40–43% B; 2.1–12.0 min, 50–46% A, 50–54% B; 12.1–18.0 min, 30–1% A, 70–99% B; 18.1–20.0 min, 60% A, 40% B.

Mass Spectrometry Conditions

After UHPLC separation, mass spectrometry was performed with a UPLC Xevo G2-XS QT of (Waters) with electrospray ionization (ESI) positive and negative ion mode detection. ESI ion source parameters were as follows: capillary voltage 2000 V (positive ion mode) or −1500 V (negative ion mode); cone hole voltage: 30 V; ion source temperature: 120°C; dissolvent gas temperature 550°C; backblast gas flow rate: 50 L/h; dissolvent gas flow rate: 900 L/h. The MSe flow rate was controlled by MassLynx V4.2 (Waters). Primary and secondary mass spectrometry data acquisition was performed using MSe mode under MassLynx V4.2 (Waters) control. In each data acquisition cycle, dual-channel data acquisition was performed simultaneously for low and high collision energies. The low collision energywas 2 V, and the high collision energy range was 10–40 V. The scan frequency was 0.2 s for one mass spectrometry map.

Lipid Identification and Quantification

Lipid molecules were identified by referring to the LIPID MAPS (http://www.lipidmaps.org/) and Lipid Bank (http://www.lipidbank.jp/) lipid databases, combined with data on retention times of chromatograms in positive and negative ion modes, primary mass spectra and secondary mass spectra.

The statistical method of orthogonal partial least squares-discriminant analysis (OPLS-DA) was used to analyze non-orthogonal and orthogonal variables separately to obtain more reliable intergroup differences in lipids concerning the degree of correlation between experimental groups information. OPLS-DA model calculations were performed mainly using the R (3.3.2) package ropls. The prediction parameters of the evaluation model are R2X, R2Y, and Q2, where R2X and R2Y indicate the explanation rate of the built model for X and Y matrices, respectively, where the X matrix is the model input, i.e. the lipid quantification matrix, and the Y matrix is the model output, i.e., the sample grouping matrix, and Q2 indicates the prediction ability of the model, i.e., whether the built model can distinguish the correct sample grouping by metabolic expressions. The closer R2Y and Q2 are to 1, the more stable and reliable is the model for screening differential lipids. In general, a model with $Q2 > 0.5$ can be considered valid, and a model with $Q2 > 0.9$ is an excellent model.

The lipidomics data were analyzed for lipid fraction differences mainly by clustering heat map analysis and difference multiplier analysis. Differential lipids were screened by combining the difference multiples, *p*-value of a *t*-test, and VIP value of the OPLS-DA model. The screening criteria were Fold Change (FC) > 2, *p*-value < 0.05 and VIP >1.

RNA Extraction, Sequencing, and *de novo* Assembly

Forsythiae Fructus were sampled for transcriptome sequencing experiments. Sequencing libraries were generated using NEBNext®Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. The mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was done using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamerprimer and M-MuLV Reverse transcriptase. Second strand cDNA synthesis was subsequently performed using DNA polymerase I and rnase H. Fragments of cDNA 240bp in lengthpreferentially selected. The clustering of the index-coded samples was performed on a cBotCluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumia). The library preparations were sequenced on an Illumina Hiseq (6,000 platform and paired-end reads were generated. About 6 GB of raw data from each sample were sequenced. Clean data were obtained by removing reads containing the adapter, reads containing ploy-N, and low quality reads from raw data. Clean data analysis was performed using Trinity.

Gene Annotation and Analysis of Differential Genes

Gene function was annotated based on the following databases: NR (NCBI nonredundant protein sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of Proteins), eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups), Swiss-Prot (a manually annotated and reviewed protein sequence database), KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO (Gene Ontology). Picard - tools v1. 41 and samtools v0. 1.18 were used to sort, remove duplicated reads and merge the bam alignment results of each sample. Gene expression levels were estimated by RSEM for each sample. Fragments per kilobase of transcript per million mapped reads (FPKM) were used to indicate the abundance of expression of the corresponding unigene.

Differential expression analysis of two conditions/groups was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting *p*-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). The *p*-value was adjusted using the q-value, which was set as the threshold for significantly differential expression. The differentially expressed genes (FDR < 0.01, FC > 2) were analyzed.

GO enrichment analysis of the differentially expressed genes (DEGs) was implemented by the top GO R packages-based Kolmogorov–Smirnov test. We used KOBAS software to test the statistical enrichment of differentially expressed genes in KEGG pathways.

The lipidome and transcriptome data of FF were fitted to analyze the regulation mechanism of lipid secondary metabolite synthesis, and candidate genes that affect the synthesis pathway were screened. According to reports, the pharmacologically active lipid secondary metabolites in FF are mainly flavonoids and terpenes. The presentstudy focuses on the analysis of these two types of compounds.

RESULTS

Data Quality Evaluation

Before the formal UHPLC-Q-Tof-MS experiment, the base peak chromatogram of the quality control sample was compared for spectral overlap. The peak response intensities and retention times of the samples overlapped, indicating the experimental instrument has excellent stability and reproducibility (**Figure 2**).

Transcriptomics experiments were grouped in the same way as lipidomics experiments. The Spearman's rank correlation coefficient (r) was used as an index to assess the correlation between sample groups in the transcriptomics experiments. When r^2 is closer to 1, a strong correlation between the two samples is indicated. The intra-group sample correlation $r^2 > 0.7$ is shown in the attached **Supplementary Figure S1**.

Lipid Composition and Content

According to the LMSD (LIPID MAPS Structure database, http://www.lipidmaps.org/data/structure) database classification, 8categories, 66 main classes, 257 lipid subclasses, and 5,802 lipid molecules were detected in FF (**Figure 3**). Glycerolipids (GL) were the most abundant, with 6 main classes and 640 lipids. GLaccounted for about 29.55 and 31.70% of the lipids in group A and B,respectively. Between the two sampling dates, the content of triacylglycerol increased significantly, and the

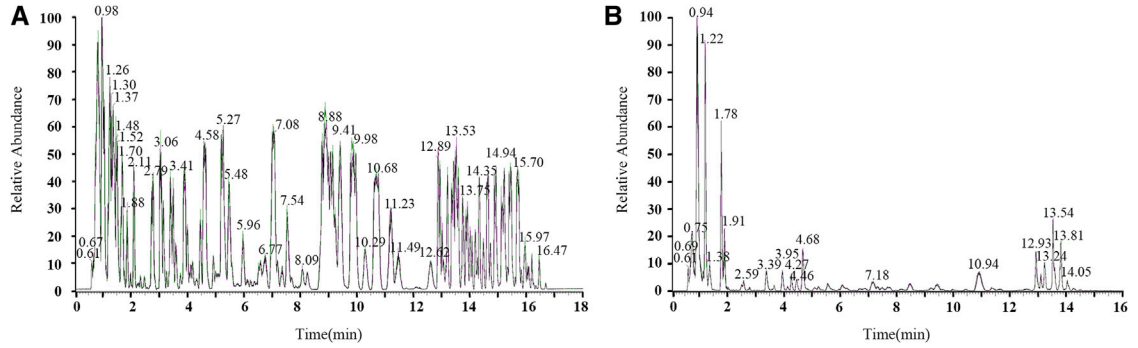


FIGURE 2 | Chromatogram of the quality control sample in ESI-positive **(A)** and ESI-negative **(B)** ion modes.

content of dialkylglycerol and glycosyl dihexylglycerol decreased. Glycerophospholipids (GP) were the next most abundant, with 15 main classes and 931 lipid substances, with 29, 163, 738.21 μg/g and 31, 718, 781.85 μg/g in groups A and B, respectively, accounting for about 16.13 and 16.57% of the total lipids. Across sampling dates, the contents of glycerophosphocholine (PC), glycerophosphoethanolamine (PE), glycerophosphoserine (PS), glycerophosphoinositol (PI) and glycerophosphate (PA) all increased and the content of oxidized glycerophospholipids decreased significantly. Sphingolipids (SP) ranked third, with seven main classes and 726 lipid substances, and the contents of the A and B groups were 17, 631, 677.45 μg/g and 17, 180, 898. 39 μg/g, respectively, accounting for about 14.77 and 13.18%. Ceramide content increased significantly from group A to B, phospholipids and acid glycosphingolipids decreased slightly, and neutral glycosphingolipids and other sphingolipids increased slightly. Fatty acyles (FA) had 13 main classes and 1,695 lipid substances and the contents of the two groups were 12, 881, 185.91 μg/g and 13, 739, 216.99 μg/g, accounting for about 10.79 and 10.54% of total lipids. From group A to B, the content of fatty acids and their conjugates, fatty esters, oxygen-containing hydrocarbons, fatty acyl glycosides and other fatty acyl groups increased whereas the higher fatty acids (octadecane, eicosane, behenyl) and hydrocarbons decreased. There were five main classes of sterol lipids (ST) and 771 lipids with 9,870,796.40 μg/g and 11, 216, 784.53 μg/g, respectively, accounting for 8.27 and 8. 61% of total lipids. There were 12 main classes of polyketides (PK) with 544 lipids, with 7,658,121.17 μg/g and 7,343,043. 54 μg/g, accounting for 6.42 and 5.63% of total lipids, respectively. In PK, the content of acetone in Annonaceae increased; the content of flavonoids and phenolic esters decreased. There were four main classes of prenol lipids (PR) with 443 lipids, with 6,366,599.90 μg/g and 7,369,347. 98 μg/g, accounting for 5.33 and 5.65% of total lipids, respectively. The content of isoprenoids, polypentenols, quinones and hydroquinones increased across sampling dates. Among these compounds were limonene-1,2-diol and (+)-exo-5-hydroxycamphor monoterpenes, selina-4 (14), 7,11-trien-9-ol and other sesquiterpenes, (-)-enunicelline diterpenes such as ximaosarcophytol A, triterpenes such as squalen-1-ol and epi-maslinic acid, and tetraterpenes such as amarouciaxanthin B/sidnyaxanthin and diatoxanthin/7,8-didehydrozeaxanthin. There were also gibberellin (GA) (including GA_{36}, GA_8, GA_9, GA_{24}, GA_{17}, GA_7), abscisic acid, zeatin, lutein, and other plant hormones, but their content did not differ between groups A and B. There were three main classes of glycolipids (SL) with 52 lipids, the contents of groups A and B were 505,912.76 μg/g and 442,897.79 μg/g accounting for 0.42 and 0.34% of total lipids, respectively. The total contents of various other types of lipids in the two groups of samples were less varied between groups A and B.

Differential Lipid Composition

The results of the orthogonal partial least-squares discriminant analysis (OPLS-DA) of differences grouping scores for each group are shown in **Supplementary Figure S2**. Among the two groups (A and B) of samples harvested at different periods, significant differences were considered according to the difference multiplier FC > 2 or FC < 0.5, p-value < 0.05 in the t-test, and VIP >1 in the OPLS-DA model, which in turn screened for variation differential lipids. A total of 139 lipid molecules were significantly different, accounting for 2.40% of the total number of lipids of which 87 (1.50%) had upregulated expression levels (**Figure 4**).

FF sampled in May and July had significant differences in the content of phospholipids, fatty acids, glycerolipids, sphingolipids, and sterol lipids. The contents of 38 types of FA were significantly changed, and 27 types (71.05%) were upregulated representing ten subgroups (**Figure 4D**). Wax monolipids, branched chain fatty acids, prostaglandins, fatty alcohols and hydroxy fatty acids all increased significantly, while rhamnolipids, lactones, and some prostaglandins decreased significantly. The contents of 29 type of GP increased representing seven subgroups, and 17 types (58.62%) were upregulated. The content of GP such as glycerophosphatidylcholine, glycerophosphoglycerol, and glycerophosphoethanolamine increased, while glycerophosphoglycerol and glycerophosphoinositol significantly decreased. The contents of 19 types SP

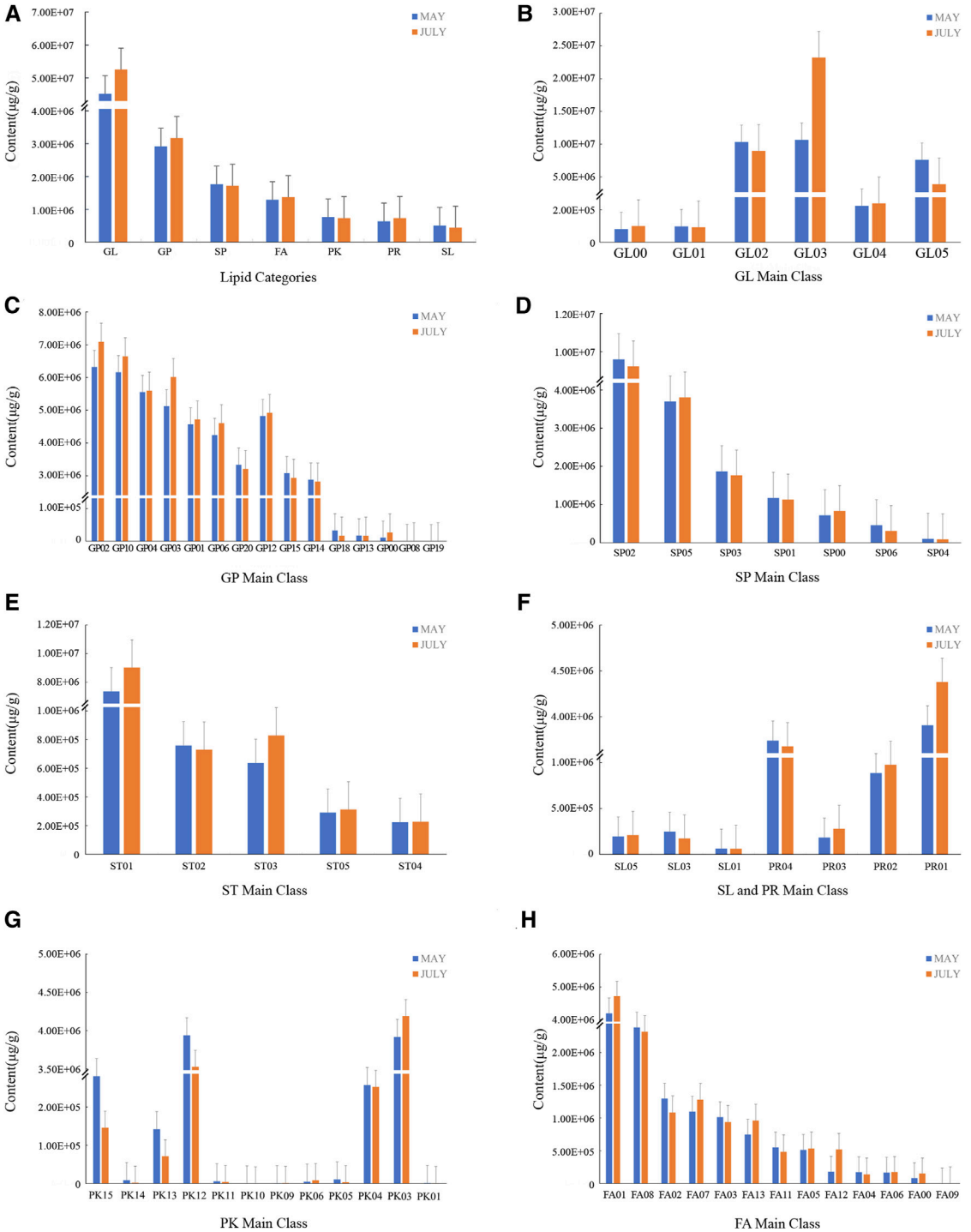


FIGURE 3 | The content of various lipid components of FF differed with growth stage. The May sample is represented in blue and the July sample in red. For each sample date, **(A)** The composition of total lipids contained in FF, **(B)** Glycerolipids (GL), **(C)** Glyceropnospholipids (GP), **(D)** Sphingolipids (SP), **(E)** Sterol Lipids (SL), **(F)** Saccharopilipids (SL) and Prenol Lipids (PR), **(G)** Polyketides (PK), **(H)** Fatty Acyls (FA).

significantly changed, and 11 types (57.89%) were upregulated representing four subgroups. Ceramide increased, and neutral glycosphingolipid substances both increased and decreased. The contents of 15 types of GL significantly changed, and 12 types (80.00%) were upregulated representing four subclasses. The diacylglycerols and triacylglycerols were the largest subgroups and their content increased from May to July. The monoalkylglycerol subgroup content decreased significantly.

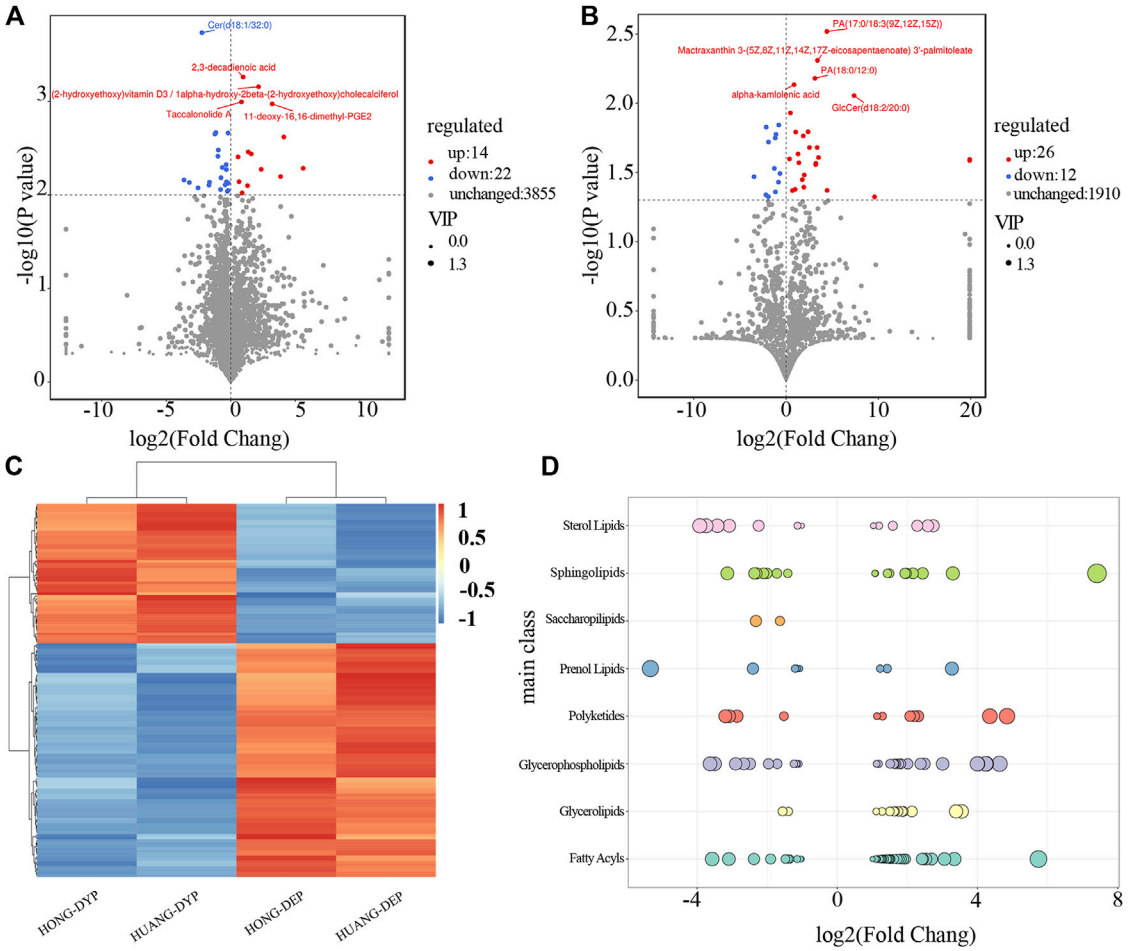


FIGURE 4 | Analysis of differential lipid molecules in FF from May and July sampling dates. Characteristics of differential lipid molecules of FF in ESI-positive **(A)** and ESI-negative **(B)** ion modes. The dots represent lipid molecules, among which the blue (down regulated) and red dots (up regulated) that satisfied FC < 0.5, FC > 2, and $p < 0.05$ indicating significant differences between the two sample dates **(C)** Hierarchical and clustering analysis **(D)** Representation of significantly different lipid distributions at different stages. Each dot represents a lipid, and the larger the dot, the greater the difference in lipids across sampling dates. Log2FC is the logarithmic value of the lipid content of the two groups. Log2FC > 0 means that the lipid content increased from the May to July sample, and Log2FC < 0 means that the lipid content decreased in from the May to July sample.

The contents of 15 types of ST were significantly changed, and six types (40.00%) representing four subgroups were upregulated. There are mainly sterols present. There were only two different glycolipid components, and the content of the two acyltlehaloses decreased significantly.

PR and PK are the pharmacologically active lipid components of FF. The contents of 12 types of PK were significantly changed, and seven types (58.33%) representing three subclasses were upregulated. Among the PK, chalcone and dihydroxychalcone increased significantly, and flavonoids and flavonols either increased or decreased. The content of nine type PR changed significantly, and they represented three subgroups. Isoprenoids accounted for the main components, mactraxanthin 3-(5Z, 8Z, 11Z, 14Z, 17Z-eicosapentaenoate) 3′-palmitoleate (tetraterpene) and (-)-5-adianene (triterpene) increased, whereas curmadione, (+)-18-hydroxy-7,16-sacculatadiene-11,12-dial, 17β,21β-epoxy-16α-ethoxyhopan-3β-ol, 3-O-acetyl-11-keto-β-boswellic acid and bisdehydro-β-carotene/tetradehydro-β-carotene decreased.

RNA Sequencing

RNA sequencing-based transcriptome profiling was performed for the FF samples. The minimum ratio of clean data to raw reads was 93.71%. A total of 92,294 single genes with an average length of 787.75 bp and an N50 value of 1,394 bp were botained, of which 20,317 were more than 1 kb in length.

Functional Annotation and Classification

The functional annotation of Forsythia unigenes was done using eight databases, including COG, GO, KO, Pfam, Wiss-Prot, eggNOG, NR. The annotation results of 7,202, 25,296, 20,193, 16,213, 19,865, 19,944, 26,079, 31,859 unigenes were obtained respectively, with a total of 32,459 items (**Figure 5**). Of the total unigenes identified, 3,810 of them had hits in all eight databases.

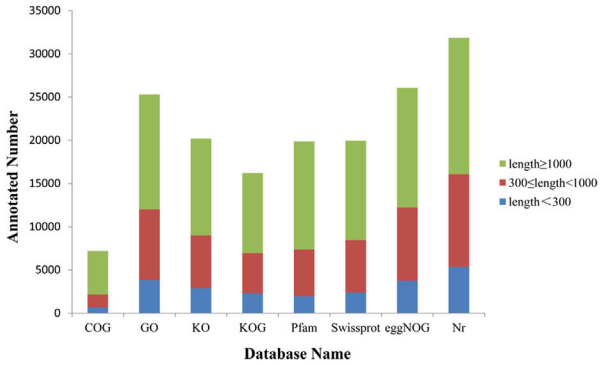


FIGURE 5 | Gene annotations in different databases. The *X*-axis is the database name. The *Y*-axis is the number of genes annotated. Unigenes length≥1000 bp represented in green, unigenes of length 300 bp to 1000 bp represented in red, and unigenes of length ≤300 bp in blue.

The sequences were assigned with GO terms based on annotation. Assignment of 25,296 unigenes fell into three GO categories, biological process, molecular function and cellular components. In the KEGG enrichment, 20,193 unigenes were assigned to 136 KEGG pathways. Plant-pathogen interaction (ID: Ko04626), plant hormone signal transduction (ID: Ko04075) and carbon metabolism (ID: Ko01200) were the top 3 pathways with the most assigned unigenes.

Clustering of Differentially Expressed Gene

The results of DEGs expression analysis based on mean FPKM showed that 1,533 DEGs were differentially expressed of which 45.27% were upregulated. The degree of difference in gene expression levels between the two groups of samples differed significantly (**Figure 6**).

In the cluster of orthologous (COG) functions, there are 497 DEGs (**Figure 7A**). Among them, 72 DEGs were related to secondary metabolite biosynthesis, transportation, decomposition and metabolism. Compared with other functions, the number of differential genes in this group was the largest. The number of differential genes related to carbohydrate transport and metabolism was 67, and the number relating to lipid transport and metabolism was 64 (**Figure 7A**). Some genes may act in multiple functions. There are three DEGs (0.8%) each involved in four functions; 13 DEGs (2.6%) each involved in three functions; 71 DEGs (14.9%) each involved in two functions and for the remainder (81.7%) each gene had a single function. Among the unigene annotated by GO, 1,289 were significantly differentially expressed, accounting for 5.10%, and the detailed classification is presented in **Figure 7B**. DEGs were enriched in the biological process module in DNA replication initiation and fatty acid biosynthetic process. In the cellular component module, DEGs were enriched in nucleus, extracellular region, host cell nucleus, among others. In the molecular function module, DEGs were enriched in DNA binding, metal ion binding, heme binding, among others.

After enriching the differentially expressed genes with the KEGG pathway, 551 differential genes were identified related to 119 pathways (**Figure 7C**). According to the ranking of the DEGs enrichment ratios, among the top 50 pathways 67.3% regulate material metabolism. The plant hormone signal transduction pathway enriched DEGs the most genes, 44 representing 8.0% of the total. The enrichment of DEGs in the phenylpropane biosynthesis pathway were second. The gene c74197. graph_c0 (0.2%) participates in the regulation of ten pathways simultaneously and the log2FC was 1.2,

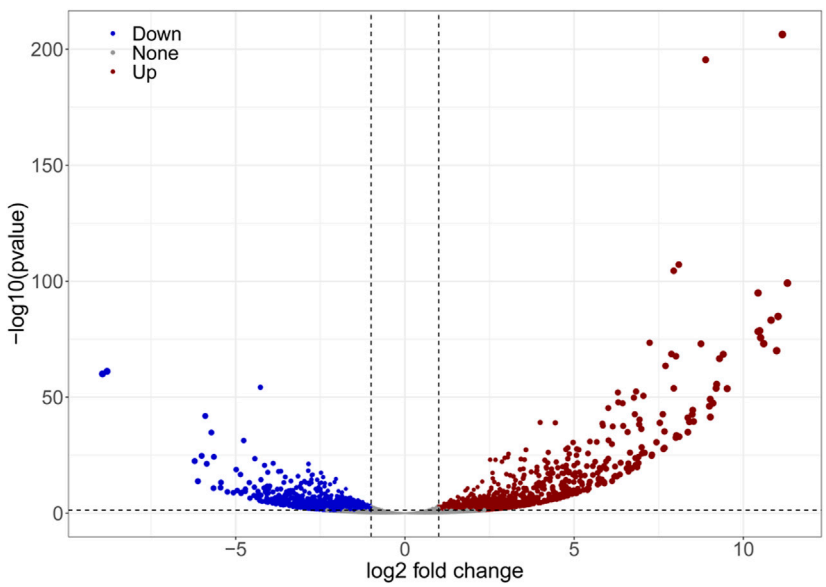


FIGURE 6 | Volcano plot showing the characteristics of differentially expressed gene. The dots represent genes, among which the blue and red dots are genes that satisfied FC < 0.5, FC > 2 and p < 0.05 and thus differed in expression between the two samples.

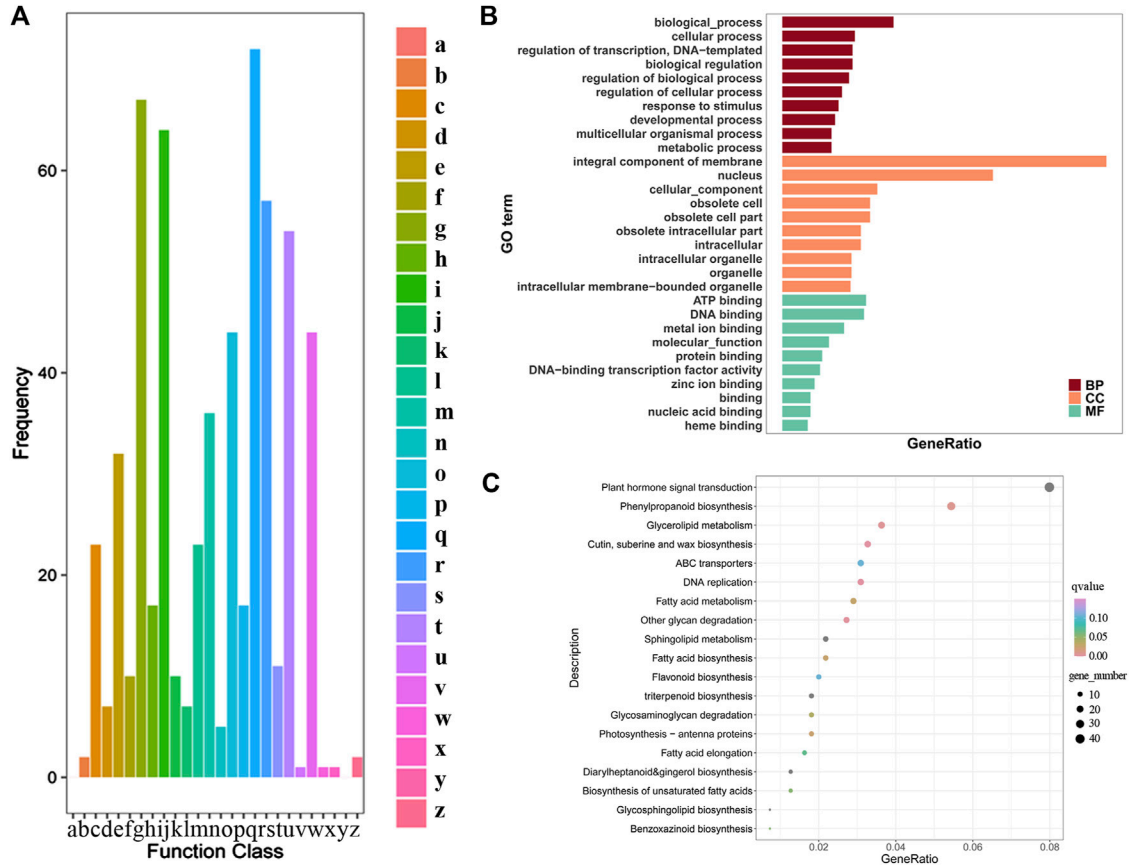


FIGURE 7 | (A) COG Function Classification of Consensus Sequence. a) RNA processing and modification; b) Chromatin structure and dynamics; c) Energy production and conversion; d) Cell cycle control, cell division, chromosome partitioning; e) Amino acid transport and metabolism; f) Nucleotide transport and metabolism; g) Carbohydrate transport and metabolism; h) Coenzyme transport and metabolism; i) Lipid transport and metabolism; j) Translation, ribosomal structure and biogenesis; k) Transcription; l) Replication, recombination and repair; m) Cell wall/membrane/envelope biogenesis; n) Cell motility; o) Posttranslational modification, protein turnover, chaperones p) Inorganic ion transport and metabolism; q) Secondary metabolite biosynthesis, transport and catabolism r) General function prediction only; s) Function unknown; t) Signal transduction mechanisms; u) Intracellular trafficking, secretion, and vesicular transport; v) Defense mechanisms; w) Extracellular structures; x) Mobilome: prophages, transposons; y) Nuclear structure; z) Cytoskeleton. Characterization of GO **(B)** and KEGG **(C)** enrichment result for transcriptome assemblies.

indicating the gene was upregulated. These include carbon metabolism (Ko01200), citrate cycle (TCA cycle) (Ko00020), glycine, serine and threonine metabolism (Ko00260), glycolysis/gluconeogenesis (Ko00010), glyoxylate and dicarboxylate metabolism (Ko00630), lysine degradation (Ko00310), propanoate metabolism (Ko00640), pyruvate metabolism (Ko00620), tryptophan metabolism (Ko00380) and valine, leucine and isoleucine degradation (Ko00280). The gene c77282. graph_c0 (0.2%) is involved in the regulation of seven pathways simultaneously and the log2FC was 8.0, indicating the gene was upregulated. Both of these two DEGs are involved in the regulation of primary metabolism. Other upregulated genes participated in more than one pathway simultaneously, including seven DEGs (1.3%) with each participating in six pathways, 14 DEGs (2.5%) with each participating in 5 pathways; 3.6% of the total DEGs participating in four pathways, 9.1% of the DEGs participating in 3 pathways, 10.2% of the DEGs participating in 2 pathways, and the remainder (72.9%) participating in a single pathway.

Prediction of Candidate Genes Involved in Material Metabolism

In the KO database annotation, there were 36 differential genes related to terpenoid metabolism (**Figures 8A**, **9A**), including the terpenoid skeleton biosynthesis pathway (Ko00900), monoterpenes (Ko00902), sesquiterpenes and triterpenes (Ko00909), diterpenes (Ko00904), and other terpenes (Ko00130). There were five pathways in total. Isoprene is produced by the mevalonate (MVA) and methyl erythritol phosphate (MEP) pathways in the terpene backbone biosynthesis pathway (Ko00900), and 9 DEGs were involved in regulation of these. The MVA pathway starts from 2 acetyl CoA which is catalyzed by AACT and HMGS to obtain 3-hydroxy-3-methylglutaryl CoA. This substance is catalyzed by 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR, EC: 1.1.1.34) to

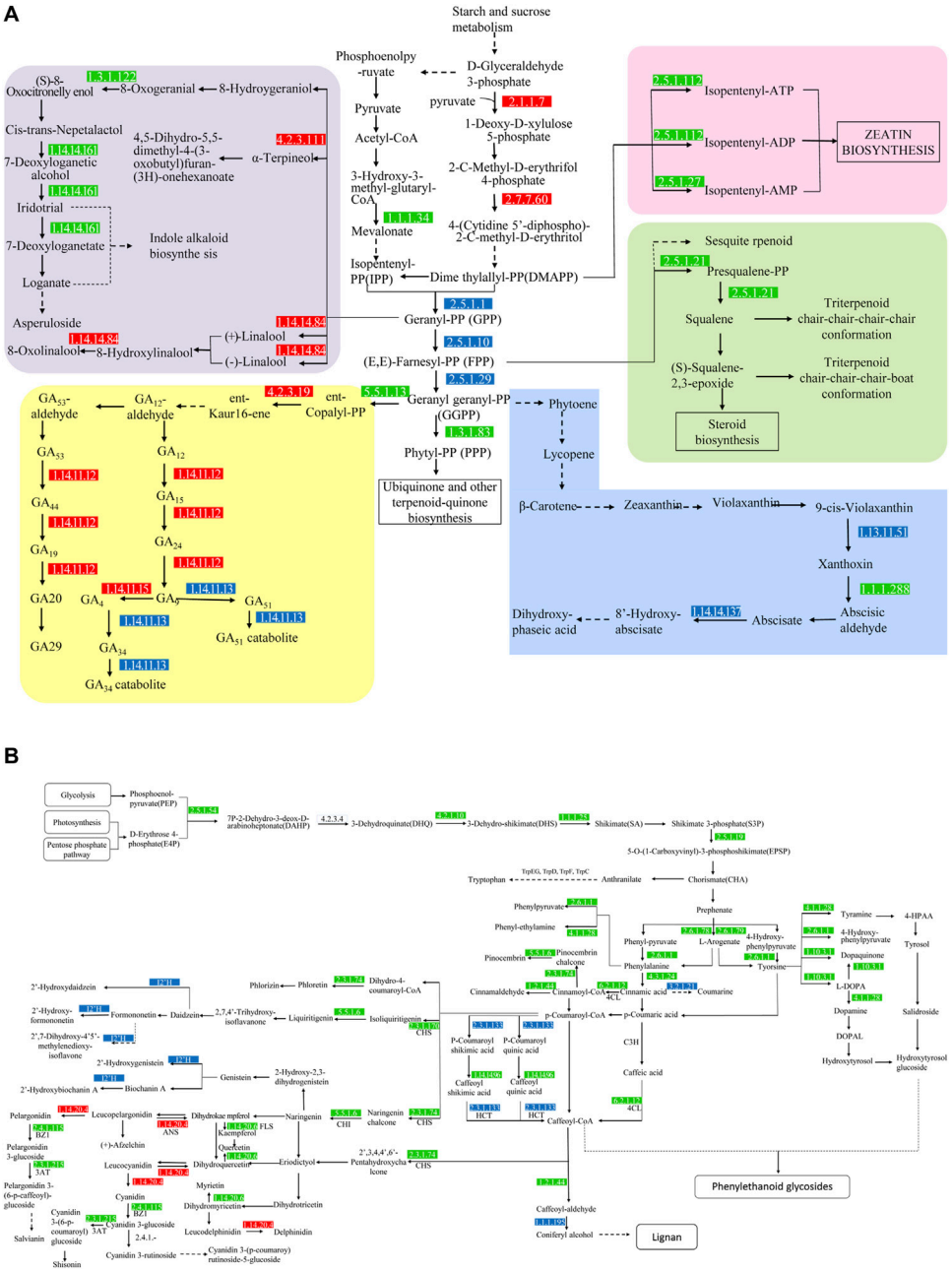


FIGURE 8 | The biosynthetic pathways and genes in the terpenoid **(A)** and flavonoid **(B)** secondary metabolite pathways. The numbers shown in the boxes are the genes/genomes that are enriched in the KEGG pathway that regulate response. The red box indicates that the gene is upregulated, the green box indicates that the gene is downregulated, and the blue box indicates that the genome has up-regulated genes and downregulated genes. The solid line represents a single-step synthesis reaction and the dashed line represents a multi-step synthesis reaction.

produce MVA, which undergoes phosphorylation and decarboxylation to produce Isopentenyl-PP (IPP). HMGR is upregulated during plant growth and promotes the formation of terpenoid skeletons through the MVA pathway. The MEP pathway, which starts with a condensation reaction between pyruvate and glyceraldehyde-3-phosphate followed by a condensation reaction catalyzed by 1-deosy-d-xylulose-5-phosphate synthase (EC:2.2.1.7), generates 1-deoxy-d-xylulose-5-phosphate, 2-C-methyl-d-erythritol 4-phosphate in the presence of 2-c-methyl-d-erythritol 4-phospphate cytidylytransferase (EC:2.7.7.60) produces 4-(Cytidine 5′-diphospho)-2-C-methyl-d-erythritol. Both genes were upregulated. Dime thylalllyl-PP (DMAPP) was generated after a four-step transformation. IPP and DMAPP are isomers, which are in equilibrium and are the source of isoprene units. DMAPP and IPP are catalyzed by EC:2.5.1.1 to generate geranyl-PP (GPP). GPP combines with an IPP to generate (E, E)-famesyl-PP

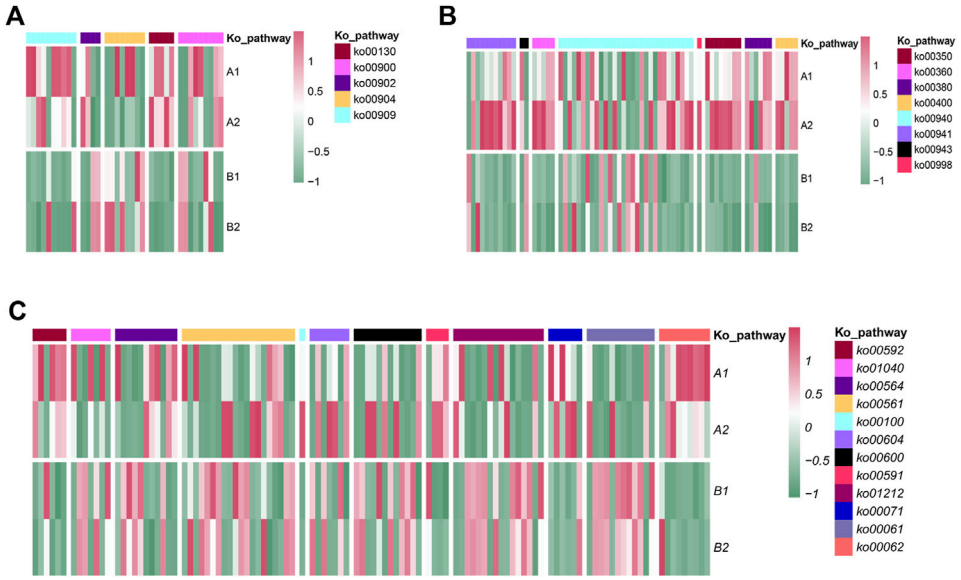


FIGURE 9 | Transcript expression analyses and the biosynthetic pathways of terpenoid metabolites **(A)**, active metabolites **(B)** and lipids **(C)**.

(FPP) under the catalysis of EC: 2.5.1.10. FPP will form geranyl geranyl-PP (GGPP) under the catalysis of EC: 2.5.1.29 with another IPP. IPP, DMAPP, FPP, GPP, GGPP, and other substances are the starting materials to produce terpenoids with different structures and are the backbone of terpenoids. GPP is the precursor of monoterpenes and enters the monoterpenoid biosynthetic pathway (Ko00902). FPP is the precursor of sesquiterpenes and triterpenes and enters the sesquiterpene and triterpenoid biosynthesis pathway (Ko00909). GGPP is a diterpene. The precursors of tetraterpene and tetraterpene enter the diterpene biosynthesis pathway (Ko00904) and the carotenoid biosynthesis pathway (Ko00906), respectively. Gibberellins (GAs) are diterpenoids, and there are five genes involved in their synthesis in the Ko00904 pathway. GGPP is regulated by the gene ent-copayl diphosphate synthase (CPS, EC: 5.5.1.13) in the plastid to generate ent-copalyl-PP, and its expression was downregulated and then regulated by the gene ent-kaurene synthase (KS, EC: 4.2.3.19) to generate ent-Kaur-16-ene. KS expression was upregulated. Ent-Kaur-16-ene undergoes a multi-step enzymatic reaction in the endoplasmic reticulum to generate GA12-aldehyde. GA12-aldehyde is regulated by different genes to generate GAs. GA_{12} produces GA_{24} and GA_9 under the regulation of EC: 1.14.11.12, GA_9 produces GA_4 under the regulation of EC: 1.14.11.15, and GA_4 is regulated by the gene EC: 1.14.11.13 to synthesize GA_{34}. Among the GAs on FF, the content of GA_{36} was the highest, and it is metabolized by two pathways, GA_{24} and GA_{37}, and can generate GA_4 under the action of P450 enzymes. GA_8 and GA_8-catabolite are metabolites of GA_1, and EC: 1.14.11.13 are genes that regulate the metabolism of active GAs. In the Ko00906 pathway, GGPP generates carotene, lutein, and other tetraterpenoids through a series of complex regulatory reactions. Carotene and lutein can absorb light energy and participate in photosynthesis and abscisic acid is a signal of seed maturation and stress resistance. Hormones are synthesized via carotenoid metabolism. Differential genes EC:1.14.14.137, EC:1.13.11.51 and EC: 1.1.1.288 are involved in the regulation of abscisic acid metabolism (**Figure 8**).

Flavonoid synthesis of secondary metabolites (**Figure 8B**) begins with d-erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP). E4P comes from the pentose phosphate pathway and PEP comes from the glycolysis pathway. The two undergo a four-step enzymatic reaction to generate shikimate acid (SA) which is catalyzed by shikimate kinase (AroK/AroL), 5-enolpyruvate shikimate-3-phosphate synthase (AroA), and chorismate synthase (AroC) to generate chorismate (CHA). CHA is a branch point, l-phenylalanine (L-Phe), l-tyrosine (L-Tyr), l-tryptophan (l-tryptophan, L-Trp) of the aromatic amino acids participate in different branch pathways. PheA catalyzes the production of phenylpyruvate, which is catalyzed by amino acid transaminase to obtain L-Phe. TyrA catalyzes the conversion of pre-benzoic acid to 4-hydroxyphenylpyruvate, which is then catalyzed by transaminase to obtain L-Tyr. The synthesis pathway of tryptophan is more complicated, starting with chorismate, and generating L-Trp under the catalysis of TrpEG, TrpD, TrpF, TrpC, TrpA and TrpB. l-phenylalanine is produced by the action of phenylalanine lyase, hydroxylase, and CoA ligase to produce 4-coumayl-CoA, which is produced by chalcone synthase (CHI) and flavanone 3-hydroxylase (F3H) to form dihydroflavonoids. Dihydroflavonols are important precursors of other flavonoids. They are catalyzed by flavonol synthase (FLS), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS). The flavonols, anthocyanins, and flavonoids are produced separately. These three substances undergo methylation, glycosylation and acylation modification to form flavonoids with

diverse structures. Five DEGs annotated with the phenylalanine acid metabolism pathway (Ko00360) were all down-regulated (**Figure 9B**). Phenylalanine ammonia-lyase (PAL, EC: 4.3.1.24), functional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase (PAT/AAT, EC: 2.6.1.1) and aromatic-l-amino-acid/L-tryptophan decarboxylase (EC: 4.1.1.28) inhibit the conversion and metabolism of phenylalanine, respectively, and inhibit its conversion to cinnamate, phenylpyruvate and phenethylamine. There are eight genes associated with enrichment in the tyrosine metabolism pathway (Ko00350). These genes (EC: 4.1.1.28, EC: 2.6.1.1, EC: 1.10.3.1 and EC: 4.1.1.28) inhibit the conversion of tyrosine to tyramine, hydroxyphenylpropionate, dopamine, and dopaquinone. The flavonoid biosynthesis pathway (Ko00941) is related to 11 genes, except for the upregulation of anthocyanidin synthase (ANS, EC: 1.14.20.4), chalcone synthase (CHS, EC: 2.3.1.74, 2.3.1.170), shikimate O-hydroxycinnamoyl transferase (HCT, EC: 2.3.1.133), chalcone isomerase (EC: 5.5.1.6), and FLS (EC: 1.14.20.6) were all downregulated to varying degrees. In the anthocyanin synthesis pathway (Ko00942), anthocyanidin 3-O-glucosyltransferase (BZ1, EC: 2.4.1.115) and anthocyanidin 3-O-glucoside 6″-O-acyltransferase (3AT, EC: 2.3.1.215) inhibit geranium glycosylation and acylation of vitamins, anthocyanins and delphiniums. In the flavonoid synthesis pathway (Ko00943), CYP81E (EC: I2′H) regulates the hydroxylation of daidzein, formononetin, pseudobaptigenin and genistein. The gene c82894. graph_c0 was upregulated, and the gene c72877. graph_c0 was downregulated.

DISCUSSION

In the present study, the lipid composition of FF was systematically determined using liquid-phase mass spectrometry. For the first time, it was determined that the FF contains 5,802 lipids and that the content of lipid components with pharmacological activity was relatively low. Comparing the lipid changes of FF during the two growth periods in May and July, the content of 139 lipid types changed significantly. Different lipid secondary metabolites are mainly concentrated in pregnenolones and polyvinyls, among which the content of chalcone, dihydrolone, flavanone, flavonol and mactraxanthin 3-(5Z,8Z,11Z,14Z, 17Z-eicosapentaenoate) 3′-palmitoleate increased, while the content of afzelechin 7-O-β-d-apiofuranoside, 4β-(2,4-dihydroxy-3-methoxyphenyl) fisetinidol, and others decreased. Transcriptome analysis of FF detected 92,294 unigenes and 1,533 DEGs. There were 551 DEGs in the KO annotation, which were enriched in 119 KEGG pathways. Plant signal transduction, interaction with pathogens, and phenylpropane biosynthesis pathways were the most enriched among the DEGs. Based on these results, the present study analyzed the lipid metabolism of FF to explore the regulatory mechanism of terpenoids and flavonoids during the growth process.

FF contains monoterpenoids, sesquiterpenoids, diterpenoids, triterpenoids and tetraterpenoids. Dong et al. (2017) found that the average content of volatile oil in FF was as high as 3.27%, containing α-pinene, β-pinene, limonene, phellandrene, cymene, linalool and terpine-4-ol (Sun et al., 2018; Yang et al., 2014). The results of the present study show that FF contains rosin alcohol, fumagillin, hydroxy camphor, and other substances, as well as plant hormones such as abscisic acid gibberellin. The terpenoid structural unit is transformed from the MVA and MEP pathways to generate IPP and its isomer DMAPP. The present study found that the synthesis of the terpenoid skeleton of FF is dominated by the MEP pathway and the pathway is regulated by genes EC: 2.2.1.7 and EC: 2.7.7.60. Plant hormones are closely related to the growth process of plants. ABA, also known as dormant or abscisin, causes dormancy of buds, shedding of leaves and inhibition of cell growth. Mass spectrometry detected no significant difference in the content of 7′-hydroxyabscisic acid and (+)-8′-hydroxyabscisic acid. Genes EC: 1.1.1.288, EC: 1.14.14.137 and EC: 1.13.11.51 jointly regulate the metabolism of abscisic acid, which may be related to the growth and development stage of FF when harvested. Abscisic acid is balanced under the dynamic regulation of the three genes state. ABA is synthesized by carotene, and the content of carotene decreases significantly, possibly due to the production of abscisic acid. Gibberellin is a diterpenoid compound with diverse structures, some of which are easily soluble in water and only a few substances have intrinsic activity controlling all aspects of plant growth and development (including seed germination, stem elongation, leaf expansion, flowering and seed development) (Hedden, 2020). GA_1, GA_3, GA_4 and GA_7 have biological activity. The 6 GAs identified were non-biologically active substances. GA_{36} had the most abundant content and is the precursor of active gibberellin GA_4, which may be transformed into an active state under the control of GA_{20}-oxidase (GA20ox, EC: 1.14.11.12) and GA_3-oxidase (GA3ox, EC: 1.14.11.15). At the same time, GA_8 and its metabolites, which are the metabolites of GA_1, may be metabolic genes in (EC: 1.14.11.13), the gene c83901. graph_c3 was upregulated to promote the catabolism of active GAs and c63052. graph_c0 was downregulated to inhibit the decomposition of active GAs. In short, gibberellin has a high content in July in Qingqiao, and its metabolism was active, which fits with it being an important substance for fruit development.

Flavonoids play are important for attracting insect pollinators, resisting ultraviolet rays, and enhancing disease resistance (Onkokesung et al., 2014) in plants (Agati et al., 2012). Flavonoids have various biological activities (Ullah et al., 2020) such as antibacterial, antiviral, antitumor, anti-inflammatory, and hypoglycemic activity (Cazarolli et al., 2006). Studies (Hoshino et al., 1999) have reported that the antibacterial effects of certain flavonoid complexes are related to their oxidative division of DNA or damage to bacterial cell membranes. Wang et al. (2004) found that baicalin inhibited HIV-1 induced syncytium formation and HIV-1 p24 antigen and RT product produced antiviral effects. Pereira et al. (2007) found that naringin can significantly reduce the nitrite content and the amount of inflammatory cells in the macrophage inflammation model (Yang et al., 2015; Nishioka et al., 1998). Flavonoids are based on the C6-C3-C6 structure, and include anthocyanins, flavonoids, flavonols and isoflavones. Pinocembrin, naringenin and eriodictyol are the precursors of other flavonoid derivatives. There are some key enzymes in the *de novo* synthesis of flavonoids. ACCase (acetyl CoA carboxylase) has a wide range of functions, which can guide the flow of carbon from photosynthesis

to primary and secondary metabolites. Among these enzymes, MF (multifunctional) guides the biosynthesis of very long-chain fatty acids and flavonoids in the cytoplasm. PAL is the entrance enzyme of phenylpropanoid biosynthesis (Sun et al., 2020). Hu et al. (2011) used a competitive inhibitor 2-aminoindan-2-phosphonic acid (AIP) to specifically inhibit PAL activity in *Cistanche* suspension cells and found that the content of phenylethanol glycosides was significantly reduced. In the phenylpropane pathway, CHS is a key enzyme in the synthesis of flavonoids, leading the central pathway of phenylpropane metabolism to the branch of flavonoid synthesis. The results of the present study showed that when the expression of anthocyanidin synthase (ANS, EC: 1.14.20.4) was upregulated, afzelechin 7-O-β-d-apiofuranoside was significantly reduced. This result may be due to the conversion of afzelechin from leucopelargonidin, and ANS can promote the metabolism of leucopelargonidin into dihydrokaempferol and pelargondin, which indirectly leads to a decrease in the synthesis of afzelechin. Perhaps ANS is an inhibitory enzyme of afzelechin synthesis. In addition, the content of quercetin 7,3′,4′-trimethyl ether 3-rhamnoside (quercetin 7,3′,4′-trimethyl ether 3-rhamnoside) increased significantly, and there were two influencing factors. On the one hand, the upregulation of ANS expression promotes the conversion of leucocyanidin to dihydroquercetin, which is the precursor of quercetin 7,3′,4′-trimethyl ether 3-rhamnoside. On the other hand, the downregulation of flavonol synthase (FLS, EC: 1.14.20.6) (Park et al., 2019) inhibits the oxidation reaction of dihydrokaempferol, dihydroquercetin and dihydromyricetin. FLS is the key enzyme for the conversion of dihydrokaempferol, dihydroquercetin, and dihydromyricetin to pelargonidin, quercetin, and myricetin. The significant reduction of myricetin 7-glucoside again confirms the effect of FLS. Park et al. isolated the OsFLS gene from rice and showed that the OsFLS gene has both flavonol synthase and flavanone 3-hydroxylase activity, which is consistent with the results of the present study.

There are also pharmacologically active substances in FF, such as phenethanol glycosides and lignans, which originate from the shikimic acid pathway and are homologous to flavonoids (Sun et al., 2018). The results showed that the phenylalanine, tryptophan, and tyrosine synthesis pathway (Ko00400), phenylalanine metabolism pathway (Ko00360), the tyrosine synthesis pathway (Ko00350) and the phenylpropane synthesis pathway (Ko00940), and the activity of anthocyanin synthesis pathway (Ko00942), both peaked and declined. The downregulation of these pathways is closely related to the synthesis of active secondary metabolites such as phenethyl alcohol glycosides, lignin, and flavonoids indicating that the accumulation of secondary metabolites reaches the peak during the maturation period of FF in July, when the crop is ready to harvest. The chemical composition and pharmacological effects of Laoqiao and Qingqiao harvested 1 month later differ greatly. Cui et al. (2010) found that the different maturity of FF determines the concentration of chemical substances. Jia et al. (2015) found that forsythin A, forsythin C, cornoside, rutin, forsythin and the gallic acid content is higher than Laoqiao. Bao et al. (2017) found that the anti-tumor activity of Laoqiao on B16-F10 mouse melanoma was significantly weaker than that of Qingqiao. Over time the content of secondary metabolites in FF will gradually decrease and the related pharmacological effects will gradually weaken. During harvesting and processing of FF, a special treatment called "shaqing" is required before Qingqiao is used as a medicine. Shaqing, via steaming, boiling and other processes can destroy the metabolic enzymes (Wu et al., 2019) in FF, so that the active substances are retained. Laoqiao is used as medicine only when dried, which is one of the reasons why Laoqiao and Qingqiao have different active ingredients and effects. The best period for harvesting medicinal plants should be the peak period of synthesis and accumulation of secondary metabolites. Traditional Chinese medicine classics and the Chinese Pharmacopoeia record July to August as the harvesting period of Qingqiao. The transcriptome results of FF in July showed that the DEGs of secondary metabolites biosynthesis, transportation, decomposition and metabolism are the most active at that time. This period is the peak of secondary metabolite accumulation, which proves that the harvesting of Qingqiao from July to August is rational for maximum medicinal benefit.

CONCLUSION

In the present study, ultra-performance liquid chromatography-mass spectrometry technology was used to compare the lipid composition and content differences of FF at different growth stages. The correlation between lipid changes in secondary metabolites and differential gene expression was analyzed. Preliminary analysis revealed the biosynthetic mechanism of the main secondary metabolites of FF, indicating that the terpenoid abscisic acid may be in the dynamic regulation of genes EC: 1.1.1.288, EC: 1.14.14.137 and EC: 1.13.11.51. In the synthesis of gibberellin, GA20-oxidase (GA20ox, EC: 1.14.11.12) and GA3-oxidase (GA3ox, EC: 1.14.11.15) catalyze the production of active GAs, and EC: 1.14.11.13 is the metabolic enzyme of active GAs. The synthesis of flavonoids in FF, MF, PAL, CHS, ANS, FLS, and others are all key enzymes. These results provide references for the analysis, development and utilization of the medicinal components of *F. suspense*. In addition, the results of the present study showed that July was the peak period for the accumulation of secondary metabolites in *Forsythia suspensa*, which provids new scientific evidence for the basis of traditional FF harvesting and processing methods.

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: Mass spectrometry data has been uploaded to EBI Metabolights, accession number MTBLS3327. Sequencing data has been uploaded to NCBI SRA BioProject, accession number PRJNA756578.

AUTHOR CONTRIBUTIONS

BW, YL and WZ designed the experiment. BW, CW, WJ and ZM performed the analysis and interpretation of the data. CW, WJ and ZM drafted the manuscript. BW, YL and WZ revised the

manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We gratefully acknowledge help from Professor Zhong Wang and Jinju Duan for improving and the English editor for editing our manuscript.

REFERENCES

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as Antioxidants in Plants: Location and Functional Significance. *Plant Sci.* 196, 67–76. doi:10.1016/j.plantsci.2012.07.014

Bao, J., Ding, R.-B., Liang, Y., Liu, F., Wang, K., Jia, X., et al. (2017). Differences in Chemical Component and Anticancer Activity of green and Ripe Forsythiae Fructus. *Am. J. Chin. Med.* 45 (07), 1513–1536. doi:10.1142/S0192415X17500823

ben Sghaier, M., Pagano, A., Mousslim, M., Ammari, Y., Kovacic, H., and Luis, J. (2016). Rutin Inhibits Proliferation, Attenuates Superoxide Production and Decreases Adhesion and Migration of Human Cancerous Cells. *Biomed. Pharmacother.* 84, 1972–1978. doi:10.1016/j.biopha.2016.11.001

Cazarolli, L. H., Zanatta, L., Jorge, A. P., de Sousa, E., Horst, H., Woehl, V. M., et al. (2006). Follow-up Studies on Glycosylated Flavonoids and Their Complexes with Vanadium: Their Anti-hyperglycemic Potential Role in Diabetes. *Chemico-Biological Interactions* 163 (3), 177–191. doi:10.1016/j.cbi.2006.07.010

Chapman, K., and Feussner, I. (2016). Plant Lipid Biology. *Biochim. Biophys. Acta (Bba) - Mol. Cel. Biol. Lipids* 1861 (9), 1205–1206. doi:10.1016/j.bbalip.2016.05.005

Chen, J.-J., Liang, X., Wang, F., Wen, Y.-H., Chen, T.-J., Liu, W.-C., et al. (2019). Combinatorial Mutation on the β-glycosidase Specific to 7-β-Xylosyltaxanes and Increasing the Mutated Enzyme Production by Engineering the Recombinant Yeast. *Acta Pharmaceutica Sinica B* 9 (3), 626–638. doi:10.1016/j.apsb.2018.11.003

Chua, L. S. (2013). A Review on Plant-Based Rutin Extraction Methods and its Pharmacological activities Research Support, Non-U.S. Gov't; Review]. *J. Ethnopharmacology* 150 (3), 805–817. doi:10.1016/j.jep.2013.10.036

Cui, Y., Wang, Q., Shi, X., Zhang, X., Sheng, X., and Zhang, L. (2010). Simultaneous Quantification of 14 Bioactive Constituents in Forsythia Suspensa by Liquid Chromatography-Electrospray Ionisation-Mass Spectrometry. *Phytochem. Anal.* 21 (3), 253–260. doi:10.1002/pca.1194

Dong, Z., Lu, X., Tong, X., Dong, Y., Tang, L., and Liu, M. (2017). Forsythiae Fructus: A Review on its Phytochemistry, Quality Control, Pharmacology and Pharmacokinetics. *Molecules* 22 (9), 1466. doi:10.3390/molecules22091466

Du, Y., You, L., Ni, B., Sai, N., Wang, W., Sun, M., et al. (2020). Phillyrin Mitigates Apoptosis and Oxidative Stress in Hydrogen Peroxide-Treated RPE Cells through Activation of the Nrf2 Signaling Pathway. *Oxid. Med. Cel. Longev.* 2020, 1–16. doi:10.1155/2020/2684672

Gong, L., Wang, C., Zhou, H., Ma, C., Zhang, Y., Peng, C., et al. (2021). A Review of Pharmacological and Pharmacokinetic Properties of Forsythiaside A. *Pharmacol. Res.* 169, 105690. doi:10.1016/j.phrs.2021.105690

Guo, H., Liu, A.-H., Li, L., and Guo, D.-A. (2007). Simultaneous Determination of 12 Major Constituents in Forsythia Suspensa by High Performance Liquid Chromatography-DAD Method. *J. Pharm. Biomed. Anal.* 43 (3), 1000–1006. doi:10.1016/j.jpba.2006.09.033

Hedden, P. (2020). The Current Status of Research on Gibberellin Biosynthesis. *Plant Cel. Physiol.* 61 (11), 1832–1849. doi:10.1093/pcp/pcaa092

Hoshino, N., Kimura, T., Yamaji, A., and Ando, T. (1999). Damage to the Cytoplasmic Membrane of *Escherichia coli* by Catechin-Copper (II) Complexes. *Free Radic. Biol. Med.* 27 (11-12), 1245–1250. doi:10.1016/s0891-5849(99)00157-4

Hu, G. S., Hur, Y. J., Jia, J. M., Lee, J. H., Chung, Y. S., Yi, Y. B., et al. (2011). Effects of 2-Aminoindan-2-Phosphonic Acid Treatment on the Accumulation of Salidroside and Four Phenylethanoid Glycosides in Suspension Cell Culture of Cistanche Deserticola. *Plant Cel. Rep.* 30 (4), 665–674. doi:10.1007/s00299-010-0997-3

Hu, K., Guan, W.-j., Bi, Y., Zhang, W., Li, L., Zhang, B., et al. (2021). Efficacy and Safety of Lianhuaqingwen Capsules, a Repurposed Chinese Herb, in Patients with Coronavirus Disease 2019: A Multicenter, Prospective, Randomized Controlled Trial. *Phytomedicine* 85, 153242. doi:10.1016/j.phymed.2020.153242

Jia, J., Zhang, F., Li, Z., Qin, X., and Zhang, L. (2015). Comparison of Fruits of Forsythia Suspensa at Two Different Maturation Stages by NMR-Based Metabolomics. *Molecules* 20 (6), 10065–10081. doi:10.3390/molecules200610065

Jiang, F., Gong, T., Chen, J., Chen, T., Yang, J., and Zhu, P. (2021). [Synthetic Biology of Plants-Derived Medicinal Natural Products]. *Sheng Wu Gong Cheng Xue Bao* 37 (6), 1931–1951. doi:10.13345/j.cjb.210138

Jiao, J., Fu, Y.-J., Zu, Y.-G., Luo, M., Wang, W., Zhang, L., et al. (2012). Enzyme-assisted Microwave Hydro-Distillation Essential Oil from Fructus Forsythia, Chemical Constituents, and its Antimicrobial and Antioxidant Activities. *Food Chem.* 134 (1), 235–243. doi:10.1016/j.foodchem.2012.02.114

Jiao, J., Gai, Q.-Y., Wei, F.-Y., Luo, M., Wang, W., Fu, Y.-J., et al. (2013). Biodiesel from Forsythia Suspense [(Thunb.) Vahl (Oleaceae)] Seed Oil. *Bioresour. Technol.* 143, 653–656. doi:10.1016/j.biortech.2013.06.018

Lee, H.-w., and Lee, H.-S. (2015). Acaricidal Abilities and Chemical Composition of Forsythia Suspense Fruit Oil against Storage and Pyroglyphid Mites. *Jabc* 58 (2), 105–108. doi:10.3839/jabc.2015.018

Liu, M., Gao, Y., Yuan, Y., Yang, K., Shi, S., Tian, J., et al. (2021). Efficacy and Safety of Herbal Medicine (Lianhuaqingwen) for Treating COVID-19: A Systematic Review and Meta-Analysis. *Integr. Med. Res.* 10 (1), 100644. doi:10.1016/j.imr.2020.100644

Ma, Q., Li, R., Pan, W., Huang, W., Liu, B., Xie, Y., et al. (2020). Phillyrin (KD-1) Exerts Anti-viral and Anti-inflammatory Activities against Novel Coronavirus (SARS-CoV-2) and Human Coronavirus 229E (HCoV-229E) by Suppressing the Nuclear Factor Kappa B (NF-Kb) Signaling Pathway. *[Journal Article] Phytomedicine* 78, 153296. doi:10.1016/j.phymed.2020.153296

Nishioka, T., Kawabata, J., and Aoyama, Y. (1998). Baicalein, an α-Glucosidase Inhibitor from Scutellaria Baicalensis. *J. Nat. Prod.* 61 (11), 1413–1415. doi:10.1021/np980163p

Onkokesung, N., Reichelt, M., van Doorn, A., Schuurink, R. C., van Loon, J. J. A., and Dicke, M. (2014). Modulation of Flavonoid Metabolites in *Arabidopsis thaliana* through Overexpression of the MYB75 Transcription Factor: Role of Kaempferol-3,7-Dirhamnoside in Resistance to the Specialist Insect Herbivore Pieris Brassicae. *J. Exp. Bot.* 65 (8), 2203–2217. doi:10.1093/jxb/eru096

Park, S., Kim, D.-H., Park, B.-R., Lee, J.-Y., and Lim, S.-H. (2019). Molecular and Functional Characterization of Oryza Sativa Flavonol Synthase (OsFLS), a Bifunctional Dioxygenase. *J. Agric. Food Chem.* 67 (26), 7399–7409. doi:10.1021/acs.jafc.9b02142

Pereira, R., Andrades, N., Paulino, N., Sawaya, A., Eberlin, M., and Marcucci, M. (2007). Synthesis and Characterization of a Metal Complex Containing Naringin and Cu, and its Antioxidant, Antimicrobial, Antiinflammatory and Tumor Cell Cytotoxicity. *Molecules* 12 (7), 1352–1366. doi:10.3390/12071352

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., and Ndungu, J. M. (2006). Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast. *Nature* 440 (7086), 940–943. doi:10.1038/nature04640

Srivastava, S., and Srivastava, A. K. (2007). Hairy Root Culture for Mass-Production of High-Value Secondary Metabolites. *Crit. Rev. Biotechnol.* 27 (1), 29–43. doi:10.1080/07388550601173918

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.758326/full#supplementary-material

Supplementary Figure S1 | Statistical table of gene expression correlation (square of Pearson's correlation coefficient) of pairwise samples.

Supplementary Figure S2 | Characteristics of differential lipid molecules with OPLA-DA and heatmap.

Stanely Mainzen Prince, P., and Priya, S. (2010). Preventive Effects of Rutin on Lysosomal Enzymes in Isoproterenol Induced Cardio Toxic Rats: Biochemical, Histological and *In Vitro* Evidences. *Eur. J. Pharmacol.* 649 (1-3), 229–235. doi:10.1016/j.ejphar.2010.08.054

Sun, L., Rai, A., Rai, M., Nakamura, M., Kawano, N., Yoshimatsu, K., et al. (2018). Comparative Transcriptome Analyses of Three Medicinal Forsythia Species and Prediction of Candidate Genes Involved in Secondary Metabolisms. *J. Nat. Med.* 72 (4), 867–881. doi:10.1007/s11418-018-1218-6

Sun, X., Li, L., Pei, J., Liu, C., and Huang, L.-F. (2020). Metabolome and Transcriptome Profiling Reveals Quality Variation and Underlying Regulation of Three Ecotypes for Cistanche Deserticola. *Plant Mol. Biol.* 102 (3), 253–269. doi:10.1007/s11103-019-00944-5

Thodey, K., Galanie, S., and Smolke, C. D. (2014). A Microbial Biomanufacturing Platform for Natural and Semisynthetic Opioids. *Nat. Chem. Biol.* 10 (10), 837–844. doi:10.1038/nchembio.1613

Ullah, A., Munir, S., Badshah, S. L., Khan, N., Ghani, L., Poulson, B. G., et al. (2020). Important Flavonoids and Their Role as a Therapeutic Agent. *Molecules* 25 (22), 5243. doi:10.3390/molecules25225243

Wang, Q., Wang, Y.-T., Pu, S.-P., and Zheng, Y.-T. (2004). Zinc Coupling Potentiates Anti-HIV-1 Activity of Baicalin. *Biochem. Biophysical Res. Commun.* 324 (2), 605–610. doi:10.1016/j.bbrc.2004.09.093

Wang, Z., Xia, Q., Liu, X., Liu, W., Huang, W., Mei, X., et al. (2018). Phytochemistry, Pharmacology, Quality Control and Future Research of Forsythia Suspensa (Thunb.) Vahl: A Review. *J. Ethnopharmacology* 210, 318–339. doi:10.1016/j.jep.2017.08.040

Wu, M. H., Shi, S. M., and Cao, H. (2019). [Herbalogical Study on merit Rating of Forsythiae Fructus Based on Near-Mature Fruit and Hyper-Mature Fruit]. *Zhongguo Zhong Yao Za Zhi* 44 (24), 5508–5512. doi:10.19540/j.cnki.cjcmm.20191104.103

Xu, T., Hu, C., Xuan, Q., and Xu, G. (2020). Recent Advances in Analytical Strategies for Mass Spectrometry-Based Lipidomics. *Analytica Chim. Acta* 1137, 156–169. doi:10.1016/j.aca.2020.09.060

Xu, X., Saadeldeen, F. S. A., Xu, L., Zhao, Y., Wei, J., Wang, H.-M. D., et al. (2019). The Mechanism of Phillyrin from the Leaves of Forsythia Suspensa for Improving Insulin Resistance. *Biomed. Res. Int.* 2019, 1–7. doi:10.1155/2019/3176483

Yadav, S., Singh, M., Singh, S. N., and Kumar, B. (2020). Tanshinone IIA Pretreatment Promotes Cell Survival in Human Lung Epithelial Cells under Hypoxia via AP-1-Nrf2 Transcription Factor. *Cell Stress and Chaperones* 25 (3), 427–440. doi:10.1007/s12192-020-01083-3

Yang, J.-R., Luo, J.-G., and Kong, L.-Y. (2015). Determination of α-glucosidase Inhibitors from ScutScutellaria Baicalensis Using Liquid Chromatography with Quadrupole Time of Flight Tandem Mass Spectrometry Coupled with Centrifugal ultrafiltrationResearch Support, Non-U.S. Gov't]. *Chin. J. Nat. Med.* 13 (3), 208–214. doi:10.1016/S1875-5364(15)30006-6

Yang, J., Wei, H., Teng, X., Zhang, H., and Shi, Y. (2014). Dynamic Ultrasonic Nebulisation Extraction Coupled with Headspace Ionic Liquid-Based Single-Drop Microextraction for the Analysis of the Essential Oil in Forsythia Suspensa. *Phytochem. Anal.* 25 (2), 178–184. doi:10.1002/pca.2490

Zhou, W., Di, L.-q., Shan, J.-j., Bi, X.-l., Chen, L.-t., and Wang, L.-c. (2011). Intestinal Absorption of Forsythoside A in Different Compositions of Shuang-Huang-Lian. *Fitoterapia* 82 (3), 375–382. doi:10.1016/j.fitote.2010.11.012

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.

Copyright © 2021 Wu, Li, Zhao, Meng, Ji and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fgene.2021.743758

Unraveling the Genetic Architecture of Two Complex, Stomata-Related Drought-Responsive Traits by High-Throughput Physiological Phenotyping and GWAS in Cowpea (*Vigna. Unguiculata* L. Walp)

Xinyi Wu[1], Ting Sun[2], Wenzhao Xu[3], Yudong Sun[3], Baogen Wang[1], Ying Wang[1], Yanwei Li[1], Jian Wang[1], Xiaohua Wu[1], Zhongfu Lu[1], Pei Xu[2] and Guojing Li[1,4]**

[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]College of Life Sciences, China Jiliang University, Hangzhou, China, [3]Huaiyin Institute of Agricultural Sciences of Xuhuai Region in Jiangsu, Huaian, China, [4]State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Jiang Libo,
Beijing Forestry University, China

Reviewed by:
Miaomiao Zhang,
Beijing Forestry University, China
Ali Raza,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Pei Xu
peixu@cjlu.edu.cn
Guojing Li
ligj@zaas.ac.cn

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

Received: 19 July 2021
Accepted: 15 October 2021
Published: 28 October 2021

Citation:
Wu X, Sun T, Xu W, Sun Y, Wang B, Wang Y, Li Y, Wang J, Wu X, Lu Z, Xu P and Li G (2021) Unraveling the Genetic Architecture of Two Complex, Stomata-Related Drought-Responsive Traits by High-Throughput Physiological Phenotyping and GWAS in Cowpea (Vigna. Unguiculata L. Walp). Front. Genet. 12:743758. doi: 10.3389/fgene.2021.743758

Drought is one of the most devasting and frequent abiotic stresses in agriculture. While many morphological, biochemical and physiological indicators are being used to quantify plant drought responses, stomatal control, and hence the transpiration and photosynthesis regulation through it, is of particular importance in marking the plant capacity of balancing stress response and yield. Due to the difficulties in simultaneous, large-scale measurement of stomatal traits such as sensitivity and speed of stomatal closure under progressive soil drought, forward genetic mapping of these important behaviors has long been unavailable. The recent emerging phenomic technologies offer solutions to identify the water relations of whole plant and assay the stomatal regulation in a dynamic process at the population level. Here, we report high-throughput physiological phenotyping of water relations of 106 cowpea accessions under progressive drought stress, which, in combination of genome-wide association study (GWAS), enables genetic mapping of the complex, stomata-related drought responsive traits "critical soil water content" (θ_{cri}) and "slope of transpiration rate declining" (K_{Tr}). The 106 accessions showed large variations in θ_{cri} and K_{Tr}, indicating that they had broad spectrum of stomatal control in response to soil water deficit, which may confer them different levels of drought tolerance. Univariate GWAS identified six and fourteen significant SNPs associated with θ_{cri} and K_{Tr}, respectively. The detected SNPs distributed in nine chromosomes and accounted for 8.7–21% of the phenotypic variation, suggesting that both stomatal sensitivity to soil drought and the speed of stomatal closure to completion were controlled by multiple genes with moderate effects. Multivariate GWAS detected ten more significant SNPs in addition to confirming eight of the twenty SNPs as detected by univariate GWAS. Integrated, a final set of 30 significant SNPs associated with stomatal closure were reported. Taken together, our work, by combining phenomics and genetics, enables forward genetic mapping of the genetic architecture of stomatal traits related to drought

tolerance, which not only provides a basis for molecular breeding of drought resistant cultivars of cowpea, but offers a new methodology to explore the genetic determinants of water budgeting in crops under stressful conditions in the phenomics era.

Keywords: GWAS, phenomic, drought, cowpea, stomatal

INTRODUCTION

Water deficiency caused by soil drought is one of the most severe agricultural problems affecting plant growth and crop yield globally (Gupta et al., 2020). To adapt to drought, plants have evolved the ability that enduring drought stress via changes at the morphological, physiological and molecular levels (Basu et al., 2016), which defining as drought resistance. In plant species, drought resistance includes three major strategies that involving several mechanisms, 1) drought escape, refers to a plant complete its life cycle before the onset of drought; 2) drought avoidance, refers to a plant maintain higher tissue water content to avoid tissue damage under water shortage situations; 3) drought tolerance, refers to a plant sustain a certain level of growth with low internal water content. For any given plant species, it is difficult to resolve the role of different mechanisms of drought resistance, because drought resistance is a dynamic process through regulation of thousands of genes and various metabolic pathways. In Arabidopsis, rice, and other plants, many drought-inducible genes with various functions have been identified, involving different molecular responses and gene pathways (Shinozaki and Yamaguchi-Shinozaki, 2007; Park et al., 2015; Wang et al., 2019; Gupta et al., 2020). In the underground plant organs, when the roots sense the changes of soil moisture, some genes like *EXO70A3*, *PIN4*, *DEEPER ROOTING1* could modulate root architecture patterning and depth to boost water absorption from soil, thereby improving drought tolerance (Uga et al., 2013; Ogura et al., 2019). In the aboveground plant organs, drought signal via *CLE25* peptide could be transmitted through vasculature to the leaves (Takahashi et al., 2018), through modulate stomatal conductance to improve water use efficiency, and then improve drought resistance. Most genes or transcript factor (TF) involved the hormone abscisic acid (ABA) signaling pathway, which could drive or fine-tune the ABA synthesis to change the ABA content in leaves and guard cells, through engineering the stomatal closure to reduce water lost, thereby promoting drought survival. For example, the Harpin-encoding gene *hrf1*, the RING-finger containing E3 ligase *OsSDIR1* modulate stomatal closure and enhance drought tolerance in rice (Gao et al., 2011; Zhang et al., 2011; You et al., 2013), the ABA receptors PYR/PYL/RCAR and SNF1-related protein kinases (SnRK2 kinases) could improve water use efficiency in A. thaliana (Park et al., 2015; Wang et al., 2019). In addition, some ABA-independent genes also could improve drought resistance, such as the tonoplast aquaporin gene *SlTIP2;2* could maintain larger stomatal aperture and higher whole-plant transpiration under drought stress in tomato (Sade et al., 2009).

For crop plants, sustaining its growth and maintaining yield stability by reducing yield loss under drought conditions is essential for future food security. Exploiting drought resistance gene and understanding the response of cellular signaling to water shortage is key for solving these agricultural problems. Until now, using forward genetics methods such as QTL mapping and genome-wide association mapping, a vast number of genes/QTLs controlling drought-related traits of plants have been identified in different crops such as rice, wheat, maize and cowpea (Mao et al., 2015; Oladosu et al., 2019; Sallam et al., 2019; Ravelombola et al., 2021). Due to the complexity of drought response phenotype, scientists often use a specific trait or several component traits, especially easy to investigate and assay, to unravel plant drought resistance, for example, whole-plant wilting, the senescence of unifoliates, stem greenness and the survival rates of seedling under drought, soluble sugar content, abscisic acid, jasmonic acid (JA), salicylic acid (SA) and ethylene concentration (Castro et al., 2008; Iehisa et al., 2014; Mao et al., 2015; Xu et al., 2015). However, these specific traits or indicators could only reflect plant part-responses at a single point in time, such as the end of drought. They cannot exactly reflect plant how to combat drought, especially plant how to conduct stomatal movements to control water loss in leaves during the progressive soil water content decrease.

As the gateway of both H_2O evaporation and CO_2 assimilation in the leaf, stomatal closure can prevent the leaves from desiccation at the cost of photosynthesis, growth, and crop productivity (Schulze et al., 2019). Therefore, investigating plants how to regulate stomatal closure to improve water efficiency and drought resistance is the most direct indicator for drought resistance and the best strategy for drought resistance research. Different species or crop genotypes exhibit different sensitivity in stomatal control, conferring them either profligate or conservative water use behaviors that are related to their water budgeting strategies (Sade et al., 2009). Critical soil water content (θ_{cri}), defined as the threshold of volumetric soil water content (VWC) at which a plant starts to restrict its transpiration through stomatal control, is an important quantitative indicator of plant responsiveness to soil drought (Jones, 2006). Some plants such as tomato and cowpea show high θ_{cri} (0.4–0.6) under given ambient environmental conditions, meaning that they respond more promptly to soil water decrease to restrict stomatal conductance; some other plants such as pepper and soybean exhibit relatively lower θ_{cri} (0.14–0.25) under the same conditions, reflecting a more lasting active transpiration until the soil drought becomes quite severe (Xu et al., 2015; Halperin et al., 2017; Dalal et al., 2019). Meanwhile, different crop genotypes also display high variation on the speed of stomatal closure (from full openness to complete closure) at the canopy level (Xu et al., 2015; Halperin et al., 2017; Dalal et al., 2019). How fast and how slow on the stomatal closure means plant could maintain parts of photosynthesis and a certain level of growth,

which relating to the yield loss. This specific trait, here we called the slope of transpiration rate declining (K_{Tr}), could be also as an accompanying parameter of θ_{cri} using for reflecting drought resistance. Clearly, θ_{cri} and K_{Tr} as two environment-related agronomic traits of plants are closely linked to the balance of drought tolerance and yield penalty in plants under natural drought stress. Understanding the genetic architecture of this two key traits and the governing genes is therefore crucial for harnessing favorable stomatal regulation traits in crop improvement and guiding irrigation management. Traditional manual measurement based on portable apparatuses can measure stomatal conductance (Gs) and transpiration rate (Tr) overtime, which, when combined with the recorded dynamic soil water content data, can be used to calculate the θ_{cri} and K_{Tr}. However, it is difficult to acquire comparable θ_{cri} and K_{Tr} data for a large number of individuals simultaneously using this method, thus mapping the genes governing this two key traits faces large challenging.

The recent emergence of several high-throughput physiological phenotyping systems opens a new venue for precise measurement of plant water relations at large scale (Li et al., 2021). One of such platforms developed by Halperin et al. (2017), known as Plantarray, combines gravimetric system, soil and atmospheric probes, controller and irrigation valves in a unit, enabling measurement and calculation of plant water relations as well as the soil-atmosphere parameters during the whole plant growth process. Using a linear regression analysis between transpiration rate and VWC during the drought stress phase, θ_{cri} and K_{Tr} can be derived at the population level (Halperin et al., 2017; Dalal et al., 2019). This revolutionary technology has been successfully used in selection of preferred genotypes in crops including tomato, cowpea and pepper (Xu et al., 2015; Halperin et al., 2017; Dalal et al., 2019). Legumes are staple foods and important vegetables for many cultures all over the world; however, these crops are generally vulnerable to drought as they are commonly grown in rainfed regions (Micheletto et al., 2007; Iglesias-García et al., 2015). Cowpea [*Vigna unguiculata* (L.) Walp], native to Africa, is an important meat alternative for the local poor in the tropics and sub-tropics (Singh et al., 1997). Drought responses of cowpea have been extensively studied, which revealed a wide spectrum of natural variation on stomatal sensitivity to soil drought (Verbree et al., 2015; Xu et al., 2015). In the current study, we established the methodology of large-scale phenotyping of whole-plant water relations in cowpea, which, combined with genome-wide association mapping, enabled forward genetic dissection of the complex, difficult-to-score traits θ_{cri} and K_{Tr}, for the first time.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

A total of 113 cowpea accessions, which was a subset of the 299-accesion cowpea mini-core previously used in trait-marker association mapping of many traits (Xu et al., 2017), were used in this study (**Supplementary Table S1**). A lysimetric-based high-throughput physiological phenotyping system, also known as Plantarray (Plant-Ditech, Israel), located in Huai'an (119°01′E, 33°35′N), Jiangsu province, China, was used for physiological phenotyping. This system contains 100 measuring units set up in a semi-controlled greenhouse (**Figure 1**) with temperature being compensated and ambient light. The experiment design generally followed Xu et al. (2015) and Halperin et al. (2017) with a few modifications. Specifically, 10 seeds from each accession were sown in a tray filled with growing medium (Huaisannong, Huai'an, China) for germination, and three uniform seedlings were transferred to a 3.5 L pot filled with the same medium 2 weeks later. After two more weeks of growth, the pots with plants were transferred to the Plantarray platform, with the pot surface wrapped with plastic film to prevent evaporation. Three replications (pots) were set for each cowpea accession.

System Setup

Before the pots transferring, weighing lysimeter for each unit was calibrated, and the initial weights for the empty green-bath, soil probes, irrigation drippers, the plastic film for cover and sticks for seedlings bracing were measured. After the pots were put on the green-bath one pot on each unit, the probes, drippers and sticks were added in the pots, and then the experiment was set to run. Because the system could hold a maximum of 100 pots each time, the assay of the 113 accessions were divided into four experiments, spanning from 2018 to 2019 (**Table 1**). A whole experiment included three stages, normal irrigation, drought stress treatment and water resumption. At the first stage, sufficient irrigation was supplied to ensure the saturation of soil, which occurred in the night during PM 9:00 to AM 3:00, lasting 120–200 s each pulse. At the second stage, irrigation was withheld until the plant's daily transpiration reached ~10% of that before drought. At the third stage, irrigation was resumed.

Raw Data Acquisition and Analysis of θ_{cri} and K_{Tr}

The environmental parameters including air temperature, humidity, VPD were recorded by an HC2-S3-L meteo probe (Rotronic, Crawley,United Kingdom) and LI-COR 190 Quantum Sensor (Lincoln, NE, United States). VWC was measured by a soil moisture, salinity and temperature sensor (5TE; Decagon Devices, Pullman, WA, United States). Plant growth and physiological parameters including plant weight, daily transpiration, transpiration rate (Tr) and normalized transpiration (E) were acquired by the Plantarray system automatically as previously detailed in Halperin et al. (2017). The measurement was made every 3 min. The data were simultaneously saved on the online web-based software SPAC analytics (Plant-Ditech, Israel). The average transpiration rate between 11:00 and 13:00 was plotted against VWC for each day of the drought treatment stage. A piecewise linear function was then used to approximatie θ_{cri} and K_{Tr} for each genotype with the following formula: yi = a1+k1*xi, where xi is the θ_{cri}, k1 is the K_{Tr} and yi is the max transpiration rate in 1 day, if x < xi, then y = a1+k1*x, else y = yi + k2*(x-xi), and a fitted value R^2

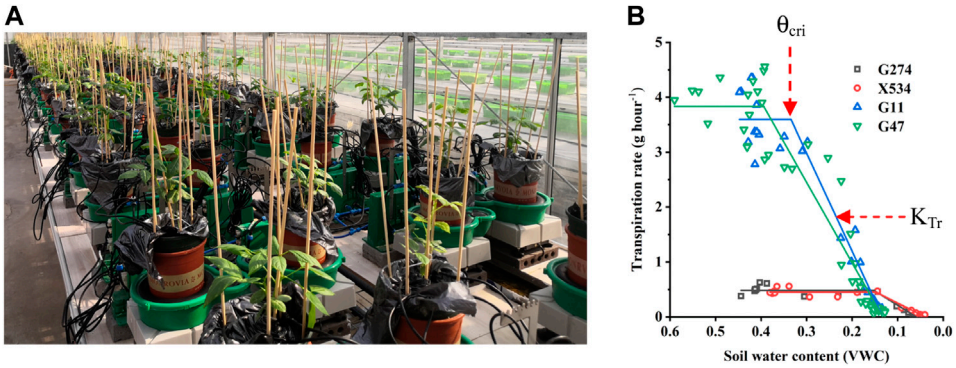


FIGURE 1 | The PlantArray system used in the current study. **(A)** Top view of the automated physiological phenotyping array with 100 measuring units loaded with cowpea seedlings. **(B)** Illustration of the θ_{cri} and K_{Tr} in four representative accessions.

Table 1 | The detailed information for four experiments.

Trial	Accessions	Start time	Normal irrigation	Drought stress	Water recovery	End time
1	73	06/07/2018	06/07–23/07	24/07–03/08	04/08–09/08	09/08/2018
2	30	18/07/2019	18/07–24/07	25/07–06/08	07/08–13/08	13/08/2019
3	31	15/08/2019	15/08–22/08	23/08–03/09	04/09–10/09	10/09/2019
4	9	23/12/2019	23/12–16/01	17/01–13/02	14/02–17/02	17/02/2020

was used to evaluate the data quality (Halperin et al., 2017). In addition, ANOVA analysis was conducted to investigate the differences among the four experiments using OriginPro 2018.

Elimination of Batch Effects

In combined data analysis using measurements from different experiments, a linear regression model was used to remove batch effects from the independent experiments (Wheeler and Chambers, 1992). Since VWC and Tr displayed a normal distribution, the specific calculation was performed using the function "lm" in the R software (R3.5 version) where the batch effect was considered as an independent variable. The calculation formula is y = a+x*b, where y is the investigated value, a is the mean value, x is the experiment batch and b is batch effect.

Population Structure Inference and GWAS

The SNP genotypic data of the 113 cowpea accessions were retrieved from Xu et al. (2017). Population structure was inferred using the software Structure 2.3.4 under the admixture model with a burn-in period of 5,000 followed by 5,000 Markov chain Monte Carlo replications. Five independent runs each were performed with the number of clusters (K) varying from 1 to 10. The optimal K for subgrouping was estimated using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). In addition, an unrooted phylogenetic tree was constructed using Tassel 5.0 under the neighbor-joining method model. Linkage disequilibrium (LD) decay was measured by calculating the square value of correlation coefficient (r^2) between each SNP pair using Tassel 5.0.

To detect the genomic regions associated with θ_{cri} and K_{Tr}, univariate GWAS was conducted on each trait using Tassel 5.0 under the generalized linear model (GLM) with accounting for population structure (Q matrix). The percentage contribution of each SNP to the total phenotypic variation was calculated based on the marker R^2 values. A multivariate GWAS accounting for population structure and relatedness was also performed on both traits in GEMMA (Zhou and Stephens 2012). For both approaches, only the SNPs showing a minus $\log_{10}$-transformed $p \geq 2.5$ were defined as significant SNPs. If two significant SNPs located in a same LD block, they were considered to represent a same QTL.

Candidate Genes Analysis

The physical locations of the detected significant SNPs were determined by aligning the marker sequences against the cowpea reference genome V1.1 (Lonardi et al., 2019). Based on the estimated LD decay distance in the genome, the genes residing in the 350 kb upstream and downstream of each SNP locus were retrieved according to the genome annotation. The putative gene functions related to drought stress were analyzed through a literature search using their orthologous genes in Arabidopsis and rice (Shinozaki and Yamaguchi-Shinozaki, 2007; Oladosu et al., 2019). Those having a putative functional relevance to drought response and/or stomatal behavior were considered as candidate causal genes.

RESULTS

Phenotypes for Whole-Plant Water Relations in the Single Experiments

Due to the limited capacity of our system (100 measuring units) and the requirement of biological replicates for physiological

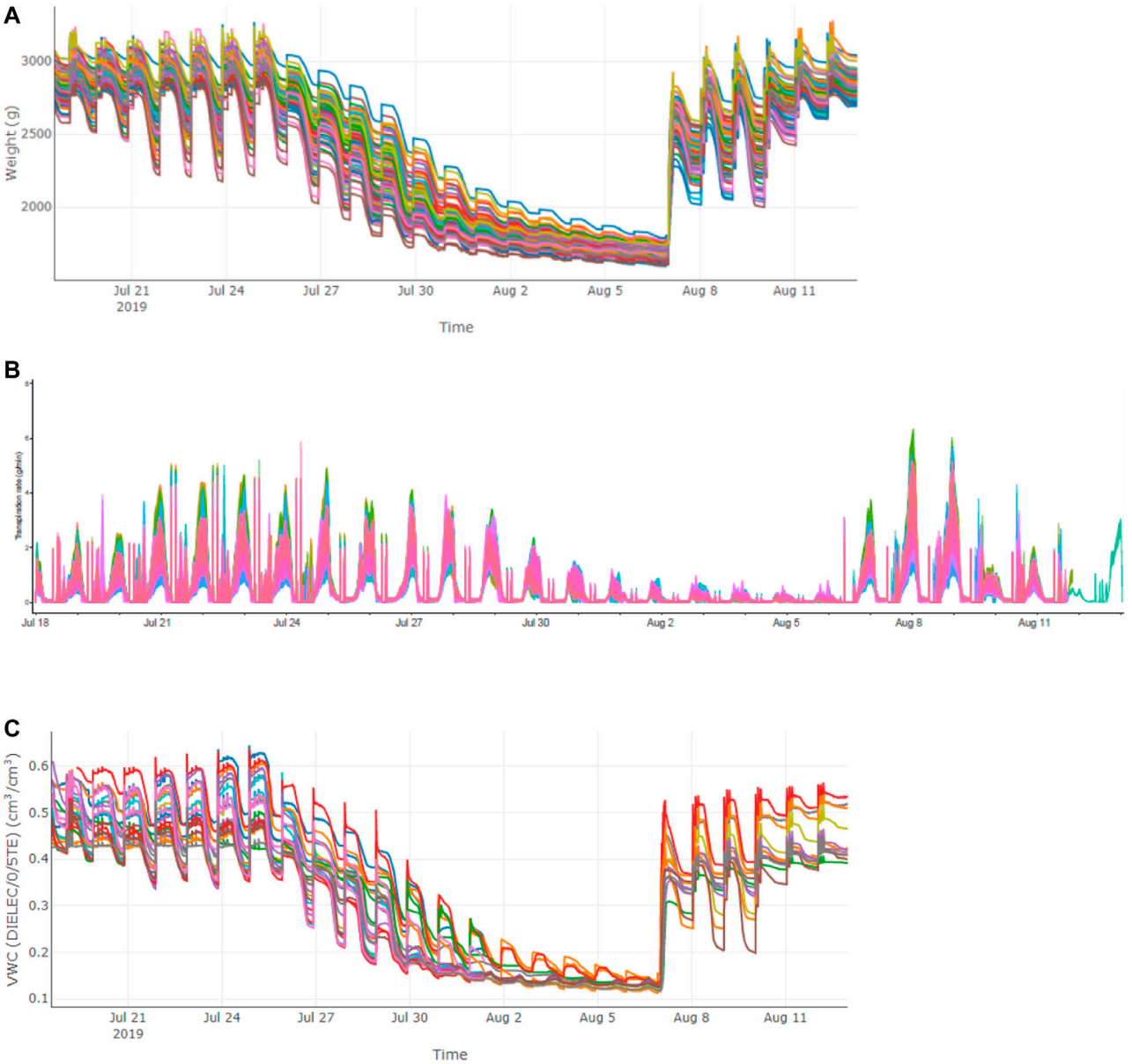


FIGURE 2 | Variations of different parameters in the population in Trial 2 during the course of experiment. **(A)** system weight; **(B)** transpiration rate; **(C)** VWC.

assay, the 113 accessions were divided into four batches for phenotyping in two consecutive years and some accessions were tested repeatedly in two trials to reduce experiment error (**Table 1**). In each experiment, plants were grown under identical ambient environmental condition and subject to the same progressive soil drought treatment imposed by water withholding, which mimicked natural field drought. As shown in **Figure 2**, the system weight in all experiment showed a similar pattern that increased gradually at the normal irrigation phase, decreased drastically with drought stress treatment and rose back rapidly with water resumption, and ranking from 1,600 to 4,000 g. The average midday transpiration rate (Trm) varied significantly among experiments, largely due to the different experimental seasons, but the CV within each experiment was relatively small, reflecting only the genotypic differences of the trait. Despite varied environmental inputs, Trm of the plants in all experiments remained stable during the initial stage of water withholding, reflecting that all the accessions could maintain the normal transpiration when the soil water is still ample (**Figure 2**). As the VWC continued to decrease, a reduction of Trm was noted due to the onset of stomata closure (**Figure 2**). By plotting the dynamic Trm data against the dynamic VWC and using the piecewise linear function, it was clear to see the turning point of Trm (θ_{cri}) and the slope of its declining (K_{Tr}), which provided a

Table 2 | ANOVA analysis for θ_{cri} and K_{Tr}.

Trait	DF	Sum of squares	Mean square	F value	Prob > F
θ_{cri}					
Model	3	0.25935	0.08645	30.89261	0
Error	102	0.28544	0.0028		
Total	105	0.54479			
K_{Tr}					
Model	3	5735.09872	1911.69957	196.75917	0
Error	102	991.02551	9.71594		
Total	105	6726.12423			

physiological measurement of the stomatal sensitivity to gradual soil drought. In **Figure 1**, the plots of four representative genotypes were shown, from which apparent genotypic variation on θ_{cri} and K_{Tr} was observed. Upon re-watering, all plants showed a rapid recovery of Trm (**Figure 2**).

Using the piecewise linear function, θ_{cri} and K_{Tr} for all the genotypes were calculated. Both the θ_{cri} and K_{Tr} showed a near-normal distribution except for Trial 4 due to the very small number (9) of accessions included (**Supplementary Figure 1**). The CV ranged from 0.12 to 0.25 for θ_{cri}, and 0.3 to 0.41 for K_{Tr}, respectively. The broad genotypic variation of these traits indicates that they could be useful in mining the genetic determinants of stomatal sensitivity to soil drought in cowpea. However, the significant between-experiment differences as detected in the four experiments according to ANOVA statistics suggested that batch effects need to be removed before merging the data from individual experiments to create a comprehensive data set for genetic mapping (**Table 2**; **Supplementary Figure 2**).

Data Quality in the Combined Population

To eliminate the batch effects within samples from different experiments, a linear regression model was used to correct the data of the two traits. After deleting data from three accessions with negative value of the fitted R^2 and four accessions lacking available genotypic data, a final mapping population with a size of 106 accessions was created. As shown in **Figure 3**, the adjusted θ_{cri} ranged from 0.09 to 0.78 with a median value of 0.21, and the adjusted K_{Tr} ranged from 0.44 to 9.17 with a median value of 3.09, after eliminating the batch effects. Both the θ_{cri} and K_{Tr} displayed a nearly normal distribution (**Figure 3**), with the CVs being 0.44 and 0.50, respectively, suggesting that this population is suitable for gene mapping. Approximately 14% of the accessions had a θ_{cri} value smaller than 0.15, and 6.6% of the accessions had a value greater than 0.35, indicating the presence of only a small portion of the genotypes with exceptionally high or low stomatal sensitivity to soil drought. On the contrary, for K_{Tr}, 49% of the accessions had a value below the median value of 3.09, suggesting that the majority of the cowpea accessions adjusted their paces of stomatal closure from full openness to complete closure slowly at the canopy scale.

Genetic Mapping of θ_{cri} and K_{Tr}

Before performing GWAS of the θcri and K_{Tr}, the genetic diversity of the 106 accessions was analyzed. Population structure analysis based on the 434 representative SNPs genotype data (Wu et al., 2021) showed that the peak of delta *K* appeared at $K = 2$ (**Figure 4**), suggesting that the 106-accession subset of the mini-core could be classified into two subpopulations, which is consistent with Xu et al. (2017) from analyzing the entire set of mini-core comprising 299 accessions. The dendrogram of neighbor-joining (NJ) tree also suggested two main branches of these genotypes (**Figure 4**). Based on their physical location of the SNPs in the reference genome (Lonardi et al., 2019), the LD decays ($r^2 = 0.3$) at about 350 kbp across the whole panel (**Supplementary Figure 3**).

In the univariate GWAS, 13 significant SNPs associated with θ_{cri} and 64 with K_{Tr} were identified, respectively (**Figure 5**; **Supplementary Table S2**). The θ_{cri}-associated SNPs were distributed on chromosomes Vu03, Vu06 and Vu11, which accounted for 8.7–21.0% of the phenotypic variation. The K_{Tr}-associated SNPs were located on chromosomes Vu01, Vu03, Vu04, Vu05, Vu07, Vu08, Vu09 and Vu11, each explaining 9.1–13.2% of the phenotypic variation. These results indicate that both stomatal sensitivity to soil drought and the speed of stomatal closure to completion in cowpea were controlled by multiple genes with moderate effects. Of the detected SNPs, some formed clusters and located in the same LD blocks, suggesting

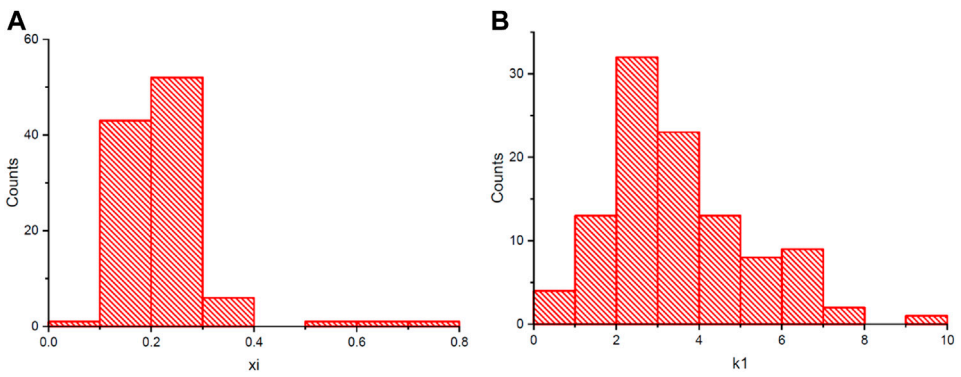


FIGURE 3 | The frequency distribution of θ_{cri} **(A)** and K_{Tr} **(B)** in 106 accessions after removal of batch effects.

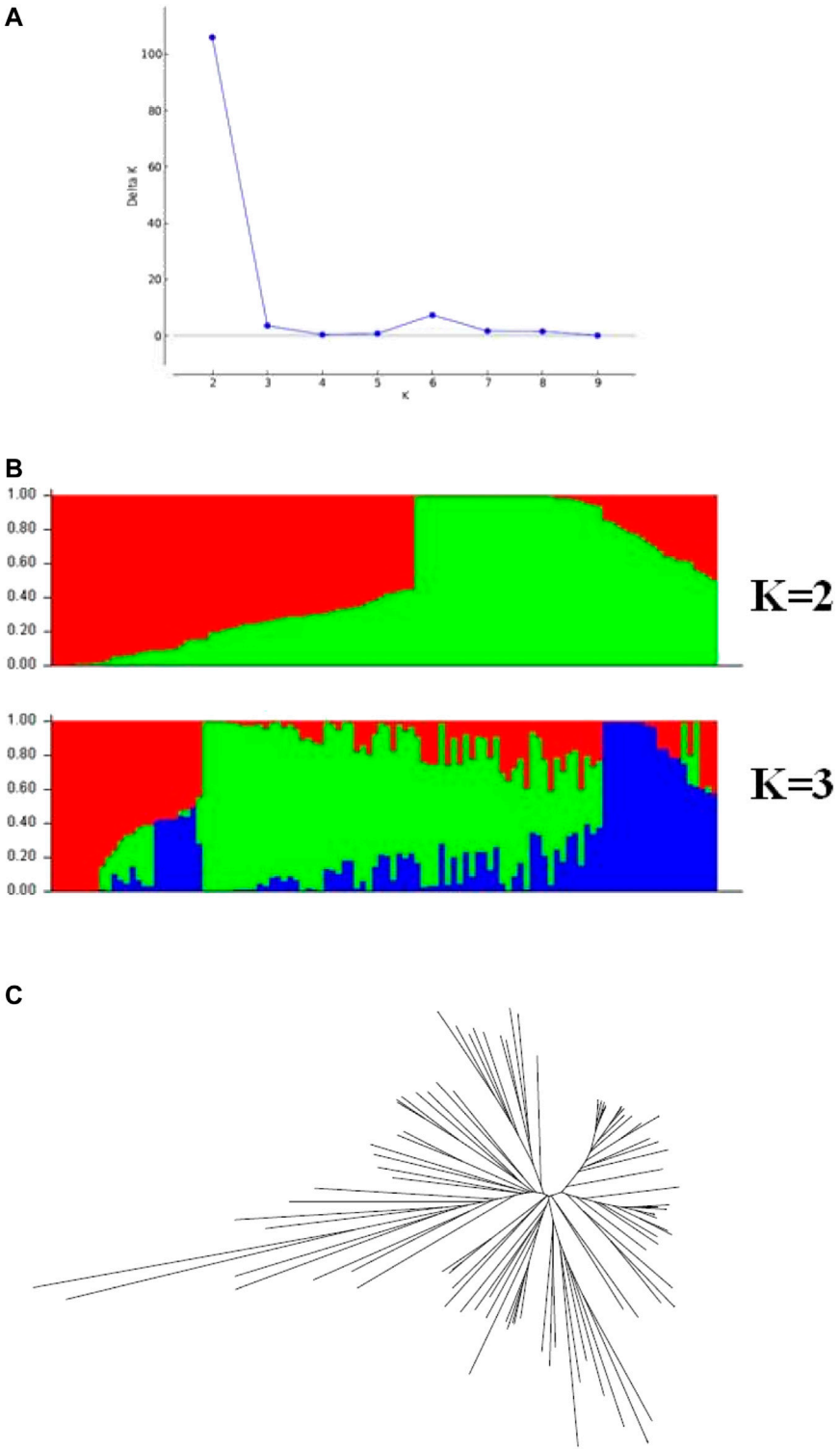


FIGURE 4 | Population structure across the 106 cowpea accessions. **(A)** Delta K values for different numbers of population assumed K values. **(B)** estimated population structure of the germplasm collection inferred at K = 2 and K = 3. **(C)** an unrooted Neighbor-joining tree showing the dendrogram of all samples.

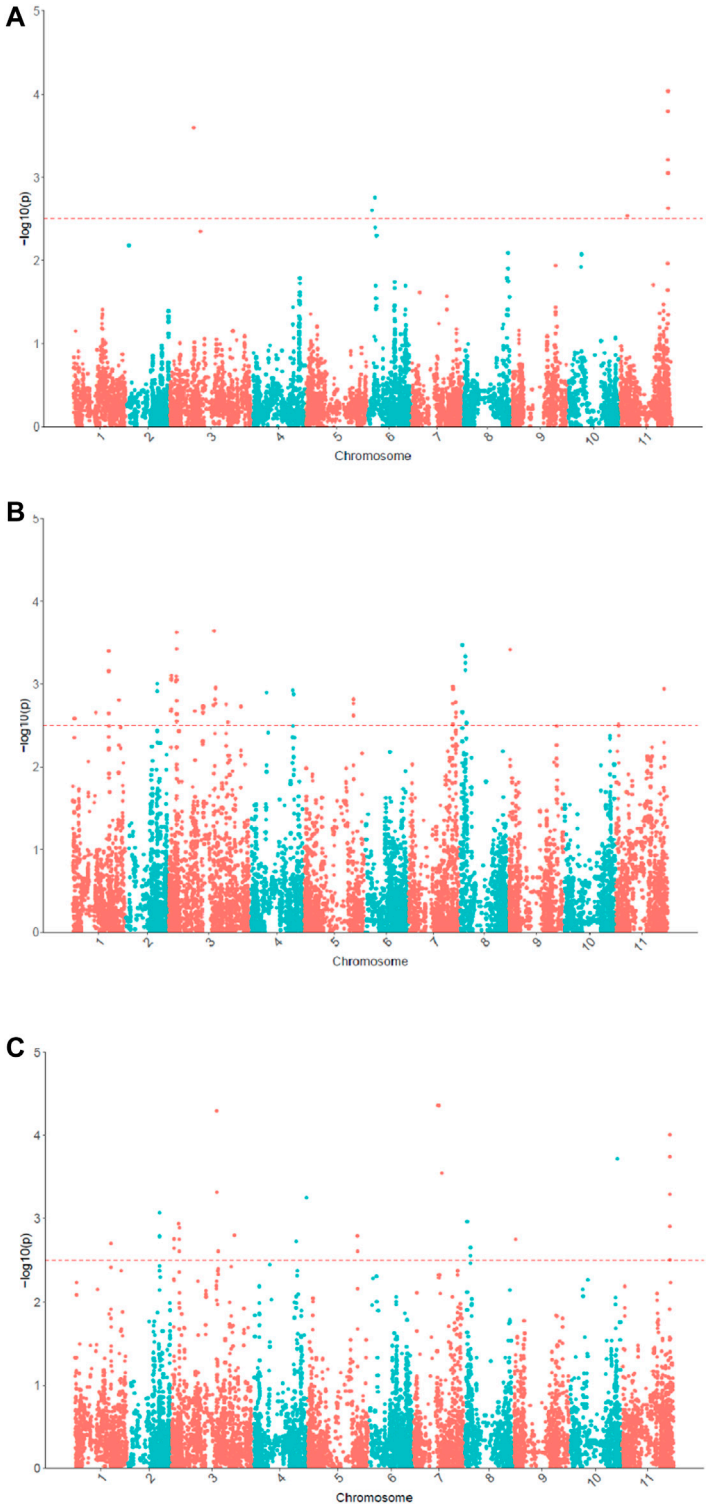


FIGURE 5 | Manhattan plots for θ_{cri} **(A)** and K_{Tr} **(B)** in univariate GWAS and multivariate GWAS for stomatal closure **(C)**.

that they may represent a single locus associated with the trait. For clarity, a final set of six significant SNPs associated with θ_{cri} and fourteen associated with K_{Tr} were reported after retaining only one representative locus in a LD block. There was no overlap between the θ_{cri}- and K_{Tr}-associated SNPs, indicating that these two traits might be governed by different set of genes.

In the multivariate GWAS, a total of 58 significant SNPs were detected (**Figure 5**; **Supplementary Table S2**), which were distributed in all eleven chromosomes except for chromosome Vu05. After merging the SNPs by LD block, 18 representative SNPs were reported in **Supplementary Table S2**. Of these 18 SNPs, 10 were only detectable in multivariate GWAS, and the remaining eight (or their very closely-neighboring SNPs) viz. 2_53558, 2_31502, 2_07162, 2_12695, 2_06424, 2_15420, 2_03550, 2_16293 were detected also in univariate GWAS. Among these eight SNPs, seven were associated with K_{Tr} and only one was associated with θ_{cri}. These results demonstrated that multivariate GWAS could detected more significant SNPs associated with the two traits. By integrating the results of univariate and multivariate GWAS, a final set of 30 SNPs associated with stomatal-closure traits were identified (**Supplementary Table S2**).

Candidate Gene Analysis

Given the LD decay distance of 350 kbp in our population, candidate genes in the 350 kb regions up- and down-stream flanking the detected SNPs were searched. This analysis discovered 28 genes as interesting candidate genes (**Supplementary Table S3**) based on their functional annotations related to drought resistance. These genes included E3 ubiquitin ligase genes, the LEA protein genes, NAC domain proteins, MYB domain protein, amino acid transporters, RING finger domain protein and Zinc finger domain proteins, which were involved in response to drought stress. These results provide a rich candidate gene reservoir for better understanding the mechanisms of stomatal behavior and physiological drought resistance in cowpea, which will help accelerate genetic improvement against drought stress.

DISCUSSION

Stomatal conductance is known to be a reliable indicator of growth-rate responses to stress (Munns et al., 2010). In response to drought, stomatal closure not only preserves water loss from the plant, but also constrains CO_2 import to the leaves and thus photosynthesis rate, thereby playing a key role in balancing plant drought resistance and yield (Gupta et al., 2020). Although stomatal control has been extensively studied by physiologists and molecular biologists, the forward genetic dissection of stomatal closure under water stress conditions still largely lagged due to the limited capacity of trait measurements, especially high-throughput measurement at the population level. In the current study, we used a high-throughput physiological phenotyping platform to monitor the dynamic water relations of 106 cowpea accessions simultaneously and continuously. Two specific indicators, θ_{cri} and K_{Tr} were measured and proved to be useful for assessing stomatal responses under progressive water stress conditions. Using phenotypic data of θ_{cri} and K_{Tr}, 30 significant SNPs associated with sensitivity or duration of stomatal closure were detected by the GWAS approaches. To our knowledge, this is the first report on genetic mapping of these stomatal closure related traits using forward genetic method in crops.

θ_{cri} and K_{Tr} indicate the critical VWC point at which the stomata start to close and the speed of stomatal closure from full openness to complete closure, respectively. The two traits can, therefore, reflect the property of stomatal response to soil drought from two complementary perspectives. To increase the power of QTL detection, here a combination of univariate GWAS and multivariate GWAS were used. Compared with traditional GWAS or univariate GWAS, multivariate GWAS or multi-trait GWAS showed higher statistical power to detect signals for complex or multiple traits, as has also been demonstrated in studies of the genomic region associated with seed fatty acid in oat and inflorescence and leaf architecture in maize (Carlson et al., 2019; Rice et al., 2020). Our results showed that eight SNPs were detected by both methods, and 10 SNPs were detected only by multivariate GWAS, which proved the greater power of this method and indicate that the 10 SNPs may represent pleiotropic quantitative trait loci for θ_{cri} and K_{Tr}.

Due to its high adaptability to drought stress and relatively small genome size (~ 620 Mb), cowpea has long been used as a model legume crop to understand the genetic basis of drought tolerance in legumes (Ravelombola et al., 2020). In previous studies, based on visually scored morphological traits such as delayed senescence (Dro), stem greenness (Stg), leaf senescence (Scu) and drought tolerance indices, hundreds of significant SNPs associated with drought resistance have been detected by GWAS (Muchero et al., 2013; Xu et al., 2015; Ravelombola et al., 2021). By comparing the locations of these earlier reported SNPs and those mapped in the present study, we found that four known drought-related SNPs/QTLs (Stg 2_25850, Stg 1_0274, Stg 2_07162, Dro-7 1_0067) were co-localized with the newly mapped stomata-associated SNPs, meaning they reside in the same LD decay blocks (distance less than 350 kb) and may represent the same QTLs. These results also imply that the genes underlying stomatal control may ultimately lead to an effect on the morphological responses of cowpea plants to soil drought.

The phytohormone abscisic acid (ABA) plays a critical role in the regulation of stomatal closure to adjust water loss (Wang et al., 2019). Under drought stress, ABA accumulates rapidly and binds its receptors belonging to the PYR1/PYL/RCAR family, which inhibits downstream protein phosphatases to initiate protective responses such as stomatal closure and gene expression reprogramming (Wang et al., 2019). A series of functional genes or transcription factors such as MYB/MYC, NAC proteins, SnRK2, E3 ubiquitin ligase are known to be involved in the ABA signaling pathway (Gao et al., 2011, Shinozaki and Yamaguchi-Shinozaki, 2007; Wang et al., 2019). In our study, 28 gene were listed as interesting candidate genes in the QTL regions based on their putative functional annotations. Most of these genes were presumably to involve in the ABA signaling pathway, including three MYB domain proteins, one NAC domain protein, two E3 ubiquitin ligase proteins, six zinc finger domain proteins and four ring finger containing proteins (**Supplementary Table S3**). Under drought conditions, these genes may increase drought resistance by adjusting the

stomatal aperture. Therefore, ABA signaling pathway is postulated to be the major pathway regulating the stomatal response under drought stress conditions in cowpea. In addition, three late embryogenesis abundant (LEA) proteins were included in the candidate gene list, which may function as chaperone-like protective molecules to combat cellular damage (Babu et al., 2004; Shinozaki and Yamaguchi-Shinozaki, 2007). To confirm the functional relevance of the candidate genes to stomatal regulation, more future work such as fine mapping and positional cloning of these genes are required.

CONCLUSION

In the current study, high-throughput physiological phenotyping was employed to quantify two stomatal-related traits that were traditionally difficult-to-score at the population level. Through combining univariate GWAS and multivariate GWAS, we detected 30 significant SNPs associated with θ_{cri} and K_{Tr}. The present study provides a so-far rare case of combining high-throughput physiological phenotyping and genetic mapping for forward genetic mapping of stomatal behaviors. Our results lay a foundation for better understanding the molecular mechanisms of stomatal regulation under drought stress conditions as well as marker-assisted breeding for more balanced drought tolerance and yield under drought scenarios in cowpea.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PX and GL conceived and designed the research. XinW performed the experiments, analyzed the data and wrote the manuscript. WX and YS contributed to Plantarray experiment, TS contributed to data analysis. BW, YW, YL, JW, XiaW, and ZL contributed to germplasm collection and Plantarray experiment, PX edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by National Natural Science Foundation (NSFC) of China (Grant No. 31772299), NSFC-Israeli Science Foundation (ISF) joint project (Grant No. 31861143044), the Key R&D Program of Guangdong Province (Grant No. 2020B020220002) and the National Key R&D Program of China (Grant No. 2017YFE0114500).

ACKNOWLEDGMENTS

We sincerely thank Ye Tao and Qin Lin (Biozeron Biotech, Shanghai) for technical assistances in Plantarray experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.743758/full#supplementary-material

Supplementary Figure S1 | The frequency distribution for θ_{cri} **(A)** and K_{Tr} **(B)** in the four experiments.

Supplementary Figure S2 | The box charts of θ_{cri} and K_{Tr} in the populations in the four experiments. **(A)** original data for θ_{cri}; **(B)** original data for K_{Tr}; **(C)** θ_{cri} after removing the batch effects; **(D)** θ_{cri} after removing the batch effects.

Supplementary Figure S3 | Decay of linkage disequilibrium (LD) in all samples. The strength of LD was measured by r^2.

Supplementary Table S1 | Germplasm lines used in the current study.

Supplementary Table S2 | The detected SNPs associated with θ_{cri} and K_{Tr} in univariate GWAS and multivariate GWAS.

Supplementary Table S3 | Twenty-eight putative candidate genes surrounding to the detected SNPs.

REFERENCES

Basu, S., Ramegowda, V., Kumar, A., and Pereira, A. (2016). Plant Adaptation to Drought Stress. *F1000Res* 5, 1554. doi:10.12688/f1000research.7678.1

Carlson, M. O., Montilla-Bascon, G., Hoekenga, O. A., Tinker, N. A., Poland, J., Baseggio, M., et al. (2019). Multivariate Genome-wide Association Analyses Reveal the Genetic Basis of Seed Fatty Acid Composition in Oat (*Avena Sativa* L.). *G* 9 (9), 2963–2975. doi:10.1534/g3.119.400228

Castro, A. M., Tacaliti, M. S., Giménez, D., Tocho, E., Dobrovolskaya, O., Vasicek, A., et al. (2008). Mapping Quantitative Trait Loci for Growth Responses to Exogenously Applied Stress Induced Hormones in Wheat. *Euphytica* 164, 719–727. doi:10.1007/s10681-008-9694-5

Chandra Babu, R., Zhang, J., Blum, A., David Ho, T.-H., Wu, R., and Nguyen, H. T. (2004). *HVA1*, a LEA Gene from Barley Confers Dehydration Tolerance in Transgenic rice (*Oryza Sativa* L.) via Cell Membrane protection. *Plant Sci.* 166, 855–862. doi:10.1016/j.plantsci.2003.11.023

Dalal, A., Bourstein, R., Haish, N., Shenhar, I., Wallach, R., and Moshelion, M. (2019). Dynamic Physiological Phenotyping of Drought-Stressed Pepper Plants Treated with "Productivity-Enhancing" and "Survivability-Enhancing" Biostimulants. *Front. Plant Sci.* 10, 905. doi:10.3389/fpls.2019.00905

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7

Gao, T., Wu, Y., Zhang, Y., Liu, L., Ning, Y., Wang, D., et al. (2011). *OsSDIR1* Overexpression Greatly Improves Drought Tolerance in Transgenic rice. *Plant Mol. Biol.* 76, 145–156. doi:10.1007/s11103-011-9775-z

Gupta, A., Rico-Medina, A., and Caño-Delgado, A. I. (2020). The Physiology of Plant Responses to Drought. *Science* 368, 266–269. doi:10.1126/science.aaz7614

Halperin, O., Gebremedhin, A., Wallach, R., and Moshelion, M. (2017). High-throughput Physiological Phenotyping and Screening System for the Characterization of Plant-Environment Interactions. *Plant J.* 89 (4), 839–850. doi:10.1111/tpj.13425

Iehisa, J. C. M., Matsuura, T., Mori, I. C., and Takumi, S. (2014). Identification of Quantitative Trait Locus for Abscisic Acid Responsiveness on Chromosome 5A and Association with Dehydration Tolerance in Common Wheat Seedlings. *J. Plant Physiol.* 171 (2), 25–34. doi:10.1016/j.jplph.2013.10.001

Iglesias-García, R., Prats, E., Fondevilla, S., Satovic, Z., and Rubiales, D. (2015). Quantitative Trait Loci Associated to Drought Adaptation in Pea (*Pisum Sativum* L.). *Plant Mol. Biol. Rep.* 33, 1768–1778. doi:10.1007/s11105-015-0872-z

Jones, H. G. (2006). Monitoring Plant and Soil Water Status: Established and Novel Methods Revisited and Their Relevance to Studies of Drought Tolerance. *J. Exp. Bot.* 58 (2), 119–130. doi:10.1093/jxb/erl118

Li, Y., Wu, X., Xu, W., Sun, Y., Wang, Y., Li, G., et al. (2021). High-Throughput Physiology-Based Stress Response Phenotyping: Advantages, Applications and Prospective in Horticultural Plants. *Hortic. Plant J.* 7 (3), 181–187. doi:10.1016/j.hpj.2020.09.004

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The Genome of Cowpea (Vigna Unguiculata [L.] Walp.). *Plant J.* 98 (5), 767–782. doi:10.1111/tpj.14349

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., et al. (2015). A Transposable Element in a *NAC* Gene Is Associated with Drought Tolerance in maize Seedlings. *Nat. Commun.* 6, 8326. doi:10.1038/ncomms9326

Micheletto, S., Rodriguez-Uribe, L., Hernandez, R., Richins, R. D., Curry, J., and O'Connell, M. A. (2007). Comparative Transcript Profiling in Roots of *Phaseolus Acutifolius* and *P. Vulgaris* under Water Deficit Stress. *Plant Sci.* 173, 510–520. doi:10.1016/j.plantsci.2007.08.003

Muchero, W., Roberts, P. A., Diop, N. N., Drabo, I., Cisse, N., Close, T. J., et al. (2013). Genetic Architecture of Delayed Senescence, Biomass, and Grain Yield under Drought Stress in Cowpea. *PLoS One* 8 (7), e70041. doi:10.1371/journal.pone.0070041

Munns, R., James, R. A., Sirault, X. R. R., Furbank, R. T., and Jones, H. G. (2010). New Phenotyping Methods for Screening Wheat and Barley for Beneficial Responses to Water Deficit. *J. Exp. Bot.* 61, 3499–3507. doi:10.1093/jxb/erq199

Ogura, T., Goeschl, C., Filiault, D., Mirea, M., Slovak, R., Wolhrab, B., et al. (2019). Root System Depth in Arabidopsis Is Shaped by *EXOCYST70A3* via the Dynamic Modulation of Auxin Transport. *Cell* 178 (2), 400–412. doi:10.1016/j.cell.2019.06.021

Oladosu, Y., Rafii, M. Y., Samuel, C., Fatai, A., Magaji, U., Kareem, I., et al. (2019). Drought Resistance in rice from Conventional to Molecular Breeding: a Review. *Ijms* 20 (14), 3519. doi:10.3390/ijms20143519

Park, S.-Y., Peterson, F. C., Mosquna, A., Yao, J., Volkman, B. F., and Cutler, S. R. (2015). Agrochemical Control of Plant Water Use Using Engineered Abscisic Acid Receptors. *Nature* 520, 545–548. doi:10.1038/nature14123

Ravelombola, W., Shi, A., Chen, S., Xiong, H., Yang, Y., Cui, Q., et al. (2020). Evaluation of Cowpea for Drought Tolerance at Seedling Stage. *Euphytica* 216, 123. doi:10.1007/s10681-020-02660-4

Ravelombola, W., Shi, A., and Huynh, B.-L. (2021). Loci Discovery, Network-Guided Approach, and Genomic Prediction for Drought Tolerance index in a Multi-Parent Advanced Generation Intercross (MAGIC) Cowpea Population. *Hortic. Res.* 8 (1), 24. doi:10.1038/s41438-021-00462-w

Rice, B. R., Fernandes, S. B., and Lipka, A. E. (2020). Multi-trait Genome-wide Association Studies Reveal Loci Associated with maize Inflorescence and Leaf Architecture. *Plant Cel Physiol* 61 (8), 1427–1437. doi:10.1093/pcp/pcaa039

Sade, N., Vinocur, B. J., Diber, A., Shatil, A., Ronen, G., Nissan, H., et al. (2009). Improving Plant Stress Tolerance and Yield Production: Is the Tonoplast Aquaporin *SlTIP2;2* a Key to Isohydric to Anisohydric Conversion. *New Phytol.* 181 (3), 651–661. doi:10.1111/j.1469-8137.2008.02689.x

Sallam, A., Alqudah, A. M., Dawood, M. F. A., Baenziger, P. S., and Börner, A. (2019). Drought Stress Tolerance in Wheat and Barley: Advances in Physiology, Breeding and Genetics Research. *Ijms* 20 (13), 3137. doi:10.3390/ijms20133137

Schulze, E.-D., Beck, E., Buchmann, N., Clemens, S., Müller-Hohenstein, K., and Scherer-Lorenzen, M. (2019). "Water Deficiency (Drought)," in *Plant Ecology* (Berlin, Heidelberg: Springer), 165–202. doi:10.1007/978-3-662-56233-8_6

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene Networks Involved in Drought Stress Response and Tolerance. *J. Exp. Bot.* 58 (2), 221–227. doi:10.1093/jxb/erl164

Singh, B, B., Chambliss, O, L., and Sharma, B. (1997). "Recent Advances in Cowpea Breeding," in *Advances in Cowpea Research, Co-publication of International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria and Japan International Research Centre for Agricultural Sciences (JIRCAS)*. Editors B. B. Singh, D. R. MohanRaj, K. E. Dashiell, and L. E. N. Jackai (Devon: Sayce Publishing), 114–128.

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., et al. (2018). A Small Peptide Modulates Stomatal Control via Abscisic Acid in Long-Distance Signalling. *Nature* 556, 235–238. doi:10.1038/s41586-018-0009-2

Uga, Y., Sugimoto, K., Ogawa, S., Rane, J., Ishitani, M., Hara, N., et al. (2013). Control of Root System Architecture by *DEEPER ROOTING 1* Increases rice Yield under Drought Conditions. *Nat. Genet.* 45, 1097–1102. doi:10.1038/ng.2725

Verbree, D. A., Singh, B. B., and Payne, W. A. (2015). Genetics and Heritability of Shoot Drought Tolerance in Cowpea Seedlings. *Crop Sci.* 55, 146–153. doi:10.2135/cropsci2014.02.0137

Wang, P., Zhao, Y., Li, Z., Hsu, C.-C., Liu, X., Fu, L., et al. (2018). Reciprocal Regulation of the TOR Kinase and ABA Receptor Balances Plant Growth and Stress Response. *Mol. Cel* 69 (1), 100–112. e6. doi:10.1016/j.molcel.2017.12.002

Wheeler, D. J., and Chambers, D. S. (1992). *Understanding Statistical Process Control*. Knoxville, TN: SPC Press.

Wu, X., Wang, B., Wu, S., Li, S., Zhang, Y., Wang, Y., et al. (2021). Development of a Core Set of Single Nucleotide Polymorphism Markers for Genetic Diversity Analysis and Cultivar Fingerprinting in Cowpea. *Legume Sci.* 3. doi:10.1002/leg3.93

Xu, P., Moshelion, M., Wu, X., Halperin, O., Wang, B., Luo, J., et al. (2015). Natural Variation and Gene Regulatory Basis for the Responses of asparagus Beans to Soil Drought. *Front. Plant Sci.* 6, 891. doi:10.3389/fpls.2015.00891

Xu, P., Wu, X., Muñoz-Amatriaín, M., Wang, B., Wu, X., Hu, Y., et al. (2017). Genomic Regions, Cellular Components and Gene Regulatory Basis Underlying Pod Length Variations in Cowpea (V. unguiculataL. Walp). *Plant Biotechnol. J.* 15, 547–557. doi:10.1111/pbi.12639

You, J., Zong, W., Li, X., Ning, J., Hu, H., Li, X., et al. (2013). The *SNAC1*-Targeted Gene *OsSRO1c* Modulates Stomatal Closure and Oxidative Stress Tolerance by Regulating Hydrogen Peroxide in rice. *J. Exp. Bot.* 64 (2), 569–583. doi:10.1093/jxb/ers349

Zhang, L., Xiao, S., Li, W., Feng, W., Li, J., Wu, Z., et al. (2011). Overexpression of a Harpin-Encoding Gene *Hrf1* in rice Enhances Drought Tolerance. *J. Exp. Bot.* 62 (12), 4229–4238. doi:10.1093/jxb/err131

Zhou, X., and Stephens, M. (2012). Genome-wide Efficient Mixed-Model Analysis for Association Studies. *Nat. Genet.* 44, 821–824. doi:10.1038/ng.2310

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.

Copyright © 2021 Wu, Sun, Xu, Sun, Wang, Wang, Li, Wang, Wu, Lu, Xu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 October 2021
doi: 10.3389/fgene.2021.763841

RNA Sequencing Reveals Phenylpropanoid Biosynthesis Genes and Transcription Factors for *Hevea brasiliensis* Reaction Wood Formation

Xiangxu Meng[1,2], *Yue Wang*[1,2], *Jia Li*[1,2], *Nanbo Jiao*[1,2], *Xiujie Zhang*[1,2], *Yuanyuan Zhang*[3,4], *Jinhui Chen*[1,2]* *and Zhihua Tu*[1,2]*

[1]*Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education/Engineering Research Center of Rare and Precious Tree Species in Hainan Province, School of Forestry, Hainan University, Haikou, China,* [2]*Hainan Key Laboratory for Biology of Tropical Ornamental Plant Germplasm, Institute of Tropical Agriculture and Forestry, School of Forestry, Hainan University, Haikou, China,* [3]*Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China,* [4]*State Centre for Rubber Breeding, Haikou, China*

Given the importance of wood in many industrial applications, much research has focused on wood formation, especially lignin biosynthesis. However, the mechanisms governing the regulation of lignin biosynthesis in the rubber tree (*Hevea brasiliensis*) remain to be elucidated. Here, we gained insight into the mechanisms of rubber tree lignin biosynthesis using reaction wood (wood with abnormal tissue structure induced by gravity or artificial mechanical treatment) as an experimental model. We performed transcriptome analysis of rubber tree mature xylem from tension wood (TW), opposite wood (OW), and normal wood (NW) using RNA sequencing (RNA-seq). A total of 214, 1,280, and 32 differentially expressed genes (DEGs) were identified in TW vs. NW, OW vs. NW, and TW vs. OW, respectively. GO and KEGG enrichment analysis of DEGs from different comparison groups showed that zeatin biosynthesis, plant hormone signal transduction, phenylpropanoid biosynthesis, and plant–pathogen interaction pathways may play important roles in reaction wood formation. Sixteen transcripts involved in phenylpropanoid biosynthesis and 129 transcripts encoding transcription factors (TFs) were used to construct a TF–gene regulatory network for rubber tree lignin biosynthesis. Among them, MYB, C2H2, and NAC TFs could regulate all the DEGs involved in phenylpropanoid biosynthesis. Overall, this study identified candidate genes and TFs likely involved in phenylpropanoid biosynthesis and provides novel insights into the mechanisms regulating rubber tree lignin biosynthesis.

Keywords: rubber tree, rubber wood, reaction wood, phenylpropanoid biosynthesis pathway, lignin biosynthesis

OPEN ACCESS

Edited by:
Jiang Libo,
Beijing Forestry University, China

Reviewed by:
Qingzhang Du,
Beijing Forestry University, China
Nan Lu,
Chinese Academy of Forestry, China

***Correspondence:**
Jinhui Chen
jinhuichen@hainanu.edu.cn
Zhihua Tu
tuzhihua@hainanu.edu.cn

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

Received: *24 August 2021*
Accepted: *08 October 2021*
Published: *29 October 2021*

Citation:
Meng X, Wang Y, Li J, Jiao N, Zhang X, Zhang Y, Chen J and Tu Z (2021) RNA Sequencing Reveals Phenylpropanoid Biosynthesis Genes and Transcription Factors for Hevea brasiliensis *Reaction Wood Formation. Front. Genet. 12:763841. doi: 10.3389/fgene.2021.763841*

INTRODUCTION

The rubber tree (*Hevea brasiliensis*) is a deciduous perennial tropical tree native to the Amazon basin that produces natural rubber as well as rubber wood. Rubber wood can be used to make a variety of products, such as pulpwood for paper production and rubber wood-based panels, furniture, and joinery products (Jahan et al., 2011; Teoh et al., 2011; Severo et al., 2016). The economic importance of the rubber tree and increasing demand have promoted its widespread domestication

(Priyadarshan, 2017). Given the increasing economic and application value of rubber wood, improving wood productivity and quality has become the focus of rubber tree breeding (Priyadarshan, 2017).

Wood mainly consists of cellulose, hemicellulose, and lignin. Lignin is a macromolecular biopolymer similar to cellulose, and is widely distributed in higher plants (Wang et al., 2018). During plant growth and development, lignin content increases, causing cell wall thickening and plant tissue lignification. As a crucial secondary metabolite of the phenylpropanoid biosynthesis pathway, lignin had profound effects on plant growth and development (Liu et al., 2018; Wang et al., 2019). It is extensively involved in the material transport of vascular bundles, cell wall structural integrity, stem strength, mechanical support, response to pathogens and other environmental stresses (You et al., 2013; Wang et al., 2019). Thus, the regulation of lignin synthesis affects not only lignin accumulation, but ultimately the growth and development of the entire plant.

Lignin polymers are primarily derived from the monolignols p-hydroxyphenyl, guaiacyl, and syringyl, which are formed by dehydrogenation of the hydroxycinnamyl alcohols p-coumaryl, coniferyl, and sinapyl, respectively (Zhao and Dixon, 2011). Monolignols are produced from phenylalanine through the phenylpropanoid pathway (Whetten and Sederoff, 1995). The first step of this pathway involves phenylalanine deamination by phenylalanine ammonia-lyase (PAL) to form cinnamic acid. This deamination step is followed by a series of hydroxylation reactions (catalyzed by cinnamate 4-hydroxylase, C4H; coumarate 3-hydroxylase; ferulate 5-hydroxylase, F5H; 4-coumarate-CoA ligase, 4CL; and hydroxycinnamoyltransferase, HCT), O-methylation reactions (carried out by caffeate O-methyltransferase, COMT; and caffeoyl-CoA O-methyltransferase, CCoAOMT), and reduction reactions (catalyzed by cinnamoyl-CoA reductase, CCR; and cinnamyl alcohol dehydrogenase, CAD). Lignin is ultimately formed through polymerization of monolignols by peroxidases (PODs) or laccases. The methoxylation levels of the three monolignols determine the amount and composition of lignin (Boerjan et al., 2003).

The genes for many key enzymes in lignin biosynthesis have been identified, with some also shown to influence environmental responses (Liu et al., 2018; Wang et al., 2018; Wang et al., 2019). For instance, up-regulation of *POD*, *PAL*, *C4H*, and *4CL* promotes lignin biosynthesis and increases muskmelon resistance to black spot disease (Han et al., 2018; Yan et al., 2019). Moreover, suppression of *Os4CL* expression leads to reduced lignin content in rice (*Oryza sativa*; Gui et al., 2011). Similarly, transgenic tobacco over-expressing the peroxidase gene *SWPA4* show high POD activity and lignin accumulation (Kim et al., 2008). Several transcription factors regulate the expression of lignin biosynthetic genes (Liu et al., 2018). In different plant species, TFs from the MYB (Ohtani and Demura, 2019), NAC (Ohtani and Demura, 2019), WRKY (Gallego-Giraldo et al., 2016), MADS (Li et al., 2016) and HSF (Zeng et al., 2016) families control lignin biosynthesis by regulating genes related to cell wall synthesis. These findings provided additional insight into the mechanism of lignin biosynthesis and accumulation, that is TFs could participate in lignin biosynthesis.

In this study, transcriptome sequencing of *H. brasiliensis* reaction wood uncovered genes involved in regulating lignin synthesis was firstly reported in rubber tree. Identifying transcriptome-level changes of lignin synthesis-related genes contributes to our understanding of rubber wood formation and may help elucidate the regulatory mechanisms of rubber tree lignin biosynthesis. Overall, our findings identify the candidate genes and serve as an important theoretical basis to develop breeding strategies that improve rubber tree quality.

MATERIALS AND METHODS

Plant Material and Sample Collection

The rubber tree clones (Reyan 7-33-97) used in this study were obtained by *in vitro* tissue culture and planted in the Hainan University experimental greenhouse (Danzhou, Hainan, China; 109°29′25″ E, 19°30′40″ N) at the end of June 2016. To explore the genes involved in reaction wood formation, rubber tree trunks of similar diameter (approximately 2 cm) were selected as experimental materials. Three three-year-old rubber trees were bent at a 30° angle for 300 days to induce reaction wood formation. The bending treatment started at 9 a.m. on August 17th, 2020, and ended at 9 a.m. on June 13th, 2021. The xylem samples were collected immediately after completion of the bending treatment. Mature xylem tissue samples from NW (normal wood), TW (tension wood), and OW (opposite wood) were isolated from the same individual to enable comparisons in the same genetic background. The bark was removed from the sampling area, after which TW (upper side of the branch) and OW (lower side of the branch) were collected from the same branch using a sharp chisel, as described in Li et al. (2013). NW, which represents the control for stem xylem tissue, was isolated from the same side of the tree at breast height, approximately 1 m above the ground. All samples were approximately 2 × 1 cm and 4–5 mm deep. Samples were collected in the morning, immediately frozen in liquid nitrogen, and stored at −80°C for RNA isolation.

RNA Extraction and Qualification

Nine samples (NW1, NW2, NW3; TW1, TW2, TW3; OW1, OW2, OW3) from 300 days rubber tree reaction wood were used for RNA extraction. A modified cetyltrimethyl ammonium bromide (CTAB) method was used to isolate total RNA according to Chang et al. (1993). Genomic DNA was eliminated by DNase treatment. RNA degradation and contamination were assessed on 1% agarose gels. A NanoPhotometer® spectrophotometer (IMPLEN, CA, United States) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2,100 system (Agilent Technologies, CA, United States) were used to evaluate RNA quality and concentration for further analysis. Only RNA samples with absorption $OD_{260/280}$ ratios from 1.9 to 2.2, $OD_{260/230}$ ratios ≥2.0, and RNA integrity number (RIN) values greater than 6.8

were used for subsequent experiments. Polyadenylated mRNA was enriched using oligo (dT) magnetic beads.

Transcriptome Profiling of Tension Wood, Opposite Wood, and Normal Wood From Rubber Tree Reaction Wood

For Illumina sequencing, fragmentation buffer was added to produce shorter mRNA strands. Single-stranded cDNA was synthesized from the mRNA using random hexamer primers. Double-stranded cDNA was synthesized by adding buffer, dNTPs, and DNA polymerase I. The double-stranded cDNA was purified using AMPure XP beads and subjected to end repair, addition of the poly-A tail, ligation of the sequencing linker, and fragment size selection. Finally, the nine cDNA libraries were subjected to PCR enrichment and sequenced on the Illumina HiSeq 2,500 platform.

Clean reads were obtained by removing low-quality sequence fragments caused by instrument errors, reads with low overall quality, 3' ends with base 10 quality score (Q) < 20 ($Q = -10\log^{error_ratio}$), reads containing N blur, any adapter sequences, and any sequences with <20 nucleotides. The clean reads were aligned to the reference (ref) genome sequence (Liu et al., 2020). The read count of each gene was obtained by mapping the clean reads to the ref genome. The read counts were converted into fragments per kilobase of exon model per million mapped reads (FPKM) values.

DEGs were selected based on the following criteria: $|\log_2 FC| \geq 1$ and P_{adjust} (P_{adj}) < 0.05. All DEGs were mapped to individual terms in the Gene Ontology (GO) database (http://www.geneontology.org/), and the number of genes per term was calculated. GO enrichment analysis was performed using GOseq software to identify significantly enriched terms. Analysis of gene regulatory pathways was conducted using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database (http://www.genome.jp/kegg/pathway.html), MapMan software (v 3.6.0RC1; http://mapman.gabipd.org) was used for the functional pathway analysis (Thimm et al., 2004).

Correlation Networks and Promoter Analysis

Co-expression among genes and TFs was assessed based on the Pearson correlation coefficient calculated in R (version 4.0.1) (**Supplementary Table S1**). The TF–gene interaction networks were visualized using Cytoscape (version 3.7.2; Shannon et al., 2003). Rubber tree gene promoter sequences (2000 bp upstream of the transcription start site in most cases) were retrieved and analyzed. Promoter analysis and prediction of TF binding sites was performed using PlantPAN 3.0 (http://plantpan.itps.ncku.edu.tw/; Chow et al., 2019).

Lignin Quantification

Quantification of lignin content was performed as in Xiong et al. (2005). The sample (1 g) was crushed and then 0.1 g was added into a centrifuge tube. Subsequently, 10 ml of 1% acetic acid solution was added; the sample was mixed and centrifuged at 4,500 rpm for 5 min. The resulting precipitate was washed with 5 ml of 1% acetic acid, followed by addition of 3–4 ml of ethanol and ether mixture (1:1). The mixture was soaked at room temperature for 3 min, and the supernatant was discarded for 3 times. The precipitate in the centrifuge tube was evaporated in a boiling water bath, and 3 ml of 72% sulfuric acid was added to the precipitate. The precipitate was stirred well with a glass rod, and cellulose was dissolved by incubating the sample at room temperature for 16 h. Then, 10 ml distilled water was added to the test tube, stirred well with a glass rod, and placed in a boiling water bath for 5 min. After cooling, 5 ml of distilled water and 0.5 ml of 10% barium chloride solution were added, mixed well, and centrifuged. The precipitate was washed twice with distilled water, and 10 ml of 10% sulfuric acid and 10 ml of 0.1 mol/L potassium dichromate solution were added to the washed lignin precipitate. The test tube was placed in a boiling water bath for 15 min, stirring constantly during the process, and allowed to cool for later use.

All materials from the cooled test tube were transferred to a beaker for titration and the remaining material was washed with 15–20 ml distilled water. Then, 5 ml of 20% KI solution and 1 ml of 0.5% starch solution were added to the beaker and titrated with 0.2 mol/L sodium thiosulfate. Note: Separate titration and add 10 ml of 10% sulfuric acid and 10 ml of 0.1 mol/L potassium dichromate solution as a blank sample.

Lignin content was calculated using the following equation: % lignin = $\frac{k(a-b)}{n\times 48}$, where 48 is the 1 mol of $C_{11}H_{12}O_4$ equivalent to sodium thiosulfate, k is the concentration of sodium thiosulfate (mol/L), a is the volume of sodium thiosulfate consumed in blank titration (ml); b is the volume of sodium thiosulfate consumed by the titration solution (ml); n is the mass of the sample (g).

Validation of Gene Expression by Reverse-Transcription Quantitative PCR (RT-qPCR)

cDNA was synthesized by reverse transcription using 1 μg total RNA from the nine samples (NW1, NW2, NW3; TW1, TW2, TW3; OW1, OW2 and OW3), respectively. Primer Premier v5 software was used to design primers specific to the selected genes (**Table 1**). Five genes were randomly selected from the NW, TW, and OW of rubber tree reaction wood. For RT-qPCR analysis, TB Green Premix Ex Taq II (Tli RNaseH Plus; Takara, Beijing, China) was used following the manufacturer's recommendations. PCR amplification was performed in a preheated (94°C) thermal cycler and incubated at 94°C for 2 min, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. For melting curve analysis, samples were denatured at 95°C for 15 s, then cooled to 60°C (4°C per second). Fluorescence signals were collected at 520 nm continuously from 60°C to 95°C (0.2°C per second). Expression stability of the *40S* (Lertpanyasampatha et al., 2014; Pramoolkit et al., 2014) and *Ubiquitin* (Li et al., 2011; Chao et al., 2016) rubber tree genes has been previously evaluated and both were confirmed as suitable internal control genes. Thus, these two genes served as internal controls for normalization. Expression levels of the DEGs were calculated using the $2^{-\triangle\triangle Ct}$ method against the internal control

TABLE 1 | Oligonucleotide primers used for RT-qPCR in this study.

Gene id	Forward primer	Reverse primer
Ubiquitin	ATGCAGCACCGGGAATAAGT	GGTCATCAGGGTTTGGAGCA
40S	TGCAGAACGAAGAGGGTCAG	TTGTTGGCGGAGTTCAACCT
gene-GH714_003817	GCTAATGTCACGGGCTCCAT	TGGGATCGACCCAGATAACG
gene-GH714_035841	ATTCTGGGTCTGGTCTCCCA	CCCCTCCCCAATTCGGTTTT
gene-GH714_043603	GCGAAATCGCTCCGATCAAC	CGAGTCACTCTGGGTTCCAC
gene-GH714_019256	GCTCATGTGGCTTTCATGGC	GAACCCTTCCCCTTGCCTAC
gene-GH714_027628	GGTGAGCACATGCCTTGTTG	GCCCTGAGTGAACGAAGTGA

TABLE 2 | Summary of RNA-seq data from normal (NW), tension (TW), and opposite wood (OW).

Sample	Raw reads	Clean reads	Mapped to genome	Mapping ratio (%)
TW1	43,083,708	42,256,214	39,392,727	93.22
TW2	40,537,540	39,876,902	37,237,975	93.38
TW3	42,349,100	41,386,550	38,704,999	93.52
OW1	40,941,484	40,039,822	37,388,130	93.38
OW2	44,907,906	43,987,054	40,848,282	92.86
OW3	41,600,314	40,578,896	37,952,472	93.53
NW1	42,924,560	41,990,498	38,750,485	92.28
NW2	41,406,942	40,674,052	37,794,460	92.92
NW3	42,488,914	41,714,582	39,000,451	93.49

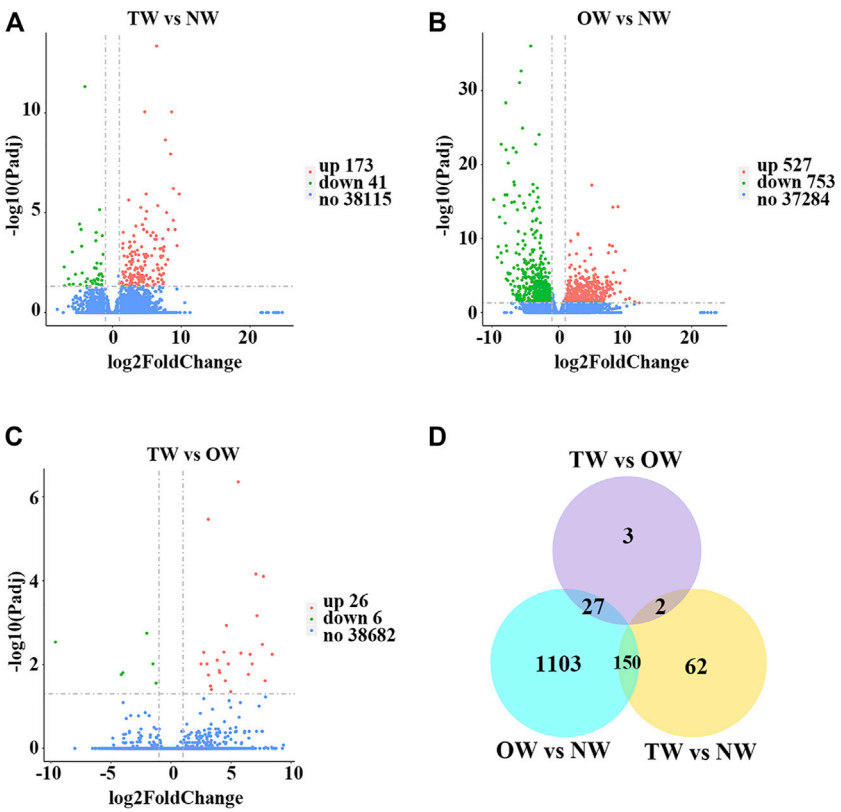


FIGURE 1 | Global analysis of gene expression in different xylem tissues of rubber tree reaction wood. Volcano plots illustrate differentially expressed genes identified in TW vs. NW **(A)**, OW vs. NW **(B)**, and TW vs. OW **(C)** as individual dots. Red represents up-regulation and green represents down-regulation. **(D)** Venn diagram showing the total number of differentially expressed genes identified in TW vs. NW, OW vs. NW, and TW vs. OW, and overlap among these comparison groups.

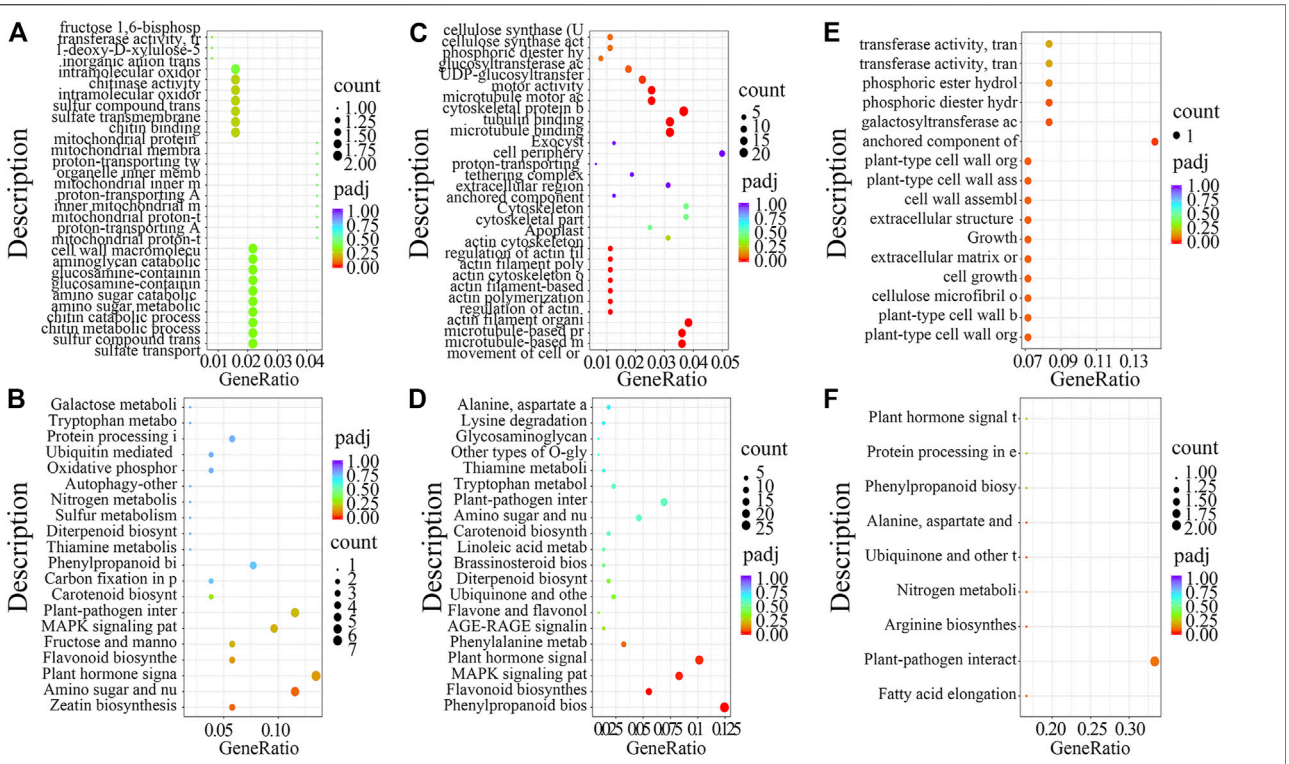


FIGURE 2 | GO and KEGG enrichment analysis of differentially expressed genes (DEGs) from the RNA-seq dataset. GO analysis of DEGs identified in TW vs. NW **(A)**, OW vs. NW **(C)**, and TW vs. OW **(E)**. KEGG analysis of DEGs identified in TW vs NW **(B)**, OW vs NW **(D)**, and TW vs OW **(F)**, respectively.

genes (Schmittgen and Livak, 2008). Three technical replicates per sample were analyzed to ensure reproducibility and reliability.

RESULTS

Global Transcriptome Analysis of RNA-Seq Data

To evaluate whether the RNA-seq data were sufficient for further analysis, we first assessed their global quality. The RNA-seq experiment generated a total of 125,970,348 (TW), 127,449,704 (OW), and 126,820,416 (NW) raw reads. After trimming, 123,519,666 (TW), 124,605,772 (OW), and 124,379,132 (NW) clean reads remained (**Table 2**). Among the total clean reads, 115,335,701 (TW), 116,188,884 (OW), and 115,545,396 (NW) were mapped to the rubber tree genome with mapping ratios of 93.37% (TW), 93.26% (OW), and 92.90% (NW) (**Table 2**). Based on previous studies (Geraldes et al., 2011; Chen et al., 2015a), these results indicated that our RNA-seq results detected most expressed genes were sufficient for subsequent quantitative analysis. To measure changes in gene expression and find key genes involved in reaction wood formation, we further selected DEGs that met the following statistical significance criteria: $|\log_2 FC| \geq 1$ and $P_{adj} \leq 0.05$. In total, 214 (TW vs. NW; 173 up-regulated and 41 down-regulated), 1,280 (OW vs. NW; 527 up-regulated and 753 down-regulated), and 32 (TW vs. OW; 26 up-regulated and 6 down-regulated) DEGs were identified in rubber tree reaction wood (**Figures 1A–C**). After removing repetitive genes, this analysis identified 1,347 genes that were significantly differentially expressed in the TW, OW and NW xylem tissues (**Figure 1D**). The DEGs obtained from three comparison groups were used for subsequent analysis.

GO and KEGG Enrichment Analyses of DEGs That Participate in Reaction Wood Formation

RNA-seq provided an overview of genes that are differentially expressed during reaction wood formation. To better understand the function of these DEGs and their associated biological processes, GO and KEGG enrichment analyses were carried out for the different comparison groups shown in **Figure 1D**.

First, for DEGs in TW vs. NW, GO analysis revealed enrichment for intramolecular oxidoreductase activity, cell wall macromolecule metabolic process, and sulfate transmembrane transporter activity (**Figure 2A**), while KEGG enrichment analysis revealed that zeatin biosynthesis and plant hormone signal transduction pathways were significantly enriched (**Figure 2B**). Next, GO analysis of the DEGs from OW vs. NW showed that biological processes including cellulose synthase (UDP-forming) activity, cellulose synthase activity,

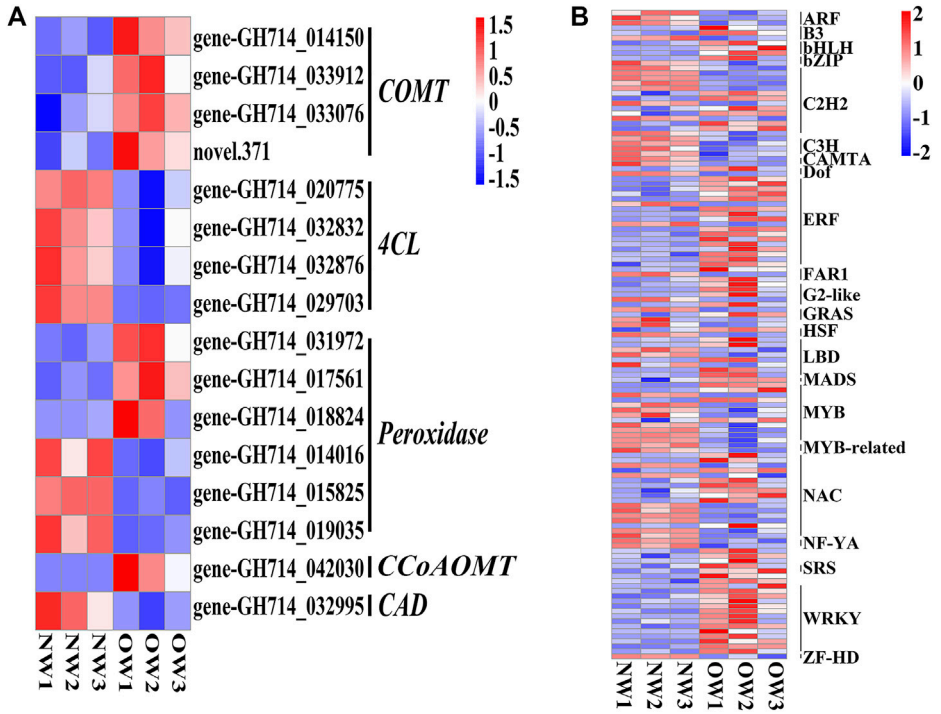


FIGURE 3 | Differentially expressed genes (DEGs) and transcription factors in OW vs. NW. Heatmap showing the expression of the DEGs **(A)** and transcription factors **(B)** identified in OW vs. NW. Columns represent three different samples per tissue type and rows represent different transcripts. Each square represents a transcript and the color indicates the level of expression; red represents up-regulation and blue represents down-regulation.

and UDP-glucosyltransferase activity were significantly enriched (**Figure 2C**), and KEGG analysis showed that phenylpropanoid biosynthesis and plant hormone signal transduction pathway were significantly enriched (**Figure 2D**). Last, GO analysis of the DEGs from TW vs. OW showed that cell growth, cell wall assembly, and cellulose microfibril organization were significantly enriched (**Figure 2E**). KEGG analysis of DEGs from TW vs. OW showed enrichment for plant–pathogen interaction pathway (**Figure 2F**).

Taken together, results of the GO and KEGG analyses of DEGs from different treatment groups suggest that zeatin biosynthesis, plant hormone signal transduction, phenylpropanoid biosynthesis, and plant–pathogen interaction pathway might function in the formation of reaction wood.

Expression Analysis of DEGs Involved in Phenylpropanoid Biosynthesis and Lignin Quantification

The phenylpropanoid biosynthesis pathway is responsible for lignin production and thus plays an important role in wood formation. KEGG enrichment analysis indicated that DEGs from OW vs. NW are involved in the phenylpropanoid biosynthesis pathway. A total of 16 transcripts of five genes from OW vs. NW are related to lignin synthesis (**Figure 3A**). The *4CL* gene was down-regulated in OW; 4CL converts coumarate, caffeate, ferulate, 5-hydroxyferulate, and sinapate to coumaroyl-CoA, caffeoyl-CoA, feruloyl-CoA, 5-hydroxyferuloyl-CoA, sinapoyl-CoA (**Figure 4**). The *CAD* gene was also down-regulated; CAD converts coumaryl aldehyde, caffeyl aldehyde, coniferaldehyde, 5-hydroxy-coniferaldehyde, and sinapaldehyde to coumaryl alcohol, caffeyl alcohol, coniferyl alcohol, 5-hydroxyconiferyl alcohol, and sinapyl alcohol (**Figure 4**). Down-regulation of *4CL* and *CAD* in OW vs. NW may imply that these two conversion steps are suppressed in reaction wood formation.

The *COMT* and *CCoAOMT* genes were significantly up-regulated in OW vs. NW. These two genes are involved in eight steps of lignin synthesis, including in the conversion of caffeyl alcohol to coniferyl alcohol and 5-hydroxyconiferyl alcohol to sinapyl alcohol (**Figure 4**). Up-regulation of *COMT* and *CCoAOMT* would promote the synthesis of G- and H-lignin (**Figure 4**), potentially leading to higher lignin content in OW than NW. Furthermore, three out of six transcripts encoding peroxidases were significantly up-regulated in OW, while the other three were significantly up-regulated in NW (**Figure 4**). These results suggest that changes in *POD* gene expression may not affect lignin content.

To link the gene expression data with effects on lignin synthesis, lignin levels were quantified in NW, TW, and OW. The average lignin content of each tissue was 263.54 mg/g in OW, 248.68 mg/g in NW, and 240.06 mg/g in TW. There were no statistically significant differences detected in the lignin content of these different tissues (**Table 3**). These results are consistent with the changes in expression levels of DEGs in the phenylpropanoid biosynthesis pathway.

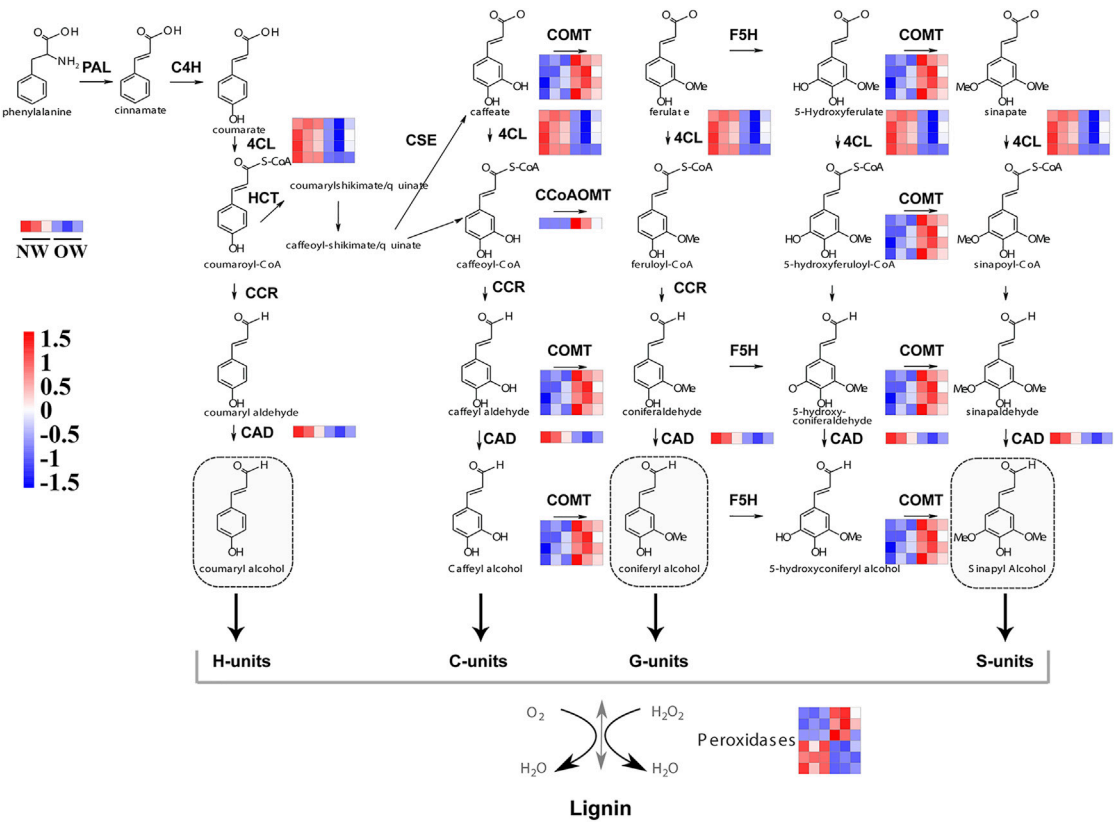


FIGURE 4 | Expression of genes in phenylpropanoid biosynthesis pathway in OW vs. NW. Red represents up-regulation and blue represents down-regulation. PAL, Phenylalanine ammonia-lyase; C4H, Cinnamate 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; F5H, Ferulate 5-hydroxylase; CCoAOMT, Caffeoyl-CoA O-methyltransferase; CCR, Cinnamoyl-CoA reductase; CAD, Cinnamyl alcohol dehydrogenase; COMT, Caffeate O-methyltransferase; HCT, hydroxycinnamoyltransferase; POD, peroxidase.

TABLE 3 | Lignin content of normal (NW), tension (TW), and opposite wood (OW).

Sample	Lignin (mg/g)
NW1	228.46
NW2	267.66
NW3	249.91
TW1	244.59
TW2	235.87
TW3	239.72
OW1	276.27
OW2	258.26
OW3	256.09

Transcription Factors Mediated Transcriptional Regulatory Networks Involved in Phenylpropanoid Biosynthesis

DEGs encoding TFs in each comparison group were identified to discover potential transcriptional regulatory networks in phenylpropanoid biosynthesis (**Figure 3B**). In total, 129 TFs and 16 DEGs associated with phenylpropanoid biosynthesis from OW vs. NW were used to construct a co-expression network using Pearson correlation coefficients.

Significantly co-expressed TF–gene pairs ($|cor| \geq 0.9$ and $p < 0.05$) were selected to construct the transcriptional regulatory network (**Figure 5**). In the network, 27 TF families regulate phenylpropanoid biosynthesis-related DEGs. *COMT*, *4CL*, *POD*, *CCoAOMT*, and *CAD* were regulated by 20, 20, 24, 10, and 12 TF families, respectively (**Figure 5**). Notably, the MYB, C2H2, C3H, and NAC families could regulate all the DEGs involved in phenylpropanoid biosynthesis (**Figure 5**). We evaluated the promoter sequence of genes co-expressed with TFs for potential TF binding sites. Promoter analysis showed that MYB, C2H2, and NAC TF families could bind to the promoter sequence of all DEGs involved in phenylpropanoid biosynthesis (**Supplementary Table S2**). These results suggest that these three TF families are likely key players in rubber tree reaction wood formation by regulating numerous downstream genes involved in phenylpropanoid biosynthesis.

Validation of RNA-Seq Gene Expression Data by RT-qPCR

The expression patterns of most genes in the NW, TW and OW showed similar trends between the high-throughput RNA-seq

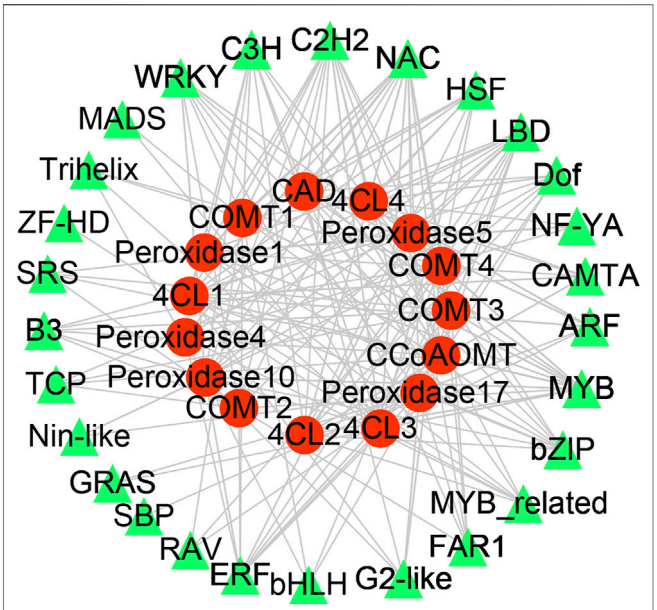


FIGURE 5 | Transcription factor-gene network in rubber tree lignin biosynthesis. Red circles represent genes and green triangles represent transcription factors.

data and RT-qPCR data (**Figure 6**). Although the relative expression level calculated by sequencing did not exactly match the expression values detected by RT-qPCR, the expression profiles were mostly consistent for the genes tested. These results confirm the reliability of the gene expression values generated from the RNA-seq data.

DISCUSSION

A Model for Reaction Wood Formation Identifies Genes Involved in Wood Formation of Rubber Tree and Other Tree Species

To explore the mechanisms of rubber tree reaction wood formation, wood formation-related genes were identified and their expression patterns were investigated. In this study, we identified 214, 1,280, and 32 DEGs from three comparison groups (TW vs. NW, OW vs. NW, and TW vs. OW), respectively (**Figure 1**). Among them, 62, 1,103, and 3 DEGs were unique to TW vs. NW, OW vs. NW, and TW vs. OW, respectively. There were no DEGs common to the three comparison groups (**Figure 1D**), which might indicate that these tissue-specific DEGs function in wood formation of different xylem tissues. In addition, GO and KEGG enrichment analyses showed that the DEGs from different comparison groups have unique functions and participate in different pathways (**Figure 2**), further confirming that these DEGs perform distinct functions in different xylem tissues.

Based on the GO and KEGG enrichment analyses, our subsequent work focused on genes involved in phenylpropanoid biosynthesis, as well as TFs identified in the list of DEGs from OW vs. NW (**Figure 2D**, **Figure 3**). The heatmap and interaction network showed that phenylpropanoid biosynthesis-related genes and TFs had divergent expression patterns in OW vs. NW, which may reflect the diverse functions of these genes and the complexity of the regulatory network associated with wood formation (**Figures 3**, **5**). These results further demonstrate that wood development is a complex biological process. Genes involved in wood formation may be effectively identified through our model of reaction wood formation.

Previous research has shown that many factors are involved in reaction wood formation, such as non-coding RNA, single nucleotide polymorphisms etc. Xiao et al. (2020) previously reported that long non-coding RNAs (lncRNAs) might participate in *Catalpa bungei* tension wood formation by regulating genes involved in indoleacetic acid (IAA) and abscisic acid (ABA) synthesis. Similarly, microRNA (miRNA)- and lncRNA-mediated regulatory networks widely participate in reaction wood formation of *Populus tomentosa (Pto)*, and single nucleotide polymorphisms of miRNAs and their target genes also influence this process (Chen et al., 2015b; Chen et al., 2016). Moreover, Liu et al. (2021) induced tension wood in stems of black cottonwood (*Populus trichocarpa*) and established a transcriptional regulatory network in which *PtrHSFB3-1* and *PtrMYB092* directly activate eight and 11 monolignol genes that participate in reaction wood formation, respectively. Future research should focus on characterizing the roles of non-coding RNAs and the candidate genes identified through our RNA-seq analysis in rubber tree reaction wood formation.

Characterization of Genes Associated With Phenylpropanoid Biosynthesis in Rubber Tree by Transcriptome Analysis

Lignin synthesis is positively associated with the expression levels of *4CL*, *COMT*, *CCoAOMT*, *CAD*, and *POD* (Do et al., 2007; Vanholme et al., 2008; Wagner et al., 2011; Hu et al., 2017; Chanoca et al., 2019). In this study, *COMT* and *CCoAOMT* were up-regulated in OW vs. NW, suggesting that the lignin biosynthetic process may be enhanced in this tissue. However, *4CL* and *CAD* were down-regulated, indicating that the reaction steps involving *4CL* and *CAD* may be inhibited in OW versus NW. Moreover, three *peroxidase* gene transcripts were up-regulated in OW, while three were up-regulated in NW. This indicates that *peroxidase* genes likely generate different transcripts in different rubber tree xylem tissues to maintain lignin biosynthesis.

Previous studies showed that repressing steps of the lignin biosynthetic pathway, from PAL to CAD, may lead to reduced lignin content (Chanoca et al., 2019). In addition, studies have shown that single or double Arabidopsis knockout mutants in several *peroxidase* genes typically exhibit minor but noticeable reductions in lignin content and/or altered lignin composition in inflorescence stems, suggesting that changes in *peroxidase* expression levels may not strongly affect lignin synthesis

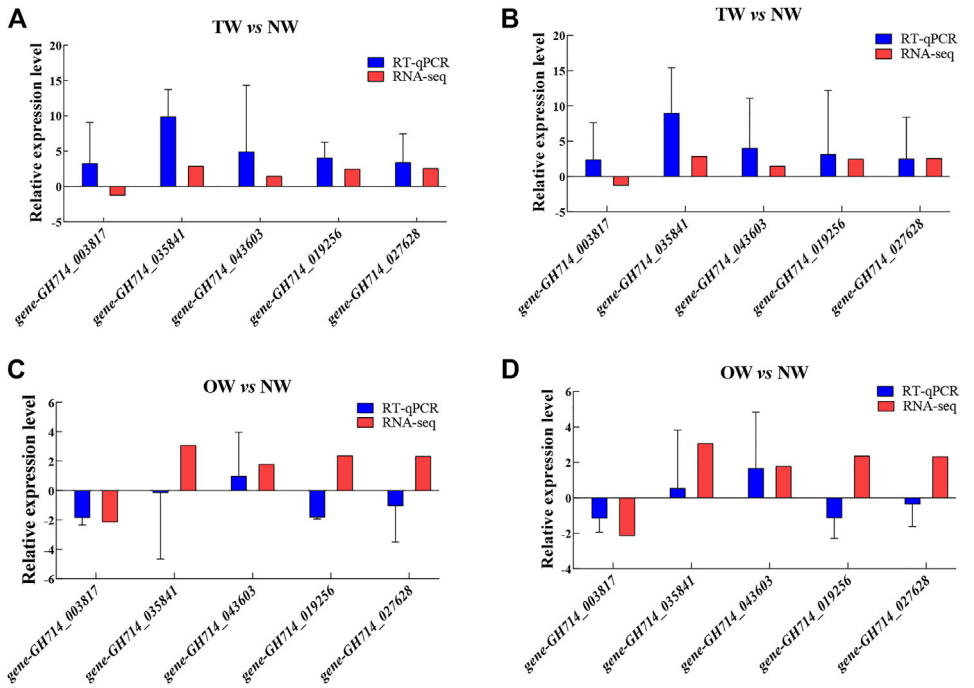


FIGURE 6 | Relative expression of five genes in TW vs. NW and OW vs. NW. Expression levels were determined by RT-qPCR using the *40S* gene **(A and C)** or the *Ubiquitin* gene **(B and D)** as the internal control for normalization. Error bars indicate standard deviation of three biological replicates.

(Herrero et al., 2013; Barros et al., 2015). Thus, changes in *peroxidase* gene levels observed in this work likely did not affect lignin content between OW and NW. Overall, the lack of a significant difference in lignin content among the different tissue types in this study may be explained by the opposite expression patterns of *4CL*, *COMT*, *CCoAOMT*, *CAD*, and by the minor contribution of the *peroxidase* genes to lignin levels.

NAC and MYB Transcription Factor Families Participate in Rubber Tree Lignin Synthesis

Wood formation is an essential yet complex biological process arising from plant secondary growth. Important transcription factors involved in secondary growth have been identified, such as members of the MYB and NAC TF families (Demura and Fukuda, 2007; Zhong and Ye, 2009). In this study, 129 TFs from 27 TF families were responsive to reaction wood treatment (**Figures 3**, **5**). Among them, TFs from the MYB and NAC families could regulate all DEGs involved in phenylpropanoid biosynthesis and might play pivotal roles in rubber tree reaction wood formation. Previous studies have shown that overexpression of MYB transcription factors *PtoMYB92*, *PtoMYB216*, and *PtoMYB74* could up-regulate genes involved in lignin biosynthesis and promote the formation of additional xylem layers, thicker xylem cell walls, as well as ectopic lignin deposition; these overexpression plants accumulated 13–50% more lignin (Tian et al., 2013; Li et al., 2015; Li et al., 2018). Similarly, overexpression of *NAC141* in Arabidopsis resulted in enhanced expression of lignin biosynthesis genes, stronger lignification, larger xylem, and higher lignin content compared with wild-type plants (Sun et al., 2021). These findings support the TF–gene regulatory network for rubber tree wood formation generated in this study, in which MYB and NAC TFs likely play important roles.

CONCLUSION

Identification of DEGs through RNA-seq analysis on mature xylem from TW, OW, and NW of rubber tree reaction wood provided insight into the molecular basis of lignin biosynthesis. Our work demonstrated that zeatin biosynthesis, plant hormone signal transduction, phenylpropanoid biosynthesis, and plant–pathogen interaction pathways likely influence reaction wood development. A transcriptional regulatory network analysis revealed that three TF families (MYB, C2H2, and NAC) could regulate all DEGs involved in phenylpropanoid biosynthesis and play important roles in lignin biosynthesis for rubber tree reaction wood formation.

Taken together, the findings presented here advance our understanding of the regulation of phenylpropanoid biosynthesis, a critical pathway for lignin production in perennial trees. The publicly available transcriptomic dataset provides essential information for further transcriptomic, genomic, and functional genomics research in rubber tree. Overall, the results generated in this study will serve as an important resource for future studies on this economically important tropical tree crop.

DATA AVAILABILITY STATEMENT

The transcriptome data for 300-day rubber tree reaction wood reported in this paper has been deposited at the Genome Sequence Archive (Wang et al., 2017) in BIG Data Center (BIG Data Center Members, 2019), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers CRA004818, and are publicly available at https://bigd.big.ac.cn/gsa.

AUTHOR CONTRIBUTIONS

JC and ZT designed the research. XM, YW, JL, NJ, XZ, and YZ performed the research. All authors analyzed and interpreted the data. JC and XM wrote the paper. All authors commented on the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31960321), the 2019 Hainan Provincial Basic and Applied Research (Natural Science), High-level Talents Program (Hainan Provincial Natural Science Foundation of China) (no. 2019RC160) and the Scientific Research Fund Project of Hainan University (KYQD (ZR)1830), Hainan Provincial Natural Science Foundation of China (No. 420RC532, No. 319QN159), Scientific Research Fund of Hainan University (No. KYQD(ZR)1950).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.763841/full#supplementary-material

REFERENCES

Barros, J., Serk, H., Granlund, I., and Pesquet, E. (2015). The Cell Biology of Lignification in Higher Plants. *Ann. Bot.* 115, 1053–1074. doi:10.1093/aob/mcv046

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin Biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi:10.1146/annurev.arplant.54.031902.134938

Chang, S., Puryear, J., and Cairney, J. (1993). A Simple and Efficient Method for Isolating RNA from Pine Trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi:10.1007/bf02670468

Chanoca, A., De Vries, L., and Boerjan, W. (2019). Lignin Engineering in Forest Trees. *Front. Plant Sci.* 10, 912. doi:10.3389/fpls.2019.00912

Chao, J., Yang, S., Chen, Y., and Tian, W.-M. (2016). Evaluation of Reference Genes for Quantitative Real-Time PCR Analysis of the Gene Expression in Laticifers on the Basis of Latex Flow in Rubber Tree (*Hevea Brasiliensis* Muell. Arg.). *Front. Plant Sci.* 7, 1–9. doi:10.3389/fpls.2016.01149

Chen, J., Chen, B., and Zhang, D. (2015a). Transcript Profiling of *Populus Tomentosa* Genes in normal, Tension, and Opposite wood by RNA-Seq. *BMC Genomics* 16, 1–16. doi:10.1186/s12864-015-1390-y

Chen, J., Quan, M., and Zhang, D. (2015b). Genome-wide Identification of Novel Long Non-coding RNAs in *Populus Tomentosa* Tension wood, Opposite wood and normal wood Xylem by RNA-Seq. *Planta* 241, 125–143. doi:10.1007/s00425-014-2168-1

Chen, J., Xie, J., Chen, B., Quan, M., Li, Y., Li, B., et al. (2016). Genetic Variations and miRNA-Target Interactions Contribute to Natural Phenotypic Variations in *Populus*. *New Phytol.* 212, 150–160. doi:10.1111/nph.14040

Chow, C.-N., Lee, T.-Y., Hung, Y.-C., Li, G.-Z., Tseng, K.-C., Liu, Y.-H., et al. (2019). Plantpan3.0: A New and Updated Resource for Reconstructing Transcriptional Regulatory Networks from Chip-Seq Experiments in Plants. *Nucleic Acids Res.* 47, D1155–D1163. doi:10.1093/nar/gky1081

Demura, T., and Fukuda, H. (2007). Transcriptional Regulation in wood Formation. *Trends Plant Sci.* 12, 64–70. doi:10.1016/j.tplants.2006.12.006

Do, C.-T., Pollet, B., Thévenin, J., Sibout, R., Denoue, D., Barrière, Y., et al. (2007). Both caffeoyl Coenzyme A 3-O-Methyltransferase 1 and Caffeic Acid O-Methyltransferase 1 Are Involved in Redundant Functions for Lignin, Flavonoids and Sinapoyl Malate Biosynthesis in Arabidopsis. *Planta* 226, 1117–1129. doi:10.1007/s00425-007-0558-3

Gallego-Giraldo, L., Shadle, G., Shen, H., Barros-Rios, J., Fresquet Corrales, S., Wang, H., et al. (2016). Combining Enhanced Biomass Density with Reduced Lignin Level for Improved Forage Quality. *Plant Biotechnol. J.* 14, 895–904. doi:10.1111/pbi.12439

Geraldes, A., Pang, J., Thiessen, N., Cezard, T., Moore, R., Zhao, Y., et al. (2011). SNP Discovery in Black cottonwood (*Populus trichocarpa*) by Population Transcriptome Resequencing. *Mol. Ecol. Resour.* 11, 81–92. doi:10.1111/j.1755-0998.2010.02960.x

Gui, J., Shen, J., and Li, L. (2011). Functional Characterization of Evolutionarily Divergent 4-Coumarate:Coenzyme A Ligases in Rice. *Plant Physiol.* 157, 574–586. doi:10.1104/pp.111.178301

Han, X.-y., Mao, L.-c., Lu, W.-j., Tao, X.-y., Wei, X.-p., and Luo, Z.-s. (2018). Abscisic Acid Induces Differential Expression of Genes Involved in Wound-Induced Suberization in Postharvest Tomato Fruit. *J. Integr. Agric.* 17, 2670–2682. doi:10.1016/S2095-3119(18)62142-2

Herrero, J., Fernández-Pérez, F., Yebra, T., Novo-Uzal, E., Pomar, F., Pedreño, M. Á., et al. (2013). Bioinformatic and Functional Characterization of the Basic Peroxidase 72 from *Arabidopsis thaliana* Involved in Lignin Biosynthesis. *Planta* 237, 1599–1612. doi:10.1007/s00425-013-1865-5

Hu, D., Liu, X. B., She, H. Z., Gao, Z., Ruan, R. W., Wu, D. Q., et al. (2017). The Lignin Synthesis Related Genes and Lodging Resistance of *Fagopyrum Esculentum*. *Biol. Plant* 61, 138–146. doi:10.1007/s10535-016-0685-4

Jahan, M. S., Haider, M. M., Rahman, M., Biswas, D., Misbahuddin, M., and Mondal, G. K. (2011). Chemical Pulping: Evaluation of Rubber wood (Hevea Brasiliensis) as a Raw Material for Kraft Pulping. *Nord. Pulp Pap. Res. J.* 26, 258–262. doi:10.3183/npprj-2011-26-03-p258-262

Kim, Y.-H., Kim, C. Y., Song, W.-K., Park, D.-S., Kwon, S.-Y., Lee, H.-S., et al. (2008). Overexpression of Sweetpotato *Swpa4* Peroxidase Results in Increased Hydrogen Peroxide Production and Enhances Stress Tolerance in Tobacco. *Planta* 227, 867–881. doi:10.1007/s00425-007-0663-3

Lertpanyasampatha, M., Viboonjun, U., Kongsawadworakul, P., Chrestin, H., and Narangajavana, J. (2014). Differential Expression of microRNAs and Their Targets Reveals a Possible Dual Role in Physiological Bark Disorder in Rubber Tree. *J. Plant Physiol.* 171, 1117–1126. doi:10.1016/j.jplph.2014.05.001

Li, C., Ma, X., Yu, H., Fu, Y., and Luo, K. (2018). Ectopic Expression of *PtoMYB74* in Poplar and Arabidopsis Promotes Secondary Cell Wall Formation. *Front. Plant Sci.* 9, 1–13. doi:10.3389/fpls.2018.01262

Li, C., Wang, X., Ran, L., Tian, Q., Fan, D., and Luo, K. (2015). PtoMYB92 Is a Transcriptional Activator of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation inPopulus Tomentosa. *Plant Cel Physiol* 56, 2436–2446. doi:10.1093/pcp/pcv157

Li, H., Qin, Y., Xiao, X., and Tang, C. (2011). Screening of Valid Reference Genes for Real-Time RT-PCR Data Normalization in *Hevea Brasiliensis* and Expression Validation of a Sucrose Transporter Gene HbSUT3. *Plant Sci.* 181, 132–139. doi:10.1016/j.plantsci.2011.04.014

Li, H., Yang, Y., Wang, Z., Guo, X., Liu, F., Jiang, J., et al. (2016). *BpMADS12* Gene Role in Lignin Biosynthesis of *Betula Platyphylla Suk* by Transcriptome Analysis. *J. For. Res.* 27, 1111–1120. doi:10.1007/s11676-016-0229-y

Li, X., Yang, X., and Wu, H. X. (2013). Transcriptome Profiling of Radiata pine Branches Reveals New Insights into Reaction wood Formation with Implications in Plant Gravitropism. *BMC Genomics* 14, 768. doi:10.1186/1471-2164-14-768

Liu, B., Liu, J., Yu, J., Wang, Z., Sun, Y., Li, S., et al. (2021). Transcriptional Reprogramming of Xylem Cell wall Biosynthesis in Tension wood. *Plant Physiol.* 186, 250–269. doi:10.1093/plphys/kiab038

Liu, J., Shi, C., Shi, C.-C., Li, W., Zhang, Q.-J., Zhang, Y., et al. (2020). The Chromosome-Based Rubber Tree Genome Provides New Insights into Spurge Genome Evolution and Rubber Biosynthesis. *Mol. Plant* 13, 336–350. doi:10.1016/j.molp.2019.10.017

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: Biosynthesis and Biological Functions in Plants. *Int. J. Mol Sci.* 19, 335. doi:10.3390/ijms19020335

Ohtani, M., and Demura, T. (2019). The Quest for Transcriptional Hubs of Lignin Biosynthesis: beyond the NAC-MYB-Gene Regulatory Network Model. *Curr. Opin. Biotechnol.* 56, 82–87. doi:10.1016/j.copbio.2018.10.002

Pramoolkit, P., Lertpanyasampatha, M., Viboonjun, U., Kongsawadworakul, P., Chrestin, H., and Narangajavana, J. (2014). Involvement of Ethylene-Responsive microRNAs and Their Targets in Increased Latex Yield in the Rubber Tree in Response to Ethylene Treatment. *Plant Physiol. Biochem.* 84, 203–212. doi:10.1016/j.plaphy.2014.09.016

Priyadarshan, P. M. (2017). Refinements to *Hevea* Rubber Breeding. *Tree Genet. Genomes* 13, 1–17. doi:10.1007/s11295-017-1101-8

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing Real-Time PCR Data by the Comparative CT Method. *Nat. Protoc.* 3, 1101–1108. doi:10.1038/nprot.2008.73

Severo, E. T. D., Calonego, F. W., Sansígolo, C. A., and Bond, B. (2016). Changes in the Chemical Composition and Decay Resistance of Thermally-Modified *Hevea Brasiliensis* Wood. *PLoS One* 11, e0151353. doi:10.1371/journal.pone.0151353

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, Y., Jiang, C., Jiang, R., Wang, F., Zhang, Z., and Zeng, J. (2021). A Novel NAC Transcription Factor from *Eucalyptus*, EgNAC141, Positively Regulates Lignin Biosynthesis and Increases Lignin Deposition. *Front. Plant Sci.* 12, 1–10. doi:10.3389/fpls.2021.642090

Teoh, Y. P., Don, M. M., and Ujang, S. (2011). Assessment of the Properties, Utilization, and Preservation of Rubberwood (*Hevea Brasiliensis*): A Case Study in Malaysia. *J. Wood Sci.* 57, 255–266. doi:10.1007/s10086-011-1173-2

Thimm, O., BlĘsing, O., Gibon, Y., Nagel, A., Meyer, S., KrĘger, P., et al. (2020). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi:10.1111/j.1365-313X.2004.02016.x

Tian, Q., Wang, X., Li, C., Lu, W., Yang, L., Jiang, Y., et al. (2013). Functional Characterization of the Poplar R2R3-MYB Transcription Factor PtoMYB216 Involved in the Regulation of Lignin Biosynthesis during Wood Formation. *PLoS One* 8, e76369. doi:10.1371/journal.pone.0076369

Vanholme, R., Morreel, K., Ralph, J., and Boerjan, W. (2008). Lignin Engineering. *Curr. Opin. Plant Biol.* 11, 278–285. doi:10.1016/j.pbi.2008.03.005

Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., et al. (2011). *CCoAOMT* Suppression Modifies Lignin Composition in *Pinus Radiata*. *Plant J.* 67, 119–129. doi:10.1111/j.1365-313X.2011.04580.x

Wang, Y.-H., Wu, X.-J., Sun, S., Xing, G.-M., Wang, G.-L., Que, F., et al. (2018). DcC4H and DcPER Are Important in Dynamic Changes of Lignin Content in Carrot Roots under Elevated Carbon Dioxide Stress. *J. Agric. Food Chem.* 66, 8209–8220. doi:10.1021/acs.jafc.8b02068

Wang, Y.-X., Teng, R.-M., Wang, W.-L., Wang, Y., Shen, W., and Zhuang, J. (2019). Identification of Genes Revealed Differential Expression Profiles and Lignin Accumulation during Leaf and Stem Development in tea Plant (*Camellia Sinensis* (L.) O. Kuntze). *Protoplasma* 256, 359–370. doi:10.1007/s00709-018-1299-9

Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., et al. (2017). GSA: Genome Sequence Archive *. *Genomics, Proteomics & Bioinformatics* 15, 14–18. doi:10.1016/j.gpb.2017.01.001

Whetten, R., and Sederoff, R. (1995). Lignin Biosynthesis. *Plant Cell* 7, 1001–1013. doi:10.1105/tpc.7.7.1001

Xiao, Y., Yi, F., Ling, J., Yang, G., Lu, N., Jia, Z., et al. (2020). Genome-wide Analysis of lncRNA and mRNA Expression and Endogenous Hormone Regulation during Tension wood Formation in *Catalpa Bungei*. *BMC Genomics* 21, 1-16. doi:10.1186/s12864-020-07044-5

Xiong, S. M., Zuo, X. F., and Zhu, Y. Y. (2005). Determination of Cellulose, Hemi-Cellulose and Lignin in Rice Hull. *Cereal Feed Ind.* 8, 40–41.

Yan, B., Zhang, Z., Zhang, P., Zhu, X., Jing, Y., Wei, J., et al. (2019). Nitric Oxide Enhances Resistance against Black Spot Disease in Muskmelon and the Possible Mechanisms Involved. *Scientia Horticulturae* 256, 108650. doi:10.1016/j.scienta.2019.108650

You, T.-T., Mao, J.-Z., Yuan, T.-Q., Wen, J.-L., and Xu, F. (2013). Structural Elucidation of the Lignins from Stems and Foliage of *Arundo donax* Linn. *J. Agric. Food Chem.* 61, 5361–5370. doi:10.1021/jf401277v

Zeng, J.-k., Li, X., Zhang, J., Ge, H., Yin, X.-r., and Chen, K.-s. (2016). Regulation of Loquat Fruit Low Temperature Response and Lignification Involves Interaction of Heat Shock Factors and Genes Associated with Lignin Biosynthesis. *Plant Cel Environ.* 39, 1780–1789. doi:10.1111/pce.12741

Zhao, Q., and Dixon, R. A. (2011). Transcriptional Networks for Lignin Biosynthesis: More Complex Than We Thought? *Trends Plant Sci.* 16, 227–233. doi:10.1016/j.tplants.2010.12.005

Zhong, R., and Ye, Z.-H. (2009). Transcriptional Regulation of Lignin Biosynthesis. *Plant Signaling Behav.* 4, 1028–1034. doi:10.4161/psb.4.11.9875

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.

Copyright © 2021 Meng, Wang, Li, Jiao, Zhang, Zhang, Chen and Tu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fgene.2021.756172

Asymmetric Divergence in Transmitted SNPs of DNA Replication/Transcription and Their Impact on Gene Expression in Polyploid *Brassica napus*

Minqiang Tang[1], Juanling Li[1], Xu Hu[1], Lu Sun[1], MMU Helal[2], Jianguo Chen[3] and Yuanyuan Zhang[2]**

[1]Key Laboratory of Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Ministry of Education), College of Forestry, Hainan University, Haikou, China, [2]The Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture, Oil Crops Research Institute of Chinese, Academy of Agricultural Sciences, Wuhan, China, [3]School of Life Sciences, Hubei University, Wuhan, China

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Qing-Yong Yang,
Huazhong Agricultural University, China
Yang Xiang,
Guizhou Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Yuanyuan Zhang
zhangyy@caas.cn
Jianguo Chen
chenjg23@hotmail.com

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

Received: *10 August 2021*
Accepted: *18 October 2021*
Published: *11 November 2021*

Citation:
Tang M, Li J, Hu X, Sun L, Helal MMU, Chen J and Zhang Y (2021) Asymmetric Divergence in Transmitted SNPs of DNA Replication/Transcription and Their Impact on Gene Expression in Polyploid Brassica napus. *Front. Genet. 12:756172. doi: 10.3389/fgene.2021.756172*

The marked increase in plant genomic data has provided valuable resources for investigating the dynamic evolution of duplicate genes in polyploidy. *Brassica napus* is an ideal model species for investigating polyploid genome evolution. The present study comprehensively analyzed DNA and RNA variation of two representative *B. napus* inbredlines, Zhongshuang11 and Zhongyou821, and we investigated gene expression levels of A_n and C_n subgenomes in multiple tissues of the two lines. The distribution of transmitted single nucleotide polymorphisms (SNPs) was significantly different in two subgenomes of *B. napus*. Gene expression levels were significantly negatively correlated with number of variations in replication and transcription of the corresponding genes, but were positively correlated with the ratios of transmitted SNPs from DNA to RNA. We found a higher density of SNP variation in A_n than that in C_n during DNA replication and more SNPs were transmitted to RNA during transcription, which may contribute to A_n expression dominance. These activities resulted in asymmetrical gene expression in polyploid *B. napus.* The SNPs transmitted from DNA to RNA could be an important complement feature in comparative genomics, and they may play important roles in asymmetrical genome evolution in polyploidy.

Keywords: polyploidization, gene expression dominance, asymmetrical evolution, *Brassica napus*, transmitted variation

INTRODUCTION

Polyploidy is an important part of the history of natural plant species, and has promoted the domestication of cultivated crop species (Chen et al., 2018). Genome sequencing and comparative genomic analysis have identified ~50 polyploidization events across plant phylogenetic trees to date (Vanneste et al., 2014; Van de Peer et al., 2017; Chen et al., 2018). Polyploidization is considered the primary driver of plant species diversification and plays an important role in plant genome evolution. Polyploidy has become a popular research topic in plant science due to the increased number of polyploidization events identified in plants. Subgenome dominance, defined as gene fractionation

bias and expression dominance between homoeologous genes from different subgenomes, is ubiquitous in polyploidy, especially in allopolyploids. The features of transposable elements (TEs) (Freeling et al., 2012; Bottaniet al., 2018), DNA methylation (Hollister and Gaut 2009), gene density (Cheng et al., 2012), alternative splicing forms of genes (Liu et al., 2014), and single nucleotide polymorphism (SNP)-marked phenotypic diversity (Renny-Byfield et al., 2017) are substantially associated with subgenome dominance. Although understanding the mechanism of subgenome dominance has progressed, further investigation is needed to fully understand the subgenome dominance and polyploidy evolution.

Asymmetric evolution has been confirmed in many plant species since it was originally identified in the mesopolyploid *Brassica oleracea* (Liu et al., 2014). Genes in the dominant subgenome (maize 1) of maize accounted for more trait variations than those in maize 2, indicating asymmetric contributions of different subgenomes (Renny-Byfield et al., 2017). A comparison of the distribution of positively selected genes and fiber-related quantitative trait loci revealed that the A_t subgenome was selected for fiber improvement genes, and the D_t subgenome was selected for stress tolerance genes (Zhang et al., 2015b). Asymmetrical changes in wheat gene expression, small RNAs, and chromatin in subgenomes have been identified in resynthesized wheat allotetraploids (Jiao et al., 2018). *B. napus,* a recent allotetraploid species derived from *B. rapa* (AA, 2n = 20) (Wang et al., 2011) and *B. oleracea* (CC, 2n = 18) (Liu et al., 2014), is ideal for studying the effects of SNPs in asymmetric subgenomes. Investigations into asymmetric subgenomes have progressed considerably, and the gaining knowledge has played an important role in guiding genetic improvements in *B. napus*. C_n has large chromosome regions associated with important seed quality traits of rapeseed and less diversity than A_n (Qian et al., 2014; Tang et al., 2019). A_n contains a higher density of SNPs and insertion/deletions (InDels) than C_n of *B. napus*, and is consistent with a faster linkage disequilibrium (LD) decay rate of A_n than that of C_n (Wu et al., 2019). Expression levels of the genes in the A_n were significantly higher than those in the C_n subgenome, besides, A_n possesses a higher level of active epigenetic marks and lower level of repressive epigenetic marks (Zhang et al., 2021). A_n had a more homoeolog expression bias and expression level dominance than those of C_n in resynthesized *B. napus* (Wu et al., 2018). Currently, asymmetric subgenomes and asymmetric evolution have become a hotspot in polyploid research and a focus of plant science research. However, the effects of SNPs transmitted from DNA to RNA during asymmetric evolution have not been reported.

Gene expression is the process through which information from a gene is used to synthesize a functional gene that enables the generation of end products, proteins, or non-coding RNAs, ultimately affecting the phenotype (Baig et al., 2016). Mutations in SNPs are caused by environmental factors and spontaneous mutations in DNA or RNA in nature (Kato et al., 2001). Gene expression levels are affected by many factors, such as SNP density, TE density, and methylation levels (Hollister and Gaut 2009). An enhanced methylation state of TEs might suppress the expression of neighboring genes in autotetraploid rice, suggesting that chromosome doubling induces methylation variation in TEs that affect gene expression (Zhang et al., 2015a). The methylation status of TEs near genes likely plays a role in the biased expression of genes depending on the subgenome location (Hollister and Gaut 2009; Springer et al., 2016). Decreased expression of genes in a submissive subgenome in the mesopolyploid crop species *B. rapa* is associated with higher transposon coverage of DNA upstream of the transcriptional unit and increased smRNA coverage of the same region (Woodhouse et al., 2014). Genes and TEs vary among subgenomes and they are associated with gene expression in *B. rapa* (Murat et al., 2014; Cheng et al., 2016). Regulatory SNPs in genes can modify gene expression levels and affect related phenotypes (Munkhtulga et al., 2010). However, little is known about correlations between SNP variations during transcription and gene expression.

Here, we investigated the asymmetric phenomenon of SNPs transmitted from DNA to RNA in two subgenomes of *B. napus* based on genomic re-sequencing data and RNA-seq data. We assessed the relationships between nucleotide variations in genomes/transcriptomes and corresponding genes expression levels and found that they showed negative correlation. Further investigation showed that transmitted SNPs had positive correlation with gene expression level, which maybe the reason why A_n had a higher SNP density and more abundant gene expression than those of C_n in *B. napus*. Our study provides novel insights with respect to how evolvability of gene expression bias between subgenomes can be boosted by transmitted SNPs in polyploidy.

MATERIALS AND METHODS

Data Collection and SNP Identification

We investigated variations in the genomic and transcriptional profiles of two *B. napus* lines, Zhongyou821 (ZY821) and Zhongshuang11 (ZS11), which had whole-genome re-sequencing and transcriptome data from two replicates of 11 tissues. These raw data were collected from the BIG Data Center under the BioProject accession codes PRJCA000376 and PRJCA001246 (Lu et al., 2019). To identify comprehensive transcription-level SNPs, we merged all RNA-seq datasets of the same sample for variation calling. Quality was controlled by using the Fastp software with default parameters (Chen et al., 2018).

A total of 6.37 and 7.56 Gb of genomic resequencing clean bases of ZY821 and ZS11 were aligned to the reference genome (*B. napus* cv. Darmor-bzh, a French winter variety) (Chalhoub et al., 2014) using BWA-MEM v0.7.15-r1140 with default parameters (Li and Durbin 2009). Mapped reads were sorted using Samtools (Li et al., 2009), and duplicate reads of the bam alignment file were marked using Picard tools (http://broadinstitute.github.io/picard). We then used the HaplotypeCaller module of the Genome Analysis Toolkit (GATK) to identify genetic variations in DNA with default parameters (McKenna et al., 2010). Variation calling and genotyping were performed across the *Darmorbzh* reference genome. We extracted SNPs using the SelectVariants module

and filtered them using the VariantFiltration module of GATK with the following parameters: clusterSize, 3; clusterWindowSize, 10; filterExpression, QUAL <30.0 || MQ < 50.0 || QD < 2.

Transcriptional variations were identified by using 111.19 and 107.85 Gb RNA-seq clean bases of ZY821 and ZS11, and variants were called in RNA-seq data using a modified pipeline (https://github.com/gatk-workflows/gatk3-4-rnaseq-germline-snps-indels). We initially used STAR (Dobin et al., 2013) to construct index files of the reference genome and mapped clean RNA-seq reads to the reference genome with default parameters. Picard tools were then used to add ReadGroup and sort the mapped reads with "AddOrReplaceReadGroups SO = coordinate RGLB = mRNA." Then, duplicate reads were marked using the MarkDuplicates module with the parameter settings with "CREATE_INDEX = true VALIDATION_STRINGENCY = SILENT." We eliminated Ns using the SplitNCigarReads module in GATK to maintain grouping information with the parameters of rf as ReassignOneMappingQuality, RMQF 255, RMQT 60, and -U ALLOW_N_CIGAR_READS. We then identified transcriptional variations using the parameters of "-dontUseSoftClippedBases-stand_call_conf 20" in HaplotypeCaller. Finally, transcription-level SNPs were extracted and filtered by GATK using the same module and parameters as the genomic SNPs. All SNP datasets were annotated using SnpEff v4.1 (Cingolani et al., 2014) with upstream and downstream interval lengths set to 0.

Statistical Analyses of Asymmetrical SNP Transmission During DNA Transcription

To investigate the differences between the two subgenomes of *B. napus*, we compared SNP transmission between ZY821 and ZS11. Two samples or two levels containing a common homozygous SNP were regarded as having been transmitted. We were initially concerned with the transmission of genomic SNPs at the transcriptional level in ZY821 and ZS11. Because ZY821 is one of parents of ZS11, some ZY821 genetic components were transferred to the ZS11 genome during breeding. We further investigated the subgenome differences in SNPs transmitted from the ancestral parent to the offspring. We analyzed two types of transmission from ZY821 to ZS11, they were genomic SNP of ZY821 transferred to ZS11 genomic level and ZY821 genomic SNP transferred to ZS11 genomic level then transferred to ZS11 transcriptional level, respectively. The transmission ratio/rate was defined as the number of transmitted SNPs divided by the total number of genomic SNPs detected in the gene. In order to study the asymmetric distribution of transmitted SNPs in two subgenomes of *B. napus*, we analyzed the transmitted SNPs of all syntenic genes between A_n and C_n subgenomes (Chalhoub et al., 2014). Ratios of transmitted SNPs on A_n and C_n in different samples/levels were compared using two-tailed t-tests. A homozygous variant that arises from DNA to RNA transmission is regarded as a novel SNP variation in DNA to RNA.

Gene Ontology Enrichment Analyses

Pathways and biological functions of genes without SNPs and genes with genomic SNPs that were all transmitted to the transcriptional level were explored via GO enrichment analyses using the PlantRegMap database (http://plantregmap.gao-lab.org/go.php) with the parameter settings of species, *B. napus*; aspects, biological process, molecular function, cellular component; threshold, and $q \leq 0.01$. The GO enrichment results are shown on the ImageGP website (http://www.ehbio.com/ImageGP/).

Gene Expression and Linear Regression Analyses

Filtered high-quality clean RNA-seq data from different tissues and periods of ZY821 and ZS11 were mapped to the "*Darmor-bzh*" genome (Chalhoub et al., 2014) using STAR with default parameters (Dobin et al., 2013). The mapped reads with mapping quality (MQ) ≤ 30 were filtered using Samtools, and bam files were sorted. As TPM (transcripts per kilobase of exon model per million mapped reads) could correct the inconsistences while comparing the RNA-seq abundance among independent samples (Wagner et al., 2012), StringTie (Pertea et al., 2015) was used to count unique and normalized mapped reads as TPM for each gene with the parameter of "-B -e -G".

We used linear regression analyses to investigate correlations between gene expression level and SNP numbers in gene regions, new SNP variations in DNA to RNA, and SNP transmission rates, respectively. Genes with different transmission rates were divided into 10 groups at 0.1 intervals. Genes with different numbers of SNP variations were divided into groups of 10 intervals in ZY821 DNA, ZY821 RNA, and ZS11 RNA, whereas those in ZS11 RNA were arranged into groups with 20 intervals since SNP variations occurred frequently. The linear model is $y_i = ax_i+b$, where y_i is the average TPM of the *i*th group, and x_i is a different group.

RESULTS

Identification and Annotation of SNPs in ZY821 and ZS11

We filtered 6.37 Gb of genome re-sequencing, and 111.19 Gb RNA-seq data were used to identify SNP variation in ZY821. In total, 1,178,526 SNPs were detected in ZY821 genomic DNA, along with 19 chromosomes and 21 scaffolds (**Supplementary Table S1**). We found that 83,179 SNPs in chromosome chrC03 had the most mutations. Among these SNPs, there were 349,509 SNPs distributed in 40,903 genic regions and 846,691 SNPs located in intergenic regions. 162,124 of the genic SNPs were located in exon regions and 139,244 SNPs in intron regions (**Supplementary Table S2**). Among the SNPs in exon regions, 69,526 missense and 91,469 synonymous variations were annotated. We also identified 792,189 SNPs at the transcription level in ZY821, and most variations were detected in chromosome chrA03 (63,925 SNPs). We found 536,701 and 144,456 variations in the exon and intron regions, respectively, and 64.05% (349,110) and 34.63% (185,846) of SNPs

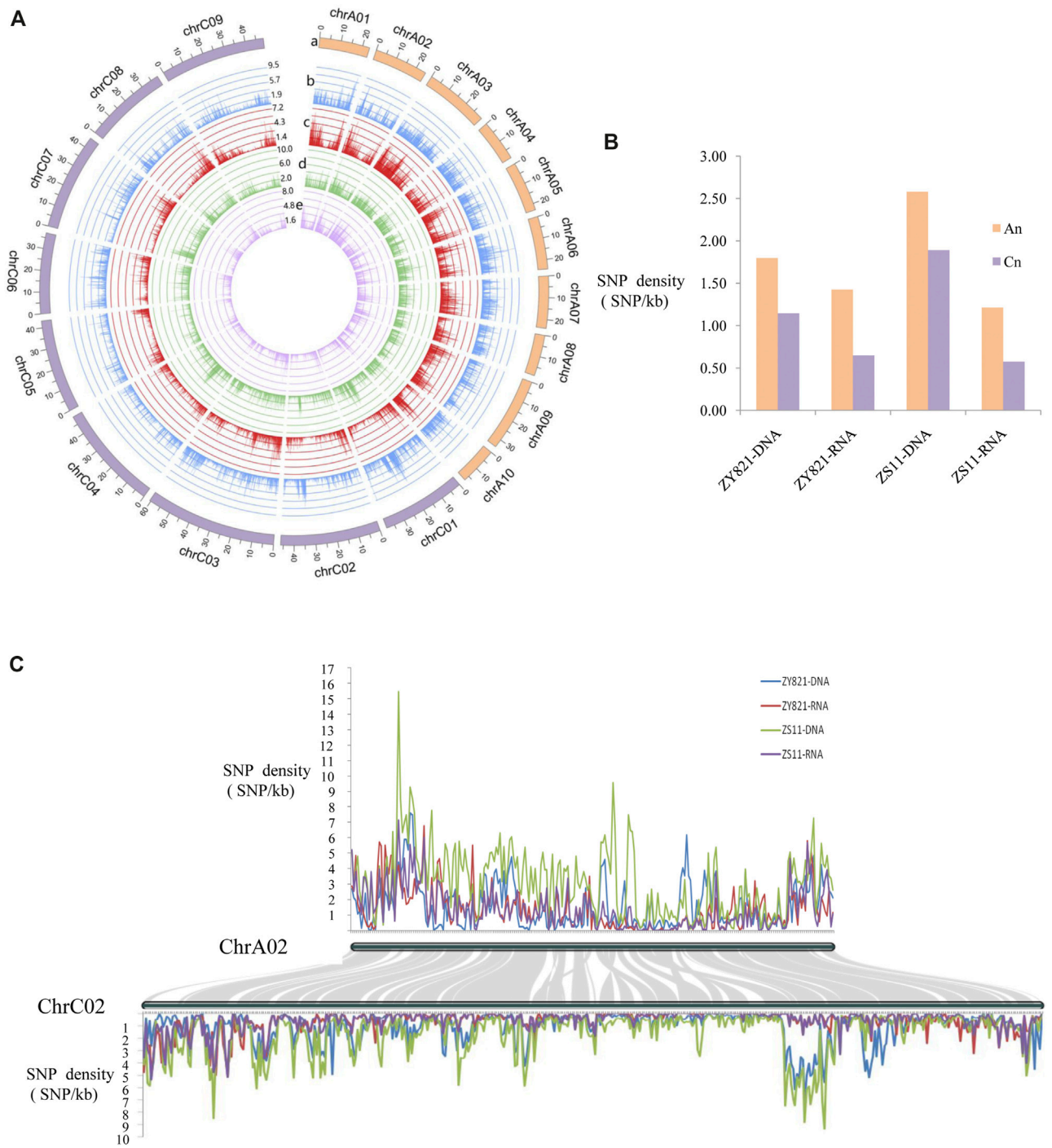


FIGURE 1 | Distribution of SNP density on DNA/RNA levels of two *B. napus* lines ZY821 and ZS11 **(A)** Density of SNPs in 19 chromosomes among four levels. a: Sizes (Mb) of 19 chromosomes in *B. napus*; b–e: line graph of SNP density (y-axis, numbers of SNPs/kb) for chromosomes at ZY821 and ZS11 genomic and transcriptional levels, respectively. **(B)** Distribution of SNP density in two subgenomes. **(C)** Local comparison of SNP density in chromosomes chrA02 and chrC02. Connecting gray lines in background highlight conserved syntenic genes between chrA02 and chrC02.

in exon regions were annotated as synonymous and missense variants, respectively (**Supplementary Table S2**).

After quality control, we identified 1,823,207 and 686,858 genomic and transcriptional SNPs, respectively, in ZS11 using 7.56 Gb genome re-sequencing and 107.85 Gb RNA-seq data. Among the genomic SNPs, 1,352,662 were located in intergenic regions with the most variations, and 228,122 and 199,500 were located in exons and introns, respectively. Overall, 97,806 (42.87%) and 128,632 (56.39%) SNPs in exons were annotated as missense and synonymous variants, respectively. Among the

SNPs at the transcription level, 464,923 and 126,513 variations occurred in the exon and intron regions, respectively. Among the SNPs in exons, 163,198 (35.10%) and 300,178 (64.57%) were missense and synonymous variants, respectively (**Supplementary Table S2**).

We identified 27,024 shared genes without SNPs at the genomic and transcriptional levels in both ZY821 and ZS11, among which 8,756 and 18,268 were located on A_n and C_n, respectively. Results of GO enrichment analysis of the genes without SNPs showed that 54 terms were enriched in cellular components, biological processes, and molecular functions. Almost all terms were basic biological processes for growth, such as an integral component of membrane and membrane part in cellular component, the regulation of nucleic acid-templated transcription and nucleoside bisphosphate metabolic processes in biological processes, enzyme inhibitor activity, and electron carrier activity in molecular function (**Supplementary Table S3**), which indicated that these genes were conserved.

Asymmetric Distribution of SNPs Between A_n- and C_n-Subgenomes

The asymmetric characteristics of the A_n and C_n subgenomes of *B. napus* have been identified (Qian et al., 2014; Lu et al., 2019; Wu et al., 2019). We found that ZY821 and ZS11 had more total SNPs in C_n than in A_n at the genomic level, whereas C_n had fewer SNPs than A_n at the transcriptional level (**Supplementary Table S1**). The total length of A_n was 315.05 Mb in the reference genome, which was less than that in the C_n (526.93 Mb) subgenome; therefore, we assessed SNP density to reevaluate differences between the two subgenomes. **Figure 1A** shows the distribution of SNP densities in the whole genome. The density of ZS11 genomic SNPs was the highest at 3.7/kb in scaffold A10_random and that of ZS11 transcriptional SNPs was the lowest (0.18/kb) in the scaffold Unn_random. These might have been due to the frequency of coding genes being the lowest (1/15.61 kb) in the scaffold Unn_random. The mean SNP density in the genome was 1.39 and 0.93/kb at the ZY821 genomic and transcriptional levels, and 2.14, and 0.81/kb at the ZS11 genomic and transcriptional levels (**Supplementary Table S1**). The genomic and transcriptional SNP density was higher in A_n than in C_n, especially at the transcriptional level, and the SNP density of A_n was almost twice that of C_n (**Figure 1B**). Taken chromosomes chrA02 and chrC02 as an example due to they had good collinearity, the SNP density of chrA02 was generally higher than that of chrC02, although that of chrA02 was occasionally lower than that of chrC02 in some chromosomal regions (**Figure 1C**).

Asymmetric SNP Transmission Between A_n- and C_n-Subgenomes

We initially detected 61,832 genomic SNPs that were transmitted at the transcriptional level in ZY821 and involved 25,583 coding genes. A total of 37,845 and 23,920 SNPs occurred in A_n and C_n, respectively (**Figure 2A**), whereas 14,258 synonymous SNPs in A_n while 6,952 synonymous SNPs in C_n. Among syntenic genes in A_n and C_n, the SNP transmission ratio from DNA to RNA was 31.05% in A_n and 30.23% in C_n (**Figure 2B**). In ZS11, 113,752 SNPs were transmitted from DNA to RNA in A_n, and 63,245 in C_n included 35,774 coding genes (**Figure 2C**). The ratios of transmitted syntenic genes in A_n and C_n were 55.43 vs. 43.21% ($p = 5.45 \times 10^{-89}$; **Figure 2D**). Parental-related transmitted SNPs differed between the two subgenomes. Among 492,916 genomic SNPs from ZY821 transmitted to ZS11, in total 249,185, 241,871, and 1,860 were in A_n, C_n, and the scaffold Unn_random, respectively. These findings indicated that 44.06% of genomic SNP in ZY821was transmitted to the A_n of ZS11, and 40.10% of SNPs in C_n of ZY821 were transmitted to C_n of ZS11 (**Figure 2E**). Moreover, 76,188 SNPs detected at the ZY821 genomic level were transmitted to the ZS11 genome, followed by the ZS11 transcriptional level; of these, 48,367 (63.48%) of 76,188 occurred in A_n and 27,751 (36.42%) in C_n, respectively (**Figure 2F**). These results indicated that SNPs in A_n were more prone to transmission than those in C_n, especially transmission SNPs from the genomic level transfer to the transcriptional level.

The results of GO enrichment analyses revealed the functions of genes with a high ratio of SNP transmission. All (100%) genomic SNPs in 3,967 genes were transmitted to the transcriptional level in ZY821. The number of transmitted genes did not substantially differ between A_n and C_n (1,987 vs. 1,971). Only three GO terms were enriched in the biological process, viz. response to endogenous stimulus, response to hormone, and response to organic substances (**Supplementary Table S4**). Excluding biological processes, no GO terms were enriched in molecular function and cellular components ($q < 0.01$). Among 9,018 completely transmitted genes detected in ZS11, 5,171 and 3,835 were located in A_n and C_n, respectively. Eight GO terms were enriched in biological processes, four in cellular components, and two in molecular functions (**Figure 3**; **Supplementary Table S5**), which are involved in fundamental biological processes and stress responses. All three GO-enriched terms in the biological process of ZY821 were also enriched in ZS11, indicating that these genes play essential roles in the speciation of ZS11.

Impact of Transmitted SNPs on Asymmetrical Gene Expression

We analyzed whole gene expression levels of ZY821 and ZS11 to determine correlations between rates of SNP transmission from DNA to RNA and gene expression. We merged RNA-seq data from different tissues and periods and found that 14,717 and 15,822 genes were not expressed in ZY821 and ZS11, respectively, and 10,931 genes of them were shared, which might be because they were tissue- and time-specific. Gene expression levels decreased as the number of genomic SNPs in genes increased (**Figures 4A,B**), indicating that the expression of a gene with more SNPs would be lower in the *B. napus* genome. The expression of genes containing more RNA variations was lower in ZY821 and ZS11 during transcription (**Figures 4C,D**), which was consistent with the number of variations at the genomic and gene expression levels. However, SNP **Figures 4E,F** density was higher in A_n (**Figure 1B**), but gene expression was higher in C_n (**Figures 4E,F**). Further results revealed that transmitted SNPs had significantly associated with gene expression level. We found that 100% of 3,967 and 9,018 genes with SNPs from the genomic level were transmitted to the transcriptional level in ZY821 and ZS11,

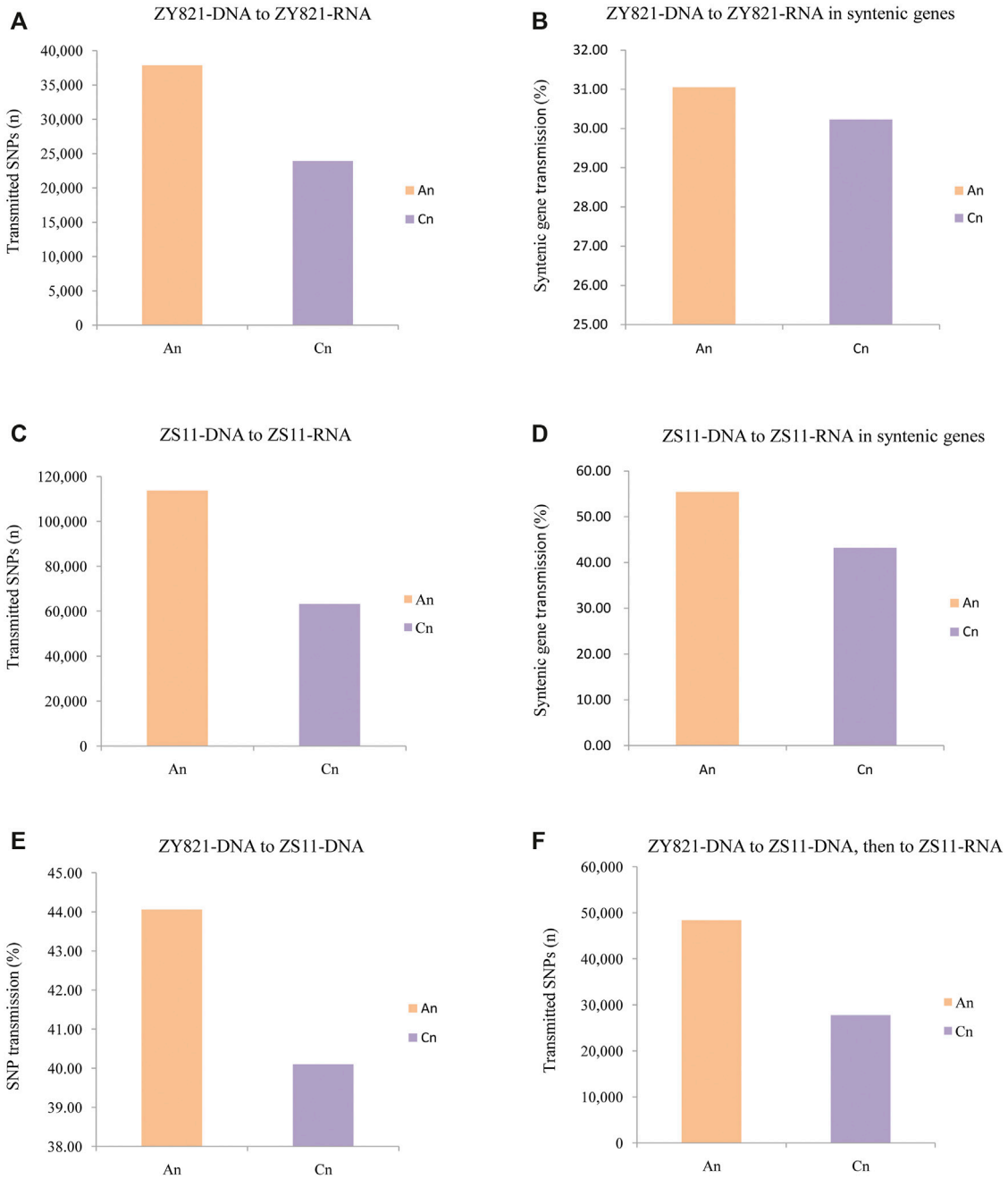


FIGURE 2 | Statistical analyses of transmitted SNPs in two subgenomes at different levels in two *B. napus* lines ZY821 and ZS11. **(A)** Rates of ZY821 genomic SNPs transmitted to the ZS11 genome. **(B)** Numbers ofZY821 genomic SNPs transmitted to theZY821 transcriptional level. **(C)** Rates of SNPsin syntenic ZY821 genes transmitted from genome to theZY821 transcriptional level. **(D)** Numbers of SNPs transmitted from the genomic to the transcriptional level in ZS11. **(E)** Rates of genomic SNPs in syntenic genes of ZS11 transmitted to the transcriptional level. **(F)** Numbers of genomic SNPs from ZY821 transmitted to the ZS11 genome and conserved at the transcriptional level in ZS11.

respectively. We established a linear regression model to fit the details of the SNP transmission ratio from DNA to RNA and gene expression. SNP transmission rates were significantly and positively associated with the gene expression levels. The R^2 values of the fitted line were 0.689 and 0.735 in ZY821 and ZS11, respectively (**Figures 4G,H**), indicating that gene expression increased as the SNP transmission ratio from the DNA to the RNA level increased. These results showed that the ratio of SNPs transmitted from DNA to RNA correlated well with gene expression. The rates and numbers of transmitted SNPs were higher in A_n than in C_n (**Figure 2**), signifying asymmetric gene expression in polyploid *B. napus*.

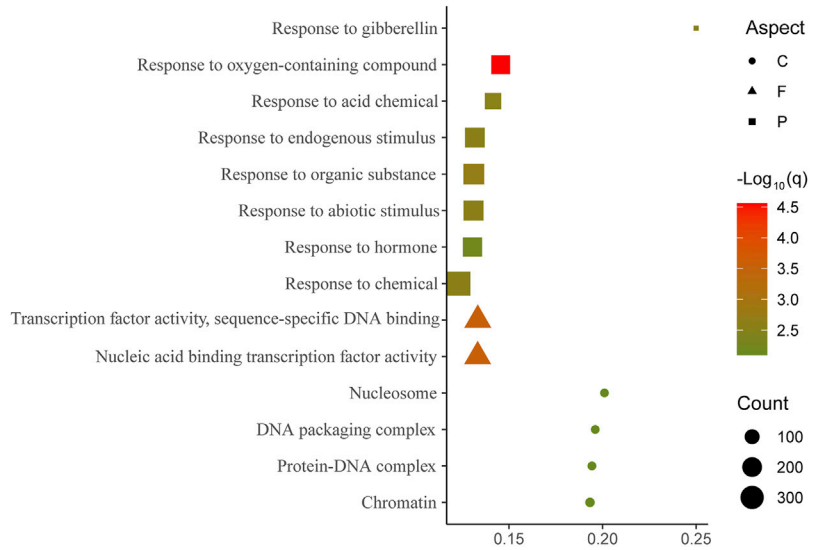


FIGURE 3 | Gene ontology (GO) enrichment findings of genes with 100% of transmitted SNPs from genomic to transcriptional level in ZS11 False discovery rate (q) of all the enriched GO terms was <0.01. ● cellular component (C); ▲ molecular function(F); ■ biological process (P). Larger symbols indicate increased gene expression levels.

DISCUSSION

Polyploidization is considered the primary driver of plant species diversification and plays an important role in plant genome evolution, adaptive selection, gene function innovation, and crop domestication (Soltis et al., 2016). Polyploidy is a state in which the cells of an organism contain two or more sets of ancestral genomes, which may not be accompanied by multiple complete sets. Therefore, redundant genes or components in polyploidy might mutate or recombine into different neomorphic mutants with a reduced risk of extinction. Fractionation, subfunctionalization, and neofunctionalization are significant outcomes for redundant genes. Subgenome dominance develops when a subgenome expresses more genes than another when coexistent within cells (Cheng et al., 2016). The rapid development of genome sequencing and analysis technologies has led to more sequenced genomes and identified polyploidization events that have facilitated comparative genomic studies on polyploid evolution. In polyploidy *B. napus*, A_n contains a higher density of coding genes, transposable elements, and more diversity than C_n (Chalhoub et al., 2014; Wu et al., 2019). We showed a higher SNP density and more transmitted SNPs from DNA to DNA/RNA in A_n, especially from the DNA level to the RNA level, which complements the asymmetric feature in polyploidy. The asymmetric features in the two subgenomes might be due to more frequent outcrossing between *B. napus* and *B. rapa* than between *B. napus* and *B. oleracea* in terms of breeding behaviors (Wu et al., 2019). The A_n subgenome of a modern *B. napus* could retain excellent variations in artificial selection according to the aims of breeders. The GO findings showed that genes with completely transmitted SNPs of ZY821 and ZS11 were commonly enriched in fundamental biological processes and stress responses. This might be an excellent stress SNP micromutation in coding genes that facilitate disease resistance. Therefore, they have been selected by breeders considering adaptability and disease resistance and they were widely planted in China during recent decades.

Gene expression is the most fundamental level at which a genotype produces observable traits. The regulation of gene expression is the basis for cellular differentiation, development, morphogenesis, versatility, and adaptability of all organisms. The amount of gene expression level was affected by SNP density, TE density, and methylation level. Regulatory factors containing a splice acceptor and promoter in intergenic regions are important for regulating gene expression (LeBowitz et al., 1993; Mette et al., 2000; Weber et al., 2007). Different profiles of methylation and altered splicing forms are also associated with gene expression levels (Liu et al., 2014; Mei et al., 2017; Renny-Byfield et al., 2017). Transposons can be repressed, and the expression of genes encoding proteins can be affected by RNA-directed DNA methylation (Rowley et al., 2017). The present findings showed that the numbers and transmission rates of SNPs were associated with gene expression. Changes in SNP bases might result in different combinations of transcription factors, which would lead to different levels of coding gene expression. Transposable elements, splice forms, methylation modifications, and SNP density contribute to gene expression, but systematic understanding remains limited. The integration of genomics, transcriptomics, epigenetics, morphology, and synthetic methods in appropriate species is essential for gaining further insight.

Transcription factors are proteins that control the rates at which genetic information is transcribed from DNA to messenger RNA by binding to a specific DNA sequence, and they are essential for regulating gene expression. Mutations in transcription factors may affect the expression of related genes. A specific binding SNP in the first intron of the *SPP1*

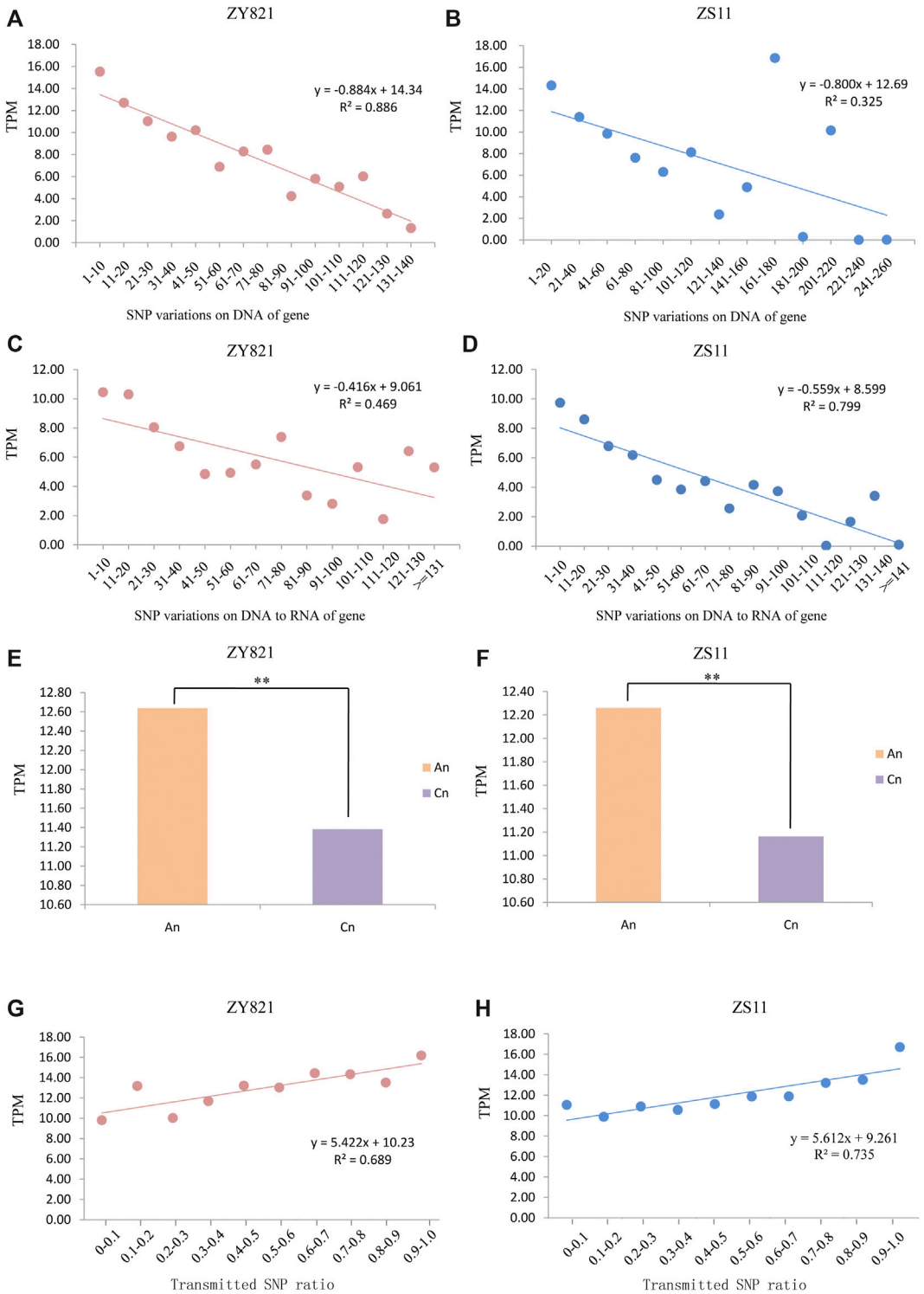


FIGURE 4 | Correlations of the gene expression levels with SNP variations and transmission ratio Correlations and linear regression between gene expression (TPM) and numbers of genomic SNP in all ZY821 **(A)** and ZS11 **(B)** genes. Correlations and linear regression analyzes between TPM and numbers of novel SNP variations on DNA to RNA of gene in ZY821 **(C)** and ZS11 **(D)**. Average expression levels of all syntenic genes in A_n and C_n in ZY821 **(E)** and ZS11 **(F)**, ** represent $p < 0.01$. Correlations and linear regression analyzes between TPM and ratios of SNPs of all genes transmitted from DNA to RNA in ZY821 **(G)** and ZS11 **(H)**.

gene affects the functional characteristics of *SPP1* at the DNA and RNA levels by activating abnormal splicing of the first intron (Muráni et al., 2009). A changed base in rs3122605-G upregulated *IL10* gene expression and increased the risk of systemic lupus erythematosus in European Americans (Sakurai et al., 2013). Our findings showed that gene expression decreased with increasing DNA (or DNA to RNA) variations in coding genes. This indicated that more variations at the level of DNA and during transcription from DNA to RNA resulted in lower gene expression. We excluded high-density variations that occur due to fewer mismatch mapping transcriptional reads, and determined the SNP frequency within 150 bp windows (the read length of RNA-seq in this study) in the entire genome (5,668,614 windows). We found that only 68 and 38 windows in ZY821 and ZS11 contained more than 10 SNPs (default mismatch mapping parameter of STAR), indicating that genes with a higher density of RNA variation did not occur due to a few mismatch reads. This may be because more genomic SNP variations in the gene would increase the risk of a transcription factor not binding to a specific DNA sequence, and nucleotide variations during the transcription process from DNA to RNA might affect RNA alternative splicing and transcriptional inhibition. Alternative splicing of precursor mRNAs from multiexon genes allows an organism to increase its coding potential and regulate gene expression through multiple mechanisms (Reddy et al., 2013). We found 19,756 and 17,509 RNA SNPs in alternative splice regions in ZY821 and ZS11, respectively. These SNPs may affect gene expression via alternative splicing.

Subgenome expression dominance might be obscured by potential homoeologous exchange between different subgenomes, for example, in wheat (Pfeifer et al., 2014; Harper et al., 2016) and *B. napus* (Chalhoub et al., 2014). We found lower expression of genes with more SNP variations, but A_n had not only a higher SNP density, but also a higher level of gene expression. Further analyses revealed that the transmitted SNP ratio was positively correlated with gene expression level. The SNPs transmitted from DNA to RNA might be excellent variations that facilitated the formation of traits that were favored and selected by breeders, and finally retained. Genome dominance is heritable; the expression of a gene in a dominant subgenome tends to be higher than that of its recessive homeolog (Woodhouse et al., 2014). Transmitted SNPs vary in homologous genes in the corresponding subgenome and might be associated with gene dominance. Given our understanding of the subgenome dominance of *B. napus*, the selected SNPs transmitted from DNA to RNA played important roles in gene expression, but the random SNPs involved in down-regulating gene expression might obscure gene expression dominance. In this study, we used the statistical methods to discover a positive correlation between transmitted SNPs and gene expression level, besides this, the effect of transmitted SNPs on gene expression level and subgenome dominance needs to be verified in more polyploidy species and direct molecular experiment evidence. Asymmetric divergence and genome evolution are popular topics in plant science, especially in polyploidy. The innovative knowledge of transmitted SNPs might be an important field in genome evolution of polyploid crops.

CONCLUSION

We identified variations in genomic and transcriptional SNPs in ZY821 and ZS11 and investigated the correlation between SNP numbers/transmission ratios and gene expression. We showed that variations in DNA or the transcription process from DNA to RNA correlated negatively with gene expression levels. The extent of gene expression increased with an increasing ratio of SNP transmission from DNA to RNA, which might explain why A_n had a higher SNP density and more abundant gene expression. SNPs transmitted from DNA to DNA or DNA to RNA might play important roles in asymmetrical evolution and subgenome dominance in polyploidy.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YZ and JC designed the study and revised the manuscript. MT carried out the study and wrote the manuscript. JL revised the manuscript. LS assisted in writing the manuscript. MH improved the language of the manuscript. XH drew the figures. All authors contributed to the manuscript and approved the final manuscript to publish.

FUNDING

This work was supported by Hainan Provincial Natural Science Foundation of China (321RC469), Hainan Province Key Research and Development Program of China (ZDYF2020104), National Natural Science Foundation of China (32060149 and 31801730), and Priming Scientific Research Foundation of Hainan University (KYQD(ZR)-21039).

ACKNOWLEDGMENTS

We sincerely thankful to Dr. Kun Lu (Southwest University) and Dr. Jiana Li (Southwest University) published and uploaded the sequencing data to the public database. We would like to thank Editage (www.editage.cn) for English language editing. We would like to thank the reviewers for their constructive comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.756172/full#supplementary-material

REFERENCES

Baig, M., Keservani, R., and Sharma, A. K. (2016). *RNA Sequencing and Gene Expression Regulation*. Chapter. Boca Raton: Computational Biology and Bioinformatics[M], 71–105.

Bottani, S., Zabet, N. R., Wendel, J. F., and Veitia, R. A. (2018). Gene Expression Dominance in Allopolyploids: Hypotheses and Models. *Trends Plant Sci.* 23, 393–402. doi:10.1016/j.tplants.2018.01.002

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early Allopolyploid Evolution in the post-Neolithic Brassica Napus Oilseed Genome. *Science* 345, 950–953. doi:10.1126/science.1253435

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M. R., Liang, J., et al. (2016). Epigenetic Regulation of Subgenome Dominance Following Whole Genome Triplication in Brassica Rapa. *New Phytol.* 211, 288–299. doi:10.1111/nph.13884

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene Retention, Fractionation and Subgenome Differences in Polyploid Plants. *Nat. Plants* 4, 258–268. doi:10.1038/s41477-018-0136-7

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., et al. (2012). Biased Gene Fractionation and Dominant Gene Expression Among the Subgenomes of Brassica Rapa. *PLoS One* 7, e36442. doi:10.1371/journal.pone.0036442

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2014). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6, 80–92. doi:10.4161/fly.19695

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Freeling, M., Woodhouse, M. R., Subramaniam, S., Turco, G., Lisch, D., and Schnable, J. C. (2012). Fractionation Mutagenesis and Similar Consequences of Mechanisms Removing Dispensable or Less-Expressed DNA in Plants. *Curr. Opin. Plant Biol.* 15, 131–139. doi:10.1016/j.pbi.2012.01.015

Harper, A. L., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S., et al. (2016). Genome Distribution of Differential Homoeologue Contributions to Leaf Gene Expression in Bread Wheat. *Plant Biotechnol. J.* 14, 1207–1214. doi:10.1111/pbi.12486

Hollister, J. D., and Gaut, B. S. (2009). Epigenetic Silencing of Transposable Elements: A Trade-Off between Reduced Transposition and Deleterious Effects on Neighboring Gene Expression. *Genome Res.* 19, 1419–1428. doi:10.1101/gr.091678.109

Jiao, W., Yuan, J., Jiang, S., Liu, Y., Wang, L., Liu, M., et al. (2018). Asymmetrical Changes of Gene Expression, Small RNA S and Chromatin in Two Resynthesized Wheat Allotetraploids. *Plant J.* 93, 828–842. doi:10.1111/tpj.13805

Kato, J., Fujikawa, K., Kanda, M., Fukuda, N., Sasaki, K., Takayama, T., et al. (2001). A Mutation, in the Iron-Responsive Element of H Ferritin mRNA, Causing Autosomal Dominant Iron Overload. *Am. J. Hum. Genet.* 69, 191–197. doi:10.1086/321261

LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M. (1993). Coupling of Poly(A) Site Selection and Trans-splicing in Leishmania. *Genes Dev.* 7, 996–1007. doi:10.1101/gad.7.6.996

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

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

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., et al. (2014). The *Brassica oleracea* Genome Reveals the Asymmetrical Evolution of Polyploid Genomes. *Nat. Commun.* 5, 3930. doi:10.1038/ncomms4930

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome Resequencing Reveals Brassica Napus Origin and Genetic Loci Involved in its Improvement. *Nat. Commun.* 10, 1154. doi:10.1038/s41467-019-09134-9

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Mei, W., Liu, S., Schnable, J. C., Yeh, C.-T., Springer, N. M., Schnable, P. S., et al. (2017). A Comprehensive Analysis of Alternative Splicing in Paleopolyploid Maize. *Front. Plant Sci.* 8, 694. doi:10.3389/fpls.2017.00694

Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A., and Matzke, A. J. M. (2000). Transcriptional Silencing and Promoter Methylation Triggered by Double-Stranded RNA. *Embo J.* 19, 5194–5201. doi:10.1093/emboj/19.19.5194

Munkhtulga, L., Nagashima, S., Nakayama, K., Utsumi, N., Yanagisawa, Y., Gotoh, T., et al. (2010). Regulatory SNP in theRBP4Gene Modified the Expression in Adipocytes and Associated with BMI. *Obesity* 18, 1006–1014. doi:10.1038/oby.2009.358

Muráni, E., Ponsuksili, S., Seyfert, H.-M., Shi, X., and Wimmers, K. (2009). Dual Effect of a Single Nucleotide Polymorphism in the First Intron of the Porcine Secreted Phosphoprotein 1 Gene: Allele-specific Binding of C/EBP Beta and Activation of Aberrant Splicing. *BMC Mol. Biol* 10, 96. doi:10.1186/1471-2199-10-96

Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., et al. (2014). Shared Subgenome Dominance Following Polyploidization Explains Grass Genome Evolutionary Plasticity from a Seven Protochromosome Ancestor with 16K Protogenes. *Genome Biol. Evol.* 6, 12–33. doi:10.1093/gbe/evt200

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., et al. (2014). Genome Interplay in the Grain Transcriptome of Hexaploid Bread Wheat. *Science* 345, 1250091. doi:10.1126/science.1250091

Qian, L., Qian, W., and Snowdon, R. J. (2014). Sub-genomic Selection Patterns as a Signature of Breeding in the Allopolyploid Brassica Napus Genome. *BMC Genomics* 15, 1170. doi:10.1186/1471-2164-15-1170

Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the Alternative Splicing Landscape in Plants. *Plant Cell* 25, 3657–3683. doi:10.1105/tpc.113.117523

Renny-Byfield, S., Rodgers-Melnick, E., and Ross-Ibarra, J. (2017). Gene Fractionation and Function in the Ancient Subgenomes of Maize. *Mol. Biol. Evol.* 34, 1825–1832. doi:10.1093/molbev/msx121

Rowley, M. J., Rothi, M. H., Böhmdorfer, G., Kuciński, J., and Wierzbicki, A. T. (2017). Long-range Control of Gene Expression via RNA-Directed DNA Methylation. *Plos Genet.* 13, e1006749. doi:10.1371/journal.pgen.1006749

Sakurai, D., Zhao, J., Deng, Y., Kelly, J. A., Brown, E. E., Harley, J. B., et al. (2013). Preferential Binding to Elk-1 by SLE-Associated IL10 Risk Allele Upregulates IL10 Expression. *Plos Genet.* 9, e1003870. doi:10.1371/journal.pgen.1003870

Soltis, D. E., Visger, C. J., Marchant, D. B., and Soltis, P. S. (2016). Polyploidy: Pitfalls and Paths to a Paradigm. *Am. J. Bot.* 103, 1146–1166. doi:10.3732/ajb.1500501

Springer, N. M., Lisch, D., and Li, Q. (2016). Creating Order from Chaos: Epigenome Dynamics in Plants with Complex Genomes. *Plant Cell* 28, 314–325. doi:10.1105/tpc.15.00911

Tang, M., Zhang, Y., Liu, Y., Tong, C., Cheng, X., Zhu, W., et al. (2019). Mapping Loci Controlling Fatty Acid Profiles, Oil and Protein Content by Genome-wide Association Study in Brassica Napus. *Crop J.* 7, 217–226. doi:10.1016/j.cj.2018.10.007

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The Evolutionary Significance of Polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi:10.1038/nrg.2017.26

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 Plant Genomes Supports a Wave of Successful Genome Duplications in Association with the Cretaceous-Paleogene Boundary. *Genome Res.* 24, 1334–1347. doi:10.1101/gr.168997.113

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent Among Samples. *Theor. Biosci.* 131, 281–285. doi:10.1007/s12064-012-0162-3

Wang, X., Wang, H., Wang, H., Wang, J., Sun, R., Wu, J., et al. (2011). The Genome of the Mesopolyploid Crop Species Brassica Rapa. *Nat. Genet.* 43, 1035–1039. doi:10.1038/ng.919

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., et al. (2007). Distribution, Silencing Potential and Evolutionary Impact of Promoter DNA Methylation in the Human Genome. *Nat. Genet.* 39, 457–466. doi:10.1038/ng1990

Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., and Wang, X. (2014). Origin, Inheritance, and Gene Regulatory Consequences of Genome Dominance in Polyploids. *Proc. Natl. Acad. Sci.* 111, 5283–5288. doi:10.1073/pnas.1402475111

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., et al. (2019). Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Mol. Plant* 12, 30–43. doi:10.1016/j.molp.2018.11.007

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., et al. (2018). Homoeolog Expression Bias and Expression Level Dominance in Resynthesized Allopolyploid Brassica Napus. *BMC Genomics* 19, 586. doi:10.1186/s12864-018-4966-5

Zhang, J., Liu, Y., Xia, E. H., Yao, Q. Y., Liu, X. D., and Gao, L. Z. (2015a). Autotetraploid rice Methylome Analysis Reveals Methylation Variation of Transposable Elements and Their Effects on Gene Expression. *Proc. Natl. Acad. Sci. U S A.* 112, E7022–E7029. doi:10.1073/pnas.1515170112

Zhang, Q., Guan, P., Zhao, L., Ma, M., Xie, L., Li, Y., et al. (2021). Asymmetric Epigenome Maps of Subgenomes Reveal Imbalanced Transcription and Distinct Evolutionary Trends in Brassica Napus. *Mol. Plant* 14, 604–619. doi:10.1016/j.molp.2020.12.020

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015b). Sequencing of Allotetraploid Cotton (Gossypium Hirsutum L. Acc. TM-1) Provides a Resource for Fiber Improvement. *Nat. Biotechnol.* 33, 531–537. doi:10.1038/nbt.3207

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.

Copyright © 2021 Tang, Li, Hu, Sun, Helal, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fgene.2021.769688

A Multilayer Interactome Network Constructed in a Forest Poplar Population Mediates the Pleiotropic Control of Complex Traits

Huiying Gong[1], Sheng Zhu[2], Xuli Zhu[3], Qing Fang[4], Xiao-Yu Zhang[1] and Rongling Wu[3,5]**

[1]College of Science, Beijing Forestry University, Beijing, China, [2]College of Biology and the Environment, Nanjing Forestry University, Nanjing, China, [3]Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [4]Faculty of Science, Yamagata University, Yamagata, Japan, [5]Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA, United States

The effects of genes on physiological and biochemical processes are interrelated and interdependent; it is common for genes to express pleiotropic control of complex traits. However, the study of gene expression and participating pathways *in vivo* at the whole-genome level is challenging. Here, we develop a coupled regulatory interaction differential equation to assess overall and independent genetic effects on trait growth. Based on evolutionary game theory and developmental modularity theory, we constructed multilayer, omnigenic networks of bidirectional, weighted, and positive or negative epistatic interactions using a forest poplar tree mapping population, which were organized into metagalactic, intergalactic, and local interstellar networks that describe layers of structure between modules, submodules, and individual single nucleotide polymorphisms, respectively. These multilayer interactomes enable the exploration of complex interactions between genes, and the analysis of not only differential expression of quantitative trait loci but also previously uncharacterized determinant SNPs, which are negatively regulated by other SNPs, based on the deconstruction of genetic effects to their component parts. Our research framework provides a tool to comprehend the pleiotropic control of complex traits and explores the inherent directional connections between genes in the structure of omnigenic networks.

Keywords: epistasis, multilayer network, omnigenic model, pleiotropic control, system mapping, variable selection

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Jiguo Cao,
Simon Fraser University, Canada
Xiangdong Ding,
China Agricultural University, China

****Correspondence:***
Xiao-Yu Zhang
xyzhang@bjfu.edu.cn
Rongling Wu
rwu@bjfu.edu.cn

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

Received: *02 September 2021*
Accepted: *19 October 2021*
Published: *12 November 2021*

Citation:
Gong H, Zhu S, Zhu X, Fang Q, Zhang X-Y and Wu R (2021) A Multilayer Interactome Network Constructed in a Forest Poplar Population Mediates the Pleiotropic Control of Complex Traits. Front. Genet. 12:769688. doi: 10.3389/fgene.2021.769688

INTRODUCTION

The study of gene pleiotropy has become a focus of genetic research in recent years. Pleiotropy describes the phenomenon that single genes can have multiple biological effects, so that an individual exhibits multiple traits (Solovieff et al., 2013). Pleiotropy is an important factor in genotype–phenotype transmission (Dudley et al., 2005; Gregory et al., 2012; Geiler-Samerotte et al., 2020), which can help us to understand how the underlying biochemical pathways determine the behavior of the cells in which they are present (Roth et al., 2021). With the development of genome-wide association statistical models, the regulatory roles of genes and their interactive effects have received sustained attention in research on pleiotropy (Sivakumaran et al., 2011; Visscher and Yang, 2016; Watanabe et al., 2019); these existing genetic studies mainly focus on the action of identified key genes, which account for only a small amount of phenotypic

variation. The current understanding of the networks of genes that actually drive the development of complex traits, and how genes throughout the genome interact remains inadequate.

In the early 20th century, reductionism had a positive impact on the development of biological understanding (Osler, 1969). Conventional reductionism is the theory that complex systems and phenomena can be understood and described by breaking them down to their fundamental parts. According to the complex network mathematical model, the salient information can be extracted from a network constructed using reductionist principles. However, unlike physico-chemical networks, in biological networks, organisms have an organic character that emerges not through the sum of all components, but the interconnection between them (Regenmortel, 2004; Roukos 2011; Mazzocchi 2012). An "omnigenic" model was therefore proposed to take into account the activity of genes in cells, which form a broad network in which each gene exerts an influence on the occurrence of disease or development of traits, including those without any obvious connection to traits or diseases in interconnected gene networks (Boyle et al., 2017; Wray et al., 2018; Liu et al., 2019).

Omnigenic network modeling has become a powerful and fundamental tool for analyzing interactomes and quantifying relationships among genes; many statistical networking methods have been established (Carter et al., 2004). However, most related gene regulatory networks have their own underlying mathematical rationale and assumptions, thus results lack robustness (Marbach et al., 2012). In addition, an omnigenic network involving large amounts of genomic data is high-dimensional, which brings inevitable challenges in computing. Clustering techniques are required to sort complicated, high-dimensional genes into communities or modules through modularity theory (Wang and Huang, 2014; Huynh-Thu and Sanguinetti, 2019). Based on the dynamic nature of gene behaviors, functional clustering has made it possible to identify the similarity of temporal genetic effects from large numbers of loci, thus resolving biological and computational complexity (Kim et al., 2008; Li et al., 2010; Wang et al., 2012).

In this article, we propose a new model to explore the multilayer interactome network mediating the pleiotropic control of complex traits (in this case tree height and diameter) by integrating system mapping (Wu et al., 2011; Bo et al., 2014; Sun and Wu, 2015), functional clustering, and differential genetic regulatory systems (Jong, 2002; Jong et al., 2003). Given that the metabolism of an organism is a network of interacting processes (Michael, 2019), we established a coupled regulatory interaction (CRI) differential equation model to describe interactions between complex traits growth (Wu et al., 2011; Jérôme et al., 2013). This differential equation can be embedded into the system mapping model to discern specific quantitative trait loci (QTLs) that control the traits, according to differences between genotypes in the equation parameters. Further, we reveal gene interactions, and describe how genes are implicated in the control of intracellular and intercellular processes through multilayer gene network modeling (Someren et al., 2002; Margolin and Califano, 2010; Yukilevich et al., 2010; Costanzo et al., 2019; Wu and Jiang, 2021); this incorporates modularity theory to resolve ultrahigh-dimensional computational complexity. From this, networks of separate modules can be constructed, and modules can be divided into submodules and sub-submodules; genome-wide epistasis can thus be interpreted from an evolutionary game theory perspective (Smith and Price, 1973) postulating that interactions between genes can lead to genetic effect payoff. Our multilayer interactome network provides a powerful computational tool in the mechanistic analysis of large, genome-wide expression datasets and revolutionizes our understanding of the pleiotropic control of complex traits.

MATERIALS AND METHODS

Plant Materials

We used published data from trees as mapping population for our study (Xu et al., 2016). It comprises a full-sib family derived from hybridization between the female clone I-69 of *Populus deltoides* and the male clone I-45 of *Populus* × *euramericana*, which were introduced from the United States in the 1970s (Wu et al., 1992). This hybridization generated 450 hybrid trees, planted with ramets in a uniform land at Zhangji Forest Farm, Xuzhou, Jiangsu, China. The two parents, I-69 and I-45, and 64 randomly-selected hybrids were used for stem growth analysis, in which annual data comprising stem height and stem diameter during the first 24 years of growth from 1987 to 2010 were measured. The trees were genotyped at single nucleotide polymorphism (SNP) sites using the Applied Biosystems QuantStudio 12K Flex Real-Time Polymerase Chain Reaction (PCR) System for genome-wide mapping. 156 362 SNPs were characterized through stringent quality-control filters segregating with different patterns, of which 94 591 SNPs belong to testcross markers and 61 771 SNPs belong to intercross markers, respectively. The testcross markers are those at which one parent is heterozygous whereas anotherr is homozygous. The intercross markers are derived from two heterozygous parents.

CRI Differential Equations of Complex Traits

The pattern of interactions between tree height and diameter is fundamental for the development and application of many growth and yield models. It is the focus of theoretical and empirical analyses indicating pleiotropic control (Dharmawardhana et al., 2010; Jiang et al., 2016). The growth relationship between diameter and height can be described by many traditional models, such as nonlinear functions and generalized height-diameter functions (Temesgen and Gadow, 2004; Ahmadi and Alavi, 2016; Roya and Tooba, 2020), and mixed-effect models (Sharma and Parton, 2007; Crecente-Campo et al., 2010; Bronisz and Mehttalo, 2020; Liao et al., 2020), which evaluate the overall change and trends among traits by establishing the relation of function. However, the internal coordination of tree height and diameter can be understood in more depth by investigating the underlying biological mechanisms of their control. From the perspective of game theory, we introduced the Lotka–Volterra differential equation

(May, 1975) to represent the specific forms of the interaction between the complex traits of growth in stem diameter and height. We present a coupled regulatory interaction (CRI) differential equation to describe the growth relationship between traits, which separates the growth of these traits into independent and dependent parts:

$$\begin{cases} \dfrac{dH}{dt} = \alpha_H\left(1 - \dfrac{H}{K_H}\right)H + \alpha_H\beta_{H\leftarrow D}HD = \hat{H} + I_{H\leftarrow D} \\ \dfrac{dD}{dt} = \alpha_D\left(1 - \dfrac{D}{K_D}\right)D + \alpha_D\beta_{D\leftarrow H}DH = \hat{D} + I_{D\leftarrow H} \end{cases} \tag{1}$$

where $\hat{H}$ or $\hat{D}$ is the independent growth of each trait, determined by its intrinsic properties, and $I_{H\leftarrow D}$ or $I_{D\leftarrow H}$ describes the interactive growth of each trait, depending on how it interacts with the other, coexisting trait. H and D in this study represent the growth of stem height and diameter, α_H and α_D represent the growth rate, and K_H and K_D represent the asymptotic values. $\beta_{H\leftarrow D}$ and $\beta_{D\leftarrow H}$ are dependent parameters; the size of the positive or negative dependent parameters $\beta_{H\leftarrow D}$ and $\beta_{D\leftarrow H}$ indicate the type of interactive relationship. Specifically, when the dependent parameter is positive, the growth is promoted by the coexistence characteristics. Conversely, when the dependent parameter is negative, the growth is hindered. If the dependent parameter is zero, the overall growth depends solely on independent growth, which means that there is no interaction between the two traits. The interaction of the two traits can be described by a strategy set, shown in **Table 1** and summarized as follows:

- neutral interaction strategy: no interaction between the two characteristics;
- cooperative interaction strategy: the growth of one characteristic is promoted by the other, without hindering the growth of the latter;
- antagonistic interaction strategy: the overall growth of at least one characteristic is inhibited by the other.

The parameters in the CRI differential equation describe the developmental mechanisms behind the formation and expression of the two traits and their interaction. Using this CRI differential equation, we can explore the dynamic changes of the growth of each trait, and quantitatively analyze the nature of the interaction between traits.

Identification of Interacting QTLs and Trait Regulation

Systems mapping is a classical approach for mapping complex traits by comparing the genotypic differences in growth equation parameters throughout the genome (Wu et al., 2011; Bo et al., 2014; Sun and Wu, 2015). We designed a mapping population of n trees. Genome-wide SNPs were genotyped in all trees using high-throughput methods, and trees were phenotyped for height and diameter at a series of time points $(1, \cdots, T)$ during development. The phenotypic values of tree i $(i = 1, \cdots, n)$ for height and diameter are expressed as:

$y_{1i} = (y_{1i}(1), \cdots, y_{1i}(T))$ and $y_{2i} = (y_{2i}(1), \cdots, y_{2i}(T))$.

The joint likelihood for n observations can be expressed as:

$$L(y_1, y_2) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}; \Theta, \psi) \tag{2}$$

where $f(y_{1i}, y_{2i}; \Theta, \psi)$ is a probability density function of bivariate normal distribution with mean vector described as:

$$\mu = (\mu_1(1), \cdots, \mu_1(T); \mu_2(1), \cdots, \mu_2(T))$$

represented by the parameters:

$$\Theta = (\alpha_H, K_H, \beta_{H\leftarrow D}, \alpha_D, K_D, \beta_{D\leftarrow H})$$

and the covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}$$

where the diagonal elements are the variance matrices of each trait, and the off-diagonal elements are the covariance matrices between a pair of traits. We used a first-order structured antedependent (SAD (1)) statistical model controlled by a set of specific parameters ψ to express the longitudinal covariance matrix (Zhao et al., 2005a; Zhao et al., 2005b).

Considering the difference in genotype on the growth of trees, we constructed the likelihood function:

$$L_1(y_1, y_2) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} f_j(y_{1i}, y_{2i}; \Theta_j, \psi) \tag{3}$$

where J is the number of QTL genotypes and n_j is the number of those trees carrying genotype j, satisfying:

$$\sum_{j=1}^{J} n_j = n.$$

TABLE 1 | A strategy set of stem growth on traits interaction. The regulation strategy table formed by the different positive and negative combinations of two interactive regulation parameters.

$\beta_{D\leftarrow H}$ \ $\beta_{H\leftarrow D}$	+	0	-
+	cooperative		
0		neutral	
-	antagonistic		

$f_j(y_{1i}, y_{2i}; \Theta_j, \psi)$ is a probability density function of bivariate normal distribution with mean vector defined as:

$$\mu_j = \left(\mu_{j1}(1), \cdots, \mu_{j1}(T); \mu_{j2}(1), \cdots, \mu_{j2}(T)\right)$$

represented by parameters:

$$\Theta_j = \left(\alpha_{Hj}, K_{Hj}, \beta_{H\leftarrow Dj}, \alpha_{Dj}, K_{Dj}, \beta_{D\leftarrow Hj}\right)$$

and covariance matrix Σ.

We incorporated the simplex (Zhao et al., 2004), expectation maximization (EM) (Dempster, 1977), and fourth-order Runge–Kutta algorithms to obtain maximum likelihood estimates (MLEs) of the parameters in the mean vector and covariance matrix, respectively. Based on the likelihood of **Eqs 2**, **3**, we can test whether a given SNP is significantly associated with trait allometry, using the following formula:

H_0: $\Theta \equiv \Theta_j$ versus H_1: $\Theta_j \neq \Theta$, for $j = 1, \cdots, J$

In which the log likelihood ratio is calculated and compared with a genome-wide critical threshold. When the null hypothesis above is rejected, this means that significantly associated QTLs have been detected. These QTLs can be further tested to determine whether they affect the independence and interdependence of traits:

$$H_0: \left(\alpha_{Hj}, K_{Hj}, \alpha_{Dj}, K_{Dj}\right) = (\alpha_H, K_H, \alpha_D, K_D), for\ j = 1, \cdots, J$$
$$H_0: \left(\beta_{H\leftarrow Dj}, \beta_{D\leftarrow Hj}\right) = (\beta_{H\leftarrow D}, \beta_{D\leftarrow H}), for\ j = 1, \cdots, J$$

Modules Detection From Bivariate Functional Clustering

For all p SNPs throughout the genome, we calculated the genetic standard deviation based on the parameters from maximum likelihood estimation for both traits, to describe the genetic effect of SNPs on trait development. We utilized functional clustering to identify distinct patterns of gene expression dynamics by dividing p SNPs into L tight-knit modules (Kim et al., 2008; Li et al., 2010; Wang et al., 2012). In our research, functional clustering is extended into bivariate functional clustering including height and diameter. The following equations denote the vectors of genetic effects of SNP k ($k = 1, \cdots, p$) on height and diameter, respectively:

$$g_{1k} = \left(g_{1k}(1), \cdots, g_{1k}(T)\right)$$
$$g_{2k} = \left(g_{2k}(1), \cdots, g_{2k}(T)\right)$$

The likelihood based on a mixture model is formulated as:

$$L_2(g_1, g_2) = \prod_{k=1}^{p} \sum_{k=1}^{L} \left[\omega_l f_l(g_{1k}, g_{2k}; \Phi_l)\right] \quad (4)$$

where ω_l is a prior probability representing the proportion of module l, and satisfying $\sum_{l=1}^{L} \omega_l = 1$, $f_l(g_{1k}, g_{2k}; \Phi_l)$ is a probability density function of bivariate normal distribution with Φ_l as a vector of unknown parameters structuring the cluster-specific mean vector:

$$u_l = (u_{l1}(1), \cdots, u_{l1}(T); u_{l2}(1), \cdots, u_{l2}(T))$$

and covariance matrix Σ_g.

We incorporated nonparametric Legendre orthogonal polynomials (LOP) mathematical equation and the SAD (1) statistical model to fit the mean-covariance structures. A hybrid EM-simplex algorithm was implemented to estimate the parameters Φ_l in likelihood of **Eq. 4**; the posterior probability that SNP k belonging to a particular module l in each iteration can be determined as:

$$\Omega_{l|k} = \frac{\omega_l f_l(g_{1k}, g_{2k}; \Phi_l)}{\sum_{l'}^{L} \omega_{l'} f_{l'}(g_{1k}, g_{2k}; \Phi_{l'})}$$

and the proportion of module l is calculated by:

$$\omega_l = \frac{\sum_{k=1}^{p} \Omega_{l|k}}{p}$$

An optimal number of SNP clusters in terms of their different genetic effects can be determined using penalized likelihood criteria, such as AIC and BIC.

Network Construction

Molecular-level genetic regulatory systems can help us to understand how genes are implicated in trait growth processes through networks. A universal property of complex networks is that the change of one component (usually expressed in the form of a rate equation) in the system is a function of other components:

$$\frac{dx_l}{dt} = G_l(\boldsymbol{x}),\ 1 \le l \le L$$

where $\boldsymbol{x} = [x_1, \cdots, x_L]'$ is the vector of system components, and G_l: $\mathbb{R}^n \to \mathbb{R}$ is the function (generally non-linear) that determines the dynamics model of the entire system (Jong, 2002).

Network theory states that the observed value of a variable is the sum of the components of its own strategy and those derived from the strategies of its interactive counterparts (Wu and Jiang, 2021). The relational structure of each component can be divided into independence and dependence, to explore how these SNPs interconnect and interdepend. Where $g_{1l} = (g_{1l}(1), \cdots, g_{1l}(T))$ and $g_{2l} = (g_{2l}(1), \cdots, g_{2l}(T))$ denote the vectors of overall genetic effects of module l on height and diameter, respectively, we can derive an ordinary differential equation (ODE)-based equation system, as follows:

$$\frac{dg_{\cdot l}}{dt} = G_l^0(g_{\cdot l}; \Theta_l) + \sum_{l'=1, l' \neq l}^{L} G_{l\leftarrow l'}(g_{\Delta l'}; \Theta_{l\leftarrow l'}), l = 1, \cdots, L \quad (5)$$

where $g_{\cdot l}$ is the net genetic effect of module l on height or diameter. This can be deconstructed into two components: $G_l^0(g_{\cdot l}; \Theta_l)$ is a time-varying function that characterizes the independent genetic effect of module l with the assumption that it occurs in isolation; and $\sum_{l'=1, l' \neq l}^{L} G_{l\leftarrow l'}(g_{\cdot l'}; \Theta_{l\leftarrow l'})$ is a time-varying function that characterizes the dependent genetic effect of module l that arises from the influence of the other module l'. Θ_l and $\Theta_{l\leftarrow l'}$ are the sets of parameters that fit the independent and dependent functions, respectively.

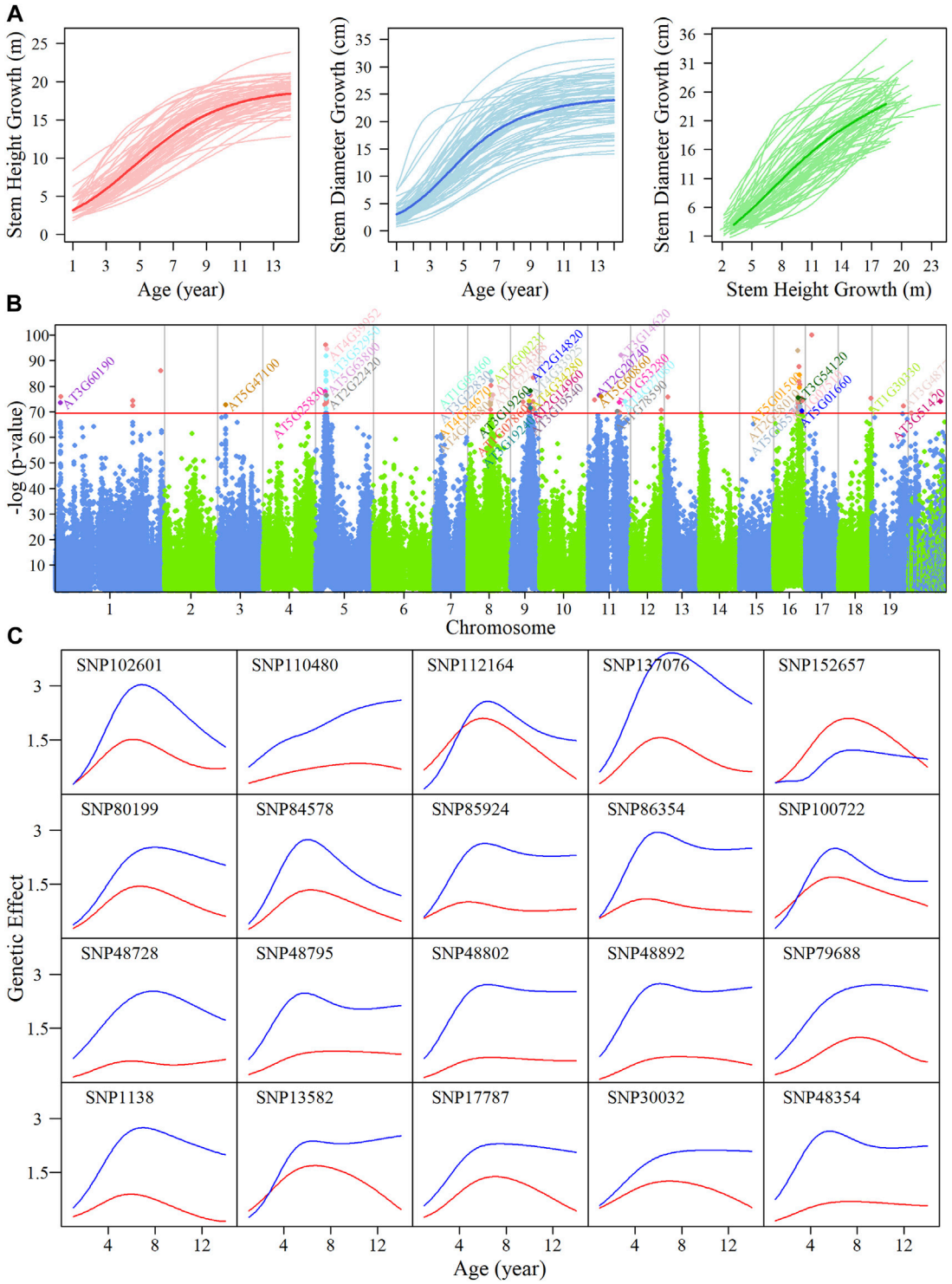


FIGURE 1 | System mapping for the identification of significant single nucleotide polymorphisms (SNPs) goverining growth in stem height and diameter in an interspecific, full-sib family of *Populus* during the first 14 years. **(A)** Overall growth trajectories for stem height (red lines) and diameter (blue lines) fitted using a CRI equation, and relationship curves between traits (green lines) in hybrid poplars. The mean curves for the two growth traits are indicated by darker colored lines. **(B)** Manhattan plot of *p* values after FDR correction over 19 chromosomes of the *Populus* genome by system mapping. Horizontal lines represent the critical threshold at the 10^{-25} significance level obtained after Bonferroni correction. The annotations in the significant region are genes with known biological function. **(C)** Genetic effect curves of 20 distinct significant SNPs identified from the Manhattan plot of stem height (red) and diameter (blue).

The established procedure for constructing genetic network consists of the following three steps: data smoothing, variable selection, and ODE solving. We used LOP to smooth the independent and dependent functions; by interpolating additional values on the curve of best fit of genetic effects over time, the case that the number of modules may be larger than the number of time points can be solved. The network sparsity theory states that there is a limit to the number of links to maintain the stability of the network structure (Liu et al., 2011; Allesina and Tang, 2012; Michailidis and Alché-Buc, 2013; Busiello et al., 2017). It is necessary to filter modules through variable selection from the regression model:

$$g_{\cdot l}(t) = a_l + \sum_{l'=1, l' \neq 1}^{L} b_{l'} g_{\cdot l'}(t) + e_l(t)$$

where a_l is the constant, $b_{l'}$ is the regression coefficient of variable l', and $e_l(t)$ is the residual error. Here, we incorporated LASSO-based variable selection (Tibshirani, 1996) to choose a set of the most significant dependent modules for a focal module. Lastly, we used the fourth-order Runge–Kutta algorithm to solve the simplified ODEs (5), and we calculated the directional, weighted interactions among modules.

RESULTS

Growth Trajectories and Temporal Patterns of Genetic Effects Identified Through System Mapping

In this paper, we used forest poplar annual growth data for stem height and diameter from 1 to 14 years. The growth curves of traits over time follow a sigmoid curve. Many classical growth equations provide quantitative assessment to capture biological rule, such as Gompertz (Gompertz, 1815), Korf (Lundqvist, 1957) and Richards (Richards, 1959). However, these classical models can only evaluate the overall change of one trait as a function of the other, but does not provide insight into the internal mechanisms of how height growth affects diameter growth or vice versa. Given the possible interactions between diameter and height, we used the CRI differential equation to fit the trait growth ($R^2 > 0.99$) (**Figure 1A**). Estimated parameters and statistical evaluation values for the Gompertz, Korf, and Richards classical growth equations and the CRI differential equation to fit the average growth curve are shown in **Supplementary Table S1**. This permitted a comparison of the accuracy and complexity of each equation, which showed that the CRI differential equation had the best fit. The residuals of the growth data were randomly distributed on the predicted value (**Supplementary Figure S1**), indicating that the CRI differential equation was robust. The growth pattern of stem height and diameter follows an antagonistic interaction strategy (**Supplementary Figure S2**). Growth in stem height is inhibited by growth in stem diameter, dramatically reducing the overall growth in stem height; conversely, growth in stem height was found to promote stem diameter growth, indicating an interactive effect between the two traits.

We implemented system mapping to quantitatively analyze the interaction between diameter and height (**Figure 1B**) based on the bivariate normal distribution model; the identified significant quantitative trait loci (QTLs) reveal the physiological mechanism of competitive or cooperative strategies. A total of 88 intercross SNPs and 17 testcross SNPs were found to significantly regulate the interaction in growth of two traits. Over half of these SNPs were within, or adjacent to, candidate genes involving the functions of plant growth-related pathways. For example, SNP 30032 on chromosome 3 was found to be in a region of the calcineurin B-like protein, CBL9 (Pandey et al., 2004), which regulates phytohormone abscisic acid (ABA) responses.

Detailed information on these SNPs that were significantly associated with our traits of interest, including segregation types and physical positions, are given in **Supplementary Table S2**. A majority of the SNPs were found to be located on chromosomes 5, 8, 9, 11, and 16. SNPs that are highly linked on the same chromosome are likely to represent the same QTL, collectively. We then explored how these QTLs percolate through the entire regulatory network structure. The temporal pattern of genetic effects exerted by the QTLs we identified was calculated, as shown in **Supplementary Figure S3**; almost all of these SNPs had a stronger effect on diameter compared with height, except for SNP 152657. The temporal pattern of QTLs effects on trait growth varied, but most QTLs had similar genetic effect patterns, in which effects increased initially and then decreased. We analyzed the dynamic genetic correlations on stem height and diameter indicating the QTLs with pleiotropy effects (**Supplementary Figure S4**). We chose 20 distinct QTLs randomly, as shown in **Figure 1C**; the effect pattern of SNP 110480 was to keep enhancing with growing time, SNP 112164 and SNP 100722 were found to be responsible for both height and diameter growth with similar intensity, and for SNP 137076, there was a marked difference in effect values between the two traits.

Network Modules of Genetic Effect Dynamics Based on Omnigenic Theory

The detection of individual QTLs by system mapping has provided the first detailed understanding of the genetic basis of complex traits, but it may provide limited insight into how common SNPs across the genome, which are below the threshold for statistical significance, act and interact to regulate growth traits (Yang et al., 2010). Some common SNPs are not necessarily significantly associated with traits by themselves, but play an important role in regulating other loci and are therefore indirectly involved in trait control. To explore the genetic contribution from these common SNPs that may have been missed, we carried out a quantitative analysis of epistatic effects among genome-wide SNPs in omnigenic networks. Based on our CRI differential model, we estimated the effect on stem height and diameter of each SNP through system mapping.

The use of high-dimensional genome-wide SNP datasets are essential to revealing how all interconnected genes generate or regulate the expression of complex traits and pleiotropic control.

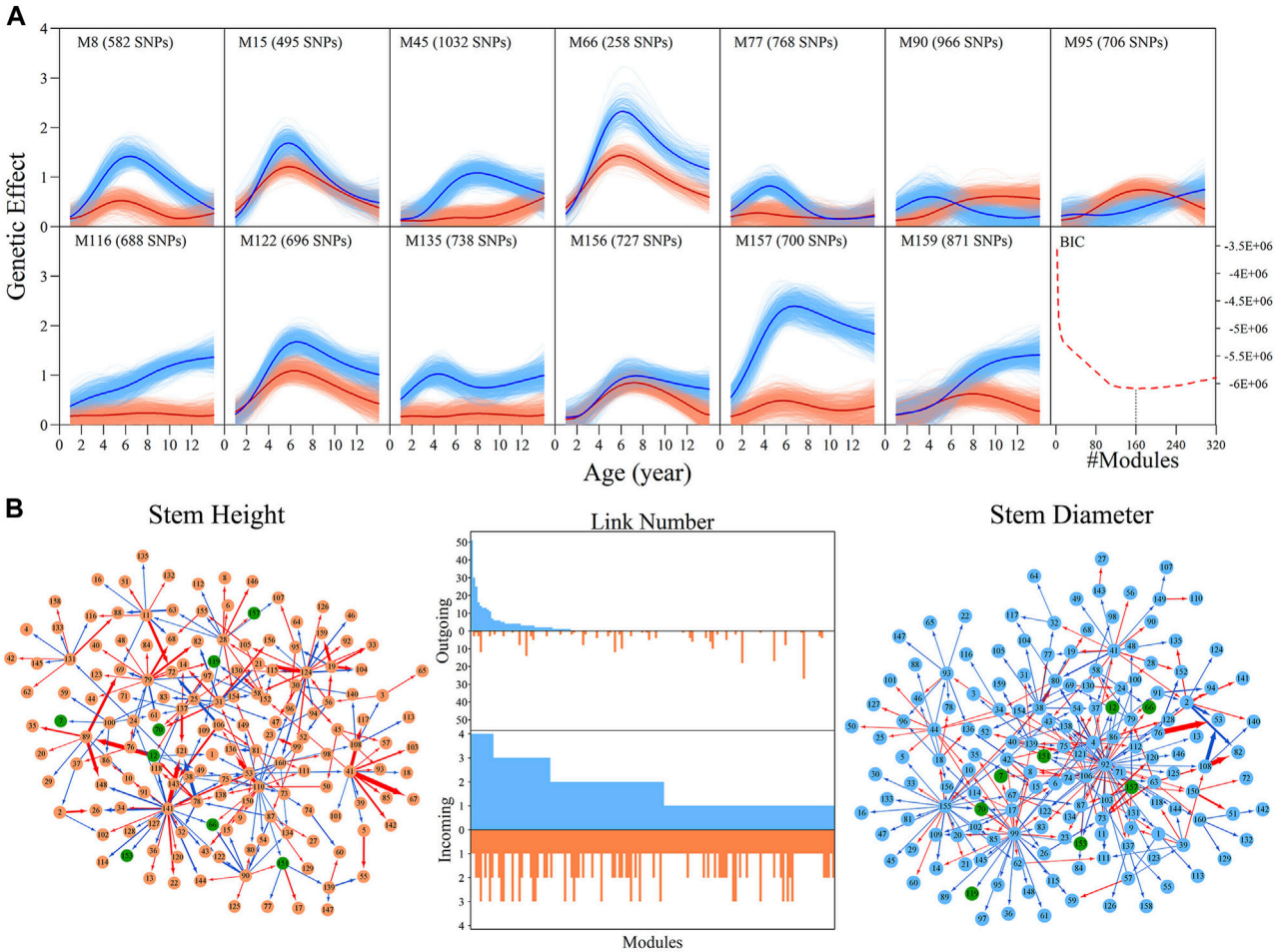


FIGURE 2 | Genetic effect clusters and metagalactic network modules for stem height and stem diameter growth in an interspecific full-sib family of *Populus*. **(A)** Genetic effect curves for 17 representative modules of stem height (red) and diameter (blue) chosen from a total of 160 gene modules detected by bivariate functional clustering; Bayesian Information Criterion analysis showed 160 as the optimal number of modules. **(B)** Metagalactic genetic networks containing 160 modules reconstructed using the mean effect values of each module for height and diameter, where red and blue arrowed lines denote inhibition and activation effects, respectively. The thickness of lines is proportional to the strength of the regulatory interaction. Modules containing quantitative trait loci (QTLs) are highlighted by green points. The distribution of the number of outgoing links and incoming links across 160 modules for height (red, bottom) and diameter (blue, top) is enumerated between networks.

The regulatory networks of gene pleiotropy in living organisms can be usefully compared the vast networks of stars in interstellar clusters, which in turn form galaxies, and then superclusters of galaxies. Although it is computationally complex, dimensions can be reduced by cluster. We incorporated modularity theory (Newman, 2006; Cantini et al., 2015) into network modeling in which nodes are densely connected in modules, with sparser connections between modules. In this paper, using a galactic analogy, we constructed metagalactic networks, intergalactic networks, and local interstellar networks to describe the layers of structure between modules, submodules and individual SNPs, respectively.

Genome-wide SNPs can be classified into different modules based on similarities in the pattern of genetic effects. We implemented bivariate functional clustering to classify the height and diameter effects of 156,362 SNPs into different modules. According to the comparison of BIC values for different cluster numbers, the most parsimonious number of modules was found to be 160. Each module represents a specific temporal pattern of genetic effects on the growth of height and diameter, which differs from those from other modules (**Supplementary Figure S3**). In **Figure 2A**, 13 representative modules are illustrated, which suggest pronounced discrepancies exist in the temporal patterns of genetic effects on the growth of both height and diameter. Some modules, such as M90 and M95, have greater genetic effects on height growth than diameter for a period of time, but some modules display an inverse pattern. All these differences in the time-dependent change of genetic effects contribute to the pleiotropic effects of the genetic architecture on growth in stem height and diameter.

We then explored how these 160 distinct modules are interconnected. We calculated the mean genetic effect curve

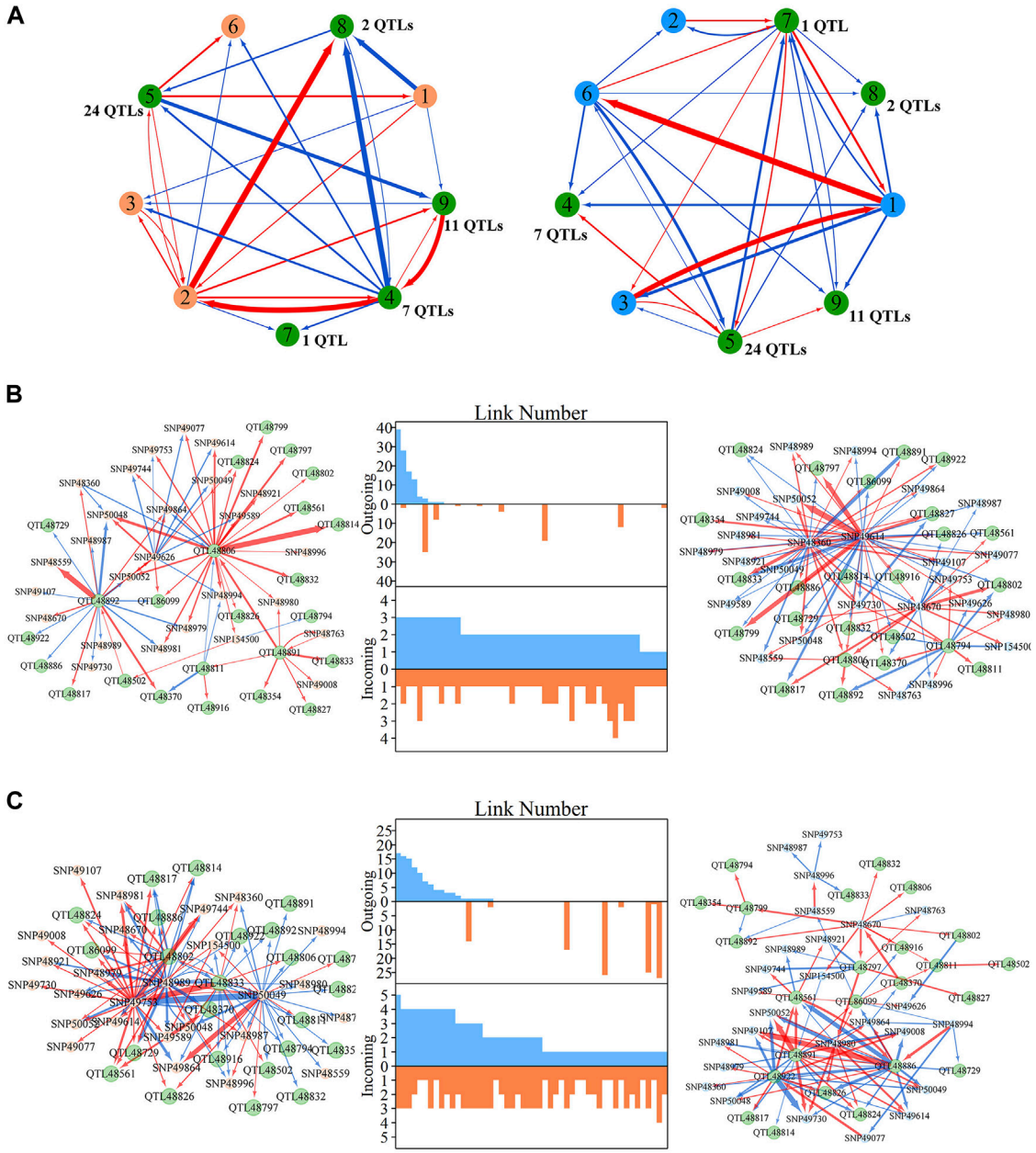


FIGURE 3 | Intergalactic genetic networks of quantitative trait loci (QTL)-containing submodules and local interstellar networks of individual single nucleotide polymorphisms (SNPs) for stem height and diameter. **(A)** Genetic networks among nine submodules of module M153. **(B)** Genetic networks among 50 SNPs from submodule 5, SM5/M153. **(C)** Independent genetic networks among 50 SNPs from submodule 5, SM5/M153. Red and blue arrowed lines represent inhibition and activation, where the thickness of lines is proportional to the strength of regulation. The submodules containing each QTL are highlighted by green points, and QTLs are highlighted within the intergalactic networks. The distribution of the number of outgoing and incoming links between SNPs for height (blue, top) and diameter (red, bottom) are counted between networks.

for each module to construct metagalactic genetic interaction networks among modules (**Figure 2B**), where nodes represent the collective effect of all SNPs within a module. This showed that both the height and diameter networks are highly sparse; directional positive and negative epistasis together dominate in the pairwise links. In both the height and diameter networks, positive epistasis constitutes a larger portion of the links: 51.04% in the height and 57.93% in the diameter growth networks, suggesting that genes tend to cooperate in the growth of these traits. Several negative links between modules indicated epistatic inhibition with a great strength, such as M41→M67 in the height network and M76→M53 in the diameter network. Outgoing and incoming links describe the activation or inhibition one module exerts on another, and the activation or inhibition of another

module on the module of interest, respectively. We counted the total number of outgoing and incoming links for each module in the network, and plotted the distribution (**Figure 2B**, middle panel). The numbers of outgoing links differed greatly across modules, ranging from 0 to 50; only a small subgroup of highly-interconnected modules predominated in the genetic network, and most modules were relatively minor nodes. The distribution of incoming links was much more consistent between modules. As shown in **Figure 2B**, we found that the character of the module is intricate, with outgoing and incoming links that varied between height and diameter networks: some nodes that were predominant in the height network, such as M141, were minor nodes in the diameter network. Conversely, modules such as M92 were predominant in the diameter network but minor in the height network. Some modules, such as M41, were also predominant in both networks.

QTL Deconstruction in Multilayer Network Architecture

Within metagalactic networks, 105 QTLs detected by system mapping resided in only eight different modules, M7 (3 QTLs), M12 (30 QTLs), M66 (6 QTLs), M70 (1 QTLs), M119 (1 QTLs), M151 (15 QTLs), M153 (45 QTLs) and M157 (4 QTLs), implying that the interplay between QTLs located in the same module may govern stem growth. Most QTL-containing modules played as minor roles, with more incoming than outgoing links (**Figure 2B**). We grouped QTL-containing modules into submodules, then constructed deep intergalactic networks to describe the connections among them. Module M153, which contained 186 SNPs, had the highest number of QTLs of all QTL-containing modules. By comparing BIC values between possible submodule arrangements, the most parsimonious number of submodules was found to be nine. In this intergalactic network organization, SM7/M153 is a minor submodule inhibited by SM2/M153 and SM4/M153 in the height network; however, SM7/M153 regulates other submodules in the diameter network. Although SM4/M153 has a predominant role in the height network, it is regulated by other submodules in the diameter network. From this intergalactic network, QTL-containing submodules exhibited considerable differences in effect on height and diameter, supporting their pleiotropic control of these complex growth traits (**Figure 3A**).

Local interstellar networks, at the individual SNP level, illustrate specific epistatic distinctions between height and diameter growth. We demonstrated how QTLs and other SNPs that were nonsignificant in the system mapping analysis interact with each other in QTL-containing, local interstellar networks by taking SM5/M153, which is the largest QTL-containing submodule involving 24 QTL in a total of 50 SNPs (**Figure 3B**), as an example. Based on the independent genetic effects calculated from the CRI differential model, independent local interstellar networks were also constructed (**Figure 3C**). In all cases, the distribution of outgoing links showed striking differences in numbers, originating from only a small portion of predominant SNPs in these networks. However, all SNPs received incoming links. In general, most QTLs essentially served as receivers, activated or inhibited by other SNPs. We also found that individual QTLs perform differently between the diameter and height networks: for example, in the SM5/M153 height network, QTLs 48806, 48811 and 48892 had the most outgoing links, tending to activate or inhibit other modules; while in the diameter network, the most predominant role was held by SNPs 49614 and 49360, followed by QTLs 48794 and 48806 (**Figure 3B**). This effect was similar in the independent local interstellar network. In **Figure 3C**, QTLs 48833 and 48802 were predominant nodes in height growth, whereas QTLs 48891, 48922 and 48886 predominated in diameter growth. In addition, there was a discrepancy in SNP organization between local interstellar networks and independent local interstellar network (**Figures 3B,C**). On one hand, regulatory roles, especially dominant nodes, change in the network structures. For example, QTL 48806 predominated in the local interstellar height network, while in the independent local interstellar height network, QTL 48833 was predominant. On the other hand, independent networks contained more frequent interactions among SNPs, indicating that the network structure of one trait may be influenced by the growth of the other traits.

We selected four QTLs from SM5/M153 to analyze the dynamics of inherent genetic effects and those influenced by other SNPs (**Figure 4**). QTL 48806 was predominant in both height and diameter networks, and had dramatic, independent effects, and was affected by other SNPs, which exerted an overall negative epistatic effect on this QTL. Thus, the net genetic effects were inferior to the independent effect. In the corresponding independent local interstellar network, a similar pattern of genetic effects was found for diameter, while in the height network, the expression of QTL 48806 is promoted by SNP 50049, and the net genetic effect is larger than the independent effect. We also found that QTL 48886 is upregulated by SNP 48892 in the height network and downregulated by SNP 49614 and SNP 48360 in the diameter network, and the same phenomenon was observed in the independent network. For QTL 86099, the overall genetic effects were similar to the independent effects for height growth, whereas the net genetic effect in diameter growth is greater than expected from its intrinsic capacity. Under independent growth, the diameter net genetic effect is smaller than the independent effect in 1–7 years, but gets stronger after the seventh year.

Nonsignificant Locus Analysis Within Modules

General analysis focused on the pleiotropic control of complex traits by significantly associated SNPs, ignoring those nonsignificant SNPs, which may also have indispensable roles in regulation and control. It is possible that some SNPs have an independent effect, but this effect may be diminished by negative epistatic effects from other SNPs. Here, we randomly chose a module without any significantly associated QTLs, M85, which comprised a total of 565 SNPs. According to the BIC values for clustering, 50 was found to be the most parsimonious number of clusters for the construction of intergalactic genetic networks (**Figure 5A**). In the height network, directional positive and

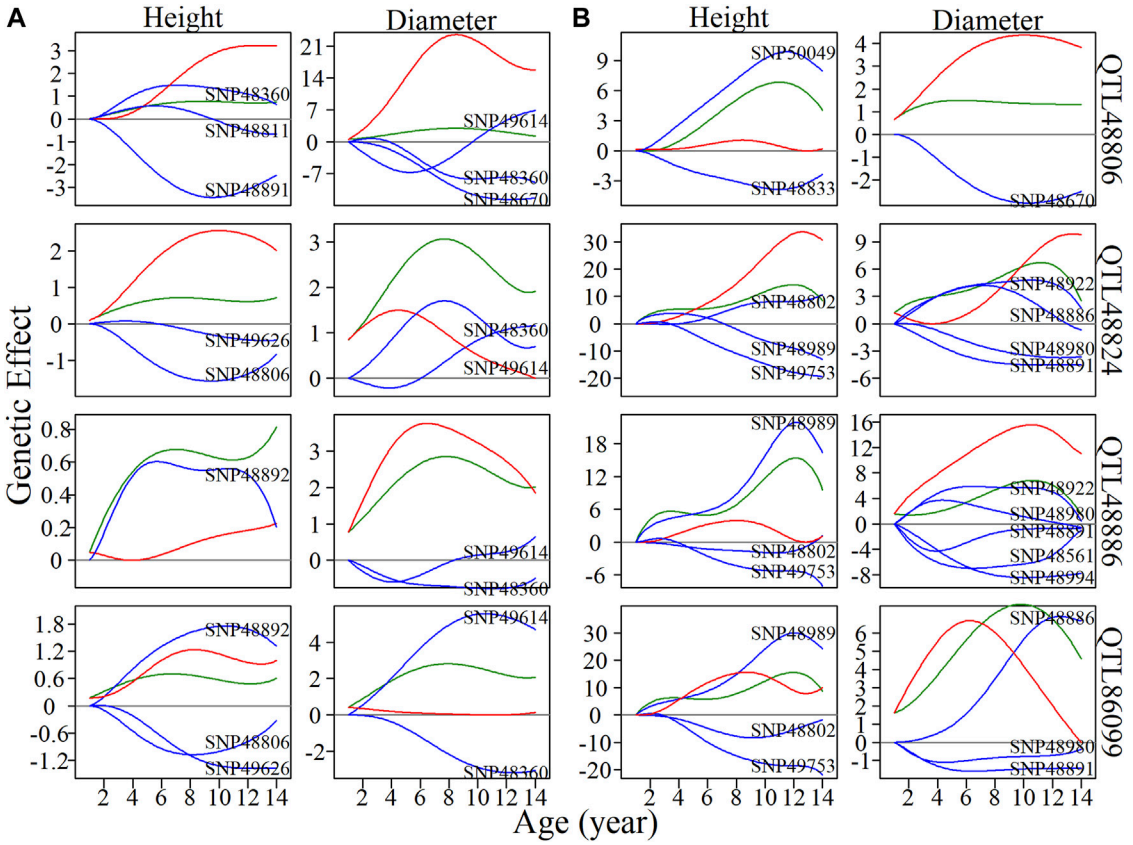


FIGURE 4 | Resolution of quantitative trait locus (QTL) overall and independent effects on stem height and diameter. **(A)** Genetic effect curves and **(B)** independent genetic effect curves of four QTLs from submodule SM5/M153. The net genetic effect of each QTL (green line) is deconstructed into the independent effects (red line) and effects that are dependent on other single nucleotide polymorphisms (SNPs; blue lines).

negative epistasis were found to be basically consistent in numbers of links, while the strength of negative regulation was much greater than for positive regulation, such as in the case of SM24↔SM26. In the diameter network, the directional positive epistatic effect was greater than that of negative epistasis. This indicates submodules tend to compete in their regulation of height growth, but reinforce each other's effects in regulating diameter growth.

SM23/M85 was found to be a minor submodule in both the height and diameter intergalactic networks, containing 20 SNPs and inhibited by SM22/M85, SM7/M85, and SM32/M85. We constructed overall and independent local interstellar networks for height and diameter (**Figures 5B,C**). These networks exhibited considerable differences in organization between height and diameter, as revealed by the distribution of outgoing and incoming links in the metagalactic networks. As with the intergalactic networks for M85, directional negative and positive links were approximately equal in the overall height network, and directional positive links outnumbered negative links in the corresponding diameter network. However, this effect disappeared in the independent diameter network, indicating the organizational structure of SNPs differs between independent and interactive perspectives of trait growth.

Four nonsignificant SNPs, 52542, 125076, 148746, and 148824 (**Figure 6**), were found to be located within the local interstellar network of SM23/M85. The overall genetic effect of SNP 52542, located in chromosome 5, on height growth is negatively regulated by SNP 4205 in the first 6 years, and is then upregulated by the same SNP. The net genetic effect on diameter growth was found to be larger than the independent effect. When considering independent effects, SNP 4205 is net-downregulated in height growth, and with net upregulation in diameter growth in the years 11–14. SNPs 125076, 148746 and 148824 were net-downregulated in height growth, in both the overall and independent networks, which offset their considerable independent effects. In contrast, the effects on diameter of SNPs 125076 and 148824 in the overall networks were found to be promoted by other SNPs. Similar effects were also observed for SNP 148824 in the independent diameter growth network: its independent effects were amplified by positive epistasis via incoming links. These examples show that SNPs may exhibit pronounced effects on trait growth if their negative regulators are silenced. However, in some cases, although an SNP may exhibit significant effects on diameter growth, it may also be negatively regulated by other SNPs in height growth, in which case the actual pleiotropic control of

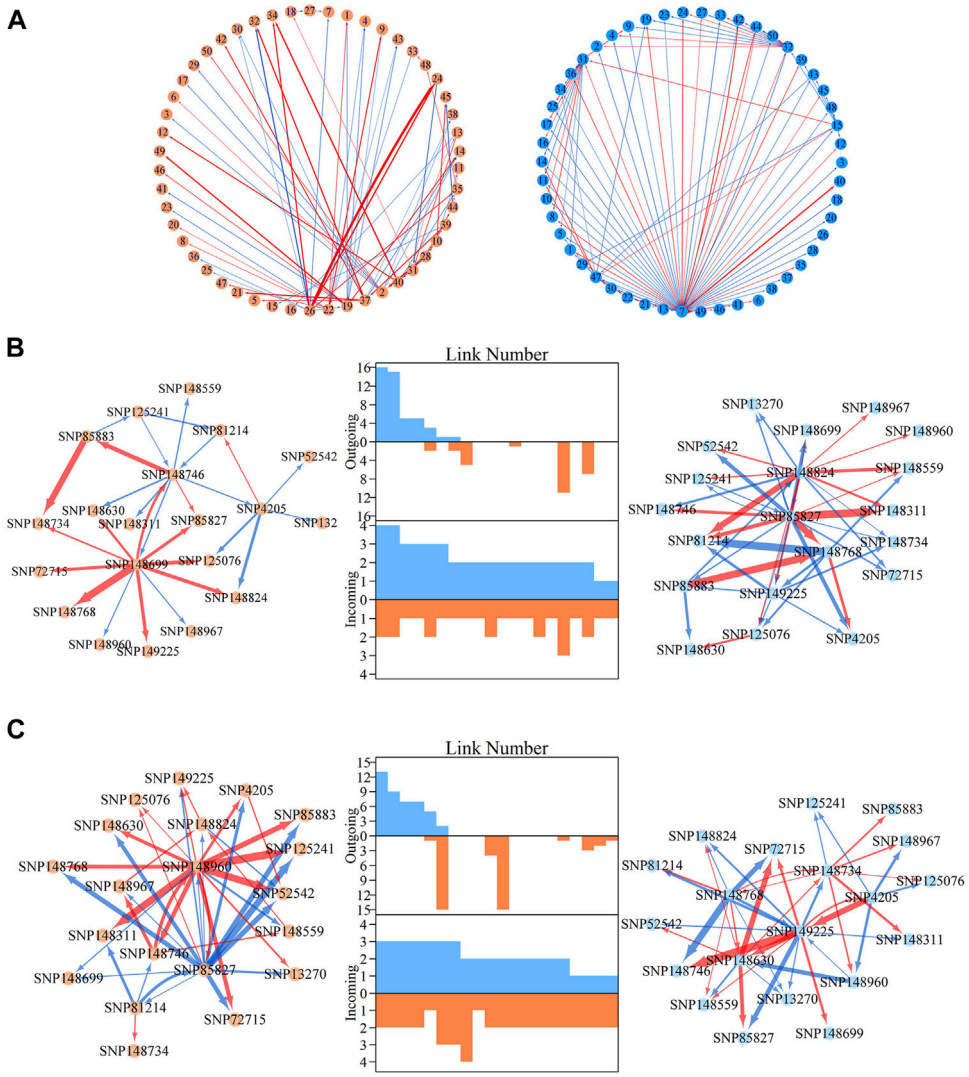


FIGURE 5 | Metagalactic genetic networks of module M85 and intergalactic networks of individual single nucleotide polymorphisms (SNPs) within module SM23/M85 for stem height and diameter. **(A)** Genetic networks among 50 submodules within module M85. **(B)** Genetic networks among 20 SNPs from submodule 23, SM23/M85. **(C)** Independent genetic networks among 20 SNPs from submodule 23, SM23/M85. Red and blue arrowed lines stand for inhibition and activation, with the thickness of lines proportional to the strength of regulation. The distribution of the number of outgoing links and incoming links across SNPs of height (blue, top) and diameter (red, bottom) is counted between networks.

complex traits by the SNP is likely to be neglected when traditional analytical approaches are used. The type of regulated SNP, as well as the positivity or negativity and strength of regulation, may vary throughout the growth of different traits, likely acting as an important driver of pleiotropic control.

DISCUSSION

Research on genetic structures has shown that genes often express pleiotropic effects over two or more traits; this is key to understanding pathways of gene action, assessing potential off-target effects of genetic manipulation with the aim of altering a specific pathway, and comprehending the effects of new mutations on evolution, potentially inducing both favorable and unfavorable effects on fitness (Hill and Zhang, 2012). Genetic mapping and association studies are highly dependent on statistical assumptions integrating reductionist thinking to detect individual, significantly-associated loci (Walling et al., 1998; Kemper et al., 2015; Vanhatalo et al., 2019). However, at the quantitative level, all genes may be pleiotropic in view of the highly interdependent and interactive nature of biological systems; however, until now the interplay pattern of genes is still unclear.

In this paper, we present a computational model integrating a novel growth equation, system mapping, functional clustering, developmental modularity theory, and evolutionary game theory.

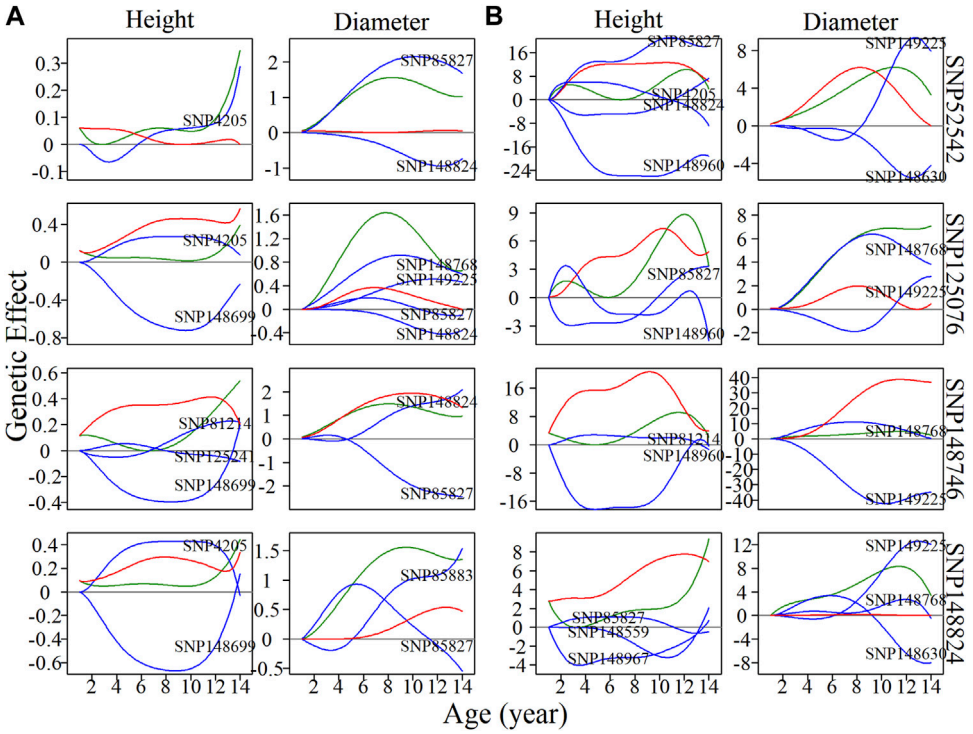


FIGURE 6 | Resolution of single nucleotide polymorphism (SNP) overall and independent effects on stem height and diameter. **(A)** Genetic effect curves and **(B)** independent genetic effect curves of four SNPs from submodule SM23/M85. The net genetic effect of each SNP (green line) is deconstructed into the independent effects (red line) and effects that are dependent on other SNPs (blue lines).

We established a CRI differential equation to describe the interaction of complex traits (stem height and diameter in forest poplar trees), which can reasonably quantify stem growth and internal structure. By deconstructing the growth of these two traits into a self-regulated part and one that is regulated by interactions between the co-existing traits in the system, our growth model can quantitatively describe the co-operative and antagonistic interactions between these traits in order to generate an in-depth understanding of the whole growth function, and reveal growth potential. We embedded the CRI differential equation into a QTL mapping model to identify important pleiotropic QTLs that play important roles in regulating the growth structure of the traits. On the other side, CRI differential equation provides a useful tool to estimate the net genetic effect of each SNP, which can resolve patterns of change over time based on the mathematical aspect of traits development.

CRI-based simulation studies show that heritability (proportion of genetic variance in the simulated phenotypic variance) and sample size affect mapping precision and power. The simulations are conducted with sample size as 66 (which is equal to the real data), 100 and 200, and heritability as 0.05, 0.1, respectively. For each simulation case, the proportion of simulation times of meaningful QTL screened out from 1,000 genetic markers of repeated simulation experiments is the mapping accuracy (Power). As seen in **Supplementary Table S3**, the mapping accuracy of QTL detection is above 0.530. The results show that system mapping can reasonably well estimate the time-varying trend of traits growth, even under a modest heritability (H^2 = 0.05) and a modest sample size (n = 66). We also found that the accuracy of effect estimation is sensitive to increasing heritability and sample size. On the other hand, in the absence of QTL expression (H^2 = 0), the same genetic sample size of 66, 100, and 200 was simulated with 1,000 genetic marker genes with false positive rates (FPRs) being generally below 0.055 (**Supplementary Table S3**). According to the QTL mapping accuracy and false positive probability at a series of different thresholds, we expressed ROC curves for different simulated sample sizes and heritability (**Supplementary Figure S5**). The area under the ROC curve (AUC) was calculated to assess the accuracy of QTL mapping. At the heritability level of 0.1, the AUC of the simulated quantities of the three sample sizes were all relatively high (>0.844). The parameter estimation results and estimated growth curves of different simulation scales are shown in **Supplementary Table S4 and Supplementary Figures S6, S7**. Our CRI-based QTL mapping has reasonably good statistical properties in interaction detection and FRP controlling.

The most important element of our framework is that the genetic architecture of complex traits is explored from omnigenic, genome-wide perspective (Boyle et al., 2017). One view proposed that the association between genes and traits was represented by a bipartite network and the presence of a modular structure detected by methods developed in physics (Barber, 2007). Our framework model constructed multilayer networks based on functional clustering to discern distinct network modules, in

which genes are linked more strongly to each other than to those in other modules. The top layer with the lowest resolution, called the metagalactic network, shows connections between modules; the next layer, the intergalactic network, has increased resolution, shows connection between submodules; and the bottom layer is the local interstellar network, which shows the interaction networks between SNPs and describes directional epistatic interactions. We deconstructed the net effect of genetic loci into independent and dependent effects, describing those in which the effect on complex traits is exerted directly through its own capacity, or indirectly through the regulation of other loci, respectively. The algorithmic aspects of the framework include curve smoothing, variable selection, matrix structuring, and ODE solving, each of which can be improved by introducing advanced theories and modern applied mathematical and statistical methods for future study.

Our model can be applied in general to reconstruct multilayer genetic networks, resolving the effects of genetic interactions and pleiotropy on the development of complex traits. The connection and regulation of the network may change with time or environment (Nie et al., 2017). There is potential extension for allowing a time-varying network instead of static model. In the context of organismal growth, our established framework can be used to further research the interaction of other multidimensional traits. For example, stem growth in trees includes the growth of some lateral organs and branches, in addition to the height and diameter of the stem that we included in our study. The multilayer interactome networks can also be extended from a two-dimensional to a multidimensional trait model, and an interactive regulation network of traits under pleiotropic control could be established, although such expansion will greatly increase the complexity of the model and the difficulty of computing. Our multilayer interactome network provides a robust and reliable modeling framework for assessing gene pleiotropy on traits and the interactions between the development of complex traits.

DATA AVAILABILITY STATEMENT

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

The computational code and data that support the findings of this study are available on request from the corresponding author.

AUTHOR CONTRIBUTIONS

HG performed data analysis and wrote the manuscript. SZ collected the data. XZ supported the scientific research funding for the publication of this study. QF participated in data analysis and simulation. X-YZ and RW conceived the study and designed the model and data analysis. All authors read and approved the manuscript.

FUNDING

This research was supported by the Fundamental Research Funds for the Central Universities (Grant Number 2017ZY45), the National Natural Science Foundation of China (Grant Number 11701028), the Fundamental Research Funds for the Central Universities (Grant Number BLX201912) (XZ) and the National Natural Science Foundation of China (Grant Number 11501032) (X-YZ). The work is partially supported by the Japan Society for the Promotion of Science (Grant Number 19K03613) (QF).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.769688/full#supplementary-material

REFERENCES

Ahmadi, K., and Sj, A. (2016). Generalized Height-Diameter Models for Fagus Orientalis Lipsky in Hyrcanian forest, Iran. *J. For. Sci.* 62 (9), 413–421. doi:10.17221/51/2016-JFS

Allesina, S., and Tang, S. (2012). Stability Criteria for Complex Ecosystems. *Nature* 483, 7388205–7388208. doi:10.1038/nature10832

Barber, M. J. (2007). Modularity and Community Detection in Bipartite Networks. *Phys. Rev. E* 76 (2), 066102. doi:10.1103/PhysRevE.76.066102

Bartholomé, J., Salmon, F., Vigneron, P., Bouvet, J.-M., Plomion, C., and Gion, J.-M. (2013). Plasticity of Primary and Secondary Growth Dynamics in Eucalyptushybrids: a Quantitative Genetics and QTL Mapping Perspective. *BMC Plant Biol.* 13 (1), 120–134. doi:10.1186/1471-2229-13-120

Bo, W., Fu, G., Wang, Z., Xu, F., Shen, Y., Xu, J., et al. (2014). Systems Mapping: How to Map Genes for Biomass Allocation toward an Ideotype. *Brief. Bioinform.* 15 (4), 660–669. doi:10.1093/bib/bbs089

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An Expanded View of Complex Traits: from Polygenic to Omnigenic. *Cell* 169 (7), 1177–1186. doi:10.1016/j.cell.2017.05.038

Bronisz, K., and Mehtätalo, L. (2020). Mixed-effects Generalized Height-Diameter Model for Young Silver Birch Stands on post-agricultural Lands. *For. Ecol. Manage.* 460, 117901. doi:10.1016/j.foreco.2020.117901

Busiello, D. M., Suweis, S., Hidalgo, J., and Maritan, A. (2017). Explorability and the Origin of Network Sparsity in Living Systems. *Sci. Rep.* 71, 12323. doi:10.1038/s41598-017-12521-1

Cantini, L., Medico, E., Fortunato, S., and Caselle, M. (2015). Detection of Gene Communities in Multi-Networks Reveals Cancer Drivers. *Sci. Rep.* 5, 17386. doi:10.1038/srep17386

Carter, G. W., Hays, M., Sherman, A., and Galitski, T. (2012). Use of Pleiotropy to Model Genetic Interactions in a Population. *Plos Genet.* 8 (10), e1003010. doi:10.1371/journal.pgen.1003010

Carter, S. L., Brechbuhler, C. M., Griffin, M., and Bond, A. T. (2004). Gene Co-expression Network Topology Provides a Framework for Molecular Characterization of Cellular State. *Bioinformatics* 20 (4), 2242–2250. doi:10.1093/bioinformatics/bth234

Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C., et al. (2019). Global Genetic Networks and the Genotype-To-Phenotype Relationship. *Cell* 177 (1), 85–100. doi:10.1016/j.cell.2019.01.033

Crecente-Campo, F., Tomé, M., Soares, P., and Diéguez-Aranda, U. (2010). A Generalized Nonlinear Mixed-Effects Height-Diameter Model for Eucalyptus Globulus L. In Northwestern Spain. *For. Ecol. Manage.* 259 (5), 943–952. doi:10.1016/j.foreco.2009.11.036

de Jong, H., Geiselmann, J., Hernandez, C., and Page, M. (2003). Genetic Network Analyzer: Qualitative Simulation of Genetic Regulatory Networks. *Bioinformatics* 19 (3), 336–344. doi:10.1093/bioinformatics/btf851

de Jong, H. (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *J. Comput. Biol.* 9 (1), 67–103. doi:10.1089/10665270252833208

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via theEMAlgorithm. *J. R. Stat. Soc. Ser. B (Methodological)* 39 (1), 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x

Dharmawardhana, P., Brunner, A. M., and Strauss, S. H. (2010). Genome-wide Transcriptome Analysis of the Transition from Primary to Secondary Stem Development in Populus Trichocarpa. *BMC Genomics* 11 (1), 150–169. doi:10.1186/1471-2164-11-150

Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R., and Church, G. M. (2005). A Global View of Pleiotropy and Phenotypically Derived Gene Function in Yeast. *Mol. Syst. Biol.* 1 (1), 1–11. doi:10.1038/msb4100004

Geiler-Samerotte, K. A., Li, S., Lazaris, C., Taylor, A., Ziv, N., Ramjeawan, C., et al. (2020). Extent and Context Dependence of Pleiotropy Revealed by High-Throughput Single-Cell Phenotyping. *Plos Biol.* 18 (8), e3000836. doi:10.1371/journal.pbio.3000836

Gompertz, B. (1833). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. In a Letter to Francis Baily, Esq. F. R. S. &c. By Benjamin Gompertz, Esq. F. R. S. *Proc. R. Soc. Lond.* 2, 252–253. doi:10.1098/rspl.1815.0271

Hill, W. G., and Zhang, X.-S. (2012). On the Pleiotropic Structure of the Genotype-Phenotype Map and the Evolvability of Complex Organisms. *Genetics* 1903, 1131–1137. doi:10.1534/genetics.111.135681

Huynh-Thu, V. A., and Sanguinetti, G. (2019). Gene Regulatory Network Inference: an Introductory Survey. *Methods Mol. Biol.* 1883, 1–23. doi:10.1007/978-1-4939-8882-2_1

Jiang, L., Ye, M., Zhu, S., Zhai, Y., Xu, M., Huang, M., et al. (2016). Computational Identification of Genes Modulating Stem Height-Diameter Allometry. *Plant Biotechnol. J.* 14 (12), 2254–2264. doi:10.1111/pbi.12579

Kearney, M. (2019). Reproductive Hyperallometry Does Not challenge Mechanistic Growth Models. *Trends Ecol. Evol.* 34, 275–276. doi:10.1016/j.tree.2018.12.006

Kemper, K. E., Reich, C. M., Bowman, P. J., vander Jagt, C. J., Chamberlain, A. J., Mason, B. A., et al. (2015). Improved Precision of QTL Mapping Using a Nonlinear Bayesian Method in a Multi-Breed Population Leads to Greater Accuracy of Across-Breed Genomic Predictions. *Genet. Sel. Evol.* 47 (1), 29–46. doi:10.1186/s12711-014-0074-4

Kim, B.-R., Zhang, L., Berg, A., Fan, J., and Wu, R. (2008). A Computational Approach to the Functional Clustering of Periodic Gene-Expression Profiles. *Genetics* 180 (2), 821–834. doi:10.1534/genetics.108.093690

Li, N., Mcmurry, T., Berg, A., Zhong, W., Berceli, S. A., and Wu, R. (2010). Functional Clustering of Periodic Transcriptional Profiles through ARMA(p Q). *PLoS One*, 5, 4e9894. doi:10.1371/journal.pone.0009894

Liao, X., Meyer, M. C., Chandler, R., and Smith, P. W. F. (2020). Estimation and Inference in Mixed Effect Regression Models Using Shape Constraints, with Application to Tree Height Estimation. *J. R. Stat. Soc. C* 69 (2), 353–375. doi:10.1111/rssc.12388

Liu, X., Li, Y. I., and Pritchard, J. K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177 (4), 1022–1034. doi:10.1016/j.cell.2019.04.014

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of Complex Networks. *Nature* 473, 7346167–7346173. doi:10.1038/nature10011

Lundqvist, B. (1957). On the Height Growth in Cultivated Stands of pine and spruce in Northern Sweden. *Medd. Fran. Statens Skogsforsk.* 47, 1–64.

Marbach, D., Costello, J. C., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., et al. (2012). Wisdom of Crowds for Robust Gene Network Inference. *Nat. Methods* 9 (8), 796–804. doi:10.1038/nmeth.2016

Margolin, A. A., and Califano, A. (2010). Theory and Limitations of Genetic Network Inference from Microarray Data. *Ann. N.Y Acad. Sci.* 1115 (1), 51–72. doi:10.1196/annals.1407.019

May, R. H. (1975). *Stability and Complexity in Model Ecosystems*. 2nd edition. Princeton: Princeton University Press.

Mazzocchi, F. (2012). Complexity and the Reductionism-Holism Debate in Systems Biology. *Wires Syst. Biol. Med.* 4 (5), 413–427. doi:10.1002/wsbm.1181

Michailidis, G., and d'Alché-Buc, F. (2013). Autoregressive Models for Gene Regulatory Network Inference: Sparsity, Stability and Causality Issues. *Math. Biosciences* 246 (2), 326–334. doi:10.1016/j.mbs.2013.10.003

Newman, M. E. J. (2006). Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Phys. Rev. E* 74 (3), 1539–3755. doi:10.1103/PhysRevE.74.036104

Nie, Y., Wang, L., and Cao, J. (2017). Estimating Time-varying Directed Gene Regulation Networks. *Biom* 73, 1231–1242. doi:10.1111/biom.12685

Osler, W. (1969). The Principles and Practice of Medicine. *Postgrad. Med. J.* 4537, 305. doi:10.1136/pgmj.45.522.305-a

Pandey, G. K., Cheong, Y. H., Kim, K.-N., Grant, J. J., Li, L., Hung, W., et al. (2004). The Calcium Sensor Calcineurin B-like 9 Modulates Abscisic Acid Sensitivity and Biosynthesis in Arabidopsis. *Plant Cell.* 16:7, 1912–1924. doi:10.1105/tpc.021311

Regenmortel, M. H. V. V. (2004). Reductionism and Complexity in Molecular Biology. *EMBO Rep.* 5 (11), 1016–1020. doi:10.1038/sj.embor.7400284

Richards, F. J. (1959). A Flexible Growth Function for Empirical Use. *J. Exp. Bot.* 10 (2), 290–301. doi:10.1093/jxb/10.2.290

Roth, C., Murray, D., Scott, A., Fu, C., Averette, A. F., Sun, S., et al. (2021). Pleiotropy and Epistasis within and between Signaling Pathways Defines the Genetic Architecture of Fungal Virulence. *Plos Genet.* 17 (1), e1009313. doi:10.1371/journal.pgen.1009313

Roukos, D. H. (2011). Networks Medicine: from Reductionism to Evidence of Complex Dynamic Biomolecular Interactions. *Pharmacogenomics* 12 (5), 695–698. doi:10.2217/pgs.11.28

Roya, T. (2020). Some Non-linear Height–Diameter Models Performance for Mixed Stand in Forests in Northwest iran. *J. Mt. Sci.* 17, 79–90. doi:10.1007/s11629-019-5870-4

Sharma, M., and Parton, J. (2007). Height-diameter Equations for Boreal Tree Species in Ontario Using a Mixed-Effects Modeling Approach. *For. Ecol. Manage.* 249 (3), 187–198. doi:10.1016/j.foreco.2007.05.006

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., et al. (2011). Abundant Pleiotropy in Human Complex Diseases and Traits. *Am. J. Hum. Genet.* 89 (5), 607–618. doi:10.1016/j.ajhg.2011.10.004

Smith, J. M., and Price, G. R. (1973). The Logic of Animal Conflict. *Nature* 246, 542715–542718. doi:10.1038/246015a0

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013). Pleiotropy in Complex Traits: Challenges and Strategies. *Nat. Rev. Genet.* 14 (7), 483–495. doi:10.1038/nrg3461

Someren, E. V., Wessels, L., Backer, E., and Reinders, M. (2002). Genetic Network Modeling. *Pharmacogenomics* 3 (4), 507–525. doi:10.1517/14622416.3.4.507

Sun, L., and Wu, R. (2015). Mapping Complex Traits as a Dynamic System. *Phys. Life Rev.* 13, 155–185. doi:10.1016/j.plrev.2015.02.007

Temesgen, H., and v. Gadow, K. (2004). Generalized Height-Diameter Models-An Application for Major Tree Species in Complex Stands of interior British Columbia. *Eur. J. For. Res* 123 (1), 45–51. doi:10.1007/s10342-004-0020-z

Tibshirani, R. (2011). Regression Shrinkage and Selection via the Lasso: a Retrospective. *J. R. Stat. Soc. B.* 73 (2), 273–282. doi:10.1111/j.1467-9868.2011.00771.x

Vanhatalo, J., Li, Z., and Sillanpää, M. J. (2019). A Gaussian Process Model and Bayesian Variable Selection for Mapping Function-Valued Quantitative Traits with Incomplete Phenotypic Data. *Bioinformatics* 35, 3684–3692. doi:10.1093/bioinformatics/btz164

Visscher, P. M., and Yang, J. (2016). A Plethora of Pleiotropy across Complex Traits. *Nat. Genet.* 48 (7), 707–708. doi:10.1038/ng.3604

Walling, G. A., Visscher, P. M., and Haley, C. S. (1998). A Comparison of Bootstrap Methods to Construct Confidence Intervals in QTL Mapping. *Genet. Res.* 71 (2), 171–180. doi:10.1017/S0016672398003164

Wang, Y. X. R., and Huang, H. (2014). Review on Statistical Methods for Gene Network Reconstruction Using Expression Data. *J. Theor. Biol.* 362, 53–61. doi:10.1016/j.jtbi.2014.03.040

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). How to Cluster Gene Expression Dynamics in Response to Environmental Signals. *Brief. Bioinform.* 13 (2), 162–174. doi:10.1093/bib/bbr032

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., et al. (2019). A Global Overview of Pleiotropy and Genetic Architecture in Complex Traits. *Nat. Genet.* 51 (9), 1339–1348. doi:10.1038/s41588-019-0481-0

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., and Visscher, P. M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* 173 (7), 1573–1580. doi:10.1016/j.cell.2018.05.051

Wu, R.-L., Wang, M., and Huang, M. (1994). Quantitative Genetics of Yield Breeding forPopulus Short Rotation Culture. III. Efficiency of Indirect Selection on Tree Geometry. *Theoret. Appl. Genet.* 88 (6-7), 803–811. doi:10.1007/BF01253989

Wu, R., Cao, J., Huang, Z., Wang, Z., Gai, J., and Vallejos, E. (2011). Systems Mapping: How to Improve the Genetic Mapping of Complex Traits through Design Principles of Biological Systems. *BMC Syst. Biol.* 5, 84. doi:10.1186/1752-0509-5-84

Wu, R., and Jiang, L. (2021). Recovering Dynamic Networks in Big Static Datasets. *Phys. Rep.* 912 (6062), 1–57. doi:10.1016/j.physrep.2021.01.003

Xu, M., Jiang, L., Zhu, S., Zhou, C., Ye, M., Mao, K., et al. (2016). A Computational Framework for Mapping the Timing of Vegetative Phase Change. *New Phytol.* 211 (2), 750–760. doi:10.1111/nph.13907

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat. Genet.* 42, 565–569. doi:10.1038/ng.608

Yukilevich, R., Lachance, J., Aoki, F., and True, J. R. (2008). Long-term Adaptation of Epistatic Genetic Networks. *Evolution* 62 (9), 2215–2235. doi:10.1111/j.1558-5646.2008.00445.x

Zhao, W., Chen, Y. Q., Casella, G., Cheverud, J. M., and Wu, R. (2005a). A Non-stationary Model for Functional Mapping of Complex Traits. *Bioinformatics* 21 (10), 2469–2477. doi:10.1093/bioinformatics/bti382

Zhao, W., Hou, W., Littell, R. C., and Wu, R. (2005b). Structured Antedependence Models for Functional Mapping of Multiple Longitudinal Traits. *Stat. Appl. Genet. Mol.* 4 (1), 1–28. doi:10.2202/1544-6115.1136

Zhao, W., Wu, R., Ma, C.-X., and Casella, G. (2004). A Fast Algorithm for Functional Mapping of Complex Traits. *Genetics* 167 (4), 2133–2137. doi:10.1534/genetics.103.024844

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.

Copyright © 2021 Gong, Zhu, Zhu, Fang, Zhang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 November 2021
doi: 10.3389/fgene.2021.758003

Integrated Lipidomic and Transcriptomic Analysis Reveals Lipid Metabolism in Foxtail Millet (*Setaria italica*)

*Haiying Zhang[1†], Junyou Wang[1†], Jing Zhao[1†], Changqing Sun[1], Jin Wang[1], Qian Wang[2], Fei Qu[1], Xiaodong Yun[1] and Zhiwei Feng[3]**

[1]College of Agriculture, Shanxi Agricultural University, Taigu, China, [2]Hebei Zhihai Technology Co., Ltd., Xingtai, China, [3]Shanxi Institute of Organic Dryland Farming, Shanxi Agricultural University, Taiyuan, China

Foxtail millet (*Setaria italica*) as the main traditional crop in China, is rich in many kinds of high quality fatty acids (FAs). In this study, Ultra-high performance liquid chromatography-time-of-flight-tandem mass spectrometer (UHPLC-Q-TOF-MS/MS) was used to determine the lipids of JG35 and JG39. A total of 2,633 lipid molecules and 31 lipid subclasses were identified, mainly including thirteen kinds of glycerophospholipids (GP), eleven kinds of glycerolipids (GL), four kinds of sphingolipids (SP), two kinds of fatty acyls (FA) and one kind of sterol (ST). Among them JG35 had higher contents of diacylglycerols (DG) and ceramides (Cer), while triacylglycerols, phosphatidyl ethanolamine, phosphatidic acid, sterol, fatty acyls and pardiolipin (TG, PE, PA, ST, FA and CL) were higher in JG39. Meantime, the correlation analysis of lipidomics and transcriptomics was used to map the main differential lipid metabolism pathways of foxtail millet. The results shown that a differentially expressed genes (DEGs) of *FATA/B* for the synthesis of FA was highly expressed in JG35, and the related genes for the synthesis DG (*ACCase*, *KAS*, *HAD*, *KCS*, *LACS* and *GAPT*), TG (*DGAT* and *PDAT*) and CL (*CLS*) were highly expressed in JG39. The results of this study will provide a theoretical basis for the future study of lipidomics, improvement of lipid quality directionally and breeding of idiosyncratic quality varieties in foxtail millet.

Keywords: foxtail millet, UPLC-Q-TOF-MS/MS, lipidomics, transcriptome, association analysis

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Dawei Xue,
Hangzhou Normal University, China
Chengsong Zhu,
University of Texas Southwestern Medical Center, United States

****Correspondence:***
Zhiwei Feng
zhiweifeng@126.com

[†]These authors have co ntributed equally to this work and share first authorship

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

Received: *13 August 2021*
Accepted: *27 October 2021*
Published: *16 November 2021*

Citation:
*Zhang H, Wang J, Zhao J, Sun C, Wang J, Wang Q, Qu F, Yun X and Feng Z (2021) Integrated Lipidomic and Transcriptomic Analysis Reveals Lipid Metabolism in Foxtail Millet (*Setaria italica*). Front. Genet. 12:758003. doi: 10.3389/fgene.2021.758003*

INTRODUCTION

Foxtail millet (*Setaria italica*) is one of the most ancient grain crops originating from China and is among the top five most important cereals worldwide (Qin et al., 2020; Zhao et al., 2021). The crop is drought and barren tolerant, with a short life cycle (Pan et al., 2018; Li et al., 2021). Dehusked foxtail millet is rich in many kinds of nutrients (Verma et al., 2015; Meherunnahar et al., 2018; Trivedi et al., 2018), and has a low-fat content (about 2.8–8.0%) with high-quality FAs. The main unsaturated fatty acids (UFAs) in dehusked foxtail millet are linoleic acid, oleic acid, and linolenic acid, whereas the main saturated fatty acids (SFAs) are palmitic acid, stearic acid, and arachidic acid. Among them, UFAs represent a large proportion, accounting for about 85.54% of the total FAs. UFAs exhibit unique biological activity, and thus, have significant physiological functions in the human body (Mccloy et al., 2004). Linoleic acid and linolenic acid are cell components and precursors of

prostaglandins that participate in phospholipid synthesis. Also, they are closely related to cholesterol transport, which is crucial in regulating human physiological functions and safeguarding physical health (Fitzsimons et al., 2019; Samson et al., 2020; Saini and Rai, 2021).

Lipids are hydrophobic or bisexual organic molecules that are the main component of biofilm structures. Besides, they function as small signaling molecules and energy substances that mediate many biological processes. The International Commission on Lipopid Classification and Nomenclature classifies lipids into eight categories: fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), saccharolipids (SL), sterol lipids (ST), prenol lipids (PR) and polyketides (PK) (Checa et al., 2015). Lipidomics is a systematic qualitative and quantitative analysis of various lipids within the body of an organism at the molecular level. It can be used to efficiently study the changes and functions of lipid families and lipid molecules in various biological processes, which is important in clarifying the processes and mechanisms of related biological activities (Wenk, 2005). Currently, the most common and convenient method for lipidomic analysis is high performance liquid chromatography-mass spectrometry (HPLC-MS/MS) and Ultra High Performance Liquid Chromatography-Time of Flight-Mass Spectrom (UHPLC-TOF MS/MS). The latter, as a high resolution mass spectrometry, has the characteristics of good stability, fast scanning speed and high sensitivity, and has been widely used in the extraction of a variety of chemical components (Wu et al., 2018; Singh et al., 2019; Yong et al., 2020).

Since its development, lipidomics has been widely applied in biomedicine, human health, and other aspects (Eggers et al., 2017; Kus et al., 2018; Wu et al., 2020; Ecker et al., 2021; Marasca et al., 2021). Lipids have been shown to play an essential role in plant growth, photosynthesis, and signal transduction (Hou et al., 2016; Rodríguez-López et al., 2017). Therefore, more and more studies on lipid metabolism have been conducted in plants. Yao et al. (2017) performed a comparative lipidomic analysis of lipid droplets in the mesocarp and seed tissues of Chinese tallow trees using LC-MS. The results showed that the most abundant triacylglycerol (TG) species in the mesocarp included one C18:1, two C16:0, and FAs. However, the three C18 FAs with higher unsaturated levels are dominant in the lipid droplets from grains. Non-targeted lipidomics has been conducted to study lipid antioxidation and galactose lipid remodeling in tomato plants under temperature stress (Spicher et al., 2016). In total, 791 lipid molecules were identified; and it was found that the remodeling of the thylakoid membrane in the chloroplast matrix is affected by fatty acid saturation in glycolipids and lipid oxidation levels at high temperatures.

With the completion of foxtail millet genome sequencing and the development of bioinformatics, it has provided convenient conditions for digging into the excellent functional genes of foxtail millet (Zhang et al., 2012). Transcriptomic sequencing can be used to sequence almost all transcripts in specific biological tissues or cells at a certain period. Additionally, it can be applied to study gene expression, structure, variable splicing and to predict new transcription (Patel et al., 2021; Sun et al., 2021; Wang et al., 2021; Zhou et al., 2021). Fei et al. (2020) employed transcriptome sequencing and RT-qPCR to determine the expression levels of 20 genes related to fatty acid synthesis in *Zanthoxylum bungeanum* seeds, the results of intergroup correlation and RDA analysis suggested that *ENR*, *ECR*, and *SAD1* are the key genes in fatty acid synthesis. Meng et al. (2021) used transcriptomics to predict the key genes of unsaturated fatty acid (UFA) synthesis and oil accumulation in *Paeonia lactiflora* seeds, including *MCAT*, *KASIII*, *FATA*, *SAD*, *FAD2*, *FAD3*, and *DGAT*, providing more comprehensive genomic resources for understanding the transcription of genes in *P. lactiflora* seeds.

Metabolomics considers organisms to be dynamic integration. It is applied to explore the process of metabolic changes caused by internal or external factors through statistical analysis (Rai et al., 2017). As a branch of metabolomics, lipidomics, in combination with transcriptology, can be used to identify and analyze the relevant genes in the metabolic pathway and precisely reflect the changes in the organism *per se* (Shen et al., 2016). However, no comprehensive analysis of lipid components and metabolism in foxtail millet has been reported. In this study, UHPLC-Q-TOF-MS/MS was used to compare and identify the lipid molecules of two different foxtail millet varieties. Further, transcriptomic analysis was used to explore the correlation between lipid metabolites and related gene expression patterns and identify the factors affecting lipid content in foxtail millet, which provides insights into the molecular biological mechanism of lipid metabolism, study of functional nutritional components and the selection of high quality foxtail millet varieties.

MATERIALS AND METHODS

Plant Materials

Two foxtail millet varieties JG35 and JG39, were bred at the Institute of Foxtail Millet Crops, Hebei Academy of Agriculture and Forestry Sciences. Both of these two varieties were grown in summer sowing areas, and were resistant to the herbicide of imazethapyr and sethoxydim. However, they had greatly difference in the crude fat content. JG35 is a high-fat variety with the content of 5.4%, whereas JG39 is a low-fat variety, with the content of 2.9%. Samples were collected from the Yulin test site (109°21′ E, 37°56′ N, with an elevation of 1,120 m, annual average precipitation of 436 mm, and an annual average temperature of 8.0°C), seeds were selected 28 days after pollination with three replicates. And the transcriptome data of seeds were obtained from the NCBI Sequence Read Archive (SRA) under accessions with the SRR number 11840494, 11840495, 11840514, 11840501, 11840502, and 11840503, which were reported by Yuan et al. (2021).

Sample Processing

Representative whole seeds without mildew, particles, or other impurities were selected and unshelled into foxtail millet using a JLGL-45 hulling mill (brand: Xinfeng, manufactured by Taizhou Grain Instruments Factory, Zhejiang Province). Next, the seeds were ground into powder using liquid nitrogen and mixed evenly

for subsequent analysis. Precisely, 50 mg sample was weighed and put into an EP tube containing 200 μL water. The sample was then blended through whirl pooling for 30 s, then ground using friction bails for 12 min at 1,000 rpm. Next, the sample was subjected to 15 min of ultrasonic processing in an ice-water bath, then 480 μL extract (MTBE: methanol = 5:1) was added, followed by 30 s of blending through whirlpooling and 10 min of ultrasonic processing in an ice-water bath. After that, the sample was left to stand for 1 h at −20°C and centrifuged for 15 min at 4°C, 10,000 rpm. A total of 380 μL supernatant was carefully extracted and added into an EP tube. The extract was dried in a vacuum concentrator then mixed with 200 μL solution (dichloromethane: methanol = 1:1) for redissolution. Next, the mixture was blended through whirlpooling for 30 s and subjected to ultrasonic processing in an ice-water bath for 10 min. The sample was then centrifuged for 15 min at 4°C, 13,000 rpm, and 180 μL supernatant was carefully extracted and added into a vial. Exactly 10 μL of each sample was taken and mixed into the QC sample for testing on the machine. All reagents used were chromatographically pure.

Chromatographic Condition

Chromatographic separation was performed on a Waters UPLC Acquity I-Class PLUS ultra-high performance liquid chromatography system. The chromatography column used was the Acquity UPLC CSH C18 (1.7 μM, 2.1 × 100 mm) column purchased from Waters. Positive ion mode: mobile phase A: 60% acetonitrile water solution, 10 mM ammonium acetate, 0.1% formic acid; mobile phase B: 90% isopropanol/acetonitrile solution, 10 mM ammonium acetate, 0.1% formic acid. Negative ion mode: mobile phase A: 60% acetonitrile water solution, 10 mM ammonium acetate, 0.1% formic acid; mobile phase B: 90% isopropanol/acetonitrile solution, 10 mM ammonium acetate, 0.1% formic acid. The gradient elution was programmed as follows: 0–2.0 min, 40% B; 2–18 min, 43%–100% B; 18–20 min, 40% B. Flow rate: 0.4 ml/min; injection volume: 5 μL; column temperature: 55°C. The sample was placed in a 10°C automatic sampler for analysis.

Mass Spectrometry Conditions

Primary and secondary mass spectrometric data were collected under software (MassLynx V4.2, Waters) controlled MSe mode using a Waters Xevo G2-XS QTOF high-resolution mass spectrometer. Data on the low and high collision energy were collected by two channels in each data collection cycle. The low collision energy was 2 V, whereas the high collision energy interval was 10–40 V. The scanning frequency was 0.2 s/mass spectrum. ESI ion source parameters: capillary voltage: 2,000 V (positive ion mode) or −1,500 V (negative ion mode); taper hole voltage: 30 V; ion source temperature: 120°C; desolvent gas temperature: 550°C; blowback air flow rate: 50 L/h; desolvent gas flow rate: 900 L/h.

Lipid Data Analysis

The original data collected by MassLynx V4.2 were put through peak recognition, peak extraction, peak alignment, and other data processing operations using Progenesis QI software. Identification was performed using Progenesis QI software's online LMSD (Lipid Maps Structure Database) database and BMK's self-built database, and theoretical fragment recognition was conducted. The Precursor Ion mass number deviation was within 100 ppm in each case, the fragment ion mass number deviation was within 50 ppm in each case (Want et al., 2010; Dunn et al., 2011; Kuhl et al., 2012). The extracted data were subjected to quantitative and lipid composition analyses, and diagram R language and SigmaPlot12.0 were drawn.

Transcriptome Data Analysis

Differential expression analysis was performed using the DESeq2. The resulting p values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted p-value < 0.01 and $|\log_2 FC| > 1$ found by DESeq2 were assigned as differentially expressed. Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GOseq R packages based Wallenius non-central hyper-geometric distribution (Young et al., 2010), which can adjust for gene length bias in DEGs. And KOBAS (Mao et al., 2005) software was used to test the statistical enrichment of DEGs in KEGG pathways.

RESULTS

Data Quality Assessment

Spectrum Comparison of the QC Sample

In order to understand the sampling situation of the tested samples and the reliability of test results, the QC sample was added to monitor the stability and repeatability of the test system. As shown in **Figure 1**, the base peak chromatograms (BPI) of the three QC samples were overlapped and compared, and the results shown that the chromatographic peak response intensity and retention time of different samples overlapped, suggesting that the data collection system was stable and reliable enough for subsequent lipidomic analysis of foxtail millet.

Univariate Analysis of Lipid Molecules

The univariate statistical methods commonly used for differential analysis of two sample groups include variation fold analysis and the t-test/non-parametric test. All metabolites detected under positive and negative ion modes (including unidentified metabolites) were subjected to differential analysis based on univariate analysis, where the red denotes the differential metabolites with $|\log_2 FC| > 1$, and adjusted p-value < 0.05. The volcano plot of differential expression was shown in **Figure 2**.

Multivariate Analysis of Lipid Molecules- Orthogonal Projections to Latent Structures-Discrimination Analysis

Orthogonal projection to latent structure-discriminant analysis (OPLS-DA) was performed to determine the lipid micromolecules with significant changes in the JG35 and JG39 millet varieties, thus building the relational model between the expression levels of lipids and the sample. The evaluation

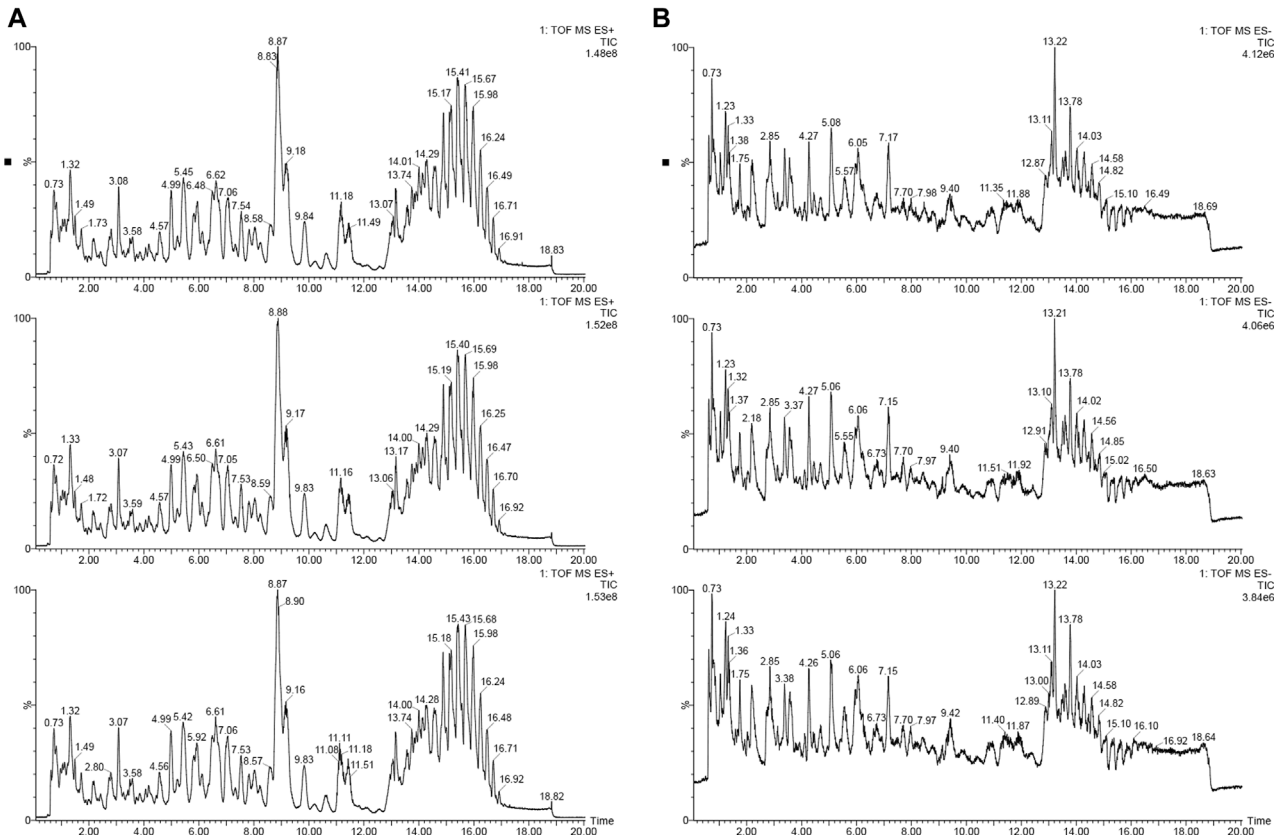


FIGURE 1 | The base peak intensity chromatogram (BPI) of positive and negative ion [**(A)**: positive iron; **(B)**: negative ion].

parameters of the model mainly included R^2X, R^2Y, and Q^2. R^2X and R^2Y denote the explanatory powers of the model built for X and Y matrices, respectively, and Q^2 denotes the prediction rate of the model. Generally, the closer the values of R^2Y and Q^2 to 1, the higher the stability and reliability of the model. Under normal circumstances, the model is considered effective when $Q^2 > 0.5$, and excellent when $Q^2 > 0.9$. According to our results, the parameters R^2Y and Q^2 of the model were 0.998 and 0.832, respectively (**Figure 3A**), indicating that the model was relatively reliable. To prevent overfitting of the model built, we performed permutation analysis to test the model and ensure its effectiveness. Normally, when the intercept between the Q^2 regression line and the Y-axis is < 0.05, there is no overfitting of the model built. In this study, the intercept was uniformly <0.05 (**Figure 3B**), indicating that there was no over fitting of the model. Overall, these results suggest that the samples tested in this study have good repeatability and stability, thus can be used for subsequent lipid metabolism analysis.

Lipid Composition Analysis

Lipid Composition Identification and Analysis in Grains

Complete lipidomic analysis was performed on grain samples from two foxtail millet varieties using UHPLC-QTOF-MS. According to the results, 31 lipid subclasses (2,633 lipid molecules) were identified, including 13 kinds of GP, 11 kinds of GL, 4 kinds of SP, 2 kinds of FA and 1 kind of ST (**Figure 4**). Among them, GP had the largest number (793) including 120 for PE, 115 for CL, 101 for PS, 98 for PG, 96 for PA, 89 for PC, 53 for PI, 32 for LPE, 26 for LPS, 21 for LPC, 21 for LPA, 14 for LPG, 7 for LPI. The followed in succession was two kinds of FA included FA (671) and WE (60). 578 kinds of GL species included 295 for TG, 216 for DG, 30 for MG, 12 for MGDG, 6 for SQDG, 4 for DGDG, 4 for DGTS, 3 for SQMD, 3 for MGMG, 2 for DGMG, 2 for and GlcADG. 224 kinds of SP included 179 for Cer, 18 for SM, 15 for MIPC, 12 for IPC. The last category included 307 kinds of ST. It can be seen that the number of lipid molecules varied significantly across different lipid subclasses, and FA had the largest number.

Lipid Content of Grains

Among the two foxtail millet varieties, JG35 had a relative lipid content of 1.21×10^8, whereas the content was 1.27×10^8 in JG39. Overall, the relative contents of various lipid compositions in JG35 and JG39 were similar, and GL content was the highest, accounting for 48.24 and 47.55% of the total lipid content, respectively (**Figure 5**). Specifically, the contents of TG and DG were relatively higher in the two foxtail millet varieties, and the relative content were 3.64×10^7vs3.94×10^7, and 1.97×10^7vs1.92×10^7, respectively. The next content was GP, accounting for 28.44 and 30.01% of the total lipid content,

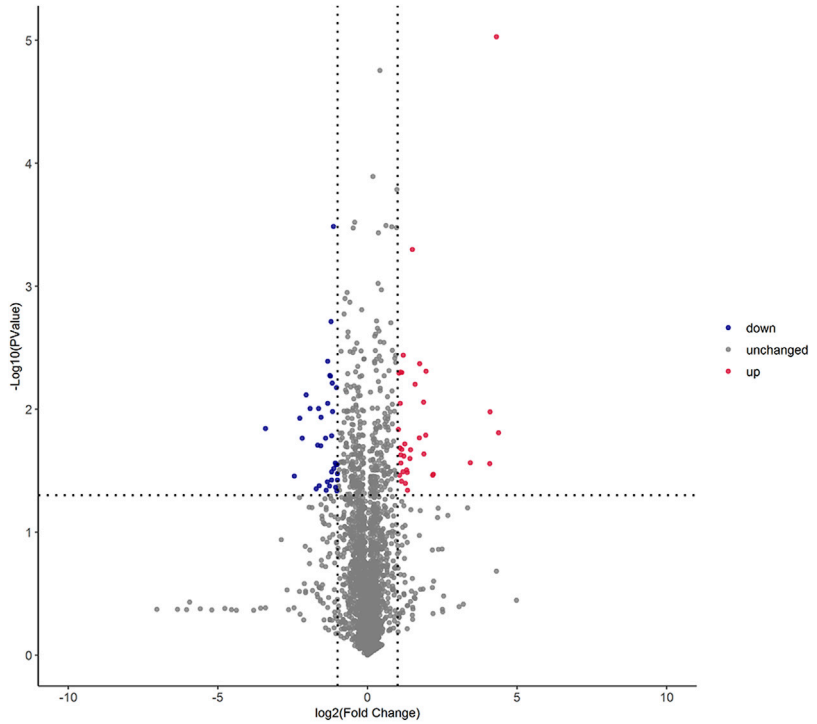


FIGURE 2 | Volcano graph of differential lipid molecules in two foxtail millet. Note: Each dot in the volcano plot denotes one lipid. The x-axis denotes the fold change of each substance under comparison in the group (adopting logarithm with a base of 2), and the y-axis denotes the *p*-value of the *t*-test (adopting logarithm with a base of 10). The scatter size denotes the VIP value of the OPLS-DA model: the larger the scatter size, the greater the VIP value, and the more reliable the differentially expressed lipid obtained through screening. The blue spots denote lipids with down-regulated differential expression; the red spots denote lipids with up-regulated differential expression; the gray spots denote the detected lipids without significant differences.

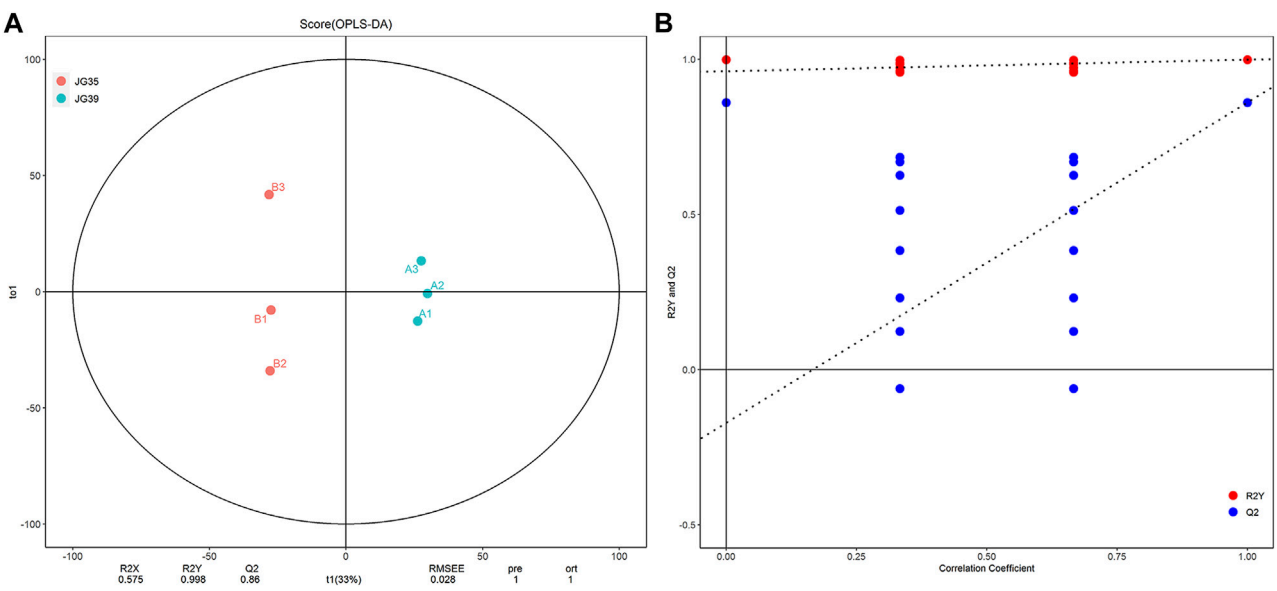


FIGURE 3 | Multivariate statistical analysis of lipid molecules. **(A)** The Score chart of OPLS-DA. **(B)** Model validation diagram of OPLS-DA.

respectively. SP (containing Cer, SM, MIPC and IPC) accounted for 13.10 and 12.31%, respectively. FA (including FA and WE) accounted for 5.46 and 5.42%, respectively. The contents of ST in JG35 and JG39 were 5.77×10^6 and 6.01×10^6, respectively. In conclusion, it was found that JG35 had higher DG and Cer, while JG39 had higher TG, PE, PA, ST, FA, and CL.

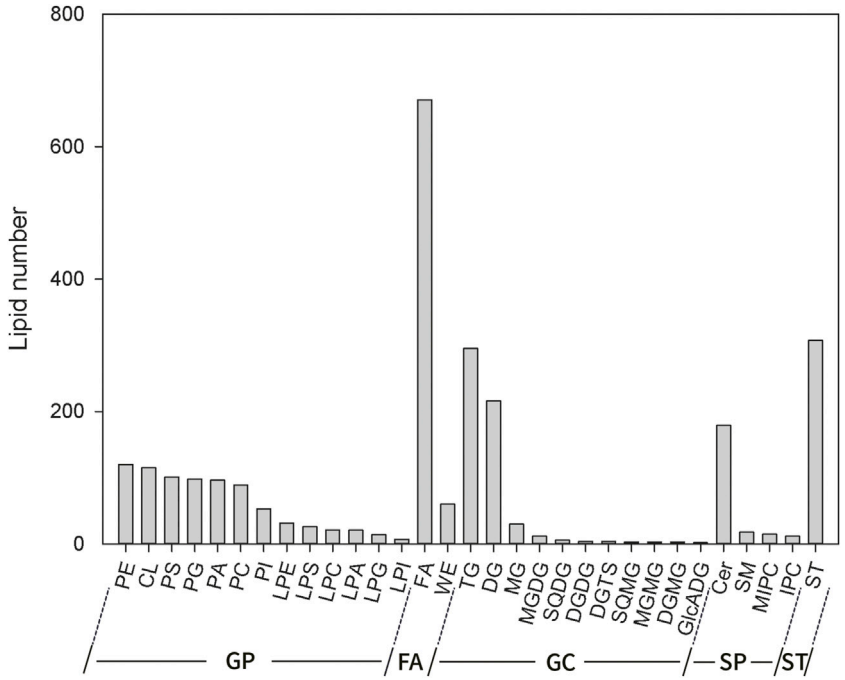


FIGURE 4 | Lipid composition of foxtail millet.

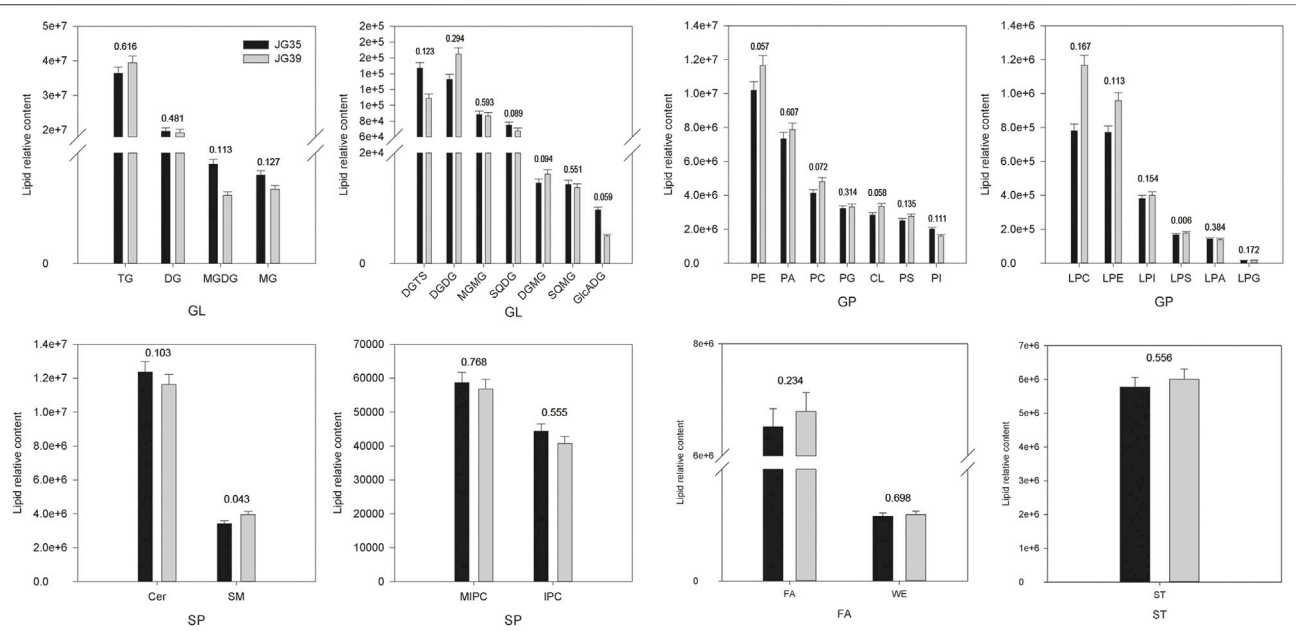


FIGURE 5 | Comparison of lipid subgroup contents between JG 35 and JG39 (The values in the figure represent *p* values).

Differential Analysis of Lipid Molecules

The number of lipid molecules in JG35 and JG39 were compared to determine their lipid differences. A total of 72 significantly differential lipid molecules ($|\log_2 FC| > 1$ and adjusted *p*-value < 0.05) were selected from the 2,633 lipid molecules identified through qualification (**Supplementary Table S1**). Particularly, the numbers of differential lipids were 11 for FA, 9 for Cer, 6 for CL, 6 for ST, 4 for DG, 4 for PA, 4 for TG, 3 for PC, 3 for PE, 3 for

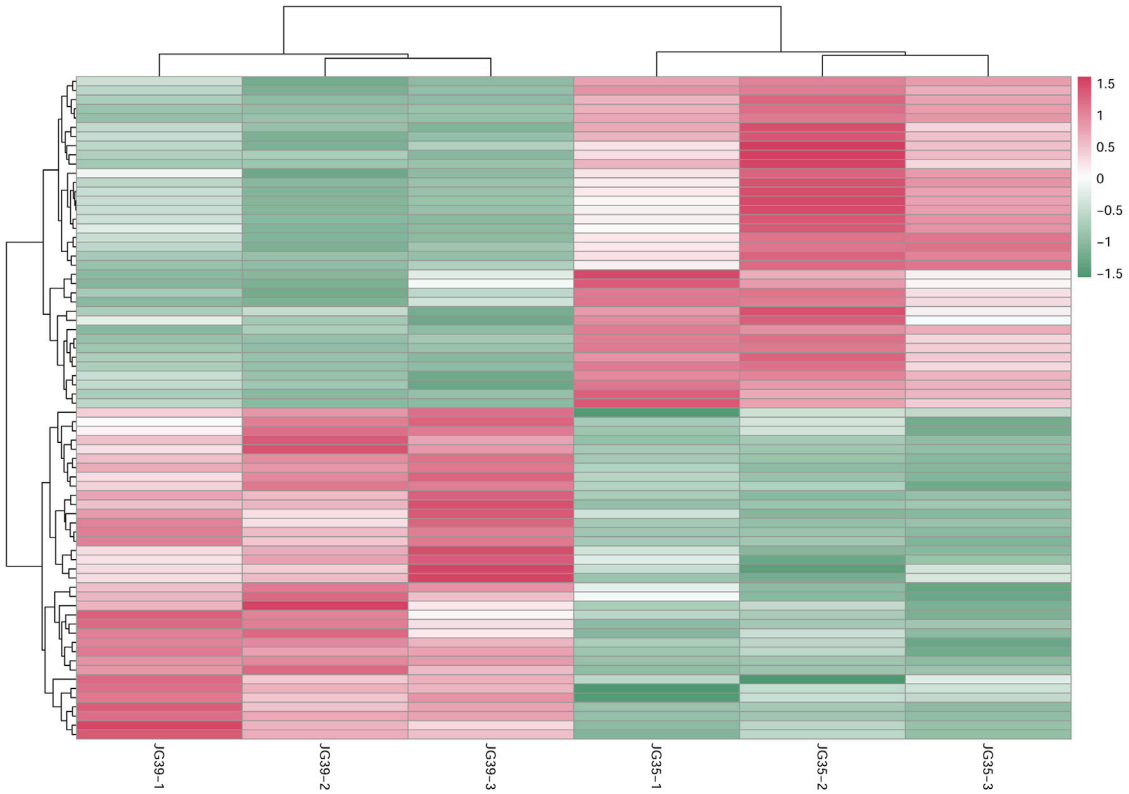


FIGURE 6 | Cluster diagram of differential lipid metabolism between JG35 and JG39.

PS, 2 for PI, 2 for IPC, 2 for SM, 2 for LPE, 2 for MG, 1 for SQDG, 1 for LPC, 1 for LPG, 1 for MIPC, 1 for LPS, 1 for LPA, 1 for PG, 1 for WE, and 1 for MGDG. The 11 FAs with the most difference were FA (22:5), FA (21:2), FA (30:2), FA (13:8), FA (19:0), FA (24:0), FA (6:2), FA (21:1), FA (14:4), FA (18:0) and FA (20:4), which most of them were unsaturated FA.

Through comparative analysis, it was found that GP had the most differential lipids, including 3 kinds of PS, 2 kinds of PI, 1 kind of PG, 3 kinds of PE, 3 kinds of PC, 4 kinds of PA, 1 kind of LPS, 1 kind of LPG, 2 kinds of LPE, 1 kind of LPC, 1 kind of LPGA and 6 kinds of CL. JG35 had higher contents of PS 35:1, 35:4, PI 21:1, PE DO-34:4, PE 22:2; O2, PC O-31:0, PC 23:2; O2, PA 36:4, PA O-28:0, LPE O-27:1; O, LPC 18:1, CL 57:3, CL 72:7, CL 76:1, CL 76:5; and JG39 had higher contains of PS 39:0, PI 35:1, PE O-34:4, PC 40:1, PA 34:1, PA 38:2, LPS O-28:1; O, LPG 17:0, LPE O-21:0; O, LPA 18:0, CL 68:4, CL 78:9. MG O-18:1, O-16:0; O, DG 40:6, DG 32:1, TG 45:0, SQDG 32:1, MGDG 34:4; O were higher in JG35, and DG 31:1, dO-34:4, TG 59:9, 64:8, 62:16 were higher in JG39. JG35 had higher contents of FA 22:5, 21:2, 19:0, 24:0, 21:1, 14:4, WE 11:2, JG39 had higher contents of FA 30:2, 13:8, 6:2, 18:0, 20:4. For the subclasses of SP, JG35 had higher contents of Cer 41:1; O3, 42:1; O4, 34:2; O3, 38:0; O4, SM 42:0; O2, IPC 36:0; O3, JG39 had higher contents of Cer 34:1; O2, 33:1; O2, 32:3; O3, 39:1; O3, 36:1; O2, SM 42:2; O2, IPC 46:0; O3, MIPC 44:0; O2. For ST, ST 31:1; O, ST 27:1; O4, ST 27:1; O3, ST 27:1; O4; S were higher in JG35, and ST 27:3; O, 27:5; O5 were higher in JG39.

The expression levels of qualified, significantly differential lipids were adopted for the hierarchical clustering of samples in different groups (**Figure 6**). This was done to evaluate the significance of differential lipid screening and comprehensively and visually display the relationships between samples and the different expression modes of lipids in different samples. According to the results, the three replicates of various samples clustered together. These differential lipid metabolites all presented similar accumulation modes in the repetitions of various samples, suggesting that they had similar reaction steps in the metabolic process. Moreover, these lipids showed obvious metabolite differences in various samples, demonstrating the significance and representativeness of the screened lipid metabolites.

Analysis of Differentially Expressed Genes

By comparing the gene expression of JG35 and JG39, 1165 DEGs were found based on $|\log_2 FC| > 1$ and FDR<0.01. Among them, 383 genes were up-regulated in JG35 and 782 genes were up-regulated in JG39 (**Supplementary Table S2**). Through GO enrichment, 736 genes of these differential genes were enriched into 48 different functional classifications (**Supplementary Figure S1**). In the biological processes, these differential genes are mainly concentrated in metabolic process, cellular process, single-organism process, etc. In the cell components, these genes are mainly concentrated in cells, cell

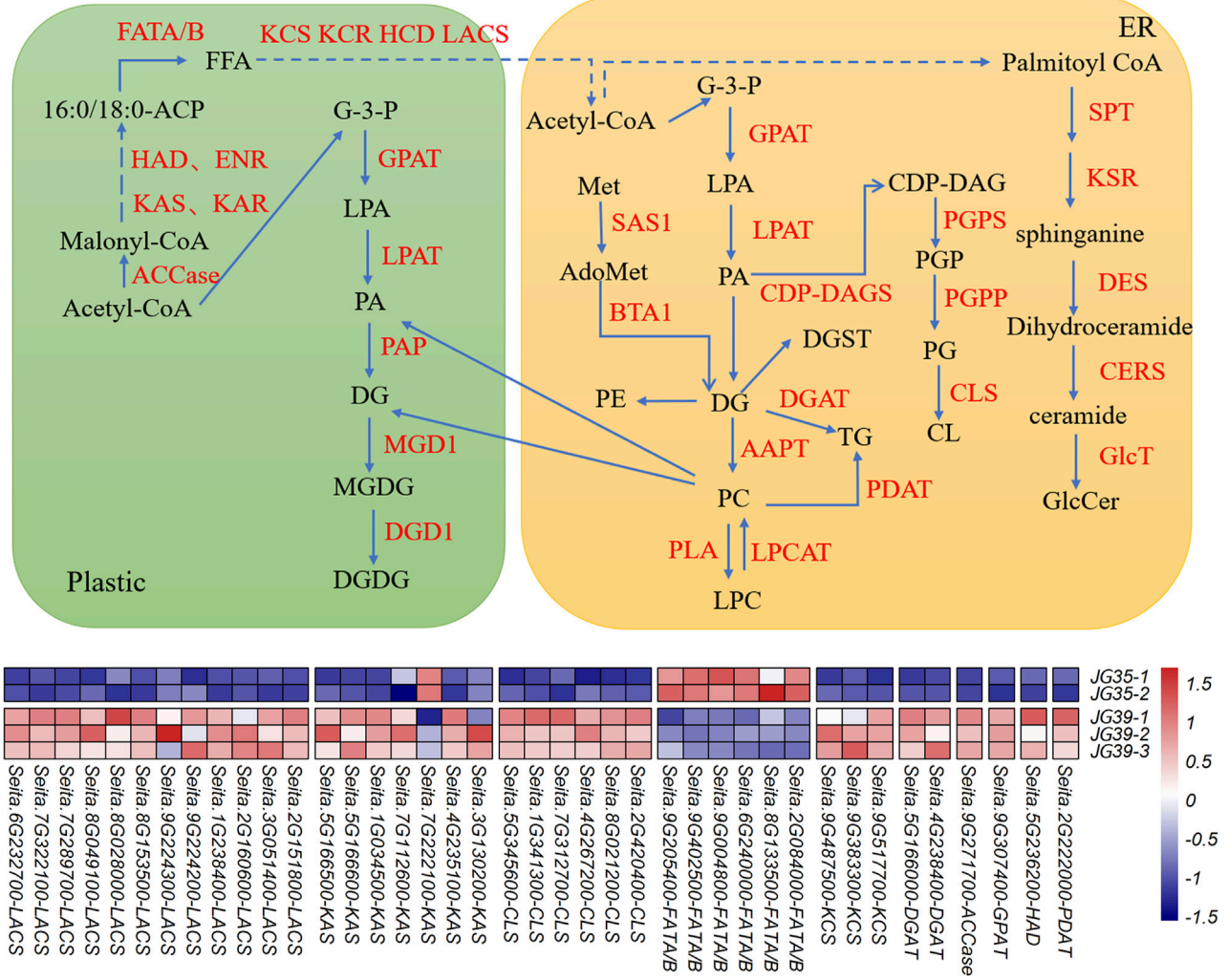


FIGURE 7 | Lipid metabolism pathway map of foxtail millet. Note: AAPT, aminoalcoholphosphotransferase; ACCase, acetyl-CoA carboxylase; Acetyl CoA, acetyl-coenzyme A; AdoMet, S-adenosylmethionine; BTA1, betaine lipid synthase; CDP-DAG, CDP-dia-cylglycerol; CDP-DAGS, CDP-DAG synthase; CERS, ceramide synthase; CL, cardiolipin; CLS, cardiolipin synthase; DES, dihydroceramide desaturase; DGD1, digalactosyldiacylglycerol synthase 1; DGAT, diacylglycerol acyltransferase; ECR, trans-2,3-enoyl-CoA reductase; ENR, enoyl-ACP-reductase; FATA/B, acyl-ACP thioesterase A/B; FFA, free fatty acids; GlcT: glycosyltransferases; G-3-P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; HAD, hydroxyacyl-ACP dehydratase; HCD, 3-hydroxacyl-CoA dehydratase; KAR, β-ketoacyl-ACP reductase; KAS, ketoacyl-ACP synthase; KCR, 3-ketoacyl-CoA reductase; KCS, 3-ketoacyl-CoA synthase; KSR, 3-ketosphinganine reductase; LACS, long-chain acyl-CoA synthetase; LPAT, lysophosphatidic acid acyltransferase; LPCAT, lysophosphatidyl choline acyltransferase; MeT, methionine; MGD1, monogalactosyldiacylglycerol synthase 1; PAP, phosphatidic acid phosphatase; PLA, phospholipase A; PDAT, phospholipid:diacylglycerol acyltransferase; PG, phosphatidylglycerol; PGP, phosphatidylglycerol phosphate; PGPP, PGP phosphatase; PGPS, phosphatidylglycerophosphate synthase; SAS1, S-adenosylmethionine synthetase; SPT, serine palmitoyltransferase.

part, organelle, membranes, membrane part, etc. In molecular function classification, these differential genes are mainly concentrated in binding, catalytic activity, transporter activity and structural molecule activity. In order to further understand the functional roles of these DEGs in the pathway, KEGG pathway enrichment analysis was conducted, and the results showed that DEGs were significantly enriched in Ribosome and Oxidative phosphorylation pathways. Furthermore, some of the DEGs were enriched in fatty acid elongation, steroid biosynthesis, fatty acid degradation and linoleic acid metabolism (**Supplementary Figure S2**), which provided a molecular basis for the further study on lipid-related genes.

Combined Analysis of Transcriptomic and Lipidomic

The transcriptome and lipidome data of JG35 and JG39 were integrated. Genes related to lipid metabolism in the transcriptome were screened out according to the functional annotation of DEGs (**Supplementary Table S3**), and the major lipid molecules in the lipid group were combined for

association analysis. Meanwhile, the molecular regulatory network of the main lipid metabolism in foxtail millet was constructed (**Figure 7**). However, according to transcriptomic analysis, JG35-1 did not show any significant difference in terms of correlation with other samples within the group or samples in the other group (**Supplementary Figure S3**). Thus, the sample was not selected for preparing the heat map of differential genes.

PA and DG are both important intermediate products in the metabolic process of lipids. Their *de novo* synthesis starts with G-3-P (3-phospho-glycerol) and FAs as initial substrates. Denovo synthesis occurs in both endoplasmic reticulum (ER) and plasmid pathways. Acetyl coenzyme A serves as the primer in the synthesis of FAs. malonyl-ACP provides a dicarbon unit for each step of the extension reaction. Malonyl thioester first engages in condensation reactions with acetyl-CoA and then goes through acceptor reactions with acyl-ACP. These reactions are catalyzed by KAS condensing enzymes, resulting in the formation of carbon-carbon bonds. It takes three different KAS condensing enzymes to generate one 18C fatty acid, containing KAS I, KAS II, and KAS III. In the condensation reaction process, the extension of the fatty acyl group chain further requires the participation of KAR, ENR, and HAD. The long-chain acyl groups are hydrolyzed by FATB, generating 18:0-ACP. 18:0-ACP is further hydroxylated under the action of FATA/B. Finally, FAs are activated into acyl coenzyme A *via* LACS, which is exported into the ER.

The UHPLC-QTOF-MS analysis showed that among the differential lipids, JG35 has a higher FA content than JG39, and the expression levels of synthesis-related genes *FATA/B* were also higher in JG35. For DG, they were exhibited a relatively high content in JG39, and the expression level of *ACCase*, *KAS*, *KCS*, *LACS* and *GPAT* corresponding to DG synthesis were also high in the transcriptomes. Similarly, JG39 had higher TG content, and the corresponding synthase genes *DGAT* and *PDAT* expression were upregulated. Besides, CL was higher in JG39 and the expression of *CLS* was also higher (**Supplementary Figure S2**). Genes related to differential lipid metabolite molecules with relatively low contents were not detected in the transcriptomes, possibly because of insufficient detection depth. However, further studies are needed to determine the precise causes.

DISCUSSIONS

Lipidomic Comparison of Two Different Foxtail Millet

In this study, the lipidomics of two different foxtail millet varieties, JG35 and JG39, were explored using the HPLC-QTOF-MS method. As a result, 31 lipid subclass and 2,633 lipid molecule were identified. By comparison, it was found that JD35 had higher DG and Cer, while TG, PE, PA, ST, FAs and CL were found to be relatively higher in JD39. DG is a structured lipid formed by glycerate after a hydroxyl replaces one fatty acid. It is deemed as a safe and healthy edible lipid, foxtail millet varieties with a high DG content have essential functions in reducing visceral fat and blood fat and inhibiting weight increase (Wan et al., 2020). As a decomposition product of sphingomyelin in biofilm bilayers, Cer is a universally recognized second messenger that plays an extensive and vital role in the growth, proliferation, differentiation, apoptosis, and damage of cells (Shikata et al., 2003). For TG, its level is closely related to the risk of cardiovascular disease with arteriosclerosis, the effective control of TG is of great significance to reduce the risk of cardiovascular system and reduce the incidence, death and disability of cardiovascular diseases (Spitler and Davies, 2020; Raposeiras-Roubin et al., 2021). GP, as a kind of phospholipid with a large content in the body, is a component of biofilm and one of the active substances on the surface of cell membrane. Meanwhile, GP participates in the recognition and signal transduction of protein by cell membrane (Chung et al., 1995; Grant and Guest, 2012). FA, the raw material for other compounds, can esterify cholesterol, lower blood cholesterol and triglycerides, improve brain cell activity and improve memory and thinking (Morton et al., 2020; Ulug and Nergiz-Unal, 2021). STs are widely found in roots, stems, leaves, fruits and seeds of plants, which plays a certain role in inhibiting tumors, promoting metabolism and regulate hormone levels (Marttinen et al., 2012; Sadek et al., 2017; Shimada et al., 2021).

Combined Analysis of Lipidome and Transcriptome

Many studies on the combination of lipid metabolism and transcriptomics have been reported, Zhang et al. (2021) combined bicolor epidermal lipidomics and transcriptomics to reveal the function of cuticular wax and cutin in maintaining drought resistance in sorghum. Also, Liang et al. (2019) applied transcriptomics and lipidomics to analyze the genes related to neutral lipid accumulation in *Nannochloropsis* under nitrogen deficiency/sufficient conditions. In this study, the composition and metabolism of lipids were comprehensively analyzed in two different foxtail millet varieties, it was found that JG35 had higher content of FA than JG39, and the expression of *FATA/B*, a key gene for the synthesis of FA (Aznar-Moreno et al., 2016; Ma et al., 2018; Correa-Aguado et al., 2021), was also increased. However, JG39 had higher levels of DG, as the key or rate-limiting enzyme in the *de novo* synthesis, the expression of *ACCase* gene was higher in JG39. And the expression levels of *KAS*, *HAD*, *KCS*, and *LACS* were also upregulated in the synthetic process from plastic to ER for acetyl-CoA in JG39. Meanwhile, the expression of *GPAT*, a related synthesis gene was also increased. As the key genes for the biosynthesis of TG (Woodfield et al., 2018), *DGAT* and *PDAT* were upregulated in JG39, and the content of TG was higher. PA can form CDP-dia-cylglycerol (CDP-DAG) by the catalysis of CDP-DAG synthase (CDP-DAGS) and then produce CL by the catalysis of cardiolipin synthase (CLS) (Zhang et al., 2020). Here, JG39 had higher CL and *CLS* was upregulated. Feng et al. (2016) identified 10 *KAS* genes involved in fatty acid synthesis in eucommia ulmoides seeds by transcriptome sequencing. Bao et al. (2021) found that fatty acid biosynthesis related genes *LACS6, DGD1*, *ACAT1*, *AGPAT*, *WSD1*, *EGY2*, *and oleosin* were highly expressed during late development of *Prunus Pedunculata* Pall seeds. Xiu et al. (2018) found key genes in oil

biosynthesis of *Paenpnia*, containing *ACCase*, *LPCAT*, *FADs*, *DGAT*. These reveals the possible reason for the high contents of differential lipids for DG, TG and CL.

China has abundant foxtail millet variety resources; however, there is limited research on the effective utilization and mining of different resources. Also, many excellent resources remain unexplored, limiting the breeding of special foxtail millet varieties. The study of investigating the composition and change in foxtail millet by lipidomics is of great significance for the development of functional foods, the utilization of germplasm resources and the breeding of special varieties of foxtail millet.

CONCLUSION

In this paper, UHPLC-Q-TOF-MS system was firstly used to compare the lipid content and composition of two different foxtail millet varieties containing JG35 and JG39, and combined with transcriptome data to construct the molecular regulatory network of major lipid metabolism preliminatively in foxtail millet. A total of 2,633 kinds of lipid molecules and 31 kinds of lipid subclasses were identified by lipid detection. By comparing, the lipid composition was not significantly different. However, there were differences in the contents of each subcategory, JG35 had higher contents of DG and Cer, and the contents of TG, PE, PA, ST, FA, and CL were higher in JG39. By the comparison of differential lipid, it was found that the key genes *FATA/B* for the synthesis of FAs was highly expressed in JG35, and the related genes (*ACCase*, *KAS*, *HAD*, *KCS*, *LACS*, and *GAPT*) for the synthesis of DG were highly expressed in JG39. Besides, in JG39 the expression of *DGAT* and *PDAT* for the synthesis of TG was also higher, and *CLS* related to CL synthesis was upregulated. The two varieties contain different contents of lipids which are beneficial to human health, so the development of functional foods and the breeding of new varieties can be carried out selectively according to different needs.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HZ performed the experiments, analyzed the metabolicdata, and drafted the manuscript. JW and JZ contributed statistical analysis and modify the draft. JW performed the experiments, prepared figures and tables. QW contributed materials and carried out the association analysis. FQ and XY contributed revising the language of this manuscript. CS and ZF designed the experiment.

FUNDING

This study was supported by Doctor Foundation of Crop Science Institute of Shanxi Academy of Agricultural Sciences (No. ZB 2002), Doctoral Research project of Shanxi Agricultural University (No. 2021BQ40) and (No. 2021BQ60), and effects of different fertilizers on nitrogen accumulation and grain quality of foxtail millet (No. 201903D221019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.758003/full#supplementary-material

REFERENCES

Aznar-Moreno, J. A., Venegas-Calerón, M., Martínez-Force, E., Garcés, R., and Salas, J. J. (2016). Acyl Carrier Proteins from sunflower (*Helianthus Annuus* L.) Seeds and Their Influence on FatA and FatB Acyl-ACP Thioesterase Activities. *Planta* 244 (2), 479–490. doi:10.1007/s00425-016-2521-7

Bao, W., Ao, D., Wang, L., Ling, Z., Chen, M., Bai, Y., et al. (2021). Dynamic Transcriptome Analysis Identifies Genes Related to Fatty Acid Biosynthesis in the Seeds of *Prunus Pedunculata* Pall. *BMC Plant Biol.* 21 (1), 152. doi:10.1186/s12870-021-02921-x

Checa, A., Bedia, C., and Jaumot, J. (2015). Lipidomic Data Analysis: Tutorial, Practical Guidelines and Applications. *Analytica Chim. Acta* 885, 1–16. doi:10.1016/j.aca.2015.02.068

Chung, S. Y., Moriyama, T., Uezu, E., Uezu, K., Hirata, R., Yohena, N., et al. (1995). Administration of Phosphatidylcholine Increases Brain Acetylcholine Concentration and Improves Memory in Mice with Dementia. *J. Nutr.* 125 (6), 1484–1489. doi:10.1093/jn/125.6.1484

Correa-Aguado, H. C., Cerrillo-Rojas, G. V., Rocha-Uribe, A., Soria-Guerra, R. E., and Morales-Domínguez, J. F. (2021). Benzyl Amino Purine and Gibberellic Acid Coupled to Nitrogen-Limited Stress Induce Fatty Acids, Biomass Accumulation, and Gene Expression in Scenedesmus Obliquus. *Phyton-International J. Exp. Bot.* 90 (2), 515–531. doi:10.32604/phyton.2021.013619

Dunn, W. B., Broadhurst, D., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., et al. (2011). Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* 6 (7), 1060–1083. doi:10.1038/nprot.2011.335

Ecker, J., Benedetti, E., Kindt, A. S. D., Höring, M., Perl, M., Machmüller, A. C., et al. (2021). The Colorectal Cancer Lipidome: Identification of a Robust Tumor-specific Lipid Species Signature. *Gastroenterology* 161 (3), 910–923. e19. doi:10.1053/j.gastro.2021.05.009

Eggers, L. F., Müller, J., Marella, C., Scholz, V., Watz, H., Kugler, C., et al. (2017). Lipidomes of Lung Cancer and Tumour-Free Lung Tissues Reveal Distinct Molecular Signatures for Cancer Differentiation, Age, Inflammation, and Pulmonary Emphysema. *Sci. Rep.* 7 (1), 11087. doi:10.1038/s41598-017-11339-1

Fei, X., Ma, Y., Hu, H., and Wei, A. (2020). Transcriptome Analysis and GC-MS Profiling of Key Genes in Fatty Acid Synthesis of *Zanthoxylum Bungeanum* Seeds. *Ind. Crops Prod.* 156, 112870. doi:10.1016/j.indcrop.2020.112870

Feng, Y.-Z., Wang, L., Fu, J.-M., Wuyun, T.-N., Du, H.-Y., Tan, X.-F., et al. (2016). Transcriptome Sequencing Discovers Genes Related to Fatty Acid Biosynthesis in the Seeds of *Eucommia Ulmoides*. *Genes Genom* 38, 275–283. doi:10.1007/s13258-015-0362-6

Fitzsimons, S., McCarthy, C., Bruen, R., and Belton, O. (2019). Microrna-155 and Conjugated Linoleic Acid in the Context of Atherosclerosis. *Atherosclerosis* 287, e24. doi:10.1016/j.atherosclerosis.2019.06.069

Grant, R., and Guest, J. (2012). Effects of Dietary Derived Antioxidants on the Central Nervous System. *Int. J. Nutr. Pharmacol. Neurol. Dis.* 2 (3), 185. doi:10.4103/2231-0738.99470

Hou, Q., Ufer, G., and Bartels, D. (2016). Lipid Signalling in Plant Responses to Abiotic Stress. *Plant Cel Environ.* 39 (5), 1029–1048. doi:10.1111/pce.12666

Joo, Y. H., Nam, M. H., Chung, N., and Lee, Y. K. (2020). UPLC-QTOF-MS/MS Screening and Identification of Bioactive Compounds in Fresh, Aged, and Browned *Magnolia Denudata* Flower Extracts. *Food Res. Int.* 133, 109192. doi:10.1016/j.foodres.2020.109192

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., and Neumann, S. (2012). CAMERA: an Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/mass Spectrometry Data Sets. *Anal. Chem.* 84 (1), 283–289. doi:10.1021/ac202450g

Kus, K., Kij, A., Zakrzewska, A., Jasztal, A., Stojak, M., Walczak, M., et al. (2018). Alterations in Arginine and Energy Metabolism, Structural and Signalling Lipids in Metastatic Breast Cancer in Mice Detected in Plasma by Targeted Metabolomics and Lipidomics. *Breast Cancer Res.* 20 (1), 148. doi:10.1186/s13058-018-1075-y

Lee, W. J., Zhang, Z., Lai, O. M., Tan, C. P., and Wang, Y. (2020). Diacylglycerol in Food Industry: Synthesis Methods, Functionalities, Health Benefits, Potential Risks and Drawbacks. *Trends Food Sci. Technol.* 97, 114–125. doi:10.1016/j.tifs.2019.12.032

Li, C., Wang, G., Li, H., Wang, G., Ma, J., Zhao, X., et al. (2021). High-depth Resequencing of 312 Accessions Reveals the Local Adaptation of Foxtail Millet. *Theor. Appl. Genet.* 134 (5), 1303–1317. doi:10.1007/s00122-020-03760-4

Liang, J., Wen, F., and Liu, J. (2019). Transcriptomic and Lipidomic Analysis of an EPA-Containing *Nannochloropsis* Sp. PJ12 in Response to Nitrogen Deprivation. *Sci. Rep.* 9, 4540. doi:10.1038/s41598-019-41169-2

Ma, R., Thomas-Hall, S. R., Chua, E. T., Alsenani, F., Eltanahy, E., Netzel, M. E., et al. (2018). Gene Expression Profiling of Astaxanthin and Fatty Acid Pathways in *Haematococcus Pluvialis* in Response to Different LED Lighting Conditions. *Bioresour. Technol.* 250, 591–602. doi:10.1016/j.biortech.2017.11.094

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 (19), 3787–3793. doi:10.1093/bioinformatics/bti430

Marasca, C., Arana, M. E. B., Protti, M., Cavalli, A., Mercolini, L., and Armirotti, A. (2021). Volumetric Absorptive Microsampling of Blood for Untargeted Lipidomics. *Molecules* 26 (2), 262. doi:10.3390/molecules26020262

Marttinen, M., Pajari, A.-M., Päivärinta, E., Storvik, M., Niku, M., Nurmi, T., et al. (2012). Plant Sterols Induce Intestinal Tumor Formation in Gender-Related Manner in ApcMin Mice. *BMC Proc.* 6 (3), 1–2. doi:10.1186/1753-6561-6-S3-P34

Mccloy, U., Ryan, M. A., Pencharz, P. B., Ross, R. J., and Cunnane, S. C. (2004). A Comparison of the Metabolism of Eighteen-Carbon 13C-Unsaturated Fatty Acids in Healthy Women. *J. Lipid Res.* 45 (3), 474–485. doi:10.1194/jlr.M300304-JLR200

Meherunnahar, M., Hoque, M., Miah, M., Abedin, N., and Islam, M. (2018). Evaluation of Nutritional and Functional Status of Three Varieties of Foxtail Millet Cultivars in Bangladesh. *Bangla. J. Bot.* 47 (2), 301–306.

Meng, J.-S., Tang, Y.-H., Sun, J., Zhao, D.-Q., Zhang, K.-L., and Tao, J. (2021). Identification of Genes Associated with the Biosynthesis of Unsaturated Fatty Acid and Oil Accumulation in Herbaceous Peony 'Hangshao' (Paeonia Lactiflora 'Hangshao') Seeds Based on Transcriptome Analysis. *BMC Genomics* 22, 94. doi:10.1186/s12864-020-07339-7

Morton, S. U., Vyas, R., Gagoski, B., Vu, C., Litt, J., Larsen, R. J., et al. (2020). Maternal Dietary Intake of omega-3 Fatty Acids Correlates Positively with Regional Brain Volumes in 1-Month-Old Term Infants. *Cereb. Cortex* 30 (4), 2057–2069. doi:10.1093/cercor/bhz222

Pan, J., Li, Z., Wang, Q., Garrell, A. K., Liu, M., Guan, Y., et al. (2018). Comparative Proteomic Investigation of Drought Responses in Foxtail Millet. *BMC Plant Biol.* 18, 315. doi:10.1186/s12870-018-1533-9

Patel, S., Pachhigar, K., Ganvit, R., Panchal, R. R., Ponnuchamy, M., Kumar, J., et al. (2021). Exploring Flowering Genes in Isabgol (*Plantago Ovata* Forsk.) through Transcriptome Analysis. *Plant Mol. Biol. Rep.* 39, 192–211. doi:10.1007/s11105-020-01237-8

Qin, L., Chen, E., Li, F., Yu, X., Liu, Z., Yang, Y., et al. (2020). Genome-Wide Gene Expression Profiles Analysis Reveal Novel Insights into Drought Stress in Foxtail Millet (Setaria Italica L.). *Int. J. Mol. Sci* 21 (22), 8520–8521. doi:10.3390/ijms21228520

Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated Omics Analysis of Specialized Metabolism in Medicinal Plants. *Plant J.* 90 (4), 764–787. doi:10.1111/tpj.13485

Raposeiras-Roubin, S., Rosselló, X., Oliva, B., Fernández-Friera, L., Mendiguren, J. M., Andrés, V., et al. (2021). Triglycerides and Residual Atherosclerotic Risk. *J. Am. Coll. Cardiol.* 77 (24), 3031–3041. doi:10.1016/j.jacc.2021.04.059

Rodríguez-López, C. E., Hernández-Brenes, C., Treviño, V., and Díaz de la Garza, R. I. (2017). Avocado Fruit Maturation and Ripening: Dynamics of Aliphatic Acetogenins and Lipidomic Profiles from Mesocarp, Idioblasts and Seed. *BMC Plant Biol.* 17, 159. doi:10.1186/s12870-017-1103-6

Sadek, N. F., Yuliana, N. D., Prangdimurti, E., Priosoeryanto, B. P., and Budijanto, S. (2017). Plant Sterol Esters in Extruded Food Model Inhibits Colon Carcinogenesis by Suppressing Inflammation and Stimulating Apoptosis. *J. Med. Food* 20, 659–666. doi:10.1089/jmf.2016.3876

Saini, S., and Rai, A. K. (2021). Linoleic Acid-A Feasible Preventive Approach for Visceral Leishmaniasis. *Front. Nutr.* 8, 649025. doi:10.3389/fnut.2021.649025

Samson, F. P., Patrick, A. T., Fabunmi, T. E., Yahaya, M. F., Madu, J., He, W., et al. (2020). Oleic Acid, Cholesterol, and Linoleic Acid as Angiogenesis Initiators. *ACS Omega* 5 (31), 20575–20585. doi:10.1021/acsomega.0c02850

Shen, Q., Fu, L., Dai, F., Jiang, L., Zhang, G., and Wu, D. (2016). Multi-Omics Analysis Reveals Molecular Mechanisms of Shoot Adaption to Salt Stress in Tibetan Wild Barley. *BMC Genomics* 17 (1), 889. doi:10.1186/s12864-016-3242-9

Shikata, K., Niiro, H., Azuma, H., Tachibana, T., and Ogino, K. (2003). Synthesis of Non-Natural C2-*Homo*-Ceramide and its Apoptotic Activity against HL-60 Cells. *Bioorg. Med. Chem. Lett.* 13 (4), 613–616. doi:10.1016/S0960-894X(02)01026-0

Shimada, T. L., Ueda, T., and Hara-Nishimura, I. (2021). Excess Sterol Accumulation Affects Seed Morphology and Physiology in *Arabidopsis Thaliana*. *Plant Signaling Behav.* 16 (4), 1872217. doi:10.1080/15592324.2021.1872217

Singh, A., Singh, P., Kumar, B., Kumar, S., Dev, K., and Maurya, R. (2019). Detection of Flavonoids from *Spinacia Oleracea* Leaves Using HPLC-ESI-QTOF-MS/MS and UPLC-QqQLIT-MS/MS Techniques. *Nat. Prod. Res.* 33 (15), 2253–2256. doi:10.1080/14786419.2018.1489395

Spicher, L., Glauser, G., and Kessler, F. (2016). Lipid Antioxidant and Galactolipid Remodeling under Temperature Stress in Tomato Plants. *Front. Plant Sci.* 7, 167. doi:10.3389/fpls.2016.00167

Spitler, K. M., and Davies, B. S. J. (2020). Aging and Plasma Triglyceride Metabolism. *J. Lipid Res.* 61 (8), 1161–1167. doi:10.1194/jlr.R120000922

Sun, S., Wang, D., Li, J., Lei, Y., Li, G., Cai, W., et al. (2021). Transcriptome Analysis Reveals Photoperiod-Associated Genes Expressed in rice Anthers. *Front. Plant Sci.* 12, 621561. doi:10.3389/fpls.2021.621561

Trivedi, A. K., Arya, L., Verma, S. K., Tyagi, R. K., Hemantaranjan, A., Verma, M., et al. (2018). Molecular Profiling of Foxtail Millet (Setaria Italica (L.) P. Beauv) from Central Himalayan Region for Genetic Variability and Nutritional Quality. *J. Agric. Sci.* 156 (3), 333–341. doi:10.1017/s0021859618000382

Ulug, E., and Nergiz-Unal, R. (2021). Dietary Fatty Acids and CD36-Mediated Cholesterol Homeostasis: Potential Mechanisms. *Nutr. Res. Rev.* 34 (1), 64–77. doi:10.1017/S0954422420000128

Verma, S., Srivastava, S., and Tiwari, N. (2015). Comparative Study on Nutritional and Sensory Quality of Barnyard and Foxtail Millet Food Products with Traditional rice Products. *J. Food Sci. Technol.* 52 (8), 5147–5155. doi:10.1007/s13197-014-1617-y

Wang, X., Zhang, J., Liu, Z., Wang, S., Huang, B., Hu, Z., et al. (2021). Comparative Transcriptome Analysis of Three Chrysanthemums Provides Insights into Flavonoid and Terpenoid Biosynthesis. *J. Plant Biol.* 64, 389–401. doi:10.1007/s12374-021-09304-7

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., et al. (2010). Global Metabolic Profiling Procedures for Urine Using UPLC-MS. *Nat. Protoc.* 5 (6), 1005–1018. doi:10.1038/nprot.2010.50

Wenk, M. R. (2005). Erratum: The Emerging Field of Lipidomics. *Nat. Rev. Drug Discov.* 4 (9), 725. doi:10.1038/nrd1847

Woodfield, H. K., Cazenave-Gassiot, A., Haslam, R. P., Guschina, I. A., Wenk, M. R., and Harwood, J. L. (2018). Using Lipidomics to Reveal Details of Lipid Accumulation in Developing Seeds from Oilseed Rape (*Brassica Napus* L.).

Biochim. Biophys. Acta (Bba) - Mol. Cel Biol. Lipids 1863 (3), 339–348. doi:10.1016/j.bbalip.2017.12.010

Wu, D., Shu, T., Yang, X., Song, J.-X., Zhang, M., Yao, C., et al. (2020). Plasma Metabolomic and Lipidomic Alterations Associated with COVID-19. *Natl. Sci. Rev.* 7 (07), 1157–1168. doi:10.1101/2020.04.05.20053819

Wu, D., Wang, H., Tan, J., Wang, C., Lin, H., Zhu, H., et al. (2018). Pharmacokinetic and Metabolism Studies of Curculigoside C by UPLC-MS/MS and UPLC-QTOF-MS. *Molecules* 24 (1), 21. doi:10.3390/molecules24010021

Xiu, Y., Wu, G., Tang, W., Peng, Z., Bu, X., Chao, L., et al. (2018). Oil Biosynthesis and Transcriptome Profiles in Developing Endosperm and Oil Characteristic Analyses in *Paeonia Ostii* Var. *Lishizhenii*. *J. Plant Physiol.* 228, 121–133. doi:10.1016/j.jplph.2018.05.011

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 (2), R14. doi:10.1186/gb-2010-11-2-r14

Yuan, Y., Liu, C., Zhao, G., Gong, X., Dang, K., Yang, Q., et al. (2021). Transcriptome Analysis Reveals the Mechanism Associated with Dynamic Changes in Fatty Acid and Phytosterol Content in Foxtail Millet (*Setaria Italica*) during Seed Development. *Food Res. Int.* 145, 110429. doi:10.1016/j.foodres.2021.110429

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, H., Jiang, C., Ren, J., Dong, J., Shi, X., Zhao, X., et al. (2020). An Advanced Lipid Metabolism System Revealed by Transcriptomic and Lipidomic Analyses Plays a central Role in Peanut Cold Tolerance. *Front. Plant Sci.* 11, 1110. doi:10.3389/fpls.2020.01110

Zhang, X., Ni, Y., Xu, D., Busta, L., Xiao, Y., Jetter, R., et al. (2021). Integrative Analysis of the Cuticular Lipidome and Transcriptome of Sorghum Bicolor Reveals Cultivar Differences in Drought Tolerance. *Plant Physiol. Biochem.* 163, 285–295. doi:10.1016/j.plaphy.2021.04.007

Zhao, Z., Tang, S., Zhang, Y., Yue, J., Xu, J., Tang, W., et al. (2021). Evolutionary Analysis and Functional Characterization of SiBRI1 as a Brassinosteroid Receptor Gene in Foxtail Millet. *BMC Plant Biol.* 21, 291. doi:10.1186/s12870-021-03081-8

Zhi, Y., Taylor, M. C., Campbell, P. M., Warden, A. C., Shrestha, P., El Tahchy, A., et al. (2017). Comparative Lipidomics and Proteomics of Lipid Droplets in the Mesocarp and Seed Tissues of Chinese Tallow (*Triadica Sebifera*). *Front. Plant Sci.* 8, 1339. doi:10.3389/fpls.2017.01339

Zhou, R., Sanz-Jimenez, P., Zhu, X.-T., Feng, J.-W., Shao, L., Song, J.-M., et al. (2021). Analysis of Rice Transcriptome Reveals the LncRNA/CircRNA Regulation in Tissue Development. *Rice* 14, 14. doi:10.1186/s12284-021-00455-2

Conflict of Interest: Author QW was employed by the company Hebei Zhihai Technology Co., Ltd.

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

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

Copyright © 2021 Zhang, Wang, Zhao, Sun, Wang, Wang, Qu, Yun and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

Cer ceramides
CL cardiolipin
DG diacylglycerols
DGDG digalactosyl-diacylglycerol
DGMG digalactosylmonoacylglycerol
DGTS diacylglyceryltrimethylhomo-Ser
FA fatty acyls
FC fold change
GlcADG glucuronosyldiacylglycerol
IPC inositolphosphoryl-ceramides
LPA lysophosphatidic-acid
LPC lyso-posphatidyl choline
LPE lyso-posphatidyl ethanolamine
LPG lyso-phosphatidylglycerol
LPI lyso-phosphatidyl inositol
LPS lipopolysaccharide
MIPC mannosyl-inositolphosphoryl-ceramides
MG monoacylglycerols
MGDG monogalactosyl-diacylglycerol
MGMG monogalactosylmonoacylglycerol
PA phosphatidic acid
PC phosphatidyl choline
PE phosphatidyl ethanolamine
PG phosphatidylglycerol
PI phosphatidyl inositol
PIP phosphatidylinositol diphosphate
PS phosphatidyl serine
SM sphingomyelin
SQDG sulfoquinovosyl-diacylglycerol
SQMG sulfoquinovosylmonoacylglycerols
ST sterol
TG Triacylglycerols
WE wax esters.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fgene.2021.758209

A Bivariate Mapping Model Identifies Major Covariation QTLs for Biomass Allocation Between Leaf and Stem Growth of *Catalpa bungei*

Miaomiao Zhang[1], *Nan Lu*[1], *Tianqing Zhu*[1], *Guijuan Yang*[1], *Guanzheng Qu*[2], *Chaozhong Shi*[1], *Yue Fei*[1], *Bingyang Liu*[1], *Wenjun Ma*[1] *and Junhui Wang*[1]*

[1]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China, [2]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China

Biomass allocation plays a critical role in plant morphological formation and phenotypic plasticity, which greatly impact plant adaptability and competitiveness. While empirical studies on plant biomass allocation have focused on molecular biology and ecology approaches, detailed insight into the genetic basis of biomass allocation between leaf and stem growth is still lacking. Herein, we constructed a bivariate mapping model to identify covariation QTLs governing carbon (C) allocation between the leaves and stem as well as the covariation of traits within and between organs in a full-sib mapping population of *C. bungei*. A total of 123 *cov*QTLs were detected for 23 trait pairs, including six leaf traits (leaf length, width, area, perimeter, length/width ratio and petiole length) and five stem traits (height, diameter at breast height, wood density, stemwood volume and stemwood biomass). The candidate genes were further identified in tissue-specific gene expression data, which provided insights into the genetic architecture underlying C allocation for traits or organs. The key QTLs related to growth and biomass allocation, which included *UVH1*, *CLPT2*, *GAD/SPL*, *COG1* and *MTERF4*, were characterised and verified via gene function annotation and expression profiling. The integration of a bivariate Quantitative trait locus mapping model and gene expression profiling will enable the elucidation of genetic architecture underlying biomass allocation and covariation growth, in turn providing a theoretical basis for forest molecular marker-assisted breeding with specific C allocation strategies for adaptation to heterogeneous environments.

Keywords: bivariate mapping model, biomass allocation, covariation, quantitative trait locus, *Catalpa bungei*

OPEN ACCESS

Edited by:
Jiang Libo,
Beijing Forestry University, China

Reviewed by:
Muhammad Adeel Hassan,
Institute of Crop Sciences (CAS), China
Jingwen Gan,
Beijing Forestry University, China

***Correspondence:**
Junhui Wang
wangjh808@sina.com
wangjh@caf.ac.cn

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

Received: *13 August 2021*
Accepted: *21 October 2021*
Published: *18 November 2021*

Citation:
Zhang M, Lu N, Zhu T, Yang G, Qu G, Shi C, Fei Y, Liu B, Ma W and Wang J (2021) A Bivariate Mapping Model Identifies Major Covariation QTLs for Biomass Allocation Between Leaf and Stem Growth of Catalpa bungei. *Front. Genet. 12:758209. doi: 10.3389/fgene.2021.758209*

INTRODUCTION

Plants are organised into various organs that have distinct functional divisions and carry out continuous material exchange. Leaves are the major organs for photosynthesis, respiration and transpiration, exhibiting tremendous diversity in shape and size (Nocera, 2017). Plant stems play key roles in nutrients and water transport, while also providing physical support (Niinemets, 2007; Brienen et al., 2017). The balanced allocation of carbon (C) to various plant organs is crucial for plant growth and development, as it allows the formation of specific plant morphology as well as phenotypic plasticity, which in turn have a profound impact on plant adaptability and

competitiveness under fluctuating environmental conditions (Poorter et al., 2019). The translocation of carbohydrates from the photosynthesising "source" leaves provide substrates required for the growth of non-photosynthesising "sink" organs. The plant architecture determined by stems and branches is closely associated with foliage photosynthetic efficiency, cultivation, yield, light assimilation and harvesting. C allocation patterns also emerge among different tissues and traits within a single organ, such as taproot-lateral root, leaf thickness and surface area as well as diameter at breast height (Chen et al., 2021). For instance, expression analysis in *Arabidopsis* suggested that excessive *SMAX1* expression suppressed rosette shoot branching, while promoting leaf and petiole elongation via regulating *TCP1* expression (Zheng et al., 2021). Currently, biomass allocation represents a hot topic in the fields of plant molecular biology, evolutionary genetics and ecology (Weraduwage et al., 2015; Berny Mier et al., 2019; Lauri, 2019). However, little is known regarding the genetic basis, which governs the interaction and coordination of leaf and stem growth in plants, especially in forest trees.

Quantitative trait locus (QTL) mapping is a valuable approach for exploring specific genes underlying complex traits (Miles and Wayne, 2008). This approach has been previously utilised for the study of growth-related traits in wheat (Semagn et al., 2021), maize (Guo et al., 2021), *Arabidopsis thaliana* (Getahun et al., 2020), *Populus trichocarpa* (Chhetri et al., 2019) and *Eucalyptus* (Ammitzboll et al., 2020). A reasonable trade-off strategy for C distribution among organs or phenotypes optimises the resource acquisition, viability and adaptability of plants (Poorter et al., 2012). For woody plants, Novaes et al. (2009) identified QTLs implicated in multiple traits, which regulated C allocation between plant growth and wood components (lignin and cellulose). The identification of pleiotropic QTLs can significantly improve the accuracy of genetic mapping and advance the molecular marker-assisted breeding process. Thus, introducing the concept of C allocation into the genetic framework is of great value for the elucidation of genetic mechanisms underlying phenotypic plasticity and pleiotropy. However, research on the genetic basis of plant C allocation has been is scarce due to the lack of suitable conceptual frameworks.

Bivariate trait correlations can be quantitatively expressed through statistical models in order to reveal C distribution patterns and dissect the underlying genetic basis. The first experimental utilisation of QTL mapping for the study of biomass allocation was reported by Wu et al. (2002), who employed a statistical model to discover the genetic origin of the allometric relationship between stem height and stem biomass in F_2 populations of poplar. Utilising a dynamic allometric QTL mapping model, Li et al. (2007) further analysed the genetic mechanism of ontogenetic C allocation in soybean RIL populations. Based on the statistical and/or dynamic models, the QTLs related to biomass distribution and balanced growth between plant organs or traits were subsequently identified for the aboveground-underground growth of *Populus euphratica* (Zhang et al., 2017), leaf number and whole dry weight of *Arabidopsis* (Gan et al., 2019) as well as for the leaf area-leaf dry weight of common bean (Zhang et al., 2020b). Despite these advances, genetic analyses of C allocation have focused mainly on annual herbs and crops. In addition, most of the studies considered relatively few traits, while plant traits are multiple and complex, generally interacting with each other. It is therefore necessary to identify QTLs in order to elucidate C allocation patterns by considering the phenotypical diversity of woody plants.

Catalpa bungei C. A. Mey. is a precious timber and garden ornamental tree species that is widely distributed in the temperate, subtropical and tropical regions of China (Wang et al., 2016). This tree species has considerable genotype variation with regard to growth performance over the long-term evolution process (Wang et al., 2016; Lu et al., 2019). To identify major covariation QTLs for biomass allocation between leaf and stem growth of *C. bungei*, we constructed a bivariate mapping model to detect covariation QTLs that govern trait covariance via C allocation among and within plant organs in a full-sib mapping population of this tree species. Candidate gene analysis of significant QTLs was further conducted, utilising tissue-specific gene expression data. The discovery of covariation QTLs related to biomass allocation between leaves and stems will help elucidate the genetic mechanism of plant morphogenesis and will provide a theoretical basis for forest molecular marker-assisted breeding with specific C allocation strategies depending on the heterogeneous environments.

MATERIALS AND METHODS

Mapping Population and SNP Genotyping

A full-sib population of 200 lines was generated from the cross between *C. bungei* "7080" (female parent) and *C. duclouxii* "16-PJ-3" (male parent). The parents and progenies were propagated via bud grafting and then planted in the experimental field in Luoyang, China (N 112.55°, E 34.71°) in 2018, following a randomised block design. For all 200 lines, a genome-wide panel of single nucleotide polymorphisms (SNPs) was sequenced through restriction-site associated DNA (RAD) technology using an Illumina HiSeq X Ten platform. The genetic map was constructed using a set of 9,593 SNPs, following several strict criteria in our previous study. The integrated genetic map included 20 linkage groups and spanned 3,151.63 cM, with an average distance of 0.32 cM between adjacent markers (Lu et al., 2019). After further filtering for duplicate markers due to imputation missing values, 6,446 SNPs were retained for this final study, including 5,362 testcross and 1,084 intercross type markers, referring to the segregation ratios of 1:1 and 1:2:1. Genotyping data were submitted to the NCBI SRA database (http://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA551333.

Phenotyping

A randomised block design was applied to the F_1 population, with two ramets per clone in each plot and five replicates. Leaf traits of the third whorl of fully expanded leaves were evaluated on 2018/9/5 at the end of the rapid growth period. Six leaf traits were

assessed, namely leaf length (LL), leaf width (LW), leaf length/width ratio (L/W), leaf area (LA), leaf perimeter (LP) and petiole length (PL). Stem growth traits were observed on 2018/10/10 sat the end of the growing season and included tree height (H) and diameter at breast height (DBH). The measurement methods for these phenotypic data were previously described in detail by Lu et al. (2019).

Wood density was measured during October 2018 after harvest. Stem segments from at least three individuals of each line were excised at 125–130 cm above ground level. After removing the bark and pith with a razor blade, wood density was determined from volumetric displacement and oven-dried mass at 103 ± 3°C for at least 6 h (Cornelissen et al., 2003). The ratio of dry weight to volume is the wood density (WD, g/cm^3). Stemwood volume (V) in forest trees is an important trait for forest productivity. This trait is determined by stem height (H) and stem diameter at breast height (DBH) through a geometric function, expressed as the below equation, which is modified from Sang et al. (2019):

$$V = 0.785 \times (DBH)^2 \times H$$

The stemwood biomass (SB) can be calculated by $SB = V \times BD$.

QTL Analysis via Bivariate Mapping

QTL Mapping

The constructed linkage map is quite dense, and we employed a multiplicative model that assumes QTLs are located at the positions of markers. The sample size of the mapping population is n, and the phenotypic values of individual i are y_i. Two phenotypic variables were selected and combined within or between leaf and stem traits. The multiplicative likelihood model is expressed as:

$$L(\mathbf{\Phi}|\mathbf{y}) \prod_{i=1}^{n_1} f_1(\mathbf{y}_i) \ldots \prod_{i=1}^{n_j} f_j(\mathbf{y}_i) \quad (1)$$

where Φ is the unknown parameter, $\mathbf{y}_i = (y_{1i}, y_{2i})$ is the phenotypic vector of progeny i for trait 1 (coded by 1) and trait 2 (coded by 2), n_j is the number of progeny with SNP genotype j, and $f_j(\mathrm{y}_i)$ is a bivariate normal distribution for progeny i with the expected mean vector for genotype j (μ_{1j}, μ_{2j}) and the variance-covariance matrix Σ.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (2)$$

where σ_1^2 and σ_2^2 are the variances of two different traits, and ρ is the trait-trait correlation. Statistical methods based on likelihood (1) have been established to estimate the model parameters $\Phi = (\mu_{1j}, \mu_{2j}, \sigma_1^2, \sigma_2^2, \rho)$.

Hypothesis Tests

After scanning SNP loci on the linkage map, we can detect whether a significant *cov*QTL affects the C allocation variation of leaf traits and stem traits as well as the covariation between leaves and stems. The tests are based on the following hypotheses:

$$\begin{aligned} &\mathbf{H}_0\text{: } (\mu_{1j}, \mu_{2j}) = (\mu_1, \mu_2), \text{for j = 1 or 2} \\ &\mathbf{H}_1\text{: At least one equality above does not hold} \end{aligned} \quad (3)$$

The log-likelihood ratio (LR) under H_0 (there is no QTL) and H_1 (there is a QTL) will be calculated. Significant QTLs were identified by comparing LR values with a critical threshold. The threshold can be empirically determined from permutation tests by reshuffling the phenotypic data 1,000 times (Churchill and Doerge, 1994). The top 5% of the maximum LR values were used as a critical threshold to indicate a significance level of 0.05.

Exploratory data analysis and visualisation, including curve fitting, correlation analysis, QTL mapping and epistasis detection, were performed using R-project version 4.0.3 (R Core Team, 2018), ggplot2 (3.3.2) and dplyr (1.0.2). Possible functions for all QTLs determined were annotated and predicted via BLAST in the "nr" database on the National Center of Biotechnology Information website (NCBI; http://blast.ncbi.nlm.nih.gov/), identified on Uniprot (http://www.uniprot.org/) and analysed for protein-protein interactions using online protein interactions analysis software STRING (http://string-db.org/).

Expression Analysis of Candidate Genes

Transcriptomic characterisation of biomass allocation *cov*QTLs identifies candidate genes that regulate the growth and development of specific organs in *C. bungei*. Different tissues (leaf, petiole, xylem and phloem) were collected from a more than one-hundred-year-old *C. bungei* in Nanyang, Henan province. RNA extraction was performed using the Plant Qiagen RNeasy kit according to the manufacturer's instructions. RNA sequencing libraries were constructed using an Illumina standard mRNA-Seq Prep Kit, and sequence data were obtained using the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, United States) to obtain 2 × 125 bp reads. Raw data were trimmed to remove adaptors, and reads of less than 100 bp were discarded to enhance the quality. The clean reads were mapped to the *C. bungei* genome using the TopHat2 package with default parameters and assembled using Cufflinks. Gene expression was expressed as fragments per kilobase of transcript per million fragments mapped reads (FPKM) values. To verify the significant QTLs, FPKM values for each candidate gene were computed to assess expression across four tissues, and the FPKM values were transformed using a log2 fold change. Fold changes in the expression of candidate genes were used to generate a gene expression heatmap using the R package "pheatmap".

RESULTS

How Trait-Trait Relationships Vary

Through plotting the correlation between traits within leaves and stem and fitting the curve by polynomial regression, we can visualise how biomass allocation influences the morphology and function of organs. In general, all growth traits exhibited considerable genotypic variation among the F_1 population. The great genetic variation was observed in the leaf area compared to the other three leaf traits (**Figure 1A**) as well as in wood density compared to the other three stem traits (**Figure 1B**). The male parent "16-PJ-3" showed a

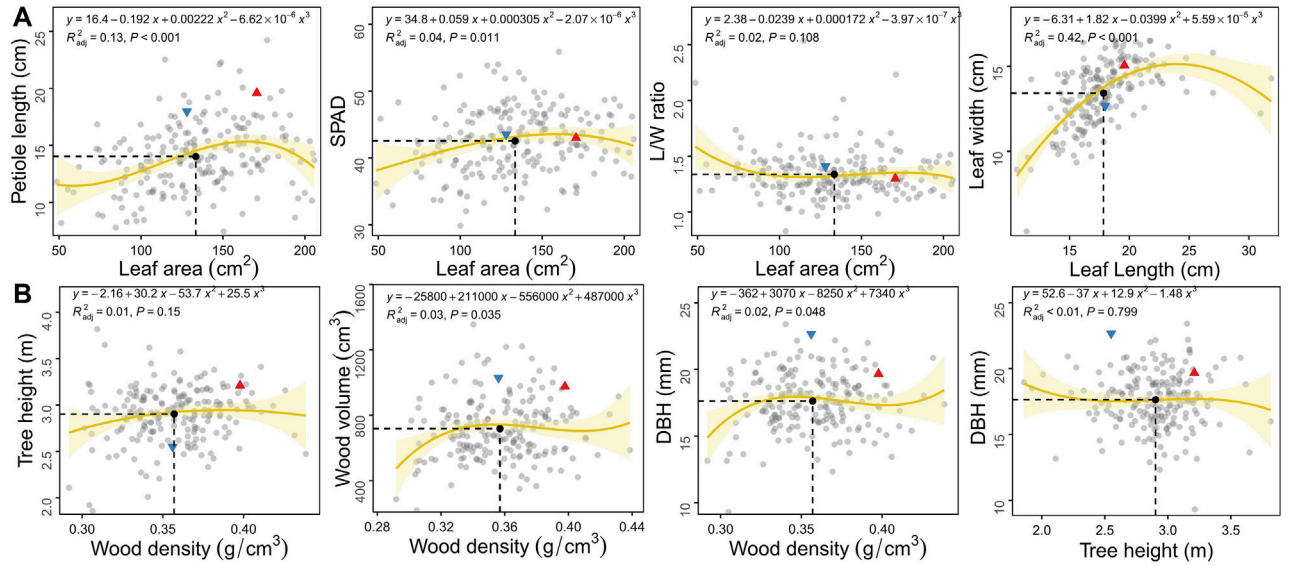


FIGURE 1 | Scatter plots of growth traits within organs of the leaf **(A)** and stem **(B)** in the F_1 population of *C. bungei*. The red triangle and blue inverted triangle denote the original parent plants "7080" and "16-PJ-3", respectively. The mean value of each trait pair among all individuals is indicated by the black solid circle crossed by vertical slash line and horizontal slash line.

smaller leaf size and wood base density than the female parent "7080". There were little differences in leaf function (SPAD), leaf morphology (L/W ratio), wood volume and stem mass between parents. Neither accession was completely distributed on the two sides on the diagonal of mean values. For each trait-trait pair, numerous transgressive segregants beyond the range of two original parents were detected, indicating that some alleles increased the phenotypic values and others decreased the values.

Relationships between leaf area and other traits (petiole length, SPAD and L/W ratio) indicated the balanced growth of leaf biomass allocation, photosynthetic function and leaf morphology, respectively. At the same mean value of leaf area (13.34 cm^2), leaves tended to maintain higher photosynthetic function instead of petiole elongation and leaf morphology (L/W ratio, **Figure 1A**). According to the fitting curve, petioles fluctuated with the increase in leaf area, gradually increased to the plateau stage and then began to decrease. Some individuals had large leaf areas (>180 cm^2) and relatively short petioles (<13 cm). The SPAD and L/W ratios changed little with leaf area, the SPAD content of small leaves was lower, and the leaf shape was longer and narrower. In terms of biomass allocation between leaf length and width, greater blade width was observed as the length increased at the beginning. However, leaf width decreased when LL > 24 cm.

By analysing the relationship between wood density and primary growth (H), wood volume (V) as well as secondary growth (DBH), the C allocation tradeoff between stem cell number and volume was demonstrated. Tree radial growth increased slightly, gradually plateaued and finally declined slightly (**Figure 1B**). Similar to WD-DBH and WD-V, thinner and smaller trees also had lower wood density. The C allocation between plant height and DBH was very important for plant morphology. For the current population, the variation between tree height and DBH generally decreased with greater plant height. Biomass was preferentially allocated to wood density, rather than tree height.

Biomass partitioning between leaves and stems could reflect the connection between source and sink as well as the functional tradeoff between sunlight interception by tree height and photosynthesis by leaves (**Figure 2**). We selected three leaf traits, LA, L/W and SPAD, which could represent the three aspects of leaf biomass, morphology and function, respectively. We then performed correlation analysis between leaf and stem traits. Stem elongation, thickening and wood density did not change with an increase in the area of a single leaf. This phenomenon demonstrates that the leaf area of the whole tree should be considered rather than that of a single leaf. Tree height increased with leaf area, leaf morphology and SPAD, indicating that larger, longer and higher SPAD leaves could promote faster axial growth. The influence of leaf morphology (L/W) on stem radial growth, biomass, wood density and volume was more dramatic than growth in height. The leaves of individuals with greater DBH were shallow rectangular (L/W < 1), while the leaves of those with a lower DBH were more elongated (L/W > 2.2). The relationship between leaf photosynthesis and stem phenotype varied in a manner similar to that for leaf morphology, yet with less fluctuation. Individuals with slower radial growth and lower wood density had more SPAD in the leaves in order to compete for light resources and enhance photosynthesis.

How *cov*QTLs Govern Trait-Trait Covariance

The bivariate mapping model identified 123 significant *cov*QTLs for all trait pairs, including 67 testcross and 57 intercross type SNPs (**Supplementary Figure S1, Supplementary Table S1**), of

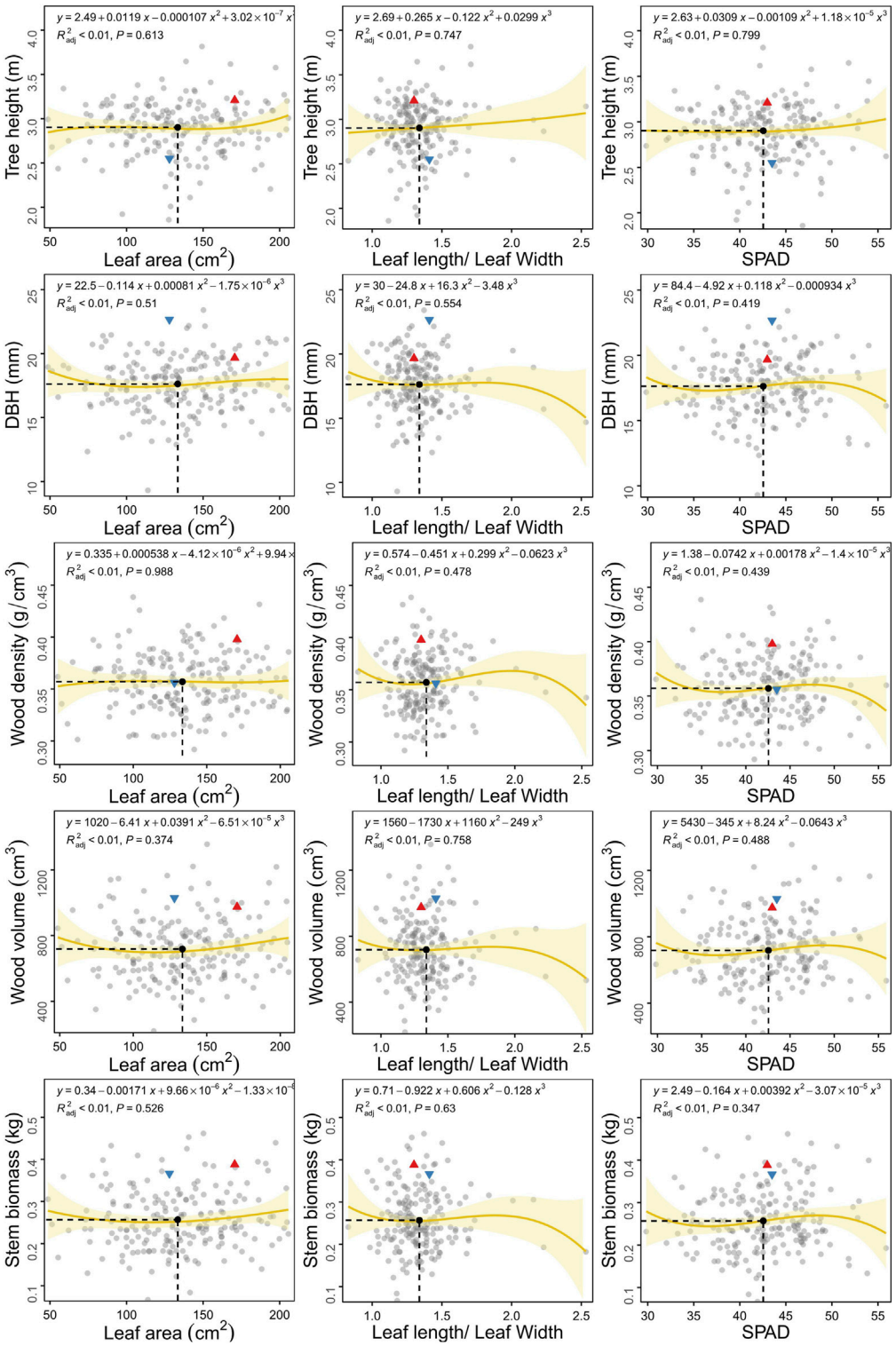


FIGURE 2 | Scatter plots of growth traits between leaf and stem in the F_1 population of *C. bungei*. The red triangle and blue inverted triangle denote the original parent plants "7080" and "16-PJ-3", respectively. The mean value of each trait pair among all individuals is indicated by the black solid circle crossed by vertical slash line and horizontal slash line.

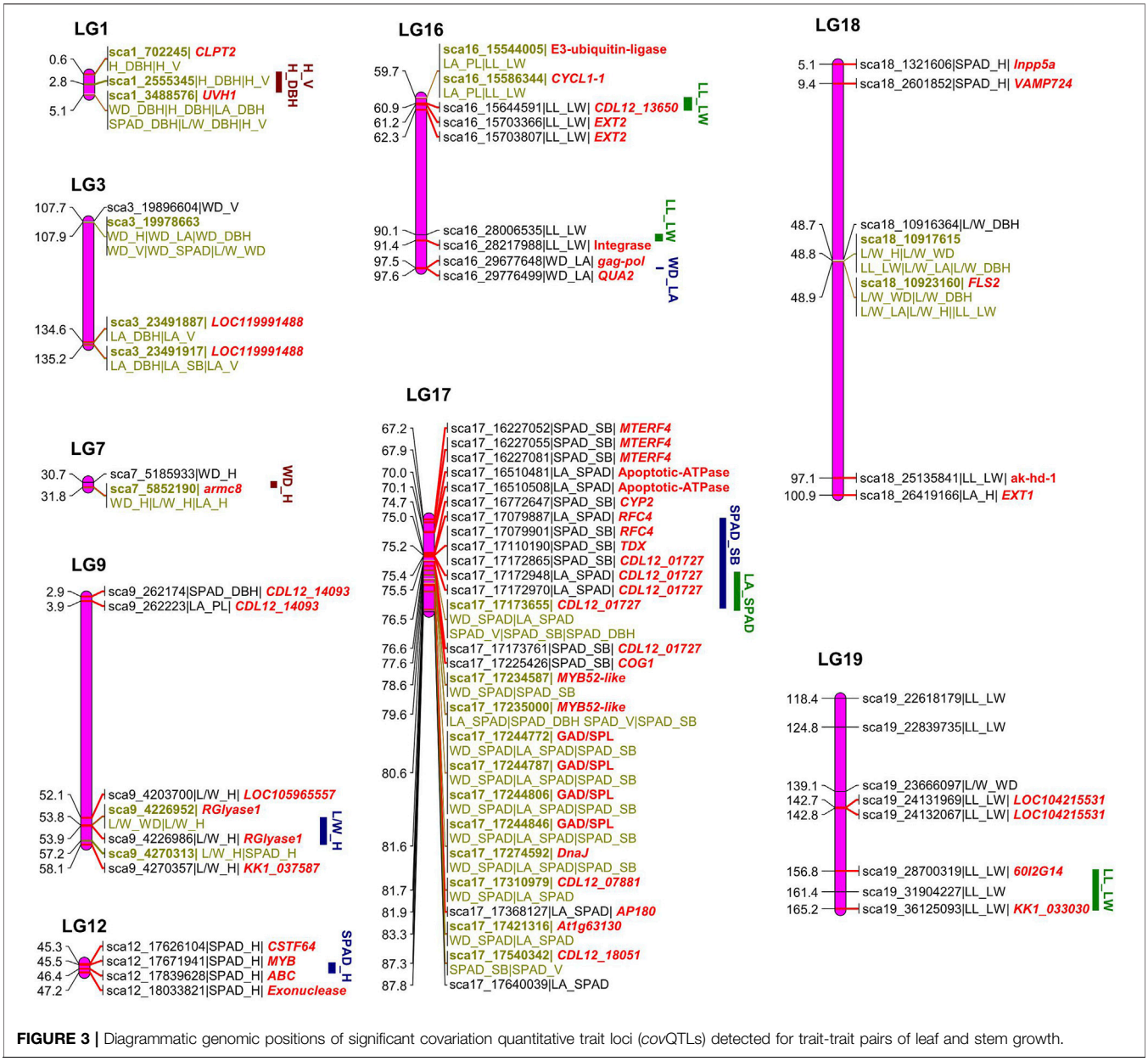


FIGURE 3 | Diagrammatic genomic positions of significant covariation quantitative trait loci (*cov*QTLs) detected for trait-trait pairs of leaf and stem growth.

which 73 unique SNPs remained after removing duplications. A total of 38 *cov*QTLs (24 testcross and 13 intercross type) were identified for trait-trait pairs in the leaf. There were 3, 16, 2 and 17 QTLs for LA-PL, LA-SPAD, L/W-LA and LL-LW, respectively, which were mainly located on lg16-lg19. The *cov*QTLs of C allocation for LL-LW were located at intervals of 59.7–91.4 cM on lg16 and 118.4–165.2 cM on lg19 (**Figure 3**). In addition, the QTLs of sca18_10917615 and sca18_10923160 were co-located with L/W-LA, indicating the reliability of mapping results based on similar traits. Coordinated variation between leaf area and leaf function (LA-SPAD) was observed relative to the interval of 70.0–87.8 cM on lg17 with peak QTL sca17_17310979 at 81.7 cM, overlapping with the stem biomass and leaf function (SB-SPAD). Regarding *cov*QTL mapping for the stem, 11 intercross type SNPs were discovered on lg1, 3, 7 and 15 for the trait pairs of H-DBH, WD-DBH, WD-V and WD-H. The C allocation of H-DBH was regulated by the QTL region of 0.6–5.1 cM on lg1 with the peak QTL sca1_3488576. Bivariate mapping was used to analyse significant QTLs for covariance between leaves and stem, resulting in a total of 71 significant *cov*QTLs (46 unique SNPs after removing duplications) related to 13 trait-trait pairs (**Supplementary Table S1**). Several significant QTL regions were detected in lg9, 12 and 17 for L/W-H, SPAD-H, and SPAD-SM. There were no QTLs for L/W-SB or L/W-V.

The functional annotations of 73 significant *cov*QTLs were identified via BLAST in the "nr" database on NCBI, with 44 gene terms after removing duplications for all trait pairs (**Figure 3**). The putative genes were mainly aligned to the

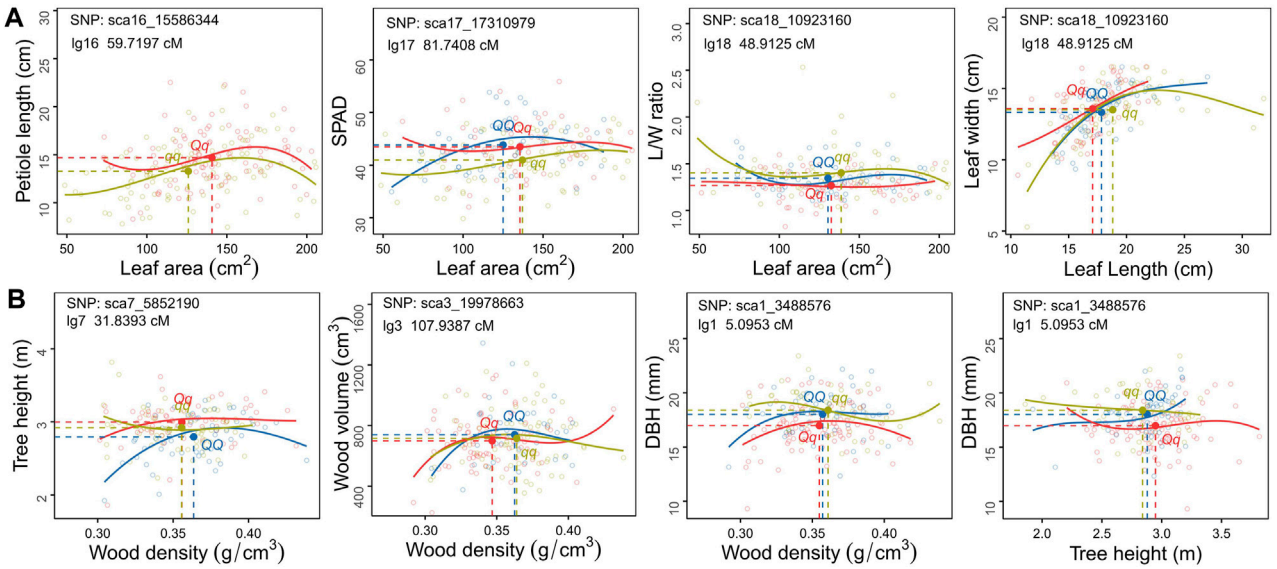


FIGURE 4 | Scatter plots of carbon allocation between traits within leaf **(A)** and stem **(B)** organs for different *C. bungei* genotypes. The mean values of each trait pair of different genotype populations is indicated by the blue (*QQ*), red (*Qq*) and ginger (*qq*) solid circles crossed by vertical slash lines and horizontal slash lines.

genomes of *Sesamum indicum*, *Populus deltoids*, *Handroanthus impetiginosus* and *Erythranthe guttata* (**Supplementary Table S1**). Several important genes were annotated by multiple QTLs (**Supplementary Table S1**; **Figure 3**). QTLs sca9_4226952 and sca9_4226986 related to L/W-H and SPAD-H, respectively, encode rhamnogalacturonate lyase B (*rglB*), which is involved in the carbohydrate metabolic process and catalytic activity. Three QTLs related to SPAD-SM in a narrow QTL region of 67.2–70.0 cM on lg17 encode the mitochondrial transcription termination factor 4 (MTERF4), which participates in the regulation of mitotic cell cycle spindle assembly checkpoint. Four QTLs between 80.6–81.6 were related to glutamate decarboxylase/sphingosine phosphate lyase (*GAD*/*SPL*), playing a key role in the carboxylic acid metabolic process. The replication factor C subunit, *RFC4*, which was related to sca17_17079887 and sca17_17079901, participates in the biological process of DNA biosynthesis, DNA replication, ATP binding and nucleotidyltransferase activity as part of the DNA polymerase III complex. More importantly, some *cov*QTLs were co-located for several trait-trait pairs, which were labelled on the linkage map (**Figure 3**). Among them, sca1_3488576, sca3_19978663, sca3_23491917, sca17_17173655, sca17_17235000, sca18_10917615 and sca18_10923160 were associated with more than three trait pairs, suggesting that they may play an important role in C allocation and coordinated growth.

How Biomass Allocation-Related covQTLs Remould the Tree Growth Model

Genotype-specific trait-trait correlation analyses were conducted for each *cov*QTL with a maximum LR value (**Figures 4**, **5**). Among 200 accessions, two or three groups could be established based on the testcross or intercross type. The cyclin-L1-1 (*CYCL1-1*) gene, located within sca16_15586344, participates in cell division, transcription regulation by RNA polymerase II and catalytic activity. Individuals with genotype *Qq* exhibited larger leaf size and longer petioles than those with genotype *qq*. For *cov*QTL sca17_17310979, the *Qq* genotype had a higher SPAD than *qq* at a similar leaf area, while having a larger leaf area than *QQ* at approximately equal SPAD, suggestive of allelic additive effects for the trade-off between leaf area and SAPD. The marker sca18_10923160 was co-located for five trait pairs, which was the peak QTL for L/W-LA and LL-LW (**Figure 1A**). LRR receptor-like serine/threonine-protein kinase *FLS2*, detected in four trait pairs, including L/W-WD, L/W-DBH, L/W-H and L/W-LA, was associated with the defense response via callose deposition in the cell wall, receptor-mediated endocytosis and anion channel activity (**Table 1**; **Supplementary Tables S2, S3**).

As for the covariation between leaf and stem, the degrees of variation among genotypes *QQ*, *Qq* and *qq* of sca1_3488576 were distinct among LL-DBH, L/W- DBH and SPAD-DBH (**Figure 5**). The marker sca17_17235000 encodes transcription factor MYB52-like, which had the maximum LR value in SPAD-V and SPAD-SB. The mean values and fitting curves of different genotypes were similar for SPAD-V and SPAD-SB at sca17_17235000 and sca3_23491917. For the trade-off between tree height and leaf traits, the co-located marker sca9_4270313 displayed different patterns among genotypes, as genotype *Qq* with the smallest L/W and medium SPAD tended to promote plant height (**Figure 5**). The most significant *cov*QTL sca3_19978663 was co-located with three pairs of WD-leaf traits (LA, L/W and SPAD). Populations with genotypes *QQ* and *qq* displayed higher wood density and DBH than *Qq* at similar leaf areas and SPAD.

The *cov*QTL sca1_3488576 with the maximum LR value was co-located for five trait pairs, including WD_DBH, H-DBH, LA-

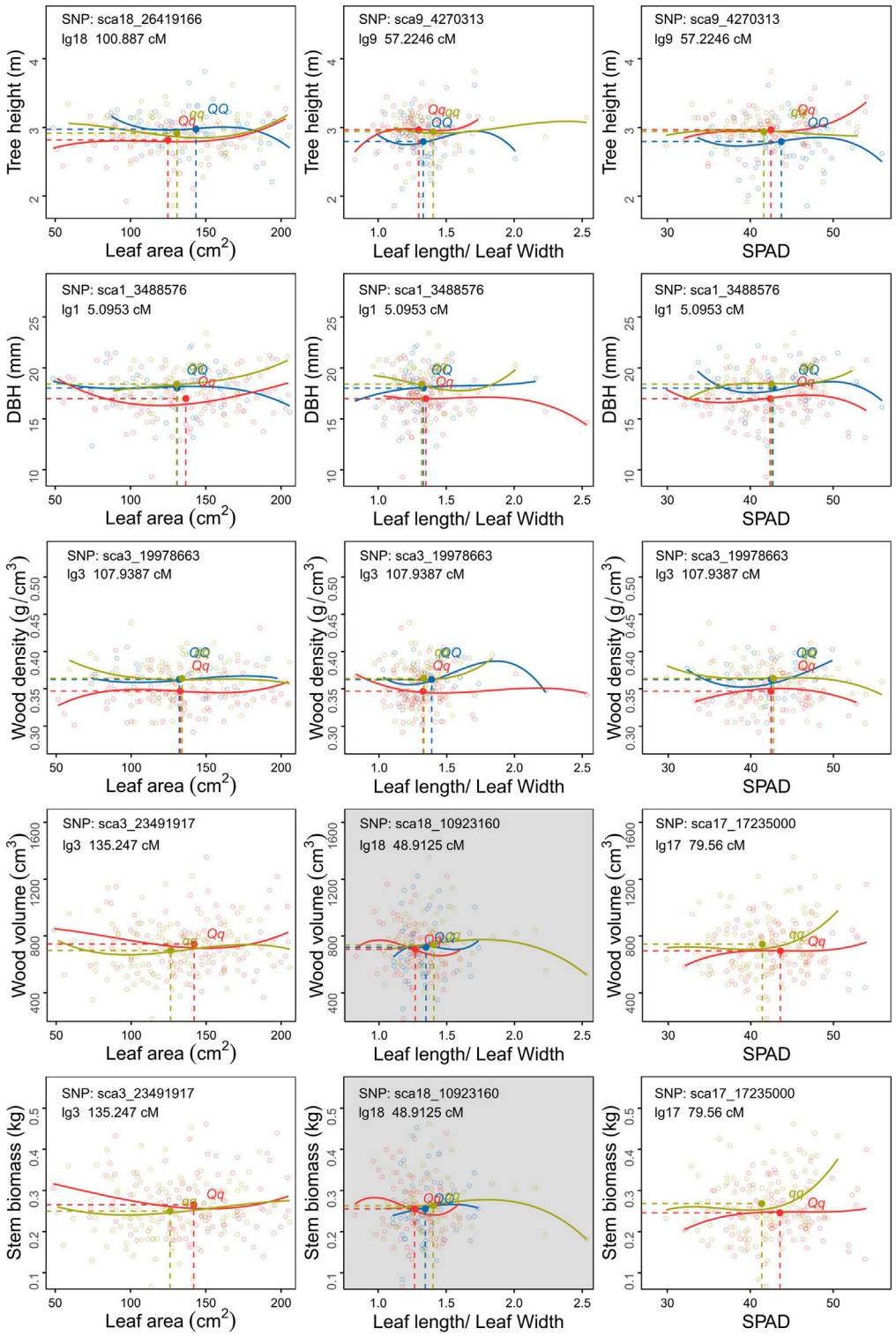


FIGURE 5 | Scatter plots of carbon allocation between leaf and stem traits of different genotype populations of *C. bungei*. The mean values of each trait pair of different genotypes populations is indicated by the blue (*QQ*), red (*Qq*) and ginger (*qq*) solid circles crossed by vertical slash lines and horizontal slash lines.

TABLE 1 | The positions, LR values and functional annotation are shown for the significant SNPs that affect biomass allocation within or between leaf and stem growth of *C. bungei*.

No	Trait pairs	SNP	LG	Position (cM)	LR	Annotation
1	WD-DBH	sca1_3488576	1	5.10	19.39	*UVH1*
2	WD-H	sca7_5852190	7	31.84	18.82	*armc8*
3	WD-LA	sca16_29677648	16	97.54	16.12	*gag-pol*
4	WD-LA	sca16_29776499	16	97.62	17.38	*QUA2*
5	WD-SPAD	sca17_17234587	17	78.56	15.80	*MYB52-like*
7	WD-SPAD	sca17_17244787	17	80.56	16.58	*GAD/SPL*
10	WD-SPAD	sca17_17274592	17	81.61	16.43	*DnaJ*
12	WD-V	sca15_9179490	15	39.33	16.76	*CFOL_v3_36453*
13	H-DBH	sca1_702245	1	0.64	18.54	*CLPT2*
14	H-DBH	sca1_3488576	1	5.10	22.47	*UVH1*
15	H-V	sca1_702245	1	0.64	18.23	*CLPT2*
16	H-V	sca1_3488576	1	5.10	21.61	*UVH1*
17	L/W-WD	sca9_4226952	9	53.81	18.72	*rglB*
18	L/W-WD	sca18_10923160	18	48.91	18.51	*FLS2*
19	L/W-DBH	sca18_10923160	18	48.91	18.57	*FLS2*
20	L/W-H	sca7_5852190	7	31.84	16.94	*armc8*
21	L/W-H	sca9_4226952	9	53.81	20.95	*rglB*
25	L/W-H	sca18_10923160	18	48.91	17.45	*FLS2*
26	L/W-LA	sca18_10923160	18	48.91	18.03	*FLS2*
27	LA-DBH	sca1_3488576	1	5.10	21.40	*UVH1*
28	LA-H	sca7_5852190	7	31.84	18.10	*armc8*
29	LA-H	sca18_26419166	18	100.89	18.25	*EXT1*
30	LA-PL	sca16_15544005	16	59.65	17.24	*E3 ubiquitin ligase*
31	LA-PL	sca16_15586344	16	59.72	17.27	*CYCL1-1*
33	LA-SPAD	sca17_16510508	17	70.08	18.29	*Apoptotic ATPase*
34	LA-SPAD	sca17_17079887	17	75.05	17.68	*RFC4*
35	LA-SPAD	sca17_17235000	17	79.56	16.50	*MYB52-like*
37	LA-SPAD	sca17_17244787	17	80.56	18.97	*GAD/SPL*
40	LA-SPAD	sca17_17274592	17	81.61	18.62	*DnaJ*
41	LA-SPAD	sca17_17368127	17	81.93	16.88	*AP180*
43	LL-LW	sca16_15544005	16	59.65	14.90	*E3 ubiquitin ligase*
44	LL-LW	sca16_15586344	16	59.72	15.72	*CYCL1-1*
46	LL-LW	sca16_15703807	16	62.35	14.67	*EXT2*
47	LL-LW	sca16_28217988	16	91.45	16.75	*Integrase*
48	LL-LW	sca18_10923160	18	48.91	18.04	*FLS2*
49	LL-LW	sca18_25135841	18	97.10	16.74	*ak-hd 1*
50	LL-LW	sca19_28700319	19	156.76	14.02	*60I2G14*
51	SPAD-DBH	sca1_3488576	1	5.10	18.43	*UVH1*
52	SPAD-DBH	sca17_17235000	17	79.56	14.34	*MYB52-like*
53	SPAD-H	sca12_17626104	12	45.27	16.72	*CSTF64*
54	SPAD-H	sca12_17671941	12	45.49	16.23	*MYB*
55	SPAD-H	sca12_17839628	12	46.39	16.48	*ABC*
56	SPAD-H	sca12_18033821	12	47.22	17.05	*Exonuclease*
57	SPAD-H	sca18_1321606	18	5.07	16.56	*Inpp5a*
58	SPAD-H	sca18_2601852	18	9.39	17.18	*VAMP724*
60	SPAD-SB	sca17_16227055	17	67.89	17.21	*MTERF4*
62	SPAD-SB	sca17_16772647	17	74.69	17.25	*CYP2*
63	SPAD-SB	sca17_17079901	17	75.19	16.80	*RFC4*
64	SPAD-SB	sca17_17110190	17	75.21	16.89	*TDX*
65	SPAD-SB	sca17_17225426	17	77.56	16.95	*COG1*
67	SPAD-SB	sca17_17235000	17	79.56	18.43	*MYB52-like*
68	SPAD-SB	sca17_17244772	17	80.56	18.05	*GAD/SPL*
72	SPAD-SB	sca17_17274592	17	81.61	17.34	*DnaJ*
73	SPAD-V	sca17_17235000	17	79.56	16.07	*MYB52-like*

DBH, SPAD-DBH and L/W-DBH (**Figures 4B**, **5**). This marker was related to the gene encoding DNA repair endonuclease UVH1 isoform X1 (*UVH1*), which is involved in nucleic acid phosphodiester bond hydrolysis and endonuclease activity. For the C distribution of WD-DBH and H-DBH, *UVH1* (sca1_3488576) plays different regulatory roles. At the same DBH value, the subpopulation of genotype *qq* tended to invest more photoassimilates to wood density than tree height, while genotype *Qq* preferentially allocated photoassimilates to tree height instead of wood density (**Figure 4B**) and also allocated more nutrition for leaf area expansion instead of SAPD increase (**Figure 5**).

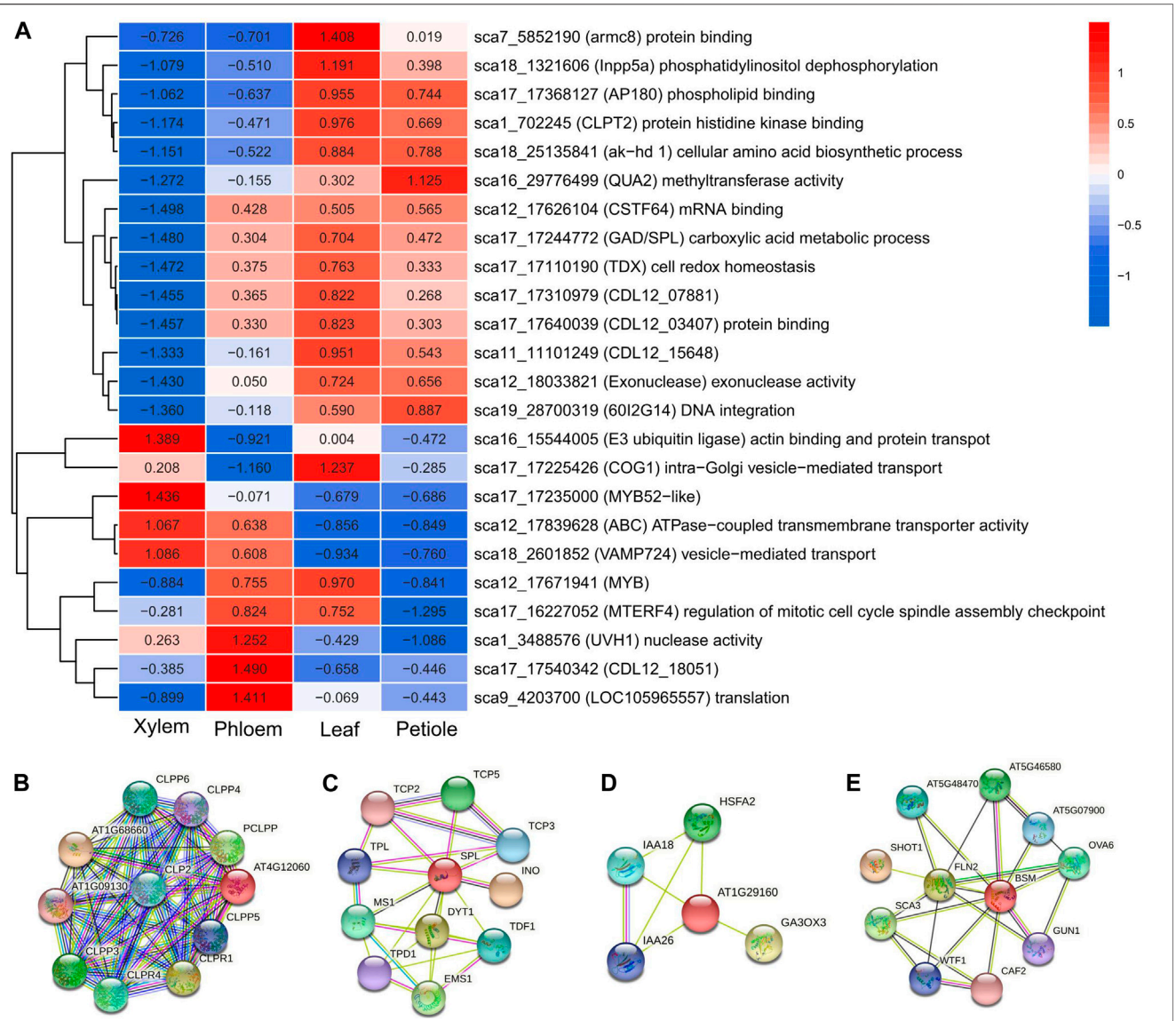


FIGURE 6 | (A) Heat map of the gene expression analysis for 24 biomass distribution-related genes using RNA-seq data from four tissues of *C. bungei*. **(B–E)** Interaction network of *CLPT2*, *SPL*, *COG1* (*At1g29160*) and *MTERF4* (*BSM*) in *Arabidopsis thaliana*.

Expression of Candidate Genes Associated With Trait-Trait Relationships

We investigated the expression profiles of the 123 *cov*QTLs to assess the overlapping genes via QTL and transcriptomic approaches. A total of 24 candidate genes were found to be significant in all trait-trait pairs (**Figure 6A**). Cluster analysis of gene expression values revealed two gene clusters. Most of the Cluster I genes were upregulated in the leaf tissues and downregulated in stem tissues, with related candidate *cov*QTLs also mainly detected in leaf trait pairs, including sca17_17368127, sca17_17640039 and sca17_17244772 for LA-SPAD, sca18_25135841 and sca19_28700319 for LL-LW. These genes are involved in the biological processes of biomacromolecule synthesis, such as phospholipid binding and endocytosis (*AP180* and *Inpp5a*), protein binding (*CDL12_03407* and *armc8*), cellular amino acid biosynthetic process (*ak–hd1*), DNA integration (*60I2G14*), exonuclease activity (Exonuclease) and methyltransferase activity (*QUA2*).

ATP-dependent Clp protease ATP-binding subunit *CLPT2* (sca1_702245) is an important component of the chloroplast, acting as an accessory protein that regulates the assembly of the plastidial Clp protease system, which is involved in the protein interaction network of proteolysis protein metabolic process and chloroplast organisation. *CLPT2* (*At4g12060*) was upregulated only in leaves (**Figure 6B**; **Supplementary Tables S2, S3**). As a subunit of the Clp core complex, *CLPT2* plays an important role in leaf growth and development, including leaf colour, shape and growth rate (Kim et al., 2015). Several genes with high expression

levels in both leaves (leaf and petiole) and stem (phloem), such as *CSTF64* (sca12_17626104), *GAD/SPL* (sca17_17244772) and *TDX* (sca17_17110190), were involved in the processes of mRNA binding, carboxylic acid metabolism and cell redox homeostasis, respectively. *GAD/SPL* was the key gene related to several trait pairs and covers, being involved in floral organ development, floral organ morphogenesis and formation, regulation of shoot apical meristem development as well as macromolecule metabolic processes such as nitrogen compound and RNA metabolism. *SPL*, *YAB*3 and *AFO* likely have transcription factor activity and were involved in the abaxial cell fate determination during embryogenesis and organogenesis (**Figure 6C**; **Supplementary Tables S2, S3**).

All Cluster II genes were downregulated in the petiole and half in the leaves. In contrast, most of these genes were upregulated in the phloem and half in the xylem. We observed that sca16_15544005 (*E3 ubiquitin ligase*), sca17_17225426 (*COG1*), sca12_17839628 (*ABC*) and sca18_2601852 (*VAMP724*), which are implicated involved in substance transportation, were highly expressed in the xylem. E3 ubiquitin ligase enables actin binding and ligase activity, thus being involved in protein transport. ABC participates in ATPase-coupled transmembrane transporter activity and ATP binding. Vesicle-associated membrane protein 724 (*VAMP724*) is involved in vesicle-mediated transport. *COG1* is upregulated in both leaves and xylem, consistent with the related trait pair of SPAD-stem biomass.

COG1 is involved in intra-Golgi vesicle-mediated transport as part of the Golgi transport complex located on the membrane of the Golgi apparatus. STRING analysis of *COG1* (*At1g29160*) indicated that it acts as a negative regulator of phytochrome-mediated light responses. The interactive protein *GA3OX3* is involved in the production of bioactive GA by converting inactive gibberellin precursors *GA9* and *GA20* in the bioactive gibberellins GA4 and GA1. *PHYA* and *PHYB* are regulatory photoreceptors that exist in two forms, Pr and Pfr, which are reversibly interconvertible by light. Photoconversion of Pr to Pfr induces an array of morphogenetic responses, whereas the reconversion of Pfr to Pr cancels the induction of these responses (**Figure 6D**; **Supplementary Tables S2, S3**). *MTERF4* (sca18_2601852) regulates mitotic cell cycle spindle assembly checkpoint and chloroplast organisation, showing high expression in the leaves and phloem at the same time. *MTERF4* (*BSM*) is required for the maturation of 16S rRNA and 23S rRNA in the chloroplast, thus maintaining appropriate transcript levels in the mitochondria and chloroplasts. *MTERF4* is involved in the regulation of cellular macromolecule biosynthetic process, nitrogen compound metabolic process, primary metabolic process, transcription and gene expression. Interactive proteins *FLN1* and *FLN2* of *BSM* were required for proper chloroplast development (**Figure 6E**; **Supplementary Tables S2, S3**).

DISCUSSION

The coordinated growth and biomass allocation among organs determine their adaptability to the environment, which is of major significance to the economic, ecological and social aspects of plant cultivation (Ambrose et al., 2009; Zhang and Xi, 2021). Plants maximise overall growth through the optimal utilisation and allocation of resources available within their environment, thus leading to the growth of different organs at distinct rates (Poorter et al., 2012). Based on the F_1 population of the woody plant *C. bungei*, different trade-off patterns were observed for various trait pairs (**Figures 1**, **2**). We evaluated leaf phenotypic data for a single leaf, which is not representative of the total leaf biomass for the whole plant. Therefore, there was no significant linear relationship between leaf and stem phenotypes. We used polynomial regression curve fitting to infer potential relationships between traits. In most trait combination pairs, the parents were clearly separated among the population, except for leaf L/W ratio, wood volume, stem biomass and SPAD. At the mean value of leaf area, *C. bungei* prioritised C allocation to increase SPAD, that is, chlorophyll content, which directly improved photosynthesis (Niinemets, 2007). Petiole elongation is the second preferred method for ensuring the appropriate spatial distribution of leaves to intercept more light energy (Levionnois et al., 2020). This C distribution strategy indicated that the plant accelerated vegetative and biomass accumulation at the sapling/early stage through leaf area expansion to increase photosynthetic efficiency (Liu et al., 2020).

Organisms tend to increase their surface area to enhance metabolic capacity in an attempt to produce sufficient matter and energy for survival and reproduction (Wang and Wu, 2004). At the same mean value of DBH (17.63 mm), individuals preferentially invested more nutrients into height (2.90 m) for increasing the surface area, rather than in wood density (0.36 g/cm^3). This allocation pattern is consistent with the growth law of woody plants (Spicer and Groover, 2010). The trees prioritise primary growth at the early juvenile stage, which refers to the elongation growth of plant organs caused by the division, differentiation and growth of the apical meristem cells of roots and stems (Zhang et al., 2017). After the completion of primary growth, most dicotyledonous plants accelerate the secondary growth process of continuously producing secondary xylem and secondary phloem with cambium activity, resulting in a thickened root and stem diameter as well as greater tissue density (Spicer and Groover, 2010). As per leaf growth and development, trees in the present study tended to develop longer leaves, with mean values of leaf length and width at 17.84 and 13.43 cm, respectively. The relationship between leaf morphology traits requires further study in different populations or growth stages.

Previous studies on genetic variation in relation to the growth traits and genotype-by-environment of *C. bungei* indicated that the height and stem volume of *C. bungei* were under strong genetic control (Xiao et al., 2019). The genetic basis of coordinated growth and trade-off strategies among growth traits have since attracted great interest among biologists. However, the conventional QTL mapping method focusing on genetic variation at a single trait may fail to reveal the genetic architecture underlying trait-trait balanced growth (Gan et al., 2019; Zhang et al., 2020b). Currently, there are two statistical approaches for QTL mapping, namely the mixture model for sparse molecular markers and the multiplicative model for dense

markers. Inclusive composite interval mapping with a mixture model was carried out to detect QTL regions related to a single growth trait in a previous study (Lu et al., 2019). Because the constructed linkage map was quite dense, we employed bivariate mapping with the multiplicative model that assumes pleiotropic QTLs were located at the positions of markers. Compared with single-trait QTL mapping, *cov*QTL analysis considers two-dimensional phenotypic data at the same time to address the complexity and multiplicity of the trait, thus improving the accuracy and efficiency of QTL mapping. We detected 123 *cov*QTLs for all trait-trait pairs, some of which were consistent with previous QTL mapping results. The *cov*QTL of C allocation trade-off of leaf traits on lg16, 17, 18 and 19 overlapped with the intervals of Q16-60, Q17-84, Q18-99 and Q19-137 reported for leaf traits in a previous study (Lu et al., 2019). Further, we also detected some important *cov*QTLs which influenced the C distribution between stem traits or stem-leaf traits, such as the QTL region of 0.6–5.1 cM on lg1 for H-DBH as well as regions on lg9, 12 and 17 for L/W-H, SPAD-H and SPAD-SM, respectively. It should be noted that the region with density SNPs between 76.5 and 87.8 CM on lg17 was associated with multiple trait pairs including WD-SPAD, SPAD-SB, SPAD-V and LA-SPAD simultaneously. It was illustrated that lg17 was not only an important region related to leaf phenotype growth and development, but was also related to C allocation between SPAD and wood density.

The binary covariate genetic mapping analysis model was used to genetically analyse the unequal growth relationships between organs and phenotypes, in an attempt to better interpret the biological significance of *cov*QTLs (Gan et al., 2019). Various members in these gene families were involved in plant development, carboxylic acid metabolism, cell division and substance transportation. *RFC4*, *gag-pol*, integrase, and *KK1_037587* were involved in DNA replication and integration (**Table 1**; **Supplementary Tables S2, S3**). *RFC4* is a subunit of *RFC*, which might have a significant in *A. thaliana* embryo development, with embryonic lethality observed in *AtRFC* mutants (Zhang et al., 2020a). The *rglB* gene is involved in the carbohydrate metabolic process and catalytic activity. *CYCL1-1* is part of a molecular thermometer fine-tuning environmental information into the alternative splicing of target genes and is required for cell cycle regulation, transcription via RNA polymerase II and mRNA splicing (Cavallari et al., 2018; Nibau et al., 2019). Flagellin-sensitive 2 (*FLS2*), which was detected for many trait pairs, functions as a potent elicitor of the pathogen defence response in the cell wall and regulates receptor-mediated endocytosis as well as anion channel activity (Gomez-Gomez and Boller, 2000; Withers and Dong, 2017). The cytochrome P450 *CYP2* subfamily has been shown to participate in plant growth, namely in internode elongation by modulating the gibberellin response in rice (Luo et al., 2006) as well as cell elongation or growth in plant height via brassinosteroid biosynthesis (Hong et al., 2003; Chakrabarti et al., 2013).

We utilised transcriptomic data to identify a set of tissue-specific genes among those underlying C allocation as a complementary approach to QTL analysis. The genes derived from the overlap of differentially expressed genes and *cov*QTLs could be used to verify candidates. Two gene clusters were obtained, and the expression patterns in different tissues were in agreement with the *covQTL* distributions in trait pairs. Most of the highly expressed genes in leaves and petioles, which were low in the stems, were usually involved in the synthesis of biological macromolecules and cellular component organisation, including phospholipid binding and endocytosis, cellular amino acid biosynthetic process, DNA integration, exonuclease activity and methyltransferase activity (**Figure 6**; **Supplementary Tables S2, S3**). *GAD*/*SPL*, *TDX* and *CSTF64* were upregulated in all tissues except the xylem. *GAD*/*SPL* was the key gene related to five trait pairs and *cov*QLTs, being involved in the regulation of floral organ morphogenesis and formation as well as shoot apical meristem development (Wei et al., 2015). *SPL* interaction partner *YAB*3 was reported to sustain adaxial-abaxial polarity by specifying the abaxial cell fate (Fritz et al., 2018). Further, YAB genes participate in expanding flat leaf development through genetic programs that are related to marginal auxin flow and the activation of a maturation schedule directing determinate growth (Machida et al., 2015). *CSTF64* regulates developmental growth through mRNA surveillance, the antisense RNA metabolic process and regulation of gene silencing (Czesnick and Lenhard, 2016). *CLPT2*, which was upregulated only in leaves, participated in proteolysis protein metabolic processes and chloroplast organisation. The *CLPT1* and *CLPT2* proteins are unique in land plants to stabilise the Clp core complex (Moreno et al., 2017). In Arabidopsis, the *clpt1*/*clpt2* double mutant showed delayed growth, pale phenotype and altered leaf shape with more serrates on the leaf margins (Kim et al., 2015).

On the other hand, some genes were highly expressed in xylem and phloem instead of leaves and petioles. These were mainly responsible for material transportation. In particular, *COG1* regulates plant phototropism and gravitropism, participates in the detection of visible light as well as the response to light intensity and is involved in the far-red and red light signalling pathways, thus influencing the circadian rhythm of plants (Park et al., 2003). *COG1* may be of significance in light intensity detection in leaves as well as stem elongation. *MYB* plays a role in the negative regulation of gene expression, cellular macromolecule biosynthetic process and nitrogen compound metabolic process. In trees, the overexpression of *MYB* transcription factors results in enhanced photosynthesis, antioxidant enzyme production and enhanced growth under water stress (Polle et al., 2019). *ABC* plays a role in the biosynthetic process of dTMP, organonitrogen compound and organophosphate as well as in organic substance transmembrane transport (**Supplementary Tables S2, S3**). The *MTERF4* (*BSM*) detected in a significant QTL region for SPAD_SB presented a high expression level in both phloem and leaf, highlighting the pleiotropy of this gene. *MTERF4* participates in the regulation of mitotic cell cycle spindle assembly checkpoint as well as chloroplast organisation and is essential for normal plant development in addition to the maintenance of adequate levels of transcripts in both mitochondria and chloroplasts (Sun et al., 2016).

In summary, we explored the genetic basis of biomass allocation among organs using a bivariate QTL mapping model and verified the function of key genes via gene expression profiling in *C. bungei*, thus providing a basis for understanding the complex genetic regulation of trait-trait

coordinated variation. The study of gene expression regulation further confirmed the reliability of *cov*QTLs obtained via the binary covariate genetic mapping model and helped explain the biological function of key genes. Narrowing down candidate genes could make functional verification feasible in future studies. Importantly, more biologically meaningful genes will be identified through multivariate (bivariate or even trivariate) mapping for multi-trait genetic dissection at different growth stages of perennial woody plants. Further, the integration of metabolomics, phenomics, genetic mapping, gene expression profiling, enzymatic activity analysis and genome-editing techniques will facilitate the identification of candidate genetic factors, thus benefiting tree breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI using accession number PRJNA551333.

AUTHOR CONTRIBUTIONS

JW and MZ conceived the ideas and designed the research; GY, CS, and WM performed the experiments and collected the data; MZ and NL analyzed the data and designed the figures; YF and BL prepared the tables; MZ and JW led the writing of the manuscript; TZ, GQ, and NL revised the manuscript. All the authors read and approved the final version of the manuscript.

FUNDING

This research is financially supported by grants from the Fundamental Research Funds of Chinese Academy of Forestry (CAFYBB2021QD001-2) and the National Natural Science Foundation of China (32001337).

ACKNOWLEDGMENTS

We would like to thank Luoyang Academy of Agriculture and Forestry for providing plant materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.758209/full#supplementary-material

REFERENCES

Ambrose, A. R., Sillett, S. C., and Dawson, T. E. (2009). Effects of Tree Height on branch Hydraulics, Leaf Structure and Gas Exchange in California Redwoods. *Plant Cel Environ.* 32 (7), 743–757. doi:10.1111/j.1365-3040.2009.01950.x

Ammitzboll, H., Vaillancourt, R. E., Potts, B. M., Harrison, P. A., Brodribb, T., Sussmilch, F. C., et al. (2020). Independent Genetic Control of Drought Resistance, Recovery, and Growth of *Eucalyptus Globulus* Seedlings. *Plant Cel Environ.* 43 (1), 103–115. doi:10.1111/pce.13649

Berny Mier, Y., Teran, J. C., Konzen, E. R., Palkovic, A., Tsai, S. M., Rao, I. M., Beebe, S., et al. (2019). Effect of Drought Stress on the Genetic Architecture of Photosynthate Allocation and Remobilization in Pods of Common Bean (*Phaseolus vulgaris* L.), a Key Species for Food Security. *BMC Plant Biol.* 19, 171. doi:10.1186/s12870-019-1774-2

Brienen, R. J. W., Gloor, E., Clerici, S., Newton, R., Arppe, L., Boom, A., et al. (2017). Tree Height Strongly Affects Estimates of Water-Use Efficiency Responses to Climate and CO_2 Using Isotopes. *Nat. Commun.* 8 (1), 288. doi:10.1038/s41467-017-00225-z

Cavallari, N., Nibau, C., Fuchs, A., Dadarou, D., Barta, A., and Doonan, J. H. (2018). The Cyclin-Dependent Kinase G Group Defines a Thermo-Sensitive Alternative Splicing Circuit Modulating the Expression of Arabidopsis *ATU2AF65A*. *Plant J.* 94 (6), 1010–1022. doi:10.1111/tpj.13914

Chakrabarti, M., Zhang, N., Sauvage, C., Munos, S., Blanca, J., Canizares, J., et al. (2013). A Cytochrome P450 Regulates a Domestication Trait in Cultivated Tomato. *Proc. Natl. Acad. Sci.* 110 (42), 17125–17130. doi:10.1073/pnas.1307313110

Chen, G., Yang, K., Zhang, J., Wang, M., Wang, L., and Xian, J. (2021). Effects of Succession Stages and Altitudinal Gradient on Leaf Surface Area and Biomass Allocation of Typical Plants in the Subalpine of Eastern Tibetan Plateau. *Glob. Ecol. Conserv.* 27, e01590. doi:10.1016/j.gecco.2021.e01590

Chhetri, H. B., Macaya-Sanz, D., Kainer, D., Biswal, A. K., Evans, L. M., Chen, J. G., et al. (2019). Multitrait Genome-wide Association Analysis of Populus trichocarpa Identifies Key Polymorphisms Controlling Morphological and Physiological Traits. *New Phytol.* 223 (1), 293–309. doi:10.1111/nph.15777

Churchill, G. A., and Doerge, R. W. (1994). Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138 (3), 963–971. doi:10.1093/genetics/138.3.963

Cornelissen, J. H. C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D. E., et al. (2003). A Handbook of Protocols for Standardised and Easy Measurement of Plant Functional Traits Worldwide. *Aust. J. Bot.* 51, 335–380. doi:10.1071/BT02124

Czesnick, H., and Lenhard, M. (2016). Antagonistic Control of Flowering Time by Functionally Specialized Poly(A) Polymerases inArabidopsis Thaliana. *Plant J.* 88 (4), 570–583. doi:10.1111/tpj.13280

Fritz, M. A., Rosa, S., and Sicard, A. (2018). Mechanisms Underlying the Environmentally Induced Plasticity of Leaf Morphology. *Front. Genet.* 9, 478. doi:10.3389/fgene.2018.00478

Gan, J., Cao, Y., Jiang, L., and Wu, R. (2019). Mapping Covariation Quantitative Trait Loci that Control Organ Growth and Whole-Plant Biomass. *Front. Plant Sci.* 10, 719. doi:10.3389/fpls.2019.00719

Getahun, B. B., Visser, R. G. F., and Van Der Linden, C. G. (2020). Identification of QTLs Associated with Nitrogen Use Efficiency and Related Traits in a Diploid Potato Population. *Am. J. Potato Res.* 97 (2), 185–201. doi:10.1007/s12230-020-09766-4

Gómez-Gómez, L., and Boller, T. (2000). FLS2: An LRR Receptor–Like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. *Mol. Cel.* 5 (6), 1003–1011. doi:10.1016/S1097-2765(00)80265-8

Guo, Z., Zhou, S., Wang, S., Li, W.-X., Du, H., and Xu, Y. (2021). Identification of Major QTL for Waterlogging Tolerance in maize Using Genome-wide Association Study and Bulked Sample Analysis. *J. Appl. Genet.* 62, 405–418. doi:10.1007/s13353-021-00629-0

Hong, Z., Ueguchi-Tanaka, M., Umemura, K., Uozu, S., Fujioka, S., Takatsuto, S., et al. (2003). A rice Brassinosteroid-Deficient Mutant, *Ebisu dwarf* (*D2*), Is Caused by a Loss of Function of a New Member of Cytochrome P450. *Plant Cell* 15, 2900–2910. doi:10.2307/387192410.1105/tpc.014712

Kim, J., Kimber, M. S., Nishimura, K., Friso, G., Schultz, L., Ponnala, L., et al. (2015). Structures, Functions, and Interactions of ClpT1 and ClpT2 in the Clp Protease System of Arabidopsis Chloroplasts. *Plant Cell* 27, 1477–1496. doi:10.1105/tpc.15.00106

Lauri, P. É. (2019). Corner's Rules as a Framework for Plant Morphology, Architecture and Functioning - Issues and Steps Forward. *New Phytol.* 221, 1679–1684. doi:10.1111/nph.15503

Levionnois, S., Coste, S., Nicolini, E., Stahl, C., Morel, H., and Heuret, P. (2020). Scaling of Petiole Anatomies, Mechanics and Vasculatures with Leaf Size in the Widespread Neotropical Pioneer Tree Species Cecropia Obtusa Trécul (Urticaceae). *Tree Physiol.* 40, 245–258. doi:10.1093/treephys/tpz136

Li, H., Huang, Z., Gai, J., Wu, S., Zeng, Y., Li, Q., et al. (2007). A Conceptual Framework for Mapping Quantitative Trait Loci Regulating Ontogenetic Allometry. *PLoS One* 2, e1245. doi:10.1371/journal.pone.0001245

Liu, P.-C., Peacock, W. J., Wang, L., Furbank, R., Larkum, A., and Dennis, E. S. (2020). Leaf Growth in Early Development Is Key to Biomass Heterosis in Arabidopsis. *J. Exp. Bot.* 71, 2439–2450. doi:10.1093/jxb/eraa006

Lu, N., Zhang, M., Xiao, Y., Han, D., Liu, Y., Zhang, Y., et al. (2019). Construction of a High-Density Genetic Map and QTL Mapping of Leaf Traits and Plant Growth in an Interspecific F1 Population of Catalpa Bungei × Catalpa Duclouxii Dode. *BMC Plant Biol.* 19, 596. doi:10.1186/s12870-019-2207-y

Luo, A., Qian, Q., Yin, H., Liu, X., Yin, C., Lan, Y., et al. (2006). *EUI1*, Encoding a Putative Cytochrome P450 Monooxygenase, Regulates Internode Elongation by Modulating Gibberellin Responses in rice. *Plant Cel Physiol.* 47, 181–191. doi:10.1093/pcp/pci233

Machida, C., Nakagawa, A., Kojima, S., Takahashi, H., and Machida, Y. (2015). The Complex of ASYMMETRIC LEAVES (AS) Proteins Plays a central Role in Antagonistic Interactions of Genes for Leaf Polarity Specification in *Arabidopsis*. *Wires Dev. Biol.* 4, 655–671. doi:10.1002/wdev.196

Miles, C., and Wayne, M. (2008). *Quantitative Trait Locus (QTL) Analysis*. Cambridge: Nature Education, 1, 208.

Moreno, J. C., Tiller, N., Diez, M., Karcher, D., Tillich, M., Schöttler, M. A., et al. (2017). Generation and Characterization of a Collection of Knock-Down Lines for the Chloroplast Clp Protease Complex in Tobacco. *J. Exp. Bot.* 68, 2199–2218. doi:10.1093/jxb/erx066

Nibau, C., Gallemí, M., Dadarou, D., Doonan, J. H., and Cavallari, N. (2019). Thermo-Sensitive Alternative Splicing of *FLOWERING LOCUS M* Is Modulated by Cyclin-dependent Kinase G2. *Front. Plant Sci.* 10, 1680. doi:10.3389/fpls.2019.01680

Niinemets, Ü. (2007). Photosynthesis and Resource Distribution through Plant Canopies. *Plant Cel Environ.* 30, 1052–1071. doi:10.1111/j.1365-3040.2007.01683.x

Nocera, D. G. (2017). Solar Fuels and Solar Chemicals Industry. *Acc. Chem. Res.* 50, 616–619. doi:10.1021/acs.accounts.6b00615

Novaes, E., Osorio, L., Drost, D. R., Miles, B. L., Boaventura-Novaes, C. R. D., Benedict, C., et al. (2009). Quantitative Genetic Analysis of Biomass and wood Chemistry of Populus under Different Nitrogen Levels. *New Phytol.* 182, 878–890. doi:10.1111/j.1469-8137.2009.02785.x

Park, D. H., Lim, P. O., Kim, J. S., Cho, D. S., Hong, S. H., and Nam, H. G. (2003). The *Arabidopsis COG1* Gene Encodes a Dof Domain Transcription Factor and Negatively Regulates Phytochrome Signaling. *Plant J.* 34, 161–171. doi:10.1046/j.1365-313x.2003.01710.x

Polle, A., Chen, S. L., Eckert, C., and Harfouche, A. (2019). Engineering Drought Resistance in Forest Trees. *Front. Plant Sci.* 9, 1875. doi:10.3389/fpls.2018.01875

Poorter, H., Niklas, K. J., Reich, P. B., Oleksyn, J., Poot, P., and Mommer, L. (2012). Biomass Allocation to Leaves, Stems and Roots: Meta-analyses of Interspecific Variation and Environmental Control. *New Phytol.* 193, 30–50. doi:10.1111/j.1469-8137.2011.03952.x

Poorter, H., Niinemets, Ü., Ntagkas, N., Siebenkäs, A., Mäenpää, M., Matsubara, S., et al. (2019). A Meta-analysis of Plant Responses to Light Intensity for 70 Traits Ranging from Molecules to Whole Plant Performance. *New Phytol.* 223, 1073–1105. doi:10.1111/nph.15754

R Core Team (2018). *R : A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/.

Sang, M., Shi, H., Wei, K., Ye, M., Jiang, L., Sun, L., et al. (2019). A Dissection Model for Mapping Complex Traits. *Plant J.* 97, 1168–1182. doi:10.1111/tpj.14185

Semagn, K., Iqbal, M., Chen, H., Perez-Lara, E., Bemister, D. H., Xiang, R., et al. (2021). Physical Mapping of QTL Associated with Agronomic and End-Use Quality Traits in spring Wheat under Conventional and Organic Management Systems. *Theor. Appl. Genet.* 134, 3699–3719. doi:10.1007/s00122-021-03923-x

Spicer, R., and Groover, A. (2010). Evolution of Development of Vascular Cambia and Secondary Growth. *New Phytol.* 186, 577–592. doi:10.1111/j.1469-8137.2010.03236.x

Sun, X., Xu, D., Liu, Z., Kleine, T., and Leister, D. (2016). Functional Relationship between *mTERF4* and *GUN1* in Retrograde Signaling. *Exbotj* 67, 3909–3924. doi:10.1093/jxb/erv525

Wang, Z., and Wu, R. (2004). A Statistical Model for High-Resolution Mapping of Quantitative Trait Loci Determining HIV Dynamics. *Statist. Med.* 23, 3033–3051. doi:10.1002/sim.1870

Wang, P., Ma, L., Li, Y., Wang, S. a., Li, L., Yang, R., et al. (2016). Transcriptome Profiling of Indole-3-Butyric Acid-Induced Adventitious Root Formation in Softwood Cuttings of the Catalpa Bungei Variety 'YU-1' at Different Developmental Stages. *Genes Genom.* 38, 145–162. doi:10.1007/s13258-015-0352-8

Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., et al. (2015). The Molecular Mechanism of Sporocyteless/nozzle in Controlling *Arabidopsis* Ovule Development. *Cell Res.* 25, 121–134. doi:10.1038/cr.2014.145

Weraduwage, S. M., Chen, J., Anozie, F. C., Morales, A., Weise, S. E., and Sharkey, T. D. (2015). The Relationship between Leaf Area Growth and Biomass Accumulation in *Arabidopsis thaliana*. *Front. Plant Sci.* 6, 167. doi:10.3389/fpls.2015.00167

Withers, J., and Dong, X. (2017). Post-translational Regulation of Plant Immunity. *Curr. Opin. Plant Biol.* 38, 124–132. doi:10.1016/j.pbi.2017.05.004

Wu, R., Ma, C.-X., Littell, R. C., and Casella, G. (2002). A Statistical Model for the Genetic Origin of Allometric Scaling Laws in Biology. *J. Theor. Biol.* 219, 121–135. doi:10.1006/jtbi.2002.3114

Xiao, Y., Ma, W., Lu, N., Wang, Z., Wang, N., Zhai, W., et al. (2019). Genetic Variation of Growth Traits and Genotype-By-Environment Interactions in Clones of *Catalpa Bungei* and *Catalpa Fargesii F. Duclouxii*. *Forests* 10, 57. doi:10.3390/f10010057

Zhang, C., and Xi, N. (2021). Precipitation Changes Regulate Plant and Soil Microbial Biomass via Plasticity in Plant Biomass Allocation in Grasslands: A Meta-Analysis. *Front. Plant Sci.* 12, 614968. doi:10.3389/fpls.2021.614968

Zhang, M., Bo, W., Xu, F., Li, H., Ye, M., Jiang, L., et al. (2017). The Genetic Architecture of Shoot-Root Covariation during Seedling Emergence of a Desert Tree, *Populus Euphratica*. *Plant J.* 90, 918–928. doi:10.1111/tpj.13518

Zhang, M., Li, W., Feng, J., Gong, Z., Yao, Y., and Zheng, C. (2020a). Integrative Transcriptomics and Proteomics Analysis Constructs a New Molecular Model for Ovule Abortion in the Female-Sterile Line of *Pinus Tabuliformis Carr*. *Plant Sci.* 294, 110462. doi:10.1016/j.plantsci.2020.110462

Zhang, M., Zhang, S., Ye, M., Jiang, L., Vallejos, C. E., and Wu, R. (2020b). The Genetic Control of Leaf Allometry in the Common Bean, *Phaseolus vulgaris*. *BMC Genet.* 21, 29. doi:10.1186/s12863-020-00838-2

Zheng, X., Yang, X., Chen, Z., Xie, W., Yue, X., Zhu, H., et al. (2021). Arabidopsis SMAX1 Overaccumulation Suppresses Rosette Shoot Branching and Promotes Leaf and Petiole Elongation. *Biochem. Biophys. Res. Commun.* 553, 44–50. doi:10.1016/j.bbrc.2021.03.006

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.

Copyright © 2021 Zhang, Lu, Zhu, Yang, Qu, Shi, Fei, Liu, Ma and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fgene.2021.757920

Analysis of Wheat Wax Regulation Mechanism by Liposome and Transcriptome

Hongwei Wen[1†], Ying Wang[1†], Bangbang Wu[1], Yanru Feng[2], Yifei Dang[1], Bin Yang[1], Xiaofei Ma[1]* and Ling Qiao[1]**

[1]State Key Laboratory of Integrative Sustainable Dryland Agriculture, Institute of Wheat Research, Shanxi Agricultural University, Linfen, China, [2]Crop Science, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany

OPEN ACCESS

Edited by:
Jiang Libo,
Beijing Forestry University, China

Reviewed by:
Fei Tao,
Northwest University, China
Vandana Jaiswal,
Institute of Himalayan Bioresource Technology (CSIR), India

***Correspondence:**
Bin Yang
sxxmsyb83@126.com
Xiaofei Ma
nongxue06123@163.com
Ling Qiao
qiaolingsmile@163.com

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

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

Received: 13 August 2021
Accepted: 02 November 2021
Published: 06 December 2021

Citation:
Wen H, Wang Y, Wu B, Feng Y, Dang Y, Yang B, Ma X and Qiao L (2021) Analysis of Wheat Wax Regulation Mechanism by Liposome and Transcriptome. Front. Genet. 12:757920. doi: 10.3389/fgene.2021.757920

As a barrier for plants to contact with the outside world, epidermal wax plays an important role in resisting biotic and abiotic stresses. In this study, we analyzed the effect of wax content on leaf permeability by measuring the wax loss rate in the leaf. To further clarify the wax composition of the wheat epidermis and its molecular regulation mechanism, we applied untargeted lipidomic and transcriptome analysis on the leaf epidermis wax of Jimai 22 low-wax mutant (*waxless*) and multi-wax mutant (*waxy*). Our research showed that the mutant *waxy* has a slow loss rate, which can maintain higher leaf water content. 31 lipid subclasses and 1,367 lipid molecules were identified. By analyzing the wax differences of the two mutants, we found that the main lipid components of leaf epidermis wax in Jimai 22 were WE (C19-C50), DG (C27-C53), MG (C31-C35), and OAHFA (C31-C52). Carbon chain length analysis showed that, in wheat epidermis wax, WE was dominated by C44 molecules, DG was mainly concentrated in C47, C45, C37, and C31 molecules, C48 played a leading role in OAHFA, and C35 and C31 played a major role in MG. Among them, DG, MG, and OAHFA were detected in wheat leaf wax for the first time, and they were closely related to stress resistance. Compared with the *waxy*, 6,840 DEGs were detected in the mutant *waxless*, 3,181 DEGs were upregulated, and 3,659 DEGs were downregulated. The metabolic pattern of main waxy components in the wheat epidermis was constructed according to KEGG metabolic pathway and 46 related genes were screened, including *KSC*, *TER*, *FAR*, *WSD1*, *CER1*, *MAH1*, *ALDH7A1*, *CYP704B1*, *ACOT1_2_4*, *CYP86*, *MGLL*, *GPAT*, *ALDH*, *DPP1*, *dgkA*, *plsC,* and *E2.3.1.158* related genes. The screened wax-related genes were confirmed to be highly reliable by qRT-PCR. In addition, we found *TER* gene *TraesCS6B03G1132900LC* in wheat mutant *waxless* leaves for the first time, which inhibited the synthesis of long-chain acyl-CoA (n+2) by downregulating its expression. These results provide valuable reference information for further study of wheat epidermis wax heredity and molecular regulation.

Keywords: wheat, wax, liposome, transcriptome, molecular regulation

INTRODUCTION

As the first barrier for plants, epidermis wax plays an important role in resisting biotic and abiotic stresses (Aharoni et al., 2004; Franke et al., 2005; Bernard and Joubes, 2013). For example, it can limit the loss of non-stomatal water, improve the drought resistance of plants, help plants reduce mechanical damage, plant diseases, and insect pests, and protect plants from high temperature and strong ultraviolet radiation (Reina-Pinto and Yephremov, 2009; Yeats et al., 2012). Epidermal wax is a complex mixture of lipids composed of very-long-chain fatty acids (VLCFAs) and their derivatives (Zhang et al., 2005; Tafolla-Arellano et al., 2018). Existing studies have shown that the biosynthesis of epidermal wax begins with a waxy forerunner transformed by very-long-chain fatty acids C16 or C18 on the outer membrane of plastid epidermal cells. The carbon chains of C16 or C18 acyl-CoA and malonyl-CoA are lengthened by β-ketoacyl-CoA synthetase (KCS), β-ketoacyl-CoA reductase (KCR), β-hydroxyacyl-CoA dehydratase (HCD), and enoloyl-CoA reductase (ECR). Then, these very-long-chain fatty acids synthesize different waxy compounds by acyl reduction and decarbonylation (Samuels et al., 2008; Dong et al., 2019). Because of the complexity of waxy biosynthesis, the accurate determination of plant waxy components and content is helpful to infer the pathway of plant waxy biosynthesis accurately. However, most of the existing studies on epidermal wax components are based on gas chromatography-mass spectrometry (GC-MS), and some trace components cannot be detected, which leads to some limits in the study of the epidermal wax synthesis pathway.

Many genes involved in the wax synthesis and regulation have been found in *Arabidopsis* and rice (Racovita et al., 2016; Shaheenuzzamn et al., 2019). For instance, transcription factor *WIN1/SHN1* in *Arabidopsis* upregulates the expression of epidermal wax synthesis genes *CER1*, *CER2*, *CER4*, *KCS*, *CYP86A7*, *CYP86A4*, *GPAT4*, *LACS2,* and *HTH* to induce epidermal wax accumulation (Kannangara et al., 2007). The overexpression of *CER1* leads to the accumulation of alkanes (Bourdenx et al., 2011), and overexpression of *CER4* (*AtFAR3*) induces the production of primary alcohols in C24:0-C30:0 (Rowland et al., 2006). Zhou et al. (2014) found that *OsWR2* in rice, as a homologue of *AtWIN1/AtSHN1*TF, controls wax synthesis and accumulation by regulating the expression of very-long-chain fatty acid biosynthesis genes *CER6/CUT1*, *FDH2*, *FAE,* and *LACS1* in the panicle. Due to the huge genomic information of wheat, there are few studies on the molecular regulation mechanism of wheat epidermal wax biosynthesis. It has been reported that *W1-W5*, *Iw1*, *Iw2,* and *Iw3* are related to the wax synthesis of the wheat epidermis (Huang et al., 2017; Li et al., 2020). Li et al. (2021) studied the wax-deficient mutant *w5* and found that the blockage of β-diketone biosynthesis inhibited waxy synthesis. Chai et al. (2018) cloned several *TaFARs* genes encoding fatty acyl-CoA reductase from wheat.

In view of the complexity of wheat epidermis wax composition and synthesis mechanism, the analysis of wheat wax deletion mutants is considered to be a tool to get many response genes. To study the regulation mechanism of wheat epidermis wax, low-wax mutants and multi-wax mutants obtained by ethyl methanesulfonate (EMS) mutagenesis of Jimai 22 were used as materials in this study. By means of mutual verification and joint analysis of untargeted liposome and transcriptome, the main components of wheat epidermis wax were identified, and new genes related to epidermal wax metabolism were excavated, such as *TraesCS1D03G0373900*, *TraesCS1D03G0374000*, *TraesCS4B03G0019500*, *TraesCS7B03G1338900*, *TraesCS5B03G0557800*, *TraesCS5D03G0511400*, and *TraesCS7A03G0874000*, which improved the possible molecular regulation mechanism of wheat epidermis wax synthesis. The purpose of this study is to provide valuable reference information for further study on the genetic and molecular mechanism of epidermis wax metabolism in wheat, which provides theoretical support for wheat breeding and genetic improvement.

MATERIALS AND METHODS

Plant Materials

Jimai 22 is a high-yielding variety selected by the Crop Research Institute of Shandong Academy of Agricultural Sciences, containing wax in wild leaf epidermis. Combined with the experimental experience of our research group for many years, 0.6% (v/w) EMS was used to mutagenize Jimai 22 in this study, and then *waxless* and *waxy* were selected from the separated high-generation population. Compared with *waxy*, *waxless* showed less wax in the whole plant (**Figures 1A,B**). In October 2020, the mutants were planted in the Hancun Experimental Base of Wheat Research Institute of Shanxi Agricultural University (36°N, 111°E), with three replicates. The length was 2 m, the row spacing was 0.3 m, and 30 seeds per row were sown evenly. Wax content and transcriptome analysis were performed on flag leaves at the heading stage. Three biological repeats were performed for both *waxless* and *waxy*, and two samples with good repeatability were selected in transcriptome analysis.

Determination of Water Loss Rate of Leaves

In order to analyze the effect of leaf epidermis wax on the water loss rate of wheat leaves, in the same period as transcriptome sequencing, the flag leaves at the heading stage of *waxy* and *waxless* mutants were fully soaked and dehydrated in the dark at room temperature for 10 h (stomata were completely closed). The leaf water loss rate was calculated by 0.001 mg analysis balance (AUW320, Japan) every hour. Each sample was repeated three times.

Untargeted Lipidomics Detection

Liposome was detected by liquid chromatogramphy-tandem mass spectrometry (LC-MS). The leaves were immersed in chloroform for 30 s to dissolve the epidermis wax and dried with nitrogen. Lipids were extracted according to the MTBE method (Pizarro et al., 2013). Briefly, samples were spiked with a suitable amount of internal lipid standards and then homogenized with 200 μl water and 240 μl methanol. After that, 800 μl of MTBE was added and the mixture was



FIGURE 1 | Phenotype of low-wax mutant *waxless* and multi-wax mutant *waxy*. **(A)** The whole plant phenotype of *waxless* and *waxy*. **(B)** Leaf phenotype of *waxless* and *waxy*.

ultrasound-sonicated 20 min at 4°C followed by keeping still for 30 min at room temperature. The solution was centrifuged at 14,000 g for 15 min at 10°C and the upper organic solvent layer was obtained and dried under nitrogen. Reverse-phase chromatography was selected for LC separation using the CSH C18 column (1.7 μm, 2.1 × 100 mm, Waters). The lipid extracts were re-dissolved in 200 μl 90% isopropanol/acetonitrile, centrifuged at 14,000 g for 15 min; finally, 3 μl of the sample was injected. Solvent A was acetonitrile-water (6:4, v/v) with 0.1% formic acid and 0.1 Mm ammonium formate and solvent B was acetonitrile-isopropanol (1:9, v/v) with 0.1% formic acid and 0.1 Mm ammonium formate. The initial mobile phase was 30% solvent B at a flow rate of 300 μl/min. It was held for 2 min and then linearly increased to 100% solvent B in 23 min, followed by equilibrating at 5% solvent B for 10 min. Mass spectra were acquired by Q-Exactive Plus in positive and negative modes, respectively. ESI parameters were optimized and preset for all measurements as follows: source temperature was set at 300°C, capillary temp at 350°C, the ion spray voltage at 3000 V, S-Lens RF level at 50% and the scan range of the instruments at m/z 200–1800.

Extraction of RNA and Transcriptome Sequencing

The clean leaves of the two mutants were taken at the heading stage, wrapped in tin foil, frozen in liquid nitrogen quickly, and then frozen in the refrigerator at −80°C. The samples were sent to Beijing Baimaike Biotechnology Co., Ltd., for transcriptome sequencing. Total RNA was extracted using the Trizol method, and the concentration and purity of RNA were measured by NanoDrop 2000 (Thermo Fisher Science, Wilmington, DE). RNA integrity was evaluated using the RNA Nano 6000 analysis kit of Agilent Biological Analyzer 2,100 system (Agilent Technologies, CA, United States). A total amount of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext Ultra TM RNA Library Prep Kit for Illumina (NEB, United States) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. In order to select cDNA fragments of preferentially 240 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, United States). Then, 3 μl USER Enzyme (NEB, United States) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2,100 system.

Real-Time Quantitative PCR Analysis

Primers were designed by Primer Premier Version 5.0 (Premier Biosoft International, Palo Alto, CA) and combined by Sangon (www.sangon.com) (**Supplementary Table S1**). LA-Taq enzyme from TaKaRa (www.takara.com.cn) was used for PCR amplification. PCR was performed in total volumes of 15 μl, including 3 pmol of each primer, 120 μM of each dNTP, 80 ng template DNA or cDNA, 0.75 unit La-Taq, and 7.5 μl of 2 × buffer (TaKaRa Biotechnology (Dalian) Co., Ltd., Product Code:

DRR20AG). PCR was performed as follows: 95°C for 4 min, followed by 35 cycles of 95°C for 30 s, annealing (55–62°C) for 30 s, extension at 72°C (30 s–3 min), and 72°C for 30 s, with a final extension of 72°C for 10 min. Annealing temperatures and extension times depended upon individual primer sets and the length of expected PCR products.

Quantitative real-time PCR was performed using SYBR® Premix Ex Taq™ II (Takara) according to the manufacturer's instructions on a 7,300 Real-time PCR System (Applied Biosystems), where the relative expression of each gene was calculated according to the $2^{-\Delta\Delta CT}$ method (Zheng et al., 2020). The glyceraldehyde-3-phosphate dehydrogenase gene was used as an endogenous reference for real-time PCR, and all analyses were performed with three technical and three biological replications.

Data Analysis

"Lipid Search" is a search engine for the identification of lipid species based on MS/MS math. Lipid Search contains more than 30 lipid classes and more than 1,500,000 fragment ions in the database. Both mass tolerances for precursor and fragment were set to 5 ppm. Under the positive and negative ion modes, the OPLS-DA model was constructed by SAMIC14.1 software, and the prediction rate Q^2 of the model was obtained by 7-fold cross-validation. $Q^2 > 0.5$ was taken as the reliable standard of the model; the multiple of variation analysis (Fold Change Analysis, FC) > 4 or < 0.25 and the variable importance for the projection (VIP) > 1, p value < 0.05, were used as the screening criteria to compare the overall differential expression multiples of lipid ions in leaves of *waxless* and *waxy*. Lipid differences were analyzed with SPSS19.0 and Microsoft Excel 2019, plotted with Origin 2018. The transcriptome data were further processed by the online platform of BMK Cloud (www.Biocloud.net). FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) method was used to standardize the gene expression level. DESeq R package was used for differential analysis, and Fold Change ≥2 and FDR < 0.01 were used as screening criteria to determine the differentially expressed genes (DEGs) between *waxless* and *waxy* mutants.

RESULTS AND ANALYSIS

Effect of Epidermis Wax on Water Loss Rate of Wheat Leaves

The leaf water loss rate of the two mutants was faster in the darkroom temperature environment. The wax content had a significant effect on the leaf dehydration rate. After natural dehydration for 1 h, the water loss rate of low-wax mutant leaves reached 24.4% and that of multi-wax mutant leaves reached 17.4%. The leaves of the two mutants were dehydrated rapidly within 1–5 h, and the water loss rate of *waxless* was significantly higher than that of *waxy*. After dehydration for 5 h, the water loss rate of *waxless* leaves reached 63.5%, while that of *waxy* was only 44.3%. There was a significant difference between the two mutants. Although the dehydration rate of the two mutants slowed down after 5 h of dehydration, the water loss rate of low-wax mutants was still significantly higher than that of wax-rich mutants at 10 h of dehydration (**Figure 2**). This shows that the mutant *waxy* has a slow dehydration rate and strong water retention capacity and can maintain higher leaf water content, indicating that leaf wax content plays an important role in maintaining leaf water content.

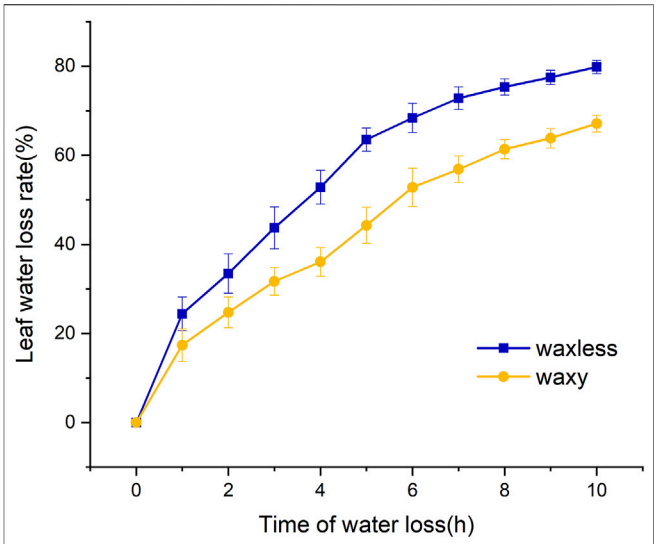


FIGURE 2 | Water loss rate of flag leaves detached from mutant *waxless* and mutant *waxy*.

Identification and Analysis of Lipid Components in Leaf Epidermis

Positive and negative ion patterns of electrospray ionization (ESI) were used in this study. By UPLC analysis, 31 lipid subclasses and 1,367 lipid molecules were identified in the leaf epidermis of the two mutants (**Figure 3A**). The lipids with a high number of lipid species are triacylglycerol (TG), ceramide (Cer), diacylglycerol (DG), wax ester (WE), (O-acyl)-1-hydroxy fatty acid (OAHFA), monogalactosyl diacylglycerol (MGDG), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), cholesterol (ChE), cardiolipin (CL), phosphatidylinositol (PI), and phosphatidylserine (PS). Among them, TG, Cer, DG, WE, and OAHFA account for 83% of the total number of molecules (**Figure 3B**). Moreover, it was found that the main lipid components of wheat leaf epidermis were TG, Cer, DG, WE, and OAHFA.

Analysis of Wax-Related Lipid Subclasses in Leaf Epidermis.

To test the repeatability of samples, we performed OPLS-DA analysis on all samples (**Figure 4B**). Each group of samples gathered closely and located in the middle of each group, indicating the liposome analysis showed high reproducibility. By comparing the contents of lipid subclasses between low-wax mutants and multi-wax mutants at the heading stage, significant

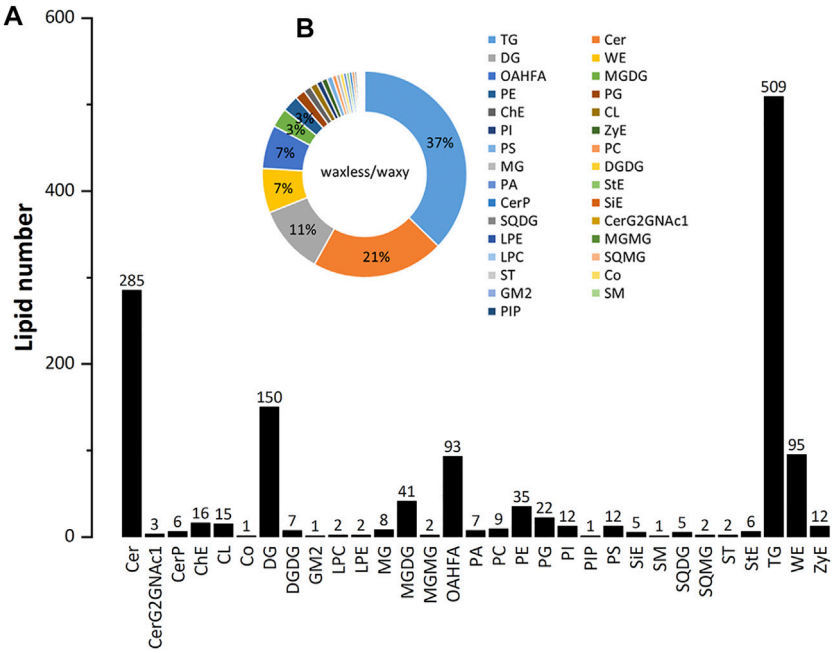


FIGURE 3 | The statistical chart of numbers of lipid subclass in wheat leaves. **(A)** Abscissa is the detected lipid subclasses; ordinate is the number of lipids. **(B)** Different lipid subclasses are represented by different colors, and the proportion is expressed by the size of the color block area. For the corresponding relationship and proportion between each lipid subclass and color, see the legend on the right.

differences in lipids were screened out. As shown in **Figure 4A**, 11 kinds of lipid substances with significant differences were found between the two mutants (VIP >1, $p < 0.05$). Among them, there are six lipid subclasses with high lipid content in *waxless*, namely, Cer, ZyE, PE, MGDG, CL, and digalactosyl diacylglycerol (DGDG). On the other hand, in *waxy*, five kinds of lipid classes were found to be DG, ChE, OAHFA, WE, and monoacylglycerol (MG) (VIP >1, $p < 0.05$). We infer that the more abundant and different lipids in the *waxy* are the main components of the wax in the epidermis of wheat compared with the *waxless*. That is the main lipid components of wheat leaf epidermis at the heading stage are DG, ChE, OAHFA, WE, and MG.

Analysis of Lipid Ions Differences in Leaf Epidermis

We analyzed the differences of 1,367 detected lipid ions (FC > 4 or <0.25, VIP >1, $p < 0.05$). Compared with the mutant *waxy*, the mutant *waxless* downregulated 177 lipid ions and upregulated 246 lipid ions. The results were shown in the form of a volcanic map (**Figure 4C**). 177 lipid ions related to wax metabolism in the epidermis of wheat leaves were screened (FC < 0.25, VIP >1, $p < 0.05$), belonging to eight lipid subclasses, including WE (C19-C50), DG (C27-C53), TG (C29-C73), MG (C31-C35), OAHFA (C31-C52), Cer (C29-C48), ChE ($C_{46}H_{86}O_2N_1$), and CL ($C_{65}H_{120}O_{17}P_2$) (**Supplementary Table S2**).

The main differential lipid ions in DG are DG(45:2)+H, DG(28:0e)+H and DG (37:4e)+H; ChE(19:0)+NH_4 in ChE; OAHFA(16:0_31:1)-H, OAHFA(30:1)-H, and OAHFA(31:0)-H in OAHFA; WE(24:0_14:1)+H, WE(28:1_22:0)+H, WE(28:1_18:0)+H, WE(26:0_21:3)+H, and WE(12:0_19:3)+NH_4 in WE; MG(32:1)+NH_4, MG(28:1)+H, and MG(32:1)+NH_4 in MG (**Table 1**). These may be the main lipid ions that affect the gloss phenotype of wheat leaf epidermis.

Analysis of Carbon Chain Length of Differential Lipids

In addition to the content of lipids, the carbon chain length of lipids is also an important factor that cannot be ignored. We added the lipid ions with the same carbon chain length and counted the lipid ions with diverse carbon chain lengths under different lipid subclasses to further analyze the main discrepant lipids in wheat epidermis wax (**Figure 5**). Compared with *waxy* mutants, the lipid ions significantly decreased in *waxless* mutants were WE (C29), WE (C31), WE (C44), WE (C46), DG (C47), DG (C45), DG (C43), DG (C37), DG (C36), DG (C31), OAHFA (C51), OAHFA (C50), OAHFA (C48), OAHFA (C31), MG (C35), MG (C31), and TG (C64).

These lipid ions affect the synthesis, transcription, and transport of wax through the change of carbon chain length, leading to the smooth green phenotype of wheat waxy deletion mutants. We found that WE in wheat leaf epidermis is mainly concentrated in C44-dominated wax ester molecules and also widely distributed in C46, C31, and C29 wax ester molecules but less in other chain lengths. DG in the wax mixture of the wheat epidermis is mainly concentrated in C47, C45, C37, and C31 molecules. In OAHFA, C48 plays a leading role, accompanied by

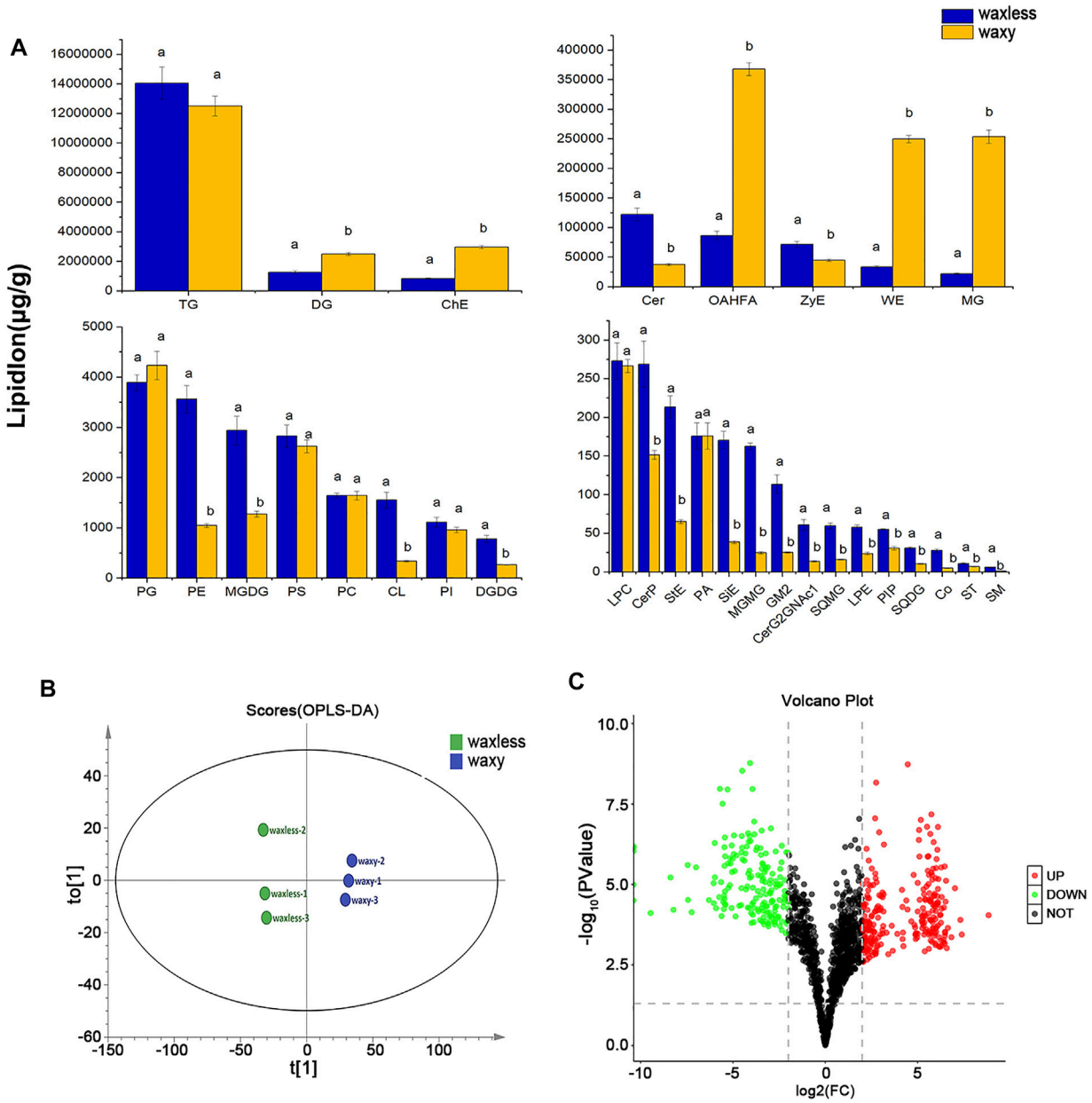


FIGURE 4 | Differential expression of wax-related lipids between *waxless* and *waxy*. **(A)** Content of lipid subclasses in the leaf epidermis of two wheat mutants, abscissa is the lipid subclass, ordinate is the lipid subclass content, blue is the low-wax mutant *waxless*, yellow is the multi-wax mutant *waxy*, and the letters indicate significant differences ($p < 0.05$). **(B)** OPLS-DA score map, "t[1]" represents principal component 1, "to[1]" represents principal component 2, and the ellipse represents 95% confidence interval. Dots of the same color represent each biological weight in the group *waxless* and *waxy*. **(C)** Volcano plot, abscissa is $\log_2$FC (*waxless*/*waxy*) and ordinate is $-\log_{10}$ (p value). Each dot represents a lipid ion, red represents upregulated lipid ions, green represents downregulated lipid ions, and black indicates undifferentiated lipid ions (FC > 4 or < 0.25, VIP>1, $p < 0.05$).

a large number of C50-C51 and C31 molecules. C35 and C31 play significant roles in MG. On the other hand, in TG, the length of the carbon chain varies from C29 to C64, in which C48 plays a major role.

Gene Differential Expression Analysis

Through the analysis of leaf transcriptional groups of low-wax mutants and multi-wax mutants at the heading stage, correlation coefficients among different biological repetitive samples ranged from 0.890 to 0.926 (**Figure 6A**). Compared with the mutant *waxy*, 6,840 DEGs were detected in the mutant *waxless*, of which 3,181 genes were upregulated and 3,659 genes were downregulated (Fold Change ≥2, FDR <0.01) (**Figure 6B**). To understand the biological significance between low-wax mutant and multi-wax mutant DEGs, these genes were enriched by Gene Ontology (GO) analysis. We found that the biological processes

TABLE 1 | Some significant differences in lipid ions between the two mutants.

Class	Lipid ion	Ion formula	Calmz	RT-(min)	VIP	FC(wxless/waxy)	*p* value
TG	TG (16:0e_11:3_16:0)+H	C46 H85 O5	717.639	11.087	1.039	0.0000	1.013E-05
TG	TG(12:1e_16:1_16:1)+H	C47 H87 O5	731.655	11.318	1.038	0.0000	7.176E-07
TG	TG(16:1e_12:2_17:1)+H	C48 H87 O5	743.655	13.602	1.039	0.0030	6.028E-06
TG	TG(16:0e_11:2_16:0)+H	C46 H87 O5	719.655	14.009	1.037	0.0033	6.086E-05
TG	TG(12:0e_16:0_18:4)+H	C49 H89 O5	757.670	16.761	1.037	0.0058	2.997E-05
DG	DG(45:2)+H	C48 H91 O5	747.686	15.278	1.039	0.0000	6.485E-07
DG	DG(28:0e)+H	C31 H63 O4	499.472	5.151	1.039	0.0000	9.089E-07
DG	DG(37:4e)+H	C40 H73 O4	617.550	15.722	1.038	0.0000	3.125E-05
DG	DG(44:3e)+H	C47 H89 O4	717.676	16.783	1.035	0.0014	7.660E-05
DG	DG(34:1e)+NH_4	C37 H76 O4 N1	598.577	6.326	1.037	0.0057	2.469E-06
Cer	Cer(d16:1_15:0)+H	C31 H62 O3 N1	496.472	6.224	1.037	0.0152	3.802E-05
Cer	Cer(m17:1_12:0)+H	C29 H58 O2 N1	452.446	8.725	1.038	0.0187	2.452E-05
Cer	Cer(d19:1_12:0)+H	C31 H62 O3 N1	496.472	8.162	1.036	0.0204	9.540E-05
Cer	Cer(d15:0_16:1)+H	C31 H62 O3 N1	496.472	5.816	1.037	0.0237	4.217E-05
Cer	Cer(m46:0)+H	C46 H94 O2 N1	692.728	15.058	1.039	0.0302	4.069E-06
ChE	ChE(19:0)+NH_4	C46 H86 O2 N1	684.665	15.942	1.036	0.0296	5.379E-05
CL	CL(56:2)-2H	C65 H120 O17 P2	617.401	7.945	1.037	0.2104	3.561E-05
OAHFA	OAHFA(16:0_31:1)-H	C47 H89 O4	717.677	16.781	1.036	0.0158	5.938E-06
OAHFA	OAHFA(30:1)-H	C30 H55 O4	479.411	10.509	1.037	0.0166	5.445E-05
OAHFA	OAHFA(31:0)-H	C31 H59 O4	495.442	6.295	1.039	0.0200	8.356E-06
OAHFA	OAHFA(15:0_16:0)-H	C31 H59 O4	495.442	8.630	1.038	0.0212	3.053E-08
OAHFA	OAHFA(31:1)-H	C31 H57 O4	493.426	21.881	1.037	0.0281	1.509E-06
MG	MG(32:1)+NH_4	C35 H72 O4 N1	570.546	6.394	1.038	0.0216	3.395E-05
MG	MG(28:1)+H	C31 H61 O4	497.456	6.322	1.038	0.0228	2.612E-06
MG	MG(32:1)+NH_4	C35 H72 O4 N1	570.546	8.707	1.035	0.0345	9.770E-05
WE	WE(24:0_14:1)+H	H75 C38 O2	563.576	15.814	1.039	0.0231	3.439E-07
WE	WE(28:1_22:0)+H	H99 C50 O2	731.764	21.557	1.037	0.0260	2.186E-06
WE	WE(28:1_18:0)+H	H91 C46 O2	675.701	19.485	1.039	0.0299	6.087E-07
WE	WE(26:0_21:3)+H	H89 C47 O2	685.686	16.346	1.039	0.0322	1.128E-06
WE	WE(12:0_19:3)+NH_4	H60 C31 O2 N1	478.462	6.817	1.037	0.0425	4.349E-05

Note: triacylglycerol (TG), diacylglycerol (DG), ceramide (Cer), cardiolipin (CL), (O-acyl)-1-hydroxy fatty acid (OAHFA), wax ester (WE), and monoacylglycerol (MG).

were mainly enriched in the metabolic process (28.89%), cellular process (24.73%), single-organism process (14.96%), biological regulation (8.35%), and response to stimulus (6.46%). These genes are mainly distributed in the membrane, membrane part, cell, component cell part, and organelle. The molecular functions mainly include binding and catalytic activity (**Supplementary Figure S1**).

Furthermore, we annotated the lipid function of 3,659 DEGs, screened 343 DEGs, and enriched them using the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (**Figures 6C,D**). It was found that these genes are mainly concentrated in glycerophospholipid metabolism, cutin, suberine and wax biosynthesis, glycerolipid metabolism, sphingolipid metabolism, terpenoid backbone biosynthesis, fatty acid degradation, and fatty acid elongation pathway, which are consistent with the main lipid components that control the glossy phenotype of wheat leaves in this study.

Joint Analysis of Waxy and Differentially Expressed Genes

Combined with the determined lipid components, we marked the annotated genes in the corresponding pathway, constructed the metabolic model of the main lipid components in wheat epidermis according to the KEGG metabolic pathway (**Figure 7**), and screened 46 related genes (**Table 2**). Moreover, compared with the multi-wax mutant *waxy*, the genes annotated to the synthesis of wax ester (WE) in the low-wax mutant *waxless* were 3-ketoacyl-CoA synthase genes (*KCS*), very-long-chain enoyl-CoA reductase genes (*TER*), alcohol-forming fatty acyl-CoA reductase genes (*FAR*), wax-ester synthase genes (*WSD1*), aldehyde decarbonylase genes (*CER1*), and midchain alkane hydroxylase genes (*MAH1*). The downregulation of *TER* inhibits the conversion of long-chain 3-oxoacyl-CoA to long-chain acyl-CoA (n + 2), the downregulation of *CER1* inhibits the transformation of upstream A long-chain aldehyde to A long-chain alkane, and the expression genes involved in *KCS*, *FAR*, *WSD1*, and *MAH1* were both upregulated and downregulated.

In the process of OAHFA synthesis, the upregulated expression of acyl-coenzyme A thioesterase 1/2/4 genes (*ACOT1_2_4*) promotes the synthesis of long-chain fatty acid, while the expression level of acylglycerol lipase genes (*MGLL*) was both upregulated and downregulated during the conversion of monoacyl-glycerol to long-chain fatty acid. Downregulation of aldehyde dehydrogenase family 7 member A1 genes (*ALDH7A1*) expression inhibited the transition between A long-chain aldehyde and long-chain fatty acid, while aldehyde dehydrogenase genes (*ALDH*) upregulated the transition between them. During the transformation of long-chain fatty acid into omega-hydroxy fatty acid, the expression of long-chain fatty acid omega-monooxygenase genes (*CYP704B1*) was

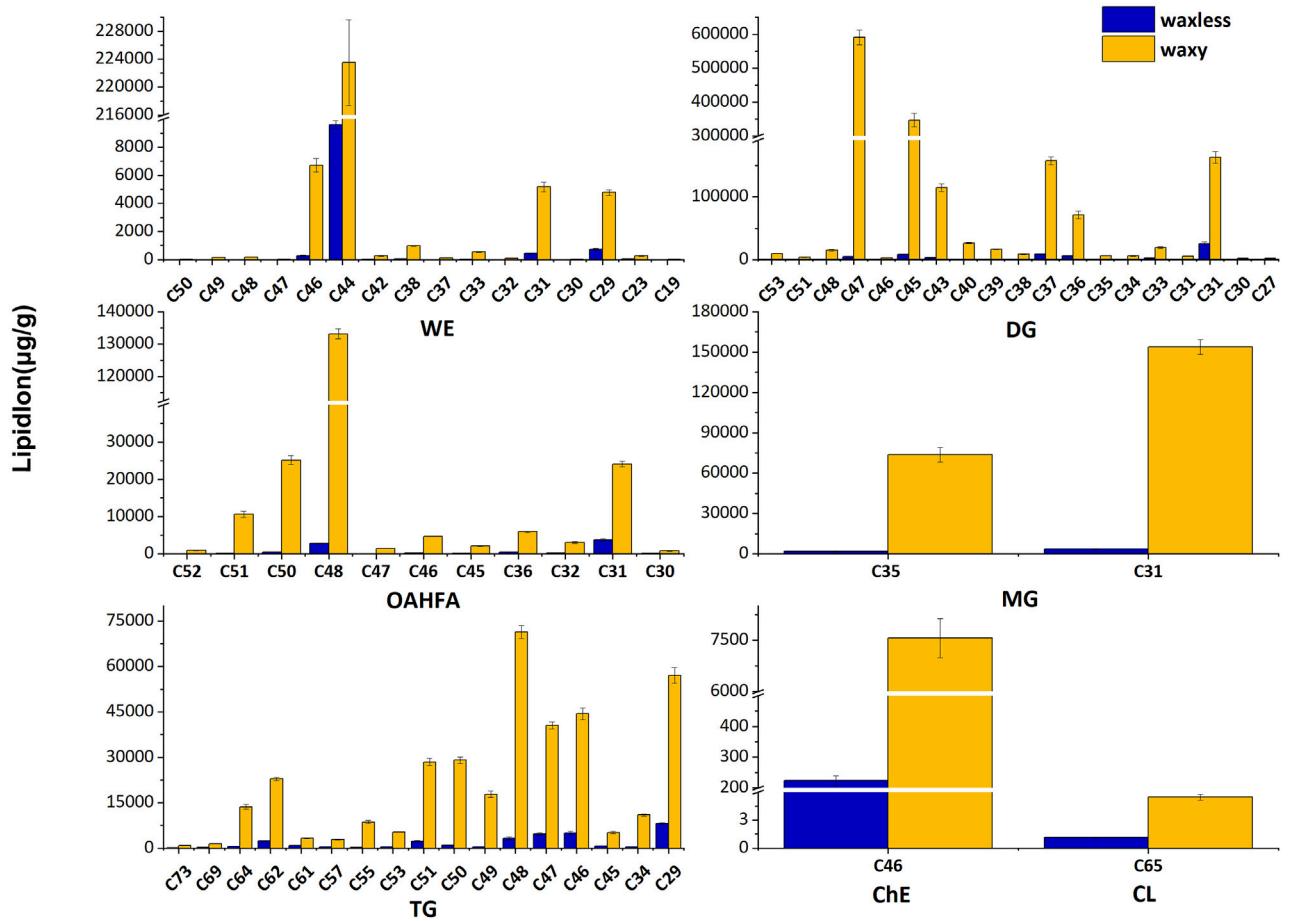


FIGURE 5 | Carbon chain length distribution. Abscissa denotes the lipid molecules with different carbon chain lengths, ordinate indicates lipid ion content, blue is the low-wax mutant *waxless*, and yellow is the multi-wax mutant *waxy*.

downregulated, while that of fatty acid omega-hydroxylase genes (*CYP86*) was upregulated in the mutant *waxless*. Glycerol-3-phosphate acyltransferase genes (*GPAT*) and *ALDH7A1* were downregulated in the process of glyceride metabolism of DG, MG, and TG. The downregulated expression of *GPAT* inhibits the synthesis of 1-acyl-sn glycerol 3-phosphate in *waxless*, which in turn reduces the synthesis of MG. The downregulated expression of *ALDH7A1* in the low-wax mutant decreases the synthesis of DG. The upregulated genes are *ALDH*, diacylglycerol diphosphate phosphatase genes (*DPP1*), and diacylglycerol kinase genes (*dgkA*), while the expressions of 1-acyl-sn-glycerol-3-phosphate acyltransferase genes (*plsC*), phospholipid: diacylglycerol acyltransferase genes (*E2.3.1.158*), *MGLL*, and alcohol dehydrogenase genes (*AKR1AI*) were both upregulated and downregulated. These genes work together to reduce the content of TG in the low-wax mutation *waxless*. In order to verify the reliability of our data for screening wax-related genes, we randomly selected 13 differentially expressed genes for real-time quantitative PCR detection, and the results of qRT-PCR of 13 genes were consistent with those of transcriptome data (**Supplementary Figure S2**), indicating that the screened wax-related genes are highly reliable, which is helpful for further functional verification and cloning of wax-related genes.

DISCUSSION

New Waxy Components of Wheat Epidermis Were Found by Untargeted Lipidomic Analysis

Epidermal wax is a complex lipid mixture composed of VLCFAs and their derivatives, which plays a special role in plant resistance to drought, diseases, and insect pests (Bernard and Joubes, 2013). At present, the GC-MS method is mainly used to determine the wax of epidermis (Rolim et al., 2015) and primary alcohols, secondary alcohols, aldehydes, alkanes, ketones, esters, triterpenes, sterols, and flavonoids (Zhang et al., 2005; Tafolla-Arellano et al., 2018). Based on the limitations of the determination method, there may still be many epidermal wax-related lipid components that have not been detected, which limits the study of the epidermal wax biosynthesis pathway. Lipidomics as a research model based on high-throughput analysis

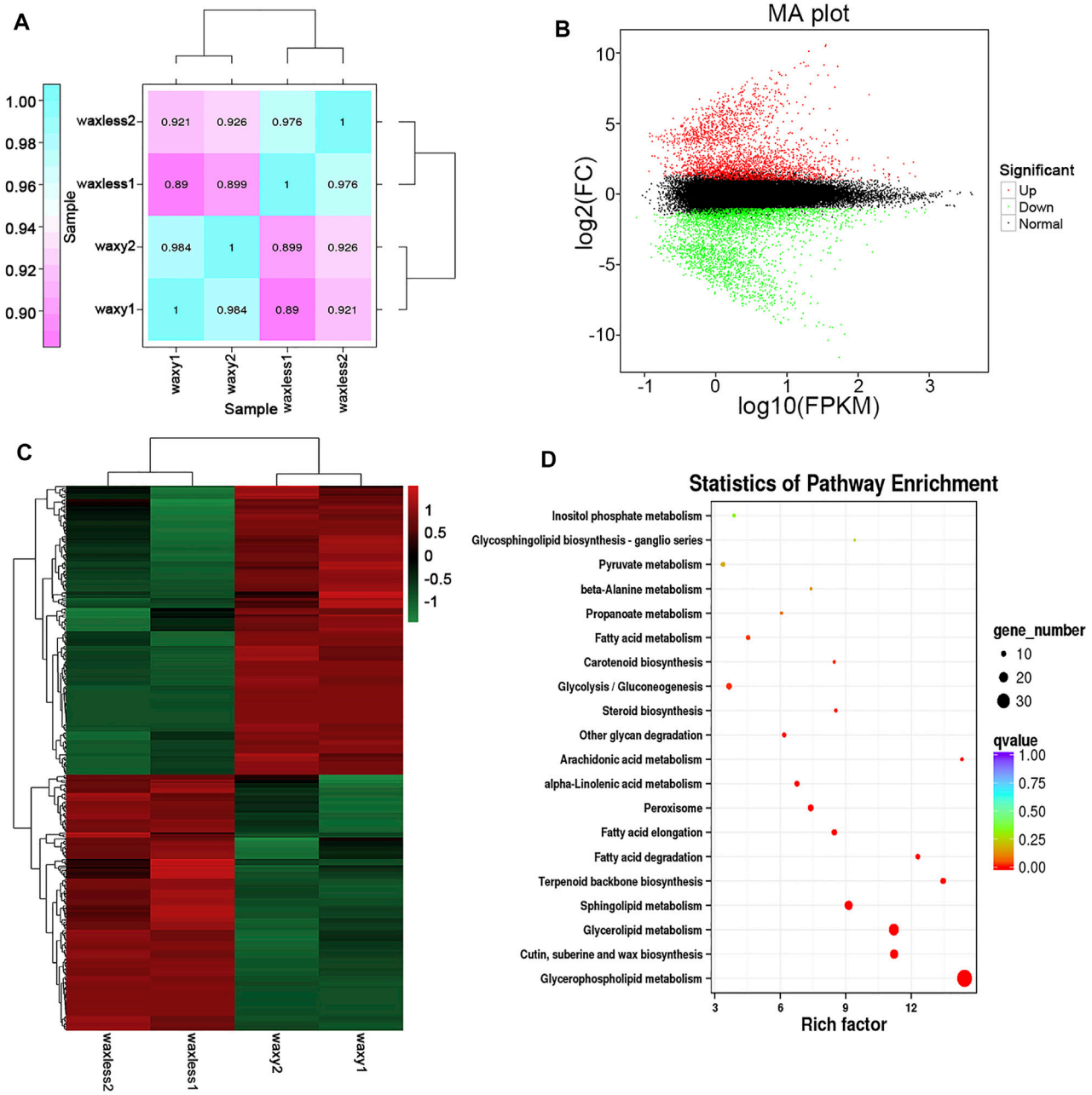


FIGURE 6 | Transcriptome data analysis. **(A)** The expression quantity correlation heatmap of the sample. The numbers in each cell represent the correlation coefficients. **(B)** MA plot of DEGs, each dot in the MA map of differentially expressed genes represents a gene. Abscissa is value A: $\log_{10}$ (FPKM), ordinate is value M: $\log_2$FC(*waxless*/*waxy*), green dots represent the downregulated DEGs, red dots represent the upregulated DEGs, and black dots represent the non-differentially expressed genes. **(C)** The heatmap presentation of fold changes of 343 DEGs obtained from RNA-seq analysis. **(D)** Enrichment and scatter map of KEGG pathway of DEGs. Each circle in the figure represents a KEGG pathway, ordinate represents the name of the pathway, and abscissa is the enrichment factor. The greater the enrichment factor, the more significant the enrichment level of DEGs in this pathway. The color of the circle represents the qvalue. The size of the circle indicates the number of genes enriched in the pathway, and the larger the circle, the more the genes.

can systematically analyze the changes in lipid composition and expression in organisms. Perez-Navarro et al. (2019) used LC-MS technology to determine grape lipids and found new information on the composition of free fatty acids such as glycerol, glycerol phospholipids, and triterpenes in grape skins and seeds. Broughton et al. (2018) used electrospray ionization tandem mass spectrometry (ESI-MS/MS) to analyze the acyl parts of wax esters (WE) and sterol esters (SE) in common and mutant sunflower oils with different fatty acid profiles and discovered the methylsterol components in sunflower oil sterol esters (SE) for the first time.

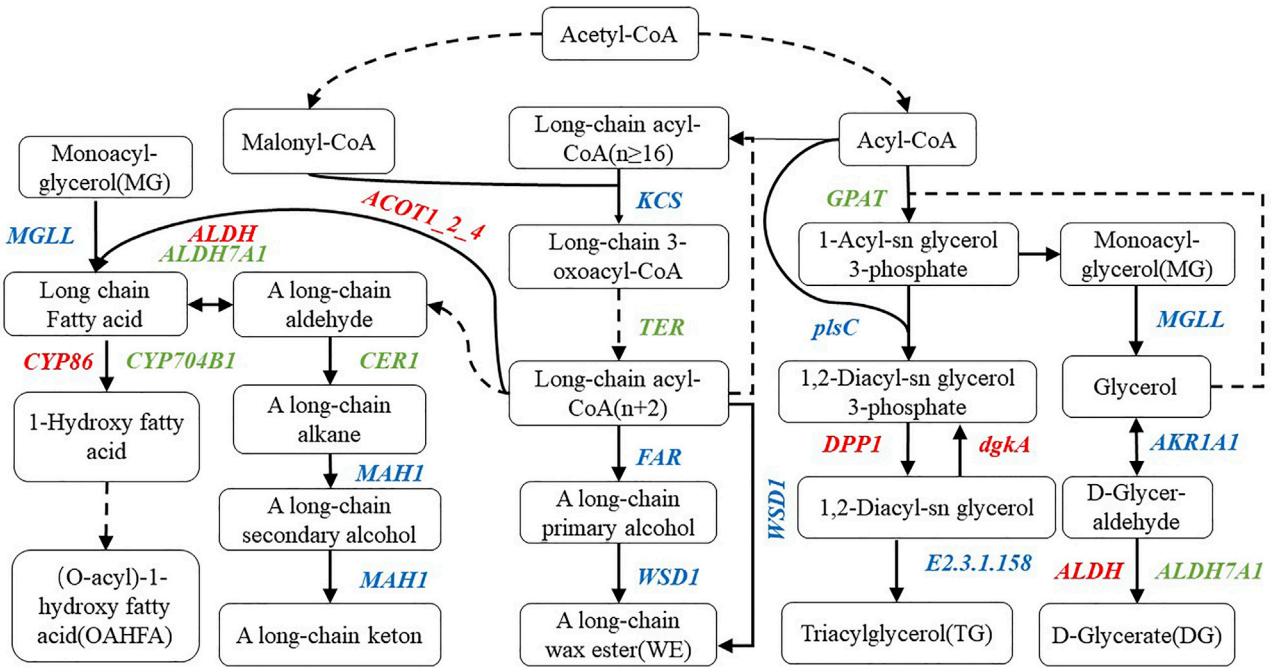


FIGURE 7 | Metabolic pathway map of wax-related lipid synthesis in wheat epidermis. There are metabolic compounds in the box, red-marked enzymes are related to upregulated genes, green-marked enzymes are related to downregulated genes, and blue-marked enzymes are related to both upregulated and downregulated genes.

Bianchi et al. (1980) studied the epidermal wax of common wheat in Chinese spring and found that its main components are n-alkanes, esters, aldehydes, free alcohols, β-diketones and hydroxy-β-diketones. Racovita et al. (2016) detected new waxy substances such as 2-alkyl alcohol, benzyl alcohol, phenylethyl alcohol, and hydroxyphenylethanol in wheat flag leaves and peduncles. Lavergne et al. (2018) analyzed the epidermis of wheat leaves and stems and found that the waxy components of the epidermis were alkanes (C20-C42), fatty acids (C7-C34), ketones (C9-C35), and primary alcohols (C22-C33). In this study, UHPLC-MS/MS analysis technique was used for the first time to detect the leaf epidermis wax of low-wax mutant *waxless* and multi-wax mutant *waxy* of Jimai 22. We found that the main lipid components of leaf epidermis wax of Jimai 22 were WE (C19-C50), DG (C27-C53), MG (C31-C35), OAHFA (C31-C52), TG (C29-C73), Cer (C29-C48), ChE ($C_{46}H_{86}O_2N_1$), and CL ($C_{65}H_{120}O_{17}P_2$). Among them, DG, MG and OAHFA were detected in wheat leaf wax for the first time, which improved the composition of wheat wax components. It is clear that the liposome analysis of wheat epidermis wax by LC-MS can more systematically study the changes and functions of related lipid subclasses and lipid molecules in the process of epidermal wax metabolism, which is helpful to improve the pathway and mechanism of epidermal wax synthesis.

New Components Such as Diacylglycerol, Monoacylglycerol, and (O-Acyl)-1-hydroxy Fatty Acid Are Closely Related to Stress Resistance

In this research, we found that the wax content was closely related to the water loss rate of leaves, which may be related to drought resistance (Zhang et al., 2013). Besides, DG, MG, and OAHFA were newly found in wheat epidermis wax. DG, as one of the main components of glycerol, plays a critical role in maintaining the stability of cell membrane (especially plasma membrane and chloroplast membrane) under high temperature and drought stress (Narayanan et al., 2016). It can maintain the fluidity of cell membrane by downregulating saturated DGs (4 and 10 double bonds) and u-regulating low unsaturated DGs (0–4 double bonds) (Navarro-Reig et al., 2019). MG, as a waxy substance, can form stable hydrated dispersions in water. Under salt stress, the ratio of MG unsaturated fatty acids to saturated fatty acids increased significantly in salt-tolerant varieties (Gogna et al., 2020). The MG in wheat epidermis wax found in this study is also mainly MG (32:1) + NH_4 and MG (28:1) + H with low saturation. OAHFA acts as a surfactant in human tear film lipids and plays a key role in stabilizing tear film (Schuett and Millar, 2013; Marshall et al., 2016). Interestingly, we found OAHFA in wheat leaf epidermis wax for the first time and C48 molecules play a leading role. Its effects on wheat stress resistance and related mechanism need to be further studied.

In addition, as an important oil substance, WE plays a significant role in plant resistance to drought because of its hydrolytic resistance (Ivarson et al., 2017; Wang et al., 2021). The WE wax found by predecessors in sunflower is mainly concentrated in C32–C48 molecules (Broughton et al., 2018). In our study, we found that the WE wax in the epidermis of wheat leaves is mainly distributed in the molecular range of C29–C46 chain length, and the wax ester molecule dominated

TABLE 2 | The differentially expressed genes related to epidermal wax synthesis pathway in wheat.

Gene ID	log2FC (*waxless*/*waxy*)	Developing	KEGG orthology	Gene symbol	Definition
TraesCS1D03G0373900	−3.361317375	Down	K15404	*CER1*	Aldehyde decarbonylase
TraesCS1D03G0374000	−2.240488466	Down	K15404	*CER1*	Aldehyde decarbonylase
TraesCS3B03G0017600	−2.077102639	Down	K15406	*WSD1*	Wax-ester synthase
TraesCS3B03G0037000	−3.159781364	Down	K15406	*WSD1*	Wax-ester synthase
TraesCS5D03G0740400	−1.28631545	Down	K15406	*WSD1*	Wax-ester synthase
TraesCS5A03G0782000	5.129181799	Up	K15406	*WSD1*	Wax-ester synthase
TraesCS5B03G0814100	2.492795173	Up	K15406	*WSD1*	Wax-ester synthase
TraesCS3B03G0228000	−1.137168478	Down	K20495	*CYP704B1*	Long-chain fatty acid omega-monooxygenase
TraesCS3B03G0938500	−1.18751546	Down	K20495	*CYP704B1*	Long-chain fatty acid omega-monooxygenase
TraesCS3D03G0167200	−2.160568355	Down	K20495	*CYP704B1*	Long-chain fatty acid omega-monooxygenase
TraesCSU03G0103800	1.359012742	Up	K15405	*MAH1*	Midchain alkane hydroxylase
TraesCS4A03G1162200	−7.768462261	Down	K15405	*MAH1*	Midchain alkane hydroxylase
TraesCS1D03G0041600	5.136041959	Up	K13356	*FAR*	Alcohol-forming fatty acyl-CoA reductase
TraesCS4B03G0019500	−1.972216449	Down	K13356	*FAR*	Alcohol-forming fatty acyl-CoA reductase
TraesCS7B03G1338900	−1.447125679	Down	K13356	*FAR*	Alcohol-forming fatty acyl-CoA reductase
TraesCS6B03G0695700	2.649123598	Up	K15402	*CYP86*	Fatty acid omega-hydroxylase
TraesCS2B03G1160500	−1.725507817	Down	K00655	*plsC*	1-acyl-sn-glycerol-3-phosphate acyltransferase
TraesCS4A03G1087000	4.199580543	Up	K00655	*plsC*	1-acyl-sn-glycerol-3-phosphate acyltransferase
TraesCS6A03G0010400	2.48094237	Up	K00655	*plsC*	1-acyl-sn-glycerol-3-phosphate acyltransferase
TraesCS4B03G0110200	1.172254409	Up	K01054	*MGLL*	Acylglycerol lipase
TraesCS4B03G0111500LC	−6.217652748	Down	K01054	*MGLL*	Acylglycerol lipase
TraesCS6B03G0004500	5.978405721	Up	K00128	*ALDH*	Aldehyde dehydrogenase (NAD+)
TraesCS5B03G0557800	−1.393721239	Down	K14085	*ALDH7A1*	Aldehyde dehydrogenase family 7 member A1
TraesCS5D03G0511400	−1.556106189	Down	K14085	*ALDH7A1*	Aldehyde dehydrogenase family 7 member A1
TraesCS6A03G0083200	−1.967010361	Down	K13508	*GPAT*	Glycerol-3-phosphate acyltransferase
TraesCS6B03G0120600	−1.672710452	Down	K13508	*GPAT*	Glycerol-3-phosphate acyltransferase
TraesCS4A03G0991200	1.398499657	Up	K00679	*E2.3.1.158*	Phospholipid:diacylglycerol acyltransferase
TraesCS6A03G0788900	1.856281232	Up	K00679	*E2.3.1.158*	Phospholipid:diacylglycerol acyltransferase
TraesCS7A03G0874000	−1.849596941	Down	K00679	*E2.3.1.158*	Phospholipid:diacylglycerol acyltransferase
TraesCS1A03G0712300	3.228966834	Up	K00002	*AKR1A1*	Alcohol dehydrogenase (NADP+)
TraesCS3B03G0102700	1.348166602	Up	K00002	*AKR1A1*	Alcohol dehydrogenase (NADP+)
TraesCS3B03G0987300	−1.189917126	Down	K00002	*AKR1A1*	Alcohol dehydrogenase (NADP+)
TraesCS7D03G0536500	−1.885493349	Down	K00002	*AKR1A1*	Alcohol dehydrogenase (NADP+)
TraesCS3D03G0571200	1.44849726	Up	K18693	*DPP1*	Diacylglycerol diphosphate phosphatase
TraesCS2A03G1233700	1.054654678	Up	K00901	*dgkA*	Diacylglycerol kinase (ATP)
TraesCS1A03G1066900	−6.470126296	Down	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS4B03G0018800LC	−3.017642109	Down	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS6A03G0053200	−2.314714201	Down	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS6D03G0655900	−2.157514317	Down	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS1B03G1176600	1.132972231	Up	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS6A03G0162400	1.030585291	Up	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS7A03G0089600LC	3.358707584	Up	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS7A03G0089800LC	3.35205506	Up	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS7D03G0081000	1.589752288	Up	K15397	*KCS*	3-ketoacyl-CoA synthase
TraesCS6B03G1132900LC	−4.230220979	Down	K10258	*TER*	Very-long-chain enoyl-CoA reductase
TraesCS2B03G1089600	1.12980861	Up	K01068	*ACOT1_2_4*	Acyl-coenzyme A thioesterase 1/2/4

by C44 plays a role, which is similar to the results of previous studies.

Molecular Regulation Mechanism of Epidermal Wax Metabolism in Wheat

Previous studies have shown that the *W1-W5* homologue of *Arabidopsis CER* protein in wheat increases the wax content of the epidermis by producing hydroxy-β-diketone and inducing the biosynthesis of β-diketone (Huang et al., 2017; Li et al., 2020). In this study, two *CER1* genes, *TraesCS1D03G0373900* and *TraesCS1D03G0374000*, were also detected in the low-wax mutants, which decreased the synthesis of A long-chain alkane by downregulating the expression of *CER1*. *TaFAR1*, *TaFAR2*, *TaFAR3*, *TaFAR4*, *TaFAR5,* and *AtCER4* are homologous genes of *CER4* in *Arabidopsis*. As a kind of alcohol-forming fatty acyl-CoA reductase, they induce the production of A long-chain primary alcohol and increase the wax content of wheat leaves (Wang W. et al., 2015; Wang et al., 2015b; Wang et al., 2015c; Wang et al., 2017). Besides, two *FAR* genes regulating primary alcohol synthesis, *TraesCS4B03G0019500,* and *TraesCS7B03G1338900*, were also found in the mutant *waxless*, which downregulated primary alcohol synthesis and promoted the smooth green appearance of the waxy wheat epidermis. In the

research of low-wax mutants of wheat, Li et al. (2021) found that the expression levels of *ACC1*, *LACS*, *KCS,* and *KCR* were downregulated, which reduced the synthesis of VLCFAs. In our research, four *KCS* genes were also found to downregulate long-chain 3-oxoacyl-CoA synthesis in *waxless*. In addition, we found *TER* gene *TraesCS6B03G1132900LC* in wheat mutant *waxless* leaves for the first time, which inhibited the synthesis of long-chain acyl-CoA (n + 2) by downregulating its expression. This provides important reference information for enriching the molecular regulation mechanism of wheat epidermis wax synthesis.

Moreover, our study also found that the downregulation of two *ALDH7A1* genes *TraesCS5B03G0557800* and *TraesCS5D03G0511400* inhibited the synthesis of long-chain fatty acid and DG, the downregulation of three *CYP704B1* genes *TraesCS3B03G0228000*, *TraesCS3B03G0938500,* and *TraesCS3D03G0167200* inhibited the synthesis of omega-hydroxy fatty acid, and the downregulation of *GPAT* genes *TraesCS6A03G0083200* and *TraesCS6B03G0120600* downregulated the expression of 1-acyl-sn glycerol 3-phosphate in DG, MG, and TG during glyceride metabolism, which inhibited the synthesis of 1-acyl-sn glycerol 3-phosphate in wax-free mutants. The downregulated expression of *plsC* gene *TraesCS2B03G1160500* reduces the synthesis of 1,2-diacyl-sn glycerol 3-phosphate. Downregulation of phospholipid: diacylglycerol acyltransferase genes *TraesCS7A03G0874000* expression inhibits the transformation from 1,2-diacyl-sn glycerol to TG. These genes are found for the first time to regulate waxy synthesis in the wheat epidermis. In-depth genomic comparison and functional identification of these genes are helpful in cloning wheat waxy functional genes and analyzing the regulatory mechanism of waxy metabolism. Our next research will focus on the functional verification of key genes to better explain the molecular regulation mechanism of wax in the leaf epidermis of wheat.

CONCLUSION

In this study, untargeted liposome detection and transcriptome analysis were carried out on the leaf epidermis wax of Jimai 22 low-wax mutant and multi-wax mutant. A total of 31 lipid subclasses and 1,367 lipid molecules were identified. The main lipid components of wheat leaf wax were identified as WE (C19-C50), DG (C27-C53), TG (C29-C73), MG (C31-C35), and OAHFA (C31-C52). DG, MG, and OAHFA were found in the epidermis wax of wheat leaf for the first time. Compared with the mutant *waxy*, a total of 6,840 DEGs were detected in the mutant *waxless*, of which 3,181 DEGs were upregulated and 3,659 DEGs were downregulated. According to KEGG metabolic pathway, the metabolic pattern of the main waxy components in the wheat epidermis was constructed and 46 related genes were screened, including *KSC*, *TER*, *FAR*, *WSD1*, *CER1*, *MAH1*, *ALDH7A1*, *CYP704B1*, *ACOT1_2_4*, *ALDH*, *CYP86*, MGLL, *GPAT*, *DPP1*, *dgkA*, *plsC,* and *E2.3.1.158* related genes. This provides valuable reference information for further study of wheat epidermis wax inheritance and molecular regulation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**. Mass spectrometry data has been uploaded to EBI Metabolights, accession number MTBLS3710. Sequencing data has been uploaded to NCBI SRA BioProject, accession number PRJNA777112.

AUTHOR CONTRIBUTIONS

BY, LQ, and XM designed the study and wrote the article. HW and YF carried out the experiments and analyzed data. YW, BW, and YD carried out field experiments for screening material. All authors contributed to the article and approved the final article for publishing.

FUNDING

This study was supported by the State Key Laboratory of Integrative Sustainable Dryland Agriculture (in preparation), Shanxi Agricultural University (No. 202105D121008-2-1), and Research Program Sponsored by the Science Research of Shanxi Academy of Agricultural Sciences (Nos. YGJPY 2008, YCX2020YQ47, and 201801D221314).

ACKNOWLEDGMENTS

We thank Mingyi Zhang for his mutated materials (Institute of Wheat Research, Shanxi Agricultural University). We thank Applied Protein Technology Company (Shanghai) and Beijing Baimaike Company for their help and valuable advice on liposome detection and transcriptome sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.757920/full#supplementary-material

REFERENCES

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., and Pereira, A. (2004). The SHINE Clade of AP2 Domain Transcription Factors Activates Wax Biosynthesis, Alters Cuticle Properties, and Confers Drought Tolerance when Overexpressed in Arabidopsis[W]. *Plant Cell* 16, 2463–2480. doi:10.1105/tpc.104.022897

Bernard, A., and Joubès, J. (2013). *Arabidopsis* Cuticular Waxes: Advances in Synthesis, export and Regulation. *Prog. Lipid Res.* 52, 110–129. doi:10.1016/j.plipres.2012.10.002

Bianchi, G., Lupotto, E., Borghi, B., and Corbellini, M. (1980). Cuticular Wax of Wheat. *Planta* 148, 328–331. doi:10.1007/bf00388119

Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., et al. (2011). Overexpression of Arabidopsis ECERIFERUM1 Promotes Wax Very-Long-Chain Alkane Biosynthesis and Influences Plant Response to Biotic and Abiotic Stresses. *Plant Physiol.* 156, 29–45. doi:10.1104/pp.111.172320

Broughton, R., Ruíz-Lopez, N., Hassall, K. L., Martínez-Force, E., Garcés, R., Salas, J. J., et al. (2018). New Insights in the Composition of Wax and Sterol Esters in Common and Mutant sunflower Oils Revealed by ESI-MS/MS. *Food Chem.* 269, 70–79. doi:10.1016/j.foodchem.2018.06.135

Chai, G., Li, C., Xu, F., Li, Y., Shi, X., Wang, Y., et al. (2018). Three Endoplasmic Reticulum-Associated Fatty Acyl-Coenzyme A Reductases Were Involved in the Production of Primary Alcohols in Hexaploid Wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 18, 41. doi:10.1186/s12870-018-1256-y

Dong, X., Ji, J., Yang, L., Fang, Z., Zhuang, M., Zhang, Y., et al. (2019). Fine-mapping and Transcriptome Analysis of *BoGL-3*, a Wax-Less Gene in Cabbage (*Brassica oleracea* L. Var. Capitata). *Mol. Genet. Genomics* 294, 1231–1239. doi:10.1007/s00438-019-01577-5

Franke, R., Briesen, I., Wojciechowski, T., Faust, A., Yephremov, A., Nawrath, C., et al. (2005). Apoplastic Polyesters in *Arabidopsis* Surface Tissues - A Typical Suberin and a Particular Cutin. *Phytochemistry* 66, 2643–2658. doi:10.1016/j.phytochem.2005.09.027

Gogna, M., Choudhary, A., Mishra, G., Kapoor, R., and Bhatla, S. C. (2020). Changes in Lipid Composition in Response to Salt Stress and its Possible Interaction with Intracellular Na+-K+ Ratio in sunflower (Helianthus Annuus L.). *Environ. Exp. Bot.* 178, 104147. doi:10.1016/j.envexpbot.2020.104147

Huang, D., Feurtado, J. A., Smith, M. A., Flatman, L. K., Koh, C., and Cutler, A. J. (2017). Long Noncoding miRNA Gene Represses Wheat β-diketone Waxes. *Proc. Natl. Acad. Sci. USA* 114, E3149–E3158. doi:10.1073/pnas.1617483114

Ivarson, E., Iven, T., Sturtevant, D., Ahlman, A., Cai, Y., Chapman, K., et al. (2017). Production of Wax Esters in the Wild Oil Species *Lepidium Campestre*. *Ind. Crops Prod.* 108, 535–542. doi:10.1016/j.indcrop.2017.07.002

Kannangara, R., Branigan, C., Liu, Y., Penfield, T., Rao, V., Mouille, G., et al. (2007). The Transcription Factor *WIN1/SHN1* Regulates Cutin Biosynthesis in *Arabidopsis thaliana*. *Plant cell* 19, 1278–1294. doi:10.1105/tpc.106.047076

Lavergne, F., Broeckling, C., Cockrell, D., Haley, S., Peairs, F., Jahn, C., et al. (2018). GC-MS Metabolomics to Evaluate the Composition of Plant Cuticular Waxes for Four *Triticum aestivum* Cultivars. *Ijms* 19 (2), 249. doi:10.3390/ijms19020249

Li, L., Qi, Z., Chai, L., Chen, Z., Wang, T., Zhang, M., et al. (2020). The Semidominant Mutation *W5* Impairs Epicuticular Wax Deposition in Common Wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 133, 1213–1225. doi:10.1007/s00122-020-03543-x

Li, L., Zhang, Z., Song, W., Su, Z., Zhang, Y., You, M., et al. (2021). The Essential Role of *W5* in Wax Metabolism in Wheat (*Triticum aestivum* L.). *J. Plant Biol.* doi:10.1007/s12374-021-09325-2

Marshall, D. L., Saville, J. T., Maccarone, A. T., Ailuri, R., Kelso, M. J., Mitchell, T. W., et al. (2016). Determination of Ester Position in Isomeric (O -Acyl)-Hydroxy Fatty Acids by Ion Trap Mass Spectrometry. *Rapid Commun. Mass. Spectrom.* 30, 2351–2359. doi:10.1002/rcm.7715

Narayanan, S., Tamura, P. J., Roth, M. R., Prasad, P. V. V., and Welti, R. (2016). Wheat Leaf Lipids during Heat Stress: I. High Day and Night Temperatures Result in Major Lipid Alterations. *Plant Cel Environ.* 39, 787–803. doi:10.1111/pce.12649

Navarro-Reig, M., Tauler, R., Iriondo-Frias, G., and Jaumot, J. (2019). Untargeted Lipidomic Evaluation of Hydric and Heat Stresses on rice Growth. *J. Chromatogr. B* 1104, 148–156. doi:10.1016/j.jchromb.2018.11.018

Pérez-Navarro, J., Da Ros, A., Masuero, D., Izquierdo-Canas, P. M., Hermosín-Gutiérrez, I., Gomez-Alonso, S., et al. (2019). LC-MS/MS Analysis of Free Fatty Acid Composition and Other Lipids in Skins and Seeds of *Vitis vinifera* Grape Cultivars. *Food Res. Int.* 125, 108556. doi:10.1016/j.foodres.2019.108556

Pizarro, C., Arenzana-Rámila, I., Pérez-del-Notario, N., Pérez-Matute, P., and González-Sáiz, J.-M. (2013). Plasma Lipidomic Profiling Method Based on Ultrasound Extraction and Liquid Chromatography Mass Spectrometry. *Anal. Chem.* 85, 12085–12092. doi:10.1021/ac403181c

Racovita, R. C., Hen-Avivi, S., Fernandez-Moreno, J.-P., Granell, A., Aharoni, A., and Jetter, R. (2016). Composition of Cuticular Waxes Coating Flag Leaf Blades and Peduncles of *Triticum aestivum* Cv. Bethlehem. *Phytochemistry* 130, 182–192. doi:10.1016/j.phytochem.2016.05.003

Reina-Pinto, J. J., and Yephremov, A. (2009). Surface Lipids and Plant Defenses. *Plant Physiol. Biochem.* 47, 540–549. doi:10.1016/j.plaphy.2009.01.004

Rolim, A. E. H., Henrique-Araujo, R., Ferraz, E. G., de Araujo Alves Dultra, F. K., and Fernandez, L. G. (2015). Lipidomics in the Study of Lipid Metabolism: Current Perspectives in the Omic Sciences. *Gene* 554, 131–139. doi:10.1016/j.gene.2014.10.039

Rowland, O., Zheng, H., Hepworth, S. R., Lam, P., Jetter, R., and Kunst, L. (2006). *CER4* Encodes an Alcohol-Forming Fatty Acyl-Coenzyme A Reductase Involved in Cuticular Wax Production in Arabidopsis. *Plant Physiol.* 142, 866–877. doi:10.1104/pp.106.086785

Samuels, L., Kunst, L., and Jetter, R. (2008). Sealing Plant Surfaces: Cuticular Wax Formation by Epidermal Cells. *Annu. Rev. Plant Biol.* 59, 683–707. doi:10.1146/annurev.arplant.59.103006.093219

Schuett, B. S., and Millar, T. J. (2013). An Investigation of the Likely Role of (O-Acyl) ω-hydroxy Fatty Acids in Meibomian Lipid Films Using (O-Oleyl) ω-hydroxy Palmitic Acid as a Model. *Exp. Eye Res.* 115, 57–64. doi:10.1016/j.exer.2013.06.016

Shaheenuzzamn, M., Liu, T. X., Shi, S. D., Wu, H. Q., and Wang, Z. H. (2019). Research Advances on Cuticular Waxes Biosynthesis in Crops: a Review. *IJAB* 21, 911–921. doi:10.17957/ijab/15.0975

Tafolla-Arellano, J. C., Baez-Sanudo, R., and Tiznado-Hernández, M. E. (2018). The Cuticle as a Key Factor in the Quality of Horticultural Crops. *Scientia Horticulturae* 232, 145–152. doi:10.1016/j.scienta.2018.01.005

Wang, M., Wu, H., Xu, J., Li, C., Wang, Y., and Wang, Z. (2017). Five Fatty Acyl-Coenzyme A Reductases Are Involved in the Biosynthesis of Primary Alcohols in *Aegilops Tauschii* Leaves. *Front. Plant Sci.* 8, 14. doi:10.3389/fpls.2017.01012

Wang, W., Zhang, Y., Xu, C., Ren, J., Liu, X., Black, K., et al. (2015a). Cucumber *ECERIFERUM1* (*CsCER1*), Which Influences the Cuticle Properties and Drought Tolerance of Cucumber, Plays a Key Role in VLC Alkanes Biosynthesis. *Plant Mol. Biol.* 87, 219–233. doi:10.1007/s11103-014-0271-0

Wang, Y., Mao, H., Lv, Y., Chen, G., and Jiang, Y. (2021). Comparative Analysis of Total Wax Content, Chemical Composition and crystal Morphology of Cuticular Wax in Korla Pear under Different Relative Humidity of Storage. *Food Chem.* 339, 128097. doi:10.1016/j.foodchem.2020.128097

Wang, Y., Wang, J., Chai, G., Li, C., Hu, Y., Chen, X., et al. (2015b). Developmental Changes in Composition and Morphology of Cuticular Waxes on Leaves and Spikes of Glossy and Glaucous Wheat (*Triticum aestivum* L.). *Plos One* 10, e0141239. doi:10.1371/journal.pone.0141239

Wang, Y., Wang, M., Sun, Y., Wang, Y., Li, T., Chai, G., et al. (2015c). FAR5, a Fatty Acyl-Coenzyme A Reductase, Is Involved in Primary Alcohol Biosynthesis of the Leaf Blade Cuticular Wax in Wheat (*Triticum aestivum* L.). *J. Exp. Bot.* 66, 1165–1178. doi:10.1093/jxb/eru457

Yeats, T. H., Buda, G. J., Wang, Z., Chehanovsky, N., Moyle, L. C., Jetter, R., et al. (2012). The Fruit Cuticles of Wild Tomato Species Exhibit Architectural and Chemical Diversity, Providing a New Model for Studying the Evolution of Cuticle Function. *Plant J.* 69, 655–666. doi:10.1111/j.1365-313X.2011.04820.x

Zhang, J.-Y., Broeckling, C. D., Blancaflor, E. B., Sledge, M. K., Sumner, L. W., and Wang, Z.-Y. (2005). Overexpression of *WXP1*, a Putative *Medicago Truncatula* AP2 Domain-Containing Transcription Factor Gene, Increases Cuticular Wax Accumulation and Enhances Drought Tolerance in Transgenic Alfalfa (*Medicago Sativa*). *Plant J.* 42, 689–707. doi:10.1111/j.1365-313X.2005.02405.x

Zhang, Z., Wang, W., and Li, W. (2013). Genetic Interactions Underlying the Biosynthesis and Inhibition of β-Diketones in Wheat and Their Impact on Glaucousness and Cuticle Permeability. *Plos One* 8, e54129. doi:10.1371/journal.pone.0054129

Zheng, X., Liu, C., Qiao, L., Zhao, J., Han, R., Wang, X., et al. (2020). The MYB Transcription Factor *TaPHR3-A1* Is Involved in Phosphate Signaling and Governs Yield-Related Traits in Bread Wheat. *J. Exp. Bot.* 71, 5808–5822. doi:10.1093/jxb/eraa355

Zhou, X., Jenks, M. A., Liu, J., Liu, A., Zhang, X., Xiang, J., et al. (2014). Overexpression of Transcription Factor *OsWR2* Regulates Wax and Cutin

Biosynthesis in rice and Enhances its Tolerance to Water Deficit. *Plant Mol. Biol. Rep.* 32, 719–731. doi:10.1007/s11105-013-0687-8

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.

Copyright © 2021 Wen, Wang, Wu, Feng, Dang, Yang, Ma and Qiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 December 2021
doi: 10.3389/fgene.2021.779557

Integration of Transcriptome and Methylome Analyses Provides Insight Into the Pathway of Floral Scent Biosynthesis in *Prunus mume*

Xi Yuan, Kaifeng Ma, Man Zhang, Jia Wang and Qixiang Zhang**

Beijing Key Laboratory of Ornamental Plants Germplasm Innovation and Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, Engineering Research Center of Landscape Environment of Ministry of Education, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, School of Landscape Architecture, Beijing Forestry University, Beijing, China

OPEN ACCESS

Edited by:
Shang-Qian Xie,
Hainan University, China

Reviewed by:
Xiao Han,
Zhejiang Agriculture and Forestry University, China
Liangsheng Wang,
Institute of Botany (CAS), China

***Correspondence:**
Kaifeng Ma
makaifeng@bjfu.edu.cn
Qixiang Zhang
zqxbjfu@126.com

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

Received: *19 September 2021*
Accepted: *19 November 2021*
Published: *15 December 2021*

Citation:
Yuan X, Ma K, Zhang M, Wang J and Zhang Q (2021) Integration of Transcriptome and Methylome Analyses Provides Insight Into the Pathway of Floral Scent Biosynthesis in Prunus mume. *Front. Genet. 12:779557. doi: 10.3389/fgene.2021.779557*

DNA methylation is a common epigenetic modification involved in regulating many biological processes. However, the epigenetic mechanisms involved in the formation of floral scent have rarely been reported within a famous traditional ornamental plant *Prunus mume* emitting pleasant fragrance in China. By combining whole-genome bisulfite sequencing and RNA-seq, we determined the global change in DNA methylation and expression levels of genes involved in the biosynthesis of floral scent in four different flowering stages of *P. mume*. During flowering, the methylation status in the "CHH" sequence context (with H representing A, T, or C) in the promoter regions of genes showed the most significant change. Enrichment analysis showed that the differentially methylated genes (DMGs) were widely involved in eight pathways known to be related to floral scent biosynthesis. As the key biosynthesis pathway of the dominant volatile fragrance of *P. mume*, the phenylpropane biosynthesis pathway contained the most differentially expressed genes (DEGs) and DMGs. We detected 97 DMGs participated in the most biosynthetic steps of the phenylpropane biosynthesis pathway. Furthermore, among the previously identified genes encoding key enzymes in the biosynthesis of the floral scent of *P. mume*, 47 candidate genes showed an expression pattern matching the release of floral fragrances and 22 of them were differentially methylated during flowering. Some of these DMGs may or have already been proven to play an important role in biosynthesis of the key floral scent components of *P. mume*, such as *PmCFAT1a/1c*, *PmBEAT36/37*, *PmPAL2*, *PmPAAS3*, *PmBAR8/9/10*, and *PmCNL1/3/5/6/14/17/20*. In conclusion, our results for the first time revealed that DNA methylation is widely involved in the biosynthesis of floral scent and may play critical roles in regulating the floral scent biosynthesis of *P. mume*. This study provided insights into floral scent metabolism for molecular breeding.

Keywords: floral scent, phenylpropane biosynthesis pathway, Prunus mume, cytosine methylation, transcriptome

INTRODUCTION

Floral fragrance is one of the most important ornamental traits of horticultural plants (Schiestl, 2010). More than 1700 floral scent substances have been identified, which are mainly biosynthesized through the terpenoid, phenylpropane, and fatty acid biosynthesis pathways. For example, *Cananga odorata* (Jin et al., 2015), *Magnolia champaca* (Dhandapani et al., 2017), and *Syringa oblata* (Zheng et al., 2015) have been found to have terpenoids as their main volatiles; *Petunia hybrida* and Damask rose (Karami et al., 2015) have phenylpropanes; and *Hedychium coronarium* (Yue et al., 2015) and *Rosa hybrida* have terpenoids and phenylpropanes. The diverse combinations of floral volatiles emitted confer the unique floral fragrances among different aromatic plants (Knudsen et al., 2006; Muhlemann et al., 2014). Genes related to the biosynthesis of substances conferring floral aromas have been comprehensively studied in model plants (Tzin and Galili, 2010), such as *Clarkia breweri*, *Antirrhinum majus*, *Rosa ssp.*, and *Petunia hybrida* (Boatright et al., 2004; Fraser and Chapple, 2011; Dudareva et al., 2013; Muhlemann et al., 2014; Widhalm and Dudareva, 2015). Some important genes for the biosynthesis of phenylpropanes have been confirmed in petunia (Boatright et al., 2004; Moerkercke et al., 2009).

Prunus mume (mei) from Rosaceae, a traditional ornamental tree, was domesticated in China more than 3000 years ago (Chen, 1996). Numerous ornamental cultivars with colorful flowers, various types of branches and flower, variety of flowering periods, and characteristic floral scents have been bred (Sun et al., 2013; Xu et al., 2014; Zhang et al., 2018; Zhuo et al., 2021). As the only species within the *Prunus* genus to produce a strong floral scent (Zhang et al., 2019b), it is important to study the molecular mechanism behind the biosynthesis in *P. mume*'s floral scent. In previous studies, benzyl acetate and eugenol were revealed to be the main components of the headspace volatiles of the unique fragrances of different *P. mume* cultivars; additionally, benzyl alcohol and cinnamyl acetate were found to be the main components of the floral volatiles of the *P. mume* cultivar 'Fenhong Zhusha' ('FZ') (Zhang et al., 2019a; Bao et al., 2020). All of these volatiles were shown to be biosynthesized by the phenylpropane biosynthesis pathway (Hao R. et al., 2014; Hao R.-J. et al., 2014; Zhang et al., 2019a). Consistent with the pattern of emission of most floral scents (Fenske et al., 2015), the emission and biosynthesis of benzyl acetate and eugenol peak when *P. mume* is in full bloom (Hao R.-J. et al., 2014; Bao et al., 2019; Zhang et al., 2019b). The intracellular levels of metabolites of floral scent components in flower buds have also been identified in five hybrids of *P. mume* (Bao et al., 2020). The encoding genes, enzymatic activities, and functions of *PmBEATs* and *PmBARs*, which are closely related to the biosynthesis of benzyl acetate, have been determined within the *P. mume* genome (Zhao et al., 2017; Bao et al., 2019; Bao et al., 2020). The gene families *PmEGSs* and *PmCFATs*, which encode key functional enzymes involved in eugenol biosynthesis, have also been identified and characterized (Zhang et al., 2019b). However the epigenetic mechanisms involved in the biosynthesis of its floral scent have rarely been reported.

DNA methylation, one of the best-studied epigenetic modifications, plays a key role in genome stability, developmental regulation, gene expression regulation, transposon silencing, and environmental adaptation (Cubas et al., 1999; Zhu, 2009; Law and Jacobsen, 2010). In plants, DNA methylation mainly occurs on cytosine, including within segments with the sequence "CG," "CHG," or "CHH" (with H representing A, T, or C) (Finnegan and Kovac, 2000). The dynamic changes of DNA methylation level (ML) have been shown to regulate the expression of genes and ultimately lead to different phenotypes. The methylation status varies among different species, cultivars of the same species, and even tissues of the same plant (Vaughn et al., 2007; Suzuki and Bird, 2008). Evidence has shown that DNA methylation is involved in regulating the expression of key genes during the ripening of tomato fruit (Zhong et al., 2013); drought resistance (Xu et al., 2018) and flowering (Xing et al., 2019) of apple; peel color formation of apple (Bai et al., 2016; Li et al., 2019), red pear (Wang et al., 2013), and sweet orange (Huang et al., 2019); and the flower color formation of *P. mume* chimera (Jiang et al., 2020). However, the role it plays in the process of floral scent biosynthesis has not yet been reported.

Here, by using transcriptome sequencing (RNA-seq) and whole-genome bisulfite sequencing (WGBS), we discuss the methylation modification associated with floral scent biosynthesis in *P. mume* cv. 'FZ'. The results provide new genome-wide evidence that deepens our understanding of the epigenetic regulatory mechanism underlying the floral scent trait.

MATERIALS AND METHODS

Plant Materials

The present study focused on 'FZ', the earliest flowering *P. mume* cultivar (Yuan et al., 2020) with a relatively long flowering time and pleasant smell, which was planted in Jiufeng International Mei Garden (40°03′53″N, 116°05′49″E, 132 m a.s.l.), Haidian, Beijing. To reflect the bud developmental state of the whole plant, all buds on one to two twigs of each branch from three different sides of each experimental tree were collected as one biological replicate. Three biological replicates (1, 2, and 3) constituted one sample. Samples were frozen in liquid nitrogen immediately after being obtained and stored at −80°C until sequencing. We selected four samples in total, and took the S2/S4/S6/S8 bud development stage as the dominant stage, from January 9th to March 13th in 2019 (**Figure 1**). The dominant stages were determined by the following reasons: some of the key floral components of *P.mume* began to evaporate/emit in the early stages of flowering and peak when *P. mume* is in full bloom (S8). Slightly earlier than that, the biosynthesis of eugenol can be extracted before blooming (S6) (Zhang et al., 2019b) and the expression of key enzyme-encoding genes closely related to floral scent biosynthesis (like *PmCFATs* and *PmBEATs*) in the budding phase (S4 or S5) (Bao et al., 2019; Zhang et al., 2019b). The sample (S2) on January 9th 2019 was also selected for sequencing, for considering that the buds of 'FZ' were



FIGURE 1 | Definition of different flowering stages of *P. mume*: bud scale tight packed (S1), bud swelling (S2), sepal tips appeared (S3), sepals clearly visible (S4), petals appeared (S5), petals clearly visible (S6), beginning of blossom (S7), full blooming (S8), and petal drop (S9).

already finished endo-dormancy on December 23th 2018, and the methylation modification may be earlier than the onset of gene expression (Yuan et al., 2020).

Whole-Genome Bisulfite Sequencing and Analysis

Genomic DNA of all four sample groups was extracted using Plant Genomic DNA Kit DP305 (Tiangen Biotech, Beijing, China) and then used for library construction using Illumina's standard DNA methylation analysis protocol and the Accel-NGS®Methyl-Seq DNA Library Kit. The sequencing process was performed on the Illumina Hiseq 6000 platform to generate raw reads, which were preprocessed with Trimmomatic v0.36 software and FastQC analysis (Langmead and Salzberg, 2012). Finally, the clean reads were obtained for subsequent analysis.

RNA Isolation and Sequencing

Total RNA was extracted with the DP422 extraction kit (Tiangen Biotech). RNA quality and quantity were determined using 1% agarose gels and a NanoPhotometer spectrophotometer (Implen, Calabasas, CA, United States), respectively. An RNA Nano 6000 Assay kit of the Bioanalyzer 2100 system (Agilent, Carpinteria, CA, United States) was used to assess RNA integrity. A total of 3 μg of RNA was used as the input material for sequencing library construction. The NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, United States) was used for reverse transcription. The libraries were also sequenced on the Illumina Hiseq 6000 platform to generate 150 bp paired-end reads (clean bases > 9G).

Sequence Mapping

Using the HISAT2 (https://daehwankimlab.github.io/hisat2/) and Bismark (https://github.com/FelixKrueger/Bismark.git) software to aligned the clean reads in transcriptome and methylome (Kim et al., 2015; Krueger and Andrews, 2011; Langmead and Salzberg, 2012; Wang et al., 2014) to the reference genome of *P. mume* (http://prunusmumegenome.bjfu.edu.cn/) (Zhang et al., 2012), respectivly. The sequencing depth and coverage were estimated according to duplicates that aligned to a unique genomic region (Krueger and Andrews, 2011; Wang et al., 2014). The non-conversion rate (r) of bisulfite treatment was defined as the rate of the number of sequenced cytosines at all cytosine reference positions divided by the number in the lambda genome.

Methylated Cytosine Site Analysis and Detection of Differentially Methylated Regions

mC sites were detected using Bismark (Krueger and Andrews, 2011) and defined using a binomial test, with thresholds of sequence depth ≥5 and q-value ≤ 0.01 (Lister et al., 2009; Gifford et al., 2013; Habibi et al., 2013). ML of C site was calculated as follows: ML = mC/(mC + umC) (C here indicate one single C site). The "mCG," "mCHG," and "mCHH" contexts (methylcytosine occurring at CG, CHG, and CHH regions, respectively) and their densities, as well as their distributions in each chromosome, were analyzed using previous methods (Chan et al., 2005; Krzywinski et al., 2009; Lister et al., 2009; Zhong et al., 2013). Between samples, the global ML and distributions were also tested (Song et al., 2013). DMRs and differentially methylated loci were identified using DSS software (Feng et al., 2014; Wu et al., 2015; Park and Wu, 2016). Among chomosomes, every 200bp were grouped in one bin. In different functional region of genes, each region is evenly divided into 50 bins. In genesbody and its up- and downstream, each region is evenly divided into 20 bins. The average ML (ML_{avg}) and methylation density (MD) of each bin was calculated as follows: MD = mCX/CX (coverage > 5X, C here indicate all C site within each bin, CX means CG, CHG or CHH) (Wang et al., 2015).

$$\mathbf{ML_{avg}} = \frac{\sum_{i=1}^{n} ML_i}{n}$$

Differentially Methylated Gene (DMG) Detection and Differentially Expressed Gene (DEG) Analysis

Differentially methylated genes were defined as genes with DMRs within 2 kb of their promoter or gene body (from TSS to TES) region. To clarify the possible regulatory effects of methylation in the natural flowering process of *P. mume*, the differentially expressed genes (DEGs) between two samples were determined using the following criteria: false discovery rate (FDR) < 0.05 and |fold change| ≥0.

Functional Annotation of DEGs and DMR-Related Genes

The GOseq R soft package was used for Gene Ontology (GO) enrichment analysis of DMR-related genes that were aligned with GO (p-value < 0.05) (Young et al., 2010). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was used to explore pathways of DMR-associated genes (Kanehisa et al., 2008). We also used KOBAS software to test DMR-related genes' statistical enrichment in the KEGG pathways (Mao et al., 2005).

Methylation in Key Pathways in the Biosynthesis of Floral Fragrance

To show the involvement of methylation in the biosynthesis of *P. mume*'s flower scent, DMGs enriched in eight pathways that are related to floral scent biosynthesis were listed. We then summarized and drew the key pathway of floral scent biosynthesis, phenylpropanoid biosynthesis, and marked the processes involving DMGs with red asterisks. For further information about the methylation involved in regulating the expression of floral scent-related genes, the FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) values of 145 summarized genes in previous reaserch that encoding the key enzymes in the biosynthesis of the dominant floral volatile of *P. mume* were used to make heat maps using the heatmap option and the normalized FPKM values of genes (row) in TBtools software. The DMGs were marked and classified according to region (promoter, gene body), context (CG, CHG, CHH), and ML change (hyper, hypo) properties on heatmap.

Quantitative Real-Time-PCR Validation

The qRT-PCR was performed on the CFX96 TouchTM Real-Time PCR Detection System (Bio-Rad, Hercules, California, United States) using the following parameters: 95°C for 5 min, 30 cycles of 95°C for 5 s, and 60°C for 30 s, concluding with a melting-curve stage for 10 s at 95°C, 5s at 65°C, and 5 s at 95°C. Each reaction consisted of 0.8 μL 1st strand cDNA, 10 μL SYBR Premix Ex Taq (Takara, Dalian, Japan), 0.8 μL each of 10 mM primer pairs, and 7.6 μL H_2O. Each sample was assessed in three biological replicates for each sample and normalized using PmPP2A as an internal control. The transcription levels were determined using the $2^{-\triangle\triangle Ct}$ method. The correlation coefficients between RNA-seq and qRT-PCR were calculated with the CORREL function in an Excel spreadsheet. The specific primers were designed online by Primer 3 plus (http://www.primer3plus.com/cgi-bin/dev/primer3plus.cgi) and are listed in **Supplementary Table S9**.

RESULTS

DNA Methylation Landscape in *Prunus mume* cv. 'FZ'

To investigate the modification of the biosynthesis of floral scent by DNA methylation during the natural flowering process of mei, we performed WGBS of *P. mume* buds at four different flowering stages: S2, S4, S6, and S8 (hereafter referred to as Fm1–Fm4) (**Figure 1**). Each stage was sequenced with three biological replicates. For each sample, at least 32 million clean reads were produced (**Supplementary Table S2**). Approximately 60% of the clean reads were mapped to the reference genome using Bismark (Krueger and Andrews, 2011) (**Supplementary Table S2**). The average genome coverage was 21.67 × (**Supplementary Table S3**). All of our sequenced methylated regions had ~21 × average coverage per chromosome and ~10 × average coverage per cytosine site, covering an average of 83.8% of the genome (**Supplementary Figure S1**; **Supplementary Table S4**).

Among the 82,368,087 C (cytosine) sites in the *P. mume* genome, the CHH context was the most common and the CG context was the least (**Figure 2A**). In each sample, ~8% (7.32–8.46%) of the total C sites were methylated (**Table 1**). The mC sites mainly featured the mCG context (42.44–46.94%), with the mCHG (28.40–30.19%) and mCHH contexts (22.86–29.14%) comprising the rest (**Figure 2B**). Among the three contexts of C sites, 33.90%–35.72% of CG regions, 15.60%–16.97% of CHG regions, and 2.21–3.26% of CHH regions were methylated (**Table 1**). Meanwhile, it was shown that chromosome Pm3 was highly methylated (**Supplementary Figure S2**). In terms of ML, most mCG and mCHG regions had higher ML (80–100%), while most mCHH regions showed lower ML (10–40%) (**Figure 2C**). The analysis of methylation within chromosomes showed that both high ML_{avg} and MD areas were enriched in pericentromeric regions, where the gene density is low (**Supplementary Figure S3**). The results are consistent with findings from the petals of *P. mume* cv. 'FT' (Jiang et al., 2020), apples, and *Arabidopsis*.

Methylation Level Distribution During Flowering of *Prunus mume* Buds

The average values of ML in the mCG, mCHG, and mCHH contexts were 38.13% (36.29–39.32%), 17.25% (16.47–17.96%), and 1.85% (1.69–2.15%), respectively, while the average global ML in flower buds of *P. mume* was 8.35% during the flowering process (**Table 2**). Among different gene functional regions, all CX contexts had the highest ML in the TE region (**Figure 3A**). Among the gene body and 2 kb regions up- and downstream of it, the CG context had higher ML in the gene body, while the CHG and CHH contexts had lower ML in the gene body than in other regions (**Figure 3B**).

In the search for DMR-related genes, we detected 5492 (3461 hyper- and hypomethylated), 4744 (2796 hyper- and 1948 hypomethylated), and 7677 (6320 hyper- and 1357 hypomethylated) DMRs within three methylome comparisons between F2v1, F3v2, and F4v3, respectively. During the flowering process, the number of hyper-DMRs was greater than that of hypo-DMRs (**Figure 4A**). As shown in this **Figure 4B**, the number of DMRs was mainly from the CHH context, which occurs in promoter regions, and the number of hyper-DMRs was larger than that of hypo-DMRs. The result is consistent with the gradual increase of ML in promoter regions in the CHH context during flowering (**Figures 3A,B**). In contrast, the number of hypo-DMRs in the final stage of flowering was mainly contributed to by hypomethylated CG-DMRs in promoter regions (**Figure 4B**).

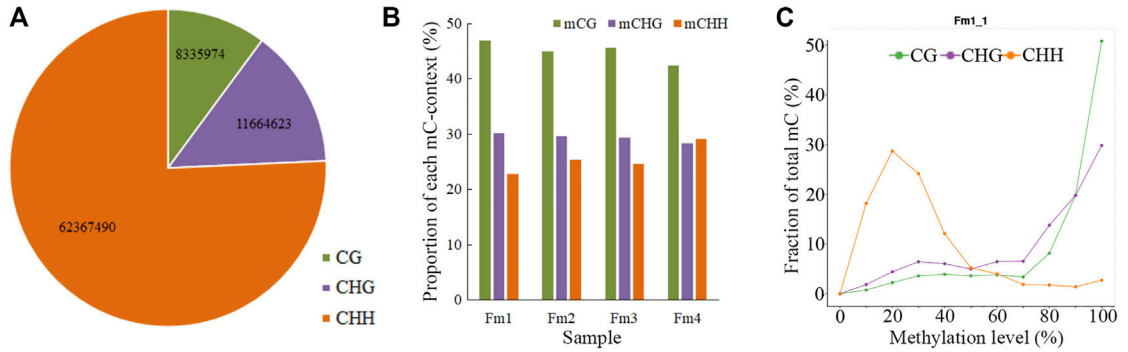


FIGURE 2 | Proportions of the three different contexts (CG, CHG, CHH) in **(A)** total cytosine **(C)** sites, **(B)** total mC (methylated cytosine) sites, and (C) ML distribution. The mC distribution diagrams of all samples are shown in the Supporting Information (Figure S4).

TABLE 1 | Proportion of mC in the three different contexts (CG, CHG, CHH) of four different flowering stages of *Prunus mume* 'Fenhong Zhusha' whole genome.

Samples	mC	mC/C	mCG	mCG/CG	mCHG	mCHG/CHG	mCHH	mCHH/CHH
Fm1	6032102	7.32%	2826502	33.90%	1820358	15.60%	1385241	2.21%
Fm2	6628481	8.04%	2978181	35.72%	1960418	16.80%	1689881	2.70%
Fm3	6226425	7.55%	2841509	34.08%	1852006	15.87%	1532909	2.45%
Fm4	6977067	8.46%	2958821	35.49%	1980533	16.97%	2037713	3.26%

TABLE 2 | Global mean ML of mC in the three different contexts (CG, CHG, CHH) of different flowering stages of *Prunus mume* 'Fenhong Zhusha'.

Samples	Mean mC(%)	Mean mCG(%)	Mean mCHG(%)	Mean mCHH(%)
Fm1	8.35	38.51	17.19	1.72
Fm2	7.93	36.29	16.47	1.69
Fm3	8.38	38.42	17.39	1.83
Fm4	8.75	39.32	17.96	2.15

DMGs Participate in the Pathway of Floral Scent Biosynthesis

Next, we detected 4005 (promoter, 3178; gene body, 1006), 3515 (promoter, 2765; gene body, 881), and 5306 (promoter, 4123; gene body, 1533) DMGs, which were associated with DMRs, in comparisons of Fm2v1, Fm3v2 and Fm4v3 ("m" here indicate methylome). In brief, the results showed that the genes related to CHH context DMR within promoter region constituted the largest proportion of DMGs among all methylome comparisons (**Figure 5**), which is consistent with previous results (**Supplementary Table S5**). We revealed the functions of those DMGs by KEGG enrichment analysis. Remarkably, some crucial DMGs were enriched in eight KEGG pathways related to floral scent biosynthesis. Overall, 97 (66 promoter, 44 gene body) and 17 DMGs were enriched in phenylpropanoid biosynthesis (ko pmum00940) and its upstream pathway phenylalanine, tyrosine, and tryptophan biosynthesis (ko pmum00400), respectively. Moreover, 20, 11, 14, 4 and 45 DMGs were enriched in the pathways of terpenoid backbone biosynthesis (ko pmum00900), monoterpenoid biosynthesis (ko pmum00902), diterpenoid biosynthesis (ko pmum00904), sesquiterpenoid and triterpenoid biosynthesis (ko pmum00909), and phenylalanine metabolism (ko pmum00360). Furthermore, 371 DMGs were enriched in the biosynthesis of secondary metabolites (ko pmum01110) (**Supplementary Table S5**).

Correlation Between DNA Methylation and Gene Expression Levels

Using the same samples of mei buds, we also performed RNA-seq to investigate the DEGs involved in floral scent biosynthesis during the natural flowering process. For each sample, at least 50 million (average 62 million, 9.32G over 30× coverage) clean reads were produced (**Supplementary Table S6**). Over 91% of the clean reads were mapped to the reference genome using Bismark (**Supplementary Table S6**) (Krueger and Andrews, 2011).

At the chromosome level, low ML and high gene expression levels were distributed in the same region on the chromosome. Higher gene density was also observed in this region (**Figure 6A**). All genes were divided into four groups according to their expression levels from low to high, and their average ML was determined in different gene regions (gene body and 2 kb up- and downstream of it, TSS, and TES). Among all methylated genes (MGs) at each flowering

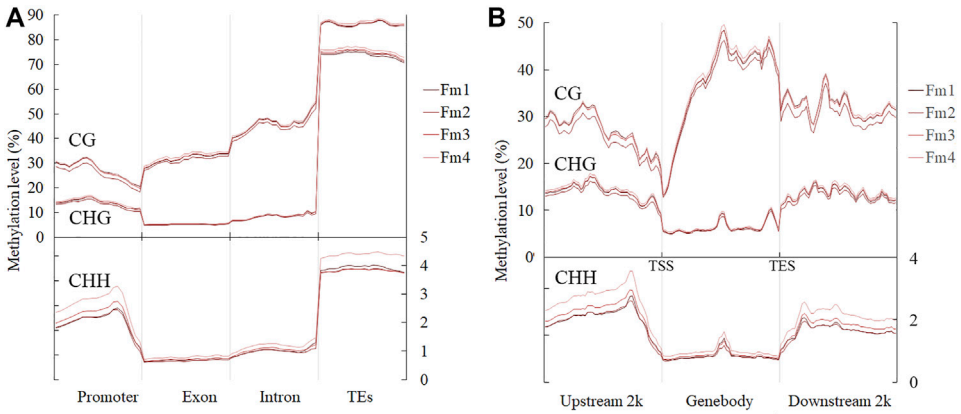


FIGURE 3 | ML in **(A)** different gene functional regions and **(B)** gene body and regions 2 kb up- and downstream of it, during flowering. TSS and TES stand for transcription start and end sites, respectively.

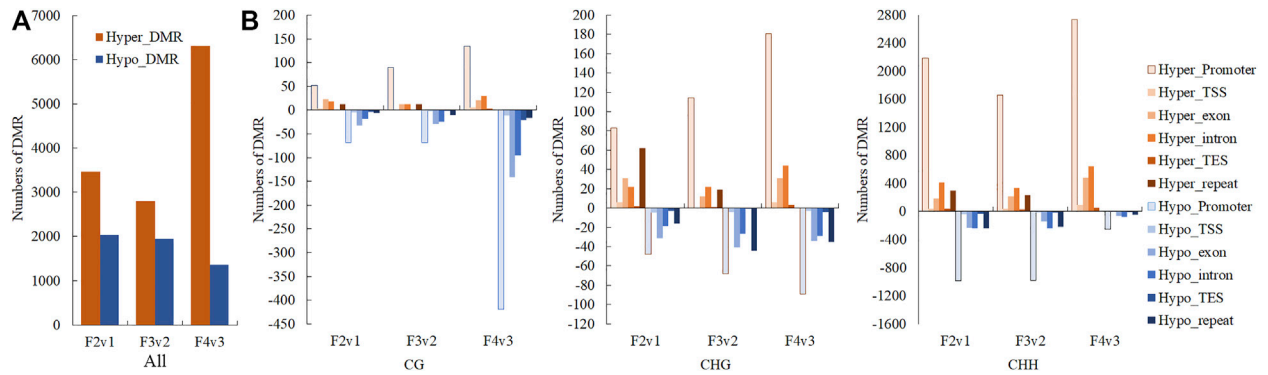


FIGURE 4 | Numbers of hyper- and hypo-DMRs during flowering, shown in **(A)** all mC site and **(B)** three different contexts (CG, CHG, CHH) and different gene region (promoter, TSS, exon, intron, TES and repeat). The terms "hyper-" and "hypo-" represent DMRs that were hypermethylated and hypomethylated, respectively.

stage, a higher expression level of the gene was associated with a lower methylation level of the TSS region. Furthermore, higher average ML was detected in the promoter regions of the CG and CHG contexts of unexpressed genes (**Figure 6B**).

The MGs were divided into five groups (groups 1–5) according to their ML from low to high, and the gene distribution frequencies (y-axis) at different expression levels (x-axis) were represented graphically for each group. In the promoter regions of the CG- and CHG-context MGs, a higher ML was associated with more genes being distributed in the low-expression region log2(FPKM+1) <1, and fewer genes being distributed in 2 < log2(FPKM+1) < 8. The CHH-context MGs also had similar distributions in frequency, although genes with the highest ML (group 5) appeared to have more genes in region 2 < log2(FPKM+1) < 8 than group 4 (**Figure 6C**). The results indicated that ML in the promoter regions was negatively correlated with expression level. In the gene body regions, the MG group with the highest ML was dominated by a low expression level in all CX contexts. From groups 1 to 3, a higher ML was associated with a higher frequency of genes in the expression level of region 2 < log2(FPKM+1) < 8. The results indicated that ML in gene body regions is positively correlated with expression level (**Figure 6C**).

The comparison of F2v1 in **Figure 6D** is here taken as an example. In the gene body region, there were 32 upregulated and 43 downregulated genes in hyper-DMG, and 28 upregulated and 41 downregulated genes in hypo-DMG. Meanwhile, in the promoter region, there were 253 upregulated and 268 downregulated genes in hyper-DMG, and 130 upregulated and 107 downregulated genes in hypo-DMG. Similar results were also found in F3v2 and F4v3. Some of the DMGs have both hyper- and hypomethylated regions (**Figure 6D**). These results showed that, when limiting the analysis to certain genes, the expression levels of many genes do not always conform to having a negative correlation with the ML in the promoter and a positive correlation with the ML in the gene body.

DMR-Associated DEGs Widely Involved in Floral Scent Biosynthesis

We detected 7706 (upregulated 3848; downregulated, 3858), 5016 (upregulated 2259; downregulated, 2457), and 12,451

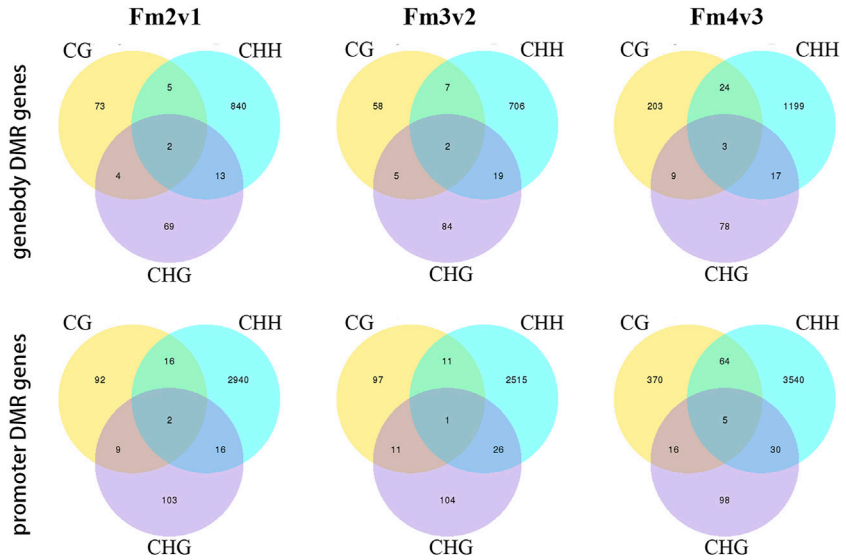


FIGURE 5 | Numbers of genes related to DMRs within promoters and gene bodies, shown for three contexts.

(upregulated 5502; downregulated, 6949) DEGs in compairsion Fr2v1, Fr3v2 and Fr4v3, respectively ("r" here indicate transcriptome) (**Figure 6D**). The results of KEGG enrichment analysis showed that a total of 214 DEGs were involved in seven floral scent biosynthesis pathways. As the key pathway for biosynthesis of the characteristic aroma of *P. mume*, phenylpropanoid biosynthesis (ko pmum00940) was enriched the most for DEGs (96 DEGs). Meanwhile, its upstream pathway, phenylalanine, tyrosine, and tryptophan biosynthesis (ko pmum00400), featured 32 DEGs. Moreover, 21, 39, 13, 14, and 10 DEGs were involved in phenylalanine metabolism (ko pmum00360), terpenoid backbone biosynthesis (ko pmum00900), monoterpenoid biosynthesis (ko pmum00902), diterpenoid biosynthesis (ko pmum00904), and sesquiterpenoid and triterpenoid biosynthesis (ko pmum00909), respectively (**Supplementary Table S7**).

Over 90% of the volatile components of *P. mume* fragrance were shown to be synthesized by the phenylpropane biosynthetic pathway. We also integrated some genes that previously characterized in the biosynthesis of other characteristic fragrance components in *P. mume* into the phenylpropane biosynthetic pathway, and then marked the processes involving DMGs with red asterisks (**Figure 7**). The results clearly showed that DMGs were widely involved in almost all steps in the phenylpropanoid biosynthesis process, including not only the dominant and key components, but also the low (FPKM value < lower quartile) or none (FPKM value < 1) part in *P. mume* fragrance.

We further summarized 145 genes that were related to the key enzyme in the phenylpropane biosynthetic pathway, as identified in our published results of *P. mume* floral fragrance research (Bao et al., 2019; Bao et al., 2020; Zhang et al., 2019b; Zhao et al., 2017), including 2 *PmPALs*, 6 *PmPAASs*, 37 *PmCNLs*, 11 *PmBARs*, 81 *PmBEATs*, 4 *PmCFATs*, and 4 *PmEGSs*. According to the heatmap results, among 47 candidate genes that exhibited an expression pattern matching the pattern of floral scent emission, 22 were differentially methylated during the flowering process (**Figure 7**, **Supplementary Table S10**). Some of these genes were proven to encode key enzymes involved in the characteristic fragrance of *P. mume* in previous studies, such as *PmCFAT1a/1c* (Zhang et al., 2019b), *PmBEAT36/37* (Bao et al., 2019), and *PmBAR8/9/10* (Bao et al., 2020); or to have the same expression pattern as in other *P. mume* cultivars, such as *PmPAL2*, *PmPAAS3*, and *PmCNL1/3/5/6/14/17/20* (Zhao et al., 2017; Bao et al., 2020) (**Supplementary Table S8**).

Among them, *PmCFAT1a*, which have been proved to produced eugenol through an enzyme assay (Zhang et al., 2019b), had four hyper-DMR in its genebody region during F1 to F2, involving all three CX contextsthe with ML changes from: 26.19–44.65% (CG), 20.05–39.34% (CHG), 2.43–5.64% and 1.06–3.45% (CHG). The ML of all hyper-DMR exceeded the average ML of the sample (**Table 2**; **Figure 7**). Pm011009 and Pm011010 (*PmBEAT36* and *PmBEAT37*) have been reported to significantly positive affect the synthesis of benzyl acetate in the petal cells of *P. mume* when overexpressed (Bao et al., 2019). Pm011009 had CG context hypo-DMR (90.49–81.02%) in the promoter region while the gene expressing were significantly increased in F3 to F4. Pm011010, which have been reported to inhibited the expression of Pm011009 in protoplasts, had CHH context hyper-DMR (22.41–39.87%) in its promoter region and CG context hypo-DMR (41.82–17.63%) in its genebody region from F3 to F4, although the expressing were also increased (**Figure 7**). We randomly selected 6 genes to do qRT-PCR to verify the transcriptome results. The qRT-PCR results showed a similar trend to the FPKM value (**Supplementary Figure S4**), proving that the transcriptome results are reliable.

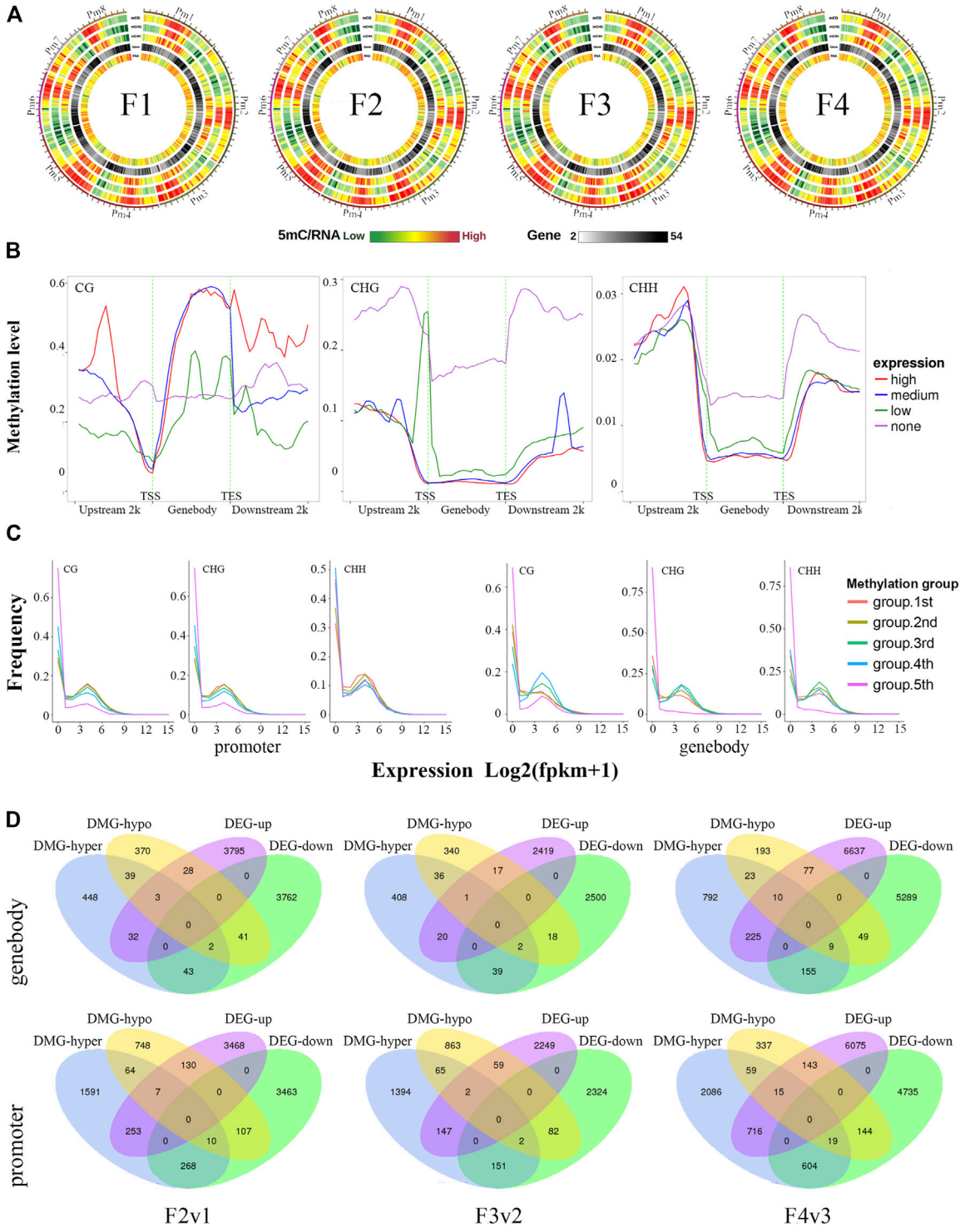


FIGURE 6 | Correlation between ML and expression level of genes, shown in three different CX contexts. **(A)** The overall relationship between ML and expression level on chromosomes. **(B)** The MLs of gene body and regions 2 kb up- and downstream of it in different CX contexts of genes under different expression levels. FPKM_25% and FPKM_75% refer to values at the boundary of the 25th and 75th percentiles of expression levels, respectively. Each region of each gene was divided equally into 50 bins, and the MLs of all CX sites in each bin were averaged as the ML of the bin. **(C)** The expression level distribution frequencies of six groups of genes with different MLs from low to high, shown separately for gene body and promoter regions. DNA methylation levels were classified into five groups: group 1 (red; 0 < methylation level < level_20%); group 2 (yellow-green; level_20% ≤ methylation level < level_40%); group 3 (green; level_40% ≤ methylation level < level_60%); group 4 (turquoise; level_60% ≤ methylation level < level_80%); and group 5 (blue; methylation level ≥ level_80%). Level_20%, _40%, _60%, and _80% represent values at the boundaries of the 20th, 40th, 60th, and 80th percentiles of methylation levels. **(D)** Venn plots showing the numbers of DEGs that were up- or downregulated in each comparison. The terms "hyper-" and "hypo-" represent DMRs within DEGs that were hypermethylated and hypomethylated, respectively. The results are displayed according to DMRs within promoter or gene body regions.

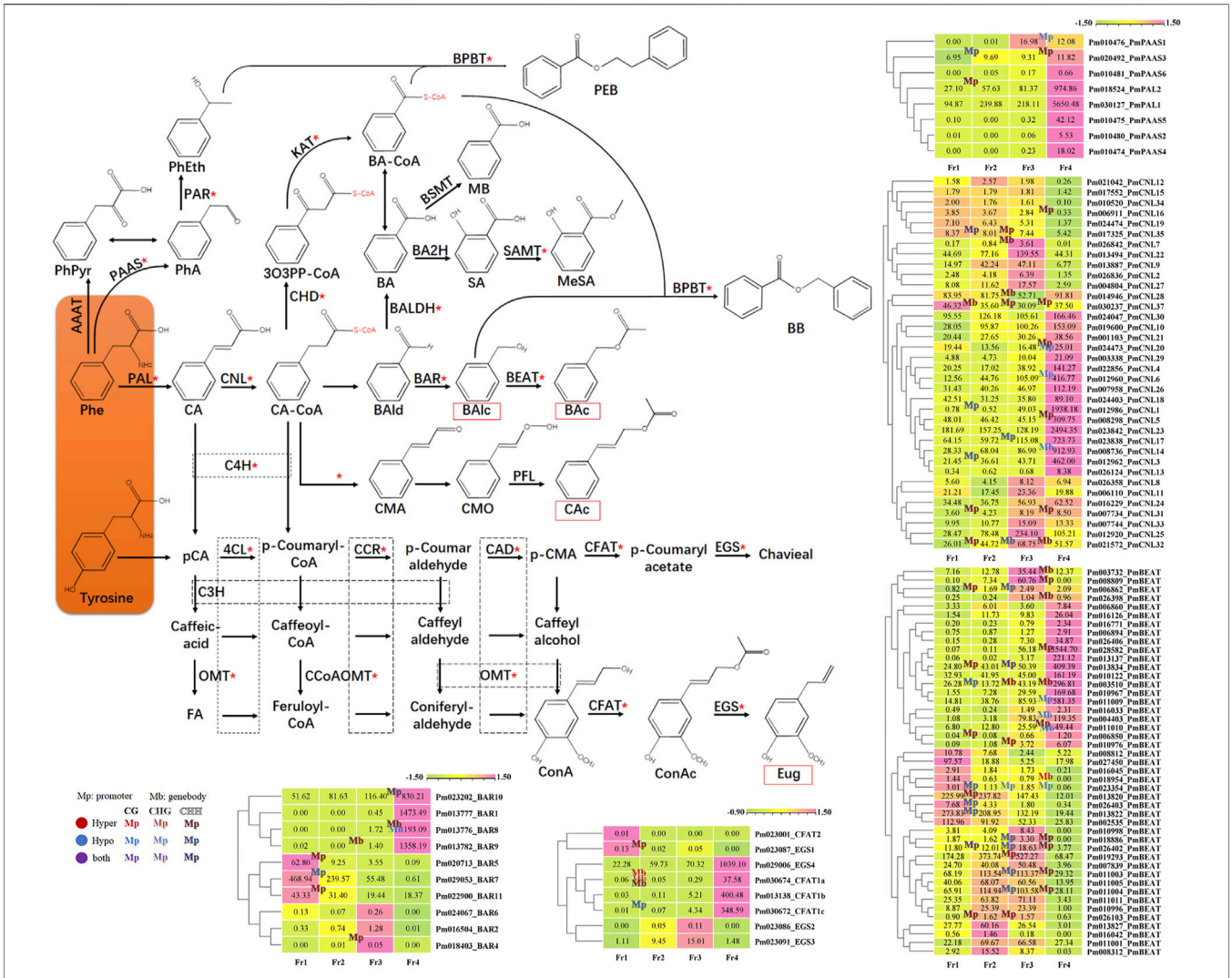


FIGURE 7 | Methylation modification and the heat maps of DEGs encoding key enzymes in the pathway of floral scent biosynthesis in *P. mume*. The key floral scent biosynthetic pathway in *P. mume* is shown. The biological compounds on an orange background are the starting substrates of the phenylpropane pathway, synthesized by the upstream pathway. The red rectangles indicate the metabolites that were detected in the headspace. Red asterisks indicate that genes encoding enzymes have undergone differential methylation during the process. The dashed boxes indicate that the processes involved share the same enzyme. In the heat maps, pink and light-green represent high and low transcript expression, respectively. The methylation statement were also marked in the heat maps with "Mp" and "Mb" to indicate the differential methylation occurred in promoter and genebody region, respectively; Text color with red, blue and purple represent ML up (hypermethylated), down (hypomethylated) and both happened respectively; Different fonts refer to different CX context, as shown in the figure.

DISCUSSION

The modification of DNA sequences by methylation is one of the heritable types of gene expression that does not change the genetic background. Research on DNA methylation thus has advantages in explaining the different phenotypes caused by differential gene expression in the same or similar genetic backgrounds. Studies have also shown variations in floral scent among different *P. mume* cultivars, despite strong similarity in their genetic background. As the only species of *Prunus* that produces a floral fragrance (Zhang et al., 2019b), research on the molecular mechanism regulating floral scent biosynthesis in *P. mume* is important.

Our results showed that DMGs were widely involved in eight recognized floral scent biosynthesis pathways during the natural flowering process of *P. mume*. Among these, the phenylpropane biosynthesis pathway was found to have the most enriched DMGs, which is consistent with the fact that over 90% of the floral volatiles emitted by the *P. mume* cv. 'FZ' are synthesized by the phenylpropane biosynthesis pathway (Zhang et al., 2019a). These DMGs participate in most of the floral biosynthetic processes in the phenylpropane biosynthetic pathway. All of these results indicated that methylation plays a role in the process of floral scent biosynthesis.

We selected 47 candidate genes that exhibit an expression pattern matching the emission of floral scent, 22 of which were

differentially methylated during flowering. Some of these genes have been proven to encode key enzymes involved in the characteristic floral scent of *P. mume* in previous studies, such as *PmCFAT1a/1c* (Zhang et al., 2019b) and *PmBEAT36/37* (Bao et al., 2019), or have the same expression pattern in other *P. mume* cultivars, such as *PmPAL2*, *PmPAAS3*, *PmBAR8/9/10*, and *PmCNL1/3/5/6/14/17/20* (Zhao et al., 2017; Bao et al., 2020). These lines of evidence indicate that the ML of DNA is involved in regulating the expression of genes encoding key enzymes in floral scent biosynthesis, and ultimately forms the unique floral scent of *P. mume*.

Notably, the DEGs involved in floral scent biosynthesis showed diverse patterns of methylation. Although our results were consistent with previous studies, hypermethylated regions were usually associated with segments showing low gene expression on chromosomes; and the gene expression level was found to be negatively and positively correlated with the ML of the promoter and gene body regions, respectively, among all methylated genes. However, when limiting the analysis to certain genes, the relationship between ML and gene expression did not always follow this rule. Similar result can be found in the result of previous research (Ma et al., 2018; Jiang et al., 2020). A more likely explanation was the complexity of the process of floral scent biosynthesis, so that methylation is not the only regulation factor. For example, the promoter of *PmBEAT36/37* has elements that respond to temperature and light (Bao et al., 2019). In the future, more research should be carried out within DMRs of key genes involved in floral scent biosynthesis that exhibit expression patterns matching the patterns of volatile emission and that differ in their expression patterns among *P. mume* cultivars with different scents.

In conclusion, our results revealed that DNA methylation is widely involved in the process of floral scent biosynthesis and may play critical roles in regulating such biosynthesis in *P. mume*. This was revealed by a comprehensive analysis of the RNA-seq and WGBS data of flower buds at four different flowering stages of the *P. mume* cv. 'FZ'. The results provide a new epigenetic perspective that deepens our understanding of the mechanism behind the biosynthesis of floral scent.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI under accession numbers PRJNA765446 and PRJNA766216.

AUTHOR CONTRIBUTIONS

XY, KM, MZ, and QZ conceived and designed the experiments. XY, and MZ performed the experiments. XY, KM, MZ, and JW analyzed the data. XY and KM wrote the manuscript. All authors have read and approved the published version of the manuscript.

FUNDING

This research was funded by The National Key R&D Program of China (2020YFD1000500) and Special Fund for Beijing Common Construction Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.779557/full#supplementary-material

Supplementary Table S10 | MR related to the 22 candidate genes.

REFERENCES

Bai, S., Tuan, P. A., Saito, T., Honda, C., Hatsuyama, Y., Ito, A., et al. (2016). Epigenetic Regulation of *MdMYB1* Is Associated with Paper Bagging-Induced Red Pigmentation of Apples. *Planta* 244 (3), 573–586. doi:10.1007/s00425-016-2524-4

Bao, F., Ding, A., Zhang, T., Luo, L., Wang, J., Cheng, T., et al. (2019). Expansion of *PmBEAT* Genes in the *Prunus Mume* Genome Induces Characteristic floral Scent Production. *Hortic. Res.* 6, 24. doi:10.1038/s41438-018-0104-4

Bao, F., Zhang, T., Ding, A., Ding, A., Yang, W., Wang, J., et al. (2020). Metabolic, Enzymatic Activity, and Transcriptomic Analysis Reveals the Mechanism Underlying the Lack of Characteristic floral Scent in Apricot Mei Varieties. *Front. Plant Sci.* 11, 574982. doi:10.3389/fpls.2020.574982

Boatright, J., Negre, F., Chen, X., Kish, C. M., Wood, B., Peel, G., et al. (2004). Understanding *In Vivo* Benzenoid Metabolism in Petunia Petal Tissue. *Plant Physiol.* 135 (4), 1993–2011. doi:10.1104/pp.104.045468

Chan, S. W.-L., Henderson, I. R., and Jacobsen, S. E. (2005). Gardening the Genome: DNA Methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 6 (5), 351–360. doi:10.1038/nrg1601

Chen, J. (1996). *Chinese Mei Flowers*. Haikou: Hainan Publishing House.

Cubas, P., Vincent, C., and Coen, E. (1999). An Epigenetic Mutation Responsible for Natural Variation in floral Symmetry. *Nature* 401 (6749), 157–161. doi:10.1038/43657

Dhandapani, S., Jin, J., Sridhar, V., Sarojam, R., Chua, N.-H., and Jang, I.-C. (2017). Integrated Metabolome and Transcriptome Analysis of *Magnolia Champaca* Identifies Biosynthetic Pathways for floral Volatile Organic Compounds. *BMC Genomics* 18 (1), 463. doi:10.1186/s12864-017-3846-8

Dudareva, N., Klempien, A., Muhlemann, J. K., and Kaplan, I. (2013). Biosynthesis, Function and Metabolic Engineering of Plant Volatile Organic Compounds. *New Phytol.* 198 (1), 16–32. doi:10.1111/nph.12145

Feng, H., Conneely, K. N., and Wu, H. (2014). A Bayesian Hierarchical Model to Detect Differentially Methylated Loci from Single Nucleotide Resolution Sequencing Data. *Nucleic Acids Res.* 42 (8), e69. doi:10.1093/nar/gku154

Fenske, M. P., Hewett Hazelton, K. D., Hempton, A. K., Shim, J. S., Yamamoto, B. M., Riffell, J. A., et al. (2015). Circadian Clock Gene LATE ELONGATED HYPOCOTYL Directly Regulates the Timing of floral Scent Emission in Petunia. *Proc. Natl. Acad. Sci. USA* 112 (31), 9775–9780. doi:10.1073/pnas.1422875112

Finnegan, E. J., and Kovac, K. A. (2000). Plant DNA Methyltransferases. *Plant Mol. Biol.* 43 (2-3), 189–201. doi:10.1023/a:1006427226972

Fraser, C. M., and Chapple, C. (2011). The Phenylpropanoid Pathway in *Arabidopsis*. *The Arabidopsis Book* 9, e0152. doi:10.1199/tab.0152

Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., et al. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell* 153 (5), 1149–1163. doi:10.1016/j.cell.2013.04.037

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., et al. (2013). Whole-genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell* 13 (3), 360–369. doi:10.1016/j.stem.2013.06.002

Hao, R.-J., Zhang, Q., Yang, W.-R., Wang, J., Cheng, T.-R., Pan, H.-T., et al. (2014b). Emitted and Endogenous floral Scent Compounds of *Prunus Mume* and Hybrids. *Biochem. Syst. Ecol.* 54, 23–30. doi:10.1016/j.bse.2013.12.007

Hao, R., Du, D., Wang, T., Yang, W., Wang, J., and Zhang, Q. (2014a). A Comparative Analysis of Characteristic floral Scent Compounds in *Prunus Mume* and Related Species. *Biosci. Biotechnol. Biochem.* 78 (10), 1640–1647. doi:10.1080/09168451.2014.936346

Huang, H., Liu, R., Niu, Q., Tang, K., Zhang, B., Zhang, H., et al. (2019). Global Increase in DNA Methylation during orange Fruit Development and Ripening. *Proc. Natl. Acad. Sci. USA* 116 (4), 1430–1436. doi:10.1073/pnas.1815441116

Jiang, L., Zhang, M., and Ma, K. (2020). Whole-Genome DNA Methylation Associated with Differentially Expressed Genes Regulated Anthocyanin Biosynthesis within Flower Color Chimera of Ornamental Tree *Prunus Mume*. *Forests* 11 (1), 90. doi:10.3390/f11010090

Jin, J., Kim, M. J., Dhandapani, S., Tjhang, J. G., Yin, J.-L., Wong, L., et al. (2015). The floral Transcriptome of Ylang Ylang (*Cananga Odorata* Var. Fruticosa) Uncovers Biosynthetic Pathways for Volatile Organic Compounds and a Multifunctional and Novel Sesquiterpene Synthase. *J. Exp. Bot.* 66 (13), 3959–3975. doi:10.1093/jxb/erv196

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

Karami, A., Niazi, A., Kavoosi, G., Khosh-Khui, M., and Salehi, H. (2015). Temporal Characterization of 2-phenylethanol in Strongly and Weakly Scented Genotypes of Damask Rose. *Physiol. Mol. Biol. Plants* 21 (1), 43–49. doi:10.1007/s12298-014-0274-y

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* 12 (4), 357–360. doi:10.1038/nmeth.3317

Knudsen, J. T., Eriksson, R., Gershenzon, J., and Ståhl, B. (2006). Diversity and Distribution of floral Scent. *Bot. Rev.* 72 (1), 1–120. doi:10.1663/0006-8101(2006)72[1:DADOFS]2.0.CO;2

Krueger, F., and Andrews, S. R. (2011). Bismark: a Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinformatics* 27 (11), 1571–1572. doi:10.1093/bioinformatics/btr167

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

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Law, J. A., and Jacobsen, S. E. (2010). Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals. *Nat. Rev. Genet.* 11 (3), 204–220. doi:10.1038/nrg2719

Li, W.-F., Ning, G.-X., Mao, J., Guo, Z.-G., Zhou, Q., and Chen, B.-H. (2019). Whole-genome DNA Methylation Patterns and Complex Associations with Gene Expression Associated with Anthocyanin Biosynthesis in Apple Fruit Skin. *Planta* 250 (6), 1833–1847. doi:10.1007/s00425-019-03266-4

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* 462 (7271), 315–322. doi:10.1038/nature08514

Ma, K.-F., Zhang, Q.-X., Cheng, T.-R., Yan, X.-L., Pan, H.-T., and Wang, J. (2018). Substantial Epigenetic Variation Causing Flower Color Chimerism in the Ornamental Tree Prunus Mume Revealed by Single Base Resolution Methylome Detection and Transcriptome Sequencing. *Ijms* 19 (8), 2315. doi:10.3390/ijms19082315

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 (19), 3787–3793. doi:10.1093/bioinformatics/bti430

Moerkercke, A. V., Schauvinhold, I., Pichersky, E., Haring, M. A., and Schuurink, R. C. (2009). A Plant Thiolase Involved in Benzoic Acid Biosynthesis and Volatile Benzenoid Production. *Plant J.* 60 (2), 292–302. doi:10.1111/j.1365-313X.2009.03953.x

Muhlemann, J. K., Klempien, A., and Dudareva, N. (2014). Floral Volatiles: from Biosynthesis to Function. *Plant Cell Environ.* 37 (8), 1936–1949. doi:10.1111/pce.12314

Park, Y., and Wu, H. (2016). Differential Methylation Analysis for BS-Seq Data under General Experimental Design. *Bioinformatics* 32 (10), 1446–1453. doi:10.1093/bioinformatics/btw026

Schiestl, F. P. (2010). The Evolution of floral Scent and Insect Chemical Communication. *Ecol. Lett.* 13 (5), 643–656. doi:10.1111/j.1461-0248.2010.01451.x

Song, Q.-X., Lu, X., Li, Q.-T., Chen, H., Hu, X.-Y., Ma, B., et al. (2013). Genome-wide Analysis of DNA Methylation in Soybean. *Mol. Plant* 6 (6), 1961–1974. doi:10.1093/mp/sst123

Sun, L., Yang, W., Zhang, Q., Cheng, T., Pan, H., Xu, Z., et al. (2013). Genome-wide characterization and linkage mapping of simple sequence repeats in mei (*Prunus mume* Sieb. et Zucc). *PLoS One* 8 (3), e59562. doi:10.1371/journal.pone.0059562

Suzuki, M. M., and Bird, A. (2008). DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nat. Rev. Genet.* 9 (6), 465–476. doi:10.1038/nrg2341

Tzin, V., and Galili, G. (2010). New Insights into the Shikimate and Aromatic Amino Acids Biosynthesis Pathways in Plants. *Mol. Plant* 3 (6), 956–972. doi:10.1093/mp/ssq048

Vaughn, M. W., Tanurdžić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., et al. (2007). Epigenetic Natural Variation in *Arabidopsis* thaliana. *Plos Biol.* 5 (7), e174. doi:10.1371/journal.pbio.0050174

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., et al. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell* 157 (4), 979–991. doi:10.1016/j.cell.2014.04.017

Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., et al. (2015). Long Noncoding RNA S and Their Proposed Functions in Fibre Development of Cotton (Gossypium spp). *New Phytol.* 207 (4), 1181–1197. doi:10.1111/nph.13429

Wang, Z., Meng, D., Wang, A., Li, T., Jiang, S., Cong, P., et al. (2013). The Methylation of the PcMYB10 Promoter Is Associated with Green-Skinned Sport in Max Red Bartlett Pear. *Plant Physiol.* 162 (2), 885–896. doi:10.1104/pp.113.214700

Widhalm, J. R., and Dudareva, N. (2015). A Familiar Ring to it: Biosynthesis of Plant Benzoic Acids. *Mol. Plant* 8 (1), 83–97. doi:10.1016/j.molp.2014.12.001

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., et al. (2015). Detection of Differentially Methylated Regions from Whole-Genome Bisulfite Sequencing Data without Replicates. *Nucleic Acids Res.* 43 (21), gkv715. doi:10.1093/nar/gkv715

Xing, L., Li, Y., Qi, S., Zhang, C., Ma, W., Zuo, X., et al. (2019). Comparative RNA-Sequencing and DNA Methylation Analyses of Apple (*Malus Domestica* Borkh.) Buds with Diverse Flowering Capabilities Reveal Novel Insights into the Regulatory Mechanisms of Flower Bud Formation. *Plant Cell Physiol.* 60 (8), 1702–1721. doi:10.1093/pcp/pcz080

Xu, J., Zhou, S., Gong, X., Song, Y., van Nocker, S., Ma, F., et al. (2018). Single-base Methylome Analysis Reveals Dynamic Epigenomic Differences Associated with Water Deficit in Apple. *Plant Biotechnol. J.* 16 (2), 672–687. doi:10.1111/pbi.12820

Xu, Z., Zhang, Q., Sun, L., Du, D., Cheng, T., Pan, H., et al. (2014). Genome-wide Identification, Characterisation and Expression Analysis of the MADS-Box Gene Family in *Prunus Mume*. *Mol. Genet. Genomics* 289 (5), 903–920. doi:10.1007/s00438-014-0863-z

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 (2), R14. doi:10.1186/gb-2010-11-2-r14

Yuan, X., Ma, K. F., Zhang, M. H., Zhang, Z. Y., Wang, J., and Zhang, Q. X. (2020). How Does an Early Flowering Cultivar of *Prunus Mume* Arrange its Bud Development in Late winter? *Acta Hortic.* 1291, 15–20. doi:10.17660/ActaHortic.2020.1291.3

Yue, Y., Yu, R., and Fan, Y. (2015). Transcriptome Profiling Provides New Insights into the Formation of floral Scent in *Hedychium Coronarium*. *BMC Genomics* 16 (1), 470. doi:10.1186/s12864-015-1653-7

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., et al. (2012). The Genome of *Prunus Mume*. *Nat. Commun.* 3 (1), 1318. doi:10.1038/ncomms2290

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., et al. (2018). The Genetic Architecture of floral Traits in the Woody Plant *Prunus Mume*. *Nat. Commun.* 9 (1), 1702. doi:10.1038/s41467-018-04093-z

Zhang, T., Bao, F., Yang, Y., Hu, L., Ding, A., Ding, A., et al. (2019a). A Comparative Analysis of Floral Scent Compounds in Intraspecific Cultivars of *Prunus Mume* with Different Corolla Colours. *Molecules* 25 (1), 145. doi:10.3390/molecules25010145

Zhang, T., Huo, T., Ding, A., Hao, R., Wang, J., Cheng, T., et al. (2019b). Genome-wide Identification, Characterization, Expression and Enzyme Activity Analysis of Coniferyl Alcohol Acetyltransferase Genes Involved in Eugenol Biosynthesis in *Prunus Mume*. *PLoS One* 14 (10), e0223974. doi:10.1371/journal.pone.0223974

Zhao, K., Yang, W., Zhou, Y., Zhang, J., Li, Y., Ahmad, S., et al. (2017). Comparative Transcriptome Reveals Benzenoid Biosynthesis Regulation as Inducer of Floral Scent in the Woody Plant *Prunus Mume*. *Front. Plant Sci.* 8, 319. doi:10.3389/fpls.2017.00319

Zheng, J., Hu, Z., Guan, X., Dou, D., Bai, G., Wang, Y., et al. (2015). Transcriptome Analysis of *Syringa Oblata* Lindl. Inflorescence Identifies Genes Associated with Pigment Biosynthesis and Scent Metabolism. *PLoS One* 10 (11), e0142542. doi:10.1371/journal.pone.0142542

Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., et al. (2013). Single-base Resolution Methylomes of Tomato Fruit Development Reveal Epigenome Modifications Associated with Ripening. *Nat. Biotechnol.* 31 (2), 154–159. doi:10.1038/nbt.2462

Zhu, J.-K. (2009). Active DNA Demethylation Mediated by DNA Glycosylases. *Annu. Rev. Genet.* 43, 143–166. doi:10.1146/annurev-genet-102108-134205

Zhuo, X., Zheng, T., Li, S., Zhang, Z., Zhang, M., Zhang, Y., et al. (2021). Identification of the *PmWEEP* Locus Controlling Weeping Traits in *Prunus Mume* through an Integrated Genome-wide Association Study and Quantitative Trait Locus Mapping. *Hortic. Res.* 8 (1), 131. doi:10.1038/s41438-021-00573-4

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.

Copyright © 2021 Yuan, Ma, Zhang, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

3O3PP-CoA 3-oxo-3-phenylpropionyl-CoA
4CL 4-coumarate CoA-ligase
BA benzoic acid
BAc Benzyl acetate
BA-CoA benzoyl-CoA
BAlc benzylalcohol
BAld benzaldehyde
BALDH benzaldehyde dehydrogenase
BB benzylbenzoate
BEAT acetyl-CoA:benzylalcohol acetyltransferase
BPBT benzoyl-CoA:benzylalcohol/2-phenylethanol benzoyltransferase
BSMT benzoic acid/salicylic acid carboxyl methyltransferase
C4H cinnamic acid 4-hydroxylase
CA cinnamic acid
CAc Cinnamyl acetate
CA-CoA cinnamoyl-CoA
CAld Cinnamyl dehyde
CCoAOMT caffeoyl-CoA O-methyltransferase
CFAT coniferylalcohol acetyltransferase
CHD cinnamoyl-CoA hydratase/dehydrogenase
CMA cinnamaldehyde
CMO cinnamyl alcohol
CNL cinnamoyl-CoA ligase
ConA coniferyl alcohol
EGS eugenol synthase
Eug eugenol
FA ferulic acid
KAT 3-ketoacyl-CoA thiolase
MB methylbenzoate
MeSA Methyl Salicylate
PAAS phenylacetaldehyde synthase
PAL phenylalanine ammonia-lyase
PAR phenylacetaldehyde reductase
pCA p-Coumaric acid
p-CMA p-Coumaryl alcohol
PhA phenylacetaldehyde
Phe L-phenylalanine
PhEth 2-phenylethanol
PhPyr phenylpyruvic acid
SA Salicylate
Tyr Tyrosine.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fgene.2022.756270

Transcript and Protein Profiling Provides Insights Into the Molecular Mechanisms of Harvesting-Induced Latex Production in Rubber Tree

*Yujie Fan[1†], Jiyan Qi[1†], Xiaohu Xiao[2†], Heping Li[1], Jixian Lan[1], Yacheng Huang[1], Jianghua Yang[2], Yi Zhang[1], Shengmin Zhang[1], Jun Tao[1] and Chaorong Tang[1]**

[1]Natural Rubber Cooperative Innovation Center of Hainan Province and Ministry of Education of PRC, Hainan University, Haikou, China, [2]Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

OPEN ACCESS

Edited by:
Lidan Sun,
Beijing Forestry University, China

Reviewed by:
Ertugrul Filiz,
Duzce University, Turkey
Jisen Zhang,
Fujian Agriculture and Forestry University, China
Chunxiang Fu,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China

***Correspondence:**
Chaorong Tang
chaorongtang@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: *10 August 2021*
Accepted: *17 January 2022*
Published: *10 February 2022*

Citation:
Fan Y, Qi J, Xiao X, Li H, Lan J, Huang Y, Yang J, Zhang Y, Zhang S, Tao J and Tang C (2022) Transcript and Protein Profiling Provides Insights Into the Molecular Mechanisms of Harvesting-Induced Latex Production in Rubber Tree. Front. Genet. 13:756270. doi: 10.3389/fgene.2022.756270

Natural rubber, an important industrial raw material with wide applications, is harvested in the form of latex (cytoplasm of rubber-producing laticifers) from *Hevea brasiliensis* (para rubber tree) by the way of tapping. Conspicuous stimulation on latex production is observed for the first few tappings conducted on virgin (untapped before) or resting (tapped before but no tapping for a period) rubber trees. To understand the underlying mechanisms, an integrative analysis of the latex transcriptome and proteome was conducted on virgin or resting *Hevea* trees for the first five tappings. A total of 505 non-redundant differentially expressed (DE) transcript-derived fragments (TDFs) were identified by silver-staining cDNA-AFLP, with 217 exhibiting patterns of upregulated, 180 downregulated and 108 irregularly-regulated. Meanwhile, 117 two dimensional gel electrophoresis DE-protein spots were isolated and subjected to mass spectrometry analysis, with 89 and 57 being successfully identified by MALDI-TOF and MALDI-TOF/TOF, respectively. About 72.5% DE-TDFs and 76.1% DE-proteins were functionally annotated and categorized. Noteworthily, most of the DE-TDFs implicated in sugar transport and metabolism as well as rubber biosynthesis were upregulated by the tapping treatment. The importance of sugar metabolism in harvesting-induced latex production was reinforced by the identification of abundant relevant DE-protein spots. About 83.8% of the randomly selected DE-TDFs were validated for expression patterns by semi-quantitative RT-PCR, and an 89.7% consistency for the 29 latex regeneration-related DE-TDFs examined by quantitative RT-PCR analysis. In brief, our results reveal extensive physiological and molecular changes in *Hevea* laticifers incurred by the tapping treatment, and the vast number of DE genes and proteins identified here contribute to unraveling the gene regulatory network of tapping-stimulated latex production.

Keywords: *Hevea brasiliensis*, cDNA-AFLP, 2-DE, gene and protein expression, harvesting, latex production

Abbreviations: 2-DE, Two dimensional electrophoresis; AFLP, Amplified fragment length polymorphism; ATP, Adenosine triphosphate; CoA, Coenzyme A; DE-proteins, Differentially expressed proteins; DE-TDF, Differentially expressed transcript derived fragment; DNA, Deoxyribonucleic acid; dNTP, Deoxy-ribonucleoside triphosphate; EDTA, Ethylene diamine tetraacetic acid; IPG, Immobilized pH gradient; NADPH, Reduced form of nicotinamide-adenine dinucleotide phosphate; NIN, Alkaline/neutral invertase; NR, Natural rubber; PCR, Polymerase chain reaction; qRT-PCR, Quantitative reverse transcription PCR; RB, Rubber biosynthesis; REF, Rubber elongation factor; sqRT-PCR, Semi-quantitative reverse transcription PCR; SRPP, Small rubber particle protein; SUT, Sucrose transporter; TDF, Transcript derived fragment.

INTRODUCTION

Natural rubber (*cis*-1, 4-polyisoprene, NR) is an elastomer with superior properties that cannot be completely replaced by petroleum-derived synthetic rubber and is used as an important industrial raw material. Due to the advantages of good quality, high yield, and easiness for harvesting, *Hevea brasiliensis* (para rubber tree, *Hevea* thereafter) has become the sole commercial NR source among the 2, 500 NR-bearing plant species (Van Beilen and Poirier, 2007). *Hevea* trees need warm and humid climate conditions for normal growth and NR production yet vulnerable to typhoon, thus confining its planting to restricted tropical areas (Huang, 2001).

Hevea rubber yield is affected mainly by three factors: duration of latex flow after tapping, the capability of latex regeneration between two consecutive tappings, and the ability of laticifer differentiation in bark cambium (d'Auzac et al., 1997; Hao et al., 2004). In the trunk bark of tapped *Hevea* trees, the number of laticifer rings is one to three times more than that observed in un-tapped trees, indicating the stimulating effect of tapping on laticifer differentiation (Hao and Wu, 1982). Meanwhile, both mechanical injury and jasmonic acid induce laticifer differentiation, latex regeneration and production (Hao and Wu, 2000; Sae-Lim et al., 2019). Cytologically, latex is the cytoplasm of the laticifers, 90% of the dry weight as rubber hydrocarbon (NR) (Chrestin et al., 1997). To sustain NR productivity, in regularly tapped *Hevea* trees, the expelled latex must be efficiently regenerated before the next tapping, usually with an interval of 2–3 days (d'Auzac et al., 1997).

After tapping, the latex flowing out of the laticifers includes cis-polyisoprene particles (rubber particles), lysosomal microvacuoles (lutoids), plastid-like Frey-Wyssling complexes, sugars, organic acids, nucleic acids, and proteins/enzymes (d'Auzac et al., 1997). A positive correlation exists between sucrose content and latex yield, and the control of sucrose metabolism has been intensively studied in relation to latex production (Tupý, 1969; Tupý, 1985; Tupý, 1989). Expression of the responsible sucrose transporter, HbSUT3, for sucrose loading into laticifers is induced by the treatments of ethylene and tapping, both bolstering the latex yield (Tang et al., 2010). In the latex, an alkaline/neutral invertase is responsible for cleaving sucrose into hexose sugars that are then exploited in subsequent latex production (Tupý, 1989). Expression and enzymatic activity of the relevant invertase gene, *HbNIN2*, correlate positively with the latex yield (Liu et al., 2015). Rubber particles are a kind of organelle suspended in latex where rubber biosynthesis takes place and subsequently stores, with proteins of rubber elongation factor (REF)/small rubber particle protein (SRPP) as important rubber-synthesizing participants (Cornish, 1993; Oh et al., 1999; Singh et al., 2003; Tang et al., 2016). Priya et al. (2007) observed a positive correlation between REF mRNA abundance and the latex yielding levels of different *Hevea* clones. After Ethrel (2-chloroethylphosphonic acid, an ethylene releaser) bark treatment, a marked increase in transtonoplast ΔpH within *Hevea* laticifers was observed consisting of one of the major mechanisms of ethylene stimulation on latex yield (Amalou et al., 1992). Together, these studies suggest that latex production, with NR biosynthesis as the core activity, involves a complex regulatory network of gene expression, multi-enzyme reaction and physio-biochemical processes. Regular tapping with intervals of 2–4 days significantly stimulates latex yield on both virgin (never tapped before) and resting (tapped before but left with no tapping for 3 months or above) *Hevea* trees in a similar way, except that the number of tappings required is larger in the former (7–10) than the latter (4–6) to yield at a relatively stable level (Pakianathan et al., 1992; Tang et al., 2010; Qi, 2011). In the latex of virgin *Hevea* trees, expressions of a number of genes involved in latex production are bolstered by the tapping treatment, e.g., *HbSUT3* (Tang et al., 2010), *HbNIN2* (Liu et al., 2015) and a farnesyl diphosphate synthase gene (Adiwilaga and Kush, 1996). Latex transcript and protein

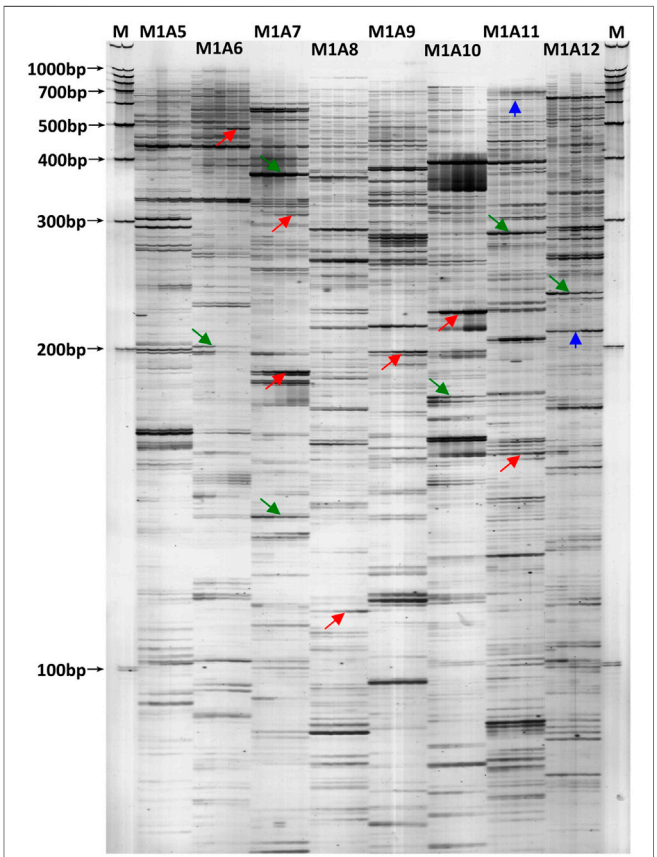


FIGURE 1 | A typical cDNA-AFLP silver-stained polyacrylamide gel revealing transcript profiles in the latex of virgin *Hevea* trees for the first five tappings. M: 100 bp DNA ladder molecular weight standard; M1A5 to M1A12: selective primer combinations. Under each primer pair, the five lanes (from left to right) represent the latex transcript profiles for the first five tappings. The arrows marked in red indicate the DE-TDFs up-regulated by tapping, the green arrows indicate down-regulated, and the blue arrows indicate irregularly-regulated. Please note there exists subtypes for each of the three regulation types.

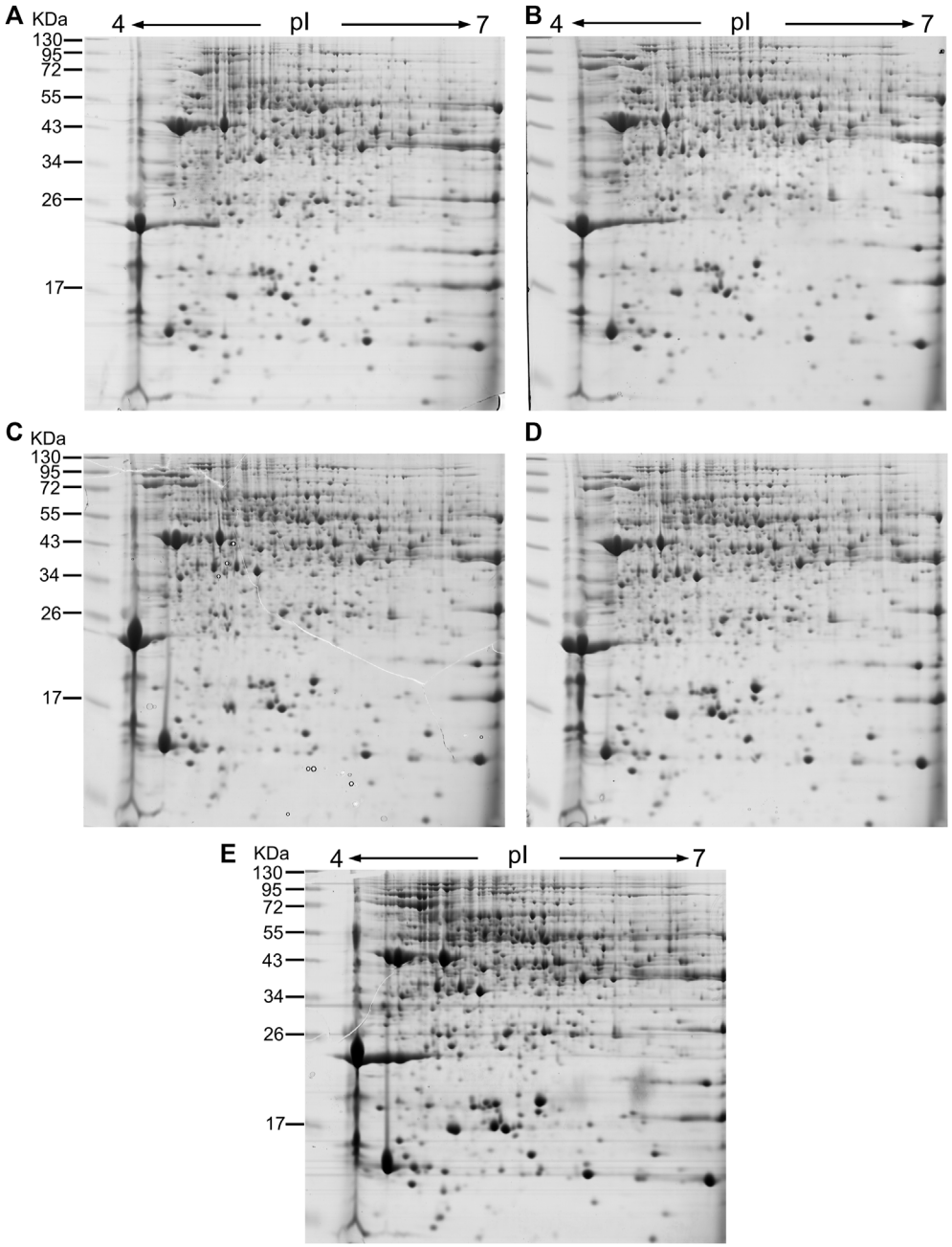


FIGURE 2 | 2-DE images of latex C-serum proteins from retapped *Hevea* trees after a four-month resting period. IGP strips of pH4-7 were used in 2-DE, and the fresh latex C-serum samples were analyzed for the first five tappings [**(A)**: first; **(B)**: second; **(C)**: third; **(D)**: fourth; **(E)**: fifth]. kDa: Molecular weight of protein. pI: Protein isoelectric point.

profiling of virgin or resting *Hevea* trees for the first few tappings could be beneficial to unraveling the mechanisms of tapping-induced latex production and to identifying the genes or proteins involved.

Since the advent of the cDNA-AFLP transcript-profiling technique in 1996 (Bachem et al., 1996), owing to its advantages of good repeatability, high sensitivity and high throughput, this technique has been successfully applied to various aspects of plant studies, such as plant's abiotic stress response (Fatemi et al., 2019), plant-microorganism interaction (Xiao et al., 2016), hormone signaling (Akihiro et al., 2006), and plant development (Leymarie et al., 2007). In this study, the latex transcriptome and proteome were compared for the first five tappings in virgin and resting *Hevea* trees, respectively, using a silver staining cDNA-AFLP (Xiao et al., 2009) and two dimensional electrophoresis coupled with mass spectrometry

TABLE 1 | Functional categories and statistics of DE-TDFs and DE-protein spots.

Category	No.	Percentage[a] (%)	Expression pattern[b] (%)		
			Up	Down	Irregular
DE-TDFs					
Primary metabolism	30	5.9	15 (50.0)	12 (40.0)	3 (10.0)
Energy	20	4.0	8 (40.0)	5 (25.0)	7 (35.0)
Cell growth and division	19	3.8	11 (57.9)	5 (26.3)	3 (15.8)
Transcription and protein synthesis	96	19.0	41 (42.7)	35 (36.5)	20 (20.8)
Protein degradation and storage	34	6.7	18 (52.9)	9 (26.5)	7 (20.6)
Transporters and intracellular transport	45	8.9	20 (44.4)	18 (40.0)	7 (15.6)
Cellular structure	13	2.6	7 (53.8)	6 (46.2)	0
Signal transduction	42	8.3	12 (28.6)	18 (42.8)	12 (28.6)
Stress and defense	46	9.1	23 (50.0)	16 (34.8)	7 (15.2)
Secondary metabolism	11	2.2	8 (72.7)	2 (18.2)	1 (9.1)
Rubber biosynthesis	10	2.0	9 (90.0)	0	1 (10.0)
Unclassified proteins	30	5.9	7 (23.3)	16 (53.4)	7 (23.3)
Predicted proteins	59	11.7	24 (40.7)	20 (33.9)	15 (25.4)
No hit sequence	50	9.9	14 (28.0)	18 (36.0)	18 (36.0)
Total	505	100.0	217 (43.0)	180 (35.6)	108 (21.4)
DE-protein spots					
Primary metabolism and energy	18	20.2	8 (44.4)	5 (27.8)	5 (27.8)
Cell growth, division and structure	12	13.5	4 (33.3)	3 (25.0)	5 (41.7)
Transcription and protein synthesis	9	10.1	5 (55.6)	3 (33.3)	1 (11.1)
Signal transduction	4	4.5	4 (100)	0	0
Stress and defense	23	25.8	4 (17.4)	4 (17.4)	15 (65.2)
Secondary metabolism	4	4.5	3 (75.0)	1 (25.0)	0
Rubber biosynthesis	8	9.0	3 (37.5)	2 (25.0)	3 (37.5)
Unclassified proteins	11	12.4	4 (36.4)	2 (18.1)	5 (45.5)
Total	89	100	35 (39.3)	20 (22.5)	34 (38.2)

[a]*The percentage in each category accounting for the total DE-TDFs and DE-protein spots.*

[b]*The number and percentage of DE-TDFs and DE-protein spots distributed in the three expression modes respectively in each category. Up: up-regulated; Down: down-regulated; Irregular: irregularly-regulated.*

analysis. A large number of differentially expressed (DE) transcripts and proteins were successfully isolated and functionally identified, and the results reveal the importance of sugar metabolism as well as rubber biosynthesis in tapping-activated latex production in *Hevea* trees.

RESULTS

Screening of Differentially Expressed Transcripts and Proteins

To extensively identify the differentially expressed transcripts responding to the tapping treatment, all the 128 cDNA-AFLP selective primer pairs of *Apo* I / *Mse* I restriction system (Xiao et al., 2009) were screened in the latex RNA for the first five tappings in virgin *Hevea* trees. On average, about 70 transcript-derived fragments (TDFs) greater than 100 bp were discernable on the silver-stained polyacrylamide gels for each primer pair (**Figure 1**). Therefore, nearly 9,000 TDFs were profiled for each latex RNA sample. In total, 651 DE-TDFs were successfully cloned and sequenced.

To identify the tapping-responsive proteins, proteomics studies were preformed on re-opened resting *Hevea* trees, the latex production of which is stimulated by tapping in a way similar to virgin *Hevea* trees (**Supplementary Figure S1**). The resting *Hevea* trees usually yield significantly higher than virgin *Hevea* trees for the first tapping, thus facilitating the conventional proteomics analysis that requires more latex than the virgin *Hevea* trees could produce (Tang et al., 2010; Qi, 2011). For each of the first five tappings in reopened resting *Hevea* trees, 1 mg of latex C-serum proteins were separated on two dimensional electrophoresis (2-DE) gels and stained by Coomassie brilliant blue R250. Most of the protein spots displayed superior resolution with clear background and obvious boundary on the 2-DE gels (**Figure 2**). On each 2-DE gel, the number of discernable protein spots ranged between 700 and 800 when pH 4–7 IPG strips were used for 1-DE (**Figure 2**), and between 900 and 1,000 when pH 5–8 IPG strips were used (**Supplementary Figure S2**). The molecular weights of latex C-serum proteins were concentrated in the range of 15–80 kDa, and the isoelectric points were in pH 4.5–7.5.

Taking the expression level at the first tapping as a reference, the DE-TDFs identified were classified into three types according to their patterns of expression across the five successive tappings: upregulation, downregulation and irregular-regulation (**Figure 1**). The upregulation type includes three subtypes: 1) increase successively; 2) increase first and then stabilize; 3) increase first and then decrease, but still higher than the first tapping. The downregulation type also includes three subtypes: 1) decrease successively; 2) decrease first and then stabilizes; 3) decrease first and then increase, but still lower than the first

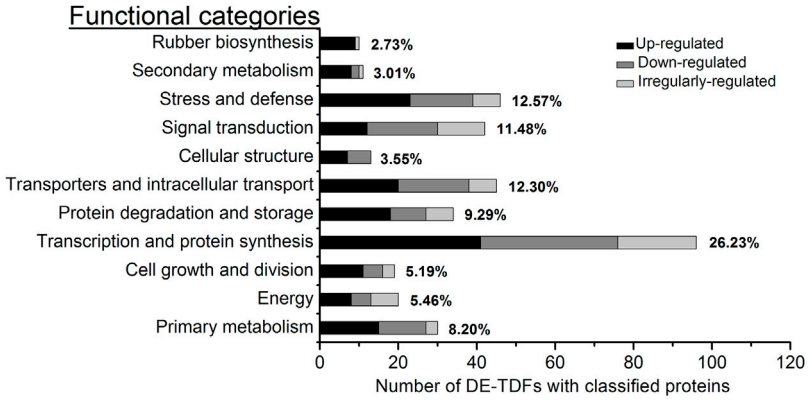


FIGURE 3 | Functional category and percentage of the DE-TDFs with known function. The 366 non-redundant DE-TDFs of annotation proteins with known function were identified in the latex for the first five tappings in rubber trees. These DE-TDFs were classified into 11 functional categories. In each category, the percentage covering the total know functional DE-TDFs was placed at its right whereas the three types of expression were shown in differing gray bars.

tapping. The irregular-regulation includes two subtypes: 1) increase first, reaching a high threshold, and then decrease to a level lower than the first tapping; 2) decrease first, reaching a low threshold, and then increase to a level higher than the first tapping. Accordingly, the DE-protein spots were also classified into the three types according to their expression dynamics across the first five tappings in reoppend resting *Hevea* trees (**Figure 2**). A protein spot was regarded as a DE one when its abundance varies ≥3-fold between the first tapping and any of the four other tappings. As a result, a total of 117 tapping-responsive DE-protein spots were identified, and picked out from the 2-DE gels for mass spectrometry analysis.

DE-TDFs and DE-Proteins Annotation and Redundancy Removal

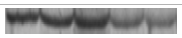
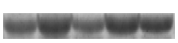
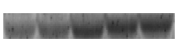
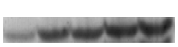
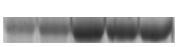
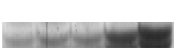
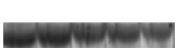
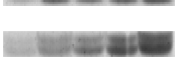
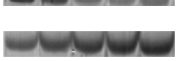
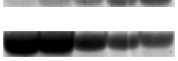
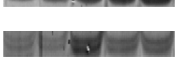
The DE-TDFs were made clean by removing the sequences of vector, primer, and adaptor at both ends and then annotated by Blastx online searching (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the NCBI non-redundant protein database (nr), with an E-value threshold of $<10^{-4}$ and a score value of >50. According to Blastx search against the *Hevea* latex transcriptome database (Tang et al., 2016), the DE-TDFs belonging to the same transcript and sharing similar expression pattern in the cDNA-AFLP analysis were clustered together, and only the longest TDF was retained. As a result, a total of 505 non-redundant DE-TDFs (GenBank accession nos MZ935745—MZ936248) were obtained, including 217 (43.0%) upregulated (**Supplementary Table S1**), 180 (35.6%) downregulated (**Supplementary Table S2**) and 108 (21.4%) irregularly-regulated (**Supplementary Table S3**). The selected DE-protein spots were identified by a Bruker's Ultraflex TOF/TOF mass spectrometer. The proteins identified as positive by peptide mass fingerprinting (PMF) analysis were further subjected with 2-4 matched peptides to peptide fragmentation fingerprinting (PFF) investigation. Of the 117 DE-protein spots examined, 89 were identified as positive by the PMF analysis, and 57 were further determined as positive by the PFF analysis. Of the 89 positively identified DE-proteins, 35 (39.3%) were classified as upregulated, 20 (22.5%) as downregulated and 34 (38.2%) as irregularly-regulated (**Supplementary Table S4**).

Detailed comparison revealed that the DE-protein spots of three small rubber particle proteins (nos. 163, 723 and 730), one tubulin alpha-3 chain-like protein (no. 552) and two latex abundant family proteins (nos. 554 and 628) had their DE-TDF counterparts (M1-A10-2, M12-A6-4, M6-A10-4, M6-A7-1 and M8-A7-2) (**Supplementary Tables S1, S2, S4, S5**), of which only one pair (protein spot 163 and TDF #M1-A10-2) revealed a consistency of upregulation. The other nine DE-TDFs overlapped in functional annotations with a number of DE- protein spots, belonging to distinct genes within the same gene families (**Supplementary Table S5**). As reported in studies covering various types of organisms (Taniguchi et al., 2010; Vogel and Marcotte, 2012; Walley et al., 2013), low overlap and poor correlation have been frequently observed between transcriptomic and proteomic data due to several possible explanations. One more plausible explanation for the low overlap revealed here is the detection limit of the proteome technique that is capable of profiling only medium to high expressed proteins (Humphery-Smith et al., 1997) in comparison with the high transcript-detecting sensitivity of the cDNA-AFLP technique (Vuylsteke et al., 2007; Xiao et al., 2009).

Functional Classification of Non-redundant DE-TDFs and DE-Proteins

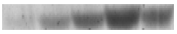
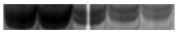
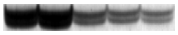
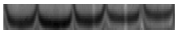
According to the results of Blastx search, the 505 non-redundant DE-TDFs were divided into four categories: 1) Proteins with clear functional annotation; 2) Unclassified proteins, the functional annotation of the proteins being multiple; 3) Predicted protein, showing high homology with a predicted protein in the database; 4) No hit, no homologous sequence in the database. Most (366, 72.5%) of these DE-TDFs were homologous to genes with known functions, whereas 30 (5.9%) and 59 (11.7%), respectively, belonged to unclassified proteins and predicted proteins and the remaining 50 (9.9%) with no hit (**Table 1**). Meanwhile,

TABLE 2 | Semi-quantitative RT-PCR validation of DE-TDFs for the profiles in virgin *Hevea* trees.

DE-TDFs	cDNA-AFLP	sqRT-PCR[a]	Consistency[b]	DE-TDFs	cDNA-AFLP	sqRT-PCR[a]	Consistency[b]
M1-A5-1		1.00, 1.09, .91, .80, .76	+	M9-A9-1		1.00, 1.42, 2.48, 3.18, 4.93	+
M1-A6-2		1.00, 1.16, .89, 2.85, 3.23	+	M9-A10-1		1.00, 1.13, .78, 0.61, .56	+
M1-A6-3		1.00, 1.10, 1.19, 1.40, 1.32	+	M9-A10-2		1.00, 1.69, 1.15, .88, .59	+
M1-A7-1		1.00, 0.94, 0.47, 0.76, 0.51	+	M9-A11-1		1.00, 1.02, .88, .61, 0.44	+
M1-A7-2		1.00, 1.00, .94, .85, 0.18	+	M9-A11-2		1.00, 1.11, 1.26, 1.33, 1.08	+
M1-A7-3		1.00, 1.13, 1.27, 1.52, 1.82	+	M9-A11-3		1.00, 1.02, .84, .75, .66	+
M1-A7-4		1.00, .94, .65, .73, .55	+	M9-A12-1		1.00, 1.08, 1.22, 1.47, 1.57	+
M1-A7-5		1.00, 1.13, 1.24, 1.26, 1.35	+	M10-A8-2		1.00, 1.18, .74, .71, .65	+
M1-A8-1		1.00, 1.29, 1.31, 1.55, 1.71	+	M10-A9-1		No band	–
M1-A8-2		No band	–	M10-A9-3		1.00, .96, .79, .77, .76	+
M1-A9-1		1.00, 1.02, 1.06, 1.24, 1.23	+	M10-A10-1		1.00, 1.13, 1.42, 1.70, 2.05	+
M1-A9-2		1.00, 1.11, .86, .92, 1.03	–	M10-A11-1		1.00, 1.11, .62, .51, .33	+
M1-A9-3		1.00, .91, 1.04, 1.23, 1.27	–	M10-A11-2		1.00, .78, .48, .40, .20	+
M1-A10-1		1.00, .76, .65, .40, .36	+	M10-A12-1		1.00, 1.56, 1.69, 2.45, 2.47	+
M1-A10-2		1.00, 1.44, 2.91, 3.42, 3.81	+	M10-A12-2		1.00, 1.12, .62, 0.59, .55	+
M1-A10-3		1.00, 1.03, 1.27, 1.40, 2.03	+	M11-A7-1		1.00, 1.14, 1.57, 1.59, 1.80	+
M1-A11-1		1.00, .95, .75, .75, .74	+	M11-A8-1		1.00, .88, .58, .55, .46	+
M1-A11-2		No band	–	M11-A8-2		1.00, 1.32, 1.69, 2.25, 2.20	+
M1-A11-3		1.00, .97, .81, .81, .78	+	M11-A9-1		1.00, .89, .67, .65, .57	+
M1-A11-4		1.00, .93, .87, .94, 1.00	–	M11-A9-2		1.00, 1.07, .77, 0.77, 0.66	+
M1-A12-2		1.00, 1.00, .82, .95, .96	+	M11-A10-1		1.00, 1.18, 1.15, 1.16, 1.19	–
M2-A5-1		1.00, 1.34, 1.64, 1.91, 1.97	+	M15-A5-1		1.00, 1.25, 1.61, 1.77, 2.27	+
M2-A5-2		1.00, 1.04, 1.07, 1.12, 1.13	+	M15-A8-1		1.00, 1.46, 1.85, 2.03, 2.48	+
M2-A6-2		1.00, 1.12, .86, .66, 0.65	+	M15-A8-4		1.00, 1.03, 1.18, 1.36, 1.67	+
M2-A7-1		1.00, 1.01, 1.18, 1.57, 2.22	+	M15-A9-1		1.00, .78, .51, .49, .49	+
M2-A7-3		1.00, 1.38, 2.51, 3.04, 3.78	+	M15-A10-1		1.00, 2.09, 1.61, 1.49, 1.19	–
M2-A7-5		1.00, 1.04, 1.23, 1.29, 1.63	+	M15-A10-3		1.00, 1.06, 1.27, 1.45, 1.66	+
M2-A8-1		1.00, 1.12, 1.22, 1.46, 1.31	+	M15-A10-4		1.00, .98, .61, .64, .40	–
M2-A8-2		1.00, .99, .94, 1.18, 1.08	–	M16-A5-1		1.00, .80, .58, 0.55, 0.51	+
M2-A9-1		1.00, 1.70, 1.85, 2.06, 2.11	+	M16-A6-1		1.00, 1.06, 1.12, 1.51, 1.72	+
M2-A9-3		1.00, .97, .86, .83, 0.81	+	M16-A7-1		1.00, 1.12, 1.07, 1.41, 2.23	–
M2-A10-1		1.00, 1.56, 2.40, 3.34, 3.94	+	M16-A8-1		1.00, 1.30, 1.41, 1.69, 1.20	+

(Continued on following page)

TABLE 2 | (*Continued*) Semi-quantitative RT-PCR validation of DE-TDFs for the profiles in virgin *Hevea* trees.

DE-TDFs	cDNA-AFLP	sqRT-PCR[a]	Consistency[b]	DE-TDFs	cDNA-AFLP	sqRT-PCR[a]	Consistency[b]
M2-A10-2		1.00, .88, .47, .68, 0.41	+	M16-A9-1		1.00, .56, .76, 1.03, 1.22	–
M2-A10-3		1.00, .99, 1.12, 1.33, 1.22	+	M16-A9-2		1.00, .99, .85, .68, .49	+
M3-A7-1		1.00, 1.06, .69, 0.68, 0.41	+	M16-A10-1		1.00, .94, .71, .65, 0.44	+
M3-A10-3		1.00, 1.60, 2.12, 2.51, 2.79	+	M16-A10-2		1.00, 1.07, 1.23, 1.54, 1.88	+
M3-A11-1		1.00, .74, .37, .45, 0.50	+	M16-A11-1		1.00, .85, .70, .62, .61	+
M9-A6-1		1.00, 1.07, 1.09, 1.30, 1.65	+	M16-A11-2		1.00, 1.12, 1.14, 1.22, 1.15	+
M9-A7-1		1.00, 1.39, 1.71, 2.30, 2.85	+	M16-A12-1		1.00, .81, 1.10, 1.22, 2.04	–
M9-A8-1		1.00, .95, .63, .61, .61	+	M16-A12-3		1.00, .83, .52, .50, .47	+

[a]*With the scanned intensity for the TDF band of the first tapping set as 1.00, the relative expression of DE-TDFs were calculated for the first five tappings.*
[b]*"+" means that the result of sqRT-PCR is consistent with that of cDNA-AFLP analysis; "–" indicates inconsistency between the two analysis.*

the 89 positive DE-protein spots were divided into two categories: 1) Proteins with clear functional annotation; 2) Unknown proteins. Most (78, 87.6%) of these DE-protein spots were functionally annotated, whereas the remaining 11 (12.4%) belonged to predicted or hypothetical proteins (**Table 1**).

With reference to the functional categories of plant genes defined by Bevan et al. (1998), the 366 DE-TDFs and 78 DE-proteins with known functions were classified, respectively, into 11 and 7 functional categories, among which a new category, rubber biosynthesis, was singled out from "secondary metabolism" (**Figure 3**; **Table 1**). Of these DE-TDFs, five categories including cell growth and division, protein degradation and storage, cellular structure, secondary metabolism, and rubber biosynthesis had a higher portion of upregulated DE-TDFs than that of the down- and irregularly-regulated DE-TDFs together (**Table 1**). Noteworthily, nine of the ten DE-TDFs implicated in rubber biosynthesis were upregulated, and the remaining one is irregularly-regulated. Of the 89 DE-protein spots, the number of upregulated (35) was significantly larger than the downregulated (20), but similar to the irregular-regulated (34). Expression patterns of the DE-protein spots varied in most functional categories, whereas all four in the signal transduction category were upregulated. It is worth noting that there were many cases that more than one protein spots corresponded to an identical gene, e.g., six spots (nos. 33, 229, 516, 534, 539, and 591) for an enolase, five (nos. 149, 163, 723, 730, and 858) for a small rubber particle protein, and five (nos. 688, 778, 779, 789, 797, and 803) for a heat shock protein (**Supplementary Table S4**), indicating extensive posttranslational modifications in *Hevea* latex proteins.

Validation of Expression Pattern by sqRT-PCR

To determine the reliability of the cDNA-AFLP results, 80 DE-TDFs, covering >10% of the DE-TDFs we identified were randomly selected from each functional category and subjected to semi-quantitative reverse transcription PCR (sqRT-PCR) analysis using specific primers for TDFs with *18S rRNA* as the reference gene. The ImageJ software was used to quantify the sqRT-PCR amplicons fractionated in agarose gel electrophoresis, with the value in the first tapping sample taken as 1.0 (**Table 2**). About 84% (67 TDFs) showed expression patterns consistent with their cDNA-AFLP gel profiles, indicating the high reliability of the cDNA-AFLP screening.

qRT-PCR Analysis of Latex Regeneration-Related DE-TDFs

A total of 29 DE-TDFs implicated in latex-regeneration were investigated by quantitative RT-PCR (qRT-PCR) analysis for their expression patterns across the five successive tappings. About 90% of the qRT-PCR results were consistent with their original cDNA-AFLP expression profiles (**Figure 4**; **Table 3**). The genes of these DE-TDFs are putatively involved in the pathways of primary metabolism, rubber biosynthesis and regulation, transporters, and intracellular transport. Of the ten rubber biosynthesis pathway DE-TDFs, nine revealed qRT-PCR patterns similar to their cDNA-AFLP results (**Figure 4**; **Table 3**).

DISCUSSION

Functional Categories With Reference to Tapping-Stimulated Latex Production

Tapping can stimulate latex production, especially in virgin and reopened resting *Hevea* trees (Pakianathan et al., 1992; Tang et al., 2010; Qi, 2011). A number of early studies have shown that the first few tappings greatly stimulate the metabolism of laticifers, accompanied by the enhanced expression of several specific genes involved in latex production (Dennis et al., 1989; Adiwilaga and Kush, 1996; Gohet, 1996; Priya et al., 2007). The latex flows out of

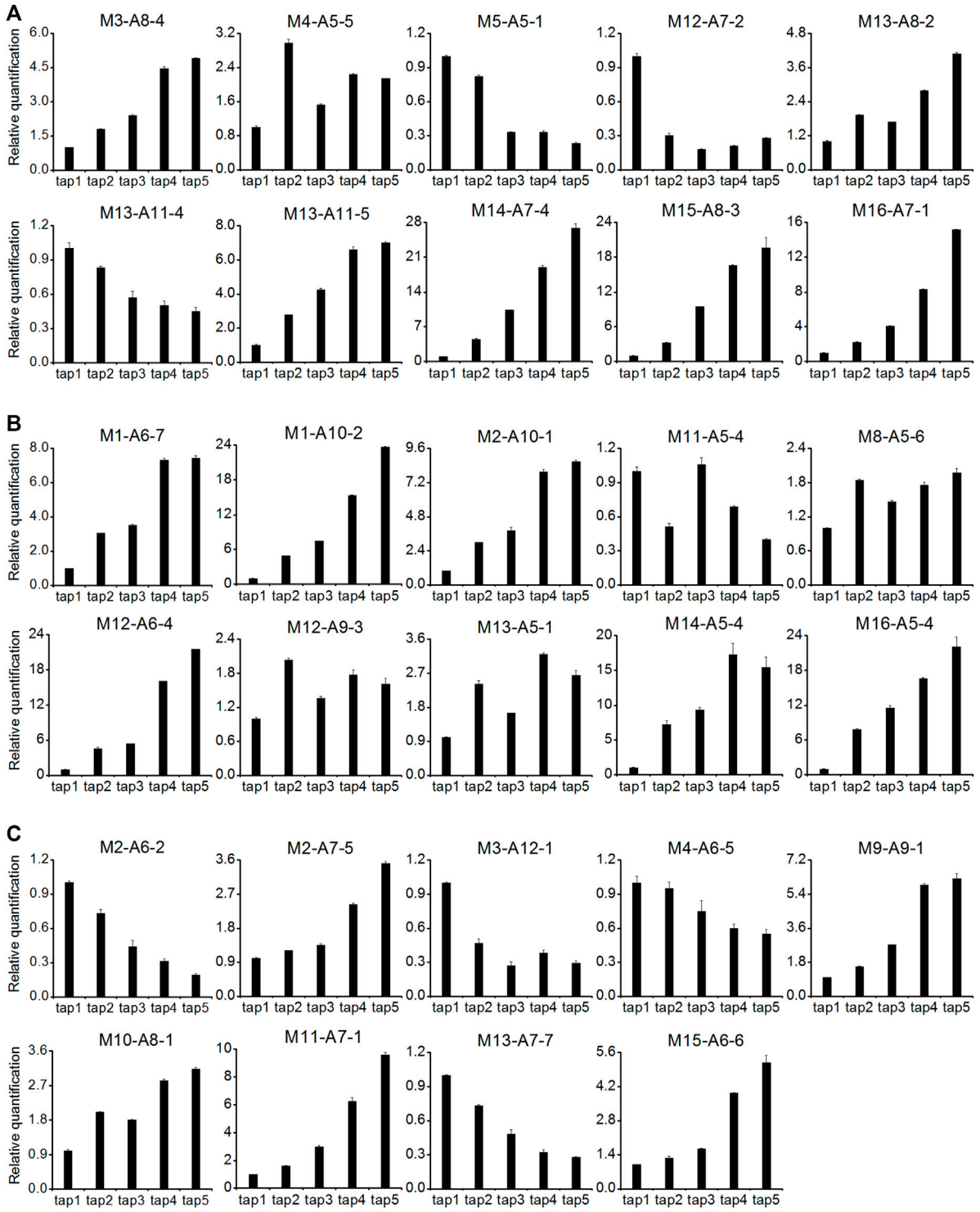


FIGURE 4 | qRT-PCR analysis for expression of latex regeneration genes in the latex of the first five tappings. **(A)**: primary metabolism pathway (ten DE-TDFs); **(B)**: rubber biosynthesis and regulation pathway (ten DE-TDFs); **(C)**: transporters and intracellular transport pathway (nine DE-TDFs). Except for M3-A8-4, M11-A5-4 and M16-A7-1, 26 of the 29 DE-TDFs shows the results consistent with their original cDNA-AFLP expression patterns and were presented here. All data were normalized to the expression level of *HbYLS8* gene. Values are averages ± STDEV of three technical replicates. The latex samples of three individual trees were analyzed with similar patterns of expression, and one representative was shown.

TABLE 3 | Information of latex regeneration related DE-TDFs for qRT-PCR analysis.

DE-TDFs	GenBank blastx annotation	Model	Consistency[a]
Primary metabolism (10)			
M3-A8-4	pyruvate dehydrogenase, putative	Irregular	–
M4-A5-5	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase, putative	Irregular	+
M5-A5-1	G6PD1 (glucose-6-phosphate dehydrogenase)	Irregular	+
M12-A7-2	Phosphofructokinase, putative	Irregular	+
M13-A8-2	pyruvate dehydrogenase, putative	Up	+
M13-A11-4	mitochondrial pyruvate dehydrogenase kinase isoform 2	Down	+
M13-A11-5	pyruvate kinase, putative	Up	+
M14-A7-4	Fructokinase, putative	Up	+
M15-A8-3	glyceraldehyde-3-phosphate dehydrogenase	Up	+
M16-A7-1	neutral/alkaline invertase	Up	–
Rubber biosynthesis and regulation (10)			
M1-A6-7	cis-prenyltransferase	Up	+
M1-A10-2	small rubber particle protein	Up	+
M2-A10-1	rubber elongation factor	Up	+
M8-A5-6	inorganic pyrophosphatase, putative	Irregular	+
M11-A5-4	farnesyl diphosphate synthase	Up	-
M12-A6-4	small rubber particle protein	Up	+
M12-A9-3	3-hydroxy-3-methylglutaryl-coenzyme A reductase 2	Up	+
M13-A5-1	rubber elongation factor	Up	+
M14-A5-4	rubber elongation factor	Up	+
M16-A5-4	hydroxymethylglutaryl coenzyme A synthase	Up	+
Transporters and intracellular transport (9)			
M2-A6-2	AGD13 (ARF-GAP domain 13); ARF GTPase activator/zinc ion binding	Irregular	+
M2-A7-5	vacuolar ATP synthase subunit G plant, putative	Up	+
M3-A12-1	RAB6A; GTP binding/protein binding	Irregular	+
M4-A6-5	small Ras-like GTP-binding protein	Down	+
M9-A9-1	AtRABA1f (Arabidopsis Rab GTPase homolog A1f); GTP binding	Up	+
M10-A8-1	sugar transporter, putative	Up	+
M11-A7-1	vacuolar ATP synthase proteolipid subunit 1, 2, 3, putative	Up	+
M13-A7-7	prenylated Rab acceptor protein, putative	Down	+
M15-A6-6	AtRABH1e (Arabidopsis Rab GTPase homolog H1e); GTP binding	Up	+

[a]*"+" means that the expression pattern of DE-TDFs by qRT-PCR are consistent with the cDNA-AFLP gel profiles, and "–" means that the expression pattern are inconsistent.*

laticifers after tapping, and in order to compensate for the loss of cytoplasm (latex) and maintain the balance of intracellular metabolism, the laticifers require large amounts of RNA and proteins to be synthesized before the next tapping. Of the 366 DE-TDFs identified in the latex with known functions (**Table 1**), 26.2% were classified into the functional category of transcription and protein synthesis (**Figure 3**), representing the largest category, 42.7% of which were upregulated by the tapping treatment. These results indicated that tapping significantly affects the ways of laticifers to synthesize RNA and proteins, providing a prerequisite for multiple subsequent biological responses to the tapping treatment. Of the 78 DE-protein spots identified in the latex with known functions, 29.5% were classified into the functional category of stress and defense (**Table 1**) and represented the largest category, corresponding well to the defense functions of *Hevea* laticifers (Sharples, 1918; Chow et al., 2007). Laticifers are believed to be a defense system for *Hevea* trees to cope with biotic and abiotic stresses, and the latex exuded after bark wounding has been found to play roles in resisting pathogen infection, insect feeding and abiotic stress (Chow et al., 2007). The tapping itself is a kind of abiotic stress upon *Hevea* trees. Consistent with the proteome study, "stress and defense" also accounted for a large portion of the functional DE-TDFs, ranking the second place among the DE-TDF functional categories (**Figure 3**; **Table 1**; **Supplementary Table S4**). The harvesting stress response has been suggested to be one of the key factors affecting latex production and rubber productivity in *Hevea* trees (Pirrello et al., 2014).

The category of transporters and intracellular transport was the third largest among the 11 functional categories, accounting for 12.3% of the total functional DE-TDFs (**Figure 3**). This agrees well with the sink effect caused by the large loss of latex after tapping. The process of regenerating the expelled latex involves the synthesis, transport, loading and subcellular localization of a large number of organelles, proteins, nucleic acids, sugars, etc., all of which require the active involvement of transporters and intracellular transport-related proteins (d'Auzac et al., 1997; Tungngoen et al., 2011). DE-TDFs involved in signal transduction were also highly represented, accounting for a proportion of 8.3% for the total DE-TDFs (**Table 1**), and all four DE-proteins involved in signal transduction were upregulated (**Table 1**; **Supplementary Table S4**). A variety of signaling pathways within *Hevea* laticifers, including ethylene, jasmonic acid and wound signaling, have been reported to be extensively participate in latex regeneration and regulation (Dennis et al., 1989; Hao et al., 2004; Duan et al., 2010; Lacote et al., 2010; Pirrello et al., 2014; Tang et al., 2016). The proportions for the two categories, protein degradation and

storage and primary metabolism were also high, covering, respectively, 9.3% and 8.2% of the total functional DE-TDFs (**Figure 3**; **Table 1**). Their high representation suggested that with the progress of tapping, in order to meet the balance of supply and demand of all substances in latex regeneration, protein turnover rate becomes faster and primary metabolism gets active. In a word, these results indicated that the latex production induced by tapping involves a complex multi-gene regulatory network, as well as multiple physiological and biochemical response processes.

Sugar Metabolism and Rubber Biosynthesis in Tapping-Stimulated Latex Production

In regularly tapped *Hevea* trees, the main metabolic activity of the laticifers is latex regeneration, which centers on the biosynthesis of NR that consists of about 90% of the latex dry weight (d'Auzac et al., 1997). Sucrose is the precursor material for rubber biosynthesis in laticifers, and provides the carbon skeleton and energy required for latex regeneration (Tupý, 1989; Duangngam et al., 2020). In *Hevea* trees that are tapped at two-four days of intervals, a tree produces dozens to hundreds of milliliters of fresh latex, and the removed latex could be effectively recovered before the next tapping (d'Auzac et al., 1997; Tang et al., 2010). Therefore, the laticifers represent an active carbon sink, and the effective supply of sucrose is a key factor determining latex production (Tupý, 1985; Chantuma et al., 2009). In this study, the genes of a sucrose transporter (M10-A9-1) and a sugar transporter (M10-A8-1) were among the DE-TDFs identified, both of which were significantly upregulated with the increase of tappings (**Supplementary Table S1**). Interestingly, the sucrose transporter (M10-A9-1) identified here is the HbSUT3 we previously reported to be critical in sucrose uptake into laticifers and *Hevea* rubber production (Tang et al., 2010). The upregulation of these two transporters indicated an active involvement of sucrose and sugar transport in tapping-stimulated latex production. Sucrose catabolism and the following pathways including glycolysis, tricarboxylic acid cycle and pentose phosphate pathway provide essential components, i.e., the carbon skeleton (acetyl CoA), the reducing power (NADPH) and the energy (ATP) for the final rubber biosynthesis pathway (Tupý, 1989; d'Auzac et al., 1997). Therefore, sugar metabolism becomes one of the core metabolic pathways contributing to latex production in *Hevea* (Tupý, 1989; d'Auzac et al., 1997; Silpi et al., 2007). This study identified multiple DE-TDFs involved in sucrose cleavage and the three above mentioned sugar metabolism pathways (**Table 3**; **Supplementary Tables S1–S3**), including neutral/alkaline invertase (M16-A7-1), fructokinase (M14-A7-4), phosphofructokinase (M12-A7-2), glyceraldehyde 3-phosphate dehydrogenase (M15-A8-3), pyruvate kinase (M13-A11-5), pyruvate dehydrogenase (M13-A8-2), and glucose-6-phosphate dehydrogenase (M5-A5-1), etc. Most of these DE-TDFs were upregulated in the latex for the first few tappings (**Figure 3**; **Supplementary Table S1**). It is worth noting that the upregulated DE-TDF (M16-A7-1) as identified by both cDNA-AFLP (**Supplementary Table S1**) and qRT-PCR (**Figure 4**) turned out to be *HbNIN2*, the neutral/alkaline invertase that is responsible for sucrose catabolism in *Hevea* laticifers (Liu et al., 2015). Here, the proteomic research also backed up the importance of sugar metabolism in tapping-stimulated latex production in reopened resting *Hevea* trees although a low overlap was observed among the DE-genes and DE-proteins we identified this study. Among the 18 DE-protein spots identified in category of primary metabolism and energy, 13 were involved in sugar metabolism (**Supplementary Table S4**). Notably, nine spots (nos. 356, 372, 593, 33, 229, 516, 535, 539, and 591) were implicated in the glycolytic pathway. By contrast, no DE-protein spots were in tricarboxylic acid (TCA) cycle, collaborating the proposition of the relative importance of glycolysis versus the TCA cycle in sugar degradation in hypoxic *Hevea* latex (d'Auzac and Jacob, 1969; Tupý and Resing, 1969) and thus in tapping-stimulated latex production.

There are 20 gene families directly involved in the NR biosynthesis and termed as rubber biosynthesis (RB) genes (Tang et al., 2016; Chow et al., 2007). In this study, a total of nine DE-TDFs involving six such families were identified, including cis-prenyltransferase (M1-A6-7), hydroxymethylglutaryl coenzyme A synthase (M16-A5-4), 3-hydroxy-3-methylglutaryl-coenzyme A reductase (M12-A9-3), farnesyl diphosphate synthase (M11-A5-4), REF/SRPP proteins (M2-A10-1, M13-A5-1, M14-A5-4, M1-A10-2, M12-A6-4) (**Table 3**). Among the nine DE-TDFs, eight were demonstrated by qRT-PCR to be upregulated with the tappings (**Figure 4**). In addition, a DE-TDF (M8-A5-6) annotated as inorganic pyrophosphatase, a vacuolar type of which being located on rubber particles and essential for IPP incorporation into elongating rubber hydrocarbon molecules (Zeng et al., 2009), was also bolstered by the tapping treatment (**Figure 4**). The tapping treatment also changes the expression of RB genes at the protein levels in reopened resting *Hevea* latex. A total of eight relevant DE-protein spots were identified, corresponding to REF/SRPP proteins and an acetyl-CoA C-acetyltransferase (**Supplementary Table S4**).

Strength and Weakness of the cDNA -AFLP Technique

The cDNA-AFLP technique has been widely applied in various eukaryotes including the *Hevea* tree for transcript profiling due to its advantages of stringency, reproducibility, cost-effectiveness, genome-wide coverage and the ability to distinguish among highly homologous genes (Ko et al., 2003; Vuylsteke et al., 2007; Ganeshan et al., 2011; Tang et al., 2013; Xiao et al., 2016). In this study, as determined by sqRT-PCR and qRT-PCR, about 84% and 90%, respectively, of the selected DE-TDFs were verified for their cDNA-AFLP profiles (**Figure 4**; **Tables 2**, **3**), reflecting a high reliability of this technique in screening tapping-responsive DE-TDFs in *Hevea* latex. According to a previous *in-silico* estimation (Xiao et al., 2009), about 84% of the genes expressed in *Hevea* latex could be visualized using the silver-staining cDNA-AFLP technique with the restriction enzyme pair of *Apo* I and *Mse* I exploited here. The sucrose transporter HbSUT3 (Tang et al., 2010) and the

neutral/alkaline invertase HbNIN2 (Liu et al., 2015) that have been reported to be upregulated in the latex of virgin *Hevea* trees by the tapping treatment were among the DE-TDFs identified in this study (**Figure 4**; **Table 3**; **Supplementary Table S1**), demonstrating a high transcript coverage of this technique. However, compared with the currently popularly used next generation RNA-sequencing technique that relies on expensive DNA sequencers and specialized bioinformatics, the cDNA-AFLP is labor-intensive. Nevertheless, the cDNA-AFLP technique still have its niche among the various transcript profiling techniques, and can be readily established in a mediocrely equipped and stringently funded lab to fulfill its customized transcript profiling task.

CONCLUSION

An integrative transcriptome and proteome analysis was conducted to identify important regulators and pathways participated in tapping stimulated latex production in *Hevea* trees. A total of 505 tapping-responsive DE-TDFs and 89 DE-proteins were identified in the rubber-producing laticifers of *Hevea* trees (virgin or reopened). The low overlap between the DE-TDFs and DE-protein spots identified is indicative of posttranscriptional regulation and the strong complementarities of transcriptome and proteome analysis. According to the 366 DE-TDFs and 78 DE-proteins with functional annotations, the tapping treatment brought about extensive physiological and molecular changes in laticifers. The integration of these changes, especially those of sugar metabolism and rubber biosynthesis, upgraded the mediocre level of laticifer metabolism in virgin or reopened trees to a high dynamic equilibrium of latex regeneration in regularly tapped trees. Further integrative studies will benefit a deeper insight into the exact relationships (synergy or antagonism) among the vast number of biological pathways implicated in tapping-stimulated latex production.

MATERIALS AND METHODS

Plant Materials

Hevea trees of Reyan7-33-97 clone for cDNA-ALFP transcript profiling (virgin trees) and 2-DE protein profiling (resting trees) analysis, were all planted in the experimental field of Chinese Academy of Tropical Agricultural Sciences (Danzhou, Hainan). i.e., removing a slice of trunk bark by a special knife, with a half spiral tapping system, every 3 days, and with no ethylene stimulation (Tang et al., 2010). The virgin trees were planted for eight years and first subjected to the tapping. The resting trees (after four months of no tapping) were planted for 10 years, and re-tapped for the third year. Because the virgin rubber trees produced little latex in the first tapping that could not satisfy the requirement of C-serum preparation for 2-DE analysis as described below, re-tapped resting trees were therefore exploited. Such *Hevea* trees produce much more latex than virgin trees in the first tapping and in the subsequent several tappings the latex production increases in a pattern similar to that observed in virgin trees albeit not striking.

Extraction of Latex Total RNA

Three batches of five rubber trees attaining the tapping standard (trunk girth >= 50 cm at 1 m above the ground) were selected for latex collection and RNA extraction. Twenty seconds after tapping, about 5 ml of latex was allowed to flow into a centrifuge tube containing 5 ml 2×RNA extraction buffer (.3 M LiCl, 10 mM EDTA, 10% SDS, 100 mM Tris-HCl, pH8.0). The collected latex was placed in ice box and brought to laboratory for RNA extraction as described in Tang et al. (2007). Electrophoresis on a 1.5% formaldehyde denaturing agarose gel was used to detect the integrity of RNA samples. RNA samples from each batch of rubber trees for the five tappings were used as one biological replicate.

cDNA-AFLP Analysis

A total of 50 μg latex total RNA taken from each of the five tapping samples was subjected to cDNA-AFLP analysis. The detailed manipulations were conducted according to the silver-staining cDNA-AFLP procedure which we previously established suitable for *Hevea* latex transcriptome profiling (Xiao et al., 2009). Briefly, the synthesized double-stranded cDNA was cut by the restriction enzymes of *Apo* I and *Mse* I (Thermo Fisher Scientific, Vilnius, Lithuania), and ligated with adaptors. The ligation product, termed primary template, was used directly for pre-amplification. The pre-amplification product, termed secondary template, was then used for selective amplification. All the 128 possible selective primer combinations with 8 *Apo* I primers and 16 *Mse* I primers as reported previously (Xiao et al., 2009) were applied in screening DE-TDFs affected by the tapping treatment in mature virgin *Hevea* trees.

DE-TDFs Extraction and Amplification

With the expression level of the TDFs at the first tapping as a reference, the DE-TDFs were identified for the tapping treatment and isolated from the polyacrylamide gel. The DE-TDFs was scraped from the polyacrylamide gel with a surgical blade, put into a sterile PCR tube containing 30 μl .1× TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH8.0), heated at 95°C for 15 min, kept overnight at 4°C, and centrifuged at 10,000 g for 5 min to collect the dissolved DNA solution for PCR amplification. A total of 25 μl PCR reaction mixture includes: 2 μl DNA solution, 2.5 μl 10× PCR buffer (plus Mg^{2+}), 1.2 μl 2.5 mM dNTPs Mix, .5 μl 50 ng/μl *Mse*Iselective primer, .5 μl 50 ng/μl *Apo*Iselective primer, .2 μl 5U/μl *Taq* DNA Polymerase (Takara, Dalian, China) and 18.1 μl ddH_2O. The PCR amplification procedure is the same as the pre-amplification in cDNA-AFLP analysis (Xiao et al., 2009). PCR products were fractionated by 1.2% agarose gel electrophoresis, and the target band was sliced and purified using AxyPrep^{TM} DNA Gel Extraction Kit (AxyGen, shanghai, China).

DE-TDFs Cloning and Sequencing

The purified DE-TDFs were ligated with the T-vectors using the pMD18-T Vector Kit (Takara, Dalian, China) in accordance with

the manufacture's manual. The ligation mixture was used to transform *E. coli* JM109 competent cells and the transformants were sent to BGI Genomics Co., Ltd. (Shenzhen, China) for sequencing.

Semi-Quantitative Reverse Transcription PCR

The first strand of cDNA was synthesized by reverse transcriptase kit (RevertAid^{TM} First Strand cDNA Synthesis Kit-K1622, Thermo Fisher Scientific, Vilnius, Lithuania), and then diluted ten times as the template for sqRT-PCR with *18S rRNA* used as the reference (Xiao et al., 2009). The annealing temperatures for the PCR amplification were set according to the primers designed for each specific gene. The PCR cycles were controlled to keep the amplification under the plateau phase. All of the amplified fragments were cloned, and were sequenced for target confirmation. The amount of cDNA samples used in sqRT-PCR was adjusted to be the same for the five tappings based on the level of *18S rRNA* expression, because it is conserved for sqRT-PCR in most of the plant species (Morris and Davila, 1996).

Quantitative RT-PCR

The expression pattern of candidate genes was detected by qRT-PCR. The first strand of cDNA was diluted 20 times as the template for qRT-PCR with *HbYLS8* as the reference gene which was recommended in our previous study as a suitable internal standard gene in qRT-PCR analysis for the tapping treatment in rubber tree (Li et al., 2011). The PCR reaction mixture includes: 2 μl template, .3 μl each for 10 μM forward and reverse primers, 10 μl 2×SYBR® Premix Ex *Taq*^{TM}II (Takara, Dalian, China) and 7.4 μl ddH_2O. Roche's LightCycler 2.0 system was used for qRT-PCR analysis with the program as follows: 95°C 30 s; 94°C 5 s, 60°C 20 s, 72°C 20 s, 45cycles. Three technical replicates were analyzed for each of the three biological samples. All the cycle threshold (Ct) values from one gene were determined at the same threshold fluorescence value of .2 using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The primers of target and reference genes were listed in **Supplementary Table S6**. Student's *t*-test was performed using Statistical analysis was performed using Student's t-test.

Extraction and Dissolution of Latex C-Serum Proteins

Three batches of five rubber trees were selected for latex collection and protein extraction. After tapping, sixty drops of latex were discarded and ~30 ml latex per tree was then collected in a 50 ml centrifuge tube placed in an ice bath. The collected latex was immediately taken back to the laboratory, and centrifuged at 22, 000 g 4°C for 2 h to acquire the middle aqueous layer, latex cytosol (C-serum). The C-serum proteins were extracted by TCA/acetone-diethyl method (Saravanan and Rose, 2004), freeze-dried, and stored at −80°C. Protein dissolution was performed by incubating for 1 h at 4°C with ultrasonic treatment for 5 min (2s on, 3s off, 15W), with a proportion of 1 mg freeze-dried protein powder added into 10 μl lysis buffer (7M urea, 2M thiourea, 4% CHAPS, 2 mM TBP, 65 mM DTT, .2% w/v IPG buffer). The insoluble precipitate was removed by centrifugation at 20°C 15, 000 g for 10 min, and the supernatant was used for protein separation by two-dimensional electrophoresis (2-DE).

Protein Separation by 2-DE

The protein content was determined according to Bradford (1976). For each of the first five tappings conducted on re-tapped resting *Hevea* trees, 1 mg of latex C-serum proteins were dissolved in hydration loading buffer [7M urea, 4% CHAPS, 65 mM DTT, .2% Bio-Lyte pH3/10 or pH5/8 ampholyte, .002% (W/V) bromophenol blue] to a final volume of 300 μl. The protein mixture was centrifuged at 20°C 15, 000 g for 20 min to remove bubbles and undissolved precipitates, and then loaded into the IPG strips (Bio-Rad, USA) as described in Sanchez et al. (1997). The strips were then subjected to isoelectric focusing and SDS-PAGE, in an Ettan IPGphor and DALT system according to the manufacturer's 2-DE manual (GE Healthcare). Three biological replicates for each tapping were conducted.

Gel Staining, Image Analysis and Mass Spectrometry

The gels were stained by a modified CBB R250 staining protocol (Wang et al., 2007), and images were scanned with a Bio-Rad GS-800TM calibrated optical density scanner at 600 dpi. Images were then analyzed with Image Master 2D Platinum 5.0 (GE Healthcare) to integrate three replicates of the protein spots for each tapping. With the expression of the protein spot at the first tapping as a reference, the DE-protein spots ($|ratio| \geq 3.0$, T value $\leq .05$)) were isolated from the SDS-PAGE gel, and sent to BPI Co., Ltd. (Beijing, China) for mass spectrometry identification on an ultraflex MALDI-TOF/TOF mass spectrometer instrument (Bruker Daltonics, Billerica, MA, United States).

Bioinformatics Analysis

For DE-TDFs analysis, sequences of vectors and adaptors were first trimmed off by using the VecScreen program on the NCBI website (https://www.ncbi.nlm.nih.gov/tools/VecScreen). Then, the clean DE-TDF sequences were subjected for homology analysis to publicly available GenBank non-redundant sequences databases (http://www.ncbi.nlm.nih.gov) using the BLASTX program. Also, the Gene Ontology (http://amigo1.geneontology.org/cgi-bin/amigo/go.cgi) database was used to investigate the molecular function of each DE-TDF in the cell, which was used as the basis for functional classification.

Statistical Analysis

At least three individual virgin or resting *Hevea* trees, with one tree serving as one biological replicate, were exploited. Data were analyzed by one-way ANOVA (SAS6.11) for comparison of TDF expressions across the five tappings. Student's t-test was performed using the software embedded in Excel 2007 for comparison of TDF or protein spot abundance between the first tapping and any of the other four tappings. Differences were accepted as significant at $p < .05$ or .01.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI under accession numbers MZ935745-MZ936248.

AUTHOR CONTRIBUTIONS

CT designed the research; YF, JQ, XX, HL, JL, and YH performed the research; YF, XX, JQ, CT, JY, YZ, SZ, and JT analyzed the data; YF and CT wrote the manuscript. The authors have read and approved the manuscript.

FUNDING

The research was supported partially by the National Natural Science Foundation of China (Nos. 31825007; 31770711) and the Starting Research Fund of Hainan University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.756270/full#supplementary-material

REFERENCES

Adiwilaga, K., and Kush, A. (1996). Cloning and Characterization of cDNA Encoding Farnesyl Diphosphate Synthase from Rubber Tree (*Hevea Brasiliensis*). *Plant Mol. Biol.* 30, 935–946. doi:10.1007/bf00020805

Akihiro, T., Umezawa, T., Ueki, C., Lobna, B. M., Mizuno, K., Ohta, M., et al. (2006). Genome Wide cDNA-AFLP Analysis of Genes Rapidly Induced by Combined Sucrose and ABA Treatment in rice Cultured Cells. *FEBS Lett.* 580, 5947–5952. doi:10.1016/j.febslet.2006.09.065

Amalou, Z., Bangratz, J., and Chrestin, H. (1992). Ethrel (Ethylene Releaser)-Induced Increases in the Adenylate Pool and Transtonoplast ΔpH within Hevea Latex Cells. *Plant Physiol.* 98, 1270–1276. doi:10.1104/pp.98.4.1270

Bachem, C. W. B., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M., and Visser, R. G. F. (1996). Visualization of Differential Gene Expression Using a Novel Method of RNA Fingerprinting Based on AFLP: Analysis of Gene Expression during Potato Tuber Development. *Plant J.* 9, 745–753. doi:10.1046/j.1365-313x.1996.9050745.x

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., et al. (1998). Analysis of 1.9 Mb of Contiguous Sequence from Chromosome 4 of *Arabidopsis thaliana*. *Nature* 391, 485–488. doi:10.1038/35140

Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* 72, 248–254. doi:10.1016/0003-2697(76)90527-3

Chantuma, P., Lacointe, A., Kasemsap, P., Thanisawanyangkura, S., Gohet, E., Clément, A., et al. (2009). Carbohydrate Storage in wood and Bark of Rubber Trees Submitted to Different Level of C Demand Induced by Latex Tapping. *Tree Physiol.* 29, 1021–1031. doi:10.1093/treephys/tpp043

Chow, K. S., Wan, K. L., Isa, M. N. M., Bahari, A., Tan, S. H., Harikrishna, K., et al. (2007). Insights into Rubber Biosynthesis from Transcriptome Analysis of *Hevea Brasiliensis* Latex. *J. Exp. Bot.* 58, 2429–2440. doi:10.1093/jxb/erm093

Chrestin, H., Gidrol, X., and Kush, A. (1997). Towards a Latex Molecular Diagnostic of Yield Potential and the Genetic Engineering of the Rubber Tree. *Euphytica* 96, 77–82. doi:10.1023/a:1002950300536

Cornish, K. (1993). The Separate Roles of Plant Cis and Trans Prenyl Transferases in Cis-1,4-Polyisoprene Biosynthesis. *Eur. J. Biochem.* 218, 267–271. doi:10.1111/j.1432-1033.1993.tb18374.x

d'Auzac, J., and Jacob, J. L. (1969). Regulation of Glycolysis in the Latex of *Hevea Brasiliensis*. *J. Rubber Res. Inst. Malaya.* 21, 417.

d'Auzac, J., Jacob, J., Prévôt, J., Clément, A., Gallois, R., Crestin, H., et al. (1997). "The Regulation of Cis-Polyisoprene Production (Natural Rubber) from Hevea Brasiliensis," in *Recent Research Developments in Plant Physiology*. Editor S. G. Pandalai (Kerala India: Research Signpost), 1, 273–332.

Dennis, M. S., Henzel, W. J., Bell, J., Kohr, W., and Light, D. R. (1989). Amino Acid Sequence of Rubber Elongation Factor Protein Associated with Rubber Particles in *Hevea* Latex. *J. Biol. Chem.* 264, 18618–18626. doi:10.1016/s0021-9258(18)51511-8

Duan, C., Rio, M., Leclercq, J., Bonnot, F., Oliver, G., and Montoro, P. (2010). Gene Expression Pattern in Response to Wounding, Methyl Jasmonate and Ethylene in the Bark of *Hevea Brasiliensis*. *Tree Physiol.* 30, 1349–1359. doi:10.1093/treephys/tpq066

Duangngam, O., Desalme, D., Thaler, P., Kasemsap, P., Sathornkich, J., Satakhun, D., et al. (2020). *In Situ* 13CO2 Labelling of Rubber Trees Reveals a Seasonal Shift in the Contribution of the Carbon Sources Involved in Latex Regeneration. *J. Exp. Bot.* 71, 2028–2039. doi:10.1093/jxb/erz551

Fatemi, F., Hashemi-Petroudi, S. H., Nematzadeh, G., Askari, H., and Abdollahi, M. R. (2019). Exploiting Differential Gene Expression to Discover Ionic and Osmotic-Associated Transcripts in the Halophyte Grass *Aeluropus Littoralis*. *Biol. Proced. Online* 21, 14. doi:10.1186/s12575-019-0103-3

Ganeshan, S., Sharma, P., Young, L., Kumar, A., Fowler, D. B., and Chibbar, R. N. (2011). Contrasting cDNA-AFLP Profiles between crown and Leaf Tissues of Cold-Acclimated Wheat Plants Indicate Differing Regulatory Circuitries for Low Temperature Tolerance. *Plant Mol. Biol.* 75, 379–398. doi:10.1007/s11103-011-9734-8

Gohet, E. (1996). *La production de latex par* Hevea brasiliensis*, relations avec la croissance. Influence de différents facteurs: origine clonale, stimulation hormonale, réserves hydrocarbonées*. PhD Thesis. Montpellier: Université Montpellier II.

Hao, B. Z., and Wu, J. L. (1982). Effects of Wound (Tapping) on Laticifer Differentiation in the *Hevea Brasiliensis*. *Acta Bot. Sin.* 24, 388–391.

Hao, B., and Wu, J. L. (2000). Laticifer Differentiation in *Hevea Brasiliensis*: Induction by Exogenous Jasmonic Acid and Linolenic Acid. *Ann. Bot.* 85, 37–43. doi:10.1006/anbo.1999.0995

Hao, B. Z., Wu, J. L., Meng, C. X., Gao, Z. Q., and Tan, H. Y. (2004). Laticifer Wound Plugging in *Hevea Brasiliensis*: the Role of a Protein-Network with Rubber Particle Aggregations in Stopping Latex Flow and Protecting Wounded Laticifers. *J. Rubber Res.* 7, 281–299.

Huang, Z. D. (2001). Challenges and Development Strategies of China's Natural Rubber. *Chin. J. Eng. Sci.* 3, 28–32.

Humphery-Smith, I., Cordwell, S. J., and Blackstock, W. P. (1997). Proteome Research: Complementarity and Limitations with Respect to the RNA and DNA Worlds. *Electrophoresis* 18 (8), 1217–1242. doi:10.1002/elps.1150180804

Ko, J. H., Chow, K. S., and Han, K. H. (2003). Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). *Plant Mol. Biol.* 53, 479–492. doi:10.1023/b:plan.0000019119.66643.5d

Lacote, R., Gabla, O., Obouayeba, S., Eschbach, J. M., Rivano, F., Dian, K., et al. (2010). Long-term Effect of Ethylene Stimulation on the Yield of Rubber Trees Is Linked to Latex Cell Biochemistry. *Field Crops Res.* 115, 94–98. doi:10.1016/j.fcr.2009.10.007

Leymarie, J., Bruneaux, E., Gibot-Leclerc, S., and Corbineau, F. (2007). Identification of Transcripts Potentially Involved in Barley Seed Germination and Dormancy Using cDNA-AFLP. *J. Exp. Bot.* 58, 425–437. doi:10.1093/jxb/erl211

Li, H., Qin, Y., Xiao, X., and Tang, C. (2011). Screening of Valid Reference Genes for Real-Time RT-PCR Data Normalization in *Hevea Brasiliensis* and Expression Validation of a Sucrose Transporter Gene *HbSUT3*. *Plant Sci.* 181, 132–139. doi:10.1016/j.plantsci.2011.04.014

Liu, S., Lan, J., Zhou, B., Qin, Y., Zhou, Y., Xiao, X., et al. (2015). H b NIN 2, a cytosolic alkaline/neutral-invertase, is responsible for sucrose catabolism in rubber-producing laticifers of H evea brasiliensis (para rubber tree). *New Phytol.* 206, 709–725. doi:10.1111/nph.13257

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262

Morris, D. L., and Davila, J. C. (1996). Analysis of Rat Cytochrome P450 Isoenzyme Expression Using Semi-quantitative Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR). *Biochem. Pharmacol.* 52, 781–792. doi:10.1016/0006-2952(96)00390-5

Oh, S. K., Kang, H., Shin, D. H., Yang, J., Chow, K. S., Yeang, H. Y., et al. (1999). Isolation, Characterization, and Functional Analysis of a Novel cDNA Clone Encoding a Small Rubber Particle Protein from Hevea Brasiliensis. *J. Biol. Chem.* 274, 17132–17138. doi:10.1074/jbc.274.24.17132

Pakianathan, S. W., Tata, S. J., Chon, L. F., and Sethuraj, M. R. (1992). Certain Aspects of Physiology and Biochemistry of Latex Production. *Dev. Crop Sci.* 23, 298–323. doi:10.1016/b978-0-444-88329-2.50020-9

Pirrello, J., Leclercq, J., Dessailly, F., Rio, M., Piyatrakul, P., Kuswanhadi, K., et al. (2014). Transcriptional and post-transcriptional Regulation of the Jasmonate Signalling Pathway in Response to Abiotic and Harvesting Stress in *Hevea Brasiliensis*. *BMC Plant Biol.* 14, 341. doi:10.1186/s12870-014-0341-0

Priya, P., Venkatachalam, P., and Thulaseedharan, A. (2007). Differential Expression Pattern of Rubber Elongation Factor (REF) mRNA Transcripts from High and Low Yielding Clones of Rubber Tree (*Hevea Brasiliensis* Muell. Arg.). *Plant Cel Rep.* 26, 1833–1838. doi:10.1007/s00299-007-0402-z

Qi, J. Y. (2011). *Large-scale screening by proteomic approaches of candidate latex-regeneration-related proteins in para rubber trees (*Hevea brasiliensis*)*. Master's Thesis. China: Hainan University.

Sae-Lim, P., Naktang, C., Yoocha, T., Nirapathpongporn, K., Viboonjun, U., Kongsawadworakul, P., et al. (2019). Unraveling Vascular Development-Related Genes in Laticifer-Containing Tissue of Rubber Tree by High-Throughput Transcriptome Sequencing. *Curr. Plant Biol.* 19, 100112. doi:10.1016/j.cpb.2019.100112

Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., et al. (1997). Improved and Simplified In-Gel Sample Application Using Reswelling of Dry Immobilized pH Gradients. *Electrophoresis* 18, 324–327. doi:10.1002/elps.1150180305

Saravanan, R. S., and Rose, J. K. C. (2004). A Critical Evaluation of Sample Extraction Techniques for Enhanced Proteomic Analysis of Recalcitrant Plant Tissues. *Proteomics* 4 (9), 2522–2532. doi:10.1002/pmic.200300789

Sharples, A. (1918). The Laticiferous System of Hevea Brasiliensis and its Protective Function. *Ann. Bot.* os-32, 247–251. doi:10.1093/oxfordjournals.aob.a089674

Silpi, U., Lacointe, A., Kasempsap, P., Thanysawanyangkura, S., Chantuma, P., Gohet, E., et al. (2007). Carbohydrate Reserves as a Competing Sink: Evidence from Tapping Rubber Trees. *Tree Physiol.* 27, 881–889. doi:10.1093/treephys/27.6.881

Singh, A. P., Wi, S. G., Chung, G. C., Kim, Y. S., and Kang, H. (2003). The Micromorphology and Protein Characterization of Rubber Particles in *Ficus Carica*, *Ficus Benghalensis* and *Hevea Brasiliensis*. *J. Exp. Bot.* 54, 985–992. doi:10.1093/jxb/erg107

Tang, C., Qi, J., Li, H., Zhang, C., and Wang, Y. (2007). A convenient and efficient protocol for isolating high-quality RNA from latex of *Hevea brasiliensis* (para rubber tree). *J. Biochem. Biophys. Methods* 70, 749–754. doi:10.1016/j.jbbm.2007.04.002

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

Tang, C., Xiao, X., Li, H., Fan, Y., Yang, J., Qi, J., et al. (2013). Comparative Analysis of Latex Transcriptome Reveals Putative Molecular Mechanisms Underlying Super Productivity of *Hevea Brasiliensis*. *PLoS One* 8, e75307. doi:10.1371/journal.pone.0075307

Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., et al. (2016). The Rubber Tree Genome Reveals New Insights into Rubber Production and Species Adaptation. *Nat. Plants* 2, 16073. doi:10.1038/nplants.2016.73

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., et al. (2010). Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* 329, 533–538. doi:10.1126/science.1188308

Tungngoen, K., Viboonjun, U., Kongsawadworakul, P., Katsuhara, M., Julien, J. L., Sakr, S., et al. (2011). Hormonal Treatment of the Bark of Rubber Trees (*Hevea Brasiliensis*) Increases Latex Yield through Latex Dilution in Relation with the Differential Expression of Two Aquaporin Genes. *J. Plant Physiol.* 168, 253–262. doi:10.1016/j.jplph.2010.06.009

Tupý, J. (1969). Stimulatory Effects of 2,4-dichlorophenoxyacetic Acid and of 1-naphthylacetic Acid on Sucrose Level, Invertase Activity and Sucrose Utilization in the Latex ofHevea Brasiliensis. *Planta* 88, 144–153. doi:10.1007/BF01391120

Tupý, J. (1985). Some Aspects of Sucrose Transport and Utilization in Latex Producing Bark ofHevea Brasiliensis Muel. Arg. *Biol. Plant* 27, 51–64. doi:10.1007/bf02894634

Tupý, J. (1989). "Sucrose Supply and Utilization for Latex Production," in *Physiology of Rubber Tree Latex*. Editors J. d'Auzac, J. L. Jacob, and H. Chrestin (Boca Raton, FL, USA: CRC Press), 179–199.

Tupý, J., and Resing, W. L. (1969). Substrate and Metabolism of Carbon Dioxide Formation in *Hevea* Latex, *In Vitro*. *J. Rubber Res. Inst. Malaya.* 21, 456.

Van Beilen, J. B., and Poirier, Y. (2007). Establishment of New Crops for the Production of Natural Rubber. *Trends Biotechnol.* 25, 522–529. doi:10.1016/j.tibtech.2007.08.009

Vogel, C., and Marcotte, E. M. (2012). Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses. *Nat. Rev. Genet.* 13, 227–232. doi:10.1038/nrg3185

Vuylsteke, M., Peleman, J. D., and van Eijk, M. J. (2007). AFLP-based Transcript Profiling (cDNA-AFLP) for Genome-wide Expression Analysis. *Nat. Protoc.* 2, 1399–1413. doi:10.1038/nprot.2007.174

Walley, J. W., Shen, Z., Sartor, R., Wu, K. J., Osborn, J., Smith, L. G., et al. (2013). Reconstruction of Protein Networks from an Atlas of maize Seed Proteotypes. *Proc. Natl. Acad. Sci. U S A.* 110, E4808–E4817. doi:10.1073/pnas.1319113110

Wang, X., Li, X., and Li, Y. (2007). A Modified Coomassie Brilliant Blue Staining Method at Nanogram Sensitivity Compatible with Proteomic Analysis. *Biotechnol. Lett.* 29, 1599–1603. doi:10.1007/s10529-007-9425-3

Xiao, X., Li, H., and Tang, C. (2009). A silver-staining cDNA-AFLP protocol suitable for transcript profiling in the latex of *Hevea brasiliensis* (para rubber tree). *Mol. Biotechnol.* 42, 91–99. doi:10.1007/s12033-008-9139-3

Xiao, D., Liu, S. T., Wei, Y. P., Zhou, D. Y., Hou, X. L., Li, Y., et al. (2016). cDNA-AFLP Analysis Reveals Differential Gene Expression in Incompatible Interaction between Infected Non-heading Chinese Cabbage and *Hyaloperonospora Parasitica*. *Hortic. Res.* 3, 16034. doi:10.1038/hortres.2016.34

Zeng, R. Z., Duan, C. F., Li, X. Y., Tian, W. M., and Nie, Z. Y. (2009). Vacuolar-type Inorganic Pyrophosphatase Located on the Rubber Particle in the Latex Is an Essential Enzyme in Regulation of the Rubber Biosynthesis in *Hevea Brasiliensis*. *Plant Sci.* 176, 602–607. doi:10.1016/j.plantsci.2009.01.009

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.

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

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