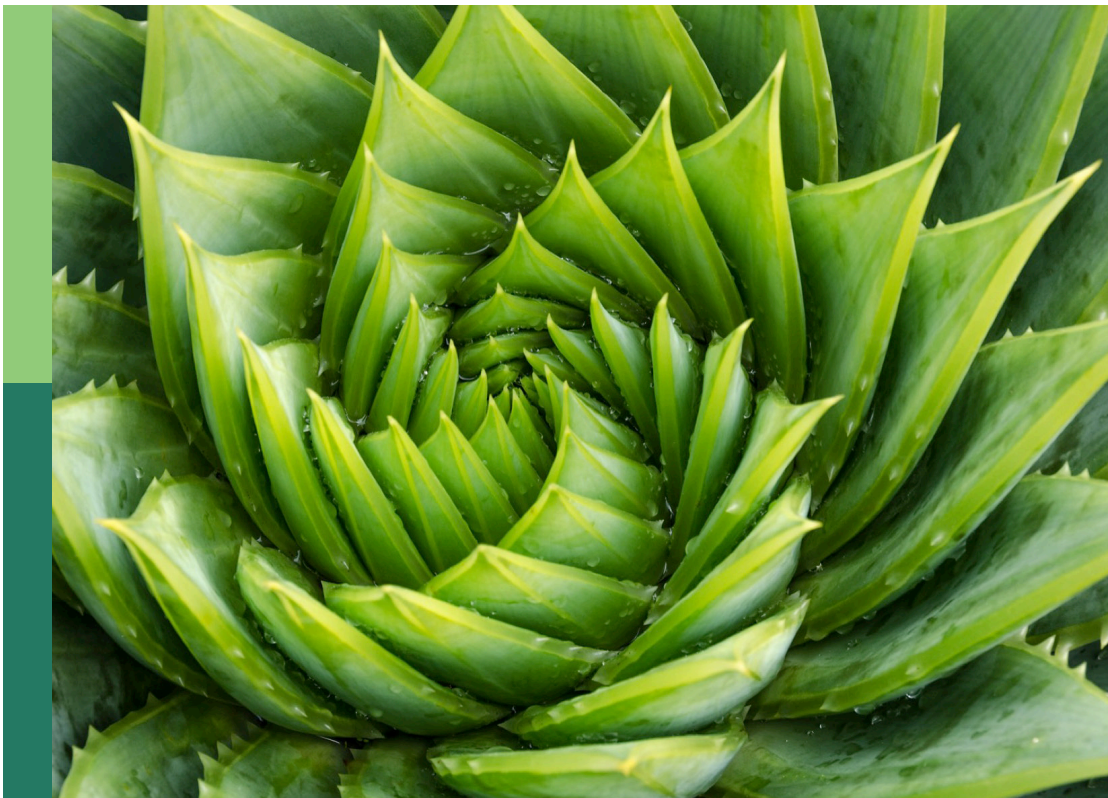
frontiers | Research Topics

Plant secondary metabolic regulation and engineering

Edited by
Fangyuan Zhang, Xueqing Fu and Yongliang Liu

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2344-5
DOI 10.3389/978-2-8325-2344-5

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

Plant secondary metabolic regulation and engineering

Topic editors

Fangyuan Zhang — Southwest University, China

Xueqing Fu — Shanghai Jiao Tong University, China

Yongliang Liu — University of Kentucky, United States

Citation

Zhang, F., Fu, X., Liu, Y., eds. (2023). *Plant secondary metabolic regulation and engineering*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2344-5

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 20 April 2023
DOI 10.3389/fpls.2023.1174717

OPEN ACCESS

EDITED AND REVIEWED BY
Kirsi-Marja Oksman-Caldentey,
VTT Technical Research Centre of Finland Ltd, Finland

*CORRESPONDENCE
Fangyuan Zhang
fyzhang@swu.edu.cn
Xueqing Fu
cathyluck@sjtu.edu.cn
Yongliang Liu
yongliangliu@uky.edu

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

RECEIVED 27 February 2023
ACCEPTED 05 April 2023
PUBLISHED 20 April 2023

CITATION
Zhang F, Fu X and Liu Y (2023) Editorial: Plant secondary metabolic regulation and engineering.
Front. Plant Sci. 14:1174717.
doi: 10.3389/fpls.2023.1174717

Editorial: Plant secondary metabolic regulation and engineering

Fangyuan Zhang[1]*, Xueqing Fu[2]* and Yongliang Liu[3]*

[1]Integrative Science Center of Germplasm Creation in Western China (Chongqing) Science City & Southwest University, School of Life Sciences, Southwest University, Chongqing, China, [2]School of Design, Shanghai Jiao Tong University, Shanghai, China, [3]Department of Plant and Soil Sciences and Kentucky Tobacco Research and Development Center, University of Kentucky, Lexington, KY, United States

KEYWORDS

plant secondary metabolites, metabolic biosynthesis, metabolic regulation, metabolic engineering, transcription factor

Editorial on the Research Topic
Plant secondary metabolic regulation and engineering

Introduction

Plant secondary metabolites (PSMs), or plant specialized metabolites attract great interest they are unique sources of drugs, nutrients, agrochemicals, and chemical additives (Kabera et al., 2014; Tiwari and Rana, 2015). PSMs include several major groups such as phenolics, terpenes, and nitrogen-containing compounds, and they play vital roles in coping with abiotic and biotic stresses (Kaushik et al., 2021; Elshafie et al., 2023). Due to the diversity of PSMs in different plant species, biosynthesis and regulation of PSMs are not fully elucidated. Moreover, most of the PSMs are low in content and metabolic engineering has been used to boost the production of valuable PSMs (Verpoorte and Memelink, 2002; Zheng et al., 2023). This Research Topic presents the most recent advances in 26 publications, including 4 reviews and 22 research articles, contributed by 197 authors. The aim of this topic is to strengthen our understanding of the biosynthesis and regulation of PSMs in non-model plants, especially medicinal plants, and provide efficient routes to elevate the production of valuable PSMs. Here we summarize these papers according to the classification of contributions, mainly including the identification of key pathway genes of PSMs, the elucidation of regulatory mechanisms, and the successful metabolic engineering of several valuable PSMs.

Identification of key pathway genes of PSMs

Although the value of specific PSMs, their biosynthetic and metabolic pathways are not fully illustrated. In this topic, several candidate genes for the biosynthesis of diterpenoids (ATDs) and phenolic acids (PAs) were identified in the medicinal plant *Salvia apiana* and its close species *S. miltiorrhiza* (Hu et al.). In this study, comparative metabolome and transcriptome analyses between *S. apiana* roots and leaves, and between two species

revealed that four cytochromes (CYPs) and clade VII laccases might contribute to the biosynthesis of specific ATDs and PAs, respectively. In another study, two borneol acetyltransferases (BAT), WvBAT3 and WvBAT4, catalyzing the last step of bornyl acetate (BA) biosynthesis were characterized in *Wurfbainia villosa* (Liang et al.). BA is an aromatic monoterpene ester mostly accumulated in the seeds of *W. villosa*. These two BATs presented *in vitro* catalytic efficiency on the substrates of BA, and their gene expression patterns well correlate with the distribution of BA. In *Trigonella foenum-graecum*, Tf3SGT2 was identified as a steroid-specific UDP-glucose 3-O-glucosyltransferase that involves in steroidal saponin biosynthesis (Gao et al.). *In vitro* enzyme assay verified the catalytic activity of Tf3SGT2. Furthermore, RNA interference (RNAi) of *Tf3SGT2* in the hairy roots of *Trigonella foenum-graecum* confirmed the involvement of Tf3SGT2 in steroidal saponin biosynthesis. In another study, DoCCD1 was characterized involving in the biosynthesis of β-ionone in *Dendrobium officinale* (Wang et al.). In both *Escerichia coli* cells and *Nicotiana benthamiana* leaves which contain carotenoid precursors, expression of DoCCD1 resulted in the production of β-ionone. These findings contribute to our understanding of the biosynthetic and metabolic pathways of PSMs and provide a basis for the elucidation of regulatory mechanisms and metabolic engineering of PSMs.

Elucidation of regulatory mechanisms of PSMs

The Biosynthesis of PSMs is commonly regulated at the transcriptional level by transcription factors (TFs). The gene expression, protein abundance, localization, or trans-activities of the TFs could be modulated by various stimuli and signaling (Vom Endt et al., 2002). In this topic, an R2R3-MYB TF TcMYB29a was identified to regulate taxol biosynthesis in *Taxus chinensis* (Cao et al.). Overexpression of *TcMYB29a* in *T. chinensis* cell suspension cultures led to an increased accumulation of taxol, and upregulated expression of several taxol biosynthetic genes. Moreover, the expression of *TcMYB29a* was strongly enhanced by the treatment of Abscisic acid (ABA), which also induced the production of taxol. In another study, a bHLH TF AabHLH112 was characterized to positively regulate the biosynthesis of three kinds of sesquiterpenes, β-caryophyllene, epi-cedrol, and β -farnesene in *Artemisia annua* (Xiang et al.). AabHLH112 directly binds to the E-box (CANNTG) motifs in the promoters of the biosynthetic genes of these three sesquiterpenes. Exogenous methyl jasmonate (MeJA) enhanced the expression of *AabHLH112*, the biosynthetic genes as well as the contents of sesquiterpenes. In apples, a B-box protein, MdBBX21 was identified to positively involve in the light-induced biosynthesis of anthocyanins in the fruit peel of red apples (Zhang et al.). Overexpression of *MdBBX21* in *Arabidopsis* and apple calli under light increased anthocyanin accumulation. Moreover, the interaction of MdBBX21 and another TF MdHY5 significantly increased their trans-activation on the promoter of a target gene. In another study, a R2R3-MYB TF, NtMIXTA1, was characterized to involve in glandular trichomes (GTs) development in the medicinal plant *Nepeta tenuifolia* (Zhou et al.). GTs are the primary storage organ for monoterpenes in *N. tenuifolia*. Knock-down of *NtMIXTA1* resulted in lower GT density, a significant reduction in monoterpene concentration, and the decreased expression of genes related to monoterpene biosynthesis. The findings enrich the the transcriptional regulatory network of PSMs.

Metabolic engineering of valuable PSMs

Due to the low concentration of valuable PSMs in plants, metabolic engineering has long been one of the most efficient approaches utilized to boost the accumulation of specific PSMs (Courdavault et al., 2021). In this topic, an *Arabidopsis* MYB-type TF AtMYB12 was overexpressed in licorice (*Glycyrrhiza inflata*) hairy roots and induced the accumulation of total flavonoids as well as the specific licochalcones, licochalcone a (LCA) and echinatin (Wu et al.). Transcriptome analyses of the *AtMYB12*-overexpressing hairy roots implied that the carbohydrate metabolism was likely reprogrammed to increase carbon flux into flavonoid biosynthesis. In another study, a *Salvia miltiorrhiza* GRAS TF *SmSCR1* was overexpressed in *S. miltiorrhiza* hairy roots and significantly induced the accumulation of tanshinone (Zhou et al.). In Tartary Buckwheat (*Fagopyrum tataricum*), FtMYB45 is an R2R3-type MYB TF that negatively regulates flavonoid biosynthesis. Knock-out of *FtMYB45* resulted in an increased accumulation of rutin, catechin, and other flavonoids (Wen et al.). Wild tomato species *Solanum habrochaites* produce various sesquiterpenes in the GTs for herbivore defense. Overexpression of a *prenyl transferase gene* and a *terpene synthase gene* in tomato leaves led to an increased accumulation of sesquiterpenes and also enhanced resistance to pests (Wang et al.). Deguchi et al. developed the vacuum agroinfiltration method to increase total CBD content and reduce the total THC content through transiently expressing *CBDAS* gene and silencing *THCAS* gene, respectively. The study suggests that metabolic engineering is an effective strategy to increase the accumulation of valuable PSMs.

PSMs have attracted the attention because of their great economic value, healthcare value and medical value. To improve the production of PSMs, the researchers focus on identification of key pathway genes and understanding regulatory mechanisms of PSMs. This Research Topic is a timely collection of advanced studies on plant secondary metabolic regulation and engineering. Despite further studies are needed, we hope that this Research Topic offers some important insight into this research area.

Author contributions

FZ, XF, and YL have made a substantial, direct, and intellectual contribution to the work, and approved it for publication in Frontiers in Plant Science.

Funding

This work was supported by the National Key Research and Development Program (2019YFE0108700, China); the National Natural Science Foundation of China (81803660).

Acknowledgments

We thank the authors for their contributions and reviewers as well as editors for their scientific reviewing and diligent work on the Research Topic.

Conflict of interest

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

Publisher's note

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

References

Courdavault, V., O'Connor, S. E., Jensen, M. K., and Papon, N. (2021). Metabolic engineering for plant natural products biosynthesis: New procedures, concrete achievements and remaining limits. *Natural Product Rep.* 38, 2145–2153. doi: 10.1039/D0NP00092B

Elshafie, H. S., Camele, I., and Mohamed, A. A. (2023). A comprehensive review on the biological, agricultural and pharmaceutical properties of secondary metabolites based-plant origin. *Int. J. Mol. Sci.* 24, 3266. doi: 10.3390/ijms24043266

Kabera, J. N., Semana, E., Mussa, A. R., and He, X. (2014). Plant secondary metabolites: Biosynthesis, classification, function and pharmacological properties. *J. Pharm. Pharmacol.* 2, 377–392.

Kaushik, B., Sharma, J., Yadav, K., Kumar, P., and Shourie, A. (2021). Phytochemical properties and pharmacological role of plants: Secondary metabolites. *Biosci. Biotechnol. Res. Asia* 18, 23–35. doi: 10.13005/bbra/2894

Tiwari, R., and Rana, C. (2015). Plant secondary metabolites: A review. *Int. J. Eng. Res. Gen. Sci.* 3, 661–670.

Verpoorte, R., and Memelink, J. (2002). Engineering secondary metabolite production in plants. *Curr. Opin. Biotechnol.* 13, 181–187. doi: 10.1016/S0958-1669(02)00308-7

Vom Endt, D., Kijne, J. W., and Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: What regulates the regulators? *Phytochemistry* 61, 107–114. doi: 10.1016/S0031-9422(02)00185-1

Zheng, H., Fu, X., Shao, J., Tang, Y., Yu, M., Li, L., et al. (2023). Transcriptional regulatory network of high-value active ingredients in medicinal plants. *Trends Plant Sci.* 28, 429–446. doi: 10.1016/j.tplants.2022.12.007

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 November 2021
doi: 10.3389/fpls.2021.774943

Identification of MBW Complex Components Implicated in the Biosynthesis of Flavonoids in Woodland Strawberry

*Pengbo Xu[1], Liang Wu[2], Minghao Cao[1], Chao Ma[1], Kun Xiao[1], Yanbang Li[1] and Hongli Lian[1]**

[1] School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States

Flavonoids belong to the family of polyphenolic secondary metabolites and contribute to fruit quality traits. It has been shown that MBW complexes (MYB-bHLH-WD40) regulate the flavonoids biosynthesis in different plants, but only a limited number of MBW complexes have been identified in strawberry species in general. In this study, we identified 112 R2R3-MYB proteins in woodland strawberry; 12 of them were found to have potential functions in regulating flavonoids biosynthesis by phylogenetic analysis. qRT-PCR assays showed that *FvMYB3*, *FvMYB9*, *FvMYB11*, *FvMYB22*, *FvMYB64,* and *FvMYB105* mostly expressed at green stage of fruit development, aligned with proanthocyanidins accumulation; *FvMYB10* and *FvMYB41* showed higher expression levels at turning and ripe stages, aligned with anthocyanins accumulation. These results suggest that different MYBs might be involved in flavonoids biosynthesis at specific stages. Furthermore, FvMYB proteins were demonstrated to interact with FvbHLH proteins and induce expression from the promoters of *CHS2* and *DFR2* genes, which encode key enzymes in flavonoids biosynthesis. The co-expression of FvMYB and FvbHLH proteins in strawberry fruits also promoted the accumulation of proanthocyanidins. These findings confirmed and provided insights into the biofunction of MBW components in the regulation of flavonoid biosynthesis in woodland strawberry.

Keywords: woodland strawberry, MBW complex, flavonoid biosynthesis, MYB, bHLH

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Yu-Qing He,
Zhejiang Academy of Agricultural Sciences, China
Andrea Moglia,
University of Turin, Italy

****Correspondence:***
Hongli Lian
hllian@sjtu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *13 September 2021*
Accepted: *20 October 2021*
Published: *08 November 2021*

Citation:
Xu P, Wu L, Cao M, Ma C, Xiao K, Li Y and Lian H (2021) Identification of MBW Complex Components Implicated in the Biosynthesis of Flavonoids in Woodland Strawberry. Front. Plant Sci. 12:774943. doi: 10.3389/fpls.2021.774943

INTRODUCTION

Strawberry is favored by consumers mainly because of its unique aroma, sweet taste, bright color and nutritional value. These quality traits are largely determined by the secondary metabolites in the fruit. Flavonoids, which are natural polyphenol compounds, consist of six major subgroups, including chalcones, flavones, flavonols, flavandiols, anthocyanins, and proanthocyanidins (Li, 2014). Flavonoids not only give plants their attractive colors for pollen and seed dispersal, but also protect plants against ultraviolet radiation (Winkel-Shirley, 2001; Emiliani et al., 2013). In addition, flavonoids may act as antioxidants or signaling molecules to exert potential beneficial effects on human health (He and Giusti, 2010), as well as being antimicrobial agents in plant-microbe interactions and plant defense response (Nijveldt et al., 2001; Battino et al., 2009).

Flavonoids are synthesized from malonyl-CoA produced by the fatty acid metabolic pathway and 4-coumaroyl-CoA produced by the phenylpropanoid metabolic pathway with the catalysis by a variety of enzymes. In general, three molecules of malonyl-CoA and one molecule of 4-coumaroyl-CoA were converted into one anthocyanin under a successive catalytic action of many enzymes, including chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H), flavonol synthase (FLS), flavonoid 3′-hydroxylase (F3′H), dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), leucoanthocyanidin reductase (LAR), anthocyanidin reductase (ANR), and 3-glycosyltransferase (3-GT). In the synthesis pathway of anthocyanins, the intermediate metabolites dihydrokaempferol and dihydroquercetin, products of F3H, can be oxidized by FLS to produce flavonols. Similarly, the products of DFR and ANS, leucocyanidin, and cyanidin, can also be converted into proanthocyanidins precursors by LAR and ANR, respectively.

Flavonoids biosynthesis has been well studied in many plants and is largely regulated at the transcriptional level by MYB-bHLH-WD40 (MBW) complex. The MYB protein family is one of the largest families of transcription factors in plants. According to the number of R motifs contained in a MYB gene, it could be divided into four subfamilies, namely 1R-MYB with one R domain, 2R-MYB (R2R3-MYB) with two R domains, and 3R-MYB or 4R-MYB with three or four R domains. Among the four subfamilies, R2R3-MYB is the most abundant, which is widely involved in regulating plant growth and development, responses to biotic and abiotic stresses, as well as environmental factors (Dubos et al., 2010). MBW comprises three classes of regulatory proteins, including R2R3-MYBs, bHLHs and TRANSPARENT TESTA GLABROUS1 (TTG1; also termed WD40) (Li, 2014). In *Arabidopsis*, it has been reported that the three regulatory proteins, namely AtTT2/AtMYB123 (Nesi et al., 2001; Dubos et al., 2010), AtTT8/AtbHLH042 (Nesi et al., 2000), and AtTTG1 (WD40-repeat protein) (Walker et al., 1999; Gonzalez et al., 2009), act together as a MBW complex to promote the production of proanthocyanidins in seed coat, by activating the expression of target gene *BAN*/*ANR* (Baudry et al., 2004). The genes involved in the synthesis and accumulation of anthocyanins are also regulated by MBW complexes, comprise at least one R2R3-MYB protein from AtMYB75, AtMYB90, AtMYB113 and AtMYB114, one bHLH protein from TT8, GLABROUS3 (GL3) and ENHANCER OF GLABRA3 (EGL3), and one WD40 protein, AtTTG1 (Dubos et al., 2010; Xu et al., 2015).

For fruit crops, such as grapevine, apple, peach and strawberry, the identification of key MYB proteins have significantly contributed to the understanding of the regulation of fruit flavonoids synthesis (Azuma et al., 2008; Terrier et al., 2009; Zhang et al., 2014; Zoratti et al., 2014; Tuan et al., 2015). Schaart et al. (2013) identified FaMYB9 and FaMYB11 as regulators of proanthocyanidins in octoploid strawberry by a strategy combining yeast-two-hybrid screening and agglomerative hierarchical clustering of transcriptomic and metabolomic data. FvMYB10 has been well studied for its function in the accumulation of anthocyanin in woodland strawberry (Lin-Wang et al., 2014; Castillejo et al., 2020). In the MBW complex, MYB protein is the decisive component for the regulation of flavonoids synthesis. The R2R3 motif at the N-terminal of MYB is responsible for the specific binding to the promoter of target genes, while the C-terminal is responsible for the activation or inhibition of the expression of target genes (Mehrtens et al., 2005; Prouse and Campbell, 2012; Franco-Zorrilla et al., 2014; Kelemen et al., 2015). bHLH protein functions by interacting with MYB (Grotewold et al., 1994). WD40 protein seems to have a more general role in the regulatory complex (Hichri et al., 2010). bHLH proteins in MBW complexes also play important roles in anthocyanin accumulation. There are 133 bHLHs in *Arabidopsis*, three of which have been proven to be related to anthocyanin formation. AtTT8 is a key regulator of anthocyanin and proanthocyanidins biosynthesis. AtEGL3 and AtGL3 mainly act in vegetative tissues (Ramsay and Glover, 2005; Gonzalez et al., 2008). In petunia, the bHLH protein AN1 and JAF13 can interact with AN2 (a R2R3-MYB protein) and WD40 protein AN11, thereby activating anthocyanin biosynthesis genes (Quattrocchio et al., 1993; deVetten et al., 1997; Spelt et al., 2000). In grape, apple and pear, several bHLH proteins, such as VvMYC1, MdbHLH3, MdbHLH33, PpbHLH3, and PpbHLH33, have been found to promote anthocyanin accumulation, and these proteins can form MBW complexes with MYB proteins (Espley et al., 2007; Hichri et al., 2010; Lin-Wang et al., 2014). Similarly, FvbHLH33 can help FvMYB10 to activate the expression of *FvDFR* and *FvUFGT*, key enzyme genes of anthocyanin synthesis. However, there is still limited information available on the transcriptional regulation of the flavonoids biosynthesis by the MBW complexes, especially in the fruits of woodland strawberry.

To identify additional regulatory proteins involved in the flavonoids biosynthesis pathways in woodland strawberry, we first identified the *R2R3-MYB* gene family using latest genome database version 4.0.2a. Novel *MYB* and *bHLH* genes that are possibly related to flavonoids biosynthesis were identified through homology alignment and comprehensive phylogenetic analysis. The corresponding MYB and bHLH proteins were confirmed to form functional complexes using molecular and biochemical experiments. The results from our study contribute to the understanding of flavonoids biosynthesis in the fruits of woodland strawberry at molecular levels.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Fragaria vesca seeds of RG (Ruegen, runnerless, everbearing, and red fruited) were grown in flowerpots and cultivated in photoperiodic conditions of 16 h of light and 8 h of dark at a temperature of 22°C, and a relative humidity of 50%.

Identification of R2R3-MYB Proteins in Woodland Strawberry

All the protein sequences of *F. vesca* were downloaded from GDR database[1] and the genome database version 4.0.2a was

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

used. An HMM search with the MYB DNA-binding domain HMM profile (PF000249) was used to blast against strawberry protein database and the *e*-value was set as 1.0×10^{-3}. Because of the different splicing models for many genes in version 4.0.2a, a total of 499 *MYB* gene transcripts were obtained as a result. For the different splicing of each *MYB* gene, the longest transcript was preserved and the remaining was removed, resulting in 242 *MYB* transcripts obtained. To confirm the presence of MYB DNA-binding domain, the protein sequences of the 242 *MYB* transcripts were used as queries to search against Pfam[2], SMART[3], and HMMER[4] database, the *e*-value was set as 1.0×10^{-5}. In the end, a total of 118 MYB protein sequences containing more than two MYB DNA-binding domains were obtained. The isoelectric points and molecular weights of the corresponding MYB proteins were obtained from the online database of ExPASY[5].

Phylogenetic Analysis

The 125 *R2R3-MYB* genes identified in *Arabidopsis* were obtained from a previous study (Dubos et al., 2010). The MYB protein sequences were downloaded from The *Arabidopsis* Information Resource (TAIR)[6]. The sequences of MYB proteins from *Arabidopsis* and strawberry were aligned using ClustalW (Thompson et al., 1994), and the BLOSUM matrix was used. The parameters of gap opening and extension penalties were 25 and 1, respectively, and the other parameters were default. The phylogenetic tree was constructed using the Neighbour Joining Tree Method in Mega5 (Tamura et al., 2011). The bootstrap was 1,000 replicates. Evolutionary distances were computed with P-distance method and the positions containing gaps and missing data were eliminated with partial deletion option (95%). For the MYB and bHLH proteins, and other corresponding proteins from different species, the phylogenetic tree was constructed as above.

Yeast Two-Hybrid Assay

Full CDS of *bHLH* and *MYB* genes were amplified and then inserted into BD vector pLexA/pGBKT7 and AD vector pB42AD/pGADT7, respectively, between EcoR I and Xho I using ClonExpress II One Step Cloning Kit (Vazyme Biotech Co., Ltd.). All sequences of the recombinant plasmids were verified by sequencing (Sangon Biotech Co., Ltd. at Shanghai). The LexA yeast two-hybrid was performed as described previously (Xu et al., 2019). For the GAL4 yeast two-hybrid assay, the bait and prey vectors were co-transformed into AH109 yeast cells as described by Lu et al. (2015). The primers used were listed in **Supplementary Table 1**.

Yeast One-Hybrid Assay

The 2.0 kb upstream sequences before the ATG of *FvCHS2* and *FvDFR2* genes were amplified from woodland strawberry of RG and inserted into the reporter vector pLacZ using homologous recombination method as above. The recombinant effector plasmids with inserted *MYB* or *bHLH* genes and the reporter plasmids were co-transformed into EGY48 yeast cells and grown at 30°C for 3 days. Then the DNA-binding activity were detected as described in the previous study (Li et al., 2020). To investigate whether bHLH can affect the bioactivity of MYB protein, the cDNA encoding the full-length *bHLH* was inserted into a modified pLexA vector, which lacks the BD domain. The two effector plasmids harboring *bHLH* and *MYB* genes, respectively, were co-expressed with the reporter plasmid in EGY48 yeast cells. Transformed colonies were selected on SD-Trp-His-Ura medium plus X-gal as described previously (Li et al., 2020). The sequences of the CHS2 and DFR2 promoters were listed in **Supplementary Table 2** and the potential *cis*-elements of the CHS2 and DFR2 promoters for MBW binding were analyzed by PlantCARE[7] (Lescot et al., 2002). The primers used were listed in **Supplementary Table 1**.

Quantitative RT-PCR Analysis

Strawberry fruits at green stage (about 15 days post anthesis, DPA), white stage (about 22 DPA), pre-turning stage (white receptacles with red achenes), turning stage (2 or 3 days after the pre-turning stage), and ripen stage (3 or 4 days after the turning stage) were harvested. The achenes were removed from all the fruits. and then the total RNA was extracted following the manufacturer's instructions of RNA Plant Plus Reagent (Tiangen, Beijing, China, Cat.DP437). DNase I was used to remove the genomic DNA contamination. First-strand cDNA was synthesized using EasyScript® One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen, China). Real-time qPCR was performed using the CFX96 Touch™ Real-Time PCR System (Bio-Rad, United States) with 2 × M5 HiPer SYBR Premix EsTaq (with Tli RNaseH) (Meibio, China, Cat. MF787-01). The qPCR reaction was performed by pre-denaturing at 95°C for 1 min, followed by 40 cycles of denaturing at 95°C for 5 s, annealing at 60°C for 15 s, and extension at 72°C for 15 s. Melting curve detection from 65 to 95°C was done after the cyclic reactions. Transcript abundance of target genes were calculated using $2^{-\Delta Ct}$ method in comparison with the internal control gene, *Fv26S* or *FvActin*. Three biological replicates were used in each assay. The experiments were repeated at least three times. The primer sequences were listed in **Supplementary Table 3**.

BiFC Assay

The vectors used to produce constructs for the BiFC assays were pXY104 and pXY106, which carry fragments encoding the C- and N-terminal halves of YFP (cYFP and nYFP), respectively. The full-length cDNAs of *MYB* and *bHLH* genes were cloned into pXY104 and/or pXY106 vector to generate corresponding recombinant plasmids. All constructs were transformed into *Agrobacterium tumefaciens* strain GV3101 and infiltrated into tobacco (*Nicotiana benthamiana*) epidermal leaves in the given combinations as described previously (Lu et al., 2015; Xu et al., 2016). After incubation in the dark for 36–40 h, the tobacco leaf samples were examined by confocal microscopy (Leica TCS

[2]https://www.ebi.ac.uk/Tools/pfa/pfamscan/

[3]http://smart.embl-heidelberg.de/

[4]https://www.ebi.ac.uk/Tools/hmmer/

[5]https://www.expasy.org/

[6]http://www.arabidopsis.org/

[7]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

SP5II) to detect the expression of various fluorescent proteins. The intensity of the laser was set at 30%, the wavelength at 514 nm, and the voltage at 700 v. All the primers used were listed in **Supplementary Table 1**.

Luciferase Reporter Assay

The pGreen0800-*Pro$_{DFR2}$:Luc* vector was described previously (Li et al., 2020). The full length CDS of *MYB* and *bHLH* genes including *FvMYB22*, *FvMYB64*, *FvMYB105,* and *FvbHLH33* were cloned into pGreen62SK vector and served as effectors. The reporter strain GV3101 harboring *Pro$_{DFR2}$:Luc* was mixed with each of the effector strain harboring pGreen62SK, pGreen62SK-*bHLH33* or one of the pGreen62SK-*FvMYB* plasmids (pGreen62SK-*FvMYB22*, pGreen62SK-*FvMYB64,* or pGreen62SK-*FvMYB105*) individually, as well as combinations of plasmids indicated. The mixed culture of each combination was infiltrated into tobacco leaves as described in the results of this article. After incubation in the dark for 36–40 h, the *N. benthamiana* leaves were treated with 1 mM l-luciferin sodium salt (Yeasen, Shanghai, China, Cat. 40903ES02) and kept in the dark for an additional 10 min before the bioluminescence signal were imaged and the intensity measured by a luminescent imaging workstation (5200; Tanon, Shanghai, China). The primers used for luciferase reporter assay are listed in **Supplementary Table 1**.

Detection of Anthocyanin and Proanthocyanidin

The method of anthocyanin detection was described in previous studies (Xu et al., 2018; Li et al., 2020). The method of proanthocyanidins detection was described in previous study (Li et al., 2018). DMACA, (−)-epicatechin and 4-hydroxy-3-methoxybenzaldehyde used in the experiments were from Bide Pharmatech. Ltd. (Shanghai, China).

Transient Overexpression in Strawberry Fruit

The construct of pGreen62SK, pGreen62SK-*bHLH33* or one of the pGreen62SK-*FvMYB* plasmids (pGreen62SK-*FvMYB22*, pGreen62SK-*FvMYB64,* or pGreen62SK-*FvMYB105*) was used for transient overexpression. The method of injection was described in previous study (Zhang et al., 2020). The octoploid strawberry of "HongYan" was used for transient overexpression. The fruit were collected at 7 days after transfection and the measurement of proanthocyanidins was described in the same previous study (Xu et al., 2018). The expression of *FvbHLH33*, *FvMYB22*, *FvMYB64*, *FvMYB105*, *FvCHS2,* and *FvDFR2* was confirmed by qRT-PCR assay.

RESULTS

Genome Wide Identification of *R2R3-MYB* Genes in *Fragaria vesca*

In order to identify *R2R3-MYB* genes in strawberry, the HMM profile (Pfam:00249) of MYB DNA-binding domain was used to blast against the woodland strawberry genome database (*F. vesca* Whole Genome v4.0.a2) from GDR (see Text Footnote 1). The blast search resulted in a total of 242 *MYB* genes after removing repetitive redundant sequences. The protein products of 118 *MYB* genes out of the 242 were verified to contain conserved MYB DNA-binding domains through searching Pfam, SMART and HMMER databases. All 118 MYB gene products contain more than two MYB repeat units. Among them, one contains four MYB repeat units, five contain three MYB repeat units. The remaining 112 contain two MYB repeat units, which are R2R3-MYBs (**Supplementary Table 4**). Through gene sequence alignment analysis, we found that FvH4_1G22020 shares the same sequence with *FvMYB10*, thus annotated as *FvMYB10*. Blast results showed that the sequences of FvH4_2g31100 and FvH4_6g34650 were almost consistent with *FaMYB9* and *FaMYB11*, respectively. We therefore named FvH4_2g31100 and FvH4_6g34650 as *FvMYB9* and *FvMYB11*. The remaining *R2R3-MYB* genes were named according to the order of their chromosomal locations in the strawberry genome browser (**Supplementary Table 4**). The physical and chemical property prediction analysis was performed for the 112 R2R3-MYB proteins using ExPASy database. The results showed that the sizes of these MYB proteins are significantly different. For example, the protein product of FvH4_7g19850 is composed of 1703 amino acids, while other MYB proteins are much smaller, containing amino acids of a number ranging from 127 to 562. The isoelectric points of MYB proteins have been predicted to range from 4.82 to 10.10, and the presumed molecular weights are in a range of 15–62.7 kD, except for the protein product of FvH4_7g19850, the molecular weight of which is 185 kD (**Supplementary Table 4**).

Identification of Potential MYB and bHLH Proteins Involved in Regulating Flavonoids Biosynthesis

To identify which *R2R3-MYB* genes in strawberry are possibly involved in flavonoids biosynthesis, the sequences of the 112 strawberry *R2R3-MYB* genes identified by us and the 125 *R2R3-MYB* genes from *Arabidopsis* were phylogenetically analyzed. The results showed that a total of 17 strawberry *MYB* genes clustered with *AtMYB*s that are involved in flavonoids biosynthesis (highlighted in black solid circles in **Figure 1**). We also included *R2R3-MYB* genes from other species with known functions in flavonoids synthesis for comparison, including those from grape, apple, cotton, tomato, maize and freesia. These *R2R3-MYB* genes were also phylogenetically analyzed together with the 17 strawberry *MYB* genes. Taken together,12 strawberry *R2R3-MYB* genes were identified as potentially involved in flavonoids biosynthesis (highlighted in black solid circles, **Supplementary Figure 1**). Specifically, *FvMYB10* and *FvMYB75* clustered with *MdMYB10* and *MdMYB1*, which promote anthocyanin biosynthesis (Espley et al., 2007). *FvMYB3*, *FvMYB9,* and *FvMYB11* clustered with *FaMYB9* and *FaMYB11*, which were found to complement the phenotype of transparent testa in *Arabidopsis tt2-1* mutant (Schaart et al., 2013), suggesting a function of regulating proanthocyanidins biosynthesis. *FvMYB64* and *FvMYB105* were homologous to *VvMYBPA1*, which is

specific to the regulation of proanthocyanidins biosynthesis in seeds (Bogs et al., 2007). *FvMYB45* and *FvMYB77* were similar to maize *P* (Grotewold et al., 1994) and *AtMYB111* (Stracke et al., 2007), respectively, suggesting that both of them might be related to flavonols synthesis. *FvMYB41* clustered with *VvMYB5b*, which contributes to the regulation of anthocyanin and proanthocyanidins biosynthesis in the developing grape fruits (Deluc et al., 2008). *AtMYB123* did not explicitly cluster with any known genes, but when *AtMYB123* was used as a query to blast against strawberry genome, *FvMYB21* and *FvMYB22* was also among the results, displaying high similarity to *AtMYB123*, suggesting that *FvMYB21* and *FvMYB22* are potentially related to proanthocyanidins biosynthesis.

To screen bHLH components of MBW complexes, we used three *Arabidopsis* bHLH proteins as queries to perform BLASTP searches against *F. vesca* protein database. The three *Arabidopsis* bHLH proteins, including AtTT8, AtEGL3, and AtGL3, have been shown to have partially overlapping expression patterns and redundant functions in flavonoids biosynthesis. By the searches, we found that no matter which one of AtTT8, AtEGL3, and AtGL3 was used for blasting, the matching proteins with higher similarity were always the products of FvH4_2g23700, FvH4_7g14230, and FvH4_5g02520 (**Supplementary Table 5**). A phylogenetic analysis of these three strawberry bHLH proteins and those from other 16 plants was performed, and we found that FvH4_2g23700, FvH4_7g14230, and FvH4_5g02520 clustered with FabHLH3, FabHLH33 and FaMYC1 of *Fragaria* × *ananassa*, respectively (**Supplementary Figure 2**). Consequently, the bHLH proteins from woodland strawberry were named as FvbHLH3, FvbHLH33, and FvMYC1, and considered potential components of MBW complexes for flavonoids biosynthesis.

The Expression Pattern of *R2R3-MYB* and *bHLH* Genes Correlates With Anthocyanin and Proanthocyanidin Accumulation in Strawberry Fruits

We studied the transcript levels of *MYBs* and *bHLHs*, and their effects on the accumulation of flavonoids in strawberry fruits at five stages of fruit development, including green stage, white stage, pre-turning stage, turning stage and ripe stage. qRT-PCR results showed that the transcript level of a *bHLH* gene (*FvMYC1*) and nine *MYB* genes (*FvMYB3*, *FvMYB9*, *FvMYB11*, *FvMYB21*, *FvMYB22*, *FvMYB45*, *FvMYB64*, *FvMYB77,* and *FvMYB105*) was higher at green stage, but significantly decreased at the other four stages (**Figure 2** and **Supplementary Figure 3**), matching proanthocyanidins biosynthesis pattern, which accumulated mainly at the early stages, especially at the green stage (**Figure 3A**). Specifically, the expression of *FvMYB11*, *FvMYB21*, *FvMYB77,* and *FvMYB105* were not detected during turning stage and ripe stage. By contrast, the expression of *FvMYB10* was extremely low in green and white stages, but significantly increased during the late development stages, especially the turning stage, resembling anthocyanin biosynthesis pattern, which accumulated in a large amount during turning and maturation stages (**Figure 3B**). The expression of *FvMYB41* and *FvbHLH3* gradually increased accompanying the development of fruit and peaked at ripe stage (**Figure 2** and **Supplementary Figure 3**), also aligned with anthocyanins accumulation (**Figure 3B**). The expression of *FvMYB75* and *FvbHLH33*, however, was detected at all five stages, with higher expression level at early stages from green to pre-turning than turning and ripe stages (**Figure 2** and **Supplementary Figure 3**). Taken together, the *bHLH* gene *FvMYC1*, and the nine *MYB* genes including *FvMYB3*, *FvMYB9*, *FvMYB11*, *FvMYB21*, *FvMYB22*, *FvMYB45*, *FvMYB64*, *FvMYB77,* and *FvMYB105,* might be involved in proanthocyanidins biosynthesis at early stages of fruit development. *FvMYB10*, *FvMYB41,* and *FvbHLH3* might be associated with the accumulation of anthocyanins at later stages. The proposed functions of woodland strawberry *MYB* and *bHLH* genes are consistent with the phylogenetical analysis (**Supplementary Figure 1**).

Interaction Between the Putative Components of Woodland Strawberry MBW Complexes *in vitro* and *in vivo*

To study whether these potential MYB and bHLH proteins can form MBW complexes, we first investigated the interaction among them *in vitro* using LexA yeast two-hybrid system. When FvbHLH3 or FvMYC1 were used as a bait (fused with DNA-binding domain, BD), they both could interact with every MYB protein tested (fused with Activation domain, AD) except for FvMYB45 or FvMYB77 (**Figure 4**). FvbHLH33, however, displayed strong self-activation when fused to LexA BD as a bait. This is also true when only the N-terminal fragment of FvbHLH33 was used (**Supplementary Figure 4**). Considering that 3-AT can be used to inhibit self-activation in GAL4 yeast two-hybrid system, the interaction between FvbHLH33 and MYB proteins were studied in GAL4 system instead. 20 mM 3-AT was introduced to inhibit self-activation using FvbHLH33 as a bait, and we were able to detect interactions between FvbHLH33 and 10 individual MYB proteins, including FvMYB3, FvMYB9, FvMYB10, FvMYB11, FvMYB21, FvMYB22, FvMYB41, FvMYB75, FvMYB77, and FvMYB105, but not FvMYB45 or FvMYB64 (**Supplementary Figure 5**).

We also performed BiFC experiments to furtherly investigate the interactions between bHLH and MYB proteins *in vivo*. As shown in **Figure 5**, when MYBs or FvbHLH3 were expressed individually, no over-lapping fluorescent signal of YFP was seen. When FvbHLH3 and any of the 10 MYB proteins, including FvMYB3, FvMYB9, FvMYB10, FvMYB11, FvMYB21, FvMYB22, FvMYB41, FvMYB64, FvMYB75, and FvMYB105, were co-expressed in tobacco cells, strong fluorescent signal of YFP was observed. However, when FvbHLH3 and FvMYB45 or FvMYB77 were co-expressed, no YFP signal was detected. Similarly, FvMYC1 or FvbHLH33 was observed to interact with the same 10 MYB proteins, but not FvMYB45 or FvMYB77 (**Supplementary Figure 6**). The interaction between FvMYB64 and FvbHLH33 was not detected in yeast two-hybrid, but the interaction was detected in BiFC assay, likely due to the difference between the *in vitro* and *in vivo* experimental systems. Taken together, the results *in vitro* and *in vivo* suggested that

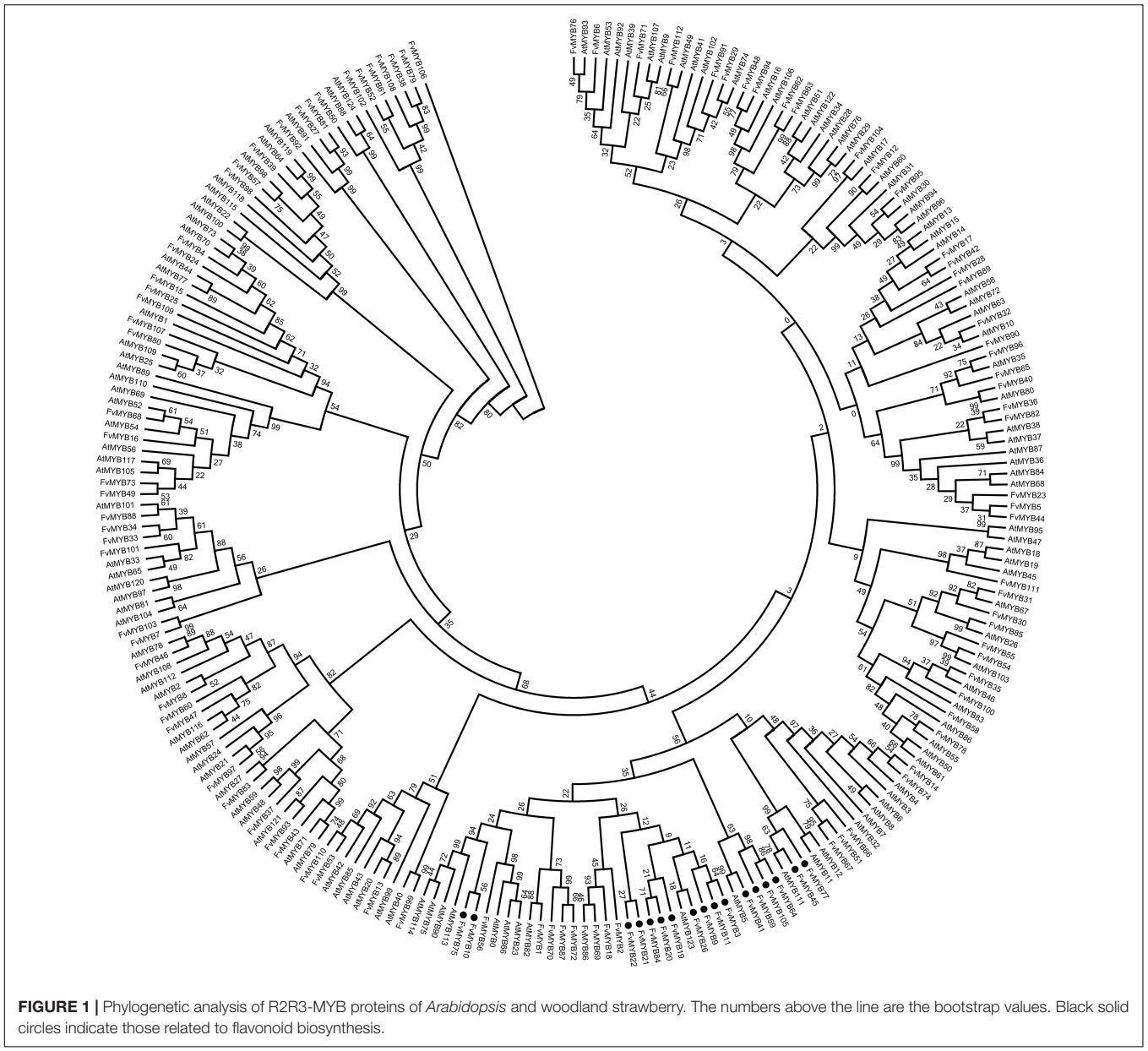


FIGURE 1 | Phylogenetic analysis of R2R3-MYB proteins of *Arabidopsis* and woodland strawberry. The numbers above the line are the bootstrap values. Black solid circles indicate those related to flavonoid biosynthesis.

the 10 MYB proteins, except for FvMYB45 and FvMYB77, could possibly form functional MBW complexes with FvbHLH3, FvbHLH33, or FvMYC1.

Most MYB Proteins Bind to the Promoters of Key Enzyme Genes in Flavonoids Biosynthesis

In a MBW complex, MYB protein is the decisive factor for the regulation of specific flavonoids synthesis genes. To confirm whether the identified MYB proteins have a potential role in regulating genes involved in flavonoids biosynthesis, we first analyzed whether the promoters of CHS2 and DFR2 contained *cis*-elements bound by MBW complexes. As shown in **Supplementary Figure 7**, the promoter regions of CHS2 and DFR2 contain binding elements of MYB proteins. Then, we tested in yeast one-hybrid system to see whether these MYB proteins are able to bind to the promoters of *CHS2* and *DFR2* genes, which encode key enzymes related to the flavonoids synthesis pathways. As shown in **Figure 6**, all tested MYB proteins, except for FvMYB75, displayed binding activity to the promoter of *CHS2* gene. However, only FvMYB11 and FvMYB21 bound to the *DFR2* promoter.

The Heterodimers Formed by MYBs and bHLHs Activate the Expression From DFR2 Promoter

We observed that most of the MYB proteins (except for FvMYB11 and FvMYB21) in strawberry could not bind to the

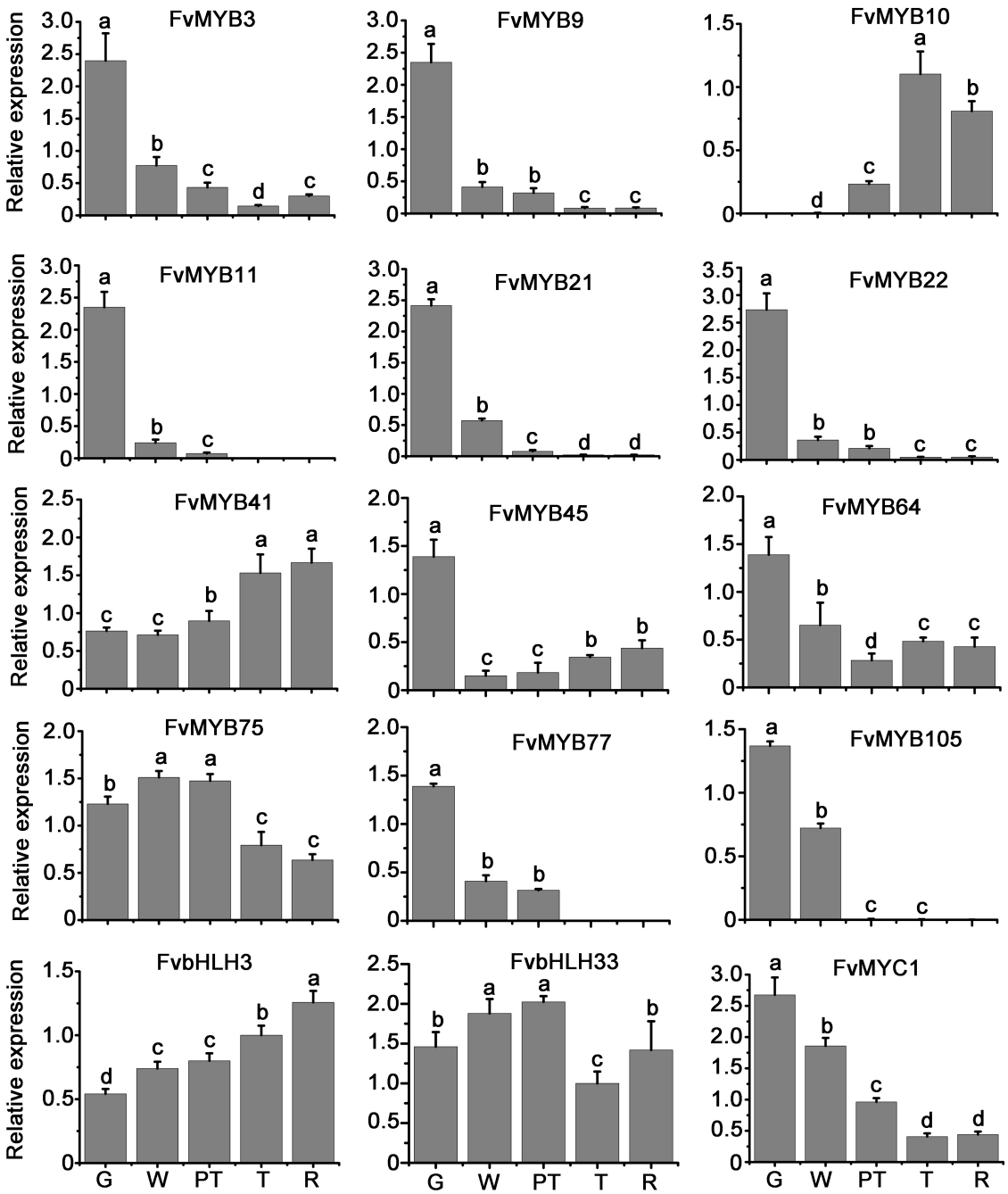


FIGURE 2 | Relative expression levels of *MYB* and *bHLH* genes in fruits of different development stages. Values were normalized to the expression level of an internal control, *Fv26S*. Data are presented as the mean of biological replicates ± SD ($n = 3$). G, green stage; W, white stage; PT, pre-turning stage; T, turning stage; R, ripe stage. The letters a–d indicate statistically significant differences, as determined by Tukey's LSD test ($P \leq 0.05$).

promotor of *DFR2* in yeast one-hybrid assay (**Figure 6**), nor the three bHLH proteins (FvbHLH3, FvbHLH33, or FvMYC1) (**Supplementary Figure 8**). To explore whether bHLH proteins are in need for those MYBs to bind to the promotor of *DFR2*, we test the DNA-binding activity of heterodimers formed between five FvMYB proteins (FvMYB10, FvMYB22, FvMYB64, FvMYB75, and FvMYB105) and three FvbHLH proteins (FvbHLH3, FvbHLH33, or FvMYC1) in yeast one-hybrid assay. As shown in **Figure 7**, transformation of yeast cells with either one of the five MYB proteins fused to AD did not show positive *DFR2* promotor binding activity. However, when either one of the three bHLH proteins, FvbHLH3, FvbHLH33 or FvMYC1, was co-expressed with AD-FvMYB22, AD-FvMYB64, or AD-FvMYB105, the report gene of *LacZ* was activated and the yeast

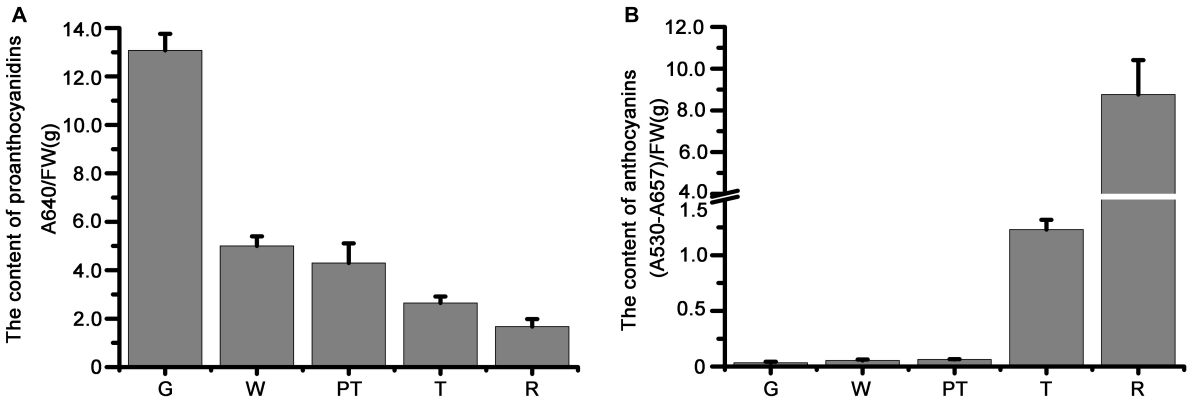


FIGURE 3 | The measurement of anthocyanins and proanthocyanidins in fruits of different development stages. **(A)** Proanthocyanidins; **(B)** anthocyanins. G, green stage; W, white stage; PT, pre-turning stage; T, turning stage; R, ripe stage.

cells displayed blue on media containing *X*-gal, demonstrating the binding of MYB proteins to the promotor of *DFR2* in the presence of bHLH proteins. For FvMYB75 and FvMYB10, the reporter gene *LacZ* can only be activated in the presence of FvMYC1, but not FvbHLH3 and FvbHLH33 (**Supplementary Figure 9**), suggesting a requirement of specific bHLH protein for the DNA-binding activity of FvMYB75 or FvMYB10.

To further confirm the requirement of FvbHLH proteins for the DNA-Binding activity of FvMYB proteins *in vivo*, we transiently expressed FvbHLH33 and three FvMYB proteins (FvMYB22, FvMYB64, and FvMYB105) separately, or as a combination of FvbHLH33 with one of the three FvMYB proteins in *N. benthamiana* leaves, along with a reporter luciferase gene under the control of *DFR2* promoter. As shown in **Figures 8A,B**, when FvMYB22, FvMYB64, or FvMYB105 was expressed alone, the reporter gene was not significantly induced, whereas these MYB proteins induced the reporter gene in the presence of FvbHLH33, leading to a strong luminescence signal when the substrate of luciferin was added. These results, together with results from yeast one-hybrid suggested that the heterodimers between FvMYBs and FvbHLHs were necessary for activating some flavonoids synthesis genes.

Co-expression of FvMYB and FvbHLH Proteins Promotes the Accumulation of Proanthocyanidins

To confirm the biofunction of strawberry MBW complexes in flavonoids synthesis, we co-expressed FvMYB and FvbHLH proteins in the fruits and measured the flavonoids accumulation. FvMYB22, FvMYB64, and FvMYB105 are homologs of MYB proteins in other plants that regulate proanthocyanidin synthesis (**Supplementary Figure 1**) and mainly expressed at green stage (**Figure 2** and **Supplementary Figure 3**). FvbHLH33 has been shown to interact with all three FvMYB proteins in yeast and tobacco cells. We transiently expressed FvbHLH33 and the three FvMYB proteins individually, or a combination of FvbHLH33 and one of the three FvMYB proteins in green stage strawberry fruit. The accumulation of proanthocyanidins was measured following the expression. As shown in **Figures 9A,B** and **Supplementary Figure 10**, the accumulation of procyanidins did not significantly increase following the expression of either FvbHLH33 or any of the three FvMYB proteins alone. However, the expression levels of flavonoid biosynthesis related genes CHS2 and DFR2 was significantly enhanced and the accumulation of proanthocyanidins was significantly increased in strawberry fruits when FvbHLH33 and one of the three FvMYB proteins were co-expressed (**Figures 9A,B** and **Supplementary Figure 10**). These results confirmed the function of the identified MBW components in strawberry flavonoids biosynthesis.

DISCUSSION

In this study, we identified 112 *R2R3-MYB* genes in woodland strawberry by using HMM profile of MYB DNA-binding domain to blast strawberry genome database and further verified the presence of the MYB DNA-binding domain in Pfam, SMART and HHMER (**Supplementary Table 4**). Recently, Liu et al. (2019) have identified *R2R3-MYB* genes in six Rosaceae species including strawberry (*F. vesca*), but they identified 105 *R2R3-MYB* genes which was seven genes fewer than what we found. The difference was likely because they used the 125 R2R3-MYB from *Arabidopsis* to perform repetitive BLASTP search against the strawberry genome version 1.1.a2, in which there were still many scaffolds (Darwish et al., 2015). In our study, we used HMM profile of MYB DNA-binding domain to search against the strawberry genome version 4.0.a2, which contains 34,007 genes, 511 more genes than the 33,496 genes in version 1.1.a2 (Li et al., 2019). We speculate that some *R2R3-MYB* genes might have been omitted or wrongly identified due to the annotation errors in the version 1.1.a2. For example, mrna32181, mrna16655, mrna01311 and mrna28755 were identified as *R2R3-MYB* genes in version 1.1.a2 by Li et al. (2019). When using these genes to blast against version 4.0.a2, each of these four genes was annotated as two or three different genes in version 4.0.a2 (**Supplementary Figure 11**), indicating that there were wrong annotations in version 1.1.2a. Given that genomic information in version 4.0 is at

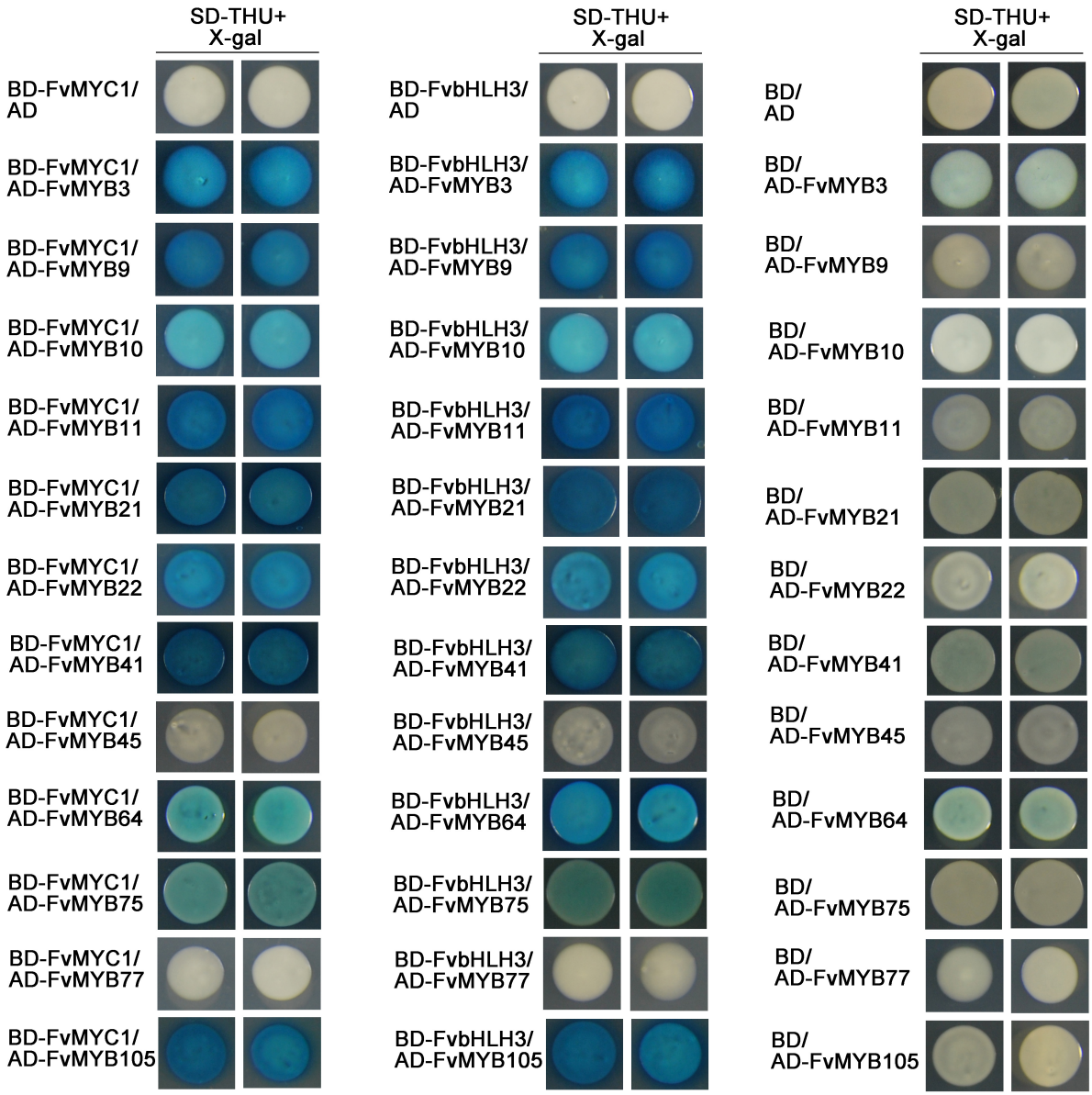


FIGURE 4 | The interaction of MYB and bHLH proteins in yeast cells. LexA yeast two-hybrid assays showing interactions between FvMYB proteins and FvbHLH proteins. Yeast cells were grown on selective medium (SD-Trp, -His, -Ura; SD-THU) with 80 mg/L *X*-gal. Blue precipitates represent cumulative β-galactosidase activity resulting from the activation of the *LacZ* reporter gene by interacting proteins. Two representative colonies were shown for each interaction.

chromosomal level, 112 *R2R3-MYB* identified in our study have higher confidence levels.

It was proposed that in a MBW complex, MYB protein is the decisive factor for the regulation of specific flavonoids synthesis (Baudry et al., 2004). In our study, through phylogenetic tree analysis of *R2R3-MYB* genes from different plants including *Arabidopsis*, grape, pear, strawberry, etc., we obtained 12 potential *MYB* genes that regulate flavonoids synthesis in strawberry. Among these 12 *MYB* genes, *FvMYB10* and *FvMYB75* clustered with *MdMYB1* and *MdMYB10* (**Supplementary Figure 1**), which were known to promote anthocyanin accumulation in apple (Espley et al., 2007; Gonzalez et al., 2008). It has been reported that over-expression of *FvMYB10* or *FaMYB10* (homolog of *FvMYB10* in cultivated strawberry *Fragaria* × *ananassa*) resulted in elevated concentrations of anthocyanin in root, leaves, petioles, stigmas, petals and fruits (Lin-Wang et al., 2014; Medina-Puche et al., 2014; Wang et al., 2020). Additionally, Hawkins et al. (2016) identified an SNP (a specific SNP mutation of G to C, leading to W12S variant) in *FvMYB10* resulting in abolished accumulation of anthocyanins in the fruits of the woodland strawberry Yellow Wonder (YW). Furthermore, in the yellow and red-fruited varieties tested, the C (W12S) in FvMYB10 is 100% associated with the yellow color of the fruit, while the G is 100% associated with the red color (Hawkins et al., 2016;

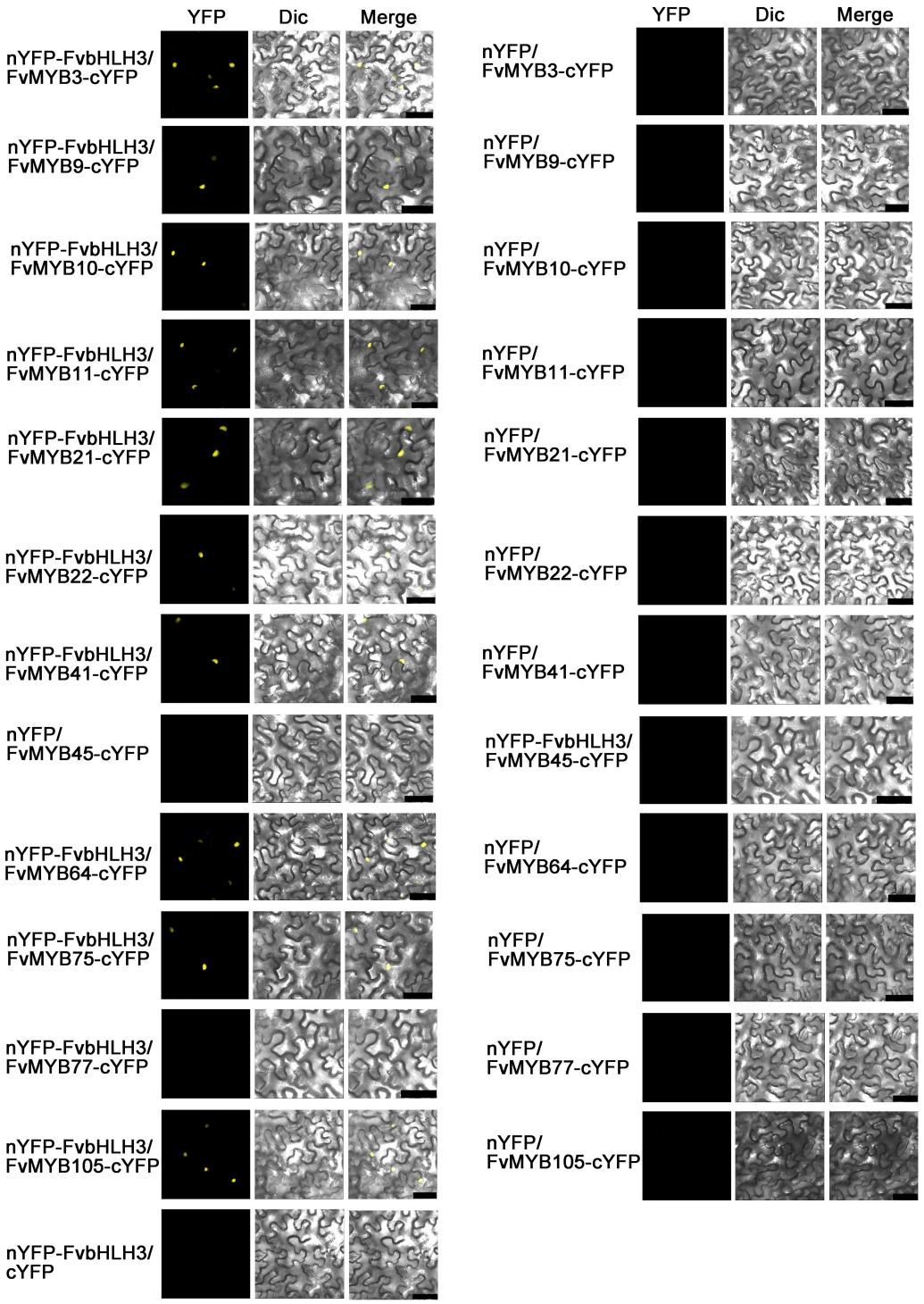


FIGURE 5 | BiFC assay showing interactions between FvMYB proteins and FvbHLH proteins in tobacco cells. Combinations of constructs encoding the indicated proteins were co-transformed into tobacco leaf epidermal cells. Overlays of fluorescence and bright field were shown. Dic, differential interference contrast. Bars, 50 μm.

Zhang et al., 2017). Interestingly in YW, there is no accumulation of anthocyanins in fruits, but the accumulation is normal in petioles (Xu et al., 2014). In addition, two *Fragaria pentaphylla* subvarieties displayed red and white skin color, the relatively low expression level of *FpMYB10* might have contributed to the white fruit variety. The achenes of both varieties, however,

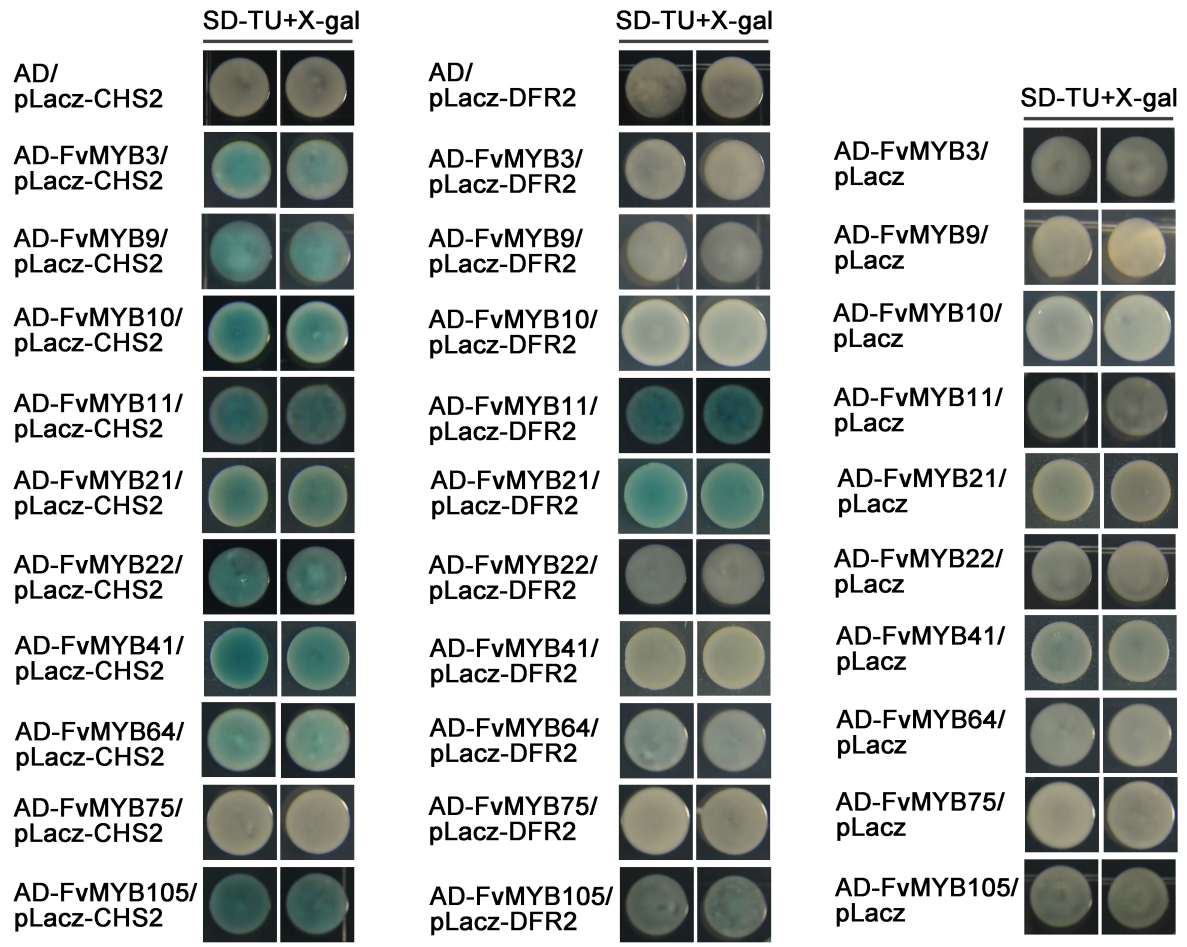


FIGURE 6 | Yeast one-hybrid assays showing the binding of MYB proteins to the promoters of *CHS2* and *DFR2*. Yeast cells were grown on selective medium (SD-Trp, -Ura; SD-TU) with 80 mg/L *X*-gal. Blue precipitates represent cumulative β-galactosidase activity resulting from the activation of the LacZ reporter gene by binding. Two representative colonies were shown for each binding.

are red (Bai et al., 2019). These results suggested that other factors besides FvMYB10 or FpMYB10 might be involved in the regulation of anthocyanin accumulation. In our study, the expression of *FvMYB41* displayed expression pattern similar to *FvMYB10* during the five development stages (**Figure 2** and **Supplementary Figure 3**), resembling anthocyanin biosynthesis pattern (**Figure 3B**), suggesting that FvMYB41 might have a role in promoting anthocyanin accumulation. Through the phylogenetic tree analysis, we found that *FvMYB41* was most genetically related to *VvMYB5b* from grape and *FhMYB5* from *Freesia*, both of which have been proven to promote the accumulation of anthocyanin and proanthocyanidins (Deluc et al., 2008; Li et al., 2018). By comparing the sequences of *VvMYB5b*, *FhMYB5*, and *FvMYB41*, we found that the similarity of all three was greater than 45% (**Supplementary Table 6**). The most conserved sequences were within the DNA-binding R2R3 domain as well as the C1 and C3 motifs in the C-terminal regions of the three genes (**Supplementary Figure 12**). Thus, it was speculated that FvMYB41 likely serves as a homologous protein of VvMYB5b and FhMYB5 in strawberry that promotes flavonoids biosynthesis. However, this notion needs to be tested in future study.

It has been known that MBW complexes composed of MYB-bHLH-WD40 play key roles in regulating flavonoids biosynthesis in plants. In this study, we confirmed the interactions between potential components of MBW in woodland strawberry. We detected the interaction between twelve MYB proteins (FvMYB3, FvMYB9, FvMYB10, FvMYB11, FvMYB21, FvMYB22, FvMYB41, FvMYB45, FvMYB64, FvMYB75, FvMYB77, and FvMYB105) and three bHLH proteins (FvbHLH3, FvbHLH33, and FvMYC1). Using the yeast two hybrid and BiFC assays, we found that except for FvMYB45 and FvMYB77, all the other MYBs can interact with either one of FvbHLH3, FvbHLH33, or FvMYC1 (**Figures 4**, **5** and **Supplementary Figures 5, 6**). In addition, we found that FvTTG1, a WD40 protein, interacted with all the three bHLH proteins (**Supplementary Figure 13**). All these results indicated that the strawberry MBW component proteins identified in this study could form MBW complexes. Interestingly, in octoploid strawberry, FabHLH3, FabHLH33, and FaMYC1 can interact

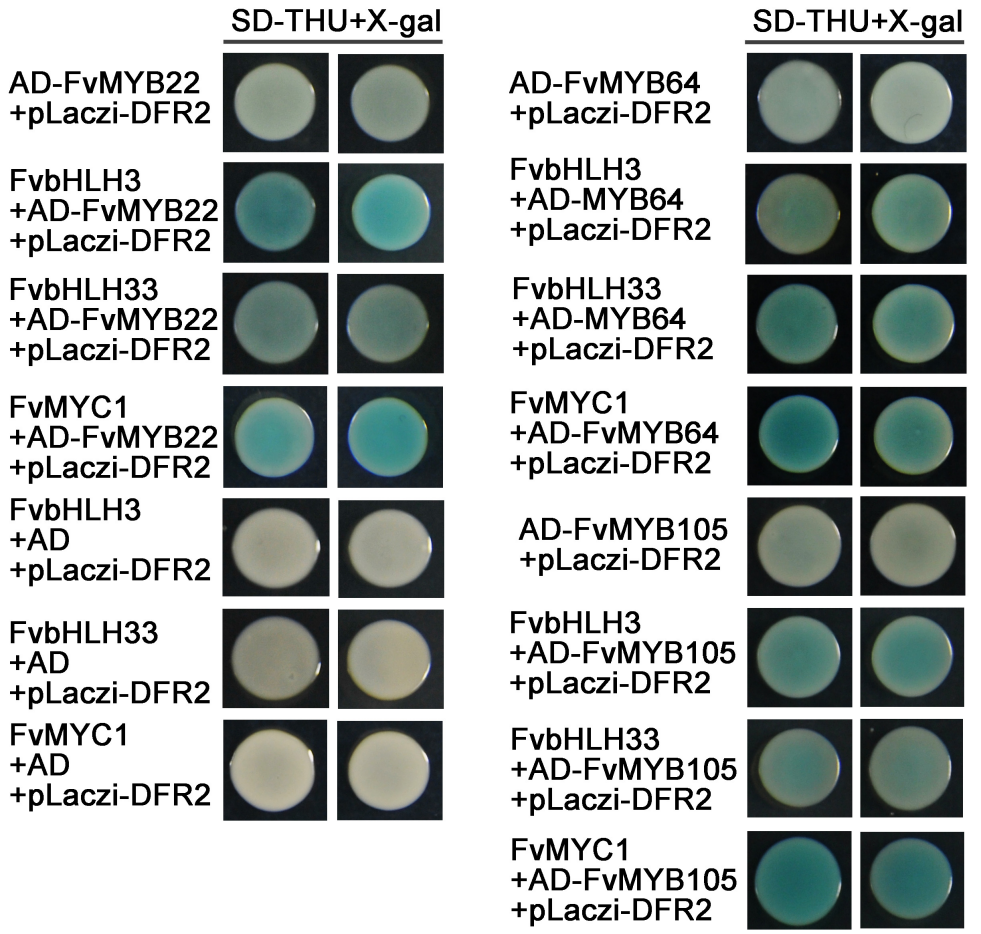


FIGURE 7 | The effect of interactions between MYB and bHLH proteins on the *DFR2* promoter fused with *LacZ* reporter gene. Yeast cells were grown on selective medium (SD-Trp, -His, -Ura; SD-THU) with 80 mg/L *X*-gal. Blue precipitates represent cumulative β-galactosidase activity resulting from the activation of the *LacZ* reporter gene by binding. Two representative colonies were shown for each binding.

with FaTTG1, but only FabHLH3 can interact with FaMYB9, FaMYB11 (Schaart et al., 2013). Whereas, we found that both FvMYB9 and FvMYB11 interacted with FvbHLH33 and FvMYC1 in woodland strawberry (**Figure 4** and **Supplementary Figure 5**). For these paradoxical results between woodland strawberry and octoploid strawberry, it is likely because that Schaart et al. (2013) used the C terminal of FvMYC1 and FvbHLH33 but not the full length of CDS for yeast two-hybrid assays (**Supplementary Figures 14, 15**). Only the full-length protein of FvbHLH33, but not the C-terminus fragment, interacted with MYB proteins in our assay (**Supplementary Figures 5, 16**). Even though Schaart et al. (2013) had demonstrated the direct interactions between FabHLH3 and the two FaMYB proteins (FaMYB9 and FaMYB11), they only did so in yeast cells, not in plant cells (Schaart et al., 2013). In our study, we provided evidence to support the interactions between FvMYB proteins and FvbHLH proteins in both yeast (**Figure 4** and **Supplementary Figure 5**) and plant cells (**Figure 5** and **Supplementary Figure 6**).

To further test whether FvMYB proteins or FvbHLH proteins are involved in the regulation of flavonoids biosynthesis-related genes, we examined the DNA-binding activity of FvMYB proteins, as well as the requirement of FvbHLH proteins, using reporter genes (*LacZ* and *Luciferase*) under the control of the promoters of *CHS2* and *DFR2*, two key enzyme genes related to the flavonoids biosynthesis pathways. Most of the FvMYB proteins except for FvMYB75 displayed DNA-binding activity to *CHS2* promoter. However, most of the FvMYB proteins and three bHLH proteins were not able to bind to *DFR2* promoter except for FvMYB11 and FvMYB21 (**Figure 6** and **Supplementary Figure 8**). It has been reported that in *Arabidopsis*, AtTT2 (AtMYB123) can't bind by itself to the promoter of *BAN*, a gene involved in proanthocyanidins production. AtTT8 (AtbHLH042) must be present for AtTT2 to bind to the *BAN* promoter in yeast (Baudry et al., 2004). Considering that MYB proteins often form complexes with bHLH proteins to regulate flavonoids biosynthesis, we further examined whether co-expression of FvbHLH proteins could promote the binding of FvMYB proteins to the promoter of *DFR2*. Our results showed that co-expression of either one of the three FvbHLH proteins (FvbHLH3, FvbHLH33, and

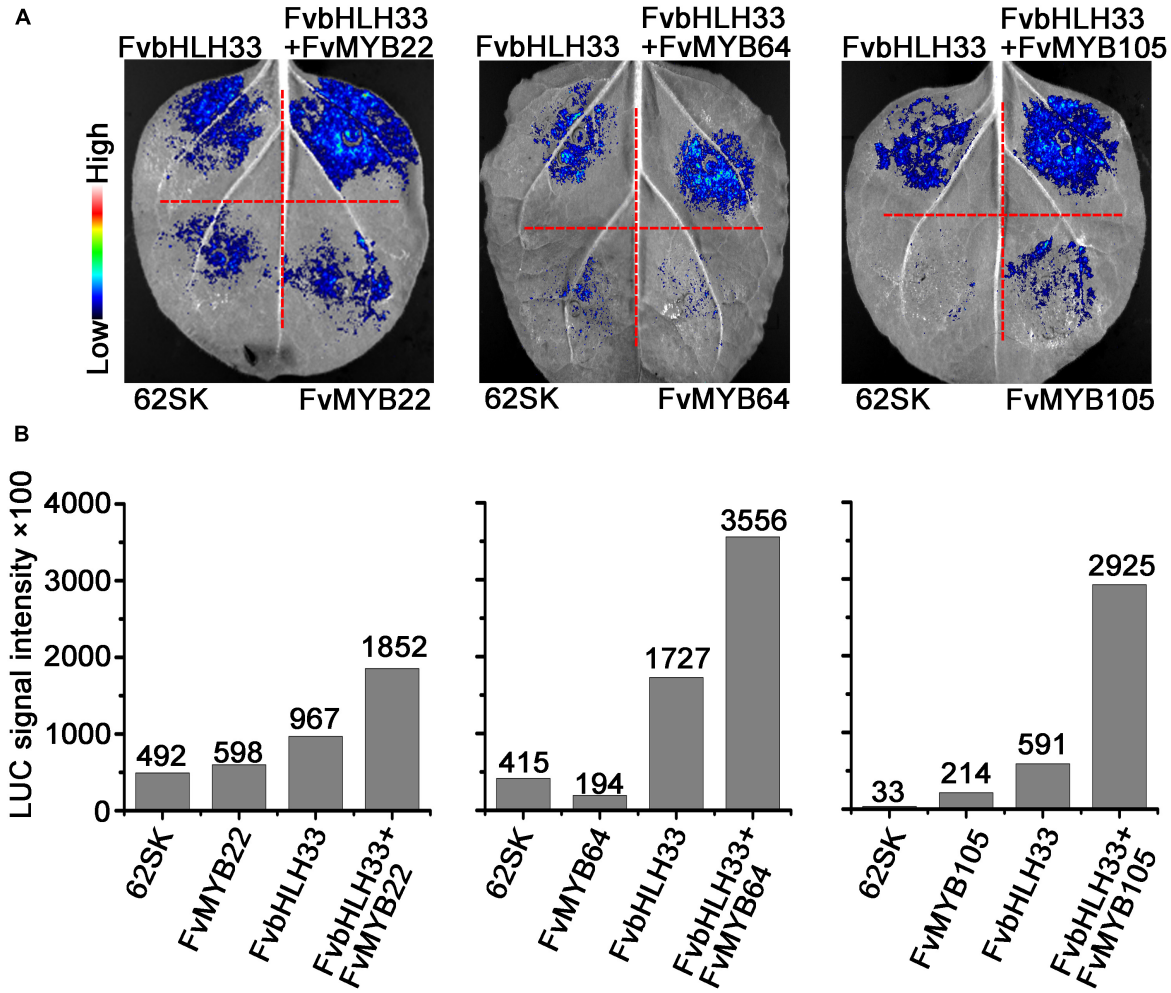


FIGURE 8 | The effect of interactions between MYB and bHLH proteins on the DFR2 promoter fused with luciferase reporter gene. The constructs encoding the indicated combinations of proteins were co-transformed into tobacco leaf epidermal cells, and the luminescence images were captured using a CCD imaging system **(A)**, the intensity of luminescence was indicated in **(B)**.

FvMYC1) with FvMYB22, FvMYB64, or FvMYB105 in yeast cells significantly increased the expression of the reporter gene *LacZ* from *DFR2* promotor, suggested that FvbHLHs can promote the binding activity of FvMYBs to the promoter of *DFR2* (**Figure 7**). The requirement of FvbHLH protein for the gene induction function of FvMYB proteins was also demonstrated in tobacco cells. FvMYB22, FvMYB64, or FvMYB105 induced the expression of luciferase from *DFR2* promoter when FvbHLH33 was co-expressed (**Figure 8**). This was further confirmed by the accumulation of flavonoids in strawberry fruits. When FvbHLH33 was co-expressed with either one of FvMYB22, FvMYB64, or FvMYB105 in green stage strawberry fruits, the accumulation of proanthocyanidins was significantly increased (**Figure 9A**). But for FvMYB75 or FvMYB10, only FvMYC1 could help increase their DNA-binding activity, leading to the induction of *LacZ* expression (**Supplementary Figure 9**). These results suggested that some FvMYB proteins only form complexes with specific FvbHLH proteins. It has been reported that the maize R2R3-MYB protein C1 co-worked with the bHLH protein R to activate the expression of anthocyanin biosynthesis genes (Grotewold et al., 2000). Later studies showed that the dimerization domain of R behaves as a switch that permits distinct configurations of the regulatory complex to be tethered to different promoters of target genes (Feller et al., 2006; Kong et al., 2012). These studies suggested that bHLH protein may determines the target gene of MBW complexes. The biofunctions of the three FvbHLH proteins in woodland strawberry, including FvbHLH3, FvbHLH33, and FvMYC1, remain to be further investigated.

Since FvMYB22, FvMYB64, and FvMYB105 clustered with VvMYBPA1, FaMYB9, and FaMYB11, which were reported to promote proanthocyanidins biosynthesis. We tested whether these three MYBs can bind to the promoter of *FvLAR*, a key enzyme gene regulating the proanthocyanidins synthesis. Our results showed that FvMYB22, FvMYB64, and FvMYB105 bound to *FvLAR* promoters only in the presence of FvMYC1

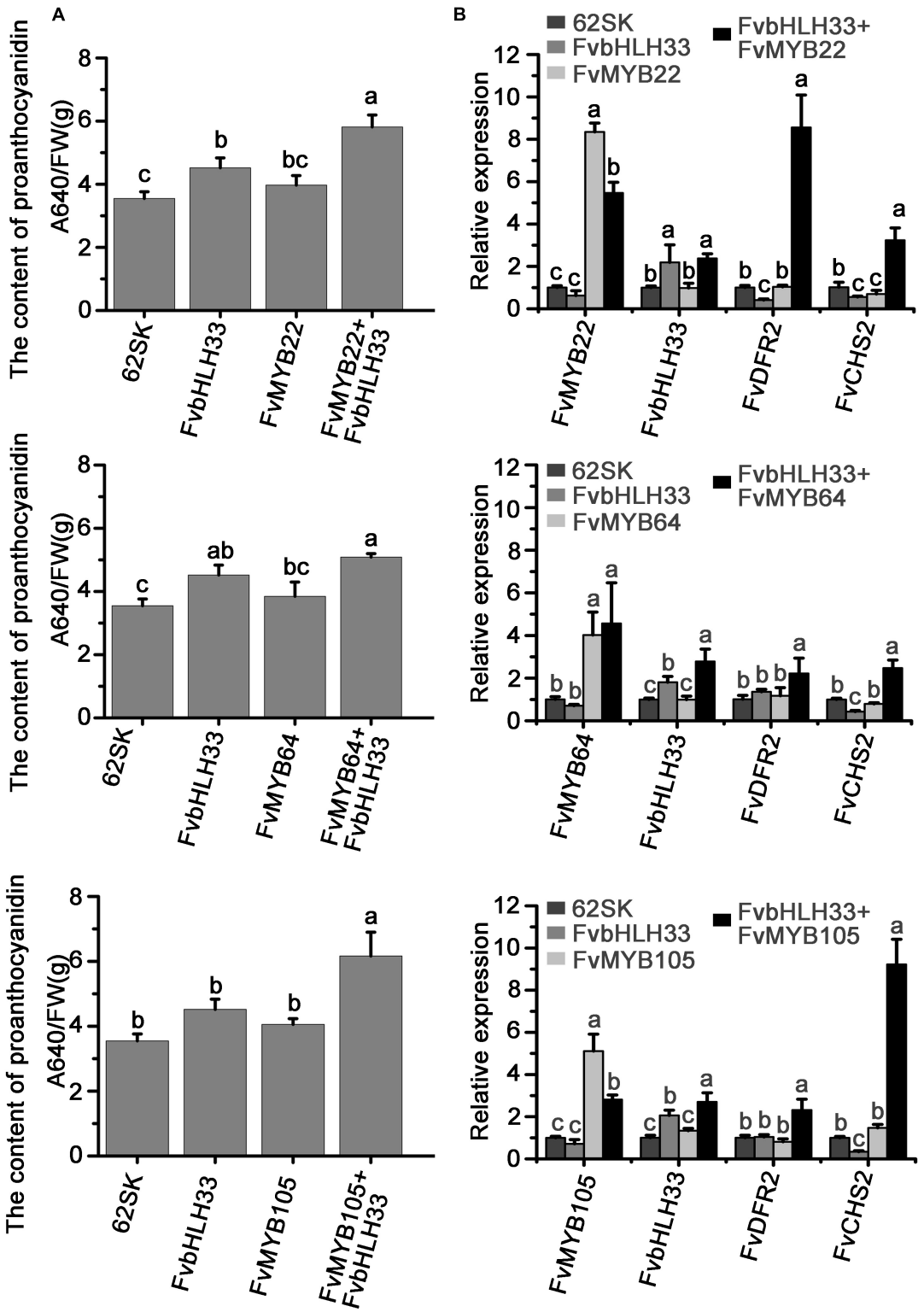


FIGURE 9 | The complex of FvMYB and FvbHLH33 can promote the accumulation of proanthocyanidin in strawberry fruits. The *Agrobacterium* containing the indicated combinations of constructs were injected into strawberry fruits and the content of proanthocyanidins was detected at 7 days after injection. **(A)** Measurement of proanthocyanidin related to the transient expression of FvMYB and FvbHLH proteins. **(B)** Relative expression levels of *FvMYBs, FvbHLH33*, *FvCHS2,* and *FvDFR2* genes from transient expression. *Fv26S* served as an internal control. Data are presented as the mean ± SD ($n = 3$). The letters a–c indicate statistically significant differences, as determined by Tukey's LSD test ($P \leq 0.05$).

(**Supplementary Figure 17**). Considering the interaction with other FvMYB proteins in activating *CHS2* and *DFR2* promoters, FvMYC1 might be a more universal bHLH partner for FvMYB proteins and play more extensive roles in flavonoids biosynthesis than the other two bHLH proteins, FvbHLH3 and FvbHLH33. In fact, even when the transcripts of *FvbHLH33* was significantly reduced in FvbHLH33 RNAi lines, the pigmentation was not significantly altered in those lines when compared with WT (Lin-Wang et al., 2014). This also indicated that other bHLH proteins, especially FvMYC1, might compensate for the lack of FvbHLH33 in this case. Recently, it was found that FvbHLH9 can promote anthocyanin accumulation in strawberry fruits, but FvMYB10 could not promote the expression of downstream target genes when only bHLH9 was present (Li et al., 2020), likely because the lack of WD40 protein. The results of existing studies strongly suggest that there are other regulatory MBW proteins in strawberry that control the synthesis and accumulation of anthocyanins.

CONCLUSION

In this study, we identified novel MBW components in woodland strawberry utilizing phylogenetic analysis, and confirmed the biofunctions of several members in flavonoids accumulation using molecular and biochemical methods. This work provided not only the basis for study of flavonoids synthesis in strawberry, but also potential economic benefit in strawberry cultivation. As transcriptional regulators, further studies of MBW complexes and their biological role in other important physiological processes and regulatory mechanism will contribute to better understanding of strawberry biology.

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

PX and HL designed the experiments. PX, MC, and YL carried out the experiments. PX, MC, and KX collected the data. PX, LW, and HL analyzed the data. PX, LW, CM, YL, and HL wrote the manuscript with input from all other authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by The National Natural Science Foundation of China grants to PX (32002003) and HL (32172536 and 31570282), the China Postdoctoral Science Foundation (2019M661488 and 2020T130415 to PX), the National Natural Science Foundation of Shanghai (20ZR1426000 to HL), and Shanghai Pujiang Program (Grant: 19PJ1406700).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.774943/full#supplementary-material

REFERENCES

Azuma, A., Kobayashi, S., Mitani, N., Shiraishi, M., Yamada, M., Ueno, T., et al. (2008). Genomic and genetic analysis of Myb-related genes that regulate anthocyanin biosynthesis in grape berry skin. *Theor. Appl. Genet.* 117, 1009–1019. doi: 10.1007/s00122-008-0840-1

Bai, L., Chen, Q., Jiang, L., Lin, Y., Ye, Y., Liu, P., et al. (2019). Comparative transcriptome analysis uncovers the regulatory functions of long noncoding RNAs in fruit development and color changes of Fragaria pentaphylla. *Hortic. Res-England* 6:42. doi: 10.1038/s41438-019-0128-4

Battino, M., Beekwilder, J., Denoyes-Rothan, B., Laimer, M., McDougall, G. J., and Mezzetti, B. (2009). Bioactive compounds in berries relevant to human health. *Nutr. Rev.* 67(Suppl. 1), S145–S150.

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Bogs, J., Jaffe, F. W., Takos, A. M., Walker, A. R., and Robinson, S. P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiol.* 143, 1347–1361. doi: 10.1104/pp.106.093203

Castillejo, C., Waurich, V., Wagner, H., Ramos, R., Oiza, N., Munoz, P., et al. (2020). Allelic variation of MYB10 is the major force controlling natural variation in skin and flesh color in strawberry (*Fragaria spp.*) fruit. *Plant Cell* 32, 3723–3749. doi: 10.1105/tpc.20.00474

Darwish, O., Shahan, R., Liu, Z. C., Slovin, J. P., and Alkharouf, N. W. (2015). Re-annotation of the woodland strawberry (*Fragaria vesca*) genome. *BMC Genom.* 16:29. doi: 10.1186/s12864-015-1221-1

Deluc, L., Bogs, J., Walker, A. R., Ferrier, T., Decendit, A., Merillon, J. M., et al. (2008). The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiol.* 147, 2041–2053. doi: 10.1104/pp.108.118919

deVetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev.* 11, 1422–1434. doi: 10.1101/gad.11.11.1422

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.

Emiliani, J., Grotewold, E., Falcone Ferreyra, M. L., and Casati, P. (2013). Flavonols protect *Arabidopsis* plants against UV-B deleterious effects. *Mol. Plant* 6, 1376–1379. doi: 10.1093/mp/sst021

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor. MdMYB10. *Plant J.* 49, 414–427.

Feller, A., Hernandez, J. M., and Grotewold, E. (2006). An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. *J. Biol. Chem.* 281, 28964–28974. doi: 10.1074/jbc.M603262200

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.

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

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

Grotewold, E., Drummond, B. J., Bowen, B., and Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* 76, 543–553. doi: 10.1016/0092-8674(94)90117-1

Grotewold, E., Sainz, M. B., Tagliani, L., Hernandez, J. M., Bowen, B., and Chandler, V. L. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc. Natl. Acad. Sci. U S A.* 97, 13579–13584. doi: 10.1073/pnas.250379897

Hawkins, C., Caruana, J., Schiksnis, E., and Liu, Z. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci. Rep.* 6:29017. doi: 10.1038/srep29017

He, J. A., and Giusti, M. M. (2010). Anthocyanins: natural colorants with health-promoting properties. *Annu. Rev. Food Sci. T* 1, 163–187.

Hichri, I., Heppel, S. C., Pillet, J., Leon, C., Czemmel, S., Delrot, S., et al. (2010). The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Mol. Plant* 3, 509–523. doi: 10.1093/mp/ssp118

Kelemen, Z., Sebastian, A., Xu, W. J., Grain, D., Salsac, F., Avon, A., et al. (2015). Analysis of the DNA-binding activities of the *Arabidopsis* R2R3-MYB transcription factor family by one-hybrid experiments in yeast. *PLoS One* 10:e0141044. doi: 10.1371/journal.pone.0141044

Kong, Q., Pattanaik, S., Feller, A., Werkman, J. R., Chai, C. L., Wang, Y. Q., et al. (2012). Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc. Natl. Acad. Sci. U S A.* 109, E2091–E2097. doi: 10.1073/pnas.1205513109

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

Li, S. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal. Behav.* 9:e27522. doi: 10.4161/psb.27522

Li, Y., Shan, X., Zhou, L., Gao, R., Yang, S., Wang, S., et al. (2018). The R2R3-MYB factor FhMYB5 from freesia hybrida contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis. *Front.Plant Sci.* 9:1935. doi: 10.3389/fpls.2018.01935

Li, Y., Xu, P., Chen, G., Wu, J., Liu, Z., and Lian, H. (2020). FvbHLH9, functions as a positive regulator of anthocyanin biosynthesis, by forming HY5-bHLH9 transcription complex in strawberry fruits. *Plant Cell Physiol.* 61, 826–837. doi: 10.1093/pcp/pcaa010

Li, Y. P., Pi, M. T., Gao, Q., Liu, Z. C., and Kang, C. Y. (2019). Updated annotation of the wild strawberry *Fragaria vesca* V4 genome. *Hortic. Res-England* 6:61. doi: 10.1038/s41438-019-0142-6

Lin-Wang, K., McGhie, T. K., Wang, M., Liu, Y., Warren, B., Storey, R., et al. (2014). Engineering the anthocyanin regulatory complex of strawberry (*Fragaria vesca*). *Front. Plant Sci.* 5:651. doi: 10.3389/fpls.2014.00651

Liu, H., Xiong, J. S., Jiang, Y. T., Wang, L., and Cheng, Z. M. (2019). Evolution of the R2R3-MYB gene family in six *Rosaceae species* and expression in woodland strawberry. *J. Integr. Agr.* 18, 2753–2770.

Lu, X.-D., Zhou, C.-M., Xu, P.-B., Luo, Q., Lian, H.-L., and Yang, H.-Q. (2015). Red-light-dependent interaction of phyB with SPA1 promotes COP1–SPA1 dissociation and photomorphogenic development in *Arabidopsis*. *Mol. Plant* 8, 467–478. doi: 10.1016/j.molp.2014.11.025

Medina-Puche, L., Cumplido-Laso, G., Amil-Ruiz, F., Hoffmann, T., Ring, L., Rodriguez-Franco, A., et al. (2014). MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of *Fragaria* x ananassa fruits. *J. Exp. Bot.* 65, 401–417. doi: 10.1093/jxb/ert377

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096.

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

Nijveldt, R. J., van Nood, E., van Hoorn, D. E., Boelens, P. G., van Norren, K., and van Leeuwen, P. A. (2001). Flavonoids: a review of probable mechanisms of action and potential applications. *Am. J. Clin. Nutr.* 74, 418–425. doi: 10.1093/ajcn/74.4.418

Prouse, M. B., and Campbell, M. M. (2012). The interaction between MYB proteins and their target DNA binding sites. *Bba-Gene Regul. Mech.* 1819, 67–77. doi: 10.1016/j.bbagrm.2011.10.010

Quattrocchio, F., Wing, J. F., Leppen, H., Mol, J., and Koes, R. E. (1993). Regulatory genes controlling anthocyaninpigmentation are functionally conserved among plant species and have distinct sets of target genes. *Plant Cell* 5, 1497–1512. doi: 10.1105/tpc.5.11.1497

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011

Schaart, J. G., Dubos, C., Fuente, I. R. D. L., Houwelingen, A. M. M. L. V., Vos, R. C. H. D., Jonker, H. H., et al. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry. *New Phytol.* 197, 454–467. doi: 10.1111/nph.12017

Spelt, C., Quattrocchio, F., Mol, J. N. M., and Koes, R. (2000). Anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* 12, 1619–1631. doi: 10.1105/tpc.12.9.1619

Stracke, R., Ishihara, H., Barsch, G. H. A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

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

Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verries, C., Cheynier, V., et al. (2009). Ectopic expression of VvMYBPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiol.* 149, 1028–1041. doi: 10.1104/pp.108.131862

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Tuan, P. A., Bai, S. L., Yaegaki, H., Tamura, T., Hihara, S., Moriguchi, T., et al. (2015). The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. *BMC Plant Biol.* 15:280. doi: 10.1186/s12870-015-0664-5

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., et al. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1349. doi: 10.1105/tpc.11.7.1337

Wang, H., Zhang, H., Yang, Y., Li, M., Zhang, Y., Liu, J., et al. (2020). The control of red colour by a family of MYB transcription factors in octoploid strawberry (*Fragaria x ananassa*) fruits. *Plant Biotechnol. J.* 18, 1169–1184. doi: 10.1111/pbi.13282

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

Xu, F., Li, T., Xu, P. B., Li, L., Du, S. S., Lian, H. L., et al. (2016). DELLA proteins physically interact with CONSTANS to regulate flowering under long days in *Arabidopsis*. *FEBS Lett.* 590, 541–549. doi: 10.1002/1873-3468.12076

Xu, P., Lian, H., Xu, F., Zhang, T., Wang, S., Wang, W., et al. (2019). Phytochrome Band AGB1 coordinately regulate photomorphogenesis by antagonistically modulating PIF3 stability in *Arabidopsis*. *Mol. Plant* 12, 229–247. doi: 10.1016/j.molp.2018.12.003

Xu, P., Zawora, C., Li, Y., Wu, J., Liu, L., Liu, Z., et al. (2018). Transcriptome sequencing reveals role of light in promoting anthocyanin accumulation of strawberry fruit. *Plant Growth Regul.* 86, 121–132. doi: 10.1007/s10725-018-0415-3

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Xu, W., Peng, H., Yang, T., Whitaker, B., Huang, L., Sun, J., et al. (2014). Effect of calcium on strawberry fruit flavonoid pathway gene expression and anthocyanin accumulation. *Plant Physiol. Biochem.* 82, 289–298. doi: 10.1016/j.plaphy.2014.06.015

Zhang, J. X., Zhang, Y. C., Dou, Y. J., Li, W. J., Wang, S. M., Shi, W. J., et al. (2017). Single nucleotide mutation in FvMYB10 may lead to the yellow fruit in *Fragaria vesca*. *Mol. Breed.* 37:35.

Zhang, Y., Butelli, E., and Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19, 81–90. doi: 10.1016/j.pbi.2014.05.011

Zhang, Z., Shi, Y., Ma, Y., Yang, X., Yin, X., Zhang, Y., et al. (2020). The strawberry transcription factor FaRAV1 positively regulates anthocyanin accumulation by activation of FaMYB10 and anthocyanin pathway genes. *Plant Biotechnol. J.* 18, 2267–2279. doi: 10.1111/pbi.13382

Zoratti, L., Karppinen, K., Luengo Escobar, A., Häggman, H., and Jaakola, L. (2014). Light-controlled flavonoid biosynthesis in fruits. *Front. Plant Sci.* 5:5345. doi: 10.3389/fpls.2014.00534

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fpls.2021.774446

MdBBX21, a B-Box Protein, Positively Regulates Light-Induced Anthocyanin Accumulation in Apple Peel

Bo Zhang[1,2†], Zhen-Zhen Zhu[1,2†], Dong Qu[3], Bo-Chen Wang[1,2], Ni-Ni Hao[1,2], Ya-Zhou Yang[1,2], Hui-Juan Yang[1,2] and Zheng-Yang Zhao[1,2*]*

[1] State Key Laboratory of Crop Stress Biology for Arid Areas, College of Horticulture, Northwest A&F University, Yangling, China, [2] Shaanxi Research Center of Apple Engineering and Technology, Yangling, China, [3] Shaanxi Key Laboratory Bio-resources, College of Bioscience and Engineering, Shaanxi University of Technology, Hanzhong, China

The red coloration of apple (*Malus* × *domestica* Borkh.) is due to the accumulation of anthocyanins in the fruit peel. Light is essential for anthocyanin biosynthesis in apple. In this study, we performed a transcriptome sequencing (RNA-seq) analysis of apple fruit exposed to light after unbagging. The identified differentially expressed genes included *MdBBX21*, which is homologous to Arabidopsis *BBX21*, suggesting it may be involved in light-induced anthocyanin biosynthesis. Additionally, MdBBX21 was localized in the nucleus and its gene was expressed earlier than *MdMYB1* in apple peel treated with light. Overexpressing *MdBBX21* in Arabidopsis and apple calli under light increased anthocyanin accumulation. Dual-luciferase and yeast one-hybrid assays confirmed that MdBBX21 binds to the *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* promoters and induces expression. At the same time, MdHY5 can also activate the expression of *MdBBX21*. Furthermore, bimolecular fluorescence complementation and yeast two-hybrid assays demonstrated that MdBBX21 can interact with MdHY5. This interaction can significantly enhance *MdMYB1* promoter activity. These findings clarify the molecular mechanism by which MdBBX21 positively regulates light-induced anthocyanin accumulation in apple.

Keywords: apple, light, anthocyanin, MdBBX21, MdMYB1

OPEN ACCESS

Edited by:
Yongliang Liu,
University of Kentucky, United States

Reviewed by:
Shaohua Zeng,
South China Botanical Garden, Chinese Academy of Sciences (CAS), China
Richard Victor Espley,
The New Zealand Institute for Plant and Food Research Ltd., New Zealand

***Correspondence:**
Hui-Juan Yang
huijuanyang@nwsuaf.edu.cn
Zheng-Yang Zhao
zhaozy@nwsuaf.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *12 September 2021*
Accepted: *26 October 2021*
Published: *12 November 2021*

Citation:
Zhang B, Zhu Z-Z, Qu D, Wang B-C, Hao N-N, Yang Y-Z, Yang H-J and Zhao Z-Y (2021) MdBBX21, a B-Box Protein, Positively Regulates Light-Induced Anthocyanin Accumulation in Apple Peel. Front. Plant Sci. 12:774446. doi: 10.3389/fpls.2021.774446

INTRODUCTION

Apples are one of the most widely cultivated fruits worldwide. Red-skinned apples are more popular with consumers than green-skinned or yellow-skinned apples. The redness of the peel is mainly determined by anthocyanins, which are natural water-soluble pigments (Kim et al., 2003; Zhang Y. et al., 2014). Under natural conditions, anthocyanins are stored in plant vacuoles in the form of glycosides (Passeri et al., 2016). The diversity in the colors of flowers, stems, leaves, fruits, and other plant organs and tissues is due to the type, content, and distribution of anthocyanins (Lawrence et al., 1939; Zhang H. et al., 2014).

Anthocyanins are produced by the flavonoid pathway in a process that is controlled by a series of structural genes and regulatory factors (Winkel-Shirley, 2001). The structural genes (e.g., early and late biosynthetic genes) encode various enzymes in the plant anthocyanin biosynthetic pathway. The early biosynthetic genes include those encoding chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3-hydroxylase (F3H), whereas the late biosynthetic genes include those encoding dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), and flavonoid 3-O-glycosyltransferase (UFGT) (Martin et al., 1991; Kubasek et al., 1992). Transcriptional regulators control the expression of structural genes (Quattrocchio et al., 1993, 1998; de Vetten et al., 1997). For example, MYB proteins, which are among the most important transcription factors, can combine with bHLH and WD40 proteins to form an MBW complex that binds to the promoter of structural genes to induce expression and regulate anthocyanin synthesis (Ramsay and Glover, 2005; Li, 2014). In apple, MdMYB1 and MdMYBA are responsible for apple skin coloration (Takos A. M. et al., 2006; Ban et al., 2007). Five direct tandem repeats of the MdMYB10-binding motif in the *MdMYB10* promoter are associated with the accumulation of anthocyanins throughout the plant, ultimately resulting in striking phenotypes (e.g., red fruit flesh and red foliage) (Espley et al., 2007, 2009). By aligning the *MdMYB1* sequence with the GDDH13 and HFTH1 genomes, a long terminal repeat retrotransposon associated with the red-skinned phenotype was detected upstream of *MdMYB1* (Zhang et al., 2019). These results strongly indicate that the *MdMYB1/A/10* alleles are the core transcriptional regulatory genes of the anthocyanin biosynthesis pathway.

In addition to being genetically regulated, anthocyanin production is also affected by various environmental factors, including drought (An et al., 2020), low temperatures (Xie et al., 2012), salt stress (Lotkowska et al., 2015), low nitrogen availability (Sun et al., 2018), and light (Cominelli et al., 2008). For the biosynthesis of anthocyanins in the apple peel, light is the most important environmental factor (Merzlyak and Chivkunova, 2000). The mechanism underlying light-induced anthocyanin synthesis has been well characterized as part of photomorphogenesis in the model plant Arabidopsis (Maier et al., 2013). Light signals are perceived by plants and then transmitted by photoreceptor proteins such as phytochrome (PHY) (Sharrock and Quail, 1989), cryptochrome (CRY) (Lin et al., 1995), phototropin (PHOT) (Kasahara et al., 2002), and ultraviolet light receptor (UVR8) (Rizzini et al., 2011). Different photoreceptor proteins can perceive and transmit light of different wavelengths. The blue light receptor gene *MdCRY2* and the UV-B photoreceptor gene *MdUVR8* have been characterized in apple (Li et al., 2013; Zhao et al., 2016). Downstream of the photoreceptors, LONG HYPOCOTYL 5 (HY5), ubiquitin E3 ligase constitutive photobiochemical enzyme 1 (COP1), and PHYTOCHROME-INTERACTING FACTORS (PIFs) are involved in the complex signaling network mediating the regulation of plant photomorphogenesis (Deng et al., 1992; Ang and Deng, 1994; Ni et al., 1998). In apple, MdHY5 binds directly to the *MdMYB1* promoter to induce expression and promote the accumulation of anthocyanins under light conditions (An et al., 2017). However, under dark conditions, MdCOP1 can interact with MdMYB1 to mediate the ubiquitination and degradation of MdMYB1 (Li et al., 2012).

The B-box (BBX) protein is a zinc finger protein containing one or two B-box motifs (Reymond et al., 2001). It is a transcription factor that regulates plant photomorphogenesis along with HY5, COP1, and PIFs. Members of the BBX family have been identified in many species, including Arabidopsis (Khanna et al., 2009), rice (Huang et al., 2012), tomato (Chu et al., 2016), apple (Liu et al., 2018), and pear (Bai et al., 2019a). In Arabidopsis, the BBX protein family is divided into five subfamilies according to the protein structures. Subfamily IV comprises eight members (BBX18–25), each with two B-box zinc finger motifs (Khanna et al., 2009). The transcription factors in this subfamily are closely related to the light signaling cascade. They can interact with the COP1 and HY5 transcription factors, and can link the light signal regulatory network with other regulatory networks (e.g., hormone and temperature) (Sarmiento, 2013; Song et al., 2020). In apple, BBX22 positively regulates UV-B-induced anthocyanin synthesis, but its function depends on the synergistic effect of HY5 (Bai et al., 2014; An et al., 2019). Additionally, MdBBX20 can form a complex with MdHY5 and bind directly to the *MdMYB1*, *MdDFR*, and *MdANS* promoters to promote anthocyanin synthesis in apple (Fang et al., 2019b). Moreover, MdCOL4 interacts with MdHY5 to inhibit the expression of *MdMYB1*, whereas it can bind directly to the *MdUFGT* and *MdANS* promoters to suppress expression (Fang et al., 2019a). However, whether there are other BBX proteins involved in the synthesis of anthocyanins in apple and the relationships among these BBX proteins remain unclear.

Although the mechanism mediating light-induced anthocyanin synthesis has been partially characterized, a thorough analysis is still needed. In this study, on the basis of transcriptome sequencing data, we revealed that *MdBBX21* is differentially expressed in the peel when unbagged fruit are exposed to light. Additionally, *MdBBX21* responds to light and reaches peak expression levels earlier than *MdBBX20*, *MdBBX22-1/2*, and *MdHY5*. The overexpression of *MdBBX21* in transgenic Arabidopsis plants and apple calli can induce anthocyanin accumulation under light. Further analyses indicated that MdBBX21 can activate the expression of *MdBBX20*, *MdBBX22-1/2*, and *MdHY5*, which subsequently promotes anthocyanin accumulation. Furthermore, MdBBX21 can interact with MdHY5 and induce the expression of *MdMYB1*. These findings clarify the transcriptional regulation that occurs upstream of MdMYB1 during light-induced anthocyanin biosynthesis.

MATERIALS AND METHODS

Plant Materials

Six-year-old "Starkrimson Delicious" apple trees growing at the Baishui Apple Experimental Farm of Northwest A&F University, Shaanxi province, China were used as experimental materials. Fruit were covered with a paper bag (Hongtai, Shanxi, China) at 45 days after blooming and harvested at 135 days after blooming.

The harvested fruit were placed in an incubator at 23°C under continuous white light (200 μmol m^{-2} s^{-1}). The peels were excised at 0, 3, 6, 9, 12, 24, 48, and 72 h after initiating the light treatment. Twelve fruit were selected at each time-point, with four fruit considered as one biological replicate.

Determination of the Anthocyanin Content

Anthocyanins were extracted from the apple peels as previously described (Xie et al., 2012). Briefly, 0.5 g peel was treated with 5 mL 1% (v/v) HCl-methanol and incubated in darkness at 4°C for 24 h. After centrifuging at 13,000 × *g* for 10 min, the anthocyanin content in the upper liquid layer was determined using an HPLC system comprising a Waters 2998 detector (Waters, Milford, MA, United States) and a C18 column (5 μm internal diameter, 250 mm × 4.6 mm; Waters) as previously described (Liu et al., 2013). The anthocyanin content of Arabidopsis was determined as previously described (Wang et al., 2016).

RNA Extraction, Library Preparation, and RNA-seq

Total RNA was extracted using the TRIzol RNA Plant Plus Reagent (Tiangen, Beijing, China). The integrity of the RNA was checked using the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States) and by agarose gel electrophoresis. The purity of the RNA was determined using the NanoPhotometer spectrophotometer (IMPLEN, Westlake Village, CA, United States). Sequencing libraries were constructed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (NE, United States). After verifying the quality of the libraries, they were sequenced using the Illumina Novaseq 6000 sequencer (150-bp paired-end sequencing) (Illumina, San Diego, CA, United States) at Novogene Bioinformatics Technology Co., Ltd., Beijing, China.

After eliminating the reads with sequencing adapters and the low-quality reads from the raw data, the remaining clean reads were aligned to the *Malus* Genome GDDH13 reference sequence (version 1.1) (Daccord et al., 2017) using the HISAT2 software. After the sequence alignment, the raw counts of the mapped reads for each *Malus* gene model in GDDH13 (version 1.1) were determined and then normalized to FPKM per million mapped reads. Using DESeq2, differentially expressed genes (DEGs) were putatively identified according to the following criteria: $|\log_2(\text{fold-change})| > 1$ and adjusted padj < 0.05.

Subcellular Localization

The *MdBBX21* coding sequence (CDS) was cloned from the cDNA derived from "Starkrimson Delicious" fruit peel. The full-length *MdBBX21* open reading frame without the stop codon was inserted into the pCAMBIA2301-eGFP vector under the control of the CaMV 35S promoter to obtain the 35S:MdBBX21-eGFP construct. The primer sequences used are listed in **Supplementary Table 4**. Tobacco (*Nicotiana benthamiana*) leaves were infiltrated with *Agrobacterium tumefaciens* strain GV3101 cells carrying 35S:MdBBX21-eGFP or the empty vector control (pCAMBIA2301-eGFP). After 3 days, the eGFP signal in the tobacco leaves were detected using the LSM 710 confocal laser-scanning microscope (Carl Zeiss) at excitation wavelengths of 488 nm for eGFP.

Analysis of mRNA Expression

First-strand cDNA was synthesized using the HiScript® II Q RT SuperMix for qPCR (Vazyme, Nanjing, China). A quantitative real-time polymerase chain reaction (qPCR) analysis was conducted using the ChamQ Universal SYBR qPCR Master Mix (Vazyme) and the ABI StepOnePlus™ Real-Time PCR System (Applied Biosystems, Waltham, MA, United States). The *MdActin* gene was used as the internal control. All expression data were examined according to the delta-delta cycle threshold method (Livak and Schmittgen, 2001). The primer sequences used are listed in **Supplementary Table 4**.

Generation of Transgenic Plant Materials

"Orin" apple calli were transformed according to a slightly modified version of a previously reported method (Xie et al., 2012). First, calli were infected with *A. tumefaciens* strain LBA4404 cells carrying 35S:MdBBX21-eGFP for 20 min. The calli were then transferred to MS medium supplemented with 1 mg L^{-1} 6-benzylaminopurine (6-BA), 1 mg L^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D), 30 g L^{-1} sucrose, and 8 g L^{-1} agar. After a 3-day incubation in darkness, the calli were transferred to screening medium (i.e., MS medium supplemented with 30 g L^{-1} sucrose, 1 mg L^{-1} 2,4-D, 1 mg L^{-1} 6-BA, 8 g L^{-1} agar, 250 mg L^{-1} carbenicillin, and 50 mg L^{-1} kanamycin) to screen for transformants. For the light treatment, wild-type (WT) and *MdBBX21*-overexpressing (MdBBX21-OX) transgenic apple calli cultivated in darkness were transferred to a light incubator and placed under constant white light (500 μmol m^{-2} s^{-1}) for 3 days.

Agrobacterium tumefaciens strain GV3101 cells carrying 35S:MdBBX21-eGFP were used to infect the Arabidopsis *bbx21* mutant (SALK_105390) according to the floral-dip method (Clough and Bent, 1998). Seeds of WT and T_3 transgenic Arabidopsis plants were chilled at 4°C for 48 h and then placed under white light (500 μmol m^{-2} s^{-1}) at 24°C under long-day conditions (16-h light/8-h dark). The anthocyanin contents of 5-day-old Arabidopsis seedlings were determined.

Yeast Two-Hybrid Assay

A yeast two-hybrid assay was performed using the Matchmaker™ Gold Yeast Two-Hybrid System (Clontech). The full-length *MdBBX21* CDS was inserted into the pGADT7 (AD) vector to construct the AD-BBX21 recombinant plasmid. The *MdHY5* CDS was inserted into the pGBKT7 (BD) vector to obtain the BD-HY5 recombinant plasmid. The primer sequences used are listed in **Supplementary Table 4**. Yeast strain Y2HGold cells were transformed with AD-BBX21 + BD-HY5, AD + BD-HY5, AD-BBX21 + BD, pGADT7-T + pGBKT7-53, or pGADT7-T + pGBKT7-Lam and then cultured on medium lacking tryptophan and leucine at 30°C. To screen for interacting proteins, the yeast cells were transferred to medium lacking

leucine, tryptophan, histidine, and adenine (−L/−T/−H/−A), but supplemented with X-α-gal.

Yeast One-Hybrid Assay

The *MdBBX20*, *MdBBX22-1/2*, *MdHY5*, and *MdMYB1* promoter fragments were inserted into the pHIS2 vector to construct the MdBBX20pro-HIS2, MdBBX22-1pro-HIS2, MdBBX22-2pro-HIS2, MdHY5pro-HIS2, and MdMYB1pro-HIS2 recombinant plasmids. The primer sequences used are listed in **Supplementary Table 4**. To determine the optimal 3-AT concentration, yeast strain Y187 cells containing the recombinant pHIS2 plasmids were grown on screening medium lacking histidine and tryptophan (−H/−T), but supplemented with different 3-AT concentrations. Next, Y187 yeast cells were co-transformed with MdBBX21-AD and individual recombinant pHIS2 plasmids. Interactions were detected on selection medium lacking histidine, tryptophan, and leucine (−H/−T/−L), but supplemented with the optimal 3-AT concentration.

Transient Dual-Luciferase Assay

Transient dual-luciferase assays were performed using tobacco (*N. benthamiana*) leaves. The *MdBBX20*, *MdBBX22-1/2*, *MdHY5*, and *MdMYB1* promoter fragments were inserted into the pGreenII 0800-LUC vector to construct the MdBBX20pro:LUC, MdBBX22-1pro:LUC, MdBBX22-2pro:LUC, MdHY5pro:LUC, and MdMYB1pro:LUC recombinant plasmids. The full-length *MdBBX21* CDS was inserted into the pGreenII 62-SK vector. The primer sequences used are listed in **Supplementary Table 4**. *A. tumefaciens* strain GV3101 cells carrying the pSoup vector were transformed with the recombinant plasmids. Leaves from 5-week-old tobacco (*N. benthamiana*) plants were injected according to a previously described method

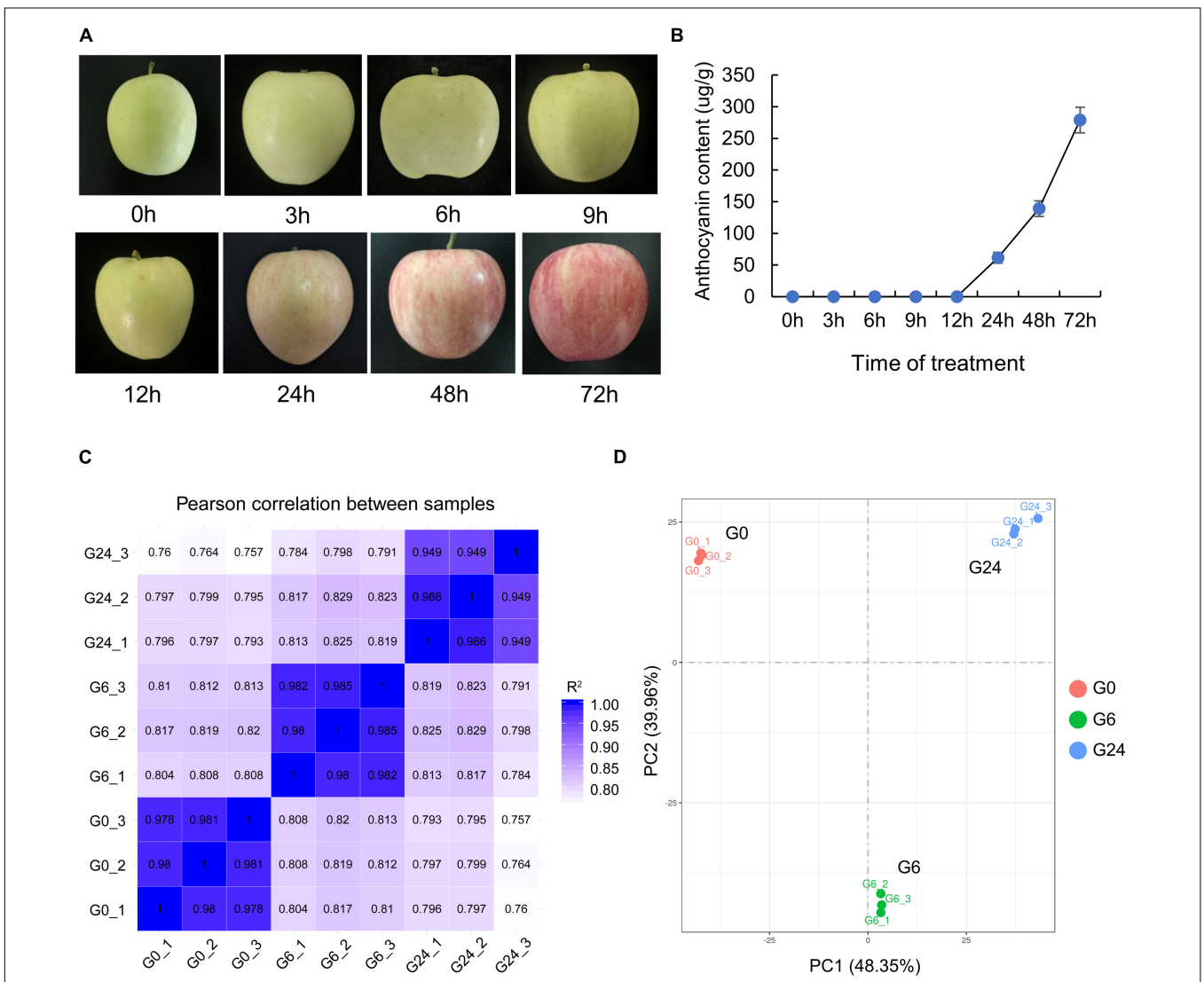


FIGURE 1 | RNA-seq data of "Starkrimson Delicious" apple fruit exposed to light after unbagging. **(A)** Change in color of "Starkrimson Delicious" fruit peel under light. **(B)** Anthocyanin content in the peel under light. **(C)** Pearson correlation between samples. **(D)** Principal component analysis of the RNA-seq data.

(Zhang et al., 2020). In order to maintain the same concentration of *Agrobacterium* solution, 50 ul of *Agrobacterium* solution MdBBX21 (OD = 0.8) and MdHY5 (OD = 0.8) were mixed with 50 ul of *Agrobacterium* solution MdMYB1pro:LUC for tobacco leaf injection, respectively. Also, equal volumes and concentrations of *Agrobacterium* solution MdBBX21 and MdHY5 were mixed, and then 50 ul of the mixture was mixed with 50 ul of *Agrobacterium* solution MdMYB1pro:LUC for tobacco leaf injection.

Bimolecular Fluorescence Complementation Assay

The *MdBBX21* CDS was inserted into the pSPYNE vector, whereas the *MdHY5* CDS was inserted into the pSPYCE vector. *A. tumefaciens* strain GV3101 cells were transformed with the MdBBX21-NE and MdHY5-CE recombinant plasmids. The primer sequences used are listed in **Supplementary Table 4**. Equal volumes of the *A. tumefaciens* cells carrying MdBBX21-NE and MdHY5-CE were mixed. Arabidopsis protoplast cells were infected with the *A. tumefaciens* cells for 15 min and then incubated at 23°C for 16–20 h. The YFP fluorescence was detected using the LSM 710 confocal laser-scanning microscope (Carl Zeiss) with an excitation wavelength of 514 nm.

RESULTS

Transcriptome Analysis of Apple Peel Exposed to Light

To clarify the response of apple to light, we first measured the anthocyanin content in the peel of unbagged apple fruit exposed to white light for 0, 3, 6, 9, 12, 24, 48, and 72 h (**Figure 1A**). Anthocyanins were basically undetectable before 12 h, but they started to accumulate significantly in the peel at 24 h (**Figure 1B**). The peel samples at the following three time-points underwent an RNA-seq analysis: 0 h (G0), 6 h (G6), and 24 h (G24). After the strict data filtering step, the number of clean reads in each library ranged from 37.69 to 45.95 million, and the Q30 values exceeded 92% (**Supplementary Table 1**). The mapping rate of the clean reads to the reference genome was 91.25–92.51%. Additionally, 88.81–90.27% of the clean reads were uniquely mapped (**Supplementary Table 2**). All biological replicates were strongly correlated ($R^2 > 0.94$) (**Figure 1C**); this correlation was confirmed by principal component analysis (**Figure 1D**).

Identification of Differentially Expressed Genes and Kyoto Encyclopedia of Genes and Genomes Analysis

To screen for early light-responsive genes, we compared the G0, G6, and G24 expression levels. In total, 11,941 DEGs were revealed in the G6 vs. G0, G24 vs. G6, and G24 vs. G0 pairwise comparisons. More DEGs were detected in the G6 vs. G0 comparison than in the G24 vs. G6 comparison (**Figures 2A,B**). These results suggest that many genes started to respond to light signals after 6 h. To further elucidate the gene expression patterns, the 11,941 DEGs were classified into eight gene expression profiles (**Supplementary Figure 1**). Specifically, 4,082 DEGs were classified into three significant profiles ($P < 0.05$), including two up-regulated profiles (profiles 6 and 7) and one down-regulated profile (profile 0) (**Figure 2C**). A KEGG enrichment analysis indicated genes related to flavonoid biosynthesis and phenylpropanoid biosynthesis were significantly enriched among the DEGs in profiles 0, 6, and 7 (**Figure 2D**).

Characterization and Expression Analysis of MdBBX21

Among the DEGs in up-regulated profile 6 (**Supplementary Table 3**), *MD08G1021000* was revealed to be highly homologous to Arabidopsis *AtBBX21*. Thus, we named this gene *MdBBX21*. To clarify the relationship between MdBBX21 and MdBBX20, MdBBX21, and MdBBX22-1/2, we compared their amino acid sequences. The sequence identities between MdBBX21 and MdBBX20, MdBBX22-1, and MdBBX22-2 were 37.58, 31.92, and 31.63%, respectively. In the constructed phylogenetic tree, MdBBX21, MdBBX20, and MdBBX22-1/2 were clustered with Arabidopsis subfamily IV members, but they were distributed in different clades (**Figure 3A**).

The correct cellular localization of a protein is critical for ensuring it functions properly. In tobacco leaf cells transiently transformed with 35S:MdBBX21-eGFP, green fluorescence was detected only in the nucleus, whereas in the control tobacco cells transiently transformed with 35S:eGFP, green fluorescence was observed throughout the cell, including in the cytoplasm and nucleus. Accordingly, MdBBX21 appears to be a nuclear protein (**Figure 3B**).

To determine whether MdBBX21 affects anthocyanin synthesis, we analyzed *MdBBX21* expression in the apple fruit peel irradiated with white light for 0, 3, 6, 9, 12, 24, 48, and 72 h. The *MdBBX21* gene was responsive to light, and its expression level started to increase at 3 h, peaking at 9 h. The expression of the regulatory genes (*MdMYB1* and *MdbHLH3*) and structural genes (*MdCHS*, *MdF3H*, *MdDFR*, *MdANS*, and *MdUFGT*) related to anthocyanin synthesis peaked after 24 h. The *MdWD40* expression level was essentially unchanged (**Figure 4**). These results imply that MdBBX21 may function upstream of MdMYB1.

Heterologous Overexpression of MdBBX21 Promotes Anthocyanin Accumulation in Arabidopsis

To confirm MdBBX21 contributes to anthocyanin biosynthesis, *MdBBX21* was overexpressed in the *bbx21* Arabidopsis mutant. The 35S:MdBBX21/*bbx21* line accumulated more anthocyanins in the cotyledons and hypocotyls than the untransformed *bbx21* Arabidopsis mutant (**Figures 5A–C**). The expression levels of anthocyanin-related genes (*AtPAP1*, *AtCHS*, *AtF3'H*, *AtDFR*, *AtLDOX*, and *AtUFGT*) also increased in the 35S:MdBBX21/*bbx21* line, which was consistent with the changes in the anthocyanin content (**Figure 5D**). We also examined the expression of *AtHY5* and Arabidopsis *BBX* subfamily IV members. The *AtHY5* and *AtBBX22* expression

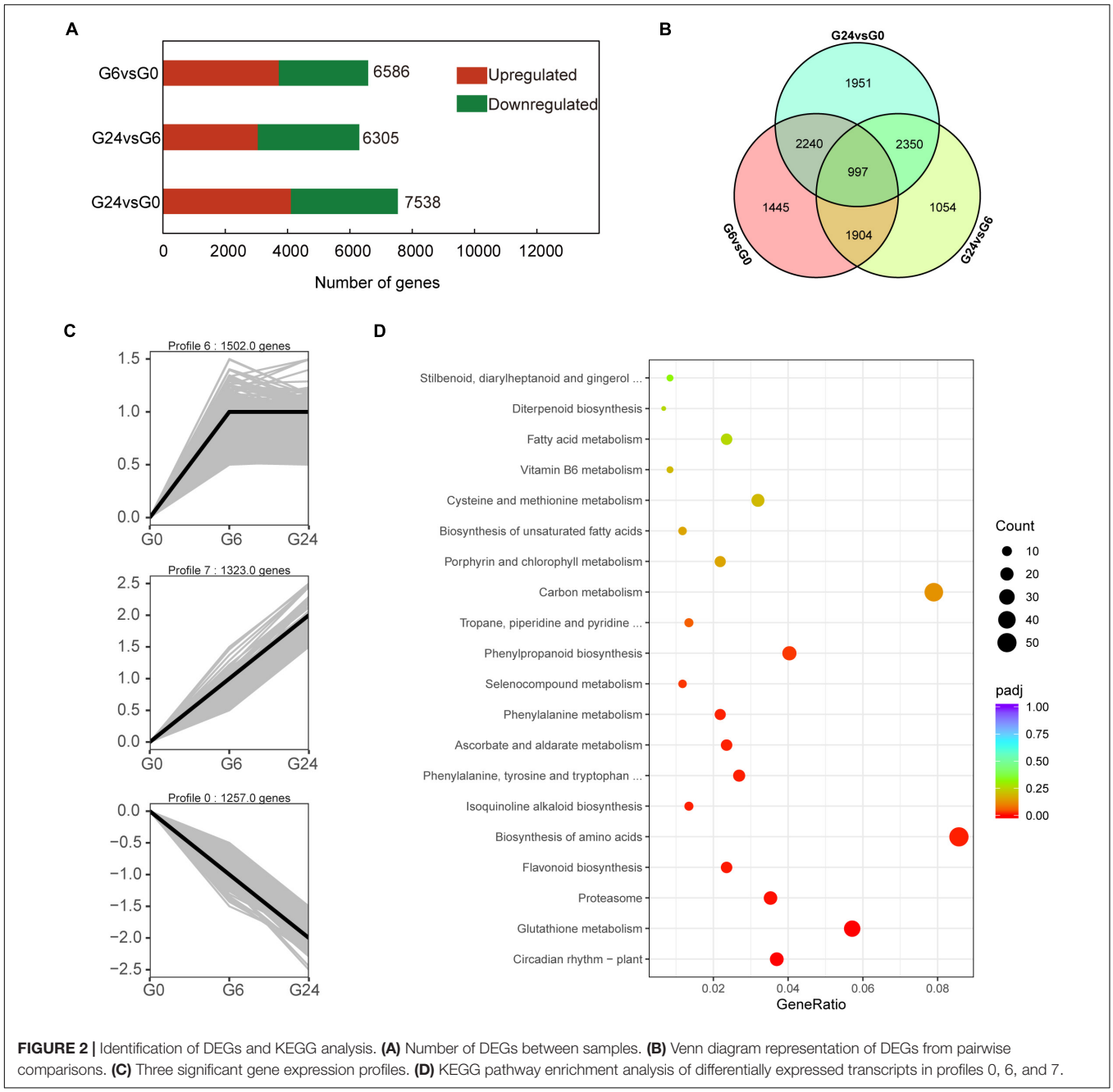


FIGURE 2 | Identification of DEGs and KEGG analysis. **(A)** Number of DEGs between samples. **(B)** Venn diagram representation of DEGs from pairwise comparisons. **(C)** Three significant gene expression profiles. **(D)** KEGG pathway enrichment analysis of differentially expressed transcripts in profiles 0, 6, and 7.

levels were higher in the 35S:MdBBX21/*bbx21* seedlings than in the *bbx21* Arabidopsis mutant seedlings. In contrast, there were no differences in the expression levels of the other subfamily IV *BBX* genes (**Figure 5E** and **Supplementary Figure 2**).

Overexpression of MdBBX21 Promotes Anthocyanin Accumulation in Apple Calli

To further confirm the MdBBX21 function related to anthocyanin biosynthesis in apple, we overexpressed *MdBBX21* in apple calli (**Figures 6A,B**). Wild-type and MdBBX21-OX transgenic apple calli were cultured on medium in darkness and then transferred to a light incubator for a 3-day exposure to constant light. The MdBBX21-OX apple calli accumulated more anthocyanins than the WT calli (**Figure 6C**). As expected, the expression levels of the anthocyanin biosynthesis-related genes (*MdMYB1*, *MdCHS*, *MdF3H*, *MdDFR*, *MdANS*, and *MdUFGT*) were significantly higher in the MdBBX21-OX calli than in the WT calli (**Figure 6D**). These results suggest that MdBBX21 might positively regulate anthocyanin accumulation by promoting the transcription of anthocyanin biosynthesis-related genes. The transcription levels of the previously characterized genes *MdBBX20*, *MdBBX22-1/2*, *MdCOL4* (*MdBBX24*), and *MdHY5* were also analyzed by qPCR

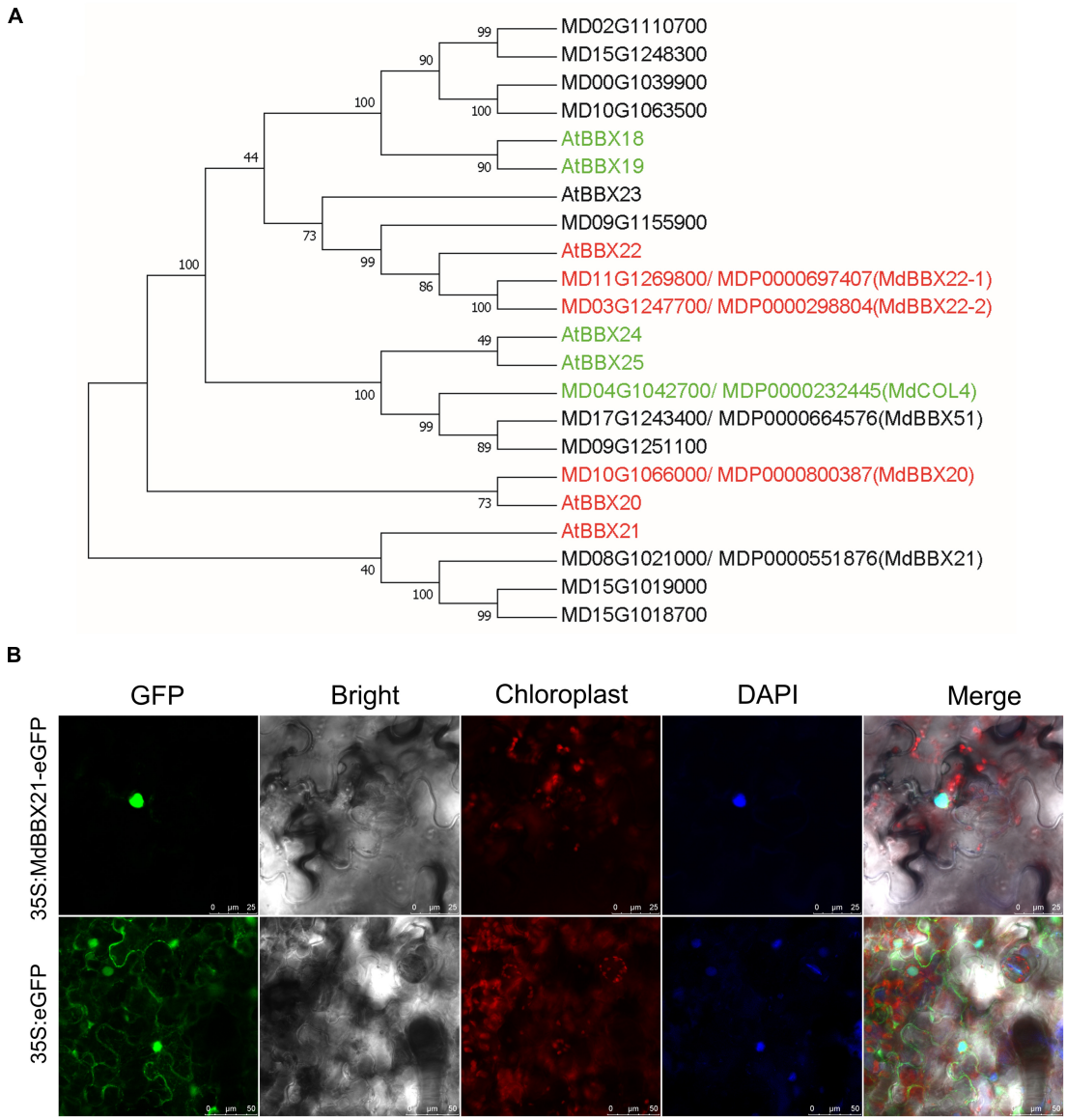


FIGURE 3 | Characterization of *MdBBX21*. **(A)** Phylogenetic relationship of apple and Arabidopsis subfamily IV members. The red and green genes represent positive and negative regulators related to light signaling, respectively. The phylogenetic tree was constructed using the neighbor-joining method by MEGA 7.0. The bootstrap values of 1000 replicates were calculated at each node. **(B)** Subcellular localization of MdBBX21 expressed in tobacco leaf cells.

(**Figure 6E** and **Supplementary Figure 3**). The data indicated that *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* were significantly more highly expressed in MdBBX21-OX calli than in WT calli. The *MdBBX24* expression level did not differ between the MdBBX21-OX and WT calli.

MdBBX21 Binds Directly to the MdHY5, MdBBX20, and MdBBX22-1/2 Promoters and Induces Expression

The increased *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* expression levels in MdBBX21-OX calli led us to speculate that MdBBX21 may directly promote the expression of these four genes. Thus, we first searched for the G-box motif, which is a BBX-binding site (Datta et al., 2008; Xu et al., 2016), in the *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* promoters. The G-box motif was detected in all three promoters (**Figure 7A**). In yeast one-hybrid assays, the Y187 yeast strains containing MdBBX21-AD + MdBBX20pro-HIS2, MdBBX21-AD + MdBBX22-1pro-HIS2, MdBBX21-AD + MdBBX22-2pro-HIS2, and MdBBX21-AD + MdHY5pro-HIS2 were able to grow on the −H/−T/−L selection medium containing 100 mM 3-AT. In contrast, the Y187 yeast strains containing AD + MdBBX20pro-HIS2, AD + MdBBX22-1pro-HIS2, AD + MdBBX22-2pro-HIS2, and AD + MdHY5pro-HIS2 did not grow under the same conditions (**Figure 7B**). This observation suggests that MdBBX21 can

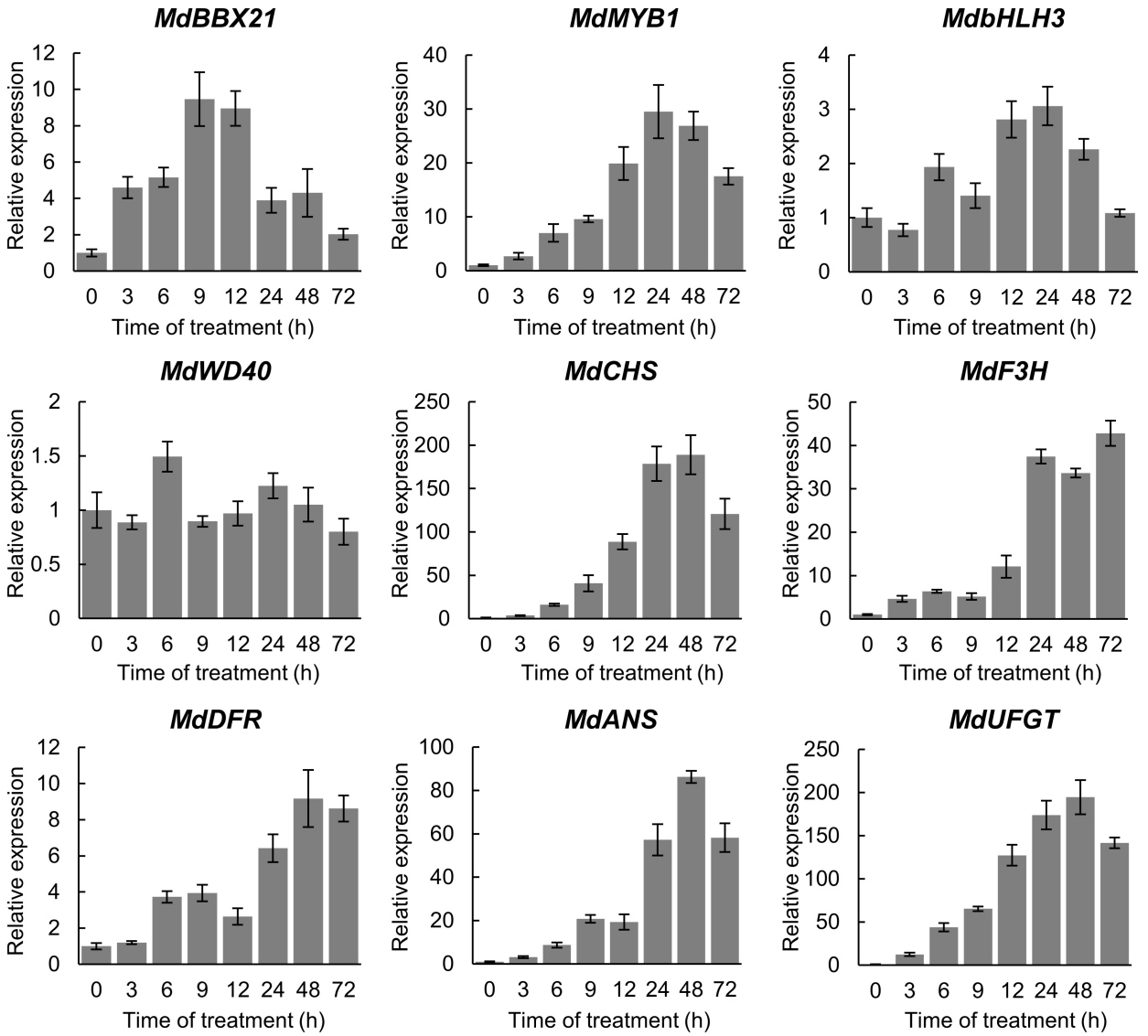


FIGURE 4 | Relative expression levels of anthocyanin biosynthesis-related genes in the apple fruit peel irradiated with white light. The *MdActin* gene was used as the internal control.

bind to the *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* promoters. The dual-luciferase assay results confirmed that MdBBX21 can promote the transcription of *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* (**Figure 7C**). We also analyzed the *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* expression patterns in the peel exposed to light. Compared with the *MdBBX21* expression pattern (**Figure 4**), the *MdBBX20*, and *MdBBX22-1/2* expression levels increased significantly at 6 h, and their peak expression levels occurred later than that of *MdBBX21* (**Supplementary Figure 4**). These findings imply that MdBBX21 functions upstream of MdBBX20, and MdBBX22-1/2 and induces the expression of *MdBBX20*, and *MdBBX22-1/2*. At the same time, we found that *MdHY5* was activated at 3 h. Considering HY5 being a transcription factor, we speculated that MdHY5, a master regulator of light signaling, might also control the expression of *MdBBX21*. Thus, we searched for the G-box motif in the *MdBBX21* promoter. There were two G-box motifs present in the *MdBBX21* promoter (**Figure 7A**). Yeast one-hybrid assay results showed that MdHY5 can bind to the *MdBBX21* promoter fragment containing these two motifs (**Figure 7B**). The dual-luciferase assay results confirmed that MdHY5 can promote the transcription of *MdBBX21* (**Figure 7C**).

The Interaction Between MdBBX21 and MdHY5 Can Significantly Enhance MdMYB1 Promoter Activity

Because MdBBX20 (Fang et al., 2019b), MdBBX22-2 (An et al., 2019), PpBBX16 (Bai et al., 2019a), and PpBBX18 (Bai et al., 2019b) can interact with HY5, we speculated that MdBBX21

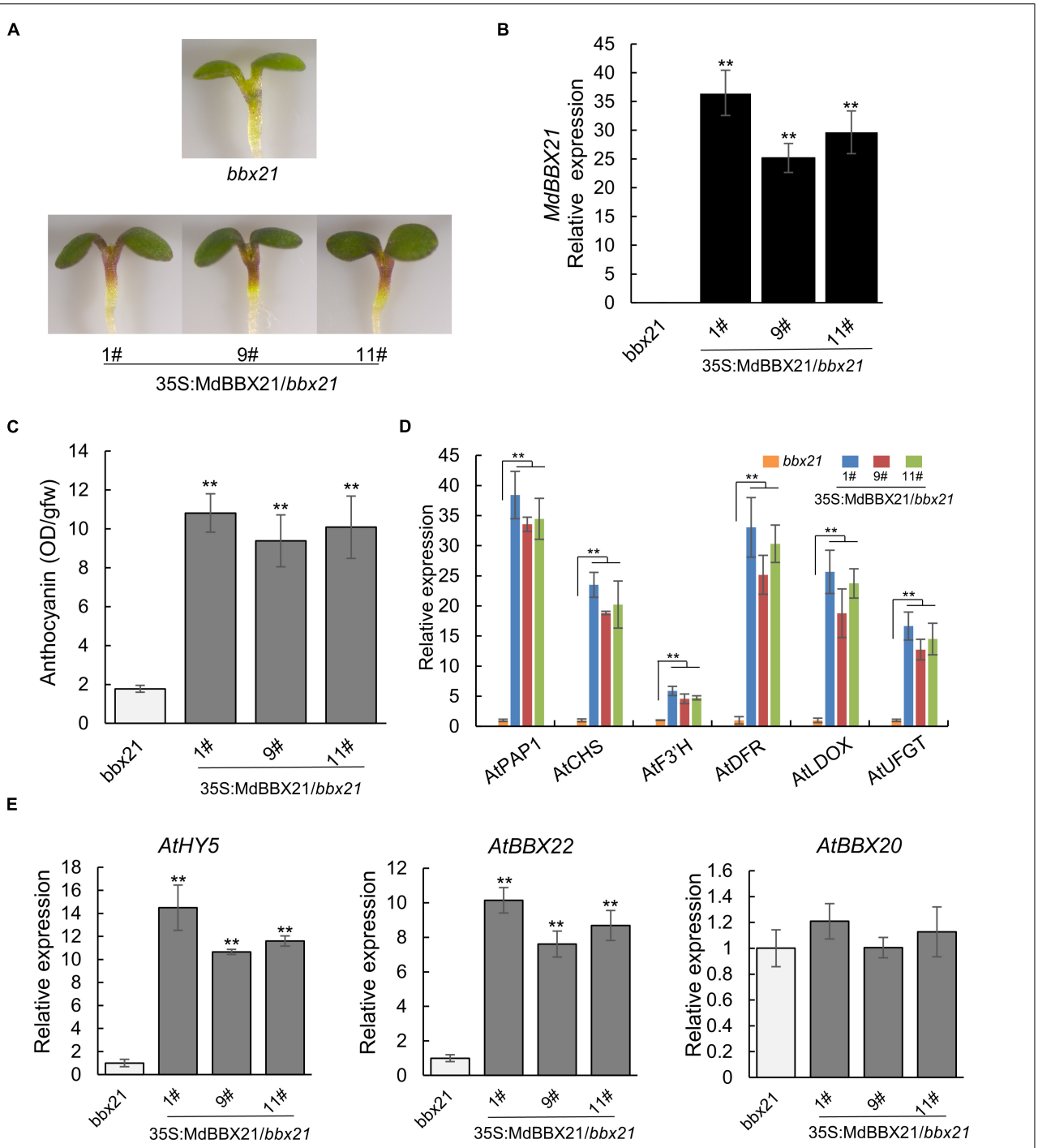


FIGURE 5 | Overexpression of *MdBBX21* promotes anthocyanin accumulation in Arabidopsis seedlings. **(A)** Phenotypes of 35S:MdBBX21/*bbx21* and *bbx21* mutant seedlings. 35S:MdBBX21/*bbx21* (#1), 35S:MdBBX21/*bbx21* (#9), and 35S:MdBBX21/*bbx 21* (#11) represent three transgenic lines. **(B)** The *MdBBX21*'s expression level in 35S:MdBBX21/*bbx21* and *bbx21* mutant seedlings. **(C)** Anthocyanin content in 35S:MdBBX21/*bbx21* and *bbx21* mutant seedlings. **(D)** Relative expression levels of anthocyanin biosynthesis-related genes in 35S:MdBBX21/*bbx21* and *bbx21* mutant seedlings. **(E)** Relative expression levels of *AtHY5*, *AtBBX22*, and *AtBBX20* in 35S:MdBBX21/*bbx21* and *bbx21* mutant seedlings. Error bars represent the standard deviation of three biological replicates. $^{**}P < 0.01$ (Student's *t*-test).

may also interact with MdHY5. Bimolecular fluorescence complementation assays were conducted to examine whether MdBBX21 can interact directly with MdHY5. Full-length *MdBBX21* and *MdHY5* CDSs were cloned into pSPYNE and pSPYCE vectors, respectively, to generate MdBBX21-NE and MdHY5-CE. Yellow fluorescence was detected in the nucleus of

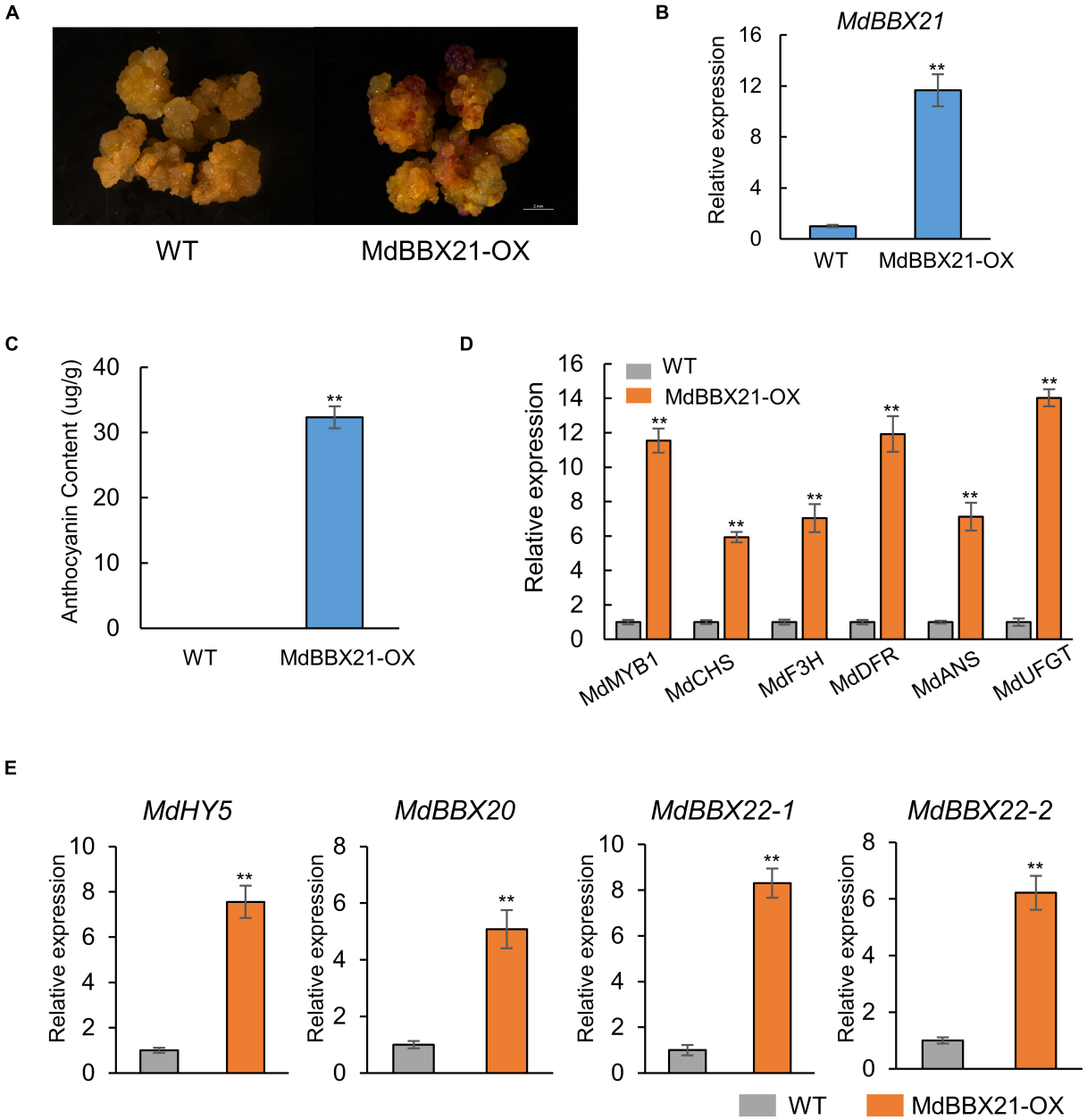


FIGURE 6 | Overexpression of *MdBBX21* promotes anthocyanin accumulation in apple calli. **(A)** Anthocyanin accumulation phenotypes of apple calli (WT: wild-type apple calli; MdBBX21-OX: *MdBBX21* overexpression apple calli). **(B)** *MdBBX21*'s expression level in WT and MdBBX21-OX. **(C)** Anthocyanin contents in WT and MdBBX21-OX. **(D)** The expression levels of anthocyanin biosynthesis-related genes in WT and MdBBX21-OX. **(E)** The expression levels of *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* in WT and MdBBX21-OX. Error bars represent the standard deviation of three biological replicates. $^{**}P < 0.01$ (Student's *t*-test).

Arabidopsis protoplasts co-transformed with MdBBX21-NE and MdHY5-CE. However, yellow fluorescence was not observed in the BBX21-NE + CE and NE + HY5-CE controls (**Figure 8A**). These results indicate that MdBBX21 can interact with MdHY5. This interaction was verified in yeast two-hybrid assays (**Figure 8B**).

MdBBX20 can bind directly to the *MdMYB1* promoter and induce expression (Fang et al., 2019b). In this study, Y187 yeast strains containing MdBBX21-AD + MdMYB1pro-HIS2 grew on −H/−T/−L selection medium containing 100 mM 3-AT, but the Y187 yeast strains containing AD + MdMYB1pro-HIS2 did not (**Figure 8C**). Accordingly, MdBBX21 appears to be able to interact with the *MdMYB1* promoter. On the basis of the dual-luciferase assay results, MdBBX21 can promote *MdMYB1* expression (**Figure 8D**). Moreover, MdHY5 can also promote *MdMYB1* expression, which is consistent with the results of an earlier study (An et al., 2017). We also determined that the MdBBX21–MdHY5 heterodimer enhances

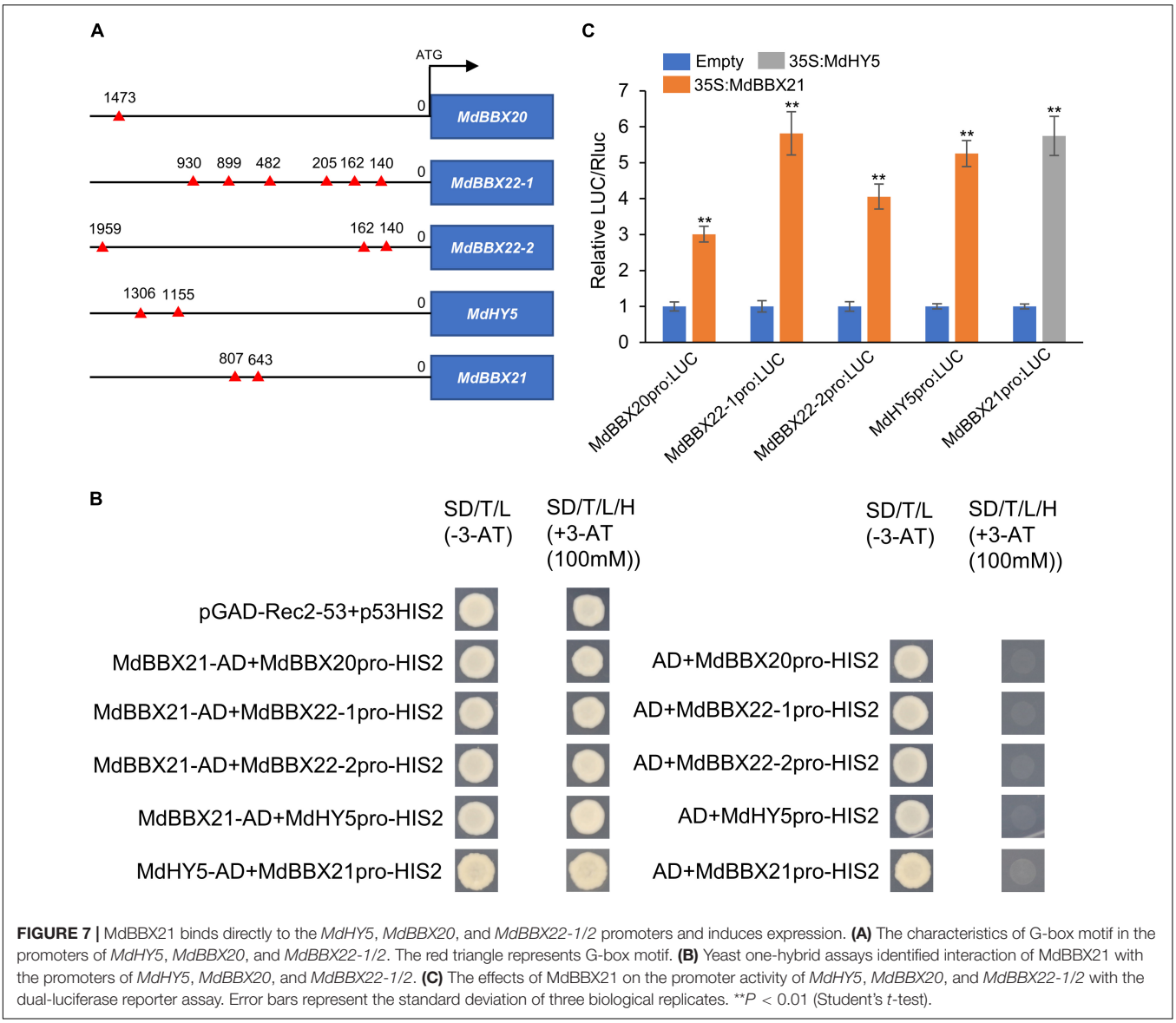


FIGURE 7 | MdBBX21 binds directly to the *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* promoters and induces expression. **(A)** The characteristics of G-box motif in the promoters of *MdHY5*, *MdBBX20*, and *MdBBX22-1/2*. The red triangle represents G-box motif. **(B)** Yeast one-hybrid assays identified interaction of MdBBX21 with the promoters of *MdHY5*, *MdBBX20*, and *MdBBX22-1/2*. **(C)** The effects of MdBBX21 on the promoter activity of *MdHY5*, *MdBBX20*, and *MdBBX22-1/2* with the dual-luciferase reporter assay. Error bars represent the standard deviation of three biological replicates. $^{**}P < 0.01$ (Student's *t*-test).

the *MdMYB1* promoter activity more than MdBBX21 or MdHY5 alone (**Figure 8D**).

DISCUSSION

MdBBX21 Responds to Light and Induces Anthocyanin Biosynthesis in Apple

Light is required for anthocyanin biosynthesis in the apple fruit peel (Saure, 1990). An exposure to light up-regulates the expression of anthocyanin biosynthesis-related structural and regulatory genes in the apple peel (Takos M. A. et al., 2006; Feng et al., 2013). However, how light signals affect the expression of these genes was unclear. There is increasing evidence that BBX proteins affect plant photomorphogenesis (Gangappa and Botto, 2014). The RNA-seq analysis in this study revealed that *MdBBX21* expression in the peel of dark-grown "Starkrimson Delicious" apples is affected by the subsequent exposure to light. Additionally, MdBBX21 and Arabidopsis AtBBX21 sequences are highly similar. Earlier research demonstrated that AtBBX21 is a positive regulator of anthocyanin biosynthesis (Datta et al., 2007). Hence, we predicted that MdBBX21 and AtBBX21 might have similar functions. We analyzed *MdBBX21* expression and observed that it increased in response to light. Moreover, *MdBBX21* expression peaked before the expression of anthocyanin-related structural and regulatory genes peaked (**Figure 4**). The overexpression of *MdBBX21* in the *bbx21* Arabidopsis mutant resulted in a significant increase in the anthocyanin contents of hypocotyls (**Figure 5**). This is consistent with the fact pear PpBBX18 (homologous to AtBBX21) regulates anthocyanin accumulation in transgenic Arabidopsis plants. In the present study, the

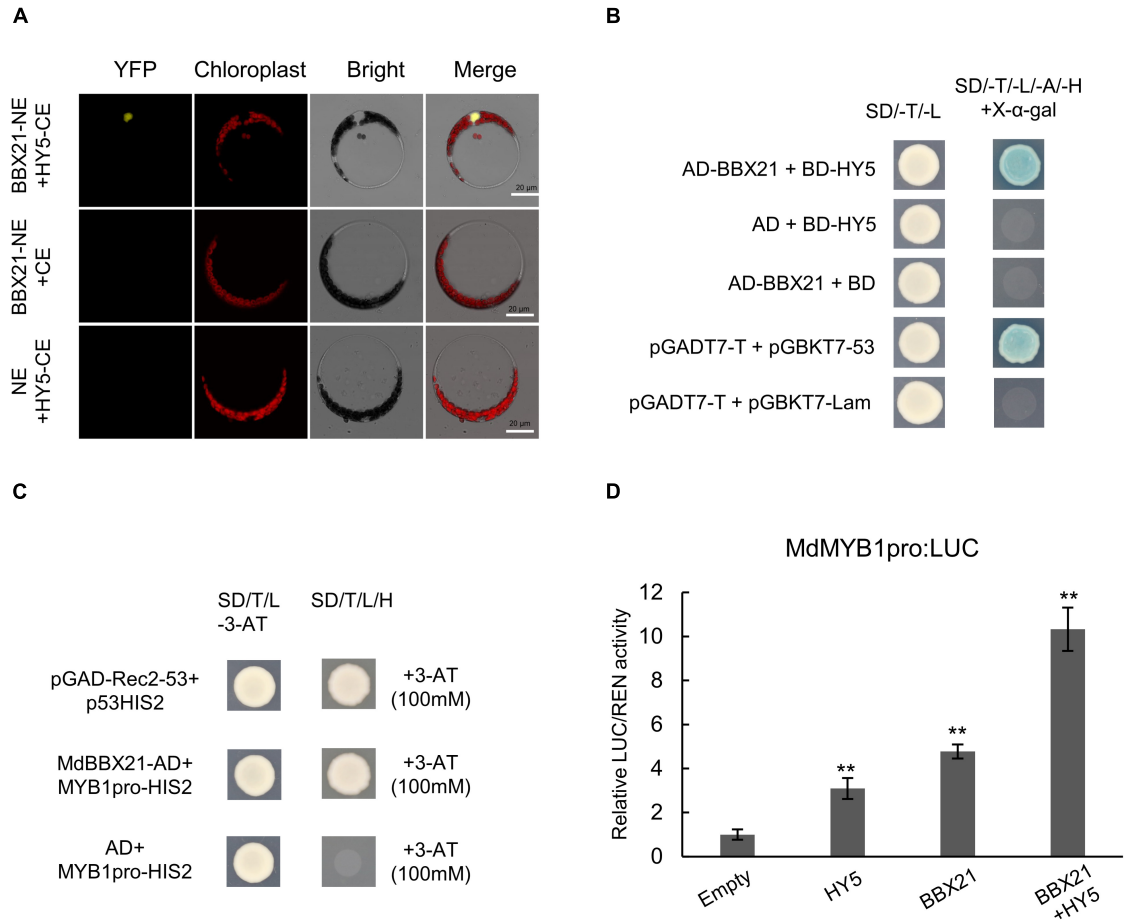


FIGURE 8 | The interaction between MdBBX21 and MdHY5 can significantly enhance *MdMYB1* promoter activity. **(A)** Bimolecular fluorescence complementation assay. **(B)** Yeast two-hybrid assay showing interaction between MdBBX21 and MdHY5. **(C)** Yeast one-hybrid assay showing interaction between MdBBX21 and *MdMYB1* promoter. **(D)** Effects of MdBBX21 and MdHY5 individually and in combination on promoter activity of *MdMYB1* as determined by luciferase reporter assay. Error bars represent the standard deviation of three biological replicates. $^{**}P < 0.01$ (Student's *t*-test).

MdBBX21-OX apple calli accumulated more anthocyanins than the WT apple calli under light (**Figure 6**). These results indicate that MdBBX21 is responsive to light and induces anthocyanin biosynthesis.

MdBBX21 Up-Regulates MdBBX20, MdBBX22-1/2, and MdHY5 Expression

The Arabidopsis BBX subfamily IV members AtBBX20, AtBBX21, AtBBX22, and AtBBX23 positively regulate plant photomorphogenic processes in response to diverse light signals (Datta et al., 2007; Chang et al., 2008; Datta et al., 2008; Fan et al., 2012; Xu et al., 2016; Zhang et al., 2017). In pear, PpBBX16 and PpBBX18, which are subfamily IV BBX proteins, positively regulate anthocyanin synthesis by interacting with PpHY5 (Bai et al., 2019a,b). In apple, MdBBX20 and MdBBX22-1/2 reportedly promote anthocyanin accumulation in response to UV-B irradiation (Bai et al., 2014; An et al., 2019; Fang et al., 2019b). In addition, Plunkett et al. (2019) found that BBX subfamily IV member *MdBBX51* and BBX subfamily I member *MdBBX1* have very high expression levels during fruit development. Further, they can activate the promoter of *MdMYB1* in the presence of some co-factors MYB and bHLH. However, the relationships among these BBX proteins are unknown. In this study, *MdBBX21* expression increased significantly after a 3-h light treatment, and subsequently peaked at 9 h. In contrast, the *MdBBX20* and *MdBBX22-1/2* expression levels peaked at 12 and 24 h, respectively. The expression of structural and regulatory genes related to anthocyanin synthesis increased rapidly and peaked after 24 h (**Figure 4** and **Supplementary Figure 2**). These findings indicate that MdBBX21 responds to light signals relatively early and likely functions upstream of MdBBX20 and MdBBX22-1/2. Dual-luciferase and yeast one-hybrid assays proved that MdBBX21 induces *MdBBX20* and *MdBBX22-1/2* expression after binding to their promoters. The *AtBBX22*, *MdBBX20*, and *MdBBX22-1/2* expression levels increased in Arabidopsis and apple calli overexpressing *MdBBX21* (**Figures 5–7**). Therefore, MdBBX21 can up-regulate the expression of *AtBBX22*, *MdBBX20*, and *MdBBX22-1/2*. However, MdBBX21 did not activate the

expression of *MdBBX1*, *MdBBX24*, and *MdBBX51* in apple calli (**Supplementary Figure 3**). In Arabidopsis, AtBBX21 can bind directly to the *AtHY5* promoter and induce expression (Xu et al., 2016). In the current study, biochemical assays proved that MdBBX21 can bind directly to the *MdHY5* promoter and induce expression (**Figure 7**). The overexpression of *MdBBX21* in Arabidopsis and apple calli resulted in up-regulated *HY5* expression (**Figures 5**, **6**). These results imply that MdBBX21 and AtBBX21 have similar functions and can directly up-regulate the expression of *HY5*. At the same time, we found that MdHY5 can also activate the expression of *MdBBX21*, indicating that a positive feedback regulation mechanism exists between MdBBX21 and MdHY5.

The MdBBX21–MdHY5 Interaction Can Significantly Enhance MdMYB1 Promoter Activity Under Light

The bZIP transcription factor AtHY5 is a positive regulator of plant photomorphogenesis that functions downstream of the photoreceptors and COP1 (Lau and Deng, 2010). Additionally, AtHY5 regulates various physiological processes, including anthocyanin synthesis, lateral root formation, and hypocotyl elongation (Oyama et al., 1997). Earlier studies regarding *in vitro* DNA–protein interactions demonstrated that AtHY5 can bind directly and specifically to the G-box motif in the promoters of the anthocyanin-related structural genes *AtF3H*, *AtCHS*, and *AtCHI* and the regulatory gene *AtPAP1* (Ang et al., 1998; Lee et al., 2007; Shin et al., 2013). In Arabidopsis, AtBBX21 and AtBBX22 can activate transcription. They interact with AtHY5 *in vivo* and induce downstream gene expression in an AtHY5-dependent and -independent manner to promote plant photomorphogenesis (Datta et al., 2007, 2008). In apples and pears, HY5 can bind directly to the *MYB1* promoter (An et al., 2017; Tao et al., 2018). Both PpBBX16 and PpBBX18 cannot bind directly to the *MYB1* promoter, but they can interact with HY5 and promote *MYB1* expression (i.e., HY5-dependent manner) (Bai et al., 2019a,b). However, in apple, MdBBX20 can bind directly to the *MdMYB1* promoter and induce expression (i.e., HY5-independent manner) (Fang et al., 2019b). In this study, we proved that MdBBX21 can bind directly to the *MdMYB1* promoter and induce expression (**Figure 8**). Additionally, the analysis of protein–protein interactions revealed that MdBBX21 interacts with MdHY5 in yeast and plant cells. Furthermore, the interaction between MdBBX21 and MdHY5 can significantly enhance *MdMYB1* promoter activity (**Figure 8**).

In conclusion, we systematically characterized the effect of light on MdBBX21 in the apple peel. We proved that MdBBX21 binds to the promoters of *MdBBX20*, *MdBBX22-1/2*, and *MdHY5*. The subsequent up-regulated expression of these genes enhances anthocyanin accumulation. Additionally, MdBBX21 can interact with MdHY5 and induce *MdMYB1* expression. The results of our study will form the basis of future investigations on the functions of BBX proteins in apple.

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 SRA BioProject, accession no: PRJNA767632.

AUTHOR CONTRIBUTIONS

BZ, H-JY, and Z-YZ conceived the original screening and research plans. BZ, Z-ZZ, B-CW, N-NH, and Y-ZY supervised the experiments. BZ, H-JY, and DQ analyzed the data. BZ and Z-ZZ wrote the manuscript. H-JY revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Earmarked Fund for the National Natural Science Foundation of China (32072555), the Modern Agro-industry Technology Research System of China (CARS-27), Science and Technology Department Youth Foundation in Shaanxi province, China (2019JQ-204), and Chinese Universities Scientific Fund (2452020053).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.774446/full#supplementary-material

REFERENCES

An, J. P., Qu, F. J., Yao, J. F., Wang, X. N., You, C. X., Wang, X. F., et al. (2017). The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. *Hortic Res.* 4:17023. doi: 10.1038/hortres.2017.23

An, J. P., Wang, X. F., Zhang, X. W., Bi, S. Q., You, C. X., and Hao, Y. J. (2019). MdBBX22 regulates UV-B-induced anthocyanin biosynthesis through regulating the function of MdHY5 and is targeted by MdBT2 for 26S proteasome-mediated degradation. *Plant Biotechnol. J.* 17, 2231–2233. doi: 10.1111/pbi.13196

An, J. P., Zhang, X. W., Bi, S. Q., You, C. X., Wang, X. F., and Hao, Y. J. (2020). The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. *Plant J.* 101, 573–589. doi: 10.1111/tpj.14555

Ang, L. H., Chattopadhyay, S., Wei, N., Oyama, T., Okada, K., Batschauer, A., et al. (1998). Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of Arabidopsis development. *Mol. Cell* 1, 213–222. doi: 10.1016/s1097-2765(00)80022-2

Ang, L. H., and Deng, X. W. (1994). Regulatory hierarchy of photomorphogenic loci: allele-specific and light-dependent interaction between the HY5 and COP1 loci. *Plant Cell* 6, 613–628. doi: 10.1105/tpc.6.5.613

Bai, S., Saito, T., Honda, C., Hatsuyama, Y., Ito, A., and Moriguchi, T. (2014). An apple B-box protein, MdCOL11, is involved in UV-B- and temperature-induced anthocyanin biosynthesis. *Planta* 240, 1051–1062. doi: 10.1007/s00425-014-2129-8

Bai, S., Tao, R., Tang, Y., Yin, L., Ma, Y., Ni, J., et al. (2019a). BBX16, a B-box protein, positively regulates light-induced anthocyanin accumulation by activating MYB10 in red pear. *Plant Biotechnol. J.* 17, 1985–1997. doi: 10.1111/pbi.13114

Bai, S., Tao, R., Yin, L., Ni, J., Yang, Q., Yan, X., et al. (2019b). Two B-box proteins, PpBBX18 and PpBBX21, antagonistically regulate anthocyanin biosynthesis via competitive association with Pyrus pyrifolia ELONGATED HYPOCOTYL 5 in the peel of pear fruit. *Plant J.* 100, 1208–1223. doi: 10.1111/tpj.14510

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol.* 48, 958–970. doi: 10.1093/pcp/pcm066

Chang, C. S., Li, Y. H., Chen, L. T., Chen, W. C., Hsieh, W. P., Shin, J., et al. (2008). LZF1, a HY5-regulated transcriptional factor, functions in Arabidopsis de-etiolation. *Plant J.* 54, 205–219. doi: 10.1111/j.1365-313X.2008.03401.x

Chu, Z., Wang, X., Li, Y., Yu, H., Li, J., Lu, Y., et al. (2016). Genomic organization, phylogenetic and expression analysis of the B-BOX gene family in tomato. *Front. Plant Sci.* 7:1552. doi: 10.3389/fpls.2016.01552

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Cominelli, E., Gusmaroli, G., Allegra, D., Galbiati, M., Wade, H. K., Jenkins, G. I., et al. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in Arabidopsis thaliana. *J. Plant Physiol.* 165, 886–894. doi: 10.1016/j.jplph.2007.06.010

Daccord, N., Celton, J. M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., et al. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* 49, 1099–1106. doi: 10.1038/ng.3886

Datta, S., Hettiarachchi, C., Johansson, H., and Holm, M. (2007). SALT TOLERANCE HOMOLOG2, a B-box protein in Arabidopsis that activates transcription and positively regulates light-mediated development. *Plant Cell* 19, 3242–3255. doi: 10.1105/tpc.107.054791

Datta, S., Johansson, H., Hettiarachchi, C., Irigoyen, M. L., Desai, M., Rubio, V., et al. (2008). LZF1/SALT TOLERANCE HOMOLOG3, an Arabidopsis B-box protein involved in light-dependent development and gene expression, undergoes COP1-mediated ubiquitination. *Plant Cell* 20, 2324–2338. doi: 10.1105/tpc.108.061747

de Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev.* 11, 1422–1434. doi: 10.1101/gad.11.11.1422

Deng, X. W., Matsui, M., Wei, N., Wagner, D., Chu, A. M., Feldmann, K. A., et al. (1992). COP1, an Arabidopsis regulatory gene, encodes a protein with both a zinc-binding motif and a G beta homologous domain. *Cell* 71, 791–801. doi: 10.1016/0092-8674(92)90555-q

Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., et al. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* 21, 168–183. doi: 10.1105/tpc.108.059329

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49, 414–427. doi: 10.1111/j.1365-313X.2006.02964.x

Fan, X. Y., Sun, Y., Cao, D. M., Bai, M. Y., Luo, X. M., Yang, H. J., et al. (2012). BZS1, a B-box protein, promotes photomorphogenesis downstream of both brassinosteroid and light signaling pathways. *Mol. Plant* 5, 591–600. doi: 10.1093/mp/sss041

Fang, H., Dong, Y., Yue, X., Hu, J., Jiang, S., Xu, H., et al. (2019b). The B-box zinc finger protein MdBBX20 integrates anthocyanin accumulation in response to ultraviolet radiation and low temperature. *Plant Cell Environ.* 42, 2090–2104. doi: 10.1111/pce.13552

Fang, H., Dong, Y., Yue, X., Chen, X., He, N., Hu, J., et al. (2019a). MdCOL4 interaction mediates crosstalk between UV-B and high temperature to control fruit coloration in apple. *Plant Cell Physiol.* 60, 1055–1066. doi: 10.1093/pcp/pcz023

Feng, F., Li, M., Ma, F., and Cheng, L. (2013). Phenylpropanoid metabolites and expression of key genes involved in anthocyanin biosynthesis in the shaded peel of apple fruit in response to sun exposure. *Plant Physiol. Biochem.* 69, 54–61. doi: 10.1016/j.plaphy.2013.04.020

Gangappa, S. N., and Botto, J. F. (2014). The BBX family of plant transcription factors. *Trends Plant Sci.* 19, 460–470. doi: 10.1016/j.tplants.2014.01.010

Huang, J., Zhao, X., Weng, X., Wang, L., and Xie, W. (2012). The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. *PLoS One* 7:e48242. doi: 10.1371/journal.pone.0048242

Kasahara, M., Swartz, T. E., Olney, M. A., Onodera, A., Mochizuki, N., Fukuzawa, H., et al. (2002). Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, rice, and Chlamydomonas reinhardtii. *Plant Physiol.* 129, 762–773. doi: 10.1104/pp.002410

Khanna, R., Kronmiller, B., Maszle, D. R., Coupland, G., Holm, M., Mizuno, T., et al. (2009). The Arabidopsis B-box zinc finger family. *Plant Cell* 21, 3416–3420. doi: 10.1105/tpc.109.069088

Kim, S. H., Lee, J. R., Hong, S. T., Yoo, Y. K., An, G., and Kim, S. R. (2003). Molecular cloning and analysis of anthocyanin biosynthesis genes preferentially expressed in apple skin. *Plant Sci.* 165, 403–413. doi: 10.1016/s0168-9452(03)00201-2

Kubasek, W. L., Shirley, B. W., McKillop, A., Goodman, H. M., Briggs, W., and Ausubel, F. M. (1992). Regulation of flavonoid biosynthetic genes in germinating arabidopsis seedlings. *Plant Cell* 4, 1229–1236. doi: 10.1105/tpc.4.10.1229

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

Lawrence, W. J. C., Price, J. R., Robinson, G. M., and Robinson, R. (1939). The distribution of anthocyanins in flowers, fruits and leaves. *Philosop. Trans. R. Soc. Lond. Series B Biol. Sci.* 230, 149–178. doi: 10.1098/rstb.1939.0006

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., et al. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19, 731–749. doi: 10.1105/tpc.106.047688

Li, S. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal Behav.* 9:e27522. doi: 10.4161/psb.27522

Li, Y. Y., Mao, K., Zhao, C., Zhao, X. Y., Zhang, H. L., Shu, H. R., et al. (2012). MdCOP1 ubiquitin E3 ligases interact with MdMYB1 to regulate light-induced anthocyanin biosynthesis and red fruit coloration in apple. *Plant Physiol.* 160, 1011–1022. doi: 10.1104/pp.112.199703

Li, Y. Y., Mao, K., Zhao, C., Zhao, X. Y., Zhang, R. F., Zhang, H. L., et al. (2013). Molecular cloning and functional analysis of a blue light receptor gene MdCRY2 from apple (Malus domestica). *Plant Cell Rep.* 32, 555–566. doi: 10.1007/s00299-013-1387-4

Lin, C., Robertson, D. E., Ahmad, M., Raibekas, A. A., Jorns, M. S., Dutton, P. L., et al. (1995). Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1. *Science* 269, 968–970. doi: 10.1126/science.7638620

Liu, X., Li, R., Dai, Y., Chen, X., and Wang, X. (2018). Genome-wide identification and expression analysis of the B-box gene family in the Apple (Malus domestica Borkh.) genome. *Mol. Genet. Genom.* 293, 303–315. doi: 10.1007/s00438-017-1386-1

Liu, Y., Che, F., Wang, L., Meng, R., Zhang, X., and Zhao, Z. (2013). Fruit coloration and anthocyanin biosynthesis after bag removal in non-red and red apples (Malus × domestica Borkh.). *Molecules* 18, 1549–1563. doi: 10.3390/molecules18021549

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

Lotkowska, M. E., Tohge, T., Fernie, A. R., Xue, G. P., Balazadeh, S., and Mueller-Roeber, B. (2015). The arabidopsis transcription factor MYB112 promotes anthocyanin formation during salinity and under high light stress. *Plant Physiol.* 169, 1862–1880. doi: 10.1104/pp.15.00605

Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., et al. (2013). Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in Arabidopsis. *Plant J.* 74, 638–651. doi: 10.1111/tpj.12153

Martin, C., Prescott, A., Mackay, S., Bartlett, J., and Vrijlandt, E. (1991). Control of anthocyanin biosynthesis in flowers of Antirrhinum majus. *Plant J.* 1, 37–49. doi: 10.1111/j.1365-313x.1991.00037.x

Merzlyak, M. N., and Chivkunova, O. B. (2000). Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples. *J. Photochem. Photobiol. B* 55, 155–163. doi: 10.1016/s1011-1344(00)00042-7

Ni, M., Tepperman, J. M., and Quail, P. H. (1998). PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* 95, 657–667. doi: 10.1016/s0092-8674(00)81636-0

Oyama, T., Shimura, Y., and Okada, K. (1997). The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev.* 11, 2983–2995. doi: 10.1101/gad.11.22.2983

Passeri, V., Koes, R., and Quattrocchio, F. M. (2016). New challenges for the design of high value plant products: stabilization of anthocyanins in plant vacuoles. *Front. Plant Sci.* 7:153. doi: 10.3389/fpls.2016.00153

Plunkett, B. J., Henry-Kirk, R., Friend, A., Diack, R., Helbig, S., Mouhu, K., et al. (2019). Apple B-box factors regulate light-responsive anthocyanin biosynthesis genes. *Sci. Rep*. 9:17762. doi: 10.1038/s41598-019-54166-2

Quattrocchio, F., Wing, J. F., Leppen, H., Mol, J., and Koes, R. E. (1993). Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *Plant Cell* 5, 1497–1512. doi: 10.1105/tpc.5.11.1497

Quattrocchio, F., Wing, J. F., van der Woude, K., Mol, J. N., and Koes, R. (1998). Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *Plant J.* 13, 475–488. doi: 10.1046/j.1365-313x.1998.00046.x

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011

Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., et al. (2001). The tripartite motif family identifies cell compartments. *Embo J.* 20, 2140–2151. doi: 10.1093/emboj/20.9.2140

Rizzini, L., Favory, J. J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., et al. (2011). Perception of UV-B by the Arabidopsis UVR8 protein. *Science* 332, 103–106. doi: 10.1126/science.1200660

Sarmiento, F. (2013). The BBX subfamily IV: additional cogs and sprockets to fine-tune light-dependent development. *Plant Signal Behav.* 8:e23831. doi: 10.4161/psb.23831

Saure, M. C. (1990). External control of anthocyanin formation in apple. *Scientia Horticulturae* 42, 181–218. doi: 10.1016/0304-4238(90)90082-P

Sharrock, R. A., and Quail, P. H. (1989). Novel phytochrome sequences in Arabidopsis thaliana: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes Dev.* 3, 1745–1757. doi: 10.1101/gad.3.11.1745

Shin, D. H., Choi, M., Kim, K., Bang, G., Cho, M., Choi, S. B., et al. (2013). HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in Arabidopsis. *FEBS Lett.* 587, 1543–1547. doi: 10.1016/j.febslet.2013.03.037

Song, Z., Bian, Y., Liu, J., Sun, Y., and Xu, D. (2020). B-box proteins: pivotal players in light-mediated development in plants. *J. Integr. Plant Biol.* 62, 1293–1309. doi: 10.1111/jipb.12935

Sun, X., Jia, X., Huo, L., Che, R., Gong, X., Wang, P., et al. (2018). MdATG18a overexpression improves tolerance to nitrogen deficiency and regulates anthocyanin accumulation through increased autophagy in transgenic apple. *Plant Cell Environ.* 41, 469–480. doi: 10.1111/pce.13110

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142, 1216–1232. doi: 10.1104/pp.106.088104

Takos, M. A., Robinson, P. S., and Walker, R. A. (2006). Transcriptional regulation of the flavonoid pathway in the skin of dark-grown 'Cripps' Red' apples in response to sunlight. *J. Horticultural Sci. Biotechnol.* 81, 735–744. doi: 10.1080/14620316.2006.11512131

Tao, R., Bai, S., Ni, J., Yang, Q., Zhao, Y., and Teng, Y. (2018). The blue light signal transduction pathway is involved in anthocyanin accumulation in 'Red Zaosu' pear. *Planta* 248, 37–48. doi: 10.1007/s00425-018-2877-y

Wang, Y., Wang, Y., Song, Z., and Zhang, H. (2016). Repression of MYBL2 by both microRNA858a and HY5 leads to the activation of anthocyanin biosynthetic pathway in arabidopsis. *Mol. Plant* 9, 1395–1405. doi: 10.1016/j.molp.2016.07.003

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

Xie, X. B., Li, S., Zhang, R. F., Zhao, J., Chen, Y. C., Zhao, Q., et al. (2012). The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant Cell Environ.* 35, 1884–1897. doi: 10.1111/j.1365-3040.2012.02523.x

Xu, D., Jiang, Y., Li, J., Lin, F., Holm, M., and Deng, X. W. (2016). BBX21, an Arabidopsis B-box protein, directly activates HY5 and is targeted by COP1 for 26S proteasome-mediated degradation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7655–7660. doi: 10.1073/pnas.1607687113

Zhang, B., Yang, H. J., Yang, Y. Z., Zhu, Z. Z., Li, Y. N., Qu, D., et al. (2020). mdm-miR828 participates in the feedback loop to regulate anthocyanin accumulation in apple peel. *Front. Plant Sci.* 11:608109. doi: 10.3389/fpls.2020.608109

Zhang, H., Jordheim, M., Lewis, D. H., Arathoon, S., Andersen, O. M., and Davies, K. M. (2014). Anthocyanins and their differential accumulation in the floral and vegetative tissues of a shrub species (Rhabdothamnus solandri A. Cunn). *Scientia Horticulturae* 165, 29–35. doi: 10.1016/j.scienta.2013.10.032

Zhang, Y., Butelli, E., and Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19, 81–90. doi: 10.1016/j.pbi.2014.05.011

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* 10:1494. doi: 10.1038/s41467-019-09518-x

Zhang, X., Huai, J., Shang, F., Xu, G., Tang, W., Jing, Y., et al. (2017). A PIF1/PIF3-HY5-BBX23 transcription factor cascade affects photomorphogenesis. *Plant Physiol.* 174, 2487–2500. doi: 10.1104/pp.17.00418

Zhao, C., Mao, K., You, C. X., Zhao, X. Y., Wang, S. H., Li, Y. Y., et al. (2016). Molecular cloning and functional analysis of a UV-B photoreceptor gene, MdUVR8 (UV Resistance Locus 8), from apple. *Plant Sci.* 247, 115–126. doi: 10.1016/j.plantsci.2016.03.006

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.

frontiers in Plant Science

REVIEW
published: 10 December 2021
doi: 10.3389/fpls.2021.781236

Effects of Light on Secondary Metabolite Biosynthesis in Medicinal Plants

Shuncang Zhang[1], Lei Zhang[1], Haiyan Zou[1], Lin Qiu[1], Yuwei Zheng[1], Dongfeng Yang[2] and Youping Wang[1]**

[1]College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, China, [2]Key Laboratory of Plant Secondary Metabolism and Regulation of Zhejiang Province, College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou, China

Secondary metabolites (SMs) found in medicinal plants are one of main sources of drugs, cosmetics, and health products. With the increase in demand for these bioactive compounds, improving the content and yield of SMs in medicinal plants has become increasingly important. The content and distribution of SMs in medicinal plants are closely related to environmental factors, especially light. In recent years, artificial light sources have been used in controlled environments for the production and conservation of medicinal germplasm. Therefore, it is essential to elucidate how light affects the accumulation of SMs in different plant species. Here, we systematically summarize recent advances in our understanding of the regulatory roles of light quality, light intensity, and photoperiod in the biosynthesis of three main types of SMs (polyphenols, alkaloids, and terpenoids), and the underlying mechanisms. This article provides a detailed overview of the role of light signaling pathways in SM biosynthesis, which will further promote the application of artificial light sources in medicinal plant production.

Keywords: secondary metabolites, medicinal plants, light quality, light intensity, photoperiod

OPEN ACCESS

Edited by:
Yongliang Liu,
University of Kentucky,
United States

Reviewed by:
Oksana Sytar,
Taras Shevchenko National University of Kyiv, Ukraine
Miroslava Konstantinova Zhiponova,
Sofia University, Bulgaria

****Correspondence:***
Dongfeng Yang
ydf807@sina.com
Youping Wang
wangyp@yzu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *22 September 2021*
Accepted: *17 November 2021*
Published: *10 December 2021*

Citation:
Zhang S, Zhang L, Zou H, Qiu L, Zheng Y, Yang D and Wang Y (2021) Effects of Light on Secondary Metabolite Biosynthesis in Medicinal Plants. Front. Plant Sci. 12:781236. doi: 10.3389/fpls.2021.781236

INTRODUCTION

Medicinal plants play pivotal roles in human development and have been used from the prehistoric times to present day. According to the World Health Organization, 80% of the human population in developing countries relies on traditional medicines, mostly plant-based drugs, for primary healthcare needs. At present, at least 25% of the drugs worldwide are directly or indirectly derived from medicinal plants, which remain the main source of drugs. For example, artemisinin, derived from *Artemisia annua*, is widely used for treating malaria (Ansari et al., 2013). *Ajuga bracteosa* is a high-value medicinal plant that has been recommended as a treatment for gout rheumatism, paralysis, and amenorrhea (Rukh et al., 2019). Secondary metabolites (SMs) serve as the material basis of the clinically curative effects of medicinal plants. SMs refer to small molecular organic compounds not directly involved in plant growth and development, but are essential for the long-term survival of plants (Fraenkel, 1959; Demain and Fang, 2000). Besides their medicinal uses, SMs are also widely used in cosmetics and healthcare products (Craker and Gardner, 2006; Mathur and Velpandian, 2009; Schmidt, 2012).

Biosynthesis and accumulation of SMs in medicinal plants are affected by environmental factors, such as water, light, temperature, soil properties, and chemical stress (Verma and Shukla, 2015; Li et al., 2020). Among these factors, light is reported to affect the accumulation of almost all types of SMs. Light quality, light intensity, and photoperiod affect the SM content of plants. For example, the composition and content of SMs in the same plant species vary greatly among different regions because of the variation in light conditions (Huang and Liu, 2015). Controlled growing systems using artificial lighting have been widely applied with the increasing demand for natural products. These systems are initially developed for the production of out of season crops and vegetables. In recently years, they are also used to enhance the SMs yields in medicinal plants (Askari-Khorasgani and Pessarakli, 2019; Moon et al., 2020; Malik et al., 2021). Artificial lights take important place in controlled growing systems and light emitting diodes (LEDs) are supposed to be optimal artificial light sources at present time (Jung et al., 2021). Compared with other types of lamps, such as fluorescent, high-pressure sodium (HPS), and metal-halide, LEDs show equivalent or higher luminous efficacy, lack of radiant heat, and longer lifespan (Loi et al., 2021). Besides, LED can also produce the monochromatic light wavelength and make it more convenient to change the light quality constitution in the controlled growing systems (Palmitessa et al., 2021). Understanding how light affects SM biosynthesis is essential for the cultivation of medicinal plants in a controlled environment as well as under open field conditions. In this review, we discuss the roles of light in the accumulation of different types of SMs, with the aim to identify the gaps in research and to provide a reference for the further investigation of the mechanism underlying light-mediated SM biosynthesis in medicinal plants.

MAIN SECONDARY METABOLITES IN MEDICINAL PLANTS

Based on their structures and biosynthetic pathways, plant SMs are mainly divided into polyphenols (phenolics), terpenoids, and alkaloids (Chiocchio et al., 2021). Polyphenols are a large and complex family of phytochemicals containing at least one aromatic ring and a hydroxyl group as functional derivatives. Over 8,000 polyphenols have been identified in plants to date. They are present in almost all plant species and have gained considerable attention because of their nutritional and pharmaceutical applications (Balasundram et al., 2006). According to the biosynthetic pathways, basic skeletons, and hydroxyl groups, polyphenols are categorized into different sub-classes, including coumarins, lignans, phenolic acids, flavonoids, and tannins (Huang and Liu, 2015; Chiocchio et al., 2021). Coumarins (C_6-C_3) are a class of lactones structurally constructed by a benzene ring fused to α-pyrone ring, such as aesculin, cnidium lactone, and alpha-Angelica lactone. Lignans [$(C_6$-$C_3)_2$] are phenolic dimers with a 2,3-dibenzylbutane skeleton, such as phyllanthin, arctiin, and podophyllotoxin. Phenolic acids present compounds containing a carboxylic group among the substituents on the benzene ring, including benzoic acid derivates (C_6-C_1) and hydroxycinnamic derivates (C_6-C_3). Flavonoids (C_6-C_3-C_6) refer to compounds consisting of two benzene rings linked by a short three carbon chain, such as chalcones, flavones, flavonols, dihydroflavones, dihydroflavonols, isoflavones, and dihydroisoflavones. Tannins are high molecular polyphenols polymerized by flavonoids units (condensed tannins) or gallic acid esterified with monosaccharide (hydrolysable tannins; King and Young, 1999). **Figure 1A** shows structures of some common bioactive polyphenols in medicinal plants, including three phenolic acids (salvianolic acid B, chlorogenic acid, and rosmarinic acid), two phenylethanol glycosides (verbascoside and salidroside), and one flavanone derivate (rutin). Polyphenols are biosynthesized *via* phenylpropanoid pathway in plants, they share a common upstream biosynthetic pathway derived from the shikimic acid pathway, but the downstream biosynthetic pathways of different polyphenols are distinct (**Figure 1B**; Dixon and Paiva, 1995).

Terpenoids are compounds with isoprene as the structural unit. According to the number of isoprene structural units, terpenoids are divided into five categories: monoterpenes, sesquiterpenes, diterpenes, triterpenes, and tetraterpenes (Bohlmann et al., 1998). **Figure 2A** shows structures of some common bioactive terpenoids in medicinal plants, including one sesquiterpene (artemisinin), three diterpenes (cryptotanshinone, tanshinone II_A, and paclitaxel), and two triterpenes (cucurbitacin I and oleanolic acid). Terpenoids are synthesized *via* two parallel upstream pathways: the mevalonate pathway (MVA) and methylerythritol-4-phosphate pathway (MEP; da Silva et al., 2017). Acetyl-CoA acts as the precursor of terpenoids in the MVA pathway, whereas pyruvate and glyceraldehyde-3-phosphate (G3P) serve as precursors in the MEP pathway. These precursors are converted to isopentenyl pyrophosphate (IPP) through a series of enzymatic reactions, which is then converted to its isomer dimethyl-allyl pyrophosphate (DMAPP) by isopentenyl diphosphate isomerase (IDI). Both DMAPP and IPP are converted to geranylgeranyl diphosphate (GGPP) by geranylgeranyl diphosphate synthase (GGPPS), and then, GGPP is used to produce diterpenes, tetraterpenes, and their derivatives (Newman and Chappell, 1999). DMAPP and IPP are also converted to farnesyl diphosphate (FPP) and geranyl diphosphate (GPP) by farnesyl diphosphate synthase (FPPS) and geranyl diphosphate synthase (GPPS), respectively. Both FPP and GPP then act as precursors of monoterpenes, sesquiterpenes, and triterpenes (**Figure 2B**; Chappell, 1995; Bergman et al., 2019).

Alkaloids form a large class of heterocyclic nitrogen organic compounds, with over 10,000 isolated to date. Given their antitumor, antibacterial, and anti-inflammatory activities, alkaloids have been widely used for the production of medications (Jain et al., 2019; Zhang et al., 2021). Depending on their biosynthetic pathways and chemical structures, alkaloids are classified into five groups: terpene indole, benzylisoquinoline, tropine, purine, and pyrrolizidine alkaloids (Bhambhani et al., 2021). **Figure 3A** shows structures of some common bioactive alkaloids in medicinal plants, including three monoterpenoid indole alkaloids (MIAs; vindoline, vinblastine, and camptothecin),

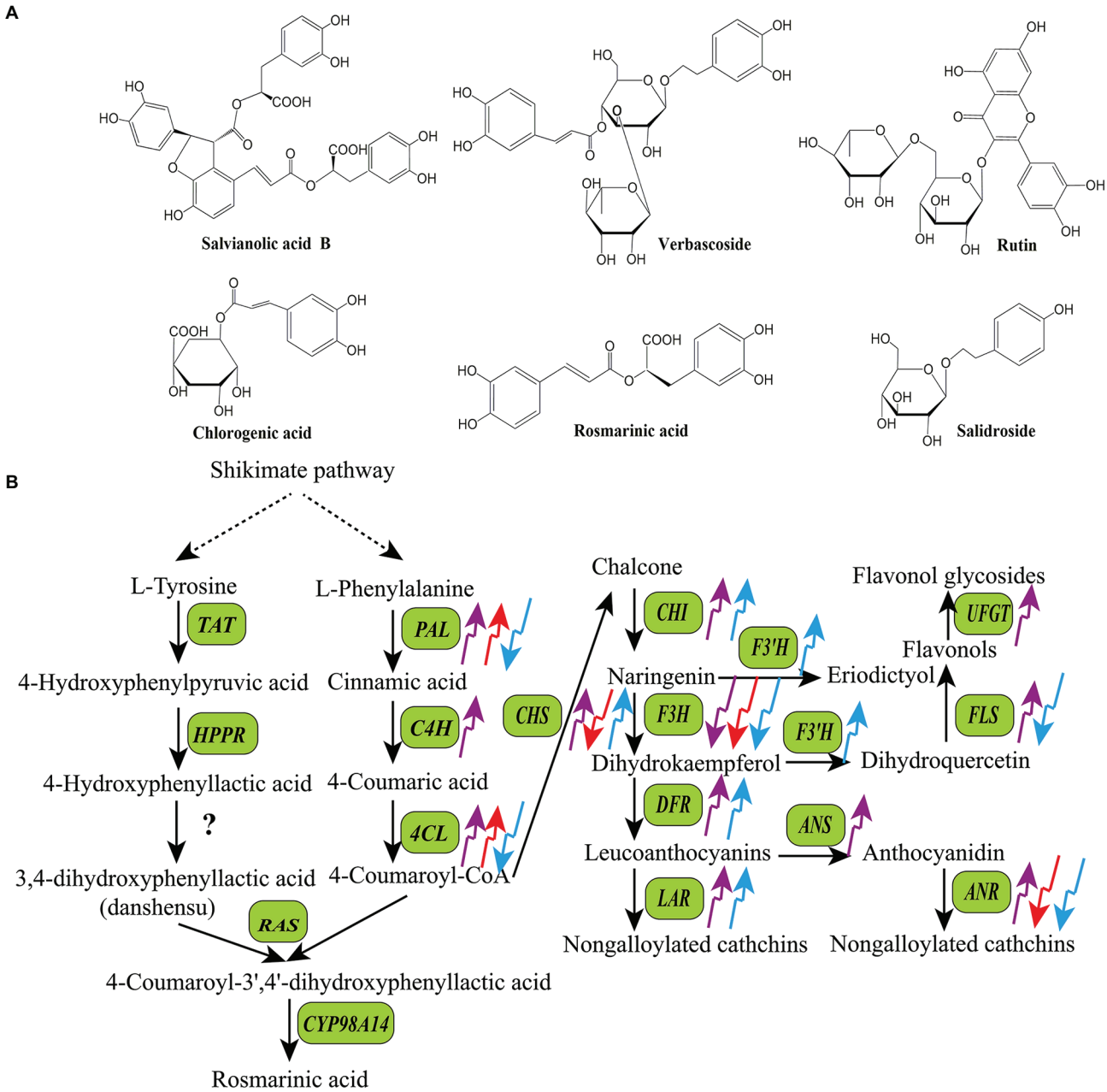


FIGURE 1 | Structures of some bioactive polyphenols in medicinal plants **(A)** and the effects of light quality on the transcripts of key enzyme genes in the biosynthetic pathways of polyphenols **(B)**. The upward arrows indicate upregulation and the downward arrows indicate downregulation. The purple arrow, red arrow, and blue arrow indicate UV-B, red light, and blue light, respectively. *ANR*, anthocyanidin reductase; *ANS*, anthocyanidin synthase; *CHS*, chalcone synthase; *CHI*, chalcone isomerase; *CYP98A14*, cytochrome p450 98A14; *C4H*, cinnamic acid 4-hydroxylase; *DFR*, dihydroflavonol 4-reductase; *FLS*, flavonol synthase; *F3′H*, flavonol 3′-hydroxylase; *F3H*, flavanone 3-hydroxylase; *HPPR*, 4-hydroxyphenylpyruvate reductase; *LAR*, leucoanthocyanidin reductase; *PAL*, phenylalanine ammonia-lyase; *RAS*, rosmarinic acid synthase; *TAT*, tyrosine aminotransferase; *UFGT*, UDP flavonoid glucosyltransferase; *4CL*, 4-coumaric acid: CoA ligase.

one benzylisoquinoline alkaloid (berberine), and two tropine alkaloids (hyoscyamine and scopolamine). The accumulations of MIAs and tropane alkaloids have been reported to be affected by light. Biosynthetic pathways of MIAs and tropane alkaloids have been characterized. MIAs are synthesized *via* two parallel upstream pathways, generating tryptamine and secologanin, which are converted to strictosidine by strictosidine synthase (STR), and then to various MIAs, such as camptothecin, serpentine, and vinblastine (Herrmann, 1995; Radwanski et al., 1996; Huang et al., 2016). Ornithine and phenylalanine, the initial precursors of tropine alkaloids, are converted to littorine, which is then converted to anisodamine and scopolamine *via*

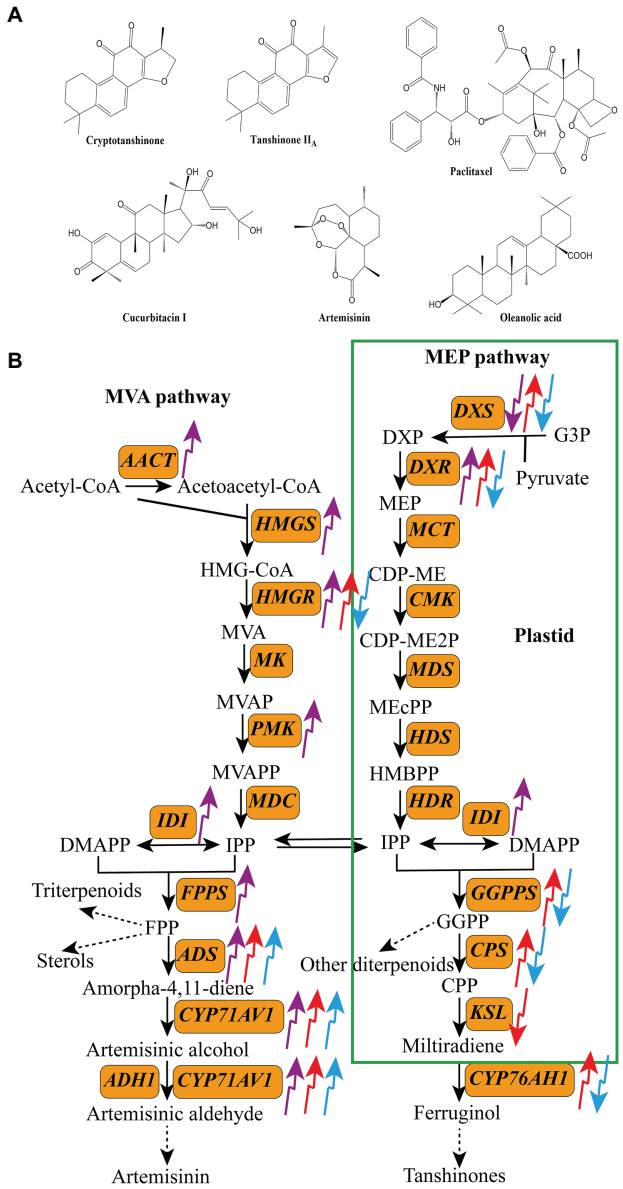


FIGURE 2 | Structures of some bioactive terpenoids in medicinal plants **(A)** and the effects of light quality on the transcripts of key enzyme genes in the biosynthetic pathways of terpenoids **(B)**. The upward arrows indicate upregulation and the downward arrows indicate downregulation. The purple arrow, red arrow, and blue arrow indicate UV-B, red light, and blue light, respectively. *AACT*, acetyl-CoA acyltransferase; *ADH1*, alcohol dehydrogenase 1; *ADS*, amorpha-4,11-diene synthase; *CMK*, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase; *CPS*, copalyl diphosphate synthase; *CYP71AV1*, cytochrome P450 71AV1; *CYP76AH1*, miltiradiene oxidase; *DXR*, 1-deoxy-D-lxylulose-5-phosphate reductoisomerase; *DXS*, 1-deoxy-D-xylulose-5-phosphate synthase; *FPPS*, farnesyl diphosphate synthase; *GGPPS*, geranylgeranyl diphosphate synthase; *HDR*, hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase; *HDS*, hydroxymethylbutenyl diphosphate synthase; *HMGR*, 3-hydroxy-3-methylglutaryl CoA reductase; *HMGS*, 3-hydroxy-3-methylglutary-1 CoA synthase; *IDI*, isopentenyl diphosphate isomerase; *KSL*, kaurene synthase-like; *MCT*, 2-C-methyl-D-erythritol-4-phosphate cytidylyltransferase; *MDC*, mevalonate 5-diphosphate decarboxylase; *MDS*, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; *MK*, mevalonate kinase; *PMK*, phosphomevalonate kinase.

several enzymatic reactions (**Figures 3B**,**C**; Nguyen et al., 2015; Qiu et al., 2020).

LIGHT QUALITY

Light Spectra and Photoreceptors

Depending on the light wavelength, the solar spectrum is divided into three parts, namely ultraviolet light (200–400 nm; UV-A, 315–400 nm; UV-B, 280–320 nm; UV-C, 200–280 nm), visible light or photosynthetically active radiation (PAR; blue light, 400–500 nm; green light, 500–600 nm; red light, 600–700 nm), and far-red light (700–800 nm). Photoreceptors are indispensable for light sensing and light signal transduction in plants. To date, at least five kinds of photoreceptors have been identified in the model plant species *Arabidopsis thaliana*: (1) phytochromes (phyA–phyE), which mainly perceive red light and far-red light (Sharrock and Quail, 1989; Sullivan and Deng, 2003; Sanchez et al., 2020); (2) cryptochromes (crys), which mainly perceive blue light and UV-A. (3) phototropins (phots), which mainly perceive blue light (Gyula et al., 2003); (4) zeitlupes, which mainly perceive blue light and green light (450–520 nm; Somers et al., 2000); and (5) UV photoreceptor UVR8, which mainly perceives UV-B (280–315 nm; Rizzini et al., 2011; Christie et al., 2012). Among the five phytochromes, phyA is the main photoreceptor that perceives far-red light, while phyB mainly perceives red light. Owing to these photoreceptors, plants can accurately detect changes in light wavelength, direction, intensity, and duration, and respond timely.

Polyphenols

The production of polyphenols is significantly affected by light in medicinal plants. UV-B is an important part of the solar spectrum; however, excessive UV-B can damage plants. To deal with UV-B-induced stress and to improve their resilience to adverse conditions, plants synthesize various types of SMs (Schreiner et al., 2012; Kumari and Prasad, 2013; Takshak and Agrawal, 2019). Among these SMs, polyphenols, owing to their high antioxidant potential, play important roles in the acclimation of plants to UV-B (Takshak and Agrawal, 2019). The biosynthesis of bioactive polyphenols is affected by UV-B in many medicinal plants (**Table 1**; Rai et al., 2011; Pandey and Pandey-Rai, 2014; Takshak and Agrawal, 2014; Hamid et al., 2019). Takshak and Agrawal (2015) showed that anthocyanin contents of *Coleus forskohlii* plants treated with supplementary UV-B (ambient +0.042 Wm^{-2}) were 1.34-, 1.67-, and 1.96-fold higher than that in control at 40, 70, and 100 days after transplantation (DAT), respectively. Zhang et al. (2017) investigated the effects of UV-B on SM accumulation in *Prunella vulgaris* and showed that UV-B radiation significantly increased the contents of total flavonoids, rosmarinic acid, caffeic acid, and hyperoside. In *Rosmarinus officinalis* plants treated with UV-B (0.359 Wm^{-2}), the contents of rosmarinic acid and carnosic acid were 2.34- and 1.78-fold higher, respectively, than that of the control (Luis et al., 2007). The total phenolic content of *Cymbopogon citratus* increased by 45% after the UV-B treatment (Kumari and Agrawal, 2010). In addition to UV-B, UV-A also affects the

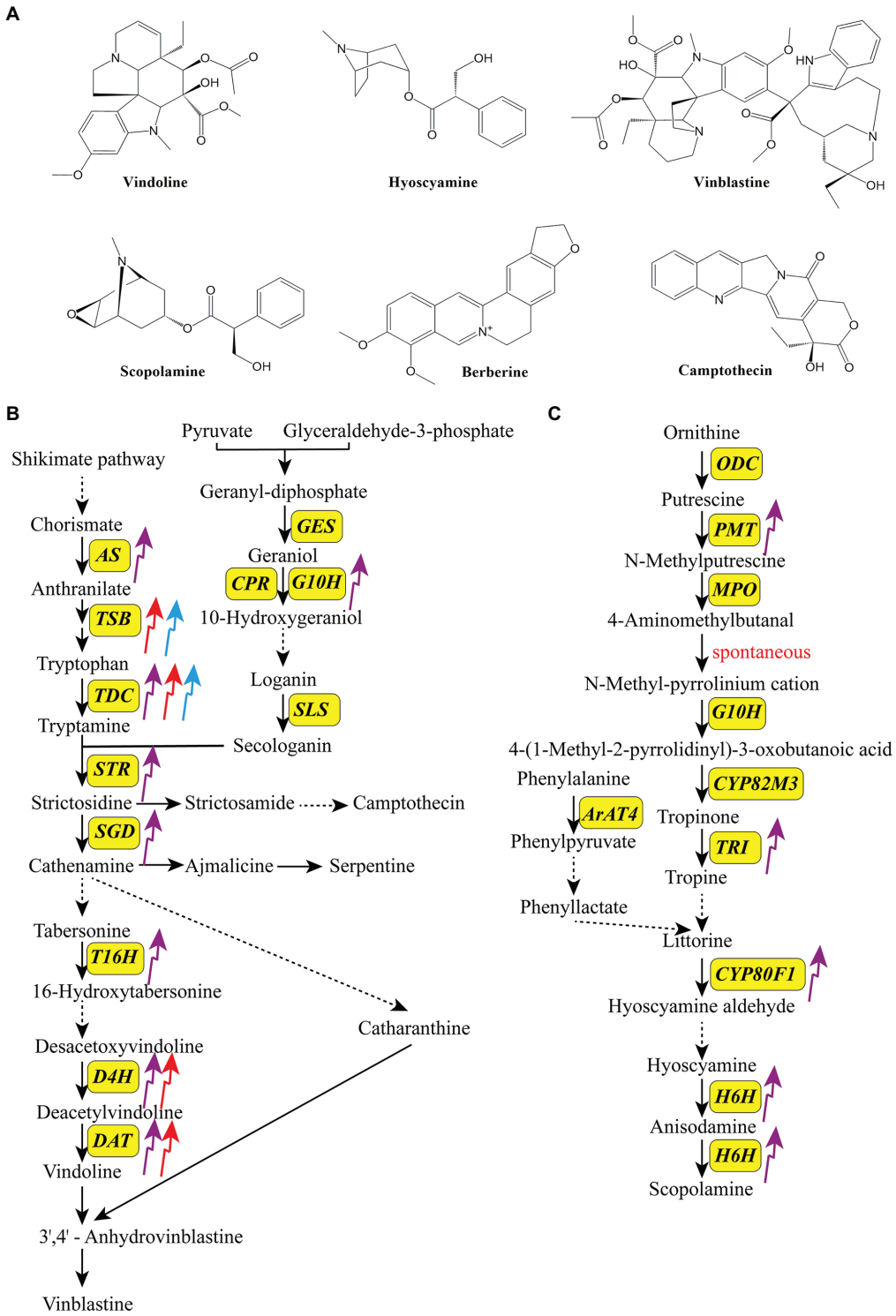


FIGURE 3 | Structures of some bioactive alkaloids in medicinal plants **(A)** and the effects of light quality on the transcripts of key enzyme genes in the biosynthetic pathways of monoterpenoid indole terpenoid alkaloids **(B)** and tropane alkaloids **(C)**. The upward arrows indicate upregulation and the downward arrows indicate downregulation. The purple arrow, red arrow, and blue arrow indicate UV-B, red light, and blue light, respectively. *ArAT4*, aromatic amino acid aminotransferase 4; *AS*, anthranilate synthase; *CPR*, cytochrome P450 reductase; *CYP80F1*, cytochrome P450 80F1; *CYP80M3*, cytochrome P450 80 M3; *DAT*, 6-17-O-deacetylvindoline O-acetyltransferase; *D4H*, deacetoxyvindoline 4-hydroxylase; *GES*, geraniol synthase; *H6H*, hyoscyamine 6b-hydroxylase; *G10H*, geraniol-10-hydroxylase; *MPO*, N-methylputrescine oxidase; *ODC*, ornithine decarboxylase; *PMT*, putrescine N-methyltransferase; *SGD*, strictosidine β-glucosidase; *SLS*, secologanin synthase; *STR*, strictosidine synthase; *TDC*, tryptophan decarboxylase; *TRI*, tropinone reductase I; *T16H*, tabersonine 16-hydroxylase.

TABLE 1 | The effects of light quality on the accumulation of phytochemicals in some medicinal plants.

Species	Class	Light quality	Treatment time	Modulation	References
Artemisia annua	Polyphenol	UV-B ($2.8\ W\ m^{-2}$)	1, 2, 3 and 4 h	↑Flavonoids ↑Anthocyanin	Pandey and Pandey-Rai, 2014
Cymbopogon citratus	Polyphenol	UV-B (ambient+0.021 and +0.042 $W\ m^{-2}$)	3 h d^{-1} for 80 days	↑Flavonoids	Kumari and Agrawal, 2010
Fagopyrum esculentum	Polyphenol	UV-B ($890\ W\ m^{-2}$)	24 h	↑Anthocyanin ↑Rutin	Tsurunaga et al., 2013
Heteropogon contortus	Polyphenol	UV-B (ambient+0.083 $W\ m^{-2}$)	3 h, d^{-1} for 110 days	↑Tannins ↑Total phenolics	Hamid et al., 2019
Perilla frutescens	Polyphenol	UV-B ($0.05\ W\ m^{-2}$)	3 h	↑Rosmarinic acid	Yoshida et al., 2021
Prunella vulgaris	Polyphenol	UV-B ($0.35\ W\ m^{-2}$)	30 min d^{-1} for 15 days	↑Total flavonoids ↑Rosmarinic acid ↑Caffeic acid	Zhang et al., 2017
Rosmarinus officinalis	Polyphenol	UV-B (0.063 and $0.359\ W\ m^{-2}$)	14 days	↑Rosmarinic acid ↑Carnosic acid ↑Total phenolics	Luis et al., 2007
Salvia verticillata	Polyphenol	UV-B ($0.127\ W\ m^{-2}$)	1 h, d^{-1} for 5 days	↑Chlorogenic acid ↑Caffeic acid ↑Rosmarinic acid	Rizi et al., 2021
Tropaeolum majus	Polyphenol	UV-B (7.5e-4 and 1.5e-3 $W\ m^{-2}$)	2 and 22 h	↑Glucotropaeolin	Schreiner et al., 2009
Withania somnifera	Polyphenol	UV-B ($0.111\ W\ m^{-2}$)	3 h d^{-1} for 100 days	↑Anthocyanins, ↑Flavonoids	Takshak and Agrawal, 2014
Crepis japonica	Polyphenol	UV-A ($6–25\ W\ m^{-2}$)	11.5 h d^{-1} for 60 days	↑Caffeic acid ↑Chlorogenic acid	Constantino et al., 2017
Ixeris dentata	Polyphenol	UV-A UV-B	24 h·d^{-1} for 5 days 4 h·d^{-1} for 3 days	↑Total phenolics ↑Total flavonoid ↑Total phenolics	Lee et al., 2013
Cichorium intybus	Polyphenol	Blue light followed by white light ($25\ \mu mol\ m^{-2}\ s^{-1}$ and $320\ \mu mol\ m^{-2}\ s^{-1}$, respectively)	14 h·d^{-1} for 30 days (blue light), 14 h·d^{-1} for 27 days (white light)	↓Chlorogenic acid ↓Caftaric acid ↑Kaempferol-3-(6″-acetyl)-glucoside ↓Isorhamnetin-3-(6″-acetyl)-glucoside ↑Isorhamnetin-3-glucoside	Sytar et al., 2019
		Orange light followed by white light ($40\ \mu mol\ m^{-2}\ s^{-1}$ and $320\ \mu mol\ m^{-2}\ s^{-1}$, respectively)	14 h·d^{-1} for 30 days (orange light), 14 h·d^{-1} for 27 days (white light)	↓Chlorogenic acid ↓Kaempferol-3-(6″-acetyl)-glucoside ↑Isorhamnetin-3-(6″-acetyl)-glucoside ↑Isorhamnetin-3-glucoside ↑Feruloyl-malate	
		Red light followed by white light ($40\ \mu mol\ m^{-2}\ s^{-1}$ and $320\ \mu mol\ m^{-2}\ s^{-1}$, respectively)	14 h·d^{-1} for 30 days (red light), 14 h·d^{-1} for 27 days (white light)	↓Chlorogenic acid ↑Quercetin-3-(6″-acetyl)-glucoside ↓Isorhamnetin-3-glucoside	
Dracocephalum forrestii	Polyphenol	Blue light ($40\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 35 days	↑Chlorogenic acid ↑Salvianolic acid B ↓Rosmarinic acid	Weremczuk-Jezyna et al., 2021
Drosera indica	Polyphenol	Blue light ($115\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 14 days	↑Plumbagin	Boonsnongcheep et al., 2019
F. esculentum	Polyphenol	Blue light	16 h d^{-1} for 7 days	↑C-glycosylflavones ↑Orientin ↑Vitexin and its isomers ↑Rutin	Nam et al., 2018

(Continued)

TABLE 1 | Continued

Species	Class	Light quality	Treatment time	Modulation	References
Ocimum basilicum	Polyphenol	Blue light ($40–50\ \mu mol\ m^{-2}\ s^{-1}$)	24 h d^{-1} for 35 days	↑Total phenolics ↑Total flavonoids	Nazir et al., 2020
Rhodiola imbricata	Polyphenol	Blue light ($25–27\ \mu mol\ m^{-2}\ s^{-1}$)	24 h·d^{-1} for 30 days	↑Total phenolics ↑Total flavonoids ↑Salidroside	Kapoor et al., 2018
		Red light ($35–37\ \mu mol\ m^{-2}\ s^{-1}$)	24 h·d^{-1} for 30 days	↑Salidroside	
Salvia przewalskii	Polyphenol	Blue light ($40\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 12 days	↑Salvianolic acid A ↑Salvianolic acid B ↑Rosmarinic acid ↓Protocatechualdehyde	Li, 2014
		Red light ($40\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 12 days	↑Salvianolic acid A ↑Salvianolic acid B ↑Rosmarinic acid ↑Protocatechualdehyde	
Sarcandra glabra	Polyphenol	Blue and red light ($80\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 60 days	↓Isofraxidin ↓Scopoletin ↓Rosmarinic acid	Xie et al., 2021
Scutellaria baicalensis	Polyphenol	Blue and red light ($80\ \mu mol\ m^{-2}\ s^{-1}$)	30 days	↑Baicalin ↑Baicalein ↑Wogonin	Stepanova et al., 2020
Verbena officinalis	Polyphenol	Blue light ($40\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 28 days	↑Verbascoside ↑Isoverbascoside	Kubica et al., 2020
Achyranthes bidentata	Terpenoid	UV-B ($0.205\ W\ m^{-2}$)	2 h and 3 h	↑Oleanolic acid ↑Ecdysterone	Li et al., 2019
A. annua	Terpenoid	UV-B ($0.047\ W\ m^{-2}$)	30 min d^{-1} for 14 days	↑Artemisinin	Rai et al., 2011
		UV-C ($0.066\ W\ m^{-2}$)	30 min d^{-1} for 14 days	↑Artemisinin	
A. annua	Terpenoid	UV-B ($0.017\ W\ m^{-2}$)	1 h d^{-1} for 10 days	↑Artemisinin	Pan et al., 2014
Salvia miltiorrhiza	Terpenoid	UV-B ($0.4\ W\ m^{-2}$)	40 min	↑Total tanshinones ↑Cryptotanshinone ↑Tanshinone II_A ↑Tanshinone I	Wang et al., 2016
A. annua	Terpenoid	Blue and red light ($50 \pm 5\ \mu mol\ m^{-2}\ s^{-1}$)	24 h d^{-1} for 2 days	↑Artemisinin	Zhang et al., 2018b
Dysphania ambrosioides	Terpenoid	Blue light ($60\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 40 days	↓Z-ascaridole	de Carvalho et al., 2020
Mentha canadensis	Terpenoid	Blue and red light (6.7 and 7.1 $\mu mol\ m^{-2}\ s^{-1}$, respectively)	6 h d^{-1} for 14 days	↑Pulegone ↑Menthofuran ↑Menthol	Ueda et al., 2021
Perovskia atriplicifolia	Terpenoid	Blue and red light ($300\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 60 days	↓α-Pinene ↑Camphene ↑δ-3-Carene ↓Camphor	Ghaffari et al., 2019
S. przewalskii	Terpenoid	Blue and red light ($40\ \mu mol\ m^{-2}\ s^{-1}$)	16 h d^{-1} for 12 days	↑Dihydrotanshinone ↑Cryptotanshinone ↑Tanshinone II_A ↓Tanshinone I	Li, 2014
Aquilaria agallocha	Terpenoid	Red light ($\sim 15\ \mu mol\ m^{-2}\ s^{-1}$)	24 h d^{-1} for 2 days	↑Cucurbitacin I ↑Cucurbitacin E	Kuo et al., 2015
		Far-red light ($\sim 15\ \mu mol\ m^{-2}\ s^{-1}$)	24 h d^{-1} for 5 days	↓Cucurbitacin I ↓Cucurbitacin E	

(Continued)

TABLE 1 | Continued

Species	Class	Light quality	Treatment time	Modulation	References
Catharanthus roseus	Alkaloid	UV-B ($13.45\ W\ m^{-2}$)	1 h	↑Strictosidine ↑Vindoline ↑Catharanthine ↑Ajmalicine	Zhu et al., 2015
Clematis terniflora	Alkaloid	UV-B ($1.208\ W\ m^{-2}$)	5 h	↑(6-Hydroxyl-1H-indol-3-yl) carboxylic acid methyl ester	Gao et al., 2016
Mahonia bealei	Alkaloid	UV-B ($1.208\ W\ m^{-2}$)	6 h	↓Berberine ↓Palmatine ↑Columbamine	Zhang et al., 2014
W. somnifera	Alkaloid	UV-B (ambient +0.042 $W\ m^{-2}$)	3 h d^{-1} for 100 days	↑Total alkaloids ↓Withanolide A	Takshak and Agrawal, 2014
Camptotheca acuminata	Alkaloid	Blue light ($1,200 \pm 50\ \mu mol\ m^{-2}\ s^{-1}$)	12 h d^{-1} for 45 days	↑Camptothecin	Liu et al., 2015
Psychotria leiocarpa	Alkaloid	Blue, red and far-red light ($30\ \mu mol\ m^{-2} s^{-1}$)	20 d	↑N,-$^{\beta\text{-}D}$-glucopyranosyl vincosamide	Matsuura et al., 2016
C. roseus	Alkaloid	Red light ($150\ \mu mol\ m^{-2} s^{-1}$)	16 h d^{-1} for 28 days	↑Vindoline ↑Catharanthine	Ohashi et al., 2013

The upward arrow in the table represents the increased content of the corresponding phytochemical, the downward arrow in the table represents the decreased content of the corresponding phytochemical.

production of polyphenols in some medicinal plants. For example, in *Crepis japonica*, treatment with UV-A (6–25 $W\ m^{-2}$) significantly promoted the production of caffeic acid and chlorogenic acid (Constantino et al., 2017). In another study, a 3-day UV-A treatment of *Ixeris dentata* increased the total phenolic and total flavonoid contents by 60 and 40%, respectively, compared with the control (Lee et al., 2013).

The mechanism underlying the regulatory effects of UV-B on the synthesis of polyphenols is not fully understood. However, it has been confirmed that UV-B affects the activities of key polyphenol biosynthetic enzymes and/or transcript levels of the corresponding genes in many medicinal plants (**Figure 1B**). For example, supplemental UV-B radiation significantly improved the activities of phenylalanine ammonia-lyase (PAL), cinnamyl alcohol dehydrogenase (CAD), 4-coumaric acid: CoA ligase (4CL), CHI, and dihydroflavonol 4-reductase (DFR), and enhanced the contents of flavonoids and phenolicp0s in the leaves of *C. forskohlii* and *Withania somnifera* (Takshak and Agrawal, 2014, 2015). In *Sinopodophyllum hexandrum*, treatment with UV-B ($1.07\ W{\cdot}m^{-2}$) significantly decreased the content of podophyllotoxin as well as transcript levels of 12 related genes, including *cinnamic acid 3-hydroxylase* (*C3H*), *caffeoyl-CoA O-methyltransferase* (*CCoAMT*), *cinnamoyl-CoA reductase* (*CCR*), *CAD*, *dirigent protein oxidase* (*DPO*), *pinoresinol-lariciresinol reductase* (*PLR*), *secoisolariciresinol dehydrogenase* (*SDH*), *cytochrome P450 719A23* (*CPY719A23*), *O-methyltransferase3* (*OMT3*), *cytochrome P450 71CU1* (*CYP71CU1*), *OMT1*, and *2-oxoglutarate/Fe(II)-dependent dioxygenase* (*2-ODD*; Lv et al., 2021). In *Glycyrrhiza uralensis*, Zhang et al. (2018a) showed that UV-B radiation stimulated the expression of several genes involved in the flavonoid biosynthetic pathway, such as *cinnamic acid 4-hydroxylase* (*C4H*), *PAL*, *CHS*, *CHI*, and *FLS*. Wulff et al. (1999) reported that UV-B stimulated the accumulation of quercetin-3-glycoside and increased the expression of *CHS* in *Betula pendula*.

Blue and red light wavelengths are two important light qualities involved in plant growth and development. Both light qualities are widely recognized as effective elicitors that regulate the accumulation of bioactive compounds in medicinal plants (**Table 1**; Dou et al., 2017). Fazal et al. (2016a) investigated the effects of monochromatic blue, green, yellow, and red light wavelengths on the production of polyphenols in *P. vulgaris* calli, and found that the calli accumulated most total phenolics (23.9 mg g^{-1} DW) and total flavonoids (1.65 mg g^{-1} DW) under blue light. Similarly, Kapoor et al. (2018) revealed that the callus cultures of *Rhodiola imbricata* accumulated the highest amounts of salidroside, total phenolics, and total flavonoids under blue light compared with those under red light, green light, RGB (40% red:40% green:20% blue), and white light. Kubica et al. (2020) found that both blue light and red light significantly stimulated the accumulation of verbascoside compared with fluorescent lamps (control) in *Verbena officinalis*. Coumarins is an important class of phenols in medicinal plants. The biosynthesis and accumulation of coumarins are significantly affected by blue and red light. For instance, Xie et al. (2021) treated *Sarcandra glabra* seedlings with different monochromatic lights for 60 days and found that the content of fraxetin and 6-methylcoumarin in red light treated group was 45 and 16% of that in control (under white light), respectively. The content of these two coumarins in blue light treated group was 51 and 11% of that in control, respectively. Khurshid et al. (2020) revealed that blue and red light stimulated the accumulation of coumarins in callus culture of *Eclipta alba*. The content of coumarin, wedelolactone, and demethylwedelolactone in the red light treated group (40–50 $\mu mol\ m^{-2}\ s^{-1}$, 28 d) was 3.07-, 1.59-, and 1.59-fold of that in control (white light), respectively.

The content of these compounds in the blue light treated group was 2.24-, 1.43-, and 1.29-fold of that in control, respectively. Combined blue and red light are often used to improve the growth and SM content of medicinal plants simultaneously. For example, Lobiuc et al. (2017) cultured *Ocimum basilicum* seedlings under different light conditions, and found that the dry mass, rosmarinic acid, and caffeic acid contents of seedlings were 1.45-, 15-, and 4-fold higher under combined red and blue light (1R:2B), respectively, than under control (white) conditions. Zhang et al. (2020) treated *Salvia miltiorrhiza* seedlings with monochromatic blue light (B), monochromatic red light (R), and combined blue and red light, and showed that seedling growth and phenolic acid production were stimulated under 7R:3B. Wei et al. (2021) found that combined red and blue LED light (1.61R:1B) improved the growth and cannabidiol content of *Cannabis sativa* seedlings, and increased the aboveground plant biomass, flower biomass, and flower cannabidiol content by 15.2, 238, and 36.53%, respectively, compared with the control (Seedlings growth under high-pressure sodium light). Silva et al. (2020) compared the effects of different light qualities on morphogenesis and SM production in *Pfaffia glomerata* and found that equal proportion of red and blue light (1R:1B) was the best light condition for the accumulation of biomass, anthocyanins, and 20-hydroxyecdisone.

Similar to UV-B, blue and red light wavelengths affect the production of phenylpropanoids by regulating the transcript levels of phenylpropanoid biosynthetic genes (**Figure 1B**; Hao et al., 2016; Alrifai et al., 2019). For example, Zhang et al. (2019b) found that the expression levels of *PAL* and *4CL*, which are required for phlorizin synthesis, are correlated with phlorizin content under red and blue light in *Lithocarpus polystachyus*. Liu et al. (2018a) reported that blue light dramatically induced flavonoid biosynthesis in *Cyclocarya paliurus* leaves, and the flavonoid content was positively correlated with the transcript levels of *PAL*, *4CL*, and *CHS*. MYB transcription factors and microRNAs are also involved in light-induced polyphenol biosynthesis in some medicinal plants. For example, in *Fagopyrum tataricum*, FtMYB16 directly binds to the promoter region of the *flavanone 3-hydroxylase* (*F3′H*) gene under red and bule light to induce its expression and enhance the flavonoid content (Zhang et al., 2019a). In *Dimocarpus longan*, miR393, miR394, and miR395 act as positive regulators of epicatechin production under blue light (Li et al., 2018b).

Terpenoids

Biosynthesis of terpenoids in medicinal plants is closely related to light conditions (**Figure 2B**). Both UV and visible light act as important elicitors of terpenoid synthesis (**Table 1**; Zhang and Bjorn, 2009; Kawoosa et al., 2010; Xie et al., 2021). Among the different light qualities, UV-B is reported to promote the accumulation of terpenoids in many plant species (Takshak and Agrawal, 2019). For instance, in *C. citratus*, treatment with supplemental UV-B increased the total essential oil yield by 25.7% (Kumari and Agrawal, 2010). Artemisinin is a typical sesquiterpene lactone that has attracted considerable attention because of its widespread application in malaria treatment (Ansari et al., 2013). Several studies showed that UV-B treatment induces the biosynthesis of artemisinin and enhances the expression levels of key enzyme-encoding genes, such as *3-hydroxy-3-methylglutaryl CoA reductase* (*HMGR*), *1-deoxy-D-xylulose-5-phosphate reductoisomerase* (*DXR*), *isopentenyl pyrophosphate isomerase* (*IPPi*), *FPS*, *amorpha-4, 11-diene synthase* (*ADS*) gene, *cytochrome P450 71AV1* (*CYP71AV1*), and dihydroartemisinic aldehyde reductase (*RED1*; Yin et al., 2008; Rai et al., 2011; Pan et al., 2014). Similarly, in *S. miltiorrhiza* hairy roots, treatment with UV-B (0.4 W m^{-2}) increased the content of total tanshinones and the transcript levels of *1-deoxy-D-xylulose-5-phosphate synthase* (*SmDXS2*) and *copalyl diphosphate synthase* (*SmCPS*) by 1.5-, 6.2-, and 7.3-fold, respectively, compared with the control (Wang et al., 2016).

Red and blue light are also effective regulators of terpenoid biosynthesis. Generally, red light enhances the accumulation of terpenoids, whereas blue light inhibits terpenoid biosynthesis. Kuo et al. (2015) planted *Aquilaria agallocha* seedlings under different light conditions and found that red light significantly enhanced the contents of cucurbitacin E and I in this species. Similarly, Wang et al. (2018) revealed that red light enhances the production of gypenoside and upregulates the expression of *squalene synthase* (*SS*) and *squalene epoxidase* (*SE*) genes in *Gynostemma pentaphyllum*. Chen et al. (2018) reported that red and blue light irradiation dramatically changes the accumulation of tanshinones in *S. miltiorrhiza* hairy roots; red light treatment upregulated the expression of *SmHMGR*, *SmDXS2*, *SmDXR*, *SmGGPPS*, *SmCPS*, and *CYP76AH1* genes and increased the content of tanshinone II_A by 1.4-fold compared with the control, whereas blue light remarkedly suppressed tanshinone II_A biosynthesis and downregulated the expression of key tanshinone II_A biosynthesis genes. de Carvalho et al. (2020) treated nodal segments of *Dysphania ambrosioides* with different light qualities and showed that blue light inhibited the accumulation of Z-ascaridole in this herb. The effects of red and blue light on terpenoid production are species-specific. For example, both red light and blue light stimulated the biosynthesis of artemisinin and artemisinic acid in *A. annua* but decreased the production of essential oils in *Melissa officinalis* (Chen, 2017; Zhang et al., 2018b).

Alkaloids

UV-B is an effective elicitor of alkaloid production and has been confirmed to promote the biosynthesis of several kinds of alkaloids in medicinal plants (**Table 1**; Peebles et al., 2009; Akula and Ravishankar, 2011). Takshak and Agrawal (2015) revealed that supplementary UV-B (ambient +0.042 Wm^{-2}) treatment increased the alkaloid content of leaves and roots of *C. forskohlii*. In *Clematis terniflora*, UV-B irradiation (1.208 W m^{-2}), followed by dark incubation, increased the content of indole alkaloid (6-hydroxyl-1H-indol-3-yl) carboxylic acid methyl ester by 7-fold (Gao et al., 2016). Similarly, 6 h of UV-B irradiation (1.208 W m^{-2}), followed by dark incubation, significantly enhanced the contents of protoberberine alkaloids, including berberine, jateorhizine, palmatine, and columbamine, in *Mahonia bealei* leaves (Zhang et al., 2014). Ramani and Jayabaskaran (2008) showed that

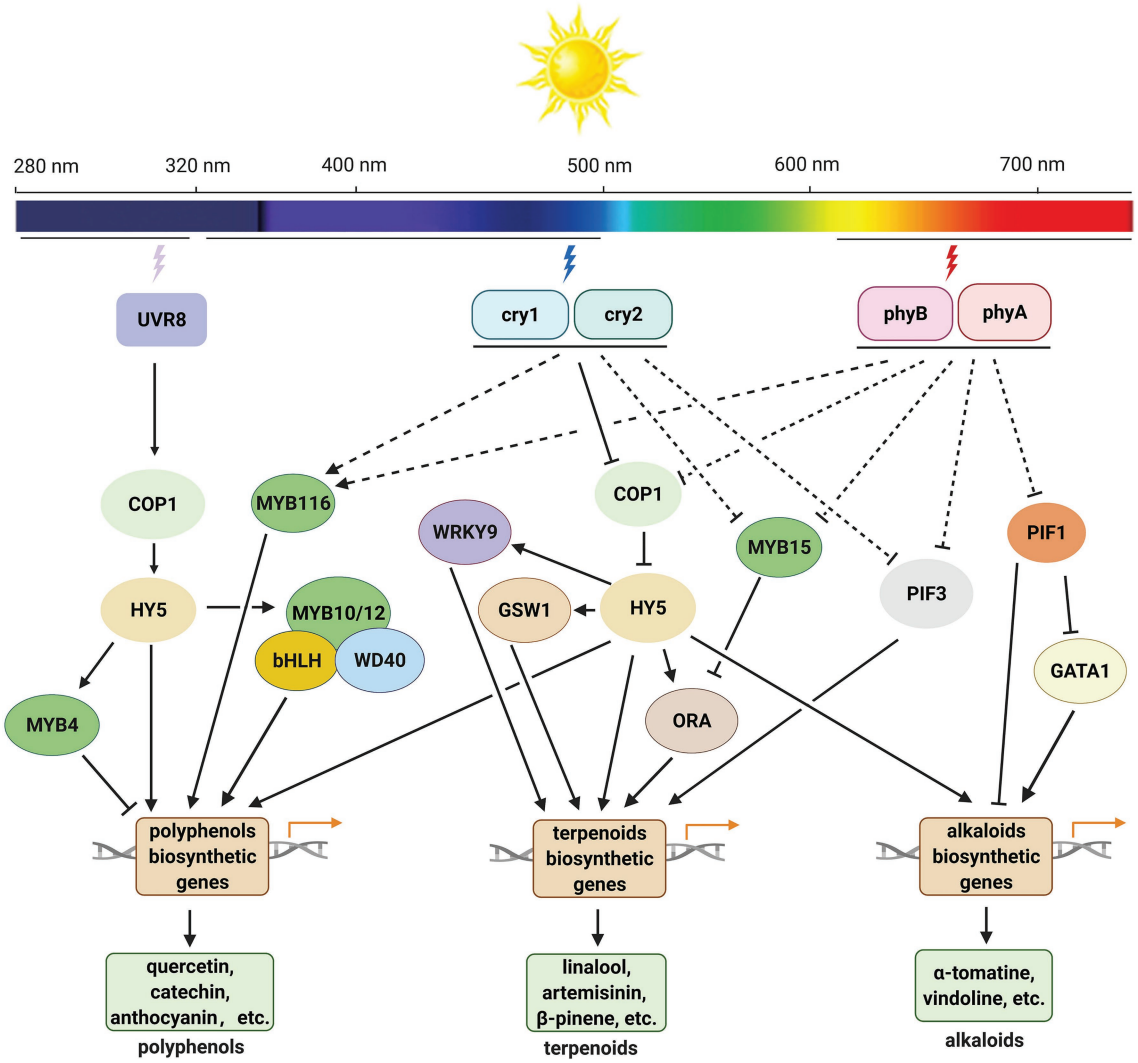


FIGURE 4 | Working model for light-mediated biosynthesis of polyphenols, terpenoids, and alkaloids in medicinal plants. Upon UV-B radiation, the UVR8 homodimer undergoes monomerization and interacts with COP1, which increases COP1 stability and induces *HY5* expression. HY5 regulates the expression of genes encoding transcription factors and/or key enzymes involved in the biosynthesis of SMs and affects the accumulation of SMs under UV-B. Under blue and red light, cryptochromes (cry1 and cry 2) and phytochromes (phyA and phyB) inhibit the E3 ligase activity of COP1, HY5 accumulates in nucleus and promotes the expression of transcription factor and/or key enzyme genes in the related biosynthetic pathways of SMs. Phytochromes may also affect the accumulation of alkaloids by mediating the function of PIF1 and PIF3 in some species. COP1, CONSTITUTIVELY PHOTOMORPHOGENIC1; cry1 and cry2, cryptochrome 1 and cryptochrome 2; GATA1, GATA-type transcription factor 1; GSW1, GLANDULAR TRICHOME-SPECIFIC WRKY 1; HY5, LONG HYPOCOTYL 5; ORA, AP2/ERF type transcription factor; phyA and phyB, phytochrome A and phytochrome B; PIF1, phytochrome interacting factor 1; PIF3, phytochrome interacting factor 3; UVR8, UV RESISTANCE LOCUS 8.

treatment of the suspension cultures of *Catharanthus roseus* with UV-B for 5 min increased the contents of catharanthine and vindoline by 3- and 12-fold, respectively. Many key enzyme-encoding genes in the alkaloid biosynthetic pathways are UV-B inducible (**Figures 3B**,**C**; Takshak and Agrawal, 2019). For example, Zhu et al. (2015) exposed *C. roseus* to binary stress (enhanced UV-B followed by dark incubation), and found that most of structural genes in the alkaloid biosynthetic pathways were upregulated, among which *10-hydroxygeranioloxidoreductase* (*10-HGO*), *tabersonine 16-hydroxylase* (*T16H*), and *strictosidine synthase* (*STR*) genes were upregulated by approximately 2-, 4-, and 4-fold, respectively, compared with the control. Gao et al. (2016) reported that UV-B irradiation remarkedly stimulated the expression of upstream genes in the indole alkaloid biosynthetic pathways in *C. terniflora*. The expression of key genes in the tropane alkaloid biosynthetic pathway is also affected by UV-B. In *Anisodus luridus* hairy roots treated with UV-B (90 W m^{-2}) for 24 h, transcript levels of *putrescine N-methyltransferase* (*PMT*), *tropinone reductase I* (*TRI*), *cytochrome P450 80F1* (*CYP80F1*), and *hyoscyamine 6b-hydroxylase* (*H6H*) genes were increased by 10-, 52-, 16-, and 9-fold, respectively, compared with the control (Qin et al., 2014).

Similar to UV-B, blue and red light are also reported to stimulate the accumulation of alkaloids as well as the transcript levels of related genes (**Figures 3B,C**; Takshak and Agrawal, 2019). Li et al. (2021) revealed that blue light irradiation (100 μmol m^{-2} s^{-1}, 30 d) dramatically increased the production of galanthamine, lycorine, and lycoramine as well as the expression of *norbelladine synthase* (*NBS*), *OMT*, and *CYP96T* in *Lycoris longituba*. Liu et al. (2015) treated *Camptotheca acuminata* seedlings with different light conditions and showed that blue light (1,200 ± 50 μmol m^{-2} s^{-1}, 45 d) promoted the camptothecin content, tryptophan decarboxylase (TDC) and tryptophan synthase (TSB) activities, and *TSB*, *TDC1*, and *TDC2* transcript levels. Matsuura et al. (2016) reported that blue light is more beneficial for the production of the monoterpene indole alkaloid N,β-D-glucopyranosyl vincosamide than other light qualities in *Psychotria leiocarp*. Red light is considered as another regulator of alkaloid biosynthesis, and its function is dependent on phytochromes as well as secondary messengers (G protein and CaM; Aerts and De Luca, 1992; Wang et al., 2010). In *C. roseus*, red light induces vindoline production by increasing the expression of the transcription factor gene *GATA1* and vindoline pathway genes *T16H2*, *tabersonine-3-oxygenase* (*T3O*), *tabersonine-3-reductase* (*T3R*), *desacetoxyvindoline-4-hydroxylase* (*D4H*), and *DAT*. In the dark, PIF1 suppresses the expression of the abovementioned genes, which dramatically decreases the accumulation of vindoline (Liu et al., 2019). Similarly, red light treatment (150 μmol m^{-2} s^{-1}) significantly enhanced the concentration and yield of vindoline and catharanthine in *C. roseus* seedlings (Ohashi et al., 2013).

LIGHT INTENSITY

Optimal light conditions required for growth and development differ among different medicinal plant species. According to their sunlight intensity requirements, medicinal plants are classified into three types: heliophytes, sciophytes, and intermediates. Similar to other physiological processes, the accumulation of SMs in medicinal plants is significantly affected by light intensity (Chen et al., 2017; Li et al., 2020). Generally, high light intensity promotes SM production in heliophytes, such as *Ginkgo biloba* (Xu et al., 2014), *Lonicera japonica* (Fang et al., 2020), *Tabernaemontana pachysiphon* (Hoft et al., 1996), and *Andrographis paniculata* (Saravanan et al., 2008), while low light intensity promotes SM production in sciophytes, such as *Glechoma longituba* (Zhang et al., 2015), *Changium smyrnioides* (Wang et al., 2017), *Polygonum minus* (Mohd Yusof et al., 2021), and *Panax ginseng* (Jung et al., 2020). Concentration and yield are two important parameters that should be considered for SM production in medicinal plants. For example, Li et al. (2018a) reported that 30 and 50% sunlight are better light conditions than 10 and 100% sunlight, for total alkaloid production in *Mahonia breviracema*, as the former light intensities result in higher biomass. Similarly, in *G. longituba*, the concentration and yield of ursolic acid and oleanolic acid were stimulated under 33% sunlight and 16% sunlight, respectively, compared with other light intensities (Zhang et al., 2015).

Light intensity also affects the chemical composition of medicinal plants. For instance, Xu et al. (2020) fond that moderate shade (38.8% of the control) promoted the accumulation of C_6C_1- and C_6C_3-type phenolics, while severe shading (16.9% of the control) stimulated the accumulation of $C_6C_3C_6$-type phenolics in *Eleutherococcus senticosus*. Santos Lazzarini et al. (2018) treated *Lippia gracilis* with different light intensities and showed that the chemical composition of seedlings in the 26 μmol m^{-2} s^{-1} group was more complex, with more monoterpene hydrocarbons and less aromatic monoterpenes, than those in other treated groups. In *M. breviracema*, light intensity had different effects on the production of alkaloids and essential oils; alkaloid content was significantly enhanced in the 30 and 50% sunlight groups, while the accumulation of essential oils increased linearly with the increase in light intensity, reaching maximum levels in the full (100%) sunlight group (Li et al., 2018a).

PHOTOPERIOD

Photoperiod is one of the critical environmental factors that regulate the growth and development of medicinal plants and is often closely related to other environmental factors, such as latitude, slope direction, and seasonal changes (Liebelt et al., 2019). Depending on the day length or amount of light required for flowering, medicinal plants are classified into long-day, short-day, and intermediate-day plants (Huang and Liu, 2015). Different plant species adapt to changes in the photoperiod through various physiological modifications, one of which is by altering the accumulation of SMs (Moyo et al., 2014; Zahir et al., 2014; Liebelt et al., 2019).

In many medicinal plants, photoperiod promotes the accumulation of SMs. For example, Fazal et al. (2016b) optimized the conditions required for *P. vulgaris* suspension culture and found that the biomass and SM content were higher under 18-h light/12-h dark (18L/12D), 16L/14D, and 14L/16D photoperiods compared with the control (16L/8D). Kumar et al. (2020) found that *Basella rubra* callus cultures under the 16:8 h photoperiod produced the highest amount of phenolics compared with those under continuous light and continuous dark conditions. Wu et al. (2007) studied the effects of photoperiod on the growth and caffeic acid derivative content of the adventitious root cultures of *Echinacea purpurea*, and found that the accumulation of caffeic acid derivatives was optimum in cultures grown under 3-h light/21-h dark conditions. There are also many studies which report that continuous light or continuous dark is more efficiency to stimulate the bioactive compounds biosynthesis compared with photoperiod in some medicinal plants. For instance, de Castro et al. (2019) found that 24-h d^{-1} light was the best light condition for enhancing the growth and essential oil content of *Lippia alba* seedlings grown *in vitro*. Anjum et al. (2017) treated the cell cultures of *Linum usitatissimum* with different photoperiods, and

showed that continuous dark conditions led to the greatest increase in total phenolics (116.85 mg L^{-1}) and total flavonoids (37.05 mg L^{-1}). Photoperiod also affects the chemical composition of medicinal plants. Tusevski et al. (2013) found that hairy roots of *Hypericum perforatum* cultured under 16-h light/8-h dark photoperiod showed *de novo* biosynthesis of two phenolic acids, three flavonol glycosides, and five xanthones compared those cultured under continuous dark. Fonseca et al. (2006) revealed that dark incubation decreased the content of parthenolide and increased the content of total phenolics in *Tanacetum parthenium*, while photoperiod showed an opposite effect on the accumulation of these compounds.

SUMMARY AND PERSPECTIVES

In this review, we summarized the regulatory roles of light quality, light intensity, and photoperiod in SM accumulation in medicinal plants and the known mechanisms underlying these roles (**Figure 4**). Generally, when plants are exposed to UV-B radiation, the UVR8 homodimer undergoes monomerization and interacts with CONSTITUTIVELY PHOTOMORPHOGENIC1 (COP1), which increases COP1 stability and induces *LONG HYPOCOTYL 5* (*HY5*) expression. HY5 is a core transcription factor in the light signaling pathway that regulates expression of genes encoding transcription factors or key enzymes involved in the biosynthesis of SMs, to affect the accumulation of SMs under UV-B (Gangappa and Botto, 2016; Liu et al., 2018b; Yang et al., 2018). Phytochromes (phyA and phyB) and cryptochromes (cry1 and cry2) regulate the accumulation of SMs in a different way (Hemm et al., 2004; Chen et al., 2006; Li et al., 2014; Fu et al., 2021a). These photoreceptors interact with COP1 or SUPPRESSOR OF PHYA-105 (SPA) in a light-dependent manner and inhibit the E3 ubiquitin ligase activity of the COP1–SPA complex. Under dark conditions, COP1 enters the nucleus and mediates the degradation of the HY5 protein through the 26S proteasome. Under blue, red, and far-red light conditions, HY5 functions its roles normally and promotes the production of SMs. In *A. annua*, HY5 binds to the promoters of some transcription factor genes, such as *AaWRKY9*, *GLANDULAR TRICHOME-SPECIFIC WRKY 1* (*AaGSW1*), and *AaORA*, to upregulates their expression. These transcription factors then stimulate the expression of related biosynthetic genes and enhance the content of artemisinin (Hao et al., 2019; Fu et al., 2021b). HY5 can also affect the artemisinin biosynthesis by directly regulate the expression of biosynthetic genes in *A. annua* (Zhou et al., 2015). Phytochrome Interacting Factors (PIFs) also mediate the biosynthesis of SMs in many plant species. In *C. roseus*, under dark conditions, PIF1 binds to the promoters of *deacetylvindoline-4-O-acetyltransferase* (*DAT*) and GATA-type transcription factor gene *GATA1*, thus repressing the expression of target genes and decreasing the content of vindoline. Under red light, PIF1 is degraded, which stimulates the accumulation of vindoline (Liu et al., 2019). PIF3 upregulates the expression of *ADS*, *CYP71AV1*, *artemisinic aldehyde Δ11(13) reductase* (*DBR2*), and *aldehyde dehydrogenase 1* (*ALDH1*), and dramatically increases the content of artemisinin under light in *A. annua* (Zhang et al., 2019c). AaMYB15 is reported as a negative regulator of artemisinin, it binds to the promoter of *AaORA* and inhibits the biosynthesis of artemisinin under light in *A. annua* (Wu et al., 2021).

As an indispensable environmental factor, light affects the contents and yields of almost all kinds of phytochemicals in medicinal plants. The responses of plants to light are species-specific, and the effects of light on SM biosynthesis are distinct among different species. Moderate UV-B and blue light irradiation improve the accumulation of phytochemicals in many medicinal plants. Compared with UV-B and blue light, red light promotes the growth and development of some medicinal plants and therefore is more efficient in enhancing the yield of target compounds. In nature, light intensity and photoperiod often act together with other environmental factors, either synergistically or antagonistically, to regulate SM biosynthesis in medicinal plants. The regulatory mechanisms underlying the effects of light intensity and photoperiod on SM biosynthesis have not yet been fully elucidated.

With the increasing demand for natural bioactive compounds, many environmentally controlled systems with artificial light sources have been used for the cultivation of medicinal plants. Light conditions optimal for SM production in different plant species have been determined, and light-responsive genes involved in the corresponding biosynthetic pathways have been characterized. However, the intact light signaling pathways in almost all medicinal plants remain unclear. Besides, the regulatory roles of light in plants are complex. Understanding how light systematically regulates the SM content and growth of medicinal plants, simultaneously affects the content and yield of target compounds is a challenge that should be tackled in the future.

AUTHOR CONTRIBUTIONS

SZ, DY, and YW conceptualized and designed the work, collected, analyzed, and interpreted the data, and drafted the manuscript. LZ and HZ collected the data and contributed to critical revision of the manuscript. LQ, YZ, and HZ approved the final version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (31700257), Key project at central government level: The ability establishment of sustainable use for valuable Chinese medicine resources (2060302), the Zhejiang Provincial Natural Science Foundation of China (LR21H280002).

REFERENCES

Aerts, R. J., and De Luca, V. (1992). Phytochrome is involved in the light-regulation of vindoline biosynthesis in *Catharanthus*. *Plant Physiol.* 100, 1029–1032. doi: 10.1104/pp.100.2.1029

Akula, R., and Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* 6, 1720–1731. doi: 10.4161/psb.6.11.17613

Alrifai, O., Hao, X. M., Marcone, M. F., and Tsao, R. (2019). Current review of the modulatory effects of LED lights on photosynthesis of secondary metabolites and future perspectives of microgreen vegetables. *J. Agric. Food Chem.* 67, 6075–6090. doi: 10.1021/acs.jafc.9b00819

Anjum, S., Abbasi, B. H., Doussot, J., Favre-Reguillon, A., and Hano, C. (2017). Effects of photoperiod regimes and ultraviolet-C radiations on biosynthesis of industrially important lignans and neolignans in cell cultures of *Linum usitatissimum* L. (Flax). *J. Photochem. Photobiol. B* 167, 216–227. doi: 10.1016/j.jphotobiol.2017.01.006

Ansari, M. T., Saify, Z. S., Sultana, N., Ahmad, I., Saeed-Ul-Hassan, S., Tariq, I., et al. (2013). Malaria and artemisinin derivatives: an updated review. *Mini-Rev. Med. Chem.* 13, 1879–1902. doi: 10.2174/13895757113136660097

Askari-Khorasgani, O., and Pessarakli, M. (2019). Shifting saffron (*Crocus sativus* L.) culture from traditional farmland to controlled environment (greenhouse) condition to avoid the negative impact of climate changes and increase its productivity. *J. Plant Nutr.* 42, 2642–2665. doi: 10.1080/01904167.2019.1659348

Balasundram, N., Sundram, K., and Samman, S. (2006). Phenolic compounds in plants and agri-industrial by-products: antioxidant activity, occurrence, and potential uses. *Food Chem.* 99, 191–203. doi: 10.1016/j.foodchem.2005.07.042

Bergman, M. E., Davis, B., and Phillips, M. A. (2019). Medically useful plant terpenoids: biosynthesis, occurrence, and mechanism of action. *Molecules* 24:3961. doi: 10.3390/molecules24213961

Bhambhani, S., Kondhare, K. R., and Giri, A. P. (2021). Diversity in chemical structures and biological properties of plant alkaloids. *Molecules* 26:3374. doi: 10.3390/molecules26113374

Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998). Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4126–4133. doi: 10.1073/pnas.95.8.4126

Boonsnongcheep, P., Sae-foo, W., Banpakoat, K., Channarong, S., Chitsaithan, S., Uafua, P., et al. (2019). Artificial color light sources and precursor feeding enhance plumbagin production of the carnivorous plants *Drosera burmannii* and *Drosera indica*. *J. Photochem. Photobiol. B* 199:111628. doi: 10.1016/j.jphotobiol.2019.111628

Chappell, J. (1995). Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46, 521–547. doi: 10.1146/annurev.pp.46.060195.002513

Chen, Z. J. (2017). Effect of red and blue light quality on the content of essential oil of *Melissa officinalis* L. master thesis. Guangzhou: South China Agricultural University.

Chen, I.-G. J., Lee, M.-S., Lin, M.-K., Ko, C.-Y., and Chang, W.-T. (2018). Blue light decreases tanshinone IIA content in *Salvia miltiorrhiza* hairy roots via genes regulation. *J. Photochem. Photobiol. B* 183, 164–171. doi: 10.1016/j.jphotobiol.2018.04.013

Chen, D. Q., Li, Z. Y., Pan, R. C., and Wang, X. J. (2006). Anthocyanin accumulation mediated by blue light and cytokinin in *Arabidopsis* seedlings. *J. Integr. Plant Biol.* 48, 420–425. doi: 10.1111/j.1744-7909.2006.00234.x

Chen, C. L., Luo, X. H., Jin, G. R., Cheng, Z., Pan, X. Y., Zhu, G. L., et al. (2017). Shading effect on survival, growth, and contents of secondary metabolites in micropropagated *Anoectochilus* plantlets. *Braz. J. Bot.* 40, 599–607. doi: 10.1007/s40415-017-0365-4

Chiocchio, I., Mandrone, M., Tomasi, P., Marincich, L., and Poli, F. (2021). Plant secondary metabolites: an opportunity for circular economy. *Molecules* 26:495. doi: 10.3390/molecules26020495

Christie, J. M., Arvai, A. S., Baxter, K. J., Heilmann, M., Pratt, A. J., O'Hara, A., et al. (2012). Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. *Science* 335, 1492–1496. doi: 10.1126/science.1218091

Constantino, L. F. D., Nascimento, L. B. D., Casanova, L. M., Moreira, N. D., Menezes, E. A., Esteves, R. L., et al. (2017). Responses of *Crepis japonica* induced by supplemental blue light and UV-A radiation. *Photochem. Photobiol. Sci.* 16, 238–245. doi: 10.1039/C6PP00343E

Craker, L. E., and Gardner, Z. E. (2006). "Medicinal plants and tomorrow's pharmacy" in *Medicinal and Aromatic Plants: Agricultural, Commercial, Ecological, Legal, Pharmacological and Social Aspects*. eds. R. J. Bogers, L. E. Craker and D. Lange (New York, NY: Springer), 29–41.

da Silva, W. M. C., Vilar, D. J., Souza, D. D., Walter, M., de Holanda, M. T., and Brigido, M. D. (2017). A terpenoid metabolic network modelled as graph database. *Int. J. Data Min. Bioinf.* 18, 74–90. doi: 10.1504/IJDMB.2017.086103

de Carvalho, A. A., Bertolucci, S. K. V., Honorato, A. D., Rocha, T. T., Silva, S. T., and Pinto, J. (2020). Influence of light spectra and elicitors on growth and ascaridole content using in vitro cultures of *Dysphania ambrosioides* L. *Plant Cell Tissue Organ Cult.* 143, 277–290. doi: 10.1007/s11240-020-01892-5

de Castro, K. M., Batista, D. S., Fortini, E. A., Silva, T. D., Felipe, S. H. S., Fernandes, A. M., et al. (2019). Photoperiod modulates growth, morphoanatomy, and linalool content in *Lippia alba* L. (Verbenaceae) cultured in vitro. *Plant Cell Tissue Organ Cult.* 139, 139–153. doi: 10.1007/s11240-019-01672-w

Demain, A. L., and Fang, A. (2000). The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* 69, 1–39. doi: 10.1007/3-540-44964-7_1

Dixon, R. A., and Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* 7, 1085–1097. doi: 10.2307/3870059

Dou, H. J., Niu, G. H., Gu, M. M., and Masabni, J. G. (2017). Effects of light quality on growth and phytonutrient accumulation of herbs under controlled environments. *Horticulturae* 3:36. doi: 10.3390/horticulturae3020036

Fang, H. L., Qi, X. W., Li, Y. M., Yu, X., Xu, D. B., Liang, C. Y., et al. (2020). De novo transcriptomic analysis of light-induced flavonoid pathway, transcription factors in the flower buds of *Lonicera japonica*. *Trees* 34, 267–283. doi: 10.1007/s00468-019-01916-4

Fazal, H., Abbasi, B. H., Ahmad, N., Ali, S. S., Akbar, F., and Kanwal, F. (2016a). Correlation of different spectral lights with biomass accumulation and production of antioxidant secondary metabolites in callus cultures of medicinally important *Prunella vulgaris* L. *J. Photochem. Photobiol. B* 159, 1–7. doi: 10.1016/j.jphotobiol.2016.03.008

Fazal, H., Abbasi, B. H., Ahmad, N., Ali, M., and Ali, S. (2016b). Sucrose induced osmotic stress and photoperiod regimes enhanced the biomass and production of antioxidant secondary metabolites in shake-flask suspension cultures of *Prunella vulgaris* L. *Plant Cell Tissue Organ Cult.* 124, 573–581. doi: 10.1007/s11240-015-0915-z

Fonseca, J. M., Rushing, J. W., Rajapakse, N. C., Thomas, R. L., and Riley, M. B. (2006). Potential implications of medicinal plant production in controlled environments: the case of feverfew (*Tanacetum parthenium*). *HortScience* 41, 531–535. doi: 10.21273/HORTSCI.41.3.531

Fraenkel, G. S. (1959). The raison d'ètre of secondary plant substances. *Science* 129, 1466–1470. doi: 10.1126/science.129.3361.1466

Fu, X. Q., He, Y. L., Li, L., Zhao, L. M., Wang, Y. T., Qian, H. M., et al. (2021a). Overexpression of blue light receptor *AaCRY1* improves artemisinin content in *Artemisia annua* L. *Biotechnol. Appl. Biochem.* 68, 338–344. doi: 10.1002/bab.1931

Fu, X. Q., Peng, B. W., Hassani, D., Xie, L. H., Liu, H., Li, Y. P., et al. (2021b). AaWRKY9 contributes to light- and jasmonate-mediated to regulate the biosynthesis of artemisinin in *Artemisia annua*. *New Phytol.* 231, 1858–1874. doi: 10.1111/nph.17453

Gangappa, S. N., and Botto, J. F. (2016). The multifaceted roles of HY5 in plant growth and development. *Mol. Plant* 9, 1353–1365. doi: 10.1016/j.molp.2016.07.002

Gao, C. X., Yang, B. X., Zhang, D. D., Chen, M., and Tian, J. K. (2016). Enhanced metabolic process to indole alkaloids in *Clematis terniflora* DC after exposure to high level of UV-B irradiation followed by the dark. *BMC Plant Biol.* 16:231. doi: 10.1186/s12870-016-0920-3

Ghaffari, Z., Rahimmalek, M., and Sabzalian, M. R. (2019). Variation in the primary and secondary metabolites derived from the isoprenoid pathway in the *Perovskia* species in response to different wavelengths generated by light emitting diodes (LEDs). *Ind. Crop. Prod.* 140:111592. doi: 10.1016/j.indcrop.2019.111592

Gyula, N., Schafer, E., and Nagy, F. (2003). Light perception and signalling in higher plants. *Curr. Opin. Plant Biol.* 6, 446–452. doi: 10.1016/S1369-5266(03)00082-7

Hamid, A., Singh, S., Agrawal, M., and Agrawal, S. B. (2019). *Heteropogon contortus* BL-1 (Pilli grass) and elevated UV-B radiation: the role of

growth, physiological, and biochemical traits in determining forage productivity and quality. *Photochem. Photobiol.* 95, 572–580. doi: 10.1111/php.12990

Hao, X. Y., Li, L. T., Hu, Y. R., Zhou, C., Wang, X. C., Wang, L., et al. (2016). Transcriptomic analysis of the effects of three different light treatments on the biosynthesis of characteristic compounds in the tea plant by RNA-Seq. *Tree Genet. Genomes* 12, 1–10. doi: 10.1007/s11295-016-1071-2

Hao, X. L., Zhong, Y. J., Nutzmann, H.-W., Fu, X. Q., Yan, T. X., Shen, Q., et al. (2019). Light-induced artemisinin biosynthesis is regulated by the bZIP transcription factor AaHY5 in *Artemisia annual*. *Plant Cell Physiol.* 60, 1747–1760. doi: 10.1093/pcp/pcz084

Hemm, M. R., Rider, S. D., Ogas, J., Murry, D. J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in *Arabidopsis* roots. *Plant J.* 38, 765–778. doi: 10.1111/j.1365-313X.2004.02089.x

Herrmann, K. M. (1995). The shikimate pathway: early steps in the biosynthesis of aromatic compounds. *Plant Cell* 7, 907–919. doi: 10.2307/3870046

Hoft, M., Verpoorte, R., and Beck, E. (1996). Growth and alkaloid contents in leaves of *Tabernaemontana pachysiphon* Stapf (Apocynaceae) as influenced by light intensity, water and nutrient supply. *Oecologia* 107, 160–169. doi: 10.1007/BF00327899

Huang, L. Q., and Liu, C. X. (2015). "Biosynthesis and regulation of secondary metabolites in medicinal plants" in *Molecular Pharmacognosy*. eds. Y. Bao and L. Y. Cao. 3rd ed (Beijing: Science Press), 251–252.

Huang, Y. X., Tan, H. X., Guo, Z. Y., Wu, X. X., Zhang, Q. L., Zhang, L., et al. (2016). The biosynthesis and genetic engineering of bioactive indole alkaloids in plants. *J. Plant Biol.* 59, 203–214. doi: 10.1007/s12374-016-0032-5

Jain, C., Khatana, S., and Vijayvergia, R. (2019). Bioactive of secondary metabolites of various plants: a review. *Int. J. Pharm. Sci. Res.* 10, 494–504. doi: 10.13040/ijpsr.0975-8232.10(2).494-04

Jung, W.-S., Chung, I.-M., Hwang, M. H., Kim, S.-H., Yu, C. Y., and Ghimire, B. K. (2021). Application of light-emitting diodes for improving the nutritional quality and bioactive compound levels of some crops and medicinal plants. *Molecules* 26:1477. doi: 10.3390/molecules26051477

Jung, J. H., Kim, H. Y., Kim, H. S., and Jung, S. H. (2020). Transcriptome analysis of *Panax ginseng* response to high light stress. *J. Ginseng Res.* 44, 312–320. doi: 10.1016/j.jgr.2018.12.009

Kapoor, S., Raghuvanshi, R., Bhardwaj, P., Sood, H., Saxena, S., and Chaurasia, O. P. (2018). Influence of light quality on growth, secondary metabolites production and antioxidant activity in callus culture of *Rhodiola imbricata* Edgew. *J. Photochem. Photobiol. B* 183, 258–265. doi: 10.1016/j.jphotobiol.2018.04.018

Kawoosa, T., Singh, H., Kumar, A., Sharma, S. K., Devi, K., Dutt, S., et al. (2010). Light and temperature regulated terpene biosynthesis: hepatoprotective monoterpene picroside accumulation in *Picrorhiza kurrooa*. *Funct. Integr. Genomics* 10, 393–404. doi: 10.1007/s10142-009-0152-9

Khurshid, R., Ullah, M. A., Tungmunnithum, D., Drouet, S., Shah, M., Zaeem, A., et al. (2020). Lights triggered differential accumulation of antioxidant and antidiabetic secondary metabolites in callus culture of *Eclipta alba* L. *PLoS One* 15:e0233963. doi: 10.1371/journal.pone.0233963

King, A., and Young, G. (1999). Characteristics and occurrence of phenolic phytochemicals. *J. Am. Diet. Assoc.* 99, 213–218. doi: 10.1016/s0002-8223(99)00051-6

Kubica, P., Szopa, A., Prokopiuk, B., Komsta, L., Pawlowska, B., and Ekiert, H. (2020). The influence of light quality on the production of bioactive metabolites - verbascoside, isoverbascoside and phenolic acids and the content of photosynthetic pigments in biomass of *Verbena officinalis* L. cultured in vitro. *J. Photochem. Photobiol. B* 203:111768. doi: 10.1016/j.jphotobiol.2019.111768

Kumar, S. S., Arya, M., Mahadevappa, P., and Giridhar, P. (2020). Influence of photoperiod on growth, bioactive compounds and antioxidant activity in callus cultures of *Basella rubra* L. *J. Photochem. Photobiol. B* 209:111937. doi: 10.1016/j.jphotobiol.2020.111937

Kumari, R., and Agrawal, S. B. (2010). Supplemental UV-B induced changes in leaf morphology, physiology and secondary metabolites of an Indian aromatic plant *Cymbopogon citratus* (D.C.) Staph under natural field conditions. *Int. J. Environ. Stud.* 67, 655–675. doi: 10.1080/00207233.2010.513828

Kumari, R., and Prasad, M. N. V. (2013). "Medicinal plant active compounds produced by UV-B exposure" in *Sustainable Agriculture Reviews*. ed. E. Lichtfouse (Dordrecht: Springer-Verlag), 225–254.

Kuo, T. C. Y., Chen, C. H., Chen, S. H., Lu, I. H., Chu, M. J., Huang, L. C., et al. (2015). The effect of red light and far-red light conditions on secondary metabolism in Agarwood. *BMC Plant Biol.* 15:139. doi: 10.1186/s12870-015-0537-y

Lee, M., Son, J. E., and Oh, M. (2013). Growth and phenolic content of sowthistle grown in a closed-type plant production system with a UV-A or UV-B lamp. *Hortic. Environ. Biotechnol.* 54, 492–500. doi: 10.1007/s13580-013-0097-8

Li, L. (2014). The influence of elicitor on the accumulation of several metabolites in the seedling *Salvia przewalskii* Maxim. master thesis. Ya'an: Sichuan Agricultural University.

Li, J., Han, X., Wang, C., Tang, L., Zhang, W., and Qi, W. (2019). The response of *Achyranthes bidentata* Blume to short-term UV-B exposure. *Russ. J. Plant Physiol.* 66, 160–170. doi: 10.1134/S1021443719010096

Li, T., Jia, K. P., Lian, H. L., Yang, X., Li, L., and Yang, H. Q. (2014). Jasmonic acid enhancement of anthocyanin accumulation is dependent on phytochrome A signaling pathway under far-red light in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 454, 78–83. doi: 10.1016/j.bbrc.2014.10.059

Li, Y. Q., Kong, D. X., Fu, Y., Sussman, M. R., and Wu, H. (2020). The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol. Biochem.* 148, 80–89. doi: 10.1016/j.plaphy.2020.01.006

Li, Y. Q., Kong, D. X., Liang, H. L., and Wu, H. (2018a). Alkaloid content and essential oil composition of *Mahonia breviracema* cultivated under different light environments. *J. Appl. Bot. Food Qual.* 91, 171–179. doi: 10.5073/jabfq.2018.091.023

Li, H. S., Lin, Y. L., Chen, X. H., Bai, Y., Wang, C. Q., Xu, X. P., et al. (2018b). Effects of blue light on flavonoid accumulation linked to the expression of miR393, miR394 and miR395 in longan embryogenic calli. *PLoS One* 13:e0191444. doi: 10.1371/journal.pone.0191444

Li, Q. Z., Xu, J. X., Yang, L. Y., Sun, Y., Zhou, X. H., Zheng, Y. H., et al. (2021). LED light quality affect growth, alkaloids contents, and expressions of amaryllidaceae alkaloids biosynthetic pathway genes in *Lycoris longituba*. *J. Plant Growth Regul.* doi: 10.1007/s00344-021-10298-2, [Epub ahead of print].

Liebelt, D. J., Jordan, J. T., and Doherty, C. J. (2019). Only a matter of time: the impact of daily and seasonal rhythms on phytochemicals. *Phytochem. Rev.* 18, 1409–1433. doi: 10.1007/s11101-019-09617-z

Liu, Y., Fang, S. Z., Yang, W. X., Shang, X. L., and Fu, X. X. (2018a). Light quality affects flavonoid production and related gene expression in *Cyclocarya paliurus*. *J. Photochem. Photobiol. B* 179, 66–73. doi: 10.1016/j.jphotobiol.2018.01.002

Liu, L. L., Li, Y. Y., She, G. B., Zhang, X. C., Jordan, B., Chen, Q., et al. (2018b). Metabolite profiling and transcriptomic analyses reveal an essential role of UVR8-mediated signal transduction pathway in regulating flavonoid biosynthesis in tea plants (*Camellia sinensis*) in response to shading. *BMC Plant Biol.* 18:233. doi: 10.1186/s12870-018-1440-0

Liu, Y. L., Patra, B., Pattanaik, S., Wang, Y., and Yuan, L. (2019). GATA and phytochrome interacting factor transcription factors regulate light-induced vindoline biosynthesis in *Catharanthus roseus*. *Plant Physiol.* 180, 1336–1350. doi: 10.1104/pp.19.00489

Liu, Y., Song, L. L., Yu, W. W., Hu, Y. Y., Ma, X. H., Wu, J. S., et al. (2015). Light quality modifies camptothecin production and gene expression of biosynthesis in *Camptotheca acuminata* Decne seedlings. *Ind. Crop. Prod.* 66, 137–143. doi: 10.1016/j.indcrop.2014.12.046

Lobiuc, A., Vasilache, V., Pintilie, O., Stoleru, T., Burducea, M., Oroian, M., et al. (2017). Blue and red LED illumination improves growth and bioactive compounds contents in acyanic and cyanic *Ocimum basilicum* L. microgreens. *Molecules* 22:2111. doi: 10.3390/molecules22122111

Loi, M., Villani, A., Paciolla, F., Mule, G., and Paciolla, C. (2021). Challenges and opportunities of light-emitting diode (LED) as key to modulate antioxidant compounds in plants. A review. *Antioxidants* 10:42. doi: 10.3390/antiox10010042

Luis, J. C., Perez, R. M., and Gonzalez, F. V. (2007). UV-B radiation effects on foliar concentrations of rosmarinic and carnosic acids in rosemary plants. *Food Chem.* 101, 1211–1215. doi: 10.1016/j.foodchem.2006.03.023

Lv, M., Su, H. Y., Li, M. L., Yang, D. L., Yao, R. Y., Li, M. F., et al. (2021). Effect of UV-B radiation on growth, flavonoid and podophyllotoxin accumulation, and related gene expression in *Sinopodophyllum hexandrum*. *Plant Biol.* 23, 202–209. doi: 10.1111/plb.13226

Malik, M., Velechovsky, J., and Tlustos, P. (2021). The overview of existing knowledge on medical cannabis plants growing. *Plant Soil Environ.* 67, 425–442. doi: 10.17221/96/2021-PSE

Mathur, R., and Velpandian, T. (2009). Medicinal plant-based health products: where is the medicinal constituent? *Indian J. Pharmacol.* 41, 205–206. doi: 10.4103/0253-7613.56063

Matsuura, H. N., Fragoso, V., Paranhos, J. T., Rau, M. R., and Fett-Neto, A. G. (2016). The bioactive monoterpene indole alkaloid N,β-D-glucopyranosyl vincosamide is regulated by irradiance quality and development in *Psychotria leiocarpa*. *Ind. Crop. Prod.* 86, 210–218. doi: 10.1016/j.indcrop.2016.03.050

Mohd Yusof, F. F., Yaacob, J. S., Osman, N., Ibrahim, M. H., Wan-Mohtar, W. A. A. Q. I., Berahim, Z., et al. (2021). Shading effects on leaf gas exchange, leaf pigments and secondary metabolites of *Polygonum minus* Huds., an aromatic medicinal herb. *Plan. Theory* 10:608. doi: 10.3390/plants10030608

Moon, K.-B., Park, J.-S., Park, Y.-I., Song, I.-J., Lee, H.-J., Cho, H. S., et al. (2020). Development of systems for the production of plant-derived biopharmaceuticals. *Plan. Theory* 9:30. doi: 10.3390/plants9010030

Moyo, M., Koetle, M. J., and Van Staden, J. (2014). Photoperiod and plant growth regulator combinations influence growth and physiological responses in *Pelargonium sidoides* DC. *In Vitro Cell. Dev. Pl.* 50, 487–492. doi: 10.1007/s11627-014-9594-4

Nam, T. G., Kim, D.-O., and Eom, S. H. (2018). Effects of light sources on major flavonoids and antioxidant activity in common buckwheat sprouts. *Food Sci. Biotechnol.* 27, 169–176. doi: 10.1007/s10068-017-0204-1

Nazir, M., Ullah, M. A., Younas, M., Siddiquah, A., Shah, M., Giglioli-Guivarc'h, N., et al. (2020). Light-mediated biosynthesis of phenylpropanoid metabolites and antioxidant potential in callus cultures of purple basil (*Ocimum basilicum* L. var purpurascens). *Plant Cell Tissue Organ Cult.* 142, 107–120. doi: 10.1007/s11240-020-01844-z

Newman, J. D., and Chappell, J. (1999). Isoprenoid biosynthesis in plants: Carbon partitioning within the cytoplasmic pathway. *Crit. Rev. Biochem. Mol. Biol.* 34, 95–106. doi: 10.1080/10409239991209228

Nguyen, T. K. O., Jamali, A., Lanoue, A., Gontier, E., and Dauwe, R. (2015). Unravelling the architecture and dynamics of tropane alkaloid biosynthesis pathways using metabolite correlation networks. *Phytochemistry* 116, 94–103. doi: 10.1016/j.phytochem.2015.03.005

Ohashi, K. K., Fukuyama, T., Nakai, A., Usami, H., Ono, E., and Watanabe, H. (2013). Growth and alkaloids production in *Madagascar periwinkle* plants grown under red LED. *IFAC Proc. Vol.* 46, 274–277. doi: 10.3182/20130327-3-JP-3017.00063

Palmitessa, O. D., Pantaleo, M. A., and Santamaria, P. (2021). Applications and development of LEDs as supplementary lighting for tomato at different latitudes. *Agronomy* 11:835. doi: 10.3390/agronomy11050835

Pan, W. S., Zheng, L. P., Tian, H., Li, W. Y., and Wang, J. W. (2014). Transcriptome responses involved in artemisinin production in *Artemisia annua* L. under UV-B radiation. *J. Photochem. Photobiol. B* 140, 292–300. doi: 10.1016/j.jphotobiol.2014.08.013

Pandey, N., and Pandey-Rai, S. (2014). Short term UV-B radiation-mediated transcriptional responses and altered secondary metabolism of in vitro propagated plantlets of *Artemisia annua* L. *Plant Cell Tissue Organ Cult.* 116, 371–385. doi: 10.1007/s11240-013-0413-0

Peebles, C. A. M., Shanks, J. V., and San, K.-Y. (2009). The role of the octadecanoid pathway in the production of terpenoid indole alkaloids in *Catharanthus roseus* hairy roots under normal and UV-B stress conditions. *Biotechnol. Bioeng.* 103, 1248–1254. doi: 10.1002/bit.22350

Qin, B. F., Ma, L. L., Wang, Y. X., Chen, M., Lan, X. Z., Wu, N. B., et al. (2014). Effects of acetylsalicylic acid and UV-B on gene expression and tropane alkaloid biosynthesis in hairy root cultures of *Anisodus luridus*. *Plant Cell Tissue Organ Cult.* 117, 483–490. doi: 10.1007/s11240-014-0454-z

Qiu, F., Zeng, J. L., Wang, J., Huang, J. P., Zhou, W., Yang, C. X., et al. (2020). Functional genomics analysis reveals two novel genes required for littorine biosynthesis. *New Phytol.* 225, 1906–1914. doi: 10.1111/nph.16317

Radwanski, E. R., Barczak, A. J., and Last, R. L. (1996). Characterization of tryptophan synthase alpha subunit mutants of *Arabidopsis thaliana*. *Mol. Gen. Genet.* 253, 353–361. doi: 10.1007/pl00008602

Rai, R., Meena, R. P., Smita, S. S., Shukla, A., Rai, S. K., and Pandey-Rai, S. (2011). UV-B and UV-C pre-treatments induce physiological changes and artemisinin biosynthesis in *Artemisia annua* L. - an antimalarial plant. *J. Photochem. Photobiol. B* 105, 216–225. doi: 10.1016/j.jphotobiol.2011.09.004

Ramani, S., and Jayabaskaran, C. (2008). Enhanced catharanthine and vindoline production in suspension cultures of *Catharanthus roseus* by ultraviolet-B light. *J. Mol. Signal.* 3:9. doi: 10.1186/1750-2187-3-9

Rizi, M. R., Azizi, A., Sayyari, M., Mirzaie-Asl, A., and Conti, L. (2021). Increased phenylpropanoids production in UV-B irradiated *Salvia verticillata* as a consequence of altered genes expression in young leaves. *Plant Physiol. Biochem.* 167, 174–184. doi: 10.1016/j.plaphy.2021.07.037

Rizzini, L., Favory, J. J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., et al. (2011). Perception of UV-B by the *Arabidopsis* UVR8 protein. *Science* 332, 103–106. doi: 10.1126/science.1200660

Rukh, G., Ahmad, N., Rab, A., Ahmad, N., Fazal, H., Akbar, F., et al. (2019). Photo-dependent somatic embryogenesis from non-embryogenic calli and its polyphenolics content in high-valued medicinal plant of *Ajuga bracteosa*. *J. Photochem. Photobiol. B* 190, 59–65. doi: 10.1016/j.jphotobiol.2018.11.012

Sanchez, S. E., Rugnone, M. L., and Kay, S. A. (2020). Light perception: a matter of time. *Mol. Plant* 13, 363–385. doi: 10.1016/j.molp.2020.02.006

Santos Lazzarini, L. E., Vilela Bertolucci, S. K., Pacheco, F. V., dos Santos, J., Silva, S. T., de Carvalho, A. A., et al. (2018). Quality and intensity of light affect *Lippia gracilis* Schauer plant growth and volatile compounds in vitro. *Plant Cell Tissue Organ Cult.* 135, 367–379. doi: 10.1007/s11240-018-1470-1

Saravanan, R., Krishti, S., Gajbhiye, N. A., and Maiti, S. (2008). Influence of light intensity on gas exchange, herbage yield and andrographolide content in *Andrographis paniculata* (Nees.). *Indian J. Hortic.* 65, 220–225. doi: 10.1007/s10658-007-9256-z

Schmidt, B. M. (2012). Responsible use of medicinal plants for cosmetics. *HortScience* 47, 985–991. doi: 10.21273/HORTSCI.47.8.985

Schreiner, M., Krumbein, A., Mewis, I., Ulrichs, C., and Huyskens-Keil, S. (2009). Short-term and moderate UV-B radiation effects on secondary plant metabolism in different organs of nasturtium (*Tropaeolum majus* L.). *Innov. Food Sci. Emerg. Technol.* 10, 93–96. doi: 10.1016/j.ifset.2008.10.001

Schreiner, M., Mewis, I., Huyskens-Keil, S., Jansen, M. A. K., Zrenner, R., Winkler, J. B., et al. (2012). UV-B-induced secondary plant metabolites - potential benefits for plant and human health. *Crit. Rev. Plant Sci.* 31, 229–240. doi: 10.1080/07352689.2012.664979

Sharrock, R. A., and Quail, P. H. (1989). Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes Dev.* 3, 1745–1757. doi: 10.1101/gad.3.11.1745

Silva, T. D., Batista, D. S., Fortini, E. A., de Castro, K. M., Sousa Felipe, S. H., Fernandes, A. M., et al. (2020). Blue and red light affects morphogenesis and 20-hydroxyecdisone content of in vitro *Pfaffia glomerata* accessions. *J. Photochem. Photobiol. B* 203:111761. doi: 10.1016/j.jphotobiol.2019.111761

Somers, D. E., Schultz, T. F., Milnamow, M., and Kay, S. A. (2000). ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* 101, 319–329. doi: 10.1016/S0092-8674(00)80841-7

Stepanova, A. Y., Solov'yova, A. I., and Salamaikina, S. A. (2020). Influence of spectral light composition on flavones formation in callus culture of *Scutellaria baicalensis* Georgi. *Pharmacogn. Mag.* 16, 156–160. doi: 10.4103/pm.pm_280_19

Sullivan, J. A., and Deng, X. W. (2003). From seed to seed: the role of photoreceptors in *Arabidopsis* development. *Dev. Biol.* 260, 289–297. doi: 10.1016/S0012-1606(03)00212-4

Sytar, O., Zivcak, M., Neugart, S., Toutounchi, P. M., and Brestic, M. (2019). Precultivation of young seedlings under different color shades modifies the accumulation of phenolic compounds in *Cichorium* leaves in later growth phases. *Environ. Exp. Bot.* 165, 30–38. doi: 10.1016/j.envexpbot.2019.05.018

Takshak, S., and Agrawal, S. B. (2014). Secondary metabolites and phenylpropanoid pathway enzymes as influenced under supplemental ultraviolet-B radiation in *Withania somnifera* Dunal, an indigenous medicinal plant. *J. Photochem. Photobiol. B* 140, 332–343. doi: 10.1016/j.jphotobiol.2014.08.011

Takshak, S., and Agrawal, S. B. (2015). Defence strategies adopted by the medicinal plant *Coleus forskohlii* against supplemental ultraviolet-B radiation: augmentation of secondary metabolites and antioxidants. *Plant Physiol. Biochem.* 97, 124–138. doi: 10.1016/j.plaphy.2015.09.018

Takshak, S., and Agrawal, S. B. (2019). Defense potential of secondary metabolites in medicinal plants under UV-B stress. *J. Photochem. Photobiol. B* 193, 51–88. doi: 10.1016/j.jphotobiol.2019.02.002

Tsurunaga, Y., Takahashi, T., Katsube, T., Kudo, A., Kuramitsu, O., Ishiwata, M., et al. (2013). Effects of UV-B irradiation on the levels of anthocyanin, rutin and radical scavenging activity of buckwheat sprouts. *Food Chem.* 141, 552–556. doi: 10.1016/j.foodchem.2013.03.032

Tusevski, O., Stanoeva, J. P., Stefova, M., and Simic, S. G. (2013). Phenolic profile of dark-grown and photoperiod-exposed *Hypericum perforatum* L. hairy root cultures. *Sci. World J.* 2013:602752. doi: 10.1155/2013/602752

Ueda, T., Murata, M., and Yokawa, K. (2021). Single wavelengths of LED light supplement promote the biosynthesis of major cyclic monoterpenes in Japanese mint. *Plan. Theory* 10:1420. doi: 10.3390/plants10071420

Verma, N., and Shukla, S. (2015). Impact of various factors responsible for fluctuation in plant secondary metabolites. *J. Appl. Res. Med. Aromat. Plants* 2, 105–113. doi: 10.1016/j.jarmap.2015.09.002

Wang, C.-L., Guo, Q.-S., Zhu, Z.-B., and Cheng, B.-X. (2017). Physiological characteristics, dry matter, and active component accumulation patterns of *Changium smyrnioides* in response to a light intensity gradient. *Pharm. Biol.* 55, 581–589. doi: 10.1080/13880209.2016.1263345

Wang, T., Tian, X. R., Wu, X. Y., Luo, Z., Li, G., Peng, X. L., et al. (2018). Effect of light quality on total gypenosides accumulation and related key enzyme gene expression in *Gynostemma pentaphyllum*. *Chin. Herb. Med.* 10, 34–39. doi: 10.1016/j.chmed.2017.12.004

Wang, W. T., Zhang, J. W., Di, W., Tao, S. H., and Wu, B. (2010). Relation between light qualities and accumulation of steroidal glycoalkaloids as well as signal molecule in cell in potato tubers. *Acta Agron. Sin.* 36, 629–635. doi: 10.3724/SP.J.1006.2010.00629

Wang, C. H., Zheng, L. P., Tian, H., and Wang, J. W. (2016). Synergistic effects of ultraviolet-B and methyl jasmonate on tanshinone biosynthesis in *Salvia miltiorrhiza* hairy roots. *J. Photochem. Photobiol. B* 159, 93–100. doi: 10.1016/j.jphotobiol.2016.01.012

Wei, X. Y., Zhao, X. L., Long, S. H., Xiao, Q. M., Guo, Y. A., Qiu, C. S., et al. (2021). Wavelengths of LED light affect the growth and cannabidiol content in *Cannabis sativa* L. *Ind. Crop. Prod.* 165:113433. doi: 10.1016/j.indcrop.2021.113433

Weremczuk-Jezyna, I., Hnatuszko-Konka, K., Lebelt, L., and Grzegorczyk-Karolak, I. (2021). The protective function and modification of secondary metabolite accumulation in response to light stress in *Dracocephalum forrestii* shoots. *Int. J. Mol. Sci.* 22. doi: 10.3390/ijms22157965

Wu, Z., Li, L., Liu, H., Yan, X., Ma, Y. N., Li, Y. P., et al. (2021). AaMYB15, an R2R3-MYB TF in *Artemisia annua*, acts as a negative regulator of artemisinin biosynthesis. *Plant Sci.* 308:110920. doi: 10.1016/j.plantsci.2021.110920

Wu, C. H., Murthy, H. N., Hahn, E. J., and Paek, K. Y. (2007). Enhanced production of caftaric acid, chlorogenic acid and cichoric acid in suspension cultures of *Echinacea purpurea* by the manipulation of incubation temperature and photoperiod. *Biochem. Eng. J.* 36, 301–303. doi: 10.1016/j.bej.2007.02.024

Wulff, A., Anttonen, S., Pellinen, R., Savonen, E. M., Sutinen, M. L., Heller, W., et al. (1999). Birch (*Betula pendula* Roth.) responses to high UV-B radiation. *Boreal Environ. Res.* 4, 77–88.

Xie, D. J., Tarin, M. W. K., Chen, L. Y., Ren, K., Yang, D. M., Zhou, C. C., et al. (2021). Consequences of LED lights on root morphological traits and compounds accumulation in *Sarcandra glabra* seedlings. *Int. J. Mol. Sci.* 22:7179. doi: 10.3390/ijms22137179

Xu, Y., Wang, G. B., Cao, F. L., Zhu, C. C., Wang, G. Y., and El-Kassaby, Y. A. (2014). Light intensity affects the growth and flavonol biosynthesis of Ginkgo (*Ginkgo biloba* L.). *New For.* 45, 765–776. doi: 10.1007/s11056-014-9435-7

Xu, M. Y., Wu, K. X., Liu, Y., Liu, J., and Tang, Z. H. (2020). Effects of light intensity on the growth, photosynthetic characteristics, and secondary metabolites of *Eleutherococcus senticosus* Harms. *Photosynthetica* 58, 881–889. doi: 10.32615/ps.2020.045

Yang, Y. J., Yang, X. L., Jang, Z. F., Chen, Z. H., Ruo, X. J., Jin, W. Y., et al. (2018). UV RESISTANCE LOCUS 8 from *Chrysanthemum morifolium* Ramat (CmUVR8) plays important roles in UV-B signal transduction and UV-B-induced accumulation of flavonoids. *Front. Plant Sci.* 9:955. doi: 10.3389/fpls.2018.00955

Yin, L. L., Zhao, C., Huang, Y., Yang, R. Y., and Zeng, Q. P. (2008). Abiotic stress-induced expression of artemisinin biosynthesis genes in *Artemisia annua* L. *Chin. J. Appl. Environ. Biol.* 14, 1–5. doi: 10.3724/SP.J.1035.2008.00091

Yoshida, H., Nishikawa, T., Hikosaka, S., and Goto, E. (2021). Effects of nocturnal UV-B irradiation on growth, flowering, and phytochemical concentration in leaves of greenhouse-grown red perilla. *Plants-Basel* 10. doi: 10.3390/plants10061252

Zahir, A., Abbasi, B. H., Adil, M., Anjum, S., Zia, M., and Ihsan-ul, H. (2014). Synergistic effects of drought stress and photoperiods on phenology and secondary metabolism of *Silybum marianum*. *Appl. Biochem. Biotechnol.* 174, 693–707. doi: 10.1007/s12010-014-1098-5

Zhang, W. J., and Bjorn, L. O. (2009). The effect of ultraviolet radiation on the accumulation of medicinal compounds in plants. *Fitoterapia* 80, 207–218. doi: 10.1016/j.fitote.2009.02.006

Zhang, X. R., Chen, Y. H., Guo, Q. S., Wang, W. M., Liu, L., Fan, J., et al. (2017). Short-term UV-B radiation effects on morphology, physiological traits and accumulation of bioactive compounds in *Prunella vulgaris* L. *J. Plant Interact.* 12, 348–354. doi: 10.1080/17429145.2017.1365179

Zhang, X., Ding, X., Ji, Y., Wang, S., Chen, Y., Luo, J., et al. (2018a). Measurement of metabolite variations and analysis of related gene expression in Chinese liquorice (*Glycyrrhiza uralensis*) plants under UV-B irradiation. *Sci. Rep.* 8:6144. doi: 10.1038/s41598-018-24284-4

Zhang, L. X., Guo, Q. S., Chang, Q. S., Zhu, Z. B., Liu, L., and Chen, Y. H. (2015). Chloroplast ultrastructure, photosynthesis and accumulation of secondary metabolites in *Glechoma longituba* in response to irradiance. *Photosynthetica* 53, 144–153. doi: 10.1007/s11099-015-0092-7

Zhang, D., Jiang, C. L., Huang, C. H., Wen, D., Lu, J. N., Chen, S., et al. (2019a). The light-induced transcription factor FtMYB116 promotes accumulation of rutin in *Fagopyrum tataricum*. *Plant Cell Environ.* 42, 1340–1351. doi: 10.1111/pce.13470

Zhang, Y., Lin, L., Long, Y., Guo, H., Wang, Z., Cui, M., et al. (2019b). Comprehensive transcriptome analysis revealed the effects of the light quality, light intensity, and photoperiod on phlorizin accumulation in *Lithocarpus polystachyus* Rehd. *Forests* 10:995. doi: 10.3390/f10110995

Zhang, S. C., Ma, J. Q., Zou, H. Y., Zhang, L., Li, S. H., and Wang, Y. P. (2020). The combination of blue and red LED light improves growth and phenolic acid contents in *Salvia miltiorrhiza* Bunge. *Ind. Crop. Prod.* 158:112959. doi: 10.1016/j.indcrop.2020.112959

Zhang, J., Morris-Natschke, S. L., Ma, D., Shang, X.-F., Yang, C.-J., Liu, Y.-Q., et al. (2021). Biologically active indolizidine alkaloids. *Med. Res. Rev.* 41, 928–960. doi: 10.1002/med.21747

Zhang, D., Sun, W., Shi, Y. H., Wu, L., Zhang, T. Y., and Xiang, L. (2018b). Red and blue light promote the accumulation of artemisinin in *Artemisia annua* L. *Molecules* 23:1329. doi: 10.3390/molecules23061329

Zhang, Q. Z., Wu, N. Y., Jian, D. Q., Jiang, R. Q., Yang, C. X., Lan, X. Z., et al. (2019c). Overexpression of *AaPIF3* promotes artemisinin production in *Artemisia annua*. *Ind. Crop. Prod.* 138:111476. doi: 10.1016/j.indcrop.2019.111476

Zhang, L., Zhu, W., Zhang, Y. Z., Yang, B. X., Fu, Z. R., Li, X. M., et al. (2014). Proteomics analysis of *Mahonia bealei* leaves with induction of alkaloids via combinatorial peptide ligand libraries. *J. Proteome* 110, 59–71. doi: 10.1016/j.jprot.2014.07.036

Zhou, F., Sun, T. H., Zhao, L., Pan, X. W., and Lu, S. (2015). The bZIP transcription factor HY5 interacts with the promoter of the monoterpene synthase gene QH6 in modulating its rhythmic expression. *Front. Plant Sci.* 6:304. doi: 10.3389/fpls.2015.00304

Zhu, W., Yang, B. X., Komatsu, S., Lu, X. P., Li, X. M., and Tian, J. K. (2015). Binary stress induces an increase in indole alkaloid biosynthesis in *Catharanthus roseus*. *Front. Plant Sci.* 6:582. doi: 10.3389/fpls.2015.00582

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 December 2021
doi: 10.3389/fpls.2021.809579

Molecular Cloning and Functional Characterization of a Sterol 3-*O*-Glucosyltransferase Involved in Biosynthesis of Steroidal Saponins in *Trigonella foenum-graecum*

Jianghong Gao, Yehan Xu, Congkun Hua, Changfu Li and Yansheng Zhang**

Shanghai Key Laboratory of Bio-Energy Crops, Research Center for Natural Products, Plant Science Center, School of Life Sciences, Shanghai University, Shanghai, China

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Wanhong Liu,
Chongqing University of Science and Technology, China
Zheyong Xue,
Northeast Forestry University, China

***Correspondence:**
Changfu Li
changfuli@shu.edu.cn
Yansheng Zhang
zhangys1@shu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *05 November 2021*
Accepted: *22 November 2021*
Published: *13 December 2021*

Citation:
Gao J, Xu Y, Hua C, Li C and Zhang Y (2021) Molecular Cloning and Functional Characterization of a Sterol 3-O-Glucosyltransferase Involved in Biosynthesis of Steroidal Saponins in Trigonella foenum-graecum. Front. Plant Sci. 12:809579. doi: 10.3389/fpls.2021.809579

Fenugreek (*Trigonella foenum-graecum*), a pharmacologically important herb, is widely known for its antidiabetic, hypolipidemic, and anticancer effects. The medicinal properties of this herb are accredited to the presence of bioactive steroidal saponins with one or more sugar moieties linked to the C-3 OH position of disogenin or its C25-epimer yamogenin. Despite intensive studies regarding pharmacology and phytochemical profiles of this plant, enzymes and/or genes involved in synthesizing the glycosidic part of fenugreek steroidal saponins are still missing so far. This study reports the molecular cloning and functional characterization of a key sterol-specific glucosyltransferase, designated as TfS3GT2 here, from fenugreek plant. The recombinant TfS3GT2 was purified *via* expression in *Escherichia coli*, and biochemical characterization of the recombinant enzyme suggested its role in transferring a glucose group onto the C-3 hydroxyl group of diosgenin or yamogenin. The functional role of TfS3GT2 in the steroidal saponin biosynthesis was also demonstrated by suppressing the gene in the transgenic fenugreek hairy roots *via* the RNA interference (RNAi) approach. Down-regulation of TfS3GT2 in fenugreek generally led to reduced levels of diosgenin or yamogenin-derived steroidal saponins. Thus, Tf3SGT2 was identified as a steroid-specific UDP-glucose 3-*O*-glucosyltransferase that appears to be involved in steroidal saponin biosynthesis in *T. foenum-graecum*.

Keywords: saponin, sterol 3-*O*-glucosyltransferase, *Trigonella foenum-graecum*, dioscin, diosgenin

INTRODUCTION

Trigonella foenum-graecum (fenugreek) is one of the oldest traditional medicinal plant species originating from the Iran and Mediterranean regions (Mostafaie et al., 2018). In ancient Egypt, fenugreek was documented for increasing milk production in lactating women and treating mummies (Nagulapalli Venkata et al., 2017). In traditional Chinese medicine, this plant has been prescribed for the treatment of many conditions, such as lung congestion, diabetes, epilepsy,

and paralysis (Nagulapalli Venkata et al., 2017). Currently, various medicinal properties of fenugreek have been revealed, such as antidiabetic (Raju et al., 2001), antiobesity (Gao et al., 2015), anticancer (Shabbeer et al., 2009), and sex-promoting (Aswar et al., 2010) activities.

Phytochemical studies of fenugreek seeds and other tissues have revealed the presence of steroidal saponins primarily based on spirostanol aglycones (e.g., diosgenin or its C25-epimer yamogenin) (Pang et al., 2012; Kang et al., 2013; Krol-Kogus et al., 2020). The steroidal saponins are formed by successively attaching more than one sugar moieties onto the C3-hydroxy groups of the alycones, and the first linked sugar group is almost completely a D-glucose in β-configuration (Pang et al., 2012; Kang et al., 2013; Krol-Kogus et al., 2020). Among the steroidal saponins, dioscin has received increasing attentions, and since 1960s, in the former soviet, this compound was used as a main active component of herbal products to treat coronary heart diseases (Li et al., 2021). Dioscin is the major active ingredient of the traditional Chinese medicine product, called "Di'ao Xinxuekang capsule" (Yu et al., 2014), which is currently utilized for the prevention of cardiovascular diseases. Despite intensive studies concerning the chemical structures and medicinal activities of the saponins, the genes and enzymes involved in biosynthesis of the steroidal saponins in fenugreek remain largely unknown.

Little is known about the biosynthetic steps leading to dioscin production in plants. The aglycone part of dioscin is diosgenin, and early labeling studies (Bennett and Heftmann, 1965; Joly et al., 1969b; Stohs et al., 1969) indicated that diosgenin is derived from cholesterol. Very recently, cytochrome P450s capable of converting cholesterol to diosgenin have been characterized from *T. foenum-graecum* and *Paris polyphylla* (Christ et al., 2019). Some researchers have proposed that dioscin is directly biosynthesized from diosgenin, simply by adding one glucose and two rhamnose groups at its C-3 OH position (see the route 1 of **Figure 1**; Ye et al., 2017; Li et al., 2018). However, this hypothesis is challenged by the natural occurrence of furostanol saponins (e.g., protodisocin, see its structure in **Figure 1**) in *T. foenum-graecum* (Kang et al., 2013; Krol-Kogus et al., 2020) and dioscorea genus (Li et al., 2010; Zhang et al., 2014). Unlike dioscin that bears the rings of E and F (see its structure in **Figure 1**), protodioscin lacks the ring F where the side chain is held open by C26-glucosylation (**Figure 1**). The widespread presence of protodioscin in the dioscin-producing species suggests that an alternative approach (see the route 2 of **Figure 1**) to biosynthesize dioscin occurs in nature. Indeed, Joly et al. ever provided direct evidence supporting a transformation of cholesterol to protodioscin in *Dioscorea floribunda* (Joly et al., 1969a). If the route 2 is followed, the addition of sugar groups must happen before the formation of diosgenin skeleton during dioscin biosynthesis. In either case, sterol glycosyltransferases (SGTs) responsible for the transfer of sugar groups during biosynthesis of steroidal saponins in *T. foenum-graecum* remain to be uncovered, as no relevant work has been made on this species.

In this study, we utilized a transcriptomics approach to screen putative SGTs involved in biosynthesis of the steroidal saponins from *T. foenum-graecum*. Biochemical characterization showed that one of the encoded enzymes (named TfS3GT2) is a steroid-specific 3-*O*-glucosyltransferase. This enzyme catalyzes the transfer of the first glucose group onto the C3-hydroxy group of diosgenin or yamogenin. The importance of this gene for steroidal saponin biosynthesis in *T. foenum-graecum* was revealed by silencing this gene using the RNA interference (RNA) approach. The probable stage (i.e., route 1 or 2 in **Figure 1**) at which TfS3GT2 catalyzes the glucosylation reaction is discussed.

MATERIALS AND METHODS

Plant Materials and Chemicals

Trigonella foenum-graecum seeds from Shanxi Province of China were used in this study. The chemical standards of UDP-glucose, UDP-rhamnose, UDP-xylose, diosgenin, yamogenin, cholesterol, β-sitosterol, ergosterol, brassicasterol, campesterol, kaempferol, quercetin, myricetin, protopanoxadiol, ursolic acid, trillin, prosapogenin A (α-L-rhamnopyranosyl-(1 → 2)-β-D-glucopyranoside-3-*O*-diosgenin), zingiberoside A1 (α-L-rhamnopyranosyl-(1 → 2)-β-D-glucopyranoside-3-*O*-yamogenin), and dioscin (α-L-rhamnopyranosyl-(1 → 4)-[α-L-rhamnopyranosyl-(1 → 2)]-β-D-glucopyranoside-3-*O*-diosgenin) were purchased from Shanghai Yuanye Biotechnology Co., Ltd. (Shanghai, China). The ruscogenin and stigmasterol standards were ordered from Chengdu HerbPurify Biotechnology Co., Ltd. (Chengdu, China). High performance liquid chromatography (HPLC)-grade acetonitrile or methanol (Thermo Fisher Scientific, United States) was used for HPLC and liquid chromatograph-mass spectrometer (LC-MS) analysis.

Isolation and Cloning of the Sterol C3-Glucosyltransferase Candidate Genes From *T. foenum-graecum*

Our previously reported *T. foenum-graecum* transcriptome (Zhou et al., 2019) was used for gene isolation. Five sterol C3-glucosyltransferase (S3GT) candidates (Cluster-2140.105632, Cluster-2140.95550, Cluster-2140.71031, Cluster-2140.131704, and Cluster-2140.319) were identified by a BLAST search against the known *Dioscorea zingiberensis* S3GTs (namely Dz3GT1 or Dz3GT2) that we previously have reported (Li et al., 2018). Except for the cluster-2140.95550, the other four gene candidates could be amplified by standard RT-PCR from the methyl jasmonate (MeJA)-treated *T. foenum-graecum* seedlings, which are the same set of plant materials that we previously used for establishing the *T. foenum-graecum* transcriptome (Zhou et al., 2019). The successfully amplified four candidates were designated as TfS3GT1 (cluster-2140.105632), TfS3GT2 (cluster-2140.71031), TfS3GT3 (cluster-2140.319), and TfS3GT4 (cluster-2140.131704), respectively, and they were subsequently cloned into an *Escherichia coli* expression vector pGEX-2T *via* *Bam*HI/*Eco*RI sites using a ClonExpress® One Step Cloning Kit (Vazyme, China). The N-terminus of TfS3GT1-4 was designed to

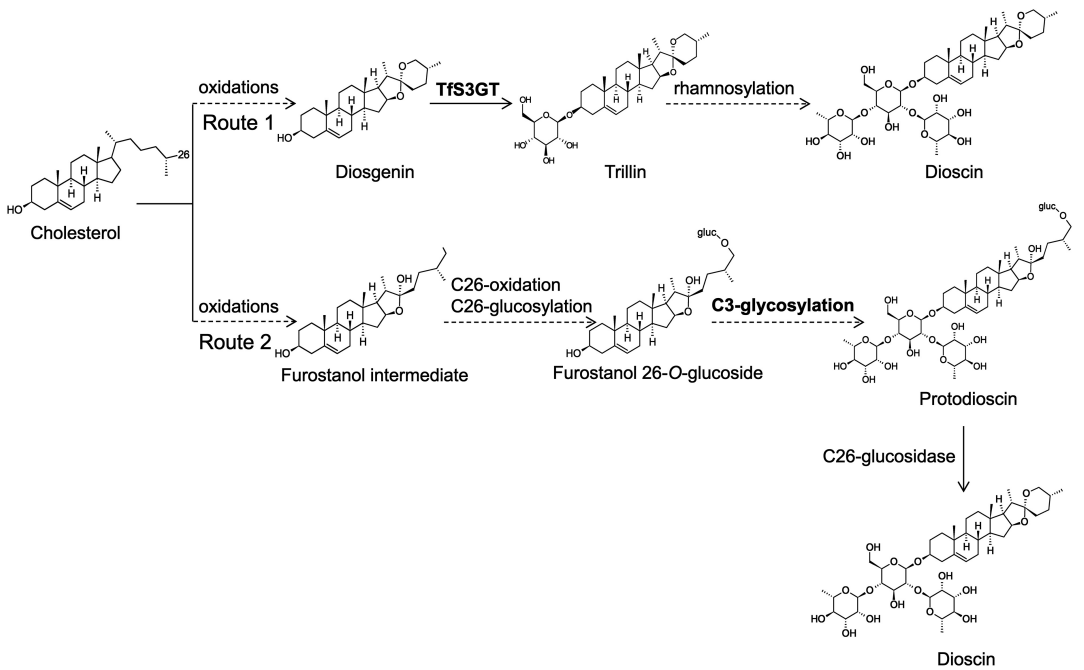


FIGURE 1 | Proposed two glycosylation schemes that may be involved in dioscin biosynthesis in *T. foenum-graecum* plants. In the route 1, the glycosidic parts (one glucose and two rhamnose moieties) of dioscin are added after the formation of diosgenin skeleton. Diosgenin is biosynthesized from multiple oxidations of cholesterol catalyzed by cytochrome P450s (Christ et al., 2019). In the route 2, the glycosidic parts of dioscin are presumed to be introduced before the formation of diosgenin skeleton, and they may be added at the furostanol level. Cholesterol is converted to proto-dioscin in which the ring F is held open by C26-glucosylation, and from the proto-dioscin intermediate cleavage of the C26-glucose group would evoke closure of the ring F, ultimately producing dioscin.

be fused in a frame with a glutathione-S-transferase (GST) tag present in the pGEX-2T vector.

Preparation of Recombinant TfS3GTs and *in vitro* Enzyme Assays

The above resulting vectors were transformed into *E. coli* strain TSsetta DE3. The cells were incubated in a Lysogeny Broth (LB) medium supplemented with 100 μg/mL ampicillin and 50 μg/mL chloramphenicol at 37°C until A_{600} reached 0.6–0.8. After addition of IPTG (isopropyl-β-D-thiogalactoside) at a final concentration of 1 mM, the cell were further cultured at 16°C for 16 h, then harvested and re-suspended in a lysis buffer (10 mM KH2PO4 pH 8.0, 140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1 mM EDTA, 2 mM dithiothreitol). Following disruption of the cells by sonication, the soluble solution was loaded onto a column filled with GST-binding magnetic beads. Recombinant TfS3GTs were then purified with an elution buffer (0.1 M potassium phosphate pH 8.0, 10 mM reduced glutathione), and concentrated into a reaction buffer (50 mM Tris–HCl; pH8.0) through a 30 kDa-desalting filter. The purity of the recombinant TfS3GTs was monitored by SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) analysis, and protein concentrations were measured by the Bradford assays.

Unless otherwise stated, the *in vitro* assays (in a total volume 100 μL) consisted of 50 mM Tris–HCl buffer (pH 8.0), 100 mM DTT, 30 mM NADPH, 50 mM sugar donor (UDP-glucose, UDP-rhamnose, or UDP-xylose), 5 mM substrate, and 7 μg purified TfS3GT. The mixture was incubated overnight at 30°C, and extracted three times with 800 μL of ethyl acetate. The ethyl acetate extracts were evaporated to dryness, and re-dissolved in 100 μL methanol for HPLC or LC-MS analysis. To measure the reaction velocities of TfSGT2 or TfS3GT4, assays were carried out for 20 min, which was proven to be a linear reaction time by preliminary experiments. For each enzyme, the assays were performed in triplicate.

Preparation of the *TfS3GT2* RNAi Construct and Generation of the Transgenic *T. foenum-graecum* Hairy Roots

To prepare the *TfS3GT2* RNAi construct, specific primers (see **Supplementary Table 1**) were designed to amplify a 483 bp-fragment from the *TfS3GT2* coding region. The amplified PCR products were first cloned into an intermediate vector pDONR201, and then introduced into a binary vector pK7GWIWG2_II-RedRoot to form a hairpin cassette using the Gateway LR recombination reaction (Thermo Fisher Scientific, United States). The vector "pK7GWIWG2_II-RedRoot" harbors a red fluorescence protein (RFP) as a selection marker. After confirmation of the construct by sequencing, the binary vector that contains the RNAi construct for *TfS3GT2*, as well as the empty vector, were transformed into the *Agrobacterium rhizogenes* ARqua1 strain (Boisson-Dernier et al., 2001) by

electroporation, and then were used for infection of *T. foenum-graecum* seedlings. The protocol, as previously described by Garagounis et al. (2020), was used to generate the *T. foenum-graecum* hairy roots with a few modifications. In brief, 3 day-old seedlings after seed germination were subjected to the infection with the ARqual strain. After cutting off the seedling radicles, the traumatized radicle surface was soaked in the *Agrobacterium* slurry harboring the target constructs for 3–5 min. The inoculated seedlings were then cultured on vertically positioned square plates, which contain half-strength MS agar media, in a growth chamber at 22°C in a 16/8 h light/dark cycle.

About 2 weeks after the infection, calluses started to form on the infected sites, while hairy roots emerged about 3 weeks post the inoculation. About 40 days later, positive roots were confirmed by examining red fluorescence signals, which resulted from expression of the RFP marker present in the RNAi construct, using a fluorescence microscope (Nikon SMZ1500, Japan). The filter sets used for excitation (Ex) and emission (Em) were as follows: RFP, 561 nm (Ex)/from 579 to 675 nm (Em); bright field, 633 nm. Signals were captured in multi-channel mode. The positive hairy roots were frozen in liquid nitrogen, and stored at –80°C for gene expression analysis and metabolite measurement later. Each biological replicate contained the hairy roots that were pooled from 30 infected plants, and data for each construct was collected from at least three biological repeats.

Metabolite Extraction From the *T. foenum-graecum* Hairy Roots

The hairy root samples were ground into a fine powder in liquid nitrogen, and were dried to constant weight in a 37°C-oven. To measure diosgenin content in the hairy root samples, the extraction was carried out as follows: 20 mg of each dried hairy root sample was extracted with 3 mL methanol, and the methanol extracts were evaporated to dryness, acid-hydrolyzed with 1.8 M sulfuric acid at 100°C for 10 h, and then extracted with 9 mL hexane. The hexane extracts were washed with water, evaporated to dryness, and then re-dissolved in 150 μL of methanol for HPLC analysis. For each sample, 60 μg of ursolic acid was included as an internal standard to normalize possible variations introduced by different extractions. To measure saponin contents in the hairy root samples, 20 mg of each dried samples was immersed in 1 mL of methanol overnight, then extracted in ultrasound bath for 1 h with addition of another 2 mL methanol. The methanol extracts were filtered through 0.22 μm nylon syringe prior to LC-MS analysis. The metabolite measurement was carried out in three biological replicates.

High Performance Liquid Chromatography and Liquid Chromatograph-Mass Spectrometer Analysis

For HPLC analyses, samples were detected on a SIL-20A HPLC system (Shimadzu, Japan) equipped with an Extend-C18 column (250 mm × 4.6 mm, 5 μm) at 30°C. The UV detection wave length was set to 203 nm. For the products from *in vitro* enzyme assays with cholesterol, β-sitosterol, ergosterol, stigmasterol, brassicasterol or campesterol, milli-Q water (solvent A), and HPLC-grade methanol (solvent B) were used as the mobile phase, and samples were separated using 98% B at a flow rate of 1.0 mL/min for 40 min. For the products from the *in vitro* assays with diosgenin, yamogenin, ruscogenin, protopanoxadiol or ursolic acid, milli-Q water (solvent A), and HPLC-grade acetonitrile (solvent B) were used as the mobile phase, and samples were separated at a flow rate of 0.8 mL/min using 60–100% B for 35 min, 100% B for 5 min, 100–60% B for 2 min, and 60% B for 8 min. For the products from the *in vitro* enzyme assays with kaempferol, quercetin, myricetin, the mobile phase was composed of water with 0.1% formic acid (solvent A) and HPLC-grade acetonitrile (solvent B), and samples were separated at a flow rate of 0.8 mL/min using 10–85% B for 10 min, 85% B for 6 min, 85–10% B for 6 min, and 10% B for 4 min. To measure diosgenin contents in the transgenic hairy roots, the mobile phase was composed of 0.1% phosphoric acid (solvent A) and 70% methanol/30% acetonitrile (solvent B), and samples were separated using 90% B for 45 min at a flow rate of 0.8 mL/min.

To monitor accumulation of the saponins in the hairy roots, LC-MS analysis was performed using a Q-Exactive Focus mass spectrometers, coupled with a VanquishTM UPLC system (Thermo Fisher Scientific, United States) and a HESI source (Thermo Fisher Scientific, United States). The column (100 mm × 2.1 mm, 1.8 μm) was used to separate the sample, the column temperature was 45°C, and the flow rate was 600 μL/min. The mobile phases contain 0.1% formic acid (solvent A) and acetonitrile (solvent B), and the solvent gradient is set as follows: 20–22% B for 1 min, 22% B for 3 min, 22–24% B for 11 min, 24–30% B for 0.1 min, 30–40% B for 4.4 min, 40–44% B for 2 min, 44–60% for 7.5 min, 60–100% B for 1 min, 100% B at 2 min. The ESI-Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, MA, United States) was operated in a full-scan and positive mode under the following conditions: resolution of 70,000; scan range of 100.0–1500.0 m/z; polarity is positive; AGC target at 3×10^6; maximum inject time is 120 ms. The parameter setting conditions of HESI source are as follows: sheath gas flow rate, aux gas flow rate and sweep gas flow rate are 40, 15, and 1, respectively; spray voltage is 3.00 kV; capillary temperature at 275°C; s-lens RF level at 55.0; aux gas heater temperature is 310°C. The instrument was controlled by Xcalibur software (Thermo Fisher Scientific, United States).

Quantitative Real-Time PCR

Total RNA was extracted from the transgenic hairy roots using an EASYspin Plus Plant RNA Extraction Kit (Aidlab, China). The first-strand cDNA was synthesized using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix Kit (TransGen Biotech, China). Quantitative RT-PCR (qRT-PCR) was performed on a ABI 7500 Fast Real-Time PCR Detection System with TOROGreen®5G qPCR PreMix Kit (TOROIVD, China). The PCR conditions were set as follows: 10 min of initial denaturation at 95°C, followed by 40 cycles of 95°C for 20 s, and then 60°C for 1 min. All real-time PCR was performed in three independent repeats. The primers used for the qRT-PCRs are listed in **Supplementary Table 1**.

Statistical Analysis

Every experiment was carried out at least in three biological replicates, and data were shown as mean ± SD. Data analysis was performed by one-way ANOVA. Difference was considered statistically significant when $p < 0.05$ (**), and extreme significant when $p < 0.01$ (***).

RESULTS

Phylogenetic Analysis and Functional Characterization of the TfS3GT Candidates

Phylogenetic Analysis

A phylogenetic analysis of the TfS3GT candidates and some previously published 3-*O*-glucosyltransferases and rhamnosyltransferases (see their GenBank accession numbers in **Supplementary Table 2**) was performed (**Figure 2**). The TfS3GT1-4 candidates all clustered into the clade composed of sterol-3-*O*-glucosyltransferases, while the flavonoid 3-*O*-glucosyltransferases formed a distinct clade in the tree, suggesting that the TfS3GT candidates are relatively more related to sterol 3-*O*-glucosyltransferases.

Functional Characterization of the TfS3GT Candidates

To characterize the catalytic functions of the TfS3GTs *in vitro*, recombinant TfS3GTs were expressed in *E. coli* cells, and purified by GST-tag affinity chromatography (**Supplementary Figure 1**). The *in vitro* assays were first performed using diosgenin as the sugar acceptor and uridine 5′-diphosphate glucose (UDP-Glc) as the sugar donor. The reaction products were then subjected to HPLC or LC-MS analysis. The negative control reaction was performed by omitting the recombinant TfS3GTs in the reaction mixture. Among the four TfS3GT candidates, only TfS3GT2 and TfS3GT4 showed an activity of converting diosgenin to a new product (**Figure 3A**). This product showed the same retention time and mass spectrum as authentic trillin (i.e., diosgenin-3-*O*-glucoside) (**Figure 3B**), supporting that either TfS3GT2 or TfS3GT4 is a functional sterol-3-*O*-glucosyltransferase that catalyzes the formation of trillin from diosgenin. By performing the assays in a linear reaction time of 20 min, the reaction velocities of TfS3GT2 and TfS3GT4 with diosgenin were calculated to be 6.50 ± 0.10 and 0.88 ± 0.29 nM/min/mg, respectively, suggesting that TfS3GT2 has a greater sterol-3-*O*-glucosyltransferase activity than TfS3GT4.

To explore the sugar donor promiscuity of TfS3GT2 or TfS3GT4, UDP-xylose and UDP-rhamnose were supplied in the *in vitro* assays, using diosgenin as an acceptor. The results demonstrated that both TfS3GT2 and TfS3GT4 could not accept UDP-xylose or UDP-rhamnose as a sugar donor. Considering that TfS3GT2 displays a much higher activity relative to TfS3GT4 (**Figure 3A**), only TfS3GT2 was further subjected to the substrate specificity analysis. The optimum pH for TfS3GT2 activity was revealed to be 8.0, which was measured by comparing the activity of TfS3GT2 with diosgenin as an acceptor and UDP-glucose as a sugar donor within the pH range of 7.0–9.0 (**Supplementary Figure 2**). Next, using Tris–HCl (pH 8.0) as a reaction buffer, we tested the activity of TfS3GT2 with four types of compounds, which all share a hydroxyl group at their C3-positions. They are spirostane sapogenins containing E- and F-rings (i.e., diosgenin, ruscogenin, tigogenin, and yamogenin), Δ5-sterols without E- and F-rings (i.e., cholesterol, β-sitosterol, ergosterol, stigmasterol, brassicasterol, and campesterol), flavonoids (i.e., kaempferol, quercetin, and myricetin), and triterpenoids (i.e., protopanoxadiol and ursolic acid) (see their structures in **Figure 4A**). TfS3GT2 showed activities toward the Δ5-sterols with or without the E- and F-rings, but it did not show glucosylation activity on triterpenoid and flavonoid substrates (**Figure 4B**), demonstrating that TfS3GT2 is a sterol-specific 3-*O*-glucosyltransferase. Generally, TfS3GT2 showed the highest activity toward spinostanol sterols with 11.5–71.3% of the activity being found with the selected Δ5-sterols without the E- and F-rings (**Figure 4B**). Interestingly, TfS3GT2 showed no activity toward ruscogenin and tigogenin. Tigogenin is a spirostanol sapogenin bearing the E- and F-rings, but it lacks a C5-C6 double bond (**Figure 4A**). Ruscogenin differs from diosgenin only by the presence of a hydroxyl group at C1-position (**Figure 4A**).

Saponin and Diosgenin Content in the *TfS3GT2*_RNAi Transgenic Hairy Roots

In order to understand the *in planta* function of TfS3GT2, *T. foenum-graecum* seedlings were transformed with the *TfS3GT2*_RNAi construct to generate the transgenic hairy roots. Transformed roots were screened by red fluorescence examination under a microscope for the presence of the transgenes (**Figure 5A**). The expression level of *TfS3GT2* gene was analyzed by qRT-PCR in the *TfS3GT2*_RNAi and the vector control transformants. Expression of *TfS3GT2* gene was decreased up to 43.4% due to the silencing in comparison with the vector control (**Figure 5B**). To investigate the effects of the *TfS3GT2* silencing on the saponin and diosgenin contents in the transgenic hairy roots, quantification of six different diosgenin- or yamogenin-derived saponins (i.e., dioscin, prosapogenin A, (**Figure 3A**) deltonin, graecunin E, diosigenin-S19-Xyl, and diosigenin-S19) (see their structures in **Figure 5C**) was carried out using LC-MS analysis. Overall, with respect to the control, a significant decrease in the contents of all the diosgenin-derived saponins, with a percentage decrease of 18.7–45.9%, was observed (**Figure 5D**). Diosgenin content was slightly decreased, but the percentage decrease was not significant (**Figure 5D**).

DISCUSSION

This study pertains to be the first report on cloning and characterization of a steroid-specific metabolism gene (*TfS3GT2*) from *T. foenum-graecum*. The *TfS3GT2* mRNA was amplified only from the methyl-jasmonate (MeJA)-treated *T. foenum-graecum* seedlings, but not from the non-treated ones (data not shown), indicating that *TfS3GT2* mRNA was up-regulated by MeJA. This is consistent with the gene expression data

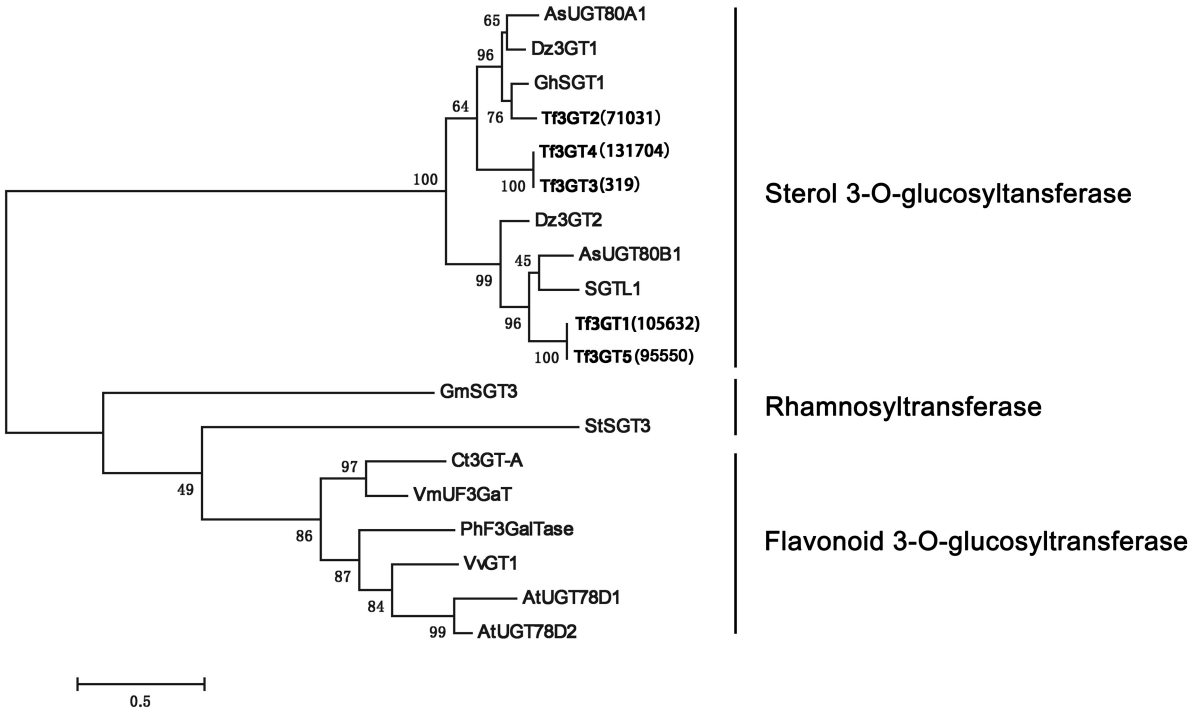


FIGURE 2 | Phylogenetic analysis of the four TfS3GTs together with the previously characterized glucose- or rhamnose- transferases. The evolution history was inferred based on the neighbor-joining method using the software MEGA 7.0 (Kumar et al., 2016). The analysis incudes 14 previously published sequences, and their Genbank accession numbers are listed in the **Supplementary Table 2**. The bootstrap values are 500, and the scale bar represents 0.5 amino acid substitutions per site.

acquired by our previously RNA-sequencing of *T. foenum-graecum* seedlings (Zhou et al., 2019), from which it appeared that *TfS3GT2* expression was strongly induced by MeJA. MeJA is a known signaling compound that can induce oxidative stress and biosynthesis of secondary metabolites in plant cells (Ho et al., 2020). Accordingly, at the beginning of this study, we speculated that *TfS3GT2* could be a defense gene leading to formation of some glycosides in response to oxidative stresses. In this study, biochemical and genetic functional analysis of TfS3GT2 demonstrate that TfS3GT2 involves in biosynthesis of dioscin and other diosgenin-derived glycosides. Interestingly, dioscin exerts its beneficial properties to human health, indeed largely *via* inhibition of oxidative stresses through different signaling pathways (Hu et al., 2018; Yang et al., 2018). TfS3GT2 also shows considerable activity in catalyzing glucosylation of several major plant-derived sterols (e.g., sitosterol, stimasterol, and campesterol) (**Figure 3**), further indicating its physiological role in adaptions to stressful environments, as glycosylation of the major plant sterols can provide increased tolerance to both abiotic and biotic stresses (Saema et al., 2016).

Our *in vitro* enzyme assays (**Figure 4**) revealed that TfS3GT2 showed strict substrate specificity with highly specific activity toward steroid aglycones, whereas non-steroidal substrates, including flavonoids and triterpenoids, were not active as sugar acceptors, strongly suggesting that TfS3GT2 is a sterol-specific glycosyltransferase. This biochemical finding well matched the result of the phylogenetic tree analysis, as in the tree TfS3GT2 showed a close relationship with sterol-3-*O*-glucosyltransferases but was distinct from others (**Figure 2**). The presence of a C5-C6 double bond in the steroid molecule seemed to be critical in glucosylation, as TfS3GT2 showed no detectable activity on tigogenin (**Figure 4**), which only differs from the best substrate diosgenin by the absence of a C5-C6 bond. Moreover, the selected triterpenoid substrates, which actually resemble the backbone structure of steroidal aglycones but lack a C5-C6 double bond, were also not glycosylated by TfS3GT2 at all (**Figure 4**). The presence of a C5-C6 double bond in steroid molecules conferring the substrate specificity was also demonstrated for a S3GT from *Withania somnifera* (Madina et al., 2007). Among the steroid substrates tested, diosgenin was the best aglycone substrate of TfS3GT2. To the best of our knowledge, in a pure form or as a crude extract, UGTs capable of catalyzing 3-*O*-glucosylation of diosgenin or yamogenin were only isolated from a few species, including *Solanum melongena* L (Pazkowski et al., 2001; Potocka and Zimowski, 2008), *W. somnifera* (Madina et al., 2007), and *D. zingiberensis* (Ye et al., 2017). In comparison with these previously reported S3GTs, TfS3GT2 was judged to be a novel diosgenin 3-*O*-glucosyltransferase based on the following observations: (1) The previously reported diosgenin 3-*O*-glucosyltransferases all give low activity on diosgenin, however, which is the best substrate here for TfS3GT2; (2) TfS3GT2 shares relatively low amino acid sequence identity (50–67%) with the previously reported diosgenin 3-*O*-glucosyltransferases. Interestingly, TfS3GT2 was

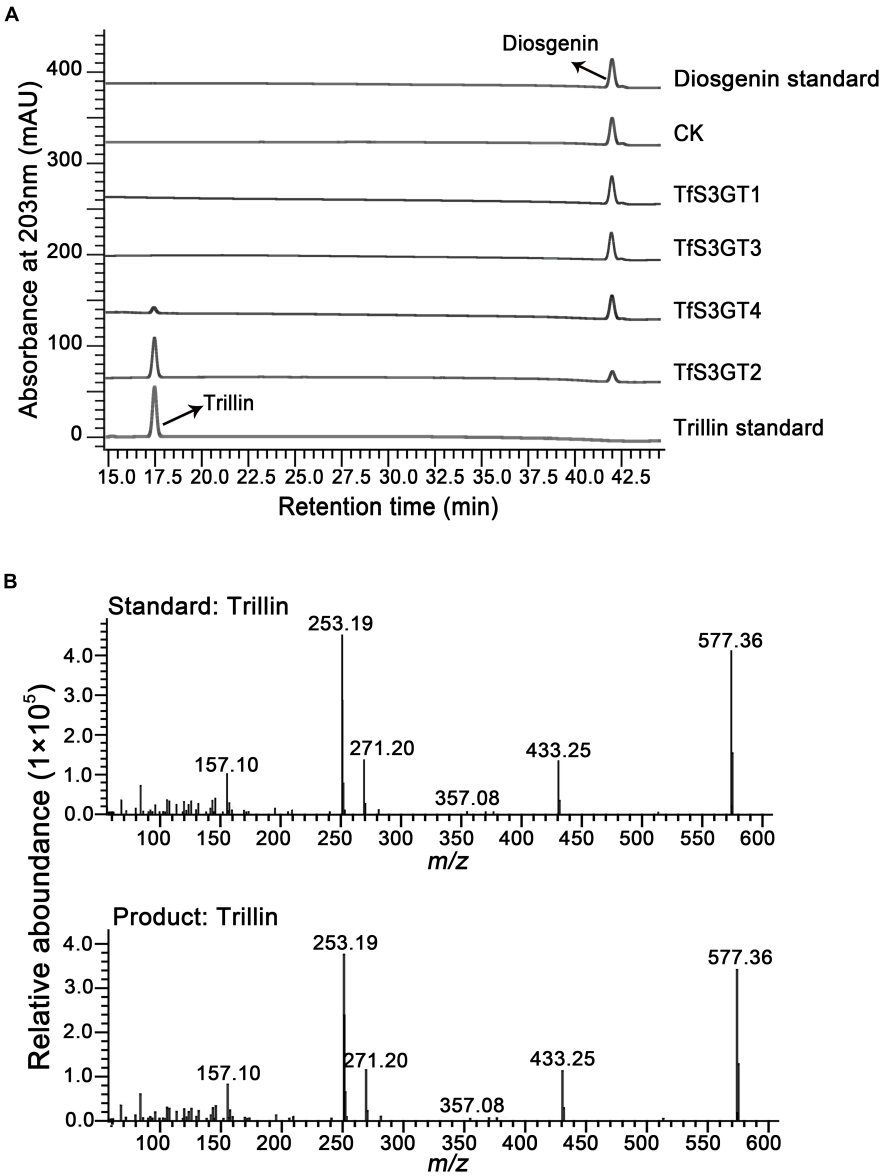


FIGURE 3 | HPLC analysis of the products extracted from the *in vivo* assays of TfS3GT2 or TfS3GT4 with diosgenin **(A)**. The mass spectrum of the trillin product in comparison with the trillin standard **(B)**.

not active with ruscogenin. It should be noted that ruscogenin has the C5-C6 double bond, and differs from diosgenin only by the presence of a hydroxyl group at C1-position (**Figure 4A**). The C1-OH of ruscogenin would increase the hydrophilicity, probably preventing it to be accommodated by TfS3GT2, since hydrophobicity is a common feature for the acceptor binding pocket of a sterol-specific glycosyltransferase (Chen et al., 2018). Indeed, diosgenin- or yamogenin-derived glycosides widely occur in *T. foenum-graecum*, but none of ruscogenin-derived saponins have been reported from this plant (Pang et al., 2012; Kang et al., 2013). To determine the substrate specificity of TfS3GT2 with respect to the sugar donor, UDP-glucose, UDP-xylose and UDP-rhamnose were tested, because they represent the dominant sugar moieties of steroidal saponins of *T. foenum-graecum* (Pang et al., 2012; Kang et al., 2013; Krol-Kogus et al., 2020). Only UDP-glucose was active as a sugar donor, whereas UDP-xylose or UDP-rhamnose could not replace UDP-glucose to facilitate the TfS3GT2 activity. The sugar donor specificity of TfS3GT2 is consistent with the fact that the first sugar moiety attached to the steroidal aglycones of *T. foenum-graecum* is almost completely the glucose group (Pang et al., 2012; Kang et al., 2013; Krol-Kogus et al., 2020).

In addition to TfS3GT2, this study has revealed another enzyme, designated as TfS3GT4, that also catalyzed the formation of diosgenin-3-*O*-glucoside (i.e., trillin) from

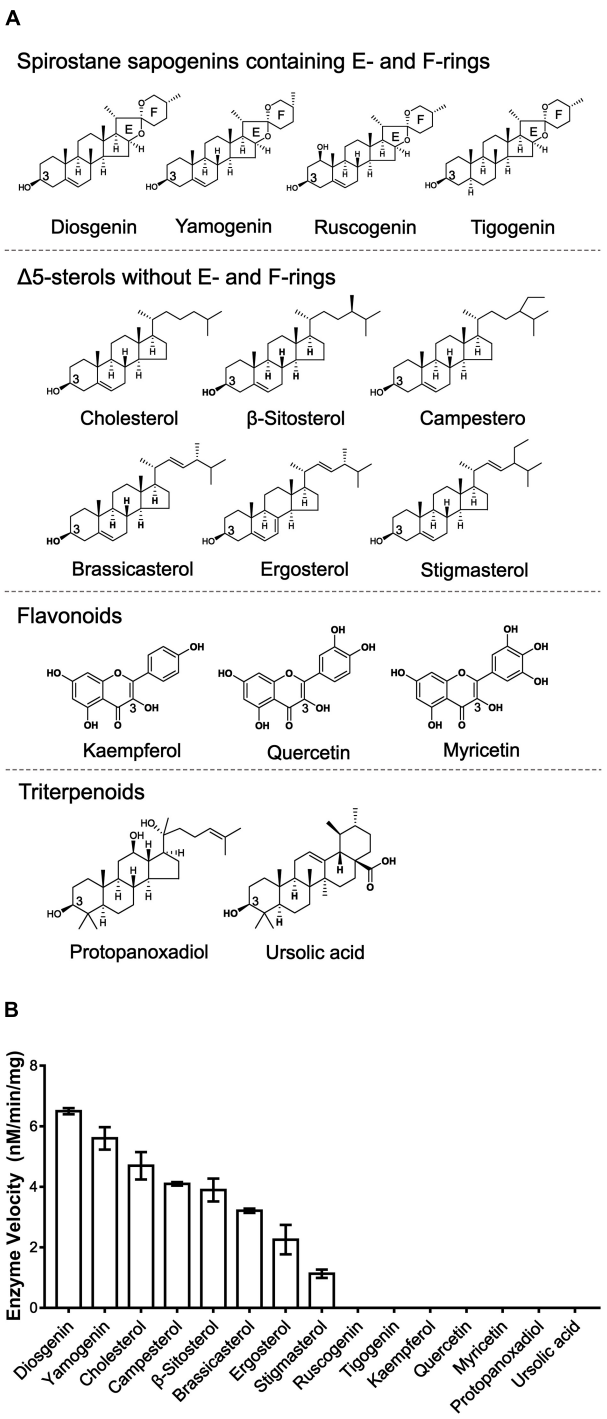


FIGURE 4 | Enzyme activities of TfS3GT2 with different types of aglycones. **(A)** Chemical structures of four types of the alycones tested. **(B)** Enzyme activities of TfS3GT2 on four types of the aglycones with UDP-glucose as a sugar donor. In the *in vitro* enzyme assays, saturate concentrations of the acceptors and the UDP-glucose donor were utilized (see the section "Materials and Methods"). The assays were carried out for 20 min. The enzyme activities of TfS3GT2 on various alycones were quantified based on the mole numbers of the aglycones consumed by the enzyme, relative to the control reactions which do not include TfS3GT2. Values represent the means ± SD from three independent assays.

disogenin (**Figure 2**). TfS3GT4 was calculated to exhibit only about 13.3% catalytic activity of TfS3GT2 in glycosylating the aglycone diosgenin. TfS3GT4 shares low sequence identity (54.4%) with TfS3GT2, with its N-terminal domain being shorter by 130 amino acid residues relative to TfS3GT2 (**Supplementary Figure 3**). The extremely low activity of TfS3GT4 toward diosgenin indicates that TfS3GT4 may have a unique substrate utilization, and diosgenin may not be its natural substrate *in vivo*. Recently, structural basis for the substrate specificity of a sterol 3β-glucosyltransferase, UGT51from *Saccharomyces cerevisiae*, has been elucidated (Chen et al., 2018). The yeast UGT51 shows considerable activities against ergosterol, cholesterol, and sitosterol (Warnecke et al., 1999), all of which can also be accepted as good substrates by TfS3GT2 (**Figure 4**). These observations tempted us to speculate that TfS3GT2 and UGT51 may share a similar structural mechanism for the activity, although there is only 14% amino acid identity between TfS3GT2 and UGT51 (**Supplementary Figure 3**). Amino acid sequence alignment showed that the nucleotide base binding motifs at the C-terminal domains of TfS3GT2, TfS3GT4, and UGT51 are highly conserved (**Supplementary Figure 3**), which is in agreement with the fact that they all display the same sugar donor specificity for the glucose. When focused on the acceptor-interacting residues in the cavity of UGT51 (Chen et al., 2018), TfS3GT2 and TfS3GT4 share the same residues at the equivalent positions with only one variable sequence at the 265 position (numbering in TfS3GT2) (**Supplementary Figure 3**). By contrast, the acceptor-interacting residues in UGT51 are distinct from the counterparts of TfS3GT2 or TfS3GT4 (**Supplementary Figure 3**), indicating that the SGTs from different organisms have evolved to accommodate diversely sterol acceptors present in their hosts, through changing only several residues within the active site of the enzyme.

This study could not provide conclusive information of which glycosylation scheme (i.e., routes 1 and 2 of **Figure 1**) that would be the major route for dioscin biosynthesis in *T. foenum-graecum*. Down-regulation of the *TfS3GT2* expression by our RNAi experiments led to a significant reduction in levels of dioscin and other diosgenin-derived saponins (**Figure 5**), strongly demonstrating that TfS3GT2 is involved in biosynthesis of steroidal saponins in *T. foenum-graecum*. If the sugar moieties are added after the formation of diosgenin, following the route 1 as shown in **Figure 1**, the reduction of the glycosides would in turn stimulate biosynthesis of the alycone backbone (i.e., diosgenin) through a feedback regulation. Through RNAi silencing of a UDP-glycosyltransferase, this type of the feedback regulation has also been observed for biosynthesis of withaferin A-derived saponins in *W. somnifera* (Saema et al., 2015) and ginsenosides in *Panax ginseng* (Lu et al., 2017). However, our data do not support this premise, as the aglycone diosgenin was not increased, and even was slightly decreased in the *TfS3GT2*-silenced lines (**Figure 5**). By contrast, if the sugar moieties are introduced before the formation of diosgenin skeleton following the route 2, the RNAi silencing of the first 3-*O*-glucosyltransferase would result in reduction in levels of both the saponins and diosgenin. Thus, the route 2 seems to give a plausible explanation of the data obtained from the RNAi experiments of this study,

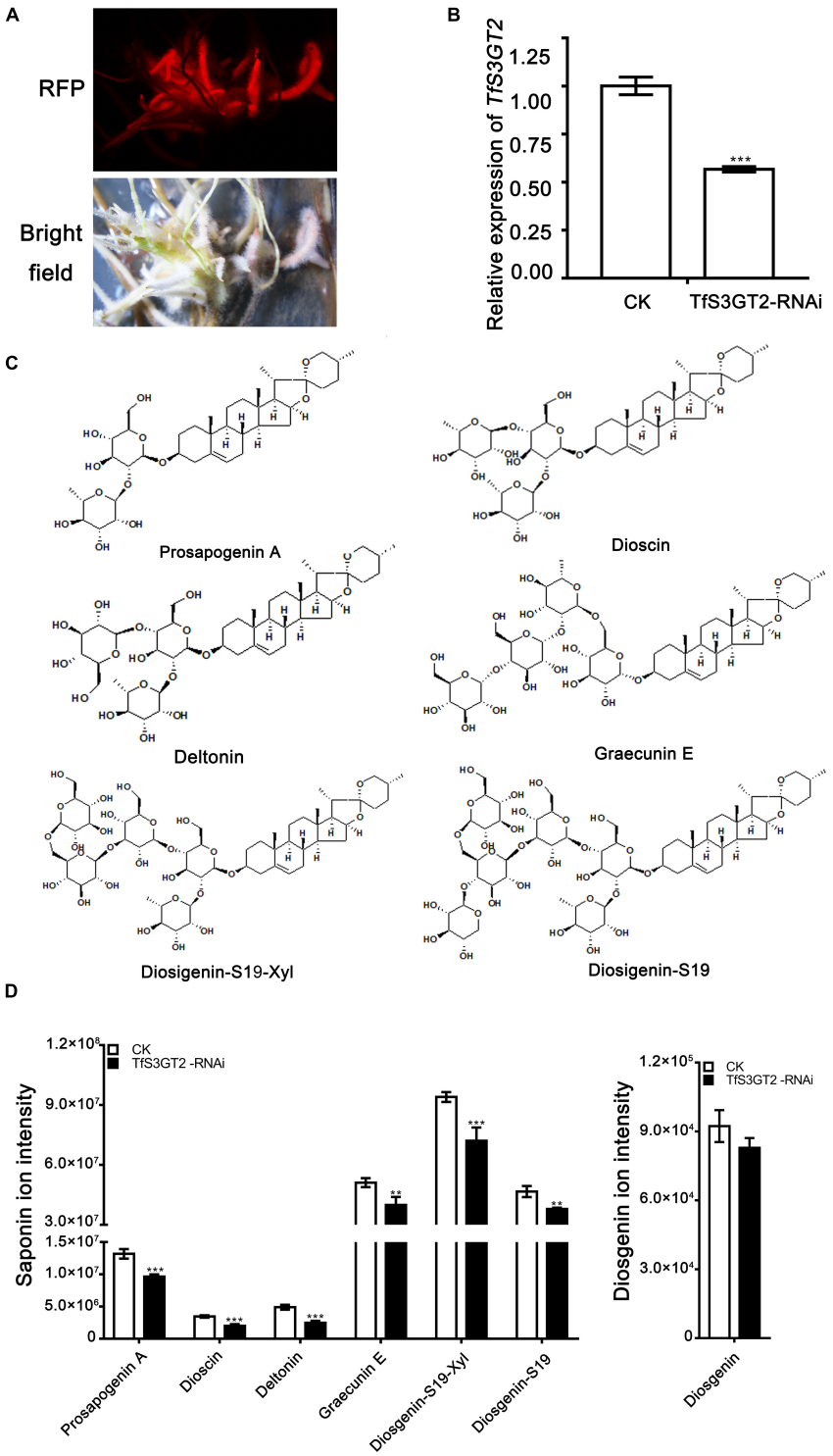


FIGURE 5 | The effects of the RNAi-based silencing of *TfS3GT2* on biosyntheses of dioscin, diosgenin, and other targeted steroidal saponins in the transgenic *T. foenum-graecum* hairy roots. **(A)** Red fluorescence signal is shown for the presence of the TfS3GT2-RNAi construct in the hairy roots. Scale bar = 1 cm. **(B)** Real-time PCR analysis of the *TfS3GT2* transcript in the TfS3GT2-RNAi compared with the corresponding vector control lines. **(C)** Chemical structures of the steroidal saponins for the targeted analysis of the transgenic hairy roots. **(D)** The LC-MS product ion-intensities of the targeted steroidal saponins and diosgenin extracted from the transgenic hairy roots. The data are expressed as the mean ± SD of three biological replicates. Asterisk indicates a significant difference by one-way ANOVA analysis, and significant difference is indicated by ** when $P < 0.05$, and extremely significant difference is shown by *** when $P < 0.01$. The mass spectra of the targeted steroidal saponins are shown in **Supplementary Figure 4**.

as the RNAi silencing of *TfS3GT2* indeed led to reduced levels of both diosgenin and its derived saponins, although the decrease in diosgenin was not statistically significant (**Figure 5**). One might argue that, in the biochemical assays of this study, diosgenin was revealed to be the best substrate of TfS3GT2 among the steroid substrates tested, thus favoring the route 1 that may exist in *T. foenum-graecum*. However, this puzzle could be explained by the catalytic plasticity of TfS3GT2 for different sterol substrates, which also seems to be a common feature of other previously reported sterol-glycosyltransferases (Pazkowski et al., 2001; Potocka and Zimowski, 2008; Stucky et al., 2015). Of course, this open scientific question deserves further investigations, which could be carried out by feeding radio-labeled glucose to the *TfS3GT2*-silenced or overexpressing lines, followed with phytochemical measurement of the radio-labeled glycosides.

CONCLUSION

This study reports the isolation and functional characterization of a novel sterol-specific 3-*O*-glucosyltransferase (TfS3GT2) that is associated with the steroidal saponin biosynthesis in *T. foenum-graecum*. First, the recombinant TfS3GT2 was purified *via* expression in *E. coli*, and *in vitro* enzyme assays showed that TfS3GT2 exhibits considerable activities toward spinostanol sapongenins (including diosgenin and yamogenin) and other Δ5-sterols, whereas it is not active with other substrates, such as triterpenoids and flavonoids, supporting that TfS3GT2 is a sterol-specific glucosyltransferase. Next, we prepared the transgenic *T. foenum-graecum* hairy roots that decreased the *TfS3GT2* expression by the RNA interference approach. The RNAi silencing of TfS3GT2 led to a significant reduction in several steroidal saponins tested, including dioscin, strongly suggesting that TfS3GT2 is involved in biosynthesis of the steroidal saponins in *T. foenum-graecum*. We discussed the glycosylation scheme by which TfS3GT2 plays its catalytic role during dioscin biosynthesis *in vivo*.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YZ designed the experiment. YZ and JG contributed to the writing. JG, YX, and CH performed the experiments. CL supervised the project and analyzed the data. All authors read and confirmed its content.

FUNDING

The work was jointly supported by a grant from the National Key R&D Program of China (2018YFC1706200) and a grant from the National Natural Science Foundation of China (31670300).

ACKNOWLEDGMENTS

We would like to thank Fang Xie from the CAS Center for Excellence in Molecular Plant Sciences, China, for providing assistance in screening the transgenic hairy roots by fluorescence examination.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.809579/full#supplementary-material

REFERENCES

Aswar, U., Bodhankar, S. L., Mohan, V., and Thakurdesai, P. A. (2010). Effect of furostanol glycosides from *Trigonella foenum*-graecum on the reproductive system of male albino rats. *Phytother. Res.* 24, 1482–1488. doi: 10.1002/ptr.3129

Bennett, R. D., and Heftmann, E. (1965). Biosynthesis of *Dioscorea sapogenins* from cholesterol. *Phytochemistry* 4, 577–586.

Boisson-Dernier, A., Chabaud, M., Garcia, F., Becard, G., Rosenberg, C., and Barker, D. G. (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant Microbe Interact.* 14, 695–700. doi: 10.1094/MPMI.2001.14.6.695

Chen, L., Zhang, Y., and Feng, Y. (2018). Structural dissection of sterol glycosyltransferase UGT51 from *Saccharomyces cerevisiae* for substrate specificity. *J. Struct. Biol.* 204, 371–379. doi: 10.1016/j.jsb.2018.11.001

Christ, B., Xu, C., Xu, M., Li, F. S., Wada, N., Mitchell, A. J., et al. (2019). Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* 10:3206. doi: 10.1038/s41467-019-11286-7

Gao, F., Du, W., Zafar, M. I., Shafqat, R. A., Jian, L., Cai, Q., et al. (2015). 4-Hydroxyisoleucine ameliorates an insulin resistant-like state in 3T3-L1 adipocytes by regulating TACE/TIMP3 expression. *Drug Des. Devel. Ther.* 9, 5727–5736. doi: 10.2147/DDDT.S92355

Garagounis, C., Beritza, K., Georgopoulou, M. E., Sonawane, P., Haralampidis, K., Goossens, A., et al. (2020). A hairy-root transformation protocol for *Trigonella foenum*-graecum L. as a tool for metabolic engineering and specialised metabolite pathway elucidation. *Plant Physiol. Biochem.* 154, 451–462. doi: 10.1016/j.plaphy.2020.06.011

Ho, T. T., Murthy, H. N., and Park, S. Y. (2020). Methyl jasmonate induced oxidative stress and accumulation of secondary metabolites in plant cell and organ cultures. *Int. J. Mol. Sci.* 21:716. doi: 10.3390/ijms21030716

Hu, Y., Mao, Z., Xu, L., Yin, L., Tao, X., Tang, Z., et al. (2018). Protective effect of dioscin against intestinal ischemia/reperfusion injury *via* adjusting miR-351-5p-mediated oxidative stress. *Pharmacol. Res.* 137, 56–63. doi: 10.1016/j.phrs.2018.09.016

Joly, R. A., Bonner, J., Bennett, R. D., and Heftmann, E. (1969a). Conversion of cholesterol to an open-chain saponin by *Dioscorea floribunda*. *Phytochemistry* 8, 857–859.

Joly, R. A., Bonner, J., Bennett, R. D., and Heftmann, E. (1969b). The biosynthesis of steroidal sapogenins in *Dioscorea floribunda* from doubly labelled cholesterol. *Phytochemistry* 8, 1709–1711.

Kang, L. P., Zhao, Y., Pang, X., Yu, H. S., Xiong, C. Q., Zhang, J., et al. (2013). Characterization and identification of steroidal saponins from the seeds of *Trigonella foenum*-graecum by ultra high-performance liquid chromatography and hybrid time-of-flight mass spectrometry. *J. Pharm. Biomed. Anal.* 74, 257–267. doi: 10.1016/j.jpba.2012.11.005

Krol-Kogus, B., Glod, D., and Krauze-Baranowska, M. (2020). Qualitative and quantitative HPLC-ELSD-ESI-MS analysis of steroidal saponins in fenugreek seed. *Acta Pharm.* 70, 89–99. doi: 10.2478/acph-2020-0013

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

Li, H., Huang, W., Wen, Y., Gong, G., Zhao, Q., and Yu, G. (2010). Anti-thrombotic activity and chemical characterization of steroidal saponins from *Dioscorea zingiberensis* C.H. Wright. *Fitoterapia* 81, 1147–1156. doi: 10.1016/j.fitote.2010.07.016

Li, J., Liang, Q., Li, C. F., Liu, M. D., and Zhang, Y. S. (2018). Comparative transcriptome analysis identifies putative genes involved in dioscin biosynthesis in *Dioscorea zingiberensis*. *Molecules* 23:454. doi: 10.3390/Molecules23020454

Li, X. F., Liu, S. L., Qu, L. P., Chen, Y., Yuan, C. Q., Qin, A. Q., et al. (2021). Dioscin and diosgenin: Insights into their potential protective effects in cardiac diseases. *J. Ethnopharmacol.* 274:114018. doi: 10.1016/J.Jep.2021.114018

Lu, C., Zhao, S., Wei, G., Zhao, H., and Qu, Q. (2017). Functional regulation of ginsenoside biosynthesis by RNA interferences of a UDP-glycosyltransferase gene in Panax ginseng and Panax quinquefolius. *Plant Physiol. Biochem.* 111, 67–76. doi: 10.1016/j.plaphy.2016.11.017

Madina, B. R., Sharma, L. K., Chaturvedi, P., Sangwan, R. S., and Tuli, R. (2007). Purification and physico-kinetic characterization of 3 beta-hydroxy specific sterol glucosyltransferase from *Withania somnifera* (L) and its stress response. *Biochim. Biophys. Acta Proteins Proteom.* 1774, 392–402. doi: 10.1016/j.bbapap.2006.12.009

Mostafaie, A., Kahrizi, D., Mohammadi, M., Yari, K., Rostami-Ahmadvandi, H., Yaghotipoor, A., et al. (2018). Effect of planting time and vermicompost on the proteomic pattern of fenugreek (*Trigonella foenum*-graecum). *Cell Mol. Biol.* 64, 46–51. doi: 10.14715/cmb/2018.64.9.7

Nagulapalli Venkata, K. C., Swaroop, A., Bagchi, D., and Bishayee, A. (2017). A small plant with big benefits: Fenugreek (*Trigonella foenum*-graecum Linn.) for disease prevention and health promotion. *Mol. Nutr. Food Res.* 61:950. doi: 10.1002/mnfr.201600950

Pang, X., Kang, L., Yu, H., Zhao, Y., Xiong, C., Zhang, J., et al. (2012). Rapid isolation of new furostanol saponins from fenugreek seeds based on ultra-performance liquid chromatography coupled with a hybrid quadrupole time-of-flight tandem mass spectrometry. *J. Sep. Sci.* 35, 1538–1550. doi: 10.1002/jssc.201200020

Pazkowski, C., Kalinowska, M., and Wojciechowski, Z. A. (2001). Phospholipids modulate the substrate specificity of soluble UDP-glucose:steroid glucosyltransferase from eggplant leaves. *Phytochemistry* 58, 663–669. doi: 10.1016/s0031-9422(01)00292-8

Potocka, A., and Zimowski, J. (2008). Metabolism of conjugated sterols in eggplant. Part 1. UDP-glucose : sterol glucosyltransferase. *Acta Biochim. Pol.* 55, 127–134.

Raju, J., Gupta, D., Rao, A. R., Yadava, P. K., and Baquer, N. Z. (2001). *Trigonella foenum* graecum (fenugreek) seed powder improves glucose homeostasis in alloxan diabetic rat tissues by reversing the altered glycolytic, gluconeogenic and lipogenic enzymes. *Mol. Cell. Biochem.* 224, 45–51. doi: 10.1023/A:1011974630828

Saema, S., Rahman, L. U., Niranjan, A., Ahmad, I. Z., and Misra, P. (2015). RNAi-mediated gene silencing of WsSGTL1 in W. somnifera affects growth and glycosylation pattern. *Plant Signal. Behav.* 10:e1078064. doi: 10.1080/15592324.2015.1078064

Saema, S., Rahman, L. U., Singh, R., Niranjan, A., Ahmad, I. Z., and Misra, P. (2016). Ectopic overexpression of WsSGTL1, a sterol glucosyltransferase gene in *Withania somnifera*, promotes growth, enhances glycowithanolide and provides tolerance to abiotic and biotic stresses. *Plant Cell Rep.* 35, 195–211. doi: 10.1007/s00299-015-1879-5

Shabbeer, S., Sobolewski, M., Anchoori, R. K., Kachhap, S., Davidson, N., Carducci, M. A., et al. (2009). Fenugreek A naturally occurring edible spice as an anticancer agent. *Cancer Biol. Ther.* 8, 268–274. doi: 10.4161/Cbt.8.3.7443

Stohs, S. J., Kaul, B., and Staba, E. J. (1969). The metabolism of 14C-cholesterol by *Dioscorea deltoidea* suspension cultures. *Phytochemistry* 8, 1679–1686.

Stucky, D. F., Arpin, J. C., and Schrick, K. (2015). Functional diversification of two UGT80 enzymes required for steryl glucoside synthesis in *Arabidopsis*. *J. Exp. Bot.* 66, 189–201. doi: 10.1093/jxb/eru410

Warnecke, D., Erdmann, R., Fahl, A., Hube, B., Muller, F., Zank, T., et al. (1999). Cloning and functional expression of UGT genes encoding sterol glucosyltransferases from *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia pastoris*, and *Dictyostelium discoideum*. *J. Biol. Chem.* 274, 13048–13059. doi: 10.1074/jbc.274.19.13048

Yang, B., Xu, B., Zhao, H., Wang, Y. B., Zhang, J., Li, C. W., et al. (2018). Dioscin protects against coronary heart disease by reducing oxidative stress and inflammation *via* Sirt1/Nrf2 and p38 MAPK pathways. *Mol. Med. Rep.* 18, 973–980. doi: 10.3892/mmr.2018.9024

Ye, T., Song, W., Zhang, J. J., An, M., Feng, S., Yan, S., et al. (2017). Identification and functional characterization of DzS3GT, a cytoplasmic glycosyltransferase catalyzing biosynthesis of diosgenin 3-O-glucoside in *Dioscorea zingiberensis*. *Plant Cell Tissue Organ Cult.* 129, 399–410. doi: 10.1007/s11240-017-1187-6

Yu, Y., Hu, S., Li, G., Xue, J., Li, Z., Liu, X., et al. (2014). Comparative effectiveness of Di'ao Xin Xue Kang capsule and Compound Danshen tablet in patients with symptomatic chronic stable angina. *Sci. Rep.* 4:7058. doi: 10.1038/srep07058

Zhang, X. X., Liang, J. R., Liu, J. L., Zhao, Y., Gao, J., Sun, W. J., et al. (2014). Quality control and identification of steroid saponins from *Dioscorea zingiberensis* C. H. Wright by fingerprint with HPLC-ELSD and HPLC-ESI-Quadrupole/Time-of-fight tandem mass spectrometry. *J. Pharm. Biomed. Anal.* 91, 46–59. doi: 10.1016/j.jpba.2013.11.023

Zhou, C., Li, X., Zhou, Z., Li, C., and Zhang, Y. (2019). Comparative transcriptome analysis identifies genes involved in diosgenin biosynthesis in *Trigonella foenum*-graecum L. *Molecules* 24:140. doi: 10.3390/molecules24010140

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 22 December 2021
doi: 10.3389/fpls.2021.793313

Epidermis-Specific Metabolic Engineering of Sesquiterpene Formation in Tomato Affects the Performance of Potato Aphid *Macrosiphum euphorbiae*

Fumin Wang, Yong-Lak Park and Michael Gutensohn*

Division of Plant and Soil Sciences, Davis College of Agriculture, Natural Resources and Design, West Virginia University, Morgantown, WV, United States

Tomato produces a number of terpenes in their glandular trichomes that contribute to host plant resistance against pests. While glandular trichomes of cultivated tomato *Solanum lycopersicum* primarily accumulate a blend of monoterpenes, those of the wild tomato species *Solanum habrochaites* produce various sesquiterpenes. Recently, we have identified two groups of sesquiterpenes in *S. habrochaites* accessions that negatively affect the performance and choice behavior of the potato aphid (*Macrosiphum euphorbiae*). Aphids are piercing-sucking herbivores that use their mouthpart to penetrate and probe plant tissues in order to ultimately access vascular tissue and ingest phloem sap. Because secondary metabolites produced in glandular trichomes can affect the initial steps of the aphid feeding behavior, introducing the formation of defensive terpenes into additional plant tissues *via* metabolic engineering has the potential to reduce tissue penetration by aphids and in consequence virus transmission. Here, we have developed two multicistronic expression constructs based on the two sesquiterpene traits with activity toward *M. euphorbiae* previously identified in *S. habrochaites*. Both constructs are composed of sequences encoding a prenyl transferase and a respective *S. habrochaites* terpene synthase, as well as enhanced green fluorescent protein as a visible marker. All three coding sequences were linked by short nucleotide sequences encoding the foot-and-mouth disease virus 2A self-processing oligopeptide which allows their co-expression under the control of one promoter. Transient expression of both constructs under the epidermis-specific *Arabidopsis CER5*-promoter in tomato leaves demonstrated that formation of the two sets of defensive sesquiterpenes, β-caryophyllene/α-humulene and (–)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene, can be introduced into new tissues in tomato. The epidermis-specific transgene expression and terpene formation were verified by fluorescence microscopy and tissue fractionation with subsequent analysis of terpene profiles, respectively. In addition, the longevity and fecundity of *M. euphorbiae* feeding on these engineered tomato leaves were significantly reduced, demonstrating the efficacy of this novel aphid control strategy.

Keywords: sesquiterpenes, metabolic engineering, multicistronic expression constructs, prenyl transferases, terpene synthases, epidermis, *Solanum lycopersicum*, potato aphid

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Monica Borghi,
Utah State University, United States
Dinesh A. Nagegowda,
Council of Scientific and Industrial Research (CSIR), India

***Correspondence:**
Michael Gutensohn
michael.gutensohn@mail.wvu.edu

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *11 October 2021*
Accepted: *29 November 2021*
Published: *22 December 2021*

Citation:
Wang F, Park Y-L and Gutensohn M (2021) Epidermis-Specific Metabolic Engineering of Sesquiterpene Formation in Tomato Affects the Performance of Potato Aphid Macrosiphum euphorbiae. Front. Plant Sci. 12:793313. doi: 10.3389/fpls.2021.793313

INTRODUCTION

In nature, plants are an integral part of a complex system of antagonistic and mutualistic biotic interactions. Due to their sessile lifestyle, plants have adapted to the resulting challenges by evolving specific strategies for the defense against attacking as well as the attraction of beneficial organisms. Many of these strategies include the formation of secondary metabolites, such as volatile organic compounds (VOCs), that belong to three major categories: phenylpropanoids/benzenoids, fatty acid derivatives, and terpenes (Dudareva et al., 2013). Volatile mono- and sesquiterpenes synthesized by plants are known to contribute to the direct and indirect defense against phytophagous insects. While terpenes accumulated in plant tissues can be toxic to biting-chewing and piercing-sucking herbivores, their emission into the surrounding atmosphere can act repellent to these herbivores and attractive toward their natural enemies (Degenhardt et al., 2003; Gershenzon and Dudareva, 2007; Unsicker et al., 2009). To facilitate their role in the antagonistic interactions, terpenes are often produced in specific plant tissues including internal ducts, extracellular cavities, and glandular trichomes (Gershenzon and Dudareva, 2007; Zulak and Bohlmann, 2010). Volatile terpenes, like all other terpenoids, are synthesized from the building blocks isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which in plants originate from two parallel pathways: the mevalonic acid (MVA) pathway and the methylerythritol phosphate (MEP) pathway (Ashour et al., 2010; Hemmerlin et al., 2012). IPP and DMAPP are subsequently utilized by prenyl transferases to form larger *trans* and *cis* prenyl diphosphate intermediates. While geranyl diphosphate (GPP) and its *cis* isomer neryl diphosphate (NPP) are used for monoterpene synthesis, *trans*- and *cis*-farnesyl diphosphate (*E*,*E*- and *Z*,*Z*-FPP) serve as precursors for sesquiterpenes. The formation of terpenes is then catalyzed by terpene synthases (TPSs) which utilize one or several of the prenyl diphosphate substrates, and frequently have the ability to form multiple different terpene products from one prenyl diphosphate substrate (Degenhardt et al., 2009).

While defensive terpene traits are under positive selection pressure to ensure survival in wild plants, it appears that they have been compromised in cultivated crop plants since selective breeding has favored other agronomic traits (Köllner et al., 2008). Domestication and breeding have also resulted in the introduction of strong genetic bottlenecks in cultivated tomato (*Solanum lycopersicum*) which suffers from a higher susceptibility to various pests compared to wild tomato species (Bai and Lindhout, 2007). Although glandular trichomes are present on vegetative tissues of both cultivated and wild tomato species, these differ significantly in their profile of volatile terpenes. In *S. lycopersicum*, they primarily produce a blend of monoterpenes and only small amounts of a few sesquiterpenes (Schilmiller et al., 2009). In contrast, the systematic characterization of glandular trichome-derived terpenes in different accessions of the wild tomato species *Solanum habrochaites* (previously *Lycopersicon hirsutum*), for example, revealed several chemotypes characterized by the dominant formation of distinct blends of sesquiterpenes (Gonzales-Vigil et al., 2012), of which most are not present in *S. lycopersicum*. Numerous studies have shown that glandular trichome-derived terpenes found in *S. lycopersicum* as well as the *S. habrochaites* accessions have repellent and toxic activity against different biting-chewing herbivores including Colorado potato beetle (*Leptinotarsa decemlineata*), tobacco hornworm (*Manduca sexta*), tomato fruitworm (*Helicoverpa zea*), and beet armyworm (*Spodoptera exigua*; Carter et al., 1989a,b; Frelichowski and Juvik, 2001; Kang et al., 2010a,b; Tian et al., 2012; Gutensohn et al., 2014). In contrast, relatively little is known about the potential effects of glandular trichome-derived terpenes in cultivated and wild tomato on piercing-sucking herbivores, such as whiteflies, spider mites, and aphids. The sesquiterpene 7-*epi*-zingiberene and some of its derivatives produced in the glandular trichomes of several *S. habrochaites* accessions were shown to have repellent and/or toxic activity against silverleaf whiteflies (*Bemisia tabaci*) and two spotted-spider mites (*Tetranychus urticae*) (Bleeker et al., 2011a, 2012; Dawood and Snyder, 2020; Zabel et al., 2021). Utilizing a collection of *S. habrochaites* accessions with different glandular trichome-derived terpenes, we have recently identified two groups of sesquiterpenes that affect the potato aphid (*Macrosiphum euphorbiae*) (Wang et al., 2020). *S. habrochaites* accessions producing β-caryophyllene and α-humulene, or (−)-*endo*-α-bergamotene, (+)-α-santalene, and (+)-*endo*-β-bergamotene, respectively, not only reduced the aphid longevity and fecundity significantly, but also had repellent activity against *M. euphorbiae*. Remarkably, by utilizing two tomato trichome mutants, *hairless* and *odorless-2*, that are differently affected in mono- and sesquiterpene production (Kang et al., 2010a,b), we demonstrated that the relatively small amounts of β-caryophyllene and α-humulene in glandular trichomes of cultivated tomato still have some effect on the performance of *M. euphorbiae* (Wang et al., 2021). However, the same analysis also suggested a role of the highly abundant TPS20-derived monoterpenes in the attraction of aphids to cultivated tomato, which was further confirmed using a mixture of pure monoterpenes (Wang et al., 2021).

Macrosiphum euphorbiae is an important agricultural pest that causes economic losses to horticultural crop production including tomato (Tomescu and Negru, 2003). Damage is caused not only by the direct feeding of *M. euphorbiae*, but even more by the transmission of multiple non-persistent and persistent viruses for which these aphids serve as vectors (Blackman and Eastop, 2000; van Emden and Harrington, 2007). This results in reduced crop yield and quality, and often plant death even at low levels of aphid infestation (Lange and Bronson, 1981). Current aphid control strategies utilizing synthetic insecticides are increasingly inefficient due to emerging resistances and avoidance behavior (Silva et al., 2012; Bass et al., 2014; Fray et al., 2014). Thus, introducing the production of the two groups of sesquiterpenes that we have identified in the *S. habrochaites* accessions and that affected choice behavior and performance of *M. euphorbiae* (Wang et al., 2020) into cultivated tomato represents a promising approach toward developing a sustainable aphid control strategy. One possible way for the introduction of these defensive terpene traits is the classical genetic approach by crossing *S. lycopersicum* with

a respective *S. habrochaites* accession, followed by several backcrosses into the cultivated tomato background to create an introgression line carrying the terpene trait from *S. habrochaites*. Indeed, we have used a near isogenic line containing a small *S. habrochaites* introgression on chromosome 8 (van der Hoeven et al., 2000) that carries the genes encoding the respective prenyl transferase and terpene synthase, and was found to produce (−)-*endo*-α-bergamotene, (+)-α-santalene and (+)-*endo*-β-bergamotene (Sallaud et al., 2009). Our assays performed with this line confirmed that introgression of (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene formation into the cultivated tomato background had successfully transferred this defensive trait affecting performance and choice behavior of *M. euphorbiae* (Wang et al., 2020). However, compared to the parental *S. habrochaites* accession the amounts of (−)-*endo*-α-bergamotene, (+)-α-santalene, and (+)-*endo*-β-bergamotene, and in consequence, the effects on *M. euphorbiae* were significantly lower in the near isogenic line (Wang et al., 2020) highlighting a potential limitation of this classical genetic approach. The observed difference in sesquiterpene production is likely due to a higher expression level of MVA and MEP pathway genes (Besser et al., 2009) and a larger storage capacity in the internal cavity of the glandular trichomes (Therezan et al., 2021) in the parental *S. habrochaites* accession compared to the introgression line. The described classical genetic approach limits the introduction of defensive terpene traits into cultivated tomato to the tissue(s) where the respective biosynthetic genes are expressed in wild tomato, for example, glandular trichomes. In contrast, metabolic engineering represents an efficient approach to introduce a respective biosynthetic pathway into a plant tissue and/or species naturally devoid of a terpene compound of interest. Multiple examples for successful metabolic engineering of terpenes in different plant species have been reported (summarized in Lange and Ahkami, 2013; Vickers et al., 2014) with three types of genes being used individually or in combination: MVA/MEP pathway, prenyl transferase, and TPS genes. Two of our previous metabolic engineering studies in tomato (Gutensohn et al., 2013, 2014) demonstrated that co-expression of prenyl transferases and respective TPSs utilizing the produced prenyl diphosphates resulted in the formation of significant amounts of the expected terpenes. To avoid potential negative effects on plant growth and performance recent engineering strategies utilized specific promoters to restrict the expression of terpene biosynthetic genes to particular plant tissues or organs (Lewinsohn et al., 2001; Davidovich-Rikanati et al., 2007, 2008; Morris et al., 2011; Bleeker et al., 2012; Borghi and Xie, 2016).

Aphids as piercing-sucking herbivores use their stylet, a specialized mouthpart, to ultimately access vascular tissue and ingest phloem sap. However, to achieve sustained phloem sap ingestion, the feeding behavior of aphids progresses through several stages (Powell et al., 2006): (i) pre-alighting behavior, (ii) initial plant contact and assessment of surface cues before stylet insertion, (iii) probing of the plant epidermis, (iv) stylet pathway activity, (v) sieve element puncture and salivation, and (vi) phloem acceptance and ingestion. Blocking or hindering these feeding behaviors could lead to the development of a novel aphid control strategy. Terpenes produced in glandular trichomes of tomato are likely only affecting the initial steps of the aphid feeding behavior. However, introducing the formation of defensive terpenes into additional plant tissues *via* metabolic engineering could have the potential to reduce or even eliminate plant tissue penetration by aphids and in consequence virus transmission. In this study, we have developed two multicistronic expression constructs based on the two sesquiterpene traits previously identified in *S. habrochaites* (Wang et al., 2020) that are composed of the coding sequences for prenyl transferases and respective *S. habrochaites* terpene synthases, as well as enhanced green fluorescent protein as a visible marker. Transient expression of both constructs under the epidermis-specific *CER5*-promoter in tomato leaves demonstrated that formation of the two sets of defensive sesquiterpenes, β-caryophyllene/α-humulene and (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene, can be introduced into new tissues in tomato. The epidermis-specific transgene expression and terpene formation were verified by fluorescence microscopy and tissue fractionation with subsequent analysis of terpene profiles, respectively. In addition, the longevity and fecundity of *M. euphorbiae* that fed on these engineered tomato leaves were significantly reduced, demonstrating the efficacy of this novel aphid control strategy.

MATERIALS AND METHODS

Plant Material

Seeds of the tomato (*S. lycopersicum*) trichome mutant *odorless-2* (Kang et al., 2010a) were kindly provided by Dr. Gregg Howe (Michigan State University, MI, United States). Tomato plants used for *Agrobacterium* leaf infiltration and all subsequent analyses were grown from seeds in Sungro® soil mixture (Sun Gro Horticulture, Agawam, MA, United States) in multi-trays (288 cells, 5 ml per cell), and seedlings were transplanted into 4-inch square pots. Plants were grown under a 16-h photoperiod in a climate-controlled growth room (23–25°C, 50–60% relative humidity) without pesticide application.

Cloning and Plasmid Construction

For the cloning of the two multicistronic gene constructs under the control of a tissue-specific promoter, the binary vector pMCS:GW was used (obtained from the Arabidopsis Biological Resource Center, stock # CD3-1933) that contains a multiple cloning site upstream of the Gateway cassette (Michniewicz et al., 2015). To obtain the promoter region of the *AtCER5* gene (At1g51500; Pighin et al., 2004), a 2,614 base pair fragment was amplified by PCR from *Arabidopsis thaliana* genomic DNA using Taq polymerase (GenScript, Piscataway, NJ, United States) and a pair of oligonucleotides each carrying a specific restriction site (*AtCER5*-fwd-*Eco*RI: 5'-CGGAATTCTTTAGTTTGCTTG AGTTCTCATG-3'; *AtCER5*-rev-*Sal*I: 5'-GCGTCGACTGTTTT TGTTTGATCTTGAAAAAGATC-3'). The resulting PCR product was cloned into the vector pMD20 (Takara Bio, United States)

and sequenced to verify its correct amplification. The *AtCER5* promoter fragment was subsequently excised from the pMD20 vector by *Eco*RI and *Sal*I digestion, and then ligated between the *Eco*RI and *Xho*I sites of the pMCS:GW vector.

The two multicistronic gene constructs were obtained by gene synthesis (Twist Bioscience, San Francisco, CA, United States). These gene constructs were flanked by attL-sequences and inserted into the Gateway Entry vector pTwist ENTR (Twist Bioscience, San Francisco, CA, United States). Both of these multicistronic constructs contained three consecutive coding regions encoding a prenyltransferase, a terpene synthase, and enhanced green fluorescent protein, respectively. Stop codons of the prenyl transferase and terpene synthase coding sequences were removed and instead 60-bp wild type F2A sequences (5'- CAGCTGTTGAATTTTGACCTT CTTAAGCTTGCGGGAGACGTCGAGTCCAACCCTGGG CCC-3'; Ryan et al., 1991; Kenneth et al., 2012) inserted such that the three coding regions were linked into one ORF. In addition, the 5'-UTR sequence of the *A. thaliana Farnesyl Diposphate Synthase 2 AtFPPS2* gene (At5g47770) was fused upstream of the start codon of the prenyl transferase coding sequence in both gene constructs.

Each of the two multicistronic gene constructs was transferred from the pTwist ENTR vector into the pMCS:GW vector carrying the *AtCER5* promoter fragment by performing a standard LR recombination reaction. One µl of the destination vector (150 ng/µl) and 0.5 µl of the entry vector (30 ng/µl) were mixed with 6.5 µl TE buffer and 2 µl LR Clonase II Plus enzyme mix (Invitrogen, Thermo Fisher Scientific), and incubated at 25°C for 1 h. After the addition of 1 µl of proteinase K (Invitrogen, Thermo Fisher Scientific), reactions were incubated at 37°C for 10 min. Aliquots of each LR reaction were transformed into competent *Escherichia coli* DH5α cells (Thermo Fisher Scientific) and colonies with recombinant plasmids selected on LB agar plates with kanamycin (100 µg/ml).

Agrobacterium Leaf Infiltration

The resulting binary pMCS:GW vectors carrying the multicistronic gene constructs under the control of the *AtCER5* promoter were introduced into *Agrobacterium tumefaciens* (strain GV3101), and those were subsequently used for transient transformation *via* infiltration of tomato leaves as described previously (Norkunas et al., 2018). Selected *Agrobacterium* clones were grown in LB medium containing kanamycin (100 µg/ml), rifampicin (25 µg/ml), and gentamycin (50 µg/ml) to an OD_{600} of 0.8–1.0. Bacterial cultures were harvested by centrifugation and resuspended in infiltration buffer [10 mm MES-KOH (pH 5.7), 10 mm $MgCl_2$ and 200 µm acetosyringone]. *Agrobacterium* suspensions were adjusted to an OD_{600} of 0.6–0.8 and further incubated at room temperature for 2 h prior to leaf infiltration. *Agrobacterium*-mediated transient transformation was subsequently conducted on the second and older true leaves of 4-week-old *odorless-2* tomato plants. The *Agrobacterium* suspension was injected into leaves from the abaxial side using a 5 ml plastic syringe (without hypodermic needle) thus infiltrating the intercellular spaces of the entire leaves. After the infiltration plants were covered by a humidity dome and continued to grow under a 16-h photoperiod at 23–25°C until further analysis.

RNA Extraction and RT-PCR Analysis

Six days after the *Agrobacterium* infiltration total RNA was isolated from infiltrated tomato leaves as previously described (Eggermont et al., 1996). For the RT-PCR analyses, total RNA was pretreated with RNase-free DNase (New England Biolabs, Ipswich, MA, United States) and cDNA was synthesized using reverse transcriptase (Superscript II, Invitrogen, Carlsbad, CA, United States). To evaluate the expression of both multicistronic gene constructs, the cDNA was subsequently used for PCR utilizing three primer pairs specific for: *AtFPPS2* (*AtFPPS*-fwd: 5'-CGGATCTGAAATCAACCTTCCTCGAC-3'; *AtFPPS*-rev: 5'- CAATGCCTTAACTACCAACCAGGAGC-3'), *ShzFPPS* (*ShzFPPS*-fwd: 5' -CAAATTCACCTCTGACAGTGTCTGC-3'; *ShzFPPS*-rev: 5'-GTGTGTCCACCAAAACGTCTATGCC-3'), and *eGFP* (*eGFP*-fwd: 5'- CGACGTAAACGGCCACAAGTT CA-3'; *eGFP*-rev: 5'- ACTTGTACAGCTCGTCCATGCC-3'). The PCR conditions were as follows: 94°C for 5 min for one cycle, followed by 35 cycles of 95°C for 60 s, 57°C for 60 s and 72°C for 60 s, and a final extension at 72°C for 10 min. The amplification products were separated by agarose gel electrophoresis, stained with GelRed® (Biotium, United States), and analyzed using the ChemiDoc Gel Imaging System and Image Lab 5.1 software (Bio-Rad, Hercules, CA, United States).

Fluorescence Microscopy

Four to five days after *Agrobacterium* infiltration tomato leaves were collected, and cross- and surface sections were prepared. Leaf cross- and surface sections were analyzed with a Zeiss Axio Imager M1 compound microscope (Zeiss, Oberkochen, Germany) equipped with an X-Cite series 120Q fluorescent illuminator. Fluorescence microscopy was performed using a 485/20 excitation filter in combination with a green filter cube, and a 545/25 excitation filter in combination with a red filter cube for the analysis of GFP fluorescence and chlorophyll autofluorescence, respectively. All images were captured using a monochrome AxioCam MRm camera mounted on the Imager M1 microscope and further processed with the AxioVision software. Composite images from GFP and chlorophyll fluorescence microscopy were overlaid using the ImageJ software (National Institutes of Health).[1]

Extraction of Terpenes From Whole Tomato Leaves

Tomato leaves were collected at different time points (0, 3, 6, 9, 12, and 15 days) after the *Agrobacterium* infiltration. Leaves were photographed, and their surface areas were determined using the ImageJ software (National Institutes of Health, see footnote 1). Leaves were stirred in 50 ml of methyl *tert*-butyl ether (MTBE) for 20 min to extract terpenes as described previously (Gutensohn et al., 2013). After removing the leaves,

[1]https://imagej.nih.gov/ij/

extracts were concentrated under a gentle stream of nitrogen gas to a volume of 200 µl and centrifuged for further purification. For the analysis of terpene profiles, extracts were transferred into GC vials and supplemented with 3.33 µg of naphthalene as an internal standard.

Extraction of Terpenes From Isolated Trichomes, Epidermis, Vasculature, and Mesophyll From Tomato Leaves

To further investigate the accumulation of terpenes in different leaf tissues, glandular trichomes, epidermis, vasculature, and mesophyll were isolated from tomato leaves 15 days after *Agrobacterium* infiltration. Type VI glandular trichomes were collected from tomato leaves as described previously (Schilmiller et al., 2009) using a stretched Pasteur pipette under a stereomicroscope (SZ-ST, Olympus, Tokyo, Japan). A total of 200 trichomes from each leaf sample were accumulated into 10 ml of MTBE for the extraction of terpenes. For the isolation of the other leaf tissues, a protocol was adapted from previous studies (Endo et al., 2016; Svozil et al., 2016). Entire tomato leaflets were sandwiched between two layers of clear scotch tape, and the abaxial leaf epidermis was peeled by gently pulling off one tape. The tape with the attached abaxial leaf epidermis was immediately transferred into 10 ml MTBE for terpene extraction. The other layer of tape with the remainder of the leaf was transferred into a 50 ml tube containing 15 ml of enzyme solution [1.00% (w/v) cellulase R-10, 0.25% (w/v) macerozyme R-10, 0.4 M mannitol, 8 mm $CaCl_2$, and 5 mm MES-KOH, pH 5.7]. After a 10 min incubation, the leaf vasculature was isolated by using a dissecting needle and transferred into 10 ml MTBE. The remaining leaf tissue was further digested for an additional 15 min for removing mesophyll cells into the enzyme solution. Subsequently, the tape with the attached adaxial leaf epidermis was transferred into the vial which already contained the tape with the abaxial leaf epidermis. The mesophyll cells isolated by the enzyme treatment were further purified using a 30 µm cell strainer (pluriSelect, Leipzig, Germany), centrifuged at 1,000 *g* for 5 min, and resuspended into 10 ml MTBE. All isolated leaf tissue fractions were allowed to shake in MTBE at 100 rpm for 30 min for terpene extraction. All extracts were concentrated under a gentle stream of nitrogen gas to a volume of 200 µl and centrifuged for further purification. Concentrated and purified extracts were transferred into GC vials and supplemented with 3.33 µg of naphthalene as an internal standard.

Gas Chromatography–Mass Spectrometry Analysis

All extracts from entire tomato leaves and leaf tissue fractions were analyzed by combined gas chromatography–mass spectrometry (GC–MS) using a TRACE 1310 gas chromatograph system linked to a TSQ 8000 Triple Quadrupole mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). Two µl of each sample was injected under a spitless mode, volatilized at 220°C, and then separated on a TraceGOLD TG-5MS GC column (30 m length, 0.25 mm I.D., and 0.25 µm film; Thermo Fisher Scientific, Pittsburgh, PA, United States). The initial column temperature was held at 40°C for 3 min and then ramped at 5°C/min to 120°C, 10°C/min to 180°C, and 20°C/min to 300°C which was maintained for 2 min. The helium carrier gas flow was 1.3 ml/min. All samples were analyzed using the total ion chromatogram (TIC) mode. Individual terpene compounds were identified by comparing their mass spectra (15–300 m/z) with those deposited in the NIST/EPA/NIH Mass Spectral Library (NIST11; National Institute of Standards and Technology NIST, Scientific Instrument Services, Inc., NJ, United States),[2] as well as those reported previously (Sallaud et al., 2009). Terpenes identified in the leaf and leaf tissue extracts were quantified using a previously determined average response factor for sesquiterpenes (Wang et al., 2020) in combination with the internal naphthalene standard.

Aphid Culture

A potato aphid (*M. euphorbiae*) colony was established from apterae collected in the WVU Evansdale Greenhouse (Morgantown, WV, United States). To avoid experience on tomato plants prior to the non-choice assays on *odorless-2* plants, aphids were allowed to reproduce parthenogenetically on potted potato plants in an insect rearing room under a 16-h photoperiod at 20–22°C. The aphid species was confirmed through barcode sequencing.

Aphid Non-Choice Assays

The performance (longevity and fecundity) of *M. euphorbiae* apterae on agroinfiltrated leaves of *odorless-2* tomato plants was determined in a climate-controlled growth room (23–25°C, 50–60% relative humidity, and 16-h photoperiod) as described previously (Eichele-Nelson et al., 2018; Wang et al., 2020). Before the assay apterae (F_0) reared on potato plants were introduced on the leaf surface of *odorless-2* plants. Four days after the introduction three neonate F_1 nymphs (considered as 1 day old) were carefully transferred to the surface of a young *odorless-2* leaf (second or third fully expanded leaf) which had been infiltrated with *Agrobacterium* 2 days earlier, and subsequently enclosed in a clip cage (BioQuip Products, Rancho Dominguez, CA, United States) that was attached to the leaf. Four tomato plants each with three clip cages attached (in total 12 clip cages) were used for each of the treatments (two expression constructs and one control). Over the course of the experiment F_2 nymphs and exuviae were removed daily from the clip cages. The longevity of F_1 nymphs and their fecundity represented by the number of F_2 nymphs in each cage were recorded. The longevity and fecundity of aphids were analyzed by one-way ANOVA followed by multiple comparisons using Tukey's HSD ($\alpha = 0.05$).

[2]https://chemdata.nist.gov/mass-spc/ms-search/

RESULTS

Design of Two Multicistronic Expression Constructs for Epidermis-Specific Engineering of Sesquiterpene Formation

To engineer formation of the two groups of sesquiterpenes with activity against *M. euphorbiae* that we have identified in *S. habrochaites* (Wang et al., 2020) into the epidermis of *S. lycopersicum*, we have designed two multicistronic expression constructs (**Figure 1**) in the binary vector pMCS:GW (Michniewicz et al., 2015). Both constructs were put under the control of the promoter of the *AtCER5* gene (At1g51500; **Figure 1**), encoding a plasma membrane localized ABC transporter involved in the export of cuticular wax, that was shown to direct epidermis-specific expression (Pighin et al., 2004). To achieve high levels of terpene formation, we opted to co-express each of the selected *S. habrochaites* terpene synthases together with a respective prenyl transferase which assures sufficient availability of the required prenyl diphosphate substrate in the correct subcellular compartment. Despite the significant difference in β-caryophyllene/α-humulene formation between cultivated and wild tomato accessions (Wang et al., 2020), in both tomato species these sesquiterpenes are produced by Terpene Synthase 12 (TPS12) utilizing *E,E*-FPP as substrate (Schilmiller et al., 2010; Bleeker et al., 2011b). Thus, for the design of the respective expression construct (**Figure 1**), the coding sequence for the *S. habrochaites Terpene Synthase 12* (*ShTPS12*, GenBank accession JN402389; Bleeker et al., 2011b) was paired with that for the *A. thaliana Farnesyl Diphosphate Synthase 2* (*AtFPPS2*, At4g17190; Keim et al., 2012). The terpene synthase responsible for the formation of (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene in some *S. habrochaites* accessions, Santalene and Bergamotene Synthase (ShSBS), accepts *Z,Z*-FPP as substrate which is synthesized by *cis*-Farnesyl Diphosphate Synthase (ShzFPPS; Sallaud et al., 2009). In contrast to AtFPPS2 and ShTPS12, which are localized in the cytosol, ShzFPPS and ShSBS both carry N-terminal transit peptides that target them toward plastids. Thus, for the design of the second expression construct (**Figure 1**), the coding sequences of *ShSBS* (GenBank accession FJ194970) and *ShzFPPS* (GenBank accession FJ194969) (Sallaud et al., 2009) were paired. To obtain coordinated and stable expression of the multiple transgenes under the control of the *AtCER5* promoter, the open reading frames encoding the terpene synthase and prenyl transferase in both expression constructs (**Figure 1**) were linked by a short 60 bp nucleotide sequence encoding the foot-and-mouth disease virus 2A oligopeptide (F2A; Ryan et al., 1991; Kenneth et al., 2012). This F2A sequence represents a self-processing peptide that *via* a ribosome skipping mechanism during the translation process leads to the separation between the upstream polypeptide ending with the C-terminal 2A sequence and the next translation product downstream (Ryan and Flint, 1997; Donnelly et al., 2001). As third part, the coding region of *enhanced green fluorescent protein* (*eGFP*) was added to both multicistronic expression constructs (**Figure 1**) and was likewise linked by an F2A sequence to the 3' end of the terpene synthases, *ShTPS12* and *ShSBS*, respectively. In summary, both constructs (**Figure 1**), the pC5-FTG construct

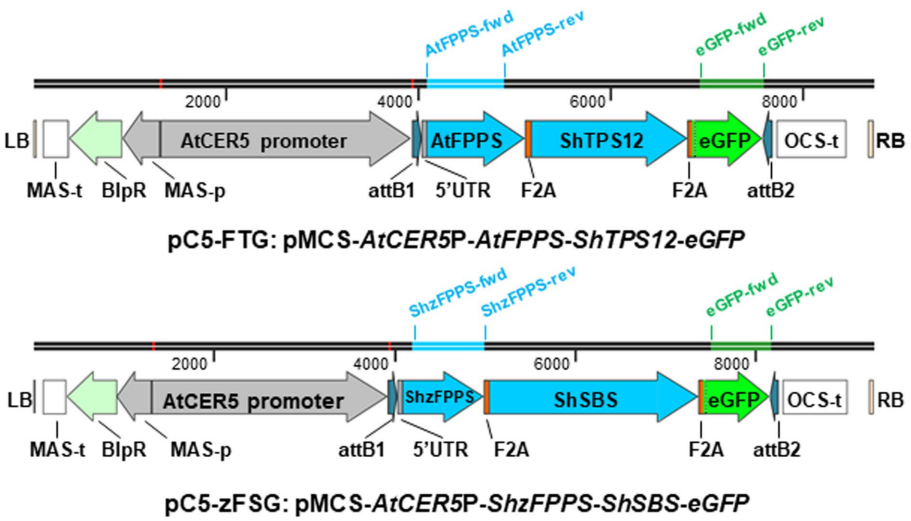


FIGURE 1 | Schematic representation of multicistronic expression constructs for epidermis-specific engineering of sesquiterpene formation. Two expression constructs were designed within the T-DNA, indicated by the left (LB) and right (RB) borders, of the binary pMCS vector. Both synthetic expression constructs are put under the control of an *AtCER5* promoter sequence and inserted between the gateway attachments sites (attB1 and attB2). Each of the multicistronic expression constructs contains the coding sequences for three proteins: a prenyl transferase (AtFPPS or ShzFPPS), a terpene synthase (ShTPS12 or ShSBS), and enhanced green fluorescent protein (eGFP). The three individual coding sequences within both multicistronic expression constructs are linked by a short nucleotide sequence encoding the self-processing foot-and-mouth disease virus 2A oligopeptide (F2A). Other elements located within the T-DNA are the octopine synthase terminator (OCS-t), mannopine synthase promoter (MAS-p), and terminator (MAS-t) and phosphinothricin acetyltransferase (BlpR). The bars above the pC5-FTG and pC5-zFSG constructs indicate their size (in base pairs) and the location of the three primer pairs used for RT-PCR analysis of *AtFPPS*, *ShzFPPS*, and *eGFP* expression.

(*AtCER5*P-*AtFPPS-ShTPS12-eGFP*) and the pC5-zFSG construct (*AtCER5*P-*ShzFPPS-ShSBS-eGFP*), will result in the formation of three separate proteins upon expression *in planta*: a prenyl transferase and terpene synthase pair catalyzing the synthesis of the desired sesquiterpenes, as well as eGFP that will serve as a visual marker of the epidermis-specific expression.

Transient Expression of the Multicistronic Constructs in Tomato Leaves

We tested the function of the newly designed multicistronic expression constructs by infiltrating leaves of cultivated tomato with *Agrobacterium* carrying the pMCS binary vector with the inserted pC5-FTG and pC5-zFSG constructs (**Figure 1**), respectively, as well as the empty pMCS vector as a negative control. For these transient transformation assays, we used the tomato trichome mutant *odorless-2* (Kang et al., 2010a) since it is deficient in the formation of TPS20-derived monoterpenes and TPS12-derived sesquiterpenes that are naturally found in *S. lycopersicum* trichomes and thus could interfere with the analysis of the engineered sesquiterpenes. To determine the expression of both multicistronic constructs upon transient transformation of the tomato leaves reverse transcription-PCR (RT-PCR) analyses were performed using primer pairs (**Figure 1**) specific for the first coding region (*AtFPPS* and *ShzFPPS*) and the third coding region (*eGFP*) in the pC5-FTG and pC5-zFSG constructs, respectively. An *AtFPPS* specific 788 base pair cDNA fragment could be amplified from leaves infiltrated with *Agrobacterium* carrying the pC5-FTG construct (**Figure 2A**).

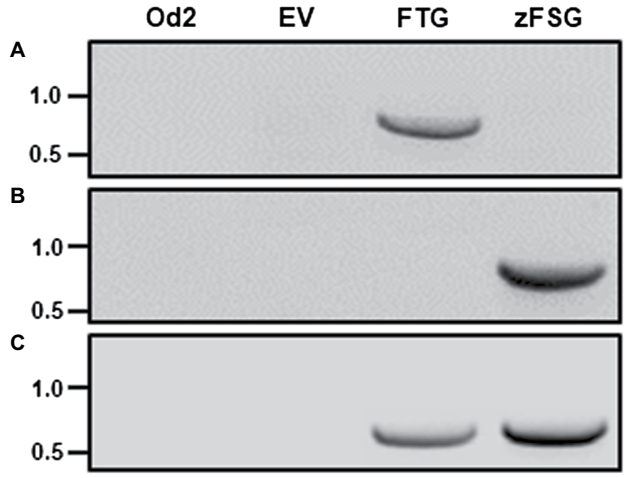


FIGURE 2 | Transient expression of the multicistronic constructs in tomato leaves. Leaves of the tomato *odorless-2* mutant were infiltrated with *Agrobacterium* carrying the pC5-FTG construct, the pC5-zFSG construct, or the empty pMCS vector (EV). Transcript levels in *Agrobacterium* infiltrated leaves and *odorless-2* control (Od2) leaves were analyzed by RT-PCR utilizing *AtFPPS* **(A)**, *ShzFPPS* **(B)**, and *eGFP* **(C)** specific primer pairs (see **Figure 1** for location). The amplification products indicating *AtFPPS* (788 bp), *ShzFPPS* (794 bp), and *eGFP* (656 bp) expression were separated by agarose gel electrophoresis (size marker in kb indicated with each panel).

However, this *AtFPPS* fragment was not detected with untransformed *odorless-2* control leaves or leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector and the pC5-zFSG construct, respectively. In contrast, a *ShzFPPS*-specific 794 base pair cDNA fragment was only amplified from leaves infiltrated with *Agrobacterium* carrying the pC5-zFSG construct (**Figure 2B**) but was absent from the *odorless-2* and empty vector controls as well as leaves infiltrated with *Agrobacterium* carrying the pC5-FTG construct. The *eGFP*-specific 656 base pair cDNA fragment could be amplified from leaves infiltrated with *Agrobacterium* carrying either the pC5-FTG or pC5-zFSG construct (**Figure 2C**) but was not found with the *odorless-2* and empty vector controls. The fact that transcripts of the prenyl transferases and *eGFP* representing the first and last coding region in both constructs were detected by RT-PCR upon the transient transformation of tomato leaves (**Figure 2**) suggests that the entire multicistronic constructs were expressed under the control of the *AtCER5* promoter.

Epidermis-Specific Expression of the Multicistronic Constructs

While the RT-PCR analysis (**Figure 2**) in general demonstrated expression of the multicistronic constructs in tomato leaves upon transient transformation, the tissue specificity of their expression under the control of the *AtCER5* promoter remained to be shown. Toward this goal, we performed confocal fluorescence microscopy of tomato leaves that had been infiltrated with *Agrobacterium* carrying the empty pMCS vector, pC5-FTG construct, and pC5-zFSG construct, respectively, to determine the tissue-specific accumulation of eGFP which is encoded in both multicistronic constructs. When cross-sections of tomato leaves were analyzed, GFP fluorescence was exclusively detected in both epidermal layers of leaves transiently transformed with the pC5-FTG and pC5-zFSG constructs (**Figure 3A**), while no respective fluorescence was observed with the empty vector control. Moreover, the GFP fluorescence did not overlap with the chlorophyll fluorescence detected in the chloroplast containing parenchyma cells of the leaf cross-sections (**Figure 3A**). In addition, we analyzed surface sections of the transiently transformed tomato leaves by fluorescence microscopy to further verify the expression of *eGFP* in epidermis cells. Upon the transient expression of the pC5-FTG and pC5-zFSG constructs in tomato leaves, GFP fluorescence could be observed in the cytosolic rim of the epidermal pavement cells (**Figure 3B**). The results of these fluorescence microscopy analyses not only provide further evidence that the entire pC5-FTG and pC5-zFSG constructs including *eGFP* are expressed, but also indicate that their expression under the *AtCER5* promoter is indeed restricted to the epidermis of transformed tomato leaves.

Sesquiterpene Formation in the Epidermis of Tomato Leaves Transiently Expressing the pC5-FTG and pC5-zFSG Constructs

To determine if the transient expression of the multicistronic constructs, both encoding pairs of prenyl transferases and terpene synthases, in the leaf epidermis, resulted in the formation

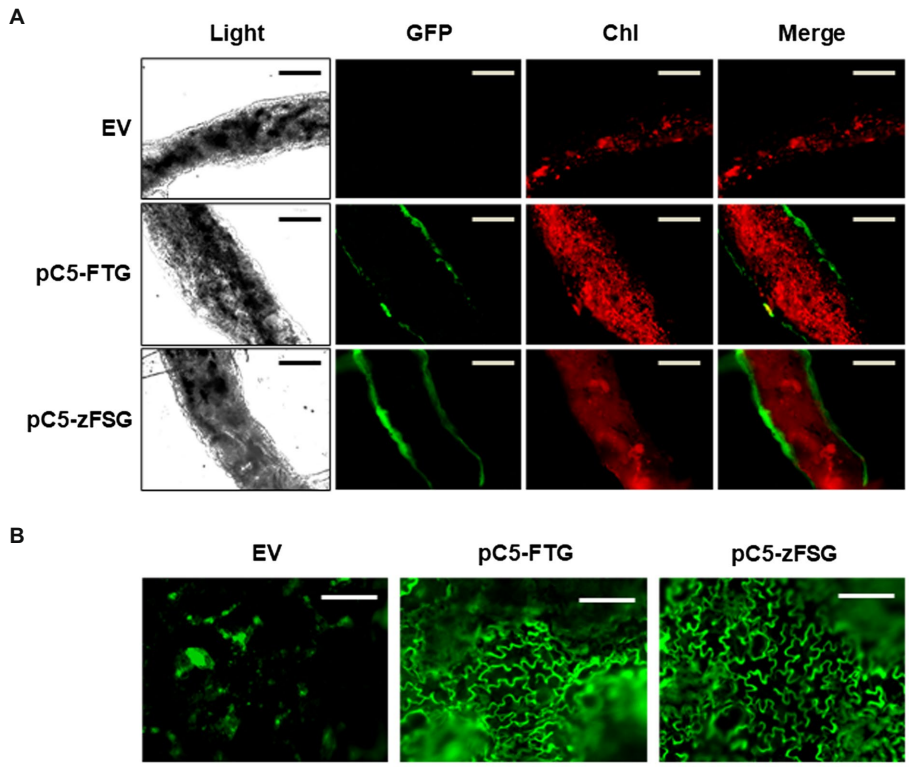


FIGURE 3 | Tissue-specific expression of the multicistronic constructs. Cross-sections **(A)** and surface sections **(B)** of tomato leaves infiltrated with *Agrobacterium* carrying the pC5-FTG construct, the pC5-zFSG construct, or the empty pMCS vector (EV) were analyzed by light and confocal laser scanning microscopy. Panels show fluorescence of green fluorescent protein (GFP) and chlorophyll autofluorescence (Chl). Scale bars represent 100 μm.

of the expected sesquiterpenes, tomato leaves were extracted with methyl *tert*-butyl ether (MTBE) 15 days after *Agrobacterium* infiltration. The subsequent analysis of the leaf extracts by combined gas chromatography–mass spectrometry (GC–MS) demonstrated that leaves of the *odorless-2* tomato mutant expressing the pC5-FTG construct had accumulated the expected ShTPS12 products β-caryophyllene and α-humulene (**Figure 4A; Supplementary Figure 1**). A similar analysis of leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector found no β-caryophyllene and α-humulene accumulation (**Figure 4C**) which is in line with the previous characterization of the *odorless-2* tomato mutant (Kang et al., 2010a; Wang et al., 2021) showing the absence of these two sesquiterpenes in this trichome mutant. In contrast, the analysis of tomato leaves expressing the pC5-zFSG construct revealed a different profile of accumulated terpenes (**Figure 4B; Supplementary Figure 2**) including (−)-*endo*-α-bergamotene, (+)-α-santalene, (−)-*exo*-α-bergamotene, (−)-*epi*-β-santalene, and (+)-*endo*-β-bergamotene. These five sesquiterpenes have been observed previously in *in vitro* enzyme assays as well as in a transgenic tobacco line as products of ShSBS when *Z,Z*-FPP was provided as substrate by ShzFPPS (Sallaud et al., 2009).

The quantitative analysis of the terpene accumulation in tomato leaves expressing the pC5-FTG and pC5-zFSG constructs showed that the ShTPS12- and ShSBS-derived sesquiterpene products, respectively, could be detected for the first time 6 days after the *Agrobacterium* infiltration (**Figures 4D,E; Supplementary Table 1**). Subsequently, the amounts of the sesquiterpene products in the tomato leaves continued to increase until 12 days after the *Agrobacterium* infiltration and appeared to remain constant afterward (**Figures 4D,E; Supplementary Table 1**). Remarkably, the total amount of sesquiterpenes produced after 15 days in leaves expressing the plastid localized ShzFPPS and ShSBS were 3.1-fold higher than in leaves expressing the cytosolic AtFPPS and ShTPS12 (**Figures 4D,E; Supplementary Table 1**).

To further verify the tissue specificity of the novel metabolic engineering approach described here, we studied the accumulation of sesquiterpenes in different tissues of tomato leaves transiently expressing the pC5-FTG and pC5-zFSG constructs. Fifteen days after *Agrobacterium* infiltration tomato leaves were separated into epidermis, mesophyll and vasculature fractions which were subsequently extracted with MTBE and analyzed for their terpene content by GC–MS. The ShTPS12-derived sesquiterpenes β-caryophyllene and α-humulene were found in the epidermis and mesophyll fractions of tomato leaves expressing the pC5-FTG construct (**Table 1**), while they were absent in the vasculature. Likewise, three of the ShSBS-derived

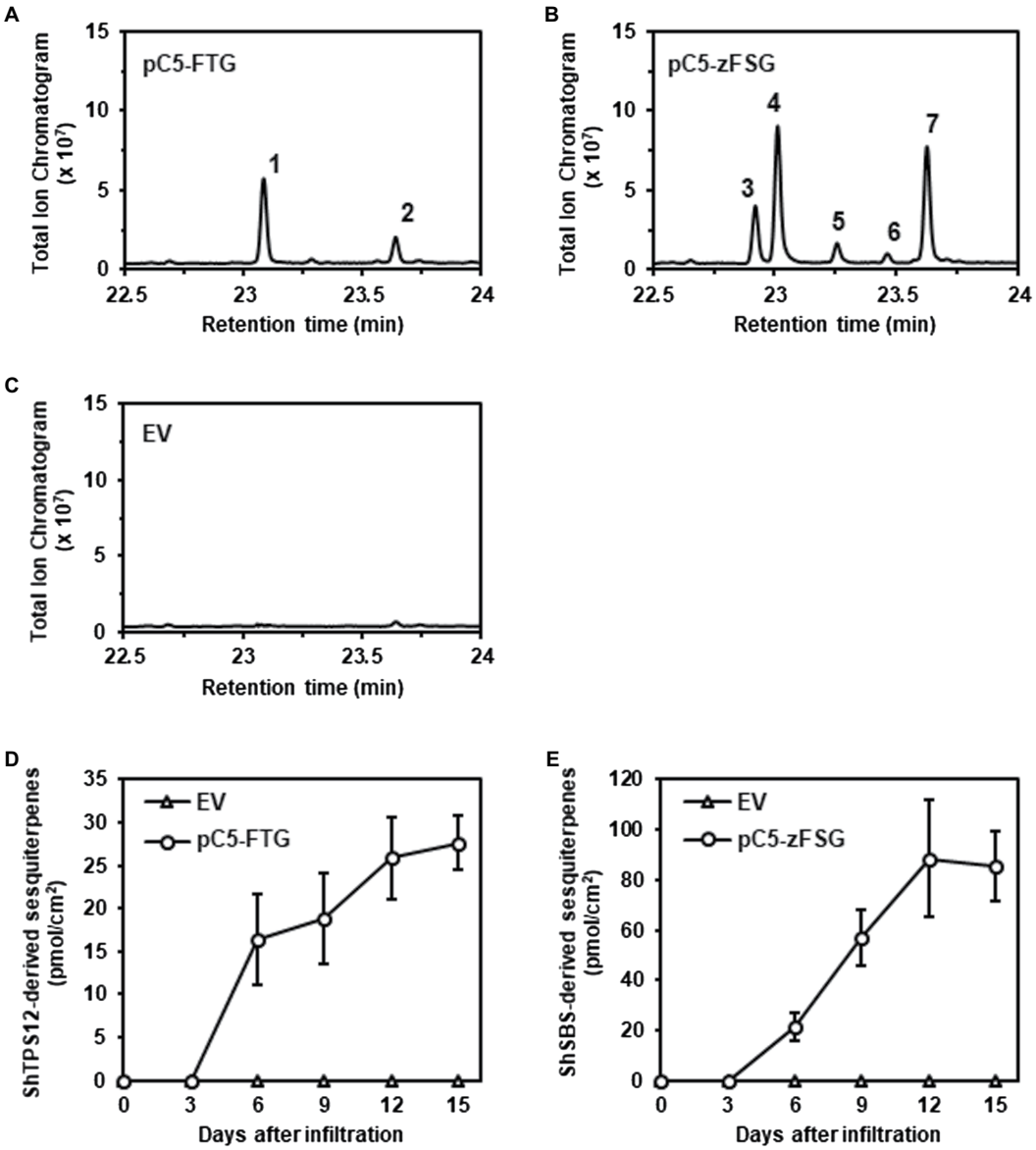


FIGURE 4 | Accumulation of ShTPS12- and ShSBS-derived sesquiterpenes in tomato leaves expressing the multicistronic constructs. Terpenes were extracted from tomato leaves infiltrated with *Agrobacterium* carrying the pC5-FTG construct **(A)**, the pC5-zFSG construct **(B)**, or the empty pMCS vector (EV) **(C)** and were analyzed by GC–MS (total ion chromatograms are shown). ShTPS12-derived sesquiterpenes: 1, β-caryophyllene; 2, α-humulene. ShSBS-derived sesquiterpenes: 3, (−)-*endo*-α-bergamotene; 4, (+)-α-santalene; 5, (−)-*exo*-α-bergamotene; 6, (−)-*epi*-β-santalene; and 7, (+)-*endo*-β-bergamotene. The total amounts (pmol/cm^2 leaf area) of ShTPS12-derived **(D)** and ShSBS-derived **(E)** sesquiterpenes were determined in tomato leaves at different time points after the *Agrobacterium* infiltration. Data are means ± SEM ($n = 3$).

sesquiterpenes, (−)-*endo*-α-bergamotene, (+)-α-santalene, and (+)-*endo*-β-bergamotene, were found in the epidermis fraction of leaves expressing the pC5-zFSG construct (**Table 1**), while only (+)-α-santalene and (+)-*endo*-β-bergamotene were detected in the mesophyll fraction and no ShSBS-derived sesquiterpenes were present in the vasculature of these leaves. In addition to the epidermis, mesophyll, and vasculature fractions, we also analyzed the terpene content of glandular trichomes collected from tomato leaves expressing the pC5-FTG and pC5-zFSG constructs, however, did not observe any accumulation of ShTPS12- and ShSBS-derived sesquiterpenes, respectively (**Table 1**). In summary, these analyses revealed that the vast majority of the ShTPS12- and ShSBS-derived sesquiterpenes, 92.99 and 92.24%, respectively, accumulate in the epidermis (**Table 1**) of the transiently transformed tomato leaves, thus indicating that expression of the pC5-FTG and pC5-zFSG constructs under the control of the *AtCER5* promoter indeed results in the epidermis-specific production of the engineered sesquiterpenes.

TABLE 1 | Accumulation of sesquiterpenes in different tissues of tomato leaves transiently transformed with the pC5-FTG and pC5-zFSG expression constructs.

Terpenes	pC5-FTG				pC5-zFSG			
	Trichomes*	Epidermis‡	Vasculature‡	Mesophyll‡	Trichomes*	Epidermis‡	Vasculature‡	Mesophyll‡
β-caryophyllene	nd	7.68 (±1.46)	nd	0.81 (±0.22)	nd	nd	nd	nd
α-humulene	nd	2.9 (±0.73)	nd	nd	nd	nd	nd	nd
(−)-*endo*-α-bergamotene	nd	nd	nd	nd	nd	6.67 (±1.22)	nd	nd
(+)-α-santalene	nd	nd	nd	nd	nd	13.95 (±1.29)	nd	1.46 (±0.11)
(+)-*endo*-β-bergamotene	nd	nd	nd	nd	nd	10.06 (±0.87)	nd	1.12 (±0.08)

**Terpenes were extracted from 200 glandular trichomes collected from tomato leaves.*

‡Epidermis, vasculature, and mesophyll fractions were prepared from the same leaves, and the amounts of terpenes (±SEM, n = 3) extracted from these fractions were normalized by the leaf surface area (pmol/cm²).

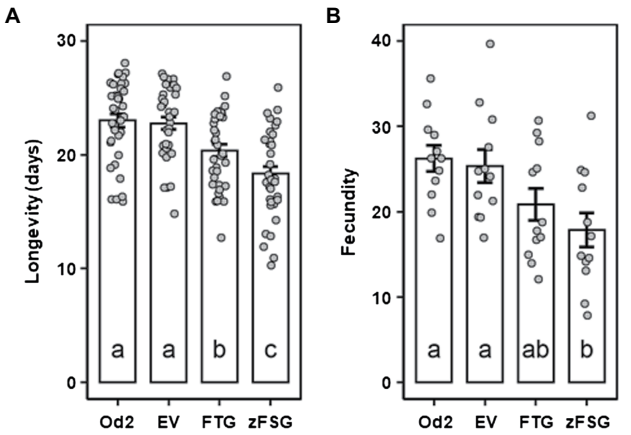


FIGURE 5 | Longevity and fecundity of potato aphids on tomato leaves expressing the multicistronic constructs. Longevity **(A)** and fecundity **(B)** of *Macrosiphum euphorbiae* on leaves of the *odorless-2* tomato mutant (Od2), and leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector (EV), the pC5-FTG construct, or the pC5-zFSG construct. Newly emerged aphid nymphs were arrested onto tomato leaves 2 days after *Agrobacterium* infiltration. Values for longevity ($n = 36$) and fecundity ($n = 12$) are presented as means ± SEM. Values of different leaf samples were compared by ANOVA and Tukey's HSD test, and different letters indicate significant differences ($p < 0.05$).

Engineered Sesquiterpene Formation in the Epidermis Affects the Longevity and Fecundity of Aphids

As a first approach to characterize the potential of the sesquiterpene formation engineered into the leaf epidermis to affect the potato aphid (*M. euphorbiae*), we performed non-choice assays utilizing tomato leaves that transiently express the pC5-FTG and pC5-zFSG constructs. Newly emerged *M. euphorbiae* nymphs were reared in clip cages on the surface of tomato leaves that previously have been infiltrated with *Agrobacterium* carrying the pC5-FTG and pC5-zFSG constructs or the empty pMCS vector control, and their longevity and fecundity (represented by the number of offspring) were determined. Compared to the non-infiltrated *odorless-2* control, infiltration of leaves with *Agrobacterium* carrying the empty pMCS vector did not significantly affect longevity ($t = 0.201$, $p = 0.997$) or fecundity ($t = 0.368$, $p = 0.983$) of *M. euphorbiae*. In contrast, the longevity of *M. euphorbiae* on tomato leaves expressing the pC5-FTG construct (20.39 ± 0.55 days), characterized by β-caryophyllene and α-humulene production in their epidermis (**Figure 4; Table 1**), was significantly decreased ($t = 5.420$, $p = 0.001$) compared to that on leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector (22.78 ± 0.53 days; **Figure 5A**). The longevity of *M. euphorbiae* (**Figure 5A**) on tomato leaves expressing the pC5-zFSG construct (18.33 ± 0.65 days), which accumulated the ShSBS-derived sesquiterpenes in their epidermis (**Figure 4; Table 1**), was also significantly decreased ($t = 2.678$, $p = 0.048$) compared to that on leaves infiltrated with the empty pMCS vector control, and even further decreased compared to that on leaves expressing the pC5-FTG construct ($t = 2.742$, $p = 0.035$). Similar effects as observed for the longevity were also found for the fecundity of *M. euphorbiae* (**Figure 5B**) on tomato leaves expressing the pC5-FTG and pC5-zFSG constructs, while their fecundity on leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector was not significantly affected ($t = 0.368$, $p = 0.983$). The number of *M. euphorbiae* offspring was significantly reduced ($t = 3.160$, $p = 0.015$) on leaves expressing the pC5-zFSG construct (17.83 ± 2.01 nymphs) compared to leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector (25.33 ± 1.94 nymphs; **Figure 5B**). A similar trend toward a reduced number of offspring (20.83 ± 1.86 nymphs) was observed with *M. euphorbiae* on leaves expressing the pC5-FTG construct (**Figure 5B**), although their fecundity was not significantly different to that of aphids on leaves infiltrated with *Agrobacterium* carrying the empty pMCS vector ($t = 1.679$, $p = 0.347$) or the pC5-zFSG construct ($t = 1.481$, $p = 0.457$).

DISCUSSION

It is well known that wild tomato species, such as *S. habrochaites*, have glandular trichome-derived resistance traits against numerous pests (Simmons and Gurr, 2005). In particular, some terpenes produced in the glandular trichomes of wild tomato

accessions have been shown to act repellent and/or toxic against pests (Carter et al., 1989a,b; Frelichowski and Juvik, 2001; Bleeker et al., 2009, 2011a). In our previous study (Wang et al., 2020), we have identified two groups of *S. habrochaites* accessions producing β-caryophyllene/α-humulene and (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene, respectively, that significantly reduced the longevity and fecundity of *M. euphorbiae*, and also had repellent activity against the aphids. Thus, introducing these defensive sesquiterpene traits identified in *S. habrochaites* into cultivated tomato represents a logical step toward developing a novel aphid control strategy. One avenue toward achieving this goal is the classical genetic approach by crossing *S. lycopersicum* and respective *S. habrochaites* accessions, followed by backcrosses into the cultivated tomato background to obtain an introgression line carrying the *S. habrochaites* sesquiterpene trait. A near isogenic tomato line with a small *S. habrochaites* introgression carrying the *ShzFPPS* and *ShSBS* genes was previously isolated (van der Hoeven et al., 2000) and found to produce (−)-*endo*-α-bergamotene, (+)-α-santalene, and (+)-*endo*-β-bergamotene (Sallaud et al., 2009). While our assays demonstrated that introgression of the (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene formation into the cultivated tomato background indeed affected the performance and choice behavior of *M. euphorbiae*, it became obvious that the sesquiterpene levels and the resulting effects on *M. euphorbiae* were significantly lower in the introgression line (Wang et al., 2020). In contrast to the limitations of the genetic approach, metabolic engineering has been shown to offer an efficient approach to introduce the biosynthesis of terpene compounds of interest into plants (Lange and Ahkami, 2013; Vickers et al., 2014) that in addition can be steered toward specific tissues through the choice of respective promoters. To engineer high levels of terpene formation in many cases multiple biosynthetic genes have to be introduced into the host plant including MVA/MEP pathway, prenyl transferase, and terpene synthase genes. However, the introduction of multiple individual transgenes and their combination in one plant line through subsequent crosses is a time-consuming process. In addition, the stacking of several transgenes that are all expressed under the identical type of promoter bears the risk of gene silencing. In contrast, the utilization of the viral self-processing 2A sequences circumvents these problems and allows the co-expression of multiple genes under the control of a single promoter (de Felipe et al., 2006).

Here, we designed two multicistronic expression constructs, each composed of the coding sequences for a prenyl transferase, a respective *S. habrochaites* terpene synthase, and enhanced green fluorescent protein linked by short nucleotide sequences encoding the foot-and-mouth disease virus 2A self-processing oligopeptide (**Figure 1**). Both constructs are under the control of the *AtCER5* promoter that directs epidermis-specific gene expression (Pighin et al., 2004). Infiltration of tomato leaves with *Agrobacterium* carrying the pC5-FTG and pC5-zFSG constructs resulted in the transient expression of all three genes included in each expression construct. The RT-PCR analyses (**Figure 2**) demonstrated the expression of the first coding region, *AtFPPS* and *ShzFPPS*, respectively, and the third coding region, *eGFP*, from each of the two multicistronic constructs. Moreover, the expression of *eGFP* was further verified through fluorescence microscopy that detected GFP fluorescence in the epidermis (**Figure 3**). The formation of the expected sesquiterpenes, β-caryophyllene and α-humulene (**Figure 4A**), and (−)-*endo*-α-bergamotene, (+)-α-santalene, (−)-*exo*-α-bergamotene, (−)-*epi*-β-santalene, and (+)-*endo*-β-bergamotene (**Figure 4B**), upon leaf infiltration with *Agrobacterium* carrying the pC5-FTG and pC5-zFSG construct, respectively, provided further evidence for the expression of the prenyl transferases and terpene synthases included in these constructs. Our observation that the total sesquiterpene amounts produced in leaves expressing the plastid localized ShzFPPS and ShSBS were higher than in leaves expressing the cytosolic AtFPPS and ShTPS12 (**Supplementary Table 1**) is in line with earlier studies showing that the plastidic MEP pathway is often metabolically more active than the cytosolic MVA pathway (Ashour et al., 2010; Hemmerlin et al., 2012). A similar co-expression system based on viral 2A sequences has previously been used to engineer the formation of a precursor of artemisinin, a plant derived sesquiterpene lactone highly effective in the treatment of malaria, into tobacco leaves (van Herpen et al., 2010). Transient expression of a multicistronic construct, containing the reading frames for amorpha-4,11-diene synthase, 3-hydroxy-3-methylglutaryl-CoA reductase, and farnesyl diphosphate synthase linked by 2A sequences, in *Nicotiana benthamiana* leaves resulted in the formation of the artemisinin precursor amorpha-4,11-diene. Moreover, a viral 2A sequence system has been used for the co-expression of the carotenoid biosynthetic genes encoding phytoene synthase and carotene desaturase in rice endosperm to obtain an improved version of the β-carotene producing Golden Rice (Ha et al., 2010; Jeong et al., 2017). Another study (Møldrup et al., 2012) utilized the viral 2A sequence co-expression system to engineer the six-step benzylglucosinolate pathway from *A. thaliana* into *Nicotiana tabacum*, thus converting the resulting tobacco lines into a trap crop for the pest diamondback moth (*Plutella xylostella*).

In vegetative parts of plants, the formation of terpenes is often restricted to specific tissues, such as glandular trichomes on the leaf surface (Gershenzon and Dudareva, 2007; Zulak and Bohlmann, 2010). Therefore, the goal of this study was to test if the formation of terpenes with activity against aphids could be engineered into new vegetative tissues, specifically the epidermis, where terpenes are naturally not found. The fluorescence microscopy analyses (**Figure 3**) of tomato leaves infiltrated with *Agrobacterium* carrying the pC5-FTG and pC5-zFSG constructs detected GFP fluorescence exclusively in both epidermal layers, thus confirming the tissue specificity of the expression under the control of the *AtCER5* promoter. Moreover, the analysis of the terpene content (**Table 1**) in different tissue fractions of the tomato leaves transiently expressing the multicistronic constructs revealed that the vast majority of the ShTPS12- and ShSBS-derived sesquiterpenes is indeed produced in the epidermis. These results suggest that sufficient pools of IPP and DMAPP are available in the cytosol and plastids of these epidermis cells that can serve as substrates for the cytosolic AtFPPS2 and ShTPS12, and the

plastid localized ShzFPPS and ShSBS, respectively. Although the minor amounts of sesquiterpenes (**Table 1**) found in the mesophyll fractions of leaves expressing the pC5-FTG and pC5-zFSG constructs are likely the consequence of contamination by epidermis cells, we cannot exclude that there might be a symplastic transport of some sesquiterpenes produced in the epidermis cells toward neighboring mesophyll cells. While to the best of our knowledge this is the first report on the metabolic engineering of terpene formation in the epidermis of leaves, there are examples of natural terpene formation in epidermis cells. The flowers of *Clarkia breweri* are strongly scented and one of the major volatile compounds emitted is the monoterpene *S*-linalool. *In situ* localization studies of the *S*-linalool synthase transcripts demonstrated that this terpene synthase is mainly expressed in the epidermal cell layers of the *C. breweri* flower petals (Dudareva et al., 1996). Likewise, both epidermal layers of the petals in rose (*Rosa* x *hybrida*) flowers were found to produce, accumulate, and emit a number of monoterpenes including geraniol, citronellol, and nerol (Bergougnoux et al., 2007).

Remarkably, the formation of the ShTPS12- and ShSBS-derived sesquiterpenes in the epidermis of tomato leaves expressing the pC5-FTG and pC5-zFSG constructs significantly affected the longevity and fecundity of *M. euphorbiae* (**Figure 5**). Recently, we observed similar effects on the longevity and fecundity of *M. euphorbiae* when the aphid performance was tested on the leaf surface of *S. habrochaites* accessions producing β-caryophyllene/α-humulene and (−)-*endo*-α-bergamotene/(+)-α-santalene/(+)-*endo*-β-bergamotene in their glandular trichomes (Wang et al., 2020). The fact that the effect on the performance of *M. euphorbiae* was less severe on the engineered leaves with the epidermis-specific sesquiterpene formation (**Figure 5**) compared to that of the glandular trichome-derived sesquiterpenes in *S. habrochaites* accessions (Wang et al., 2020) could be due to a difference in the amounts of sesquiterpenes produced. On the other hand, the reduced longevity and fecundity of *M. euphorbiae* observed in this study are clearly due to the sesquiterpene formation engineered into the leaf epidermis, since we have used the *odorless-2* mutant that is deficient in the formation of the glandular trichome-derived terpenes normally found in tomato leaves (Kang et al., 2010a; Wang et al., 2021). This result of our study is in line with previous studies which have revealed that the host plant selection by aphids is not only affected by glandular trichomes, but also by factors located in the epidermis including epicuticular lipids, cell wall barriers, and the presence or absence of certain metabolites that serve as gustatory cues upon probing (Alvarez et al., 2006; Schwarzkopf et al., 2013).

CONCLUSION

In this study, we have taken a novel approach toward developing a sustainable aphid control strategy that specifically considers the feeding behavior of these piercing-sucking pests. By utilizing the viral 2A sequence system, we co-expressed two pairs of prenyl transferases and *S. habrochaites* terpene synthases under the control of the epidermis-specific *AtCER5* promoter. This metabolic engineering approach resulted not only in the exclusive accumulation of the desired sesquiterpenes in the epidermis of tomato leaves, but also significantly affected the aphid performance. Thus, the metabolic engineering of sesquiterpenes into the leaf epidermis introduced an additional layer of defense against aphids, besides the glandular trichome-derived terpenes naturally present in cultivated and wild tomato species that in particular affect the aphid choice behavior. Future metabolic engineering approaches could now also test the effects of sesquiterpene formation in other tissues relevant to aphid feeding, such as the mesophyll and phloem, by expressing the newly designed multicistronic constructs under the control of respective tissue-specific promoters. While the outcome of our transient engineering study and the aphid bioassays highlighted the potential and efficacy of the tissue-specific metabolic engineering approach described here, further detailed characterization of respective stable transgenic tomato lines and aphids feeding on them, including electrical penetration graph analysis, will be required to verify which specific stages of the aphid feeding behavior are affected by the sesquiterpene formation in the epidermis and other tissues.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MG and Y-LP conceived and designed the research. FW performed the experiments. FW, Y-LP, and MG analyzed the data. FW and MG wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

This project was supported by the Agricultural and Food Research Initiative competitive grant number 2018-67014-28092 from the USDA National Institute of Food and Agriculture to MG and Y-LP. This study was also partially supported by the West Virginia Agricultural and Forestry Experiment Station Hatch Projects WVA00730 to MG and WVA00024 to Y-LP. In addition, the work of MG was supported by the Ray Marsh and Arthur Pingree Dye Professorship.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Gregg Howe (Michigan State University) for kindly providing seeds of the *odorless-2* mutant. The authors would like to acknowledge Dr. Scott Bowdridge

(Animal and Nutritional Sciences, WVU) for assistance with the fluorescence microscopy, the WVU Genomics Core Facility for assistance with sequencing, and the staff at the WVU Evansdale Greenhouse for support in cultivating the plant material used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.793313/full#supplementary-material

Supplementary Material 1 | Mass spectra of ShTPS12-derived sesquiterpenes. Mass spectra of β-caryophyllene **(A)** and α-humulene **(B)** corresponding to peaks 1 and 2 (**Figure 4**), respectively, are shown. Terpenes were extracted from tomato leaves infiltrated with Agrobacterium carrying the pC5-FTG construct and were analyzed by GC–MS. Mass spectra were scanned at a range of 30–500 (m/z) after electron ionization at 70 eV.

Supplementary Material 2 | Mass spectra of ShSBS-derived sesquiterpenes. Mass spectra of (–)-**endo**-α-bergamotene **(A)**, (+)-α-santalene **(B)**, (–)-**exo**-α-bergamotene **(C)**, (–)-**epi**-β-santalene **(D)**, and (+)-**endo**-β-bergamotene **(E)** corresponding to peaks 3–7 (**Figure 4**), respectively, are shown. Terpenes were extracted from tomato leaves infiltrated with Agrobacterium carrying the pC5-zFSG construct and were analyzed by GC–MS. Mass spectra were scanned at a range of 30–500 (m/z) after electron ionization at 70 eV.

REFERENCES

Alvarez, A. E., Tjallingii, W. F., Garzo, E., Vleeshouwers, V., Dicke, M., and Vosman, B. (2006). Location of resistance factors in the leaves of potato and wild tuber-bearing *Solanum* species to the aphid *Myzus persicae*. *Entomol. Exp. Appl.* 121, 145–157. doi: 10.1111/j.1570-8703.2006.00464.x

Ashour, M., Wink, M., and Gershenzon, J. (2010). Biochemistry of terpenoids: monoterpenes, sesquiterpenes, and diterpenes. *Annu. Plant Rev.* 40, 258–303. doi: 10.1002/9781444320503.ch5

Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Bot.* 100, 1085–1094. doi: 10.1093/aob/mcm150

Bass, C., Puinean, A. M., Zimmer, C. T., Denholm, I., Field, L. M., Foster, S. P., et al. (2014). The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect Biochem. Mol. Biol.* 51, 41–51. doi: 10.1016/j.ibmb.2014.05.003

Bergougnoux, V., Caissard, J. C., Jullien, F., Magnard, J. L., Scalliet, G., Cock, J. M., et al. (2007). Both the adaxial and abaxial epidermal layers of the rose petal emit volatile scent compounds. *Planta* 226, 853–866. doi: 10.1007/s00425-007-0531-1

Besser, K., Harper, A., Welsby, N., Schauvinhold, I., Slocombe, S., Li, Y., et al. (2009). Divergent regulation of terpenoid metabolism in the trichomes of wild and cultivated tomato species. *Plant Physiol.* 149, 499–514. doi: 10.1104/pp.108.126276

Blackman, R. L., and Eastop, V. F. (2000). *Aphids on the World's Crops: An Identification and Information Guide*. 2nd Edn. Chichester: Wiley.

Bleeker, P. M., Diergaarde, P. J., Ament, K., Guerra, J., Weidner, M., Schütz, S., et al. (2009). The role of specific tomato volatiles in tomato-whitefly interaction. *Plant Physiol.* 151, 925–935. doi: 10.1104/pp.109.142661

Bleeker, P. M., Diergaarde, P. J., Ament, K., Schütz, S., Johne, B., Dijkink, J., et al. (2011a). Tomato-produced 7-epizingiberene and *R*-curcumene act as repellents to whiteflies. *Phytochemistry* 72, 68–73. doi: 10.1016/j.phytochem.2010.10.014

Bleeker, P. M., Mirabella, R., Diergaarde, P. J., VanDoorn, A., Tissier, A., Kant, M. R., et al. (2012). Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20124–20129. doi: 10.1073/pnas.1208756109

Bleeker, P. M., Spyropoulou, E. A., Diergaarde, P. J., Volpin, H., De Both, M. T. J., Zerbe, P., et al. (2011b). RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from *Solanum lycopersicum* and *Solanum habrochaites* trichomes. *Plant Mol. Biol.* 77, 323–336. doi: 10.1007/s11103-011-9813-x

Borghi, M., and Xie, D. Y. (2016). Tissue-specific production of limonene in *Camelina sativa* with the Arabidopsis promoters of genes *BANYULS* and *FRUITFULL*. *Planta* 243, 549–561. doi: 10.1007/s00425-015-2425-y

Carter, C. D., Gianfagna, T. J., and Sacalis, J. N. (1989a). Sesquiterpenes in glandular trichomes of wild tomato species and toxicity to the Colorado potato beetle. *J. Agric. Food Chem.* 37, 1425–1428. doi: 10.1021/jf00089a048

Carter, C. D., Sacalis, J. N., and Gianfagna, T. J. (1989b). Zingiberene and resistance to Colorado potato beetle in *Lycopersicon hirsutum* f. *hirsutum*. *J. Agric. Food Chem.* 37, 206–210. doi: 10.1021/jf00085a047

Davidovich-Rikanati, R., Lewinsohn, E., Bar, E., Iijima, Y., Pichersky, E., and Sitrit, Y. (2008). Overexpression of the lemon basil alpha-zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit. *Plant J.* 56, 228–238. doi: 10.1111/j.1365-313X.2008.03599.x

Davidovich-Rikanati, R., Sitrit, Y., Tadmor, Y., Iijima, Y., Bilenko, N., Bar, E., et al. (2007). Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat. Biotechnol.* 25, 899–901. doi: 10.1038/nbt1312

Dawood, M. H., and Snyder, J. C. (2020). The alcohol and epoxy alcohol of zingiberene, produced in trichomes of wild tomato, are more repellent to spider mites than zingiberene. *Front. Plant Sci.* 11:35. doi: 10.3389/fpls.2020.00035

de Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C., and Ryan, M. D. (2006). *E unum pluribus*: multiple proteins from a self-processing polyprotein. *Trends Biotechnol.* 24, 68–75. doi: 10.1016/j.tibtech.2005.12.006

Degenhardt, J., Gershenzon, J., Baldwin, I. T., and Kessler, A. (2003). Attracting friends to feast on foes: engineering terpene emission to make crop plants more attractive to herbivore enemies. *Curr. Opin. Biotechnol.* 14, 169–176. doi: 10.1016/S0958-1669(03)00025-9

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. doi: 10.1016/j.phytochem.2009.07.030

Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., et al. (2001). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* 82, 1013–1025. doi: 10.1099/0022-1317-82-5-1013

Dudareva, N., Cseke, L., Blanc, V. M., and Pichersky, E. (1996). Evolution of floral scent in *Clarkia*: novel patterns of *S*-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell* 8, 1137–1148. doi: 10.1105/tpc.8.7.1137

Dudareva, N., Klempien, A., Muhlemann, J. K., and Kaplan, I. (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* 198, 16–32. doi: 10.1111/nph.12145

Eggermont, K., Goderis, I. J., and Broekaert, W. F. (1996). High-throughput RNA extraction from plant samples based on homogenisation by reciprocal shaking in the presence of a mixture of sand and glass beads. *Plant Mol. Biol. Rep.* 14, 273–279. doi: 10.1007/BF02671663

Eichele-Nelson, J., DeSutter, T., Wick, A. F., Harmon, E. L., and Harmon, J. P. (2018). Salinity improves performance and alters distribution of soybean aphids. *Environ. Entomol.* 47, 875–880. doi: 10.1093/ee/nvy072

Endo, M., Shimizu, H., and Araki, T. (2016). Rapid and simple isolation of vascular, epidermal and mesophyll cells from plant leaf tissue. *Nat. Protoc.* 11, 1388–1395. doi: 10.1038/nprot.2016.083

Fray, L. M., Leather, S. R., Powell, G., Slater, R., McIndoe, E., and Lind, R. J. (2014). Behavioral avoidance and enhanced dispersal in neonicotinoid-resistant *Myzus perisicae* (Sulzer). *Pest Manag. Sci.* 70, 88–96. doi: 10.1002/ps.3530

Frelichowski, J. E., and Juvik, J. A. (2001). Sesquiterpene carboxylic acids from a wild tomato species affect larval feeding behavior and survival of *Helicoverpa zea* and *Spodoptera exigua* (Lepidoptera: Noctuidae). *J. Econ. Entomol.* 94, 1249–1259. doi: 10.1603/0022-0493-94.5.1249

Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414. doi: 10.1038/nchembio.2007.5

Gonzales-Vigil, E., Hufnagel, D. E., Kim, J., Last, R. L., and Barry, C. S. (2012). Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Plant J.* 71, 921–935. doi: 10.1111/j.1365-313X.2012.05040.x

Gutensohn, M., Nguyen, T. T., McMahon, R. D. III, Kaplan, I., Pichersky, E., and Dudareva, N. (2014). Metabolic engineering of monoterpene biosynthesis in tomato fruits via introduction of the non-canonical substrate neryl diphosphate. *Metab. Eng.* 24, 107–116. doi: 10.1016/j.ymben.2014.05.008

Gutensohn, M., Orlova, I., Nguyen, T. T., Davidovich-Rikanati, R., Ferruzzi, M. G., Sitrit, Y., et al. (2013). Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J.* 75, 351–363. doi: 10.1111/tpj.12212

Ha, S. H., Liang, Y. S., Jung, H., Ahn, M. J., Suh, S. C., Kweon, S. J., et al. (2010). Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotechnol. J.* 8, 928–938. doi: 10.1111/j.1467-7652.2010.00543.x

Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* 51, 95–148. doi: 10.1016/j.plipres.2011.12.001

Jeong, Y. S., Ku, H. K., Kim, J. K., You, M. K., Lim, S. H., Kim, J. K., et al. (2017). Effect of codon optimization on the enhancement of the β-carotene content in rice endosperm. *Plant Biotechnol. Rep.* 11, 171–179. doi: 10.1007/s11816-017-0440-0

Kang, J. H., Liu, G., Shi, F., Jones, A. D., Beaudry, R. M., and Howe, G. A. (2010a). The tomato *odorless-2* mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.* 154, 262–272. doi: 10.1104/pp.110.160192

Kang, J. H., Shi, F., Jones, A. D., Marks, M. D., and Howe, G. A. (2010b). Distortion of trichome morphology by the *hairless* mutation of tomato affects leaf surface chemistry. *J. Exp. Bot.* 61, 1053–1064. doi: 10.1093/jxb/erp370

Keim, V., Manzano, D., Fernández, F. J., Closa, M., Andrade, P., Caudepón, D., et al. (2012). Characterization of Arabidopsis FPS isozymes and FPS gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds. *PLoS One* 7:e49109. doi: 10.1371/journal.pone.0049109

Kenneth, K. Y., Aguilar, K., Tsai, J., Galimidi, R., Gnanapragasam, P., Yang, L., et al. (2012). Use of mutated self-cleaving 2A peptides as a molecular rheostat to direct simultaneous formation of membrane and secreted anti-HIV immunoglobulins. *PLoS One* 7:e50438. doi: 10.1371/journal.pone.0050438

Köllner, T. G., Held, M., Lenk, C., Hiltpold, I., Turlings, T. C. J., Gershenzon, J., et al. (2008). A maize (*E*)-β-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell* 20, 482–494. doi: 10.1105/tpc.107.051672

Lange, B. M., and Ahkami, A. (2013). Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes - current status and future opportunities. *Plant Biotechnol. J.* 11, 169–196. doi: 10.1111/pbi.12022

Lange, W. H., and Bronson, L. (1981). Insect pests of tomatoes. *Annu. Rev. Entomol.* 26, 345–371. doi: 10.1146/annurev.en.26.010181.002021

Lewinsohn, E., Schalechet, F., Wilkinson, J., Matsui, K., Tadmor, Y., Nam, K. H., et al. (2001). Enhanced levels of the aroma and flavor compound *S*-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. *Plant Physiol.* 127, 1256–1265. doi: 10.1104/pp.010293

Michniewicz, M., Frick, E. M., and Strader, L. (2015). Gateway-compatible tissue-specific vectors for plant transformation. *BMC. Res. Notes* 8:63. doi: 10.1186/s13104-015-1010-6

Møldrup, M. E., Geu-Flores, F., de Vos, M., Olsen, C. E., Sun, J., Jander, G., et al. (2012). Engineering of benzylglucosinolate in tobacco provides proof-of-concept for dead-end trap crops genetically modified to attract *Plutella xylostella* (diamondback moth). *Plant Biotechnol. J.* 10, 435–442. doi: 10.1111/j.1467-7652.2011.00680.x

Morris, W. L., Ducreux, L. J., Shepherd, T., Lewinsohn, E., Davidovich-Rikanati, R., Sitrit, Y., et al. (2011). Utilization of the MVA pathway to produce elevated levels of the sesquiterpene α-copaene in potato tubers. *Phytochemistry* 72, 2288–2293. doi: 10.1016/j.phytochem.2011.08.023

Norkunas, K., Harding, R., Dale, J., and Dugdale, B. (2018). Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods* 14:71. doi: 10.1186/s13007-018-0343-2

Pighin, J. A., Zheng, H., Balakshin, L. J., Goodman, I. P., Western, T. L., Jetter, R., et al. (2004). Plant cuticular lipid export requires an ABC transporter. *Science* 306, 702–704. doi: 10.1126/science.1102331

Powell, G., Tosh, C. R., and Hardie, J. (2006). Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu. Rev. Entomol.* 51, 309–330. doi: 10.1146/annurev.ento.51.110104.151107

Ryan, M. D., and Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.* 78, 699–723. doi: 10.1099/0022-1317-78-4-699

Ryan, M. D., King, A. M. Q., and Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J. Gen. Virol.* 72, 2727–2732. doi: 10.1099/0022-1317-72-11-2727

Sallaud, C., Rontein, D., Onillon, S., Jabès, F., Duffé, P., Giacalone, C., et al. (2009). A novel pathway for sesquiterpene biosynthesis from *Z,Z*-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell* 21, 301–317. doi: 10.1105/tpc.107.057885

Schilmiller, A. L., Miner, D. P., Larson, M., McDowell, E., Gang, D. R., Wilkerson, C., et al. (2010). Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.* 153, 1212–1223. doi: 10.1104/pp.110.157214

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

Schwarzkopf, A., Rosenberger, D., Niebergall, M., Gershenzon, J., and Kunert, G. (2013). To feed or not to feed: plant factors located in the epidermis, mesophyll, and sieve elements influence pea aphid's ability to feed on legume species. *PLoS One* 8:e75298. doi: 10.1371/journal.pone.0075298

Silva, A. X., Jander, G., Samaniego, H., Ramsey, J. S., and Figueroa, C. C. (2012). Insecticide resistance mechanisms in the green peach aphid *Myzus persicae* (Hemiptera: Aphididae) I: a transcriptomic survey. *PLoS One* 7:e36366. doi: 10.1371/journal.pone.0036366

Simmons, A. T., and Gurr, G. M. (2005). Trichomes of *Lycopersicon* species and their hybrids: effects on pests and natural enemies. *Agric. For. Entomol.* 7, 265–276. doi: 10.1111/j.1461-9555.2005.00271.x

Svozil, J., Gruissem, W., and Baerenfaller, K. (2016). Meselect - a rapid and effective method for the separation of the main leaf tissue types. *Front. Plant Sci.* 7:1701. doi: 10.3389/fpls.2016.01701

Therezan, R., Kortbeek, R., Vendemiatti, E., Legarrea, S., de Alencar, S. M., Schuurink, R. C., et al. (2021). Introgression of the sesquiterpene biosynthesis from *Solanum habrochaites* to cultivated tomato offers insights into trichome morphology and arthropod resistance. *Planta* 254:11. doi: 10.1007/s00425-021-03651-y

Tian, D., Tooker, J., Peiffer, M., Chung, S. H., and Felton, G. W. (2012). Role of trichomes in defense against herbivores: comparison of herbivore response to *woolly* and *hairless* trichome mutants in tomato (*Solanum lycopersicum*). *Planta* 236, 1053–1066. doi: 10.1007/s00425-012-1651-9

Tomescu, A., and Negru, G. (2003). An overview on fungal diseases and pests on the field tomato crops in Romania. *Acta Hortic.* 613, 259–266. doi: 10.17660/ActaHortic.2003.613.41

Unsicker, S. B., Kunert, G., and Gershenzon, J. (2009). Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr. Opin. Plant Biol.* 12, 479–485. doi: 10.1016/j.pbi.2009.04.001

van der Hoeven, R. S., Monforte, A. J., Breeden, D., Tanksley, S. D., and Steffens, J. C. (2000). Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirstum*. *Plant Cell* 12, 2283–2294. doi: 10.1105/tpc.12.11.2283

van Emden, H. F., and Harrington, R. (2007). *Aphids as Crop Pests*. Cambridge, UK: CAB International.

van Herpen, T. W. J. M., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H. J., and Beekwilder, J. (2010). *Nicotiana benthamiana* as a production platform for artemisinin precursors. *PLoS One* 5:e14222. doi: 10.1371/journal.pone.0014222

Vickers, C. E., Bongers, M., Liu, Q., Delatte, T., and Bouwmeester, H. (2014). Metabolic engineering of volatile isoprenoids in plants and microbes. *Plant Cell Environ.* 37, 1753–1775. doi: 10.1111/pce.12316

Wang, F., Park, Y. L., and Gutensohn, M. (2020). Glandular trichome-derived sesquiterpenes of wild tomato accessions (*Solanum habrochaites*) affect aphid performance and feeding behavior. *Phytochemistry* 180:112532. doi: 10.1016/j.phytochem.2020.112532

Wang, F., Park, Y. L., and Gutensohn, M. (2021). Glandular trichome-derived mono- and sesquiterpenes of tomato have contrasting roles in the interaction with the potato aphid *Macrosiphum euphorbiae*. *J. Chem. Ecol.* 47, 204–214. doi: 10.1007/s10886-021-01243-4

Zabel, S., Brandt, W., Porzel, A., Athmer, B., Bennewitz, S., Schäfer, P., et al. (2021). A single cytochrome P450 oxidase from *Solanum habrochaites* sequentially oxidizes 7-*epi*-zingiberene to derivatives toxic to whiteflies and various microorganisms. *Plant J.* 105, 1309–1325. doi: 10.1111/tpj.15113

Zulak, K. G., and Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. *J. Integr. Plant Biol.* 52, 86–97. doi: 10.1111/j.1744-7909.2010.00910.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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fpls.2021.796248

Genome-Wide Identification of R2R3-MYB Transcription Factors: Discovery of a "Dual-Function" Regulator of Gypenoside and Flavonol Biosynthesis in *Gynostemma pentaphyllum*

Ding Huang[1,2†], Ruhong Ming[1†], Shiqiang Xu[3,4†], Shaochang Yao[1,2], Liangbo Li[1], Rongshao Huang[1] and Yong Tan[1,2*]*

[1] College of Pharmacy, Guangxi University of Chinese Medicine, Nanning, China, [2] Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine, Guangxi University of Chinese Medicine, Nanning, China, [3] Guangdong Provincial Engineering and Technology Research Center for Conservation and Utilization of the Genuine Southern Medicinal Resources, Guangzhou, China, [4] Guangdong Provincial Key Laboratory of Crops Genetics and Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China

The *R2R3-MYB* gene family participates in several plant physiological processes, especially the regulation of the biosynthesis of secondary metabolites. However, little is known about the functions of *R2R3-MYB* genes in *Gynostemma pentaphyllum* (*G. pentaphyllum*), a traditional Chinese medicinal herb that is an excellent source of gypenosides (a class of triterpenoid saponins) and flavonoids. In this study, a systematic genome-wide analysis of the *R2R3-MYB* gene family was performed using the recently sequenced *G. pentaphyllum* genome. In total, 87 *R2R3-GpMYB* genes were identified and subsequently divided into 32 subgroups based on phylogenetic analysis. The analysis was based on conserved exon–intron structures and motif compositions within the same subgroup. Collinearity analysis demonstrated that segmental duplication events were majorly responsible for the expansion of the *R2R3-GpMYB* gene family, and Ka/Ks analysis indicated that the majority of the duplicated *R2R3-GpMYB* genes underwent purifying selection. A combination of transcriptome analysis and quantitative reverse transcriptase-PCR (qRT-PCR) confirmed that *Gynostemma pentaphyllum* myeloblastosis 81 (*GpMYB81*) along with genes encoding gypenoside and flavonol biosynthetic enzymes exhibited similar expression patterns in different tissues and responses to methyl jasmonate (MeJA). Moreover, GpMYB81 could bind to the promoters of *Gynostemma pentaphyllum* farnesyl pyrophosphate synthase 1 (*GpFPS1*) and *Gynostemma pentaphyllum* chalcone synthase (*GpCHS*), the key structural genes of gypenoside and flavonol biosynthesis, respectively, and activate their expression. Altogether, this study highlights a novel transcriptional regulatory mechanism that suggests that GpMYB81 acts as a "dual-function" regulator of gypenoside and flavonol biosynthesis in *G. pentaphyllum*.

Keywords: *R2R3-MYB* gene family, *Gynostemma pentaphyllum*, gypenoside, flavonol, transcription factor

OPEN ACCESS

Edited by:
Xueqing Fu,
Shanghai Jiao Tong University, China

Reviewed by:
Hexin Tan,
Second Military Medical University, China
Praveen Awasthi,
Institute of Plant Molecular Biology, Centre for Biology, Academy of Sciences of the Czech Republic (ASCR), Czechia

***Correspondence:**
Ding Huang
hdh016@126.com
Yong Tan
xjty321@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *16 October 2021*
Accepted: *02 December 2021*
Published: *05 January 2022*

Citation:
Huang D, Ming R, Xu S, Yao S, Li L, Huang R and Tan Y (2022) Genome-Wide Identification of R2R3-MYB Transcription Factors: Discovery of a "Dual-Function" Regulator of Gypenoside and Flavonol Biosynthesis in Gynostemma pentaphyllum. Front. Plant Sci. 12:796248. doi: 10.3389/fpls.2021.796248

INTRODUCTION

The myeloblastosis (MYB) transcriptional regulators comprise one of the largest and important families in the plant kingdom (Riechmann et al., 2000). A classical characteristic of MYB proteins is that they contain conserved MYB DNA-binding domain repeats in the N-terminus. The conserved MYB domain is composed of one to four continuous and non-redundant imperfect sequence repeats, designated as R1, R2, and R3 according to their similarity to the c-MYB protein. Each conserved MYB DNA-binding domain repeat is approximately 50 amino acids in length and encodes three α-helices (Dubos et al., 2010). Depending on the number of MYB repeats in the MYB domain, MYB transcription factors (TFs) can be subdivided into R2R3-MYB (2R-MYB), R1R2R3-MYB (3R-MYB), 4R-MYB (containing four R1/R2 repeats), and the MYB-related subfamily (containing a single or a partial MYB repeat) (Stracke et al., 2001). In the R2R3-MYB family, the last two α-helices of each MYB repeat form a helix–turn–helix (HTH) structure and the third α-helix of R2 and R3 repeats are essential for DNA binding, allowing direct contact and insertion into the DNA major groove (Gabrielsen et al., 1991; Jia et al., 2004).

The R2R3-MYB TFs constitute the largest MYB subfamily in plants and the majority of them can specifically recognize the MYB-core sequence [(C/T)NGTT(G/T)] and AC-rich element [A/CCC(T/A)A(C/A)C/G] (Zhu et al., 2015; Millard et al., 2019). As important regulatory proteins involved in several crucial biological processes, the number of identified and characterized R2R3-MYB TFs in plants is continuously increasing. R2R3-MYB TFs play vital roles in plant growth and development, respond to various biotic and abiotic stresses, and regulate secondary metabolism, especially those affecting nutrition and medicinal components or appearance and quality traits (Huang D. et al., 2019; Wang et al., 2019; He et al., 2020). For example, *Ruby1* and *Ruby2* encode R2R3-MYB TFs and form a gene cluster that shows a regulatory subfunctionalization in anthocyanin biosynthesis in citrus (Huang et al., 2018). Similarly, kiwifruit R2R3-MYB TF MYB7 acts as a positive regulator to activate the promoter of the key gene lycopene beta-cyclase (*AdLCY-b*) in the carotenoid biosynthetic pathway (Ampomah-Dwamena et al., 2019). The genes *AtMYB11*, *AtMYB12*, and *AtMYB111* from subgroup 7 of the *Arabidopsis thaliana* (*A. thaliana*) *R2R3-MYB* gene family control flavonol accumulation in different parts of the *A. thaliana* seedling (Stracke et al., 2007).

Extensive study on the *R2R3-MYB* gene family members in horticultural and crop plants has increased our understanding of their functions and transcriptional regulatory mechanism. However, characteristics of this gene family in *Gynostemma pentaphyllum* (*G. pentaphyllum*), a traditional Chinese medicinal herb named jiaogulan, have not yet been declassified. As an economically valuable medicinal and edible plant, jiaogulan tea has been commercialized globally. Gypenosides are a major class of triterpenoids with a dammarane-type carbon skeleton in *G. pentaphyllum*, which exert beneficial effects on human health (Park et al., 2014; Shen et al., 2020; Wang et al., 2020). In addition, flavonoids can be divided into flavonols, flavones, flavonones, and anthocyanidins that play important roles in medicine and hygiene due to their high antioxidant activity (Chen and Chen, 2013; Wang et al., 2018). Gypenosides are the primary active components of *G. pentaphyllum*. In contrast, flavonols extracted from *G. pentaphyllum* contain mainly kaempferol and quercetin derivatives, which were considered major contributors to the beneficial properties of *G. pentaphyllum* (Xie et al., 2011). In a recent study, gypenoside biosynthetic genes, including *farnesyl pyrophosphate synthase* (*FPS*), *squalene synthase* (*SS*), *squalene epoxidase* (*SE*), *2,3-oxidosqualene cyclase* (*OSC*), and *cytochrome P450* (*CYP450*), have been well elucidated (Huang et al., 2021). In addition, structural genes of the flavonoid metabolic pathway are best understood at present (Nabavi et al., 2020). However, it remains poorly understood whether the *R2R3-MYB* gene family members are involved in the regulation of both gypenoside and flavonol biosynthesis in *G. pentaphyllum*.

The recently published *G. pentaphyllum* genome sequence provides a convenient tool to identify and characterize the *R2R3-GpMYB* gene family (Huang et al., 2021). In this study, we performed a genome-wide identification of *R2R3-MYB* genes in *G. pentaphyllum* and screened 87 *R2R3-GpMYB* genes. Next, a comprehensive analysis including phylogenetic relationship, gene structure, conserved domains and motifs, chromosomal location, gene duplication, and collinearity was performed. Based on the weighted gene co-expression network analysis (WGCNA) and expression pattern response to methyl jasmonate (MeJA) treatment, GpMYB81 was suggested as a "dual-function" TF that can regulate both gypenoside and flavonol biosynthesis. In addition, GpMYB81 could bind to the promoters of *GpFPS1* and *GpCHS* genes and activate their transcription, thus opening up the possibility for improving the yield of both gypenosides and flavonols in *G. pentaphyllum* through metabolic engineering.

MATERIALS AND METHODS

Plant Materials and Methyl Jasmonate Treatment

Plant materials were cultivated in a fully controlled climate room of Guangxi University of Chinese Medicine (Nanning, China), with a 16-h light/8-h dark cycle at 24°C temperature. *G. pentaphyllum* seedlings were culture in Hoagland's nutrient solutions. For MeJA treatment, 6-week-old *G. pentaphyllum* seedlings were cultured in Hoagland's nutrient solutions with 100 μm MeJA. For quantification of gene expression using quantitative reverse transcriptase-PCR (qRT-PCR), *G. pentaphyllum* seedlings were collected at 0, 6, 12, and 24 h after MeJA treatment; the leaves of three seedlings were randomly selected to form three biological replicates. All the plant samples were frozen with liquid nitrogen and stored at −80°C.

Identification of *Gynostemma pentaphyllum R2R3-MYB* Family Genes

The Hidden Markov Model (HMM) file of MYB DNA-binding domain (PF00249), obtained from the Pfam database,[1] was used as the query for HMM search using HMMER 3.0 (Finn et al., 2011) to identify *MYB* genes from *G. pentaphyllum* genome

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

with default parameters. To ensure the presence of two MYB DNA-binding domain repeats, candidate MYB protein sequences were further examined using the Simple Modular Architecture Research Tool (SMART) database.[2] Finally, a manual inspection was performed to confirm the reliability of our results.

Sequence Analysis and Phylogenetic Analysis of *R2R3-MYB* Genes

The exon/intron structure of all the *R2R3-GpMYB* genes was displayed using the TBtools software (Chen et al., 2020) based on gene annotation data in general feature format 3 (GFF3) format. The conserved motif of R2R3-GpMYB protein sequences was predicted using a motif-based sequence analysis tool Multiple Expectation maximizations for Motif Elicitation (MEME) version 5.1.1 program (Bailey et al., 2009). The parameters were as follows: maximum motif number of 25; other options were set to default.

Multiple sequence alignments of *G. pentaphyllum* and *A. thaliana* R2R3-MYB protein sequences were performed using molecular evolutionary genetics analysis (MEGA) version 10.1.7. Subsequently, a maximum likelihood (ML) phylogenetic tree was constructed using the FastTree version 2.1.1 (Price et al., 2009). The ML phylogenetic tree was visualized by the Interactive Tree of Life (iTOL) (Letunic and Bork, 2021). Additionally, an ML phylogenetic tree including full length of 87 R2R3-GpMYB protein sequences was constructed using the same methods. Finally, a combination of the phylogenetic tree, conserved domains, gene structures, and conserved motifs of R2R3-GpMYB protein sequences was visualized using the Tbtools software (Chen et al., 2020).

Genomic Localization and Gene Duplication of *R2R3-GpMYB* Genes

The physical positions of the identified *R2R3-GpMYB* genes were mapped to 11 chromosomes of the *G. pentaphyllum* genome using the Tbtools software (Chen et al., 2020). The orthologous *MYB* genes between *G. pentaphyllum* and *A. thaliana* as well as those between *G. pentaphyllum* and *C. sativus* were identified using OrthoVenn2 (Xu et al., 2019). Multiple Collinearity Scan toolkit (MCScanX) was used to analyze the gene duplication events with default parameters (Wang et al., 2012). Non-synonymous (ka) and synonymous (ks) substitutions of each duplicated *R2R3-MYB* gene were calculated using the Tbtools software (Chen et al., 2020).

Ribonucleic Acid Isolation and Quantitative Reverse Transcriptase-PCR Analysis

Total RNA isolation and qRT-PCR analysis were performed using the methods described by Xu et al. (2020). qRT-PCR was performed using the LightCycler 96 System (Roche, United States). The *GpActin* gene was used for quantitative gene expression normalization (Xu et al., 2020; Huang et al., 2021). The $2^{-\Delta\Delta Ct}$ analysis method was adopted to calculate the relative gene expression. Primer information is given in **Supplementary Table 1**.

Yeast One-Hybrid Assays

Yeast one-hybrid (Y1H) assays were performed as described previously (Huang et al., 2018). To construct the prey vector, the open reading frame (ORF) of the *GpMYB81* gene was cloned into the pGADT7 plasmid. To construct the bait vectors, the fragments of *GpFPS1* and *GpCHS* promoters (about 1.5 Kb) were cloned into the pAbAi plasmid. Yeast cells were grown for 3 days at 30°C on synthetic dropout (SD)/-Ura/-Leu medium added with or without aureobasidin A (AbA). Primer information is given in **Supplementary Table 1**.

Dual-Luciferase Assays

A dual-luciferase (LUC) reporter assay was conducted in *Nicotiana benthamiana* leaves according to the method described previously (Huang et al., 2018). To construct the effector vector, the ORF of the *GpMYB81* gene was cloned into the pK2GW7 plasmid. An empty vector of pK2GW7 was used as a negative control. To construct the reporter vectors, the fragments of *GpFPS1* and *GpCHS* promoters (about 1.5 Kb) were cloned into the pGreenII 0800-LUC plasmid. Fluorescence was detected using an *in vivo* imaging system (NightShade LB 985, Germany). Primer information is given in **Supplementary Table 1**.

RESULTS

Identification and Characterization of *Gynostemma pentaphyllum R2R3-MYB* Family Genes

In total, 248 candidate genes were originally obtained from the *G. pentaphyllum* genome as encoding proteins that contained MYB domains. After removing the redundant transcripts, all the candidates were further verified *via* Pfam, HMMscan, and SMART. As a result, 87 *R2R3-GpMYB* genes were identified in *G. pentaphyllum*. Among these, 86 *R2R3-GpMYB* genes were mapped to 11 chromosomes and renamed from *GpMYB1* to *GpMYB86* according to their location on the chromosomes. However, one exception was observed, an *R2R3-GpMYB* gene renamed *GpMYB87* was not located on any chromosome.

The amino acid number of R2R3-GpMYB proteins ranged from 126 to 556, with theoretical isoelectric point and molecular weight values ranging from 4.97 (*GpMYB14*) to 9.87 (*GpMYB5*) and 14.68 (*GpMYB5*) to 61.53 (*GpMYB68*) kDa, respectively. To provide possible clues for functional studies, we predicted their subcellular locations. The results indicated that all the R2R3-GpMYB proteins were located in the nucleus. These results are shown in **Supplementary Table 2**.

Phylogenetic Analysis and Classification of *R2R3-MYB* Genes in *Gynostemma pentaphyllum*

To elucidate the evolutionary relationship and gene function of the *R2R3-GpMYB* gene family, a ML tree containing 87 *R2R3-GpMYB* genes and 124 *R2R3-AtMYB* genes was constructed

[2]http://smart.embl-heidelberg.de/

using the FastTree software (**Figure 1**). These 87 *R2R3-GpMYB* genes were divided into 32 subgroups (A1–A32), among which 20 subgroups (containing 61 *R2R3-GpMYB* genes) were consistent with the previously constructed phylogenetic tree of *A. thaliana* R2R3-MYB proteins. There were 10 specific subgroups in *G. pentaphyllum*, which were not clustered with *A. thaliana*. Moreover, no *R2R3-GpMYB* gene belonged to the *A. thaliana* S6, S12, or S25 subgroup, indicating that these *R2R3-GpMYB* genes may have evolved or lost in a given subgroup after divergence. The R2R3-AtMYB proteins of the same subgroup may have similar functions. For example, *R2R3-AtMYB* genes in the S6 and S12 subgroups are known to regulate anthocyanin and glucosinolate biosynthesis, respectively (Lin-Wang et al., 2010; Yu et al., 2021). Thus, these results suggested that *G. pentaphyllum* may have lost the ability to activate the accumulation of anthocyanin and glucosinolate or contained other special regulated pathways to produce these metabolites.

Gene Structure, Conserved Domains, and Motif Composition of *Gynostemma pentaphyllum R2R3-MYB* Gene Family

The typical *R2R3-MYB*-encoded proteins were characterized by R2 and R3 repeats (Dubos et al., 2010). As shown in **Figures 2A,B**, the 87 identified R2R3-GpMYB proteins from 32 subgroups contained two conserved MYB repeats and were separated by approximately 108 basic residues. The exon–intron structure analysis indicated that the number of exons in *R2R3-GpMYB* genes varied from 0 to 11, the majority of which contained two

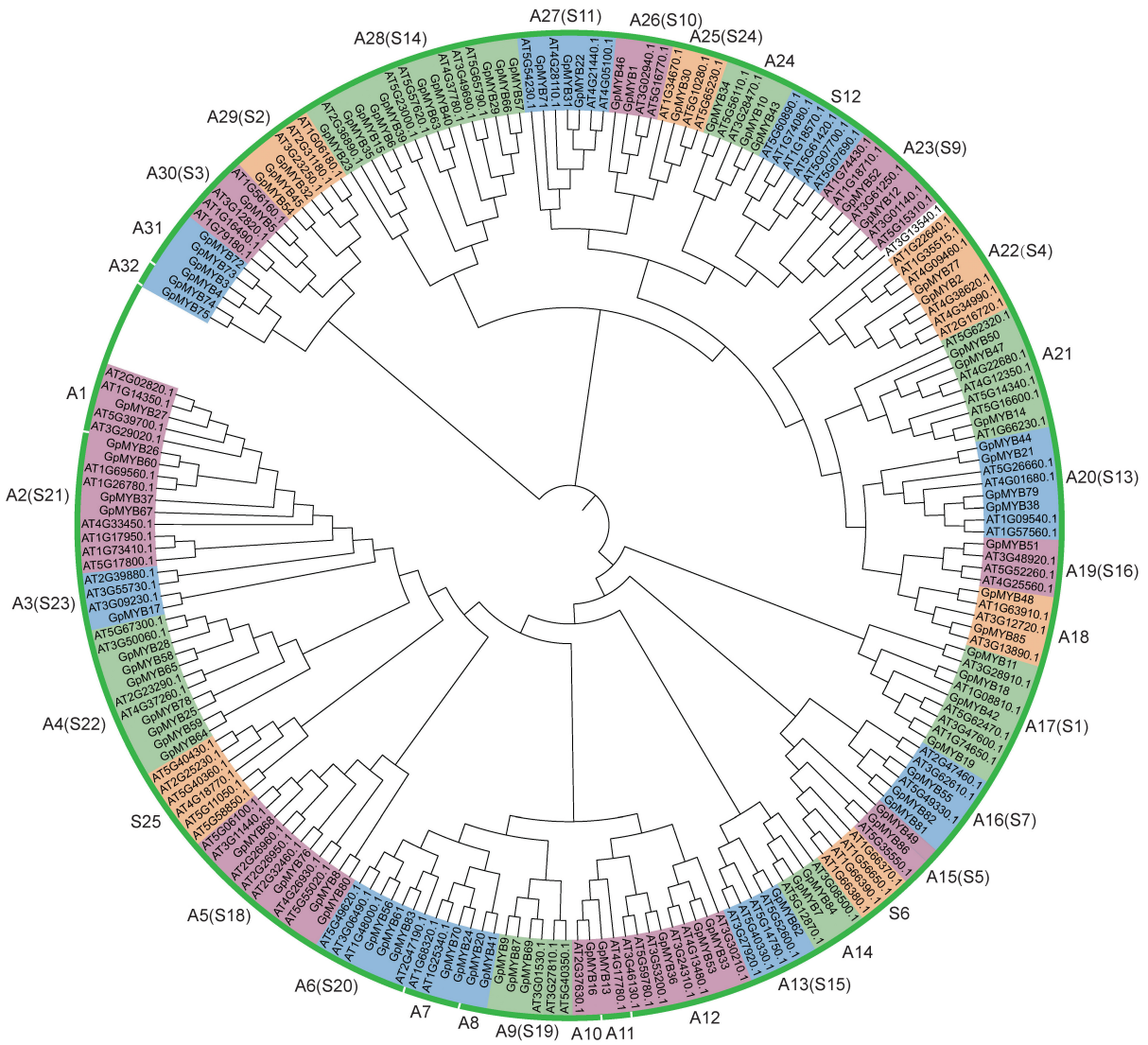


FIGURE 1 | Phylogenetic tree comparison of R2R3-MYB proteins between *Gynostemma pentaphyllum* (*G. pentaphyllum*) and *Arabidopsis thaliana* (*A. thaliana*). A maximum likelihood (ML) phylogenetic tree, including full length of 87 R2R3-GpMYB and 124 R2R3-AtMYB protein sequences, was generated using the Jones–Taylor–Thornton (JTT) algorithm *via* the FastTree software. The genes of the *R2R3-GpMYB* family were divided into 32 subgroups (designated as A1–A32). In addition, the classification method of *A. thaliana* (Stracke et al., 2001) was adopted.

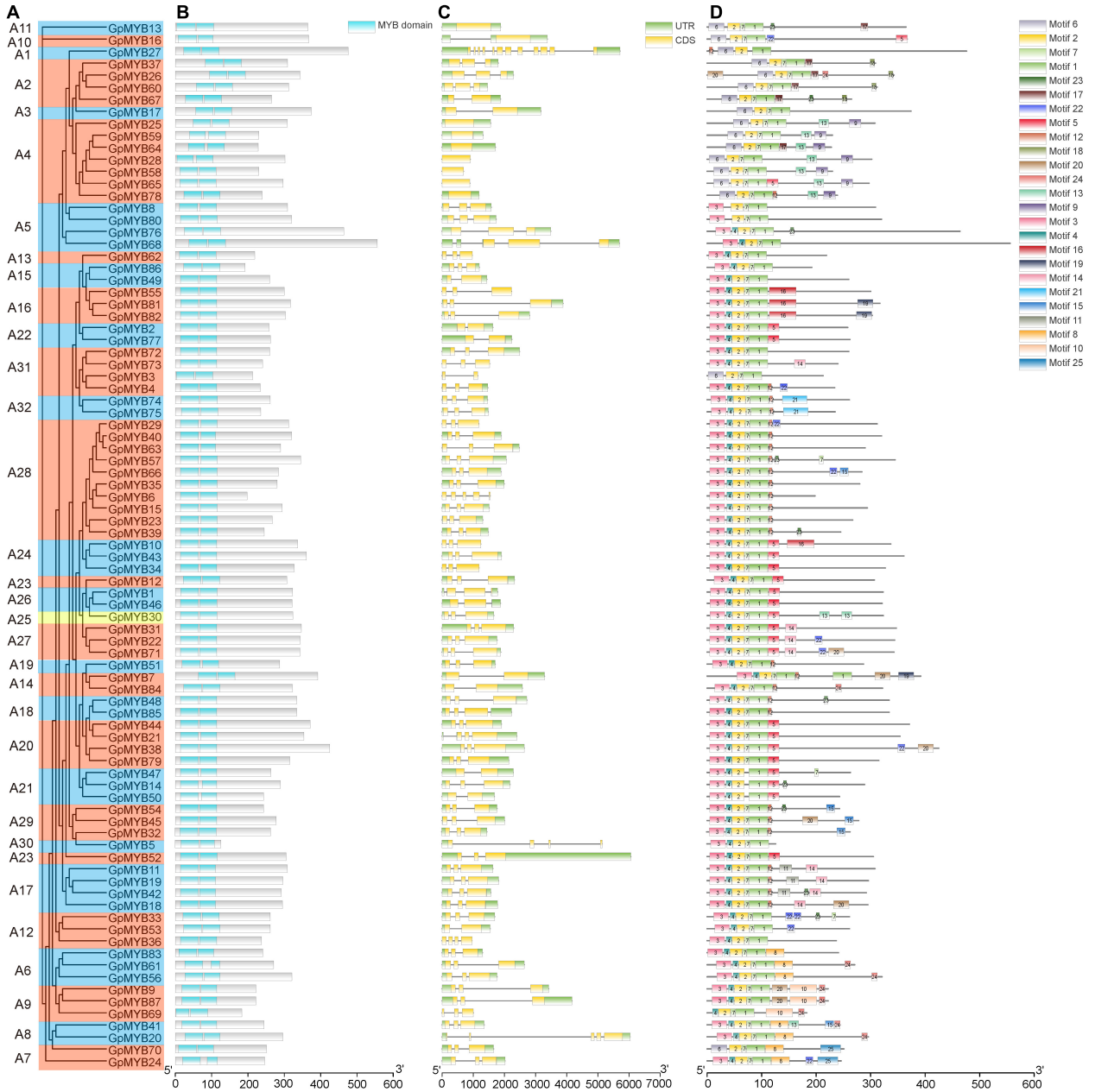


FIGURE 2 | Analysis of conserved MYB DNA-binding domains, gene structure, and conserved motifs depending on the phylogenetic relationships in *R2R3-GpMYB* genes. **(A)** A phylogenetic tree was built using 87 R2R3-GpMYB proteins with the ML method. The phylogenetic tree contains 32 subgroups represented by different colors. **(B)** The conserved MYB DNA-binding domains were examined using Pfam and the Simple Modular Architecture Research Tool (SMART). Green boxes indicate conserved MYB DNA-binding domains. **(C)** Exon/intron structure analysis of *R2R3-GpMYB* genes. Gray lines, yellow boxes, and green boxes indicate introns, exons, and untranslated regions (UTRs), respectively. **(D)** Conserved motifs of *R2R3-GpMYB* genes elucidated by Multiple Expectation maximizations for Motif Elicitation (MEME). The conserved motifs are represented by the different colored boxes. The scale bar of each *R2R3-GpMYB* gene is shown at the bottom.

introns, accounting for about 65.5%. Generally, similar structures of exon/intron were observed among the genes in the same subgroup, especially the number of introns. For example, the *R2R3-GpMYB* genes in the A4 subgroup contained no intron, whereas the A17 subgroup harbored two introns (**Figures 2A,C**).

The conserved motifs of all the R2R3-GpMYB proteins were studied using a motif-based sequence analysis tool (**Supplementary Table 3**). As shown in **Figure 2D**, motif 1, motif 2, motif 3, motif 4, motif 6, and motif 7 in the N-terminus encoded the conserved MYB DNA-binding domain, whereas motifs in the C-terminus were highly variable. The majority of *R2R3-GpMYB* genes belonging to the same subgroup with similar functions exhibited similar motif compositions outside the MYB domain (**Figures 2A,D**). For example, the A2 subgroup contained

motif 17 and motif 18, which played important roles in the development of axillary meristem (Lee et al., 2009), whereas motif 9 and motif 13 in the A4 subgroup participated in providing resistance to biotic and abiotic stresses (Jung et al., 2008). These results indicated that these motifs were conserved in specific subgroups, and proteins sharing these motifs within a group in the phylogenetic tree likely had similar functions.

Chromosomal Distribution and Synteny Analyses of *Gynostemma pentaphyllum R2R3-MYB* Family

The *G. pentaphyllum* genomic database and genome chromosomal location results revealed that 86 out of 87 *R2R3-GpMYB* genes were unevenly distributed on 11 chromosomes (**Figure 3**). In detail, chromosome 11 had 16 *R2R3-GpMYB* genes, accounting for the largest number of *R2R3-GpMYB* genes, followed by chromosome 7 (11 *R2R3-GpMYB* genes), whereas chromosome 8 contained only two genes and had the minimum number of *R2R3-GpMYB* genes. The majority of *R2R3-GpMYB* genes were located on both ends of the chromosome. In addition, no correlation was found between the chromosome length and the distribution of *R2R3-GpMYB* gene family members on the chromosome. According to a previous study, if two or more genes are present within 200 kb, the elements are considered a tandem repeat event (Holub, 2001). In total, six *R2R3-GpMYB* genes underwent five tandem repeat events (**Supplementary Table 4**).

We employed basic local alignment search tool for proteins (BLASTP) and MCScanX to construct the collinearity of the *R2R3-MYB* gene family in *G. pentaphyllum* and identify the possible relationship and potential duplication events between them. Intrachromosomal duplications of the *R2R3-MYB* gene family were observed in the *G. pentaphyllum* genome (**Supplementary Table 4**). In detail, 34 pairs of *R2R3-GpMYB* genes duplicated tandemly on all the 11 chromosomes (**Figure 4**).

To further illustrate the potential evolutionary patterns of the *R2R3-GpMYB* gene family, a comparative orthologous analysis was performed between *G. pentaphyllum* and other two representative species, namely, *A. thaliana* and *Cucumis sativus* (*C. sativus*), which belong to the Brassicaceae and Cucurbitaceae families, respectively (**Figure 5**). The orthologous gene pairs between *G. pentaphyllum* and *A. thaliana* and *G. pentaphyllum* and *C. sativus* were 44 and 70, respectively (**Supplementary Tables 5, 6**). These results revealed that the identified orthologous events of *GpMYB-CsMYB* were considerably more than those of *GpMYB-AtMYB* based on the close evolutionary relationship between *G. pentaphyllum* and *C. sativus*. An extensive level of synteny conservation and increased number of orthologous events of *GpMYB-CsMYB* indicated that *R2R3-GpMYB* genes in *G. pentaphyllum* shared a similar structure and function with *R2R3-CsMYB* genes in *C. sativus*.

To further investigate the driving force behind the duplication of *R2R3-MYB* gene pairs in *G. pentaphyllum*, Ka/Ks (non-synonymous/synonymous substitution ratio) calculation of the duplicated *R2R3-MYB* gene pairs was performed to determine whether a selective pressure acted on the *R2R3-GpMYB* genes (**Supplementary Tables 5, 6**). Interestingly, all the Ka/Ks values of orthologous *R2R3-GpMYB* gene pairs were less than 1, indicating that these genes were subjected to purifying selection with limited functional divergence during evolution after duplication events.

Identification of *R2R3-MYB* Was Related to Both Gypenoside and Flavonol Biosynthesis in *Gynostemma pentaphyllum*

Transcriptional activators usually present similar expression patterns to the downstream structural genes of the metabolic pathway, narrowing the scope of screening candidate regulators and providing functional prediction. In a previous study, the early biosynthesis genes (EBGs) and late biosynthesis genes (LBGs) corresponding to gypenoside biosynthesis were elucidated (Huang et al., 2021). Furthermore, based on the gene expression profiles of different tissues (tendril, young leaf, mature leaf, root, stem, flower, and fruit), a WGCNA was performed to identify the potential upstream regulators of gypenoside biosynthetic pathway genes (Huang et al., 2021).

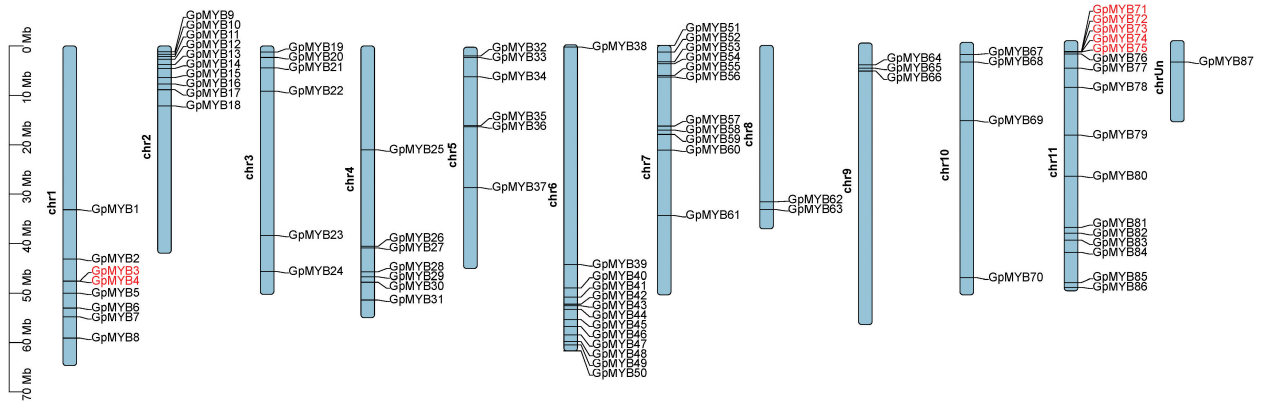


FIGURE 3 | Chromosomal locations of *G. pentaphyllum R2R3-MYB* genes. Chr 1–11 represent chromosomes 1–11 and chrUn indicates an unanchored linkage group. Chromosomal locations of *R2R3-GpMYB* genes were mapped based on the *G. pentaphyllum* genome. The names of genes highlighted in red on chromosomes indicate tandem duplications.

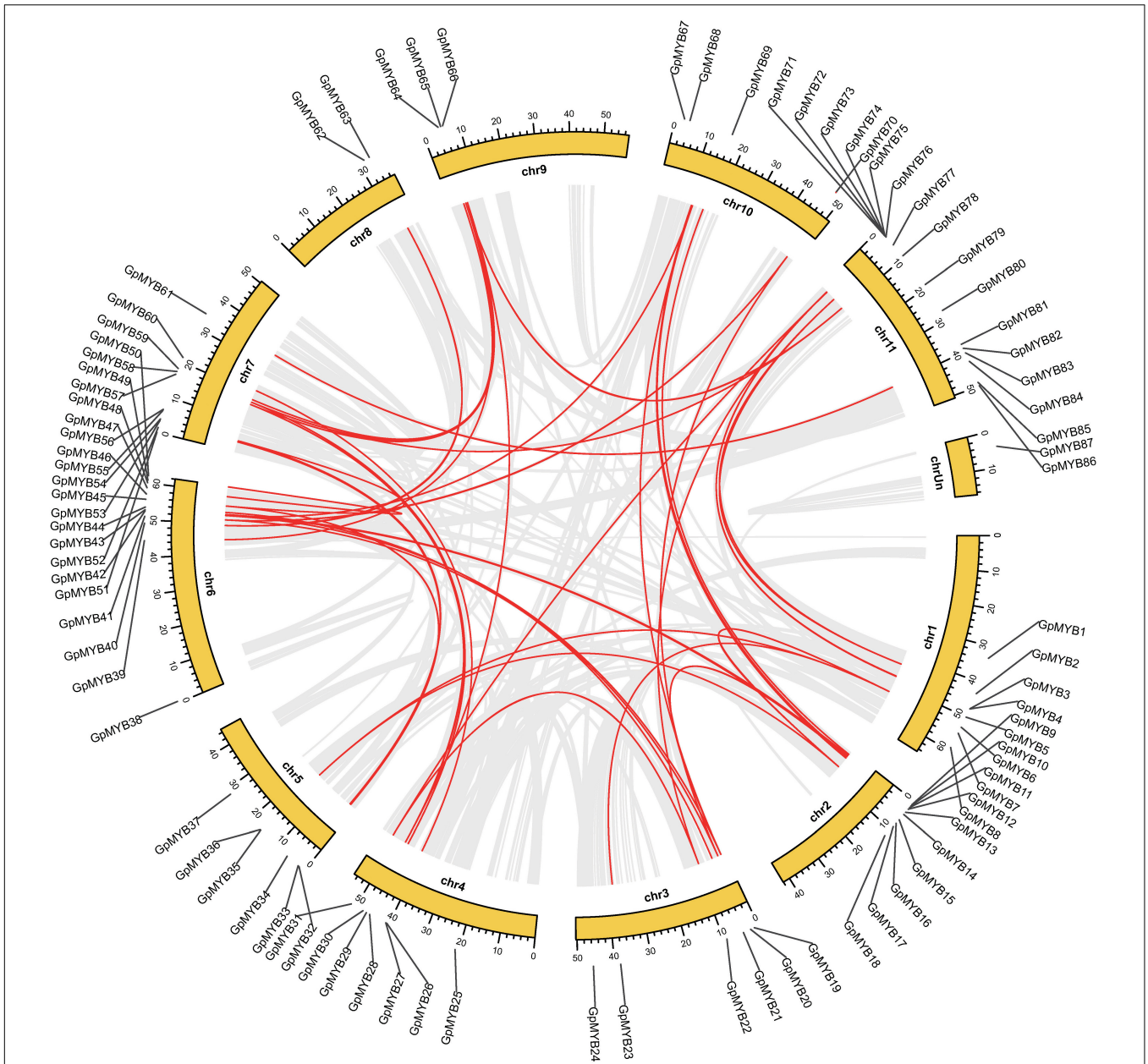


FIGURE 4 | Synteny analysis of *G. pentaphyllum R2R3-MYB* genes. Chr 1–11 represent chromosome-scale scaffolds and ChrUn indicates an unanchored linkage group. All the synteny gene pairs and duplicated *MYB* gene pairs were presented by gray lines and red lines, respectively.

In the constructed gypenoside biosynthesis regulatory network, five *GpMYB* genes were identified, three of which were *R2R3-MYB* genes, including *GpMYB60*, *GpMYB80*, and *GpMYB81*. Among these, *GpMYB60* was related to *AtMYB105* and *AtMYB117*, suggesting that the function of *GpMYB60* was related to the development of floral organs and the initiation of ovule outgrowth (Lee et al., 2009). *GpMYB80* was the closest homolog of *AtMYB97* and *AtMYB120*, acting as a transcriptional activator to control the differentiation of the pollen tube required for sperm release (Liang et al., 2013), and *GpMYB81* was highly close to other positive regulators of flavonol biosynthesis, such as *AtMYB11*, *AtMYB12*, and *AtMYB111* (Tan et al., 2019). In addition, flavonol biosynthetic pathway genes, including *Gynostemma pentaphyllum* 4-coumarate-CoA ligase (*Gp4CL*), *GpCHS*, *Gynostemma pentaphyllum* chalcone isomerase (*GpCHI*), *Gynostemma pentaphyllum* flavanone 3-hydroxylase (*GpF3H*), and *Gynostemma pentaphyllum* flavonol synthase (*GpFLS*), were identified in the gypenoside-related module (**Supplementary Table 7**). As confirmed by qRT-PCR, *GpMYB81*, along with genes encoding gypenoside and flavonol biosynthetic enzymes, showed similar expression patterns, i.e., a high expression in young leaf tissue (**Figure 6A**).

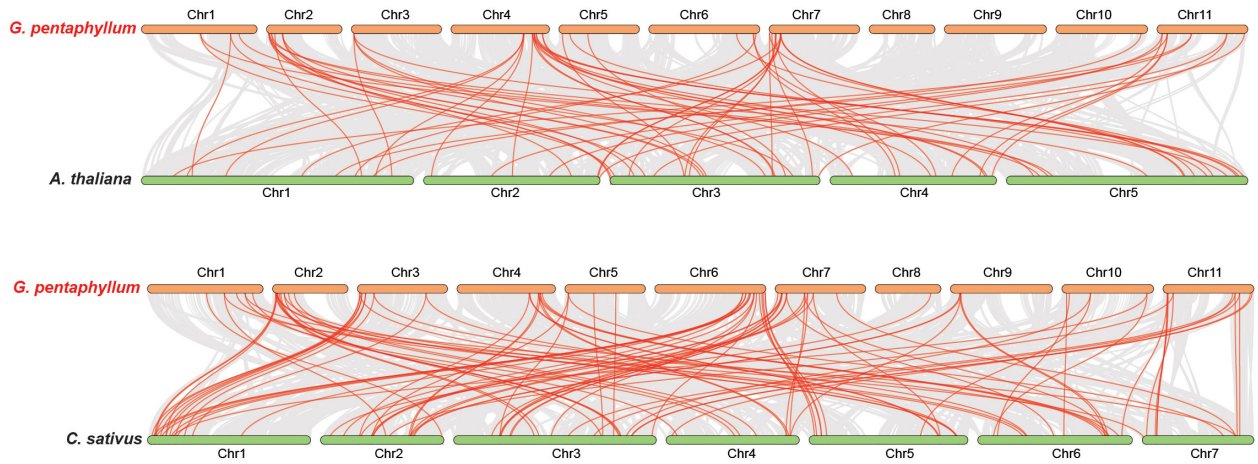


FIGURE 5 | Collinearity analysis of *R2R3-MYB* genes between *G. pentaphyllum* and two representative plant species [*A. thaliana* and *Cucumis sativus* (*C. sativus*)]. Chr represents chromosome-scale scaffolds. All the gene pairs and syntenic *R2R3-MYB* gene pairs were presented by gray lines and red lines, respectively.

As an effective elicitor, MeJA can intensify the accumulation of several secondary metabolites in various medicinal plants (Zhan et al., 2018; Deng et al., 2020). Candidate genes including TFs and biosynthetic pathway genes showed similar expression patterns in response to MeJA, further narrowing down the number of candidate genes. The qRT-PCR results showed that the expression of gypenoside biosynthetic pathway genes comprising *GpFPS1*, *Gynostemma pentaphyllum* squalene synthase 1 (*GpSS1*), *Gynostemma pentaphyllum* squalene epoxidase 2 (*GpSE2*), and *Gynostemma pentaphyllum* 2,3-oxidosqualene cyclases 1 (*GpOSC1*), and flavonol biosynthetic pathway genes, including *Gp4CL*, *GpCHS*, *GpCHI*, *GpF3H*, and *GpFLS*, was significantly increased after the MeJA treatment. Moreover, among the three candidate *R2R3-GpMYB* genes, the expression of *GpMYB81* increased gradually with the increase in the MeJA treatment time (**Figure 6B**). According to functional predictions and co-expression patterns, we speculated that *GpMYB81* most likely functions as a "dual-function" activator of both gypenoside and flavonol biosynthesis.

Gypenoside and Flavonol Biosynthetic Pathway Genes Were Transcriptionally Activated by GpMYB81

To investigate the mechanism underlying similar expression patterns among gypenoside, flavonol biosynthetic pathway genes, and their potential regulator GpMYB81, the transcriptional *cis*-elements of *GpFPS1* and *GpCHS* were analyzed (**Figure 7A**). The results revealed that the promoters of *GpFPS1* and *GpCHS* contained conserved MYB-recognition elements (MREs) or AC-rich elements, suggesting that GpMYB81 might bind to the promoters of *GpFPS1* and *GpCHS*. To prove this hypothesis, Y1H assays were performed. As shown in **Figure 7B**, GpMYB81 could bind to the promoters of *GpFPS1* and *GpCHS in vivo*. Moreover, the transient expression of the promoter activity assays revealed that GpMYB81 could activate the expression of *GpFPS1* and *GpCHS* (**Figure 7C**). These results confirmed that GpMYB81 can simultaneously activate gypenoside and flavonol biosynthetic pathway genes, thereby parallelly promoting the accumulation of gypenosides and flavonols.

DISCUSSION

Secondary metabolism not only endows plants with the ability to adapt to the ecological environment, but also provides abundant pharmaceutical ingredients of considerable health benefits to humans (Ramakrishna and Ravishankar, 2011; Thimmappa et al., 2014; Li et al., 2015). Terpenoid saponins and flavonoids are two important secondary metabolites, several of which have been proven to exhibit antiviral, antifungal, and anticancer properties (Zhang et al., 2018; Bonta, 2020; Russo et al., 2020; Yang et al., 2021). *G. pentaphyllum* is a traditional Chinese medicinal herb and known for its industrial use. It often serves as a natural source of gypenosides and flavonoids in medicinal products. As reported in our previous study (Huang et al., 2021), the genome sequencing of *G. pentaphyllum* was completed in 2021, providing the opportunity to perform a genome-wide analysis of the *R2R3-GpMYB* gene family. However, the *R2R3-GpMYB* gene family has not been comprehensively analyzed and their dual functionality in regulating distinct pathways of synthesis has not been well studied.

In this study, 87 members of the *R2R3-GpMYB* family were identified and characterized in *G. pentaphyllum*. Although the number of *R2R3-GpMYB* was higher than that in *C. sativus* (69) (Cheng et al., 2020), it was significantly less than that in cotton (205) (Huang J. et al., 2019), *Populus trichocarpa* (196) (Wilkins et al., 2009), and banana (285) (Pucker et al., 2020). Whole-genome duplication (WGD) is a crucial event for the rapid expansion and evolution of gene families, e.g., the recent gene duplication in several angiosperms (Pucker et al., 2020). However, WGD events and tandem duplications are rare in the *G. pentaphyllum* genome. Conversely, numerous segmental duplication events were identified in

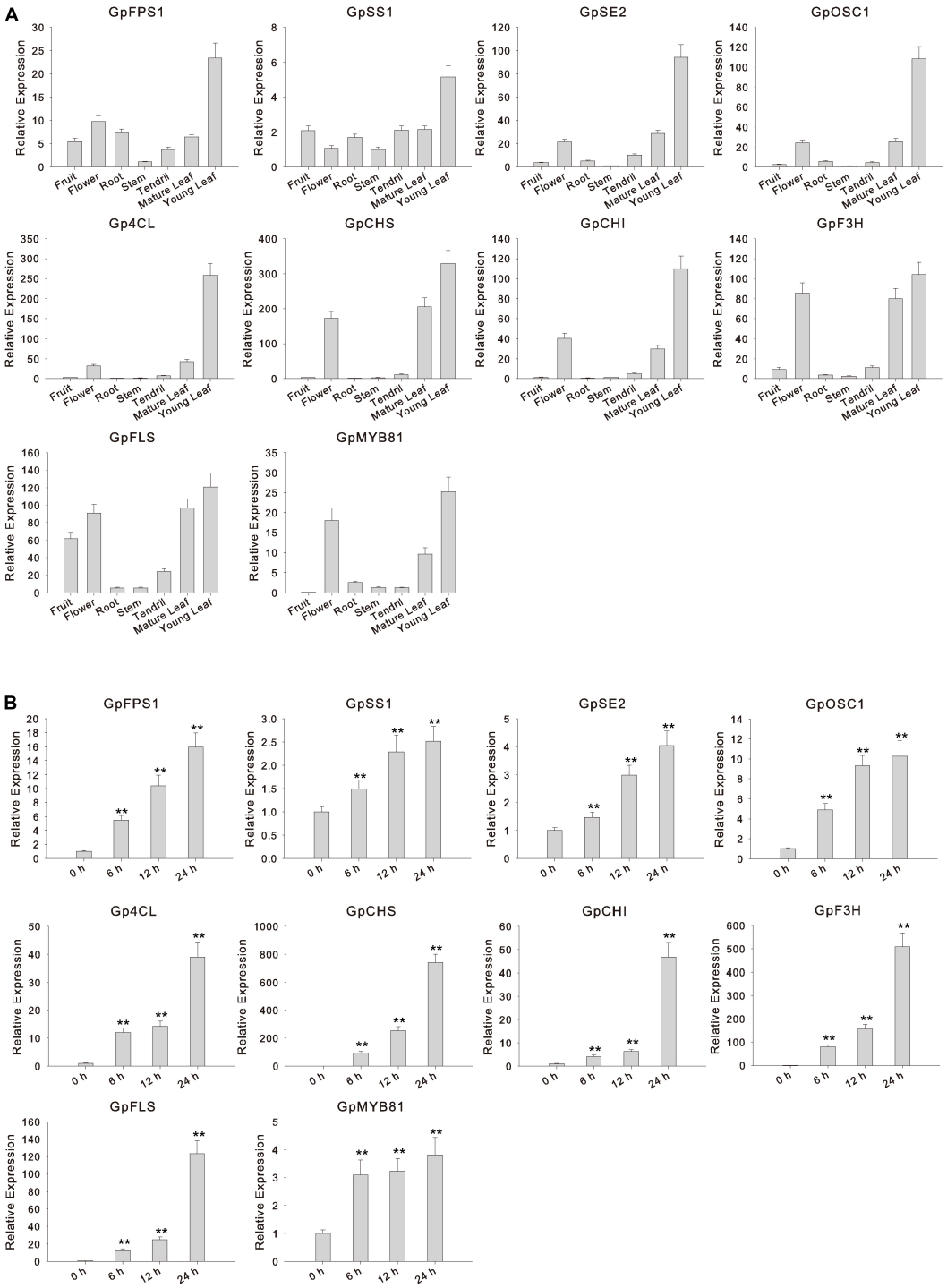


FIGURE 6 | Expression profile of 10 selected candidate genes in different tissues and response to methyl jasmonate (MeJA) treatment. **(A)** Expression analysis of gypenoside and flavonol biosynthetic genes and candidate regulator *GpMYB81* in different tissues of *G. pentaphyllum*. Values represent the mean ± SE ($n = 3$ biological replicates). **(B)** Expression analysis of gypenoside and flavonol biosynthetic genes and candidate regulator *GpMYB81* with 0 (negative control), 6, 12, and 24 h treatment with MeJA. $^{**}p < 0.01$ (two-tailed Student's *t*-test). Values represent the mean ± SE ($n = 3$ biological replicates).

G. pentaphyllum, indicating that segmental duplication events were the most important factor in the expansion of the *R2R3-GpMYB* gene family, which exhibited an evolutionary pattern similar to that of *MYB* genes in other plants (Cannon et al., 2004). In addition, certain *R2R3-GpMYB* genes were not associated with syntenic gene pairs in *A. thaliana* or

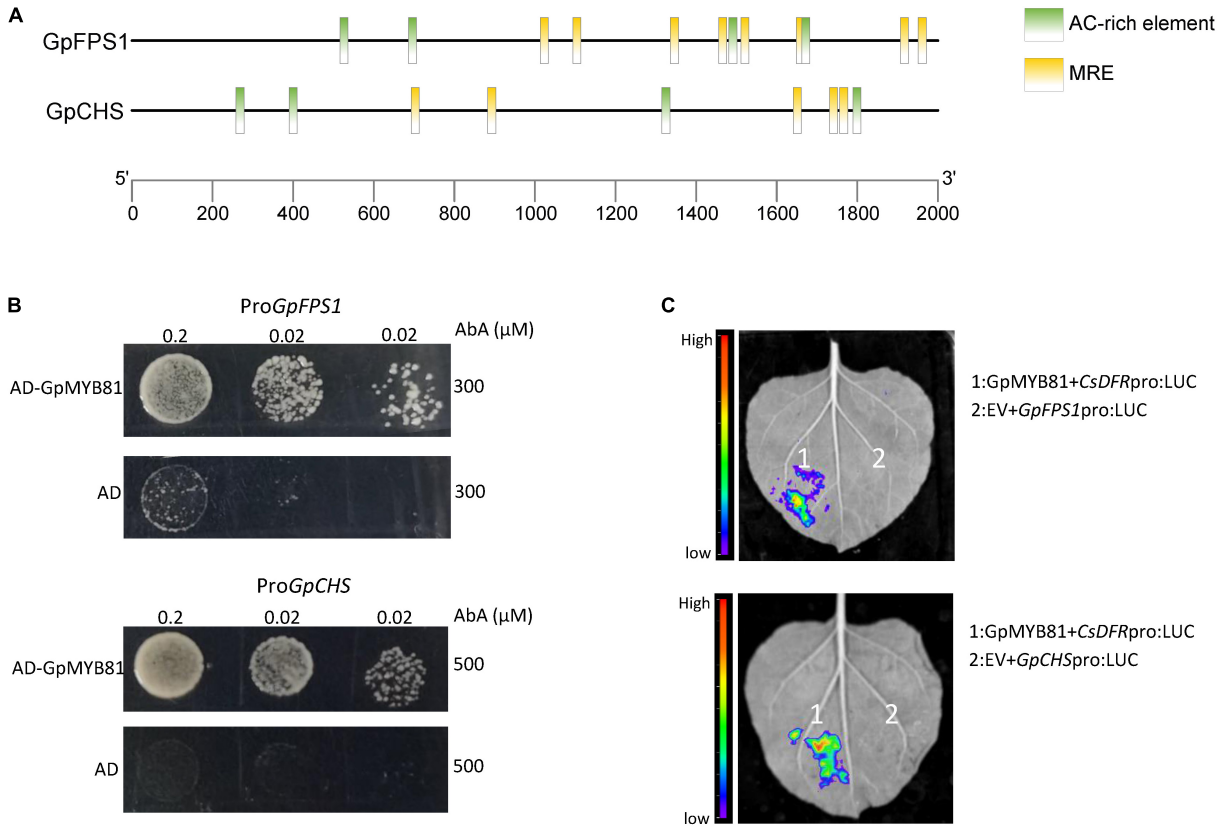


FIGURE 7 | Binding of GpMYB81 to *GpFPS1* and *GpCHS* promoters and activation of their expression. **(A)** Schematic diagram of *GpFPS1* and *GpCHS* promoters. The MYB-recognition elements (MREs) and AC-rich elements, identified with manual inspection, are indicated by their respective labels. **(B)** Yeast one-hybrid assays reveal that GpMYB81 can bind to the promoters of *GpFPS1* and *GpCHS*. **(C)** Dual-luciferase (LUC) assays in *Nicotiana benthamiana* leaves were performed using *GpFPS1* and *GpCHS* promoters driving *LUC* as a reporter along with effector GpMYB81; the empty vector (EV) served as an internal control. These results showed that GpMYB81 can activate both *GpFPS1* and *GpCHS* promoters.

C. sativus, indicating their possible specificity to *G. pentaphyllum* during the course of evolution, providing new insight or interest to explore those special *R2R3-GpMYB* genes in the future.

Flavonoids and terpenoids, the two largest groups of specialized plant metabolites, are derived from two distinct pathways. Notably, there are several examples where MYB proteins act to regulate terpenoid and flavonoid biosynthesis. In *A. thaliana*, the R2R3-MYB TFs that are currently known to regulate flavonol biosynthesis belong to subgroup 7 (Stracke et al., 2007). Some MYB TFs involved in the regulation of triterpene saponins have also been characterized. For instance, *Panax ginseng* myeloblastosis 2 (PgMYB2) in *Panax ginseng* was characterized as a positive regulator of ginsenoside metabolism (Liu et al., 2019). However, TFs act in a coordinated manner to simultaneously regulate different pathways of specialized metabolism in a novel regulatory mechanism. In this study, *GpMYB81* belongs to subgroup 7 of the *R2R3-MYB* gene family from *A. thaliana*, which was identified in the gypenoside-related module. In addition, the expression patterns of *GpMYB81* were highly correlated with gypenoside and flavonol biosynthesis genes and with those involved in response to MeJA, further supporting the finding that *GpMYB81* can regulate the accumulation of both gypenosides and flavonols. The chalcone synthase (*CHS*) gene plays a vital role in flavonol biosynthesis and greatly impacts the content of flavonols, whereas the overexpression or RNA interference of the *FPS* gene can significantly increase or decrease the biosynthesis of gypenoside (Kim et al., 2014; Zhang et al., 2017). In this study, GpMYB81 could bind to the promoters of *GpFPS1* and *GpCHS* and increase the transcriptional activities of these promoters. These findings suggested that *GpMYB81* acts as a "dual-function" TF that regulates both gypenoside and flavonol pathways in *G. pentaphyllum*.

CONCLUSION

In summary, this study presented a detailed genome-wide analysis of the *R2R3-GpMYB* gene family. A total of 87 *R2R3-GpMYB* genes were identified in *G. pentaphyllum* and divided into 32 subgroups, with an uneven distribution on 11 chromosomes. Similar exon–intron structures and conserved motif compositions of *R2R3-GpMYB* genes were observed in the same subgroup, which provided additional support for phylogenetic analysis. Synteny analysis indicated that segmental duplication events primarily contributed to the expansion of the *R2R3-GpMYB* gene family. The Ka/Ks analysis suggested that

the *R2R3-GpMYB* gene family underwent purifying selection. A combination of similar gene expression patterns, Y1H, and dual-LUC assay results verified that *GpMYB81* acted as a "dual-function" activator in gypenoside and flavonol biosynthesis by directly binding to the promoters of *GpFPS1* and *GpCHS*. These results provide novel insights into the parallel transcriptional regulation of gypenoside and flavonol biosynthesis in *G. pentaphyllum*.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the transcriptome sequencing data can be found in NCBI under accession codes PRJNA720501 and PRJNA631355.

AUTHOR CONTRIBUTIONS

DH and YT conceived this project. DH and RM designed the experiments. DH and SX prepared the samples and wrote the manuscript. DH, RM, and SX analyzed the bioinformatics data. SY, LL, RH, and YT provided valuable suggestions on the research design and the improvement of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Natural Science Foundation of Guangxi Zhuang Autonomous Region (2020GXNSFBA297025), the Guangxi Middle-aged and Young Teachers' Basic Ability Promotion Project (2020KY07039), the Guangxi University of Chinese Medicine Scientific Research Fund (2019BS007), the Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine Open Project Fund (20-065-14), and the Special Fund for Introducing Scientific and Technological Talents of Guangdong Academy of Agricultural Sciences (R2020YJ-YB3003).

ACKNOWLEDGMENTS

The authors would like to thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.796248/full#supplementary-material

REFERENCES

Ampomah-Dwamena, C., Thrimawithana, A. H., Dejnoprat, S., Lewis, D., Espley, R. V., and Allan, A. C. (2019). A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* 221, 309–325. doi: 10.1111/nph.15362

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

Bonta, R. K. (2020). Dietary phenolic acids and flavonoids as potential anti-cancer agents: current state of the art and future perspectives. *Anticancer Agents Med. Chem.* 20, 29–48. doi: 10.2174/1871520619666191019112712

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi: 10.1186/1471-2229-4-10

Chen, A. Y., and Chen, Y. C. (2013). A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem.* 138, 2099–2107. doi: 10.1016/j.foodchem.2012.11.139

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

Cheng, C., Li, Q., Wang, X., Li, Y., and Qian, C. (2020). Identification and expression analysis of the CsMYB gene family in root knot nematode-resistant and susceptible cucumbers. *Front. Genet.* 11:550677. doi: 10.3389/fgene.2020.550677

Deng, C., Wang, Y., Huang, F., Lu, S., Zhao, L., Ma, X., et al. (2020). SmMYB2 promotes salvianolic acid biosynthesis in the medicinal herb *Salvia miltiorrhiza*. *J. Integr. Plant Biol.* 62, 50–64. doi: 10.1111/jipb.12943

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

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Gabrielsen, O., Sentenac, A., and Fromageot, P. (1991). Specific DNA binding by c-Myb: evidence for a double helix-turn-helix-related motif. *Science* 253, 1140–1143. doi: 10.1126/science.1887237

He, J., Liu, Y., Yuan, D., Duan, M., Liu, Y., Shen, Z., et al. (2020). An R2R3 MYB transcription factor confers brown planthopper resistance by regulating the phenylalanine ammonia-lyase pathway in rice. *Proc. Natl. Acad. Sci. U.S.A.* 117, 271–277. doi: 10.1073/pnas.1902771116

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527. doi: 10.1038/35080508

Huang, D., Ming, R., Xu, S., Wang, J., Yao, S., Li, L., et al. (2021). Chromosome-level genome assembly of *Gynostemma pentaphyllum* provides insights into gypenoside biosynthesis. *DNA Res.* 28:dsab018. doi: 10.1093/dnares/dsab018

Huang, D., Tang, Z., Fu, J., Yuan, Y., Deng, X., and Xu, Q. (2019). CsMYB3 and CsRuby1 form an 'activator-and-repressor' loop for regulation of anthocyanin biosynthesis in citrus. *Plant Cell Physiol.* 61, 318–330. doi: 10.1093/pcp/pcz198

Huang, J., Guo, Y., Sun, Q., Zeng, W., Li, J., Li, X., et al. (2019). Genome-wide identification of R2R3-MYB transcription factors tegulating secondary cell wall thickening in cotton fiber development. *Plant Cell Physiol.* 60, 687–701. doi: 10.1093/pcp/pcy238

Huang, D., Wang, X., Tang, Z., Yuan, Y., Xu, Y., He, J., et al. (2018). Subfunctionalization of the Ruby2-Ruby1 gene cluster during the domestication of citrus. *Nat. Plants* 4, 930–941. doi: 10.1038/s41477-018-0287-6

Jia, L., Clegg, M. T., and Jiang, T. (2004). Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.* 134, 575–585. doi: 10.1104/pp.103.027201

Jung, C., Seo, J. S., Han, S. W., Koo, Y. J., Kim, C. H., Song, S. I., et al. (2008). Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiol.* 146, 623–635. doi: 10.1104/pp.107.110981

Kim, Y. K., Kim, Y. B., Uddin, M. R., Lee, S., Kim, S. U., and Park, S. U. (2014). Enhanced triterpene accumulation in *Panax* ginseng hairy roots overexpressing mevalonate-5-pyrophosphate decarboxylase and farnesyl pyrophosphate synthase. *ACS Synth. Biol.* 3, 773–779. doi: 10.1021/sb400194g

Lee, D. K., Geisler, M., and Springer, P. S. (2009). LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in *Arabidopsis*. *Development* 136, 2423–2432. doi: 10.1242/dev.031971

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, C., Ng, K. Y., and Fan, L. M. (2015). MYB transcription factors, active players in abiotic stress signaling. *Environ. Exp. Bot.* 114, 80–91. doi: 10.1016/j.envexpbot.2014.06.014

Liang, Y., Tan, Z. M., Zhu, L., Niu, Q. K., Zhou, J. J., Li, M., et al. (2013). MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization. *PLoS Genet.* 9:e1003933. doi: 10.1371/journal.pgen.1003933

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T. K., et al. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in *Rosaceae*. *BMC Plant Biol.* 10:50. doi: 10.1186/1471-2229-10-50

Liu, T., Luo, T., Guo, X., Zou, X., Zhou, D., Afrin, S., et al. (2019). PgMYB2, a MeJA-responsive transcription factor, positively regulates the dammarenediol synthase gene expression in *Panax* Ginseng. *Int. J. Mol. Sci.* 20:2219. doi: 10.3390/ijms20092219

Millard, P. S., Kragelund, B. B., and Burow, M. (2019). R2R3 MYB transcription factors - functions outside the DNA-binding domain. *Trends Plant Sci.* 24, 934–946. doi: 10.1016/j.tplants.2019.07.003

Nabavi, S. M., Šamec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S., et al. (2020). Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. *Biotechnol. Adv.* 38:107316. doi: 10.1016/j.biotechadv.2018.11.005

Park, S. H., Huh, T. L., Kim, S. Y., Oh, M. R., Tirupathi Pichiah, P. B., Chae, S. W., et al. (2014). Antiobesity effect of *Gynostemma pentaphyllum* extract (actiponin): a randomized, double-blind, placebo-controlled trial. *Obesity* 22, 63–71. doi: 10.1002/oby.20539

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Pucker, B., Pandey, A., Weisshaar, B., and Stracke, R. (2020). The R2R3-MYB gene family in banana (*Musa acuminata*): genome-wide identification, classification and expression patterns. *PLoS One* 15:e0239275. doi: 10.1371/journal.pone.0239275

Ramakrishna, A., and Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* 6, 1720–1731. doi: 10.4161/psb.6.11.17613

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105

Russo, M., Moccia, S., Spagnuolo, C., Tedesco, I., and Russo, G. L. (2020). Roles of flavonoids against coronavirus infection. *Chem. Biol. Interact* 328:109211. doi: 10.1016/j.cbi.2020.109211

Shen, C. Y., Ma, P. Y., Zhu, J. J., Jiang, J. G., Liu, L., Yi, Y. K., et al. (2020). Saponin extracts from *Gynostemma pentaphyllum* (Thunb.) Makino display sedative-hypnotic and anxiolytic effects. *Ind. Crops Prod.* 157:112893. doi: 10.1016/j.indcrop.2020.112893

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol* 4, 447–456. doi: 10.1016/s1369-5266(00)00199-0

Tan, H., Man, C., Xie, Y., Yan, J., Chu, J., and Huang, J. (2019). A crucial role of GA-regulated flavonol biosynthesis in root growth of *Arabidopsis*. *Mol Plant.* 12, 521–537. doi: 10.1016/j.molp.2018.12.021

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* 65, 225–257. doi: 10.1146/annurev-arplant-050312-120229

Wang, J., Zhao, M., Cheng, X., Han, Y., Zhao, T., Fan, M., et al. (2020). Dammarane-type saponins from *Gynostemma pentaphyllum* prevent hypoxia-induced neural injury through activation of ERK, Akt, and CREB pathways. *J. Agric. Food Chem.* 68, 193–205. doi: 10.1021/acs.jafc.9b06659

Wang, L., Lu, W., Ran, L., Dou, L., Yao, S., Hu, J., et al. (2019). R2R3-MYB transcription factor MYB6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. *Plant J.* 99, 733–751. doi: 10.1111/tpj.14364

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wang, Z. L., Wang, S., Kuang, Y., Hu, Z. M., Qiao, X., and Ye, M. (2018). A comprehensive review on phytochemistry, pharmacology, and flavonoid biosynthesis of *Scutellaria baicalensis*. *Pharm Biol.* 56, 465–484. doi: 10.1080/13880209.2018.1492620

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. (2009). Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* 149, 981–993. doi: 10.1104/pp.108.132795

Xie, Z., Zhao, Y., Chen, P., Jing, P., Yue, J., and Yu, L. L. (2011). Chromatographic fingerprint analysis and rutin and quercetin compositions in the leaf and whole-plant samples of di- and tetraploid *Gynostemma pentaphyllum*. *J. Agric. Food Chem.* 59, 3042–3049. doi: 10.1021/jf104329v

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Xu, S., Yao, S., Huang, R., Tan, Y., and Huang, D. (2020). Transcriptome-wide analysis of the AP2/ERF transcription factor gene family involved in the regulation of gypenoside biosynthesis in *Gynostemma pentaphyllum*. *Plant Physiol. Biochem.* 154, 238–247. doi: 10.1016/j.plaphy.2020.05.040

Yang, G., Zhang, J., Wang, S., Wang, J., Wang, J., Zhu, Y., et al. (2021). Gypenoside inhibits bovine viral diarrhea virus replication by interfering with viral attachment and internalization and activating apoptosis of infected cells. *Viruses* 13:1810. doi: 10.3390/v13091810

Yu, Y., Xu, M., Ding, X., Chu, Z., and Liu, H. (2021). Activating the MYB51 and MYB122 to upregulate the transcription of glucosinolates biosynthesis genes by copper ions in *Arabidopsis*. *Plant Physiol. Biochem.* 162, 496–505. doi: 10.1016/j.plaphy.2021.03.025

Zhan, X., Liao, X., Luo, X., Zhu, Y., Feng, S., Yu, C., et al. (2018). Comparative metabolomic and proteomic analyses reveal the regulation mechanism underlying MeJA-induced bioactive compound accumulation in cutleaf groundcherry (*Physalis angulata* L.) hairy roots. *J. Agric. Food Chem.* 66, 6336–6347. doi: 10.1021/acs.jafc.8b02502

Zhang, X., Abrahan, C., Colquhoun, T. A., and Liu, C. J. (2017). A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in *Arabidopsis*. *Plant Cell* 29, 1157–1174. doi: 10.1105/tpc.16.00855

Zhang, X., Shi, G., Wu, X., and Zhao, Y. (2018). Gypensapogenin H from hydrolyzate of total *Gynostemma pentaphyllum* saponins induces apoptosis in human breast carcinoma cells. *Nat. Prod. Res.* 34, 1642–1646. doi: 10.1080/14786419.2018.1525370

Zhu, Z., Wang, H., Wang, Y., Guan, S., Wang, F., Tang, J., et al. (2015). Characterization of the cis elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. *J. Exp. Bot.* 66, 3775–3789. doi: 10.1093/jxb/erv173

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

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

frontiers in Plant Science

REVIEW
published: 11 January 2022
doi: 10.3389/fpls.2021.808228

Research Advances in Multi-Omics on the Traditional Chinese Herb *Dendrobium officinale*

*Yue Wang[1,2], Yan Tong[1,2], Oluwaniyi Isaiah Adejobi[1,2], Yuhua Wang[1,2] and Aizhong Liu[3]**

[1] Key Laboratory of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [2] Bio-Innovation Center of DR PLANT, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [3] Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming, China

Dendrobium officinale Kimura et Migo is an important epiphytic plant, belonging to the Orchidaceae family. There are various bioactive components in *D. officinale* plants, mainly including polysaccharides, alkaloids, and phenolic compounds. These compounds have been demonstrated to possess multiple functions, such as anti-oxidation, immune regulation, and anti-cancer. Due to serious shortages of wild resources, deterioration of cultivated germplasm and the unstable quality of *D. officinale*, the study has been focused on the biosynthetic pathway and regulation mechanisms of bioactive compounds. In recent years, with rapid developments in detection technologies and analysis tools, omics research including genomics, transcriptomics, proteomics and metabolomics have all been widely applied in various medicinal plants, including *D. officinale*. Many important advances have been achieved in *D. officinale* research, such as chromosome-level reference genome assembly and the identification of key genes involved in the biosynthesis of active components. In this review, we summarize the latest research advances in *D. officinale* based on multiple omics studies. At the same time, we discuss limitations of the current research. Finally, we put forward prospective topics in need of further study on *D. officinale*.

Keywords: *Dendrobium officinale*, multi-omics, bioactive compounds, biosynthesis, regulation

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Zhichao Xu,
Northeast Forestry University, China
Li Gu,
Fujian Agriculture and Forestry University, China

****Correspondence:***
Aizhong Liu
liuaizhong@mail.kib.ac.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *03 November 2021*
Accepted: *07 December 2021*
Published: *11 January 2022*

Citation:
Wang Y, Tong Y, Adejobi OI, Wang Y and Liu A (2022) Research Advances in Multi-Omics on the Traditional Chinese Herb Dendrobium officinale. Front. Plant Sci. 12:808228. doi: 10.3389/fpls.2021.808228

INTRODUCTION

Dendrobium officinale Kimura et Migo, an important epiphytic plant in the Orchidaceae family, has been used as a traditional Chinese medicine herb for thousands of years in China, and widely as a medicinal herb in many Asian countries (Ng et al., 2012). Wild *D. officinale* is often distributed in tropical regions, and it has been widely introduced to subtropical regions for cultivation. Generally, *D. officinale* can nourish the stomach, moisten the lung, relieve coughs and strengthen the body (Ng et al., 2012; Tang et al., 2017; Chen W. et al., 2021). Chemical isolation of compounds has identified diverse active compounds including polysaccharides, flavonoids, alkaloids, and multiple amino acids. In particular, the polysaccharides, alkaloids, and flavonoids extracted from *D. officinale* are thought to be the main bioactive ingredients for medicine (Meng et al., 2016; Tang et al., 2017; Wang et al., 2020; Wang, 2021). Pharmacological studies have confirmed the positive effects of polysaccharides, alkaloids and flavonoids extracted from *D. officinale* on immune regulation and

liver protection (Wang et al., 2020), on anti-cancer and neuroprotection (Zhang et al., 2003; Wang et al., 2020), and on anti-oxidant activity (Yuan et al., 2020a). Phenols extracted from *D. officinale* (such as erianin and gigantol) are active in enhancing immune activity while also exhibiting anti-tumor and anti-oxidant properties (Zhang et al., 2018; He et al., 2020). Owning to the increasing demands for diverse healthy products based on *D. officinale*, the cultivation of *D. officinale* is expanded throughout south China and south Asian countries with an annual output value of 10 billion Yuan per year (Wang and Shi, 2019).

Studies have revealed that the content of active compounds often varies among different tissues in *D. officinale*. For instance, stems accumulate the most abundant polysaccharides (Wang et al., 2020; Yuan et al., 2020a), while leaves contain higher contents of flavonoids and alkaloids than other tissues (Shen et al., 2017; Chen et al., 2019; Yuan et al., 2020b; Wang, 2021). Bibenzyls (belonging to phenolic compounds) including erianin and gigantol are mainly detected in roots and stems (Adejobi et al., 2021). Moreover, increasing evidence shows that the biosynthesis and accumulation of diverse active compounds are often influenced by various environmental factors such as different cultivation substrates (Zuo et al., 2020) and habitats (Lei et al., 2018). Therefore, it is important to dissect the physiological and molecular mechanisms underlying accumulation of those active compounds during the growth and development of different tissues for breeding by genetic improvement, cultivation and management practices in agriculture. In the last decade, technical development of DNA sequencing and molecular identification has led to the extensive application of diverse omics such as genomics, transcriptomics, proteomics and metabolomics, which are used to dissect the physiological and molecular mechanisms of active compound biosynthesis. These multi-omics studies have provided vital insights into the understanding of the biosynthesis pathways of active compounds and their molecular regulation networks during the growth and development of *D. officinale*. In this review, we focus on the genetic and molecular bases of active compounds identified from *D. officinale* based on the latest research on multi-omics in *D. officinale* to aid more studies in the discovery of the physiological and molecular mechanisms that underlie those active compounds accumulated in *D. officinale*, facilitating directional breeding, cultivation and management, and medicinal usage.

GENOMIC RESEARCH IN *DENDROBIUM OFFICINALE*

Genome data can provide basic and comprehensive information on genetic backgrounds and bridge gaps between genetic bases and active compounds in medicinal plants. To date, three versions of *D. officinale* genomic data (2n = 38) have been reported, with the genome size varying from 1.01 to 1.36 Gb and heterozygosity of 0.48–1.27% (Yan et al., 2015; Zhang G. Q. et al., 2016; Niu et al., 2021). A comparison between them is listed in **Table 1**. The first two versions were assembled at the scaffold level, while the third version was assembled onto 19 chromosomes with third-generation and Hi-C sequencing platforms (Niu et al., 2021). The number of protein-coding genes varied from 27631 to 35567, with 629–1462 gene families specific to *D. officinale*. There are differences in genomic data between three versions, which could result from the sequencing platform, assembly strategies and genome annotation methods. The genome of *D. officinale* characteristically features abundant repetitive sequences and two rounds of whole-genome duplication (WGD; Zhang G. Q. et al., 2016; Niu et al., 2021). In the future, with the aid of genetic and physical maps and more transcriptomic data, genome quality and annotation integrity can be further improved.

TABLE 1 | A quality comparison of three *D. officinale* genome versions.

Items	v.1.0 (Yan et al., 2015)	v.2.0 (Zhang G. Q. et al., 2016)	v.3.0 (Niu et al., 2021)
Sequencing platform	Illumina HiSeq 2000 PacBio RS II	Illumina HiSeq 2000	Illumina HiSeq 2500 PacBio Sequel II
Genome size	1.36 Gb	1.01 Gb	1.23 Gb
Heterozygosity	0.48%	0.628%	1.27%
Contig N50	25.12 kb	33.09 kb	1.44 Mb
Contig number	814,881	105,732	2,430
Scaffold N50	76.49 kb	391.46 kb	63.07 Mb
Assembly level	Scaffold	Scaffold	Chromosome
WGD	–	2	2
Repetitive ratio	63.33%	78.1%	64.39%
SNPs	5,432,657	5,758,781	–
Protein-coding genes	35,567	28,910	27,631
Specific gene families	1,462	629	1,196
Functionally annotated genes	34,699	–	25,894

For medicinal plants, one of the most important studies in genomics is to annotate and identify candidate genes related to the biosynthesis of active compounds. Three versions of the *D. officinale* genome showed marked expansion of some genes encoding the enzymes responsible for the biosynthesis of active compounds. According to the earlier two versions, some key enzyme genes involved in the biosynthesis of active compounds were identified (Yan et al., 2015; Zhang G. Q. et al., 2016). Based on the third version, the researchers comprehensively identified the genes in the biosynthetic pathways of polysaccharides, alkaloids, and flavonoids, including 418 genes (Niu et al., 2021; **Figures 1–3**). For the biosynthesis of polysaccharides, a total of 268 genes encoding 56 enzymes were identified (Niu et al., 2021). Among them, there are 13, 10, and 15 genes that encoded cellulose synthase-like A (CLSA), sucrose-phosphate synthase and sucrose synthase, respectively, which showed expansion (Yan et al., 2015; Zhang G. Q. et al., 2016). In particular, the galacturonosyltransferase and β-galactosidase gene families are uniquely evolved in *D. officinale* (Yan et al., 2015). The large-scale expansion of these genes may partly explain the abundant polysaccharides in *D. officinale*. For the

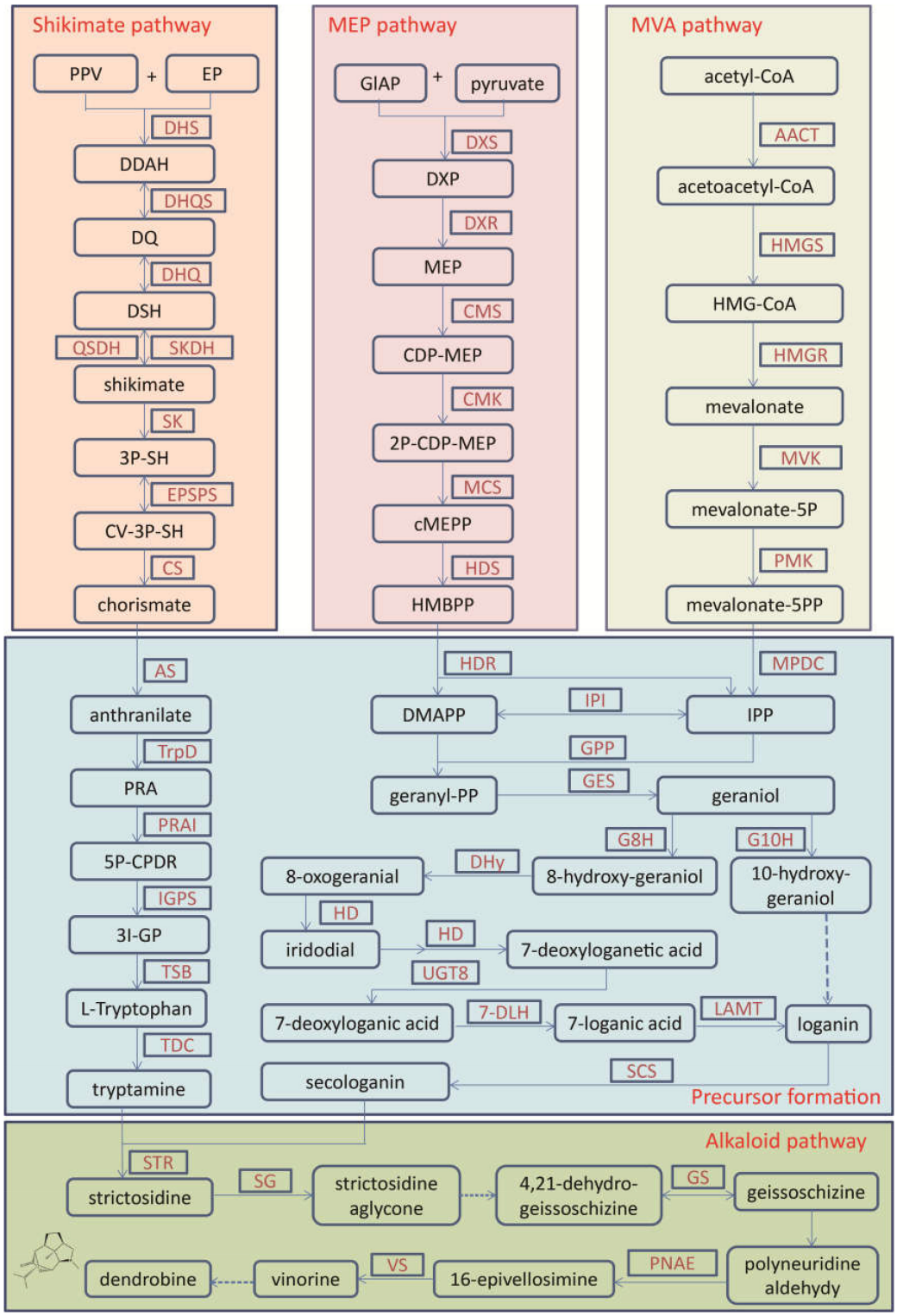


FIGURE 1 | The biosynthetic pathway for terpenoid indole alkaloids in *Dendrobium officinale*. The alkaloid biosynthesis contains the precursor formation from the shikimate, MEP, and MVA pathways and the alkaloid pathway. 2P-CDP-MEP, 2-phospho-4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol; 3I-GP, (3-indoyl)-glycerolphosphate; 3P-SH, shikimate 3-phosphate; 5P-CPDRC, 1-(2-carboxyphenylamino)-1′-deoxy-D-ribulose 5-phosphate; CDP-MEP, 4-(cytidine 5′-diphoshpo)-2-C-methyl-D-erythritol; cMEPP, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; CV-5P-SH, 5-O-(1-carboxyvinyl)-3-phosphoshikimate; DDAH, 7P-2-dehydro-3-deoxy-D-arabino-heptonate; DMAPP, dimethylallyl-PP; DQ, 3-dehydroquinate; DSH, 3-dehydroshikimate; DXP, 1-deoxy-D-xylulose-5-phosphate; *(Continued)*

FIGURE 1 | EP, D-erythrose 4-phosphate; HMBPP, 1-hydroxy-2-methyl-2-butenyl-4-diphosphate; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; IPP, isopentenyl-PP; MEP, 2-C-methyl-D-erythritol-4-phosphate; PPV, phosphoenolpyruvate; PRA, N-(5-phospho-β-D-ribosyl)-anthranilate; 7-DLH, 7-deoxyloganate 7-hydroxylase; AACT, acetyl-CoA acetyltransferase; AS, anthranilate synthase; CMK, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase; CMS, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase; CPR, NADPH-hemoprotein reductase; CS, chorismate synthase; DHQ, 3-dehydroquinate dehydratase; DHQS, 3-dehydroquinate synthase; DHS, 3-deoxy-7-phosphoheptulonate synthase; DHy, dehydrogenase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; DXS, 1-deoxyxylulose-5-phosphate synthetase; EPSPS, 3-phosphoshikimate 1-carboxyvinyltransferase; G10H, geraniol 10-hydroxylase; G8H, geraniol 8-hydroxylase; GPP, geranyl diphosphate diphosphatase; GES, geranyl diphosphate synthase; GS, geissoschizine synthase; HD, hydroxylase; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; HDS, 1-hydroxy-2-methyl-2-(E)-butenyl-4- diphosphate synthase; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; IGPS, indole-3-glycerolphosphate synthase; IPI, IPP isomerase; LAMT, loganate O-methyltransferase; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MPDC, mevalonate diphosphate decarboxylase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; PNAE, polyneuridine-aldehyde esterase; PRAI, phosphoribosylanthranilate isomerase; QSDH, quinate/shikimate dehydrogenase; SCS, secologanin synthase; SG, strictosidine β-D-glucosidase; SK, shikimate kinase; SKDH, shikimate dehydrogenase; STR, strictosidine synthase; TDC, tryptophan decarboxylase; TrpD, anthranilate phosphoribosyltransferase; TSB, tryptophan synthase; UGT8, 7-deoxyloganetic acid glucosyltransferase; VS, vinorine synthase.

two other types of active compounds (alkaloids and flavonoids), there are 98 and 52 genes that could be involved in the biosynthesis pathways, respectively (Niu et al., 2021). A putative enzyme (polyneuridine-aldehyde esterase) may be responsible for the extension of 16-epivellosimine (Yan et al., 2015; **Figure 1**). Moreover, other genes related to distinctive traits were also annotated and identified, including fungus symbiosis, drought resistance, photosynthesis, and flower development (Yan et al., 2015; Zhang G. Q. et al., 2016; Niu et al., 2021). The analysis of those genes suggests their adaptation to certain environments.

Moreover, *D. officinale* has distinct flower architecture, which has attracted researchers' attention. Genomic research provides important insights on flower development in *D. officinale*. The ABCDE model is a classical model for flower architecture. Although there are fewer MADS-box genes (63) in *D. officinale* than that in rice and Arabidopsis, five classes of MADS-box genes all exist, suggesting presence of complete flower genes in *D. officinale* (Yan et al., 2015; Zhang G. Q. et al., 2016). Unique genes, gene expansion and gene loss were observed in MADS-box gene family. Among them, MADS-box ZMM17 gene family, belong to B class, is unique in *D. officinale*, which could be related to the distinct flower architecture (Yan et al., 2015). ANR1, StMADS11, and MIKC* genes generally participate in regulation of growth and development, and their expansion could result in the diversity (Zhang G. Q. et al., 2016). The less number of type I MADS box genes may lead to the failure of endosperm development (Zhang G. Q. et al., 2016). The genomic investigation provides important hues for understanding the distinct and various flowers in *D. officinale*.

The genomic research of *D. officinale* provides abundant information for understanding the potential genetic resources of active component accumulation and insights for some important traits. In the future, confirming the key genes related to the above traits will be a research focus, and Genome-Wide Association Studies (GWAS) could be one of the principle methods for mapping candidate genes.

TRANSCRIPTOMIC RESEARCH IN *DENDROBIUM OFFICINALE*

Transcriptomic sequencing (RNA-seq) has been widely applied to research in medicinal plants including *D. officinale* due to low expenditure, high throughout, high sensitivity and no limitation from genome sequencing requirement. Since the first transcriptome of *D. officinale* stems was reported (Guo et al., 2013), there are now up to 30 transcriptome data of *D. officinale* available in the NCBI (**Table 2**). This research characterizes the biosynthetic pathways of active compounds in *D. officinale*, which provides an important basis for its uses as a medicinal material (**Table 1**; Yuan et al., 2020a). Transcriptomic analysis was also used to investigate the effects of various factors on the accumulation of active compounds, including the material location, hormones and more (Chen et al., 2019; Ren et al., 2020; Zuo et al., 2020). Moreover, the potential molecular mechanisms of a variety of physiological processes were analyzed, such as response to stress, flowering and symbiotic relationships (Wang et al., 2018; Zou et al., 2018; Li H. et al., 2021).

Tissue-specifically expressed genes can be identified through analysis of comparative transcriptomics. A total of 2645, 256, 42, and 54 transcripts were highly expressed in four organs (flower, root, leaf, and stem), respectively (Meng et al., 2016). Those differential expression genes (DEGs) were shown to be involved in organ-specific functions. Among them, eight genes with high transcripts in stem were annotated to the secondary metabolic pathways and carbohydrate transport and metabolism, which was consistent with its abundant metabolites. Moreover, 25 genes were identified to participate in the regulation of flowering, which suggests the presences of classical flowering pathways (Shen et al., 2017).

Alkaloids are one type of the earliest-identified active compounds in *D. officinale* (Tang et al., 2017). Although there are different types of alkaloids accumulated in *D. officinale* (Jiao et al., 2018), the main classes of alkaloids are terpenoid indole alkaloids (TIAs; Guo et al., 2013; Niu et al., 2021). Alkaloids are present in multiple tissues, including protocorm-like bodies (PLB), flower, root, leaf and stem, and the content in PLBs was the highest (Wang et al., 2021). There were 25 genes that participated in the shikimate, mevalonate (MVA) and methylerythritol phosphate (MEP) and strictosidine pathways, and other genes were potentially related to the alkaloid synthesis, such as cytochrome P450s, aminotransferases, methyltransferases, multidrug resistance protein (MDR), transporters and transcription factors (Guo et al., 2013; **Figure 1**). Further expression analysis confirmed that the genes responsible for 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 5-enolpyruvylshikimate-3-phosphate synthase,

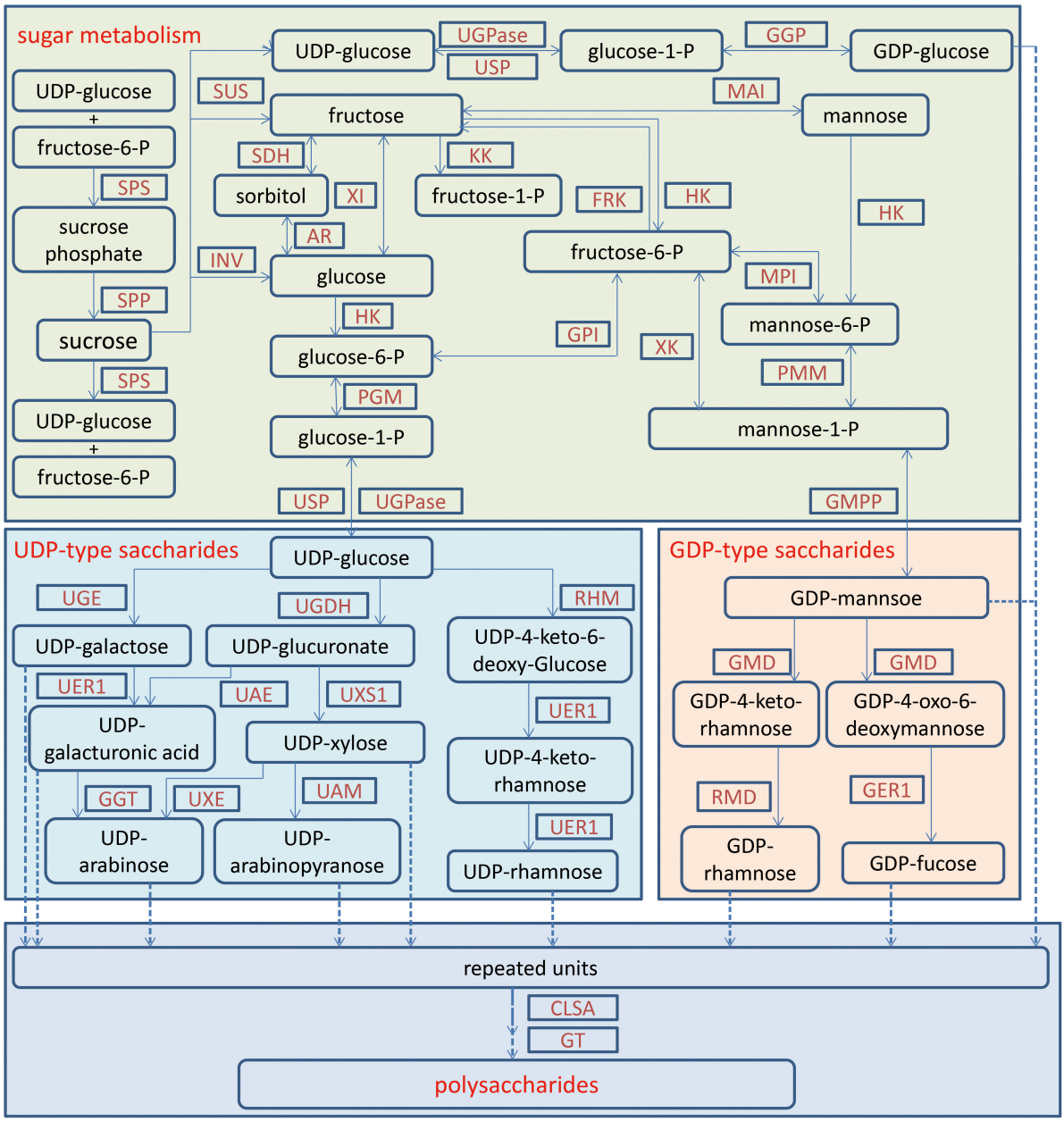


FIGURE 2 | The biosynthetic pathway for polysaccharides in *Dendrobium officinale*. The biosynthesis of polysaccharides contains sugar metabolism, UDP-saccharide metabolism, GDP-saccharide metabolism and polysaccharide formation. AR, aldose reductase; CSLA, cellulose synthase-like A; FRK, fructokinase; GAE, UDP-glucose A-4-epimerase; GER1, GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase; GGP, glucose-1-phosphate guanylyltransferase; GGT, UDP-galacturonate decarboxylase; GMD, GDP-D-mannose-4,6-dehydratase; GMPP, mannose-1-phosphate guanylyl transferase; GPI, glucose-6-phosphate isomerase; GT, glycosyltransferases; HK, hexokinase; INV, invertase; KK, ketohexokinase; MAI, mannose isomerase; MPI, mannose-6-phosphate isomerase; PGM, phosphoglucomutase; PMM, phosphomannomutase; RHM, UDP-glucose 4, 6-dehydratase; RMD, GDP-4-dehydro-D-rhamnose reductase; SDH, sorbitol dehydrogenase; SPP, sucrose phosphate phosphatase; SPS, sucrose phosphate synthase; SuS, sucrose synthase; TSTA3, GDP-l-fucose synthase; UAE, UDP-glucuronate 4-epimerase; UAM, UDP-arabinopyranose mutase; UER1, 3,5-epimerase/4-reductase; UGDH, UDP-glucose dehydrogenase; UGE, UDP-glucuronate 4-epimerase; UGPase, UDP-glucose pyrophosphorylase; USP, UDP-sugar pyrophosphorylase; UXE, UDP-arabinose 4-epimerase; UXS1, UDP-glucuronate decarboxylase; XI, xylose isomerase; XK, xylulokinase.

shikimate dehydrogenase, mevalonate kinase (MVK), and aminotransferases were mainly expressed in leaves, which could be the main reason for higher alkaloid content in leaves (Shen et al., 2017). Recently, 41 genes showing significant differences in expression levels were detected between PLBs and leaves, including genes in the strictosidine biosynthesis (Wang et al., 2021). The candidate genes encoding strictosidine β-D-Glucosidase, geissoschizine synthase and vinorine synthase

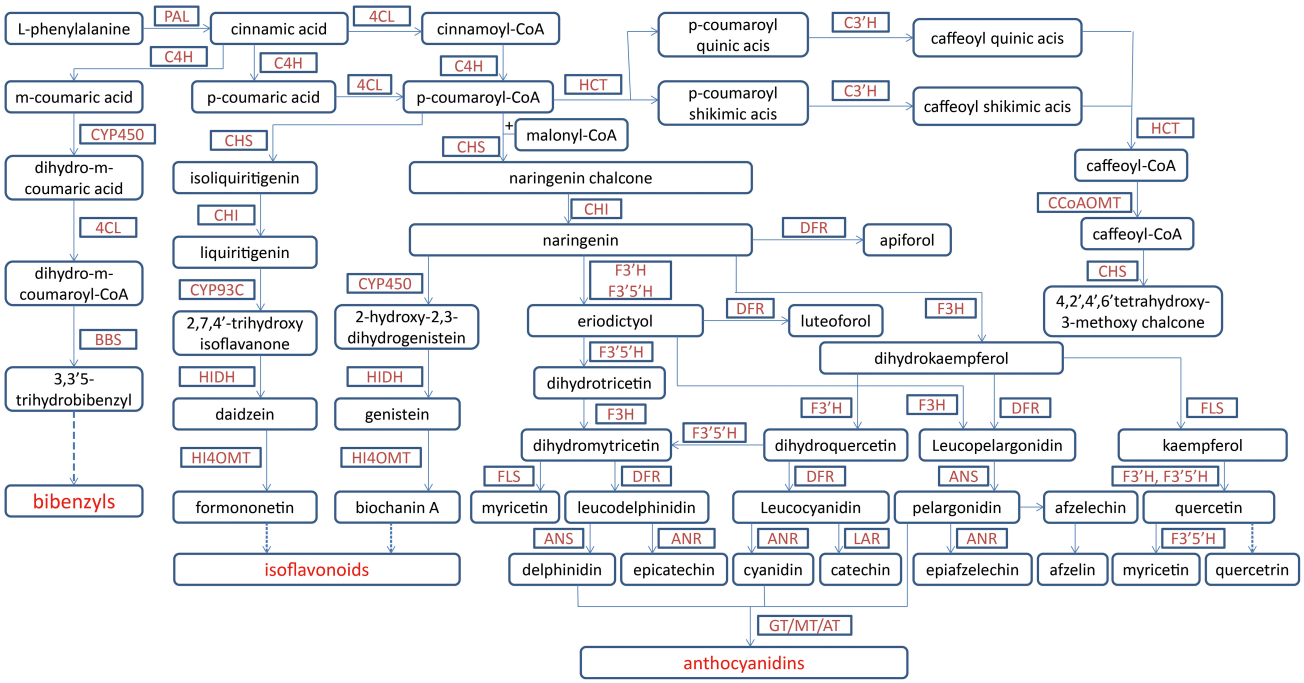


FIGURE 3 | The biosynthetic pathway for phenols in *Dendrobium officinale*. The biosynthesis for main phenols includes anthocyanidins, isoflavonoids, flavones, flavonols and bibenzyls from phenylalanine. 4CL, 4-coumarate-CoA ligase; ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; AT, acyltransferase; BBS, bibenzyl synthase; C3′H, 5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase; C4H, cinnamate 4-hydroxylase; CCoAOMT, caffeoyl-CoA O-methyltransferase; CHI, chalcone isomerase; CHS, chalcone synthase; CYP450, cytochrome P450-dependent monooxygenase; DFR, dihydroflavonol-4-reductase; F3′5′H, flavonoid 3′,5′-hydroxylase; F3′H, flavonoid 3′-monooxygenase; F3H, flavonoid 3-hydroxylase; FLS, flavonol synthase; GT, glycosyltransferase; HCT, quinate O-hydroxycinnamoyltransferase; HI4OMT, hydroxy-isoflavanone-4- O-methyltransferases; HIDH, tri-hydroxy-isoflavanone dehydratase; LAR, leucoanthocyanidin dioxygenase; MT, methyltransferase; PAL, phenylalanine ammonia lyase.

in the alkaloid biosynthesis enzymes were first identified, which facilitates the analysis of its metabolite biosynthesis in future studies.

Polysaccharides in *D. officinale* are the most abundant active compounds in various tissues and adult stems show the highest content (He et al., 2015; Zhang J. et al., 2016; Shen et al., 2017). A metabolic pathway for fructose and mannose biosynthesis has been developed through comparative transcriptomic analysis of different tissues, including 44 genes (He et al., 2015; **Figure 2**). The expression patterns of eight *CELLULOSE SYNTHASE-LIKE A* (*CSLA*) genes were closely linked to the content of mannose in stems (He et al., 2015). Similarly, other putative genes responsible for higher polysaccharides in adult stems were identified, including mannose-1-phosphate guanylyltransferase, GDP-L-fucose synthase and mannose-6-phosphate isomerase (Shen et al., 2017). Using the PacBio sequencing technique, alternative splicing forms of two glycosyltransferase and four cellulose synthase genes have been detected, and two genes encoding SWEET and sucrose transporter showed higher expression in stems than in leaves (He et al., 2017). Although the biosynthesis for polysaccharides is complex, these studies provide hints for further investigation on the molecular mechanism of rich polysaccharides in *D. officinale*.

There are abundant phenols in *D. officinale*. Among them, flavonoids were found in leaves, stems, and roots of *D. officinale*, and the content in leaves was the highest (Yuan et al., 2020b). The biosynthetic pathway of flavonoids was proposed and contained 26 genes including chalcone synthase (CHS), flavanone 3-dioxygenase (F3H), dihydroflavonol reductase (DFR), flavonol synthase (FLS), *trans*-cinnamate 4-monooxygenase, leucoanthocyanidin dioxygenase (LAR), anthocyanidin reductase (ANR), and shikimate O-hydroxycinnamoyl transferase (Yuan et al., 2020b; **Figure 3**). DEGs were enriched in phenylpropanoid and flavonoid biosynthesis. Multiple structural genes in the flavonoid pathway were significantly upregulated in leaves, such as one *LAR1*, one *DFR3*, one *F3H,* and three *CHS* genes, which facilitate the accumulation of flavonoids in leaves and could be the main reason for different flavonoid contents among tissues. Recently, it has been reported that the contents of flavonoids and anthocyanins were different in the stems of *D. officinale* from three provinces (Lei et al., 2018; Ren et al., 2020). Different anthocyanin contents could result from the different activities of some key enzymes, such as hydroxyl cinnamomum acyltransferase 1, UDP-glycosyltransferase (UGT)-83A1, chalcone flavone isomerase-like, and UGT-88B1 serine carboxypeptidase-like 1 (Ren et al., 2020).

Exogenous application of plant hormones can induce accumulation of metabolites in medicinal plants, including in *D. officinale*. Treatments with methyl jasmonate (MeJA) and salicylic acid (SA) resulted in more active compounds

TABLE 2 | A summary of transcriptomic reports in *D. officinale*.

Tissues	Treatments or conditions	Research focus	References
S	–	Alkaloid, genetic markers	Guo et al., 2013
F, R, L, and S	–	Organ-specific study	Meng et al., 2016
L, S, R, and F	–	Polysaccharide synthesis, alkaloid synthesis	Shen et al., 2017
S and L	–	Variant splicing, sugar translocation	He et al., 2017
Eight tissues	–	Expression	Zhang et al., 2017
R, S, and L	–	Flavonoid biosynthesis	Yuan et al., 2020b
L, S, R, and F	–	Flowering	Chen Y. et al., 2017
L, R, S, and F	–	Polysaccharide synthesis	Zhang G. Q. et al., 2016
PLB, L	–	Alkaloid synthesis	Wang et al., 2021
Seedlings	Juvenile and adult	Polysaccharide synthesis	Zhang J. et al., 2016
S	Four stages	Mannan polysaccharides	He et al., 2015
R	Cadmium stress	Stress response	Jiang et al., 2020
L	MeJA	Accumulation of alkaloids	Chen et al., 2019
Seedlings	Far-red light	Shade-avoidance	Li H. et al., 2021
Seeds	Symbiotic and asymbiotic	Germination	Chen J. et al., 2020
L	Drought	Drought stress	Wan et al., 2018
L	CO_2	Crassulacean acid metabolism	Zou et al., 2018
Germinated seeds	Symbiotic and asymbiotic seeds	Symbiotic association	Wang et al., 2018
L	Drought stress and stress removal	Drought stress	Zou et al., 2018
Plants	SA	Polysaccharide, flavonoid, and alkaloid synthesis	Niu et al., 2021
R	MeJA	Bibenzyl biosynthesis	Adejobi et al., 2021
S	Three provinces	Flavonoids	Lei et al., 2018
S	Three provinces	Anthocyanins	Ren et al., 2020
S, L, and R	Three species	Polysaccharide and alkaloid synthesis	Yuan et al., 2020a
Styles	Non-, self-, and cross-pollinated	Self-incompatibility	Chen Y. et al., 2021

S, stems; L, leaves; F, flowers; R, roots; PLB, protocorm-like bodies; Eight tissues – L, S, column, flower buds, lip, sepal, white root, and green root tip.

accumulated in various tissues (Jiao et al., 2018; Chen et al., 2019; Adejobi et al., 2021; Niu et al., 2021). After MeJA treatment the alkaloids were enhanced in leaves of *D. officinale* through upregulating expression levels of the multiple genes in the pathways of MVA and MEP, such as P450 genes, transaminase genes and methyltransferase (Chen et al., 2019). The bibenzyl compounds were mainly accumulated in the roots and greatly increased after MeJA treatment because some key genes involved in the flavonoid and bibenzyl pathways were abundantly expressed after treatment, including P450 and putative bibenzyl synthase genes (Adejobi et al., 2021; **Figure 3**). Presence of SA in the culture medium can increase the accumulation of active components in *D. officinale* seedlings including alkaloids, polysaccharides, and flavonoids (Niu et al., 2021). As many as 107 genes involved in biosynthesis of active components were upregulated. Among them, two key enzyme genes (*F3′H* and *DFR*) involved in the anthocyanin synthesis were significantly upregulated, which may result in color differences between individual plants (Niu et al., 2021). These studies could help us improve the contents of bioactive compounds and supply information for further functional investigation of putative genes.

Transcriptomics can also be used to analyze other physiological processes. In nature *D. officinale* generally forms few seeds potentially because of self-incompatibility. A total of 41 putative genes involved in self-incompatibility were identified, including six Ca^{2+} signal genes and 11 S-locus receptor kinase (SRK) related genes (Chen Y. et al., 2021), which supply helpful insights for genetic mechanisms and preservation of *D. officinale* resources. Moreover, *D. officinale* seeds are small and have limited energy reserves, and the symbiotic relationship between *D. officinale* and mycorrhizal fungi is beneficial for *D. officinale*. RNA-Seq showed that endogenous hormones play a vital role in the seed germination of *D. officinale* (Wang et al., 2018). The relatively low concentrations of JA, abscisic acid (ABA) and strigolactones (SLs) may promote the growth of *D. officinale* (Wang et al., 2018; Chen J. et al., 2020). After inoculation of mycorrhizal fungi, genes associated with gibberellic acid (GA) biosynthesis were upregulated, and GA3 may be a key signaling molecule for germination (Chen J. et al., 2020). These results provide valuable insights for orchid-fungal symbiosis and seed germination in *Orchidaceae*. Under cadmium stress, *D. officinale* showed regulatory responses in roots through metal transporters, sulfate glutathione metabolism, cell wall metabolism, and phenylpropanoid metabolism (Jiang et al., 2020). Furthermore, as an epiphytic plant, *D. officinale* shows strong resistances to environments and can partly use crassulacean acid metabolism under stress conditions (Zhang et al., 2018; Zou et al., 2018). Although transcriptomic data on drought treatments have been obtained, further analysis still needs be performed to explore the mechanisms behind drought resistance (Wan et al., 2018; Zou et al., 2018).

As summarized above, many transcriptomic researches have been carried out in *D. officinale*, including various tissues, treatments or conditions. Tissue-specific patterns and key genes in the biosynthesis of bioactive compounds and responses to stresses have been extensively investigated. This research provides valuable molecular information for understanding these distinct physiological processes. At present, it is urgent to build a public, open database of transcriptional resources for *D. officinale* as the established database for *Arabidopsis thaliana* (The Arabidopsis

Information Resource, TAIR), which can provide easy access to researchers, especially for those who understand the function of genes but are not familiar with bioinformatics tools. In the future, the newest sequencing techniques, especially the third generation platform, should be used for further analysis, such as alternative splicing.

PROTEOMIC RESEARCH IN *DENDROBIUM OFFICINALE*

Because the changes of gene expression profiles at the mRNA level are not always the same as the changes at the protein level, proteomic research in plants can help to reveal the molecular mechanisms of plant growth, development, metabolites and responses to environments. At present, the methods in proteomics include isobaric tags for relative and absolute quantification (iTRAQ), data-independent acquisition (DIA), and tandem mass tags (TMT), which can detect protein expression levels, post-translational modifications and protein-protein interactions of all proteins.

Proteomic research into *D. officinale* is a new avenue of research. The first report on *D. officinale* focused on analysis of root induced by *Mycena dendrobii* (Xu et al., 2015), which shows important proteins related to defense and stress responses, the formation of mycorrhizal fungi and the biosynthesis of bioactive components. These data explain why *M. dendrobii* can promote the growth of *D. officinale* seedlings. The lysine succinylation sites on *D. officinale* proteins were identified (Feng et al., 2017), and five key enzymes in the glycolytic pathway exhibit succinylation of lysine, which will promote understanding the functions of lysine succinylation in plants. Recently, to discriminate *D. officinale* from other *Dendrobium* species, among 343 measurable peptides, 29 peptides were chosen as putative biomarkers, and the short peptide ALGLELDLSER can be a biomarker for the identification of *D. officinale* plants from different geographical areas (Fang et al., 2020).

The aim of proteomics research in medicinal plants is usually to evaluate the active components and proteomes across different samples and dissect the molecular mechanisms of bioactive compounds. Compared with the gene expression profiles at the mRNA level, protein expression patterns can more reliably and accurately reflect the expression level of key enzymes in the biosynthesis pathways of active components. The deep research needs to be performed on the proteome of *D. officinale*, especially on the changes that occur at different stages and under various stresses.

METABOLOMIC RESEARCH IN *DENDROBIUM OFFICINALE*

As metabolites represent the final products of gene expression and protein function, they can reflect changes in plant growth, development and responses to environments. Metabolomics has become a widely used tool in plant research, including wide-targeted metabolomics and targeted metabolomics. Due to bioactive functions of multiple metabolites in *D. officinale*, it is of great significance to analyze those metabolites (**Table 3**).

The different metabolites in different *Dendrobium* species can be used to discriminate between them. There were 11 secondary metabolites that were considered as biomarkers of *D. officinale* and *Dendrobium huoshanense* (Jin et al., 2016). Moreover, in these two species, as many as 133 nitrogenous compounds were identified, and allyl alkaloid is an important medicinal component worth further investigation (Song et al., 2020). For *D. officinale* plants, the third year is the most suitable harvest time based on the accumulation of metabolites (Jin et al., 2016). The metabolites vary often in different tissues of *D. officinale*. The stems were composed of higher levels of multiple metabolites than the leaves, while leaves included higher concentrations levels of polyphenols and lipids (Cao et al., 2019). Moreover, 649 different metabolites were found in the stem and leaf of 6-month-old *D. officinale* (Cao et al., 2019), including organic acids, amino acids, nucleotides, and flavonoids.

Multiple factors influence the accumulation of metabolites. There were 101 volatile compounds detected in stems of *D. officinale* from four regions, and some distinctive compounds were found in certain region (Hu et al., 2020). UV-B radiation can induce accumulation of polysaccharides, alkaloids, and flavonoids in *D. officinale* stem (Chen Y. et al., 2020). The stems of *D. officinale* grown on different substrates contained significantly different metabolites (Zuo et al., 2020), and the mainly changed metabolites were the flavonoids. The stems from plants grown in pine bark showed a higher content of flavonoids, which provides the practical basis of substrate selection for *D. officinale* plants. Furthermore, the accumulation of alkaloids was increased when protocorm-like bodies of *D. officinale* were treated with tryptophan (secologanin) and MeJA (Jiao et al., 2018), and among them changes in 29 metabolites were confirmed, including carapanaubine, a kind of TIAs.

TABLE 3 | A summary of metabolomic studies in *D. officinale*.

Materials	Treatments or conditions	Method	References
Stems	Three different substrates	UPLC-MS/MS	Zuo et al., 2020
Protocorm-like bodies	Precursors and methyl jasmonate	GC-MS, LC-MS	Jiao et al., 2018
Stems	Different growth years of two species	GC-MS	Jin et al., 2016
Stems	UV-B treatment	UPLC-MS/MS	Chen Y. et al., 2020
Caulis	Two *Dendrobium* plants	LC-MS	Song et al., 2020
Leaves and stems	–	LC-ESI-MS/MS	Cao et al., 2019
Stems	Different regions	GC–MS	Hu et al., 2020

TABLE 4 | A summary of omics integration in *D. officinale*.

Samples	Treatments or conditions	Combinations of omics	Research focus	References
Stems	Purple and normal varieties	UPLC-ESI-MS/MS, RNA-seq	Pigmentation	Zhan et al., 2020
Flowers	Buds and opened flower	RNA-seq, HPLC-MS	Flower development	He et al., 2020
Seeds	Asymbiotic and symbiotic at different stages	RNA-Seq, iTRAQ	Symbiotic germination	Chen J. et al., 2017
Leaves	at 0°C and at 20°C	GC-MS, RNA-seq	Cold acclimation	Wu et al., 2016
Leaves	NaCl treatment	RNA-seq, LC–MS/MS	Salt stress	Zhang et al., 2021
Stems	Fungus MF23	RNA-seq, UHPLC-LTQ	Fungus on growth	Shan et al., 2021
Flowers	Two cultivars	RNA-seq, GC-MS	Volatile terpenoids	Li N. et al., 2021

To date, metabolomics studies into *D. officinale* remain limited. However, with technological development, more attention could be placed on this area because it is home to important medicinal properties in need of elucidation. Metabolomics studies not only provide us insights of metabolite accumulation under different conditions, but also important potential management methods that can improve the contents of target metabolites. Moreover, although wide-targeted metabolomics can provide huge information regarding metabolites, more focuses needs be paid to the active compounds through targeted metabolomics in *D. officinale*, which can provide more accurate information for certain type of metabolites.

INTEGRATION RESEARCH OF MULTI-OMICS IN *DENDROBIUM OFFICINALE*

The process of plant growth and development is complex and changeable. The plant status is the result of interactions between the genetic information and environmental factors. Integration of multi-omics can provide more comprehensive information to understand the dynamic changes and the potential mechanisms of various physiological processes. At present, there are some reports on the integration of different omics in *D. officinale* as shown in **Table 4**. Mainly, transcriptomics and metabolomics are associated to dissect changes in gene expression and metabolite accumulation in different tissues or under special conditions, which can be used to construct a regulatory network.

The combination of metabolomic and transcriptomic analysis has also widely been used to analyze various physiological processes in *D. officinale*, such as flower development (He et al., 2020), cold acclimation (Wu et al., 2016), stem color (Zhan et al., 2020), salt stress (Zhang et al., 2021), and flower volatile terpenoids (Li N. et al., 2021). Some key factors have been identified to explain why the metabolites were changed, including the key enzymes in the biosynthetic pathway and potential transcription factors (Li N. et al., 2021). For example, the increase in content of delphinidin and quercetin derivatives was responsible for the purple stem of *D. officinale* and *F3H* and leucoanthocyanidin dioxygenase (*LDOX*) genes were found to be highly expressed (Zhan et al., 2020). Through analysis of seed germination by transcriptomics and proteomics (Chen J. et al., 2017), 32 genes showed consistent levels at mRNA and protein level and some of them were related to lipid and glycometabolism, which may be induced by fungi during germination, offering valuable insights into seed germination for orchid plants. *Mycena* sp. can promote the growth of *D. officinale* due to better nitrogen uptake and NH_4^+ assimilation (Shan et al., 2021), which supplies the molecular basis for cultivation.

Reports on combination of different omics are presently limited in *D. officinale*. However, the ample advantages of the approach are amply demonstrated (**Table 4**). The molecular status of plants can comprehensively be reflected at different levels, which is conducive to screening for trait-related genes and for further investigation.

PERSPECTIVES

Due to its high medicinal and economic values, *D. officinale* has attracted significant attention from researchers. Thanks to the rapid development of sequencing and analysis methods, close relationships have been built between traditional Chinese medicine and genetic basis (**Tables 1–3**). Various biological processes of medicinal plants including *D. officinale* are now increasingly being investigated. Among them, the omics methods have been considered as a powerful tool to obtain a huge amount of information, including gene sequences, gene expression patterns, protein expression profiles, and metabolites in *D. officinale*. Therefore, in-depth studies of the molecular mechanisms underlying the biological processes are now possible and credible, such as the accumulation of active components. In recent years, many breakthroughs have been made in the study of *D. officinale*. Of particular note, the third version of the *D. officinale* genome at the chromosome level has recently been reported (Niu et al., 2021), which facilitates future genomic research. High-quality genomic data of *D. officinale* can have a vital influence on further studies into the accumulation and regulation of active components and molecular breeding of *D. officinale*. Transcriptomics of *D. officinale* is mainly focused on the key enzyme genes in the pathway of active component synthesis and analyzing its metabolic pathway.

Although great strides have been made in deepening our understanding of *D. officinale*, limitations remain. First, many genes related to active components in *D. officinale* have been identified by genomic and transcriptomic studies (**Figures 1–3**), but the functions of most these genes have not been confirmed through functional investigation, and their regulatory mechanisms have not yet been elucidated. For example, MeJA has been shown to induce the accumulation of polysaccharides,

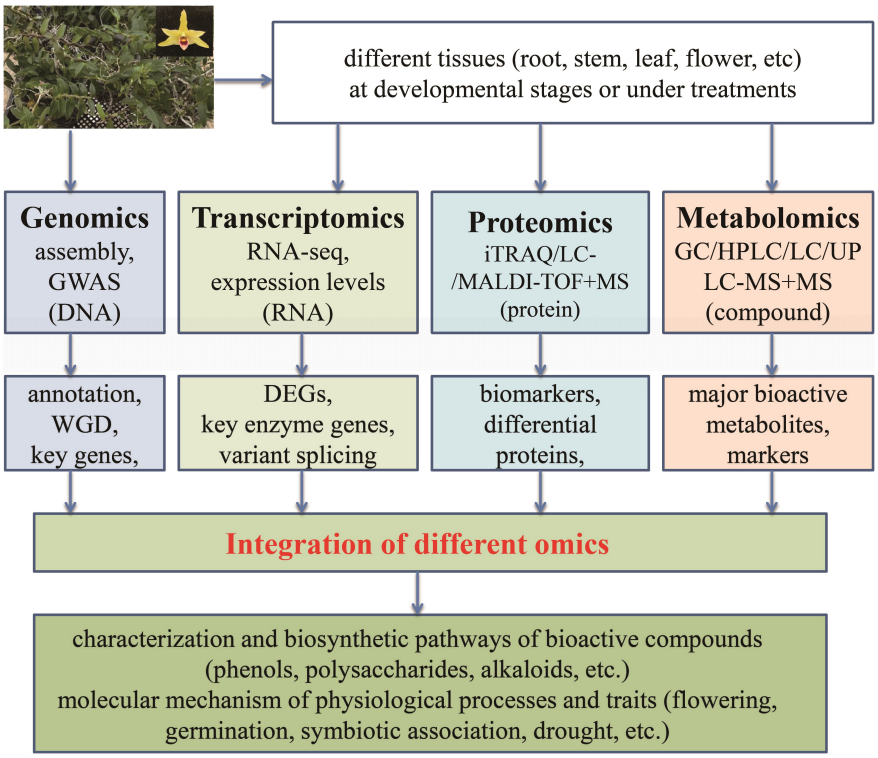


FIGURE 4 | Application of multiple omics in *Dendrobium officinale*. The integration of data from different levels (DNA, RNA, protein, and metabolites) can make it feasible and reliable to discover the potentially key factors (transcription factor, gene, protein, or compound) responsible for bioactive compounds as well as physiological processes and traits.

flavonoids, and alkaloids in *D. officinale* (Yuan et al., 2017; Jiao et al., 2018; Chen et al., 2019; Adejobi et al., 2021). However, the key factors in the JA signal pathway still need to be identified and clarified. Second, *D. officinale* plants generally live in extremely harsh environments and show strong drought resistance, the molecular mechanism of which is still limitedly clarified. There is a large amount of transcriptomic data available on *D. officinale* treated under different drought conditions (Wan et al., 2018; Zou et al., 2018). However, these data have not yet been fully and deeply analyzed. Therefore, there is an urgent need to excavate the data to reveal the molecular mechanisms of drought resistance in *D. officinale*.

In order to promote further researches into *D. officinale*, attentions should be paid to the following work in the future: (1) Although current omics technologies have been used to study *D. officinale*, there are limitedly comprehensive multi-omics studies. Obviously, one single omics method such as transcriptomics, proteomics, or metabolomics cannot satisfy the necessary deep research into *D. officinale*. Therefore, it is imperative to study the mechanisms that underlie biosynthesis and the regulation of active compounds in *D. officinale via* the integration of multi-omics (**Figure 4**). In particular, the combination of proteomics and metabolomics has not yet been carried out for *D. officinale*. (2) The genuine resource of medicinal plants is essential for their functions, including *D. officinale*. Therefore, it is important to comprehensively investigate why differences exist in active compounds between plants from different cultivation regions or conditions at the genome, mRNA and protein levels, which can provide propitiate conditions for metabolite accumulation. (3) Based on the whole genome of *D. officinale*, resequencing of different germplasm resources can help to screen the key genes responsible for interesting traits for further functional dissection. In particular, studies on the mechanisms of biosynthesis and regulation of growth and development of active ingredients should be continued. (4) Although some genes belonging to the biosynthetic pathway of the main active ingredients in *D. officinale* have been identified, the accurate functions of only a few genes have been confirmed. Studies on key genes related to the biosynthesis of active components, growth regulation, and stress resistance of *D. officinale* require further investigated. Recently, the potential transformation system for *Dendrobium* plants was reported (Li Y. et al., 2021), which could benefit the functional investigation into *D. officinale*.

AUTHOR CONTRIBUTIONS

AL: conceptualization and writing – review and revising. YueW, OA, YT, and YuhW: literature search and data analysis. YueW: writing – original draft. All authors have read and agreed to the published version of the manuscript.

FUNDING

We are grateful for funds supported from the National Natural Science Foundation of China (grant number 31701465), Yunnan Provincial Science and Technology Department (grant number 202101AT070189), and Beijing DR PLANT Biotechnology Co., Ltd. (E0514832C1). This study received funding from Beijing DR PLANT Biotechnology Co. Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. All authors declare no other competing interests.

ACKNOWLEDGMENTS

We appreciate Austin G. Smith (ICRAF) for language improvement.

REFERENCES

Adejobi, O. I, Guan, J., Yang, L., Hu, J. M., Yu, A., Muraguri, S., et al. (2021). Transcriptomic analyses shed light on critical genes associated with bibenzyl biosynthesis in *Dendrobium officinale*. *Plants* 10:633. doi: 10.3390/plants10040633

Cao, H., Ji, Y., Li, S., Lu, L., Tian, M., Yang, W., et al. (2019). Extensive metabolic profiles of leaves and stems from the medicinal plant *dendrobium officinale* kimura et migo. *Metabolites* 9:215. doi: 10.3390/metabo9100215

Chen, J., Liu, S. S., Kohler, A., Yan, B., Luo, H. M., Chen, X. M., et al. (2017). iTRAQ and RNA-Seq analyses provide new insights into regulation mechanism of symbiotic germination of *Dendrobium officinale* Seeds (Orchidaceae). *J. Proteome Res.* 16, 2174–2187. doi: 10.1021/acs.jproteome.6b00999

Chen, J., Yan, B., Tang, Y., Xing, Y., Li, Y., Zhou, D., et al. (2020). Symbiotic and asymbiotic germination of *Dendrobium officinale* (Orchidaceae) respond differently to exogenous gibberellins. *Int. J. Mol. Sci.* 21:6104. doi: 10.3390/ijms21176104

Chen, W., Lu, J., Zhang, J. Wu, J., Yu, L., Qin, L., et al. (2021). Traditional Uses, phytochemistry, pharmacology, and quality control of *dendrobium officinale* kimura et. migo. *Front. Pharmacol.* 12:726528. doi: 10.3389/fphar.2021.726528

Chen, Y., Hu, B., Zhang, F., Luo, X., and Xie, J. (2021). Cytological observation and transcriptome comparative analysis of self-pollination and cross-pollination in *dendrobium officinale*. *Genes* 12:432. doi: 10.3390/genes12030432

Chen, Y., Shen, Q., Lin, R., Zhao, Z., Shen, C., and Sun, C. (2017). De novo transcriptome analysis in Dendrobium and identification of critical genes associated with flowering. *Plant Physiol. Biochem.* 119, 319–327. doi: 10.1016/j.plaphy.2017.09.008

Chen, Y., Shen, Q., Lv, P., and Sun, C. (2020). Comparative metabolomic analyses of *Dendrobium officinale* Kimura et Migo responding to UV-B radiation reveal variations in the metabolisms associated with its bioactive ingredients. *PeerJ* 8:e9107. doi: 10.7717/peerj.9107

Chen, Y., Wang, Y., Lyu, P., Chen, L., Shen, C., and Sun, C. (2019). Comparative transcriptomic analysis reveal the regulation mechanism underlying MeJA-induced accumulation of alkaloids in *Dendrobium officinale*. *J. Plant Res.* 132, 419–429. doi: 10.1007/s10265-019-01099-6

Fang, C., Xin, G. Z., Wang, S. L., Wei, M. M., Wu, P., Dong, X. M., et al. (2020). Discovery and validation of peptide biomarkers for discrimination of Dendrobium species by label-free proteomics and chemometrics. *J. Pharm. Biomed. Anal.* 182:113118. doi: 10.1016/j.jpba.2020.113118

Feng, S., Jiao, K., Guo, H., Jiang, M., Hao, J., Wang, H., et al. (2017). Succinyl-proteome profiling of *Dendrobium officinale*, an important traditional Chinese orchid herb, revealed involvement of succinylation in the glycolysis pathway. *BMC Genomics* 18:598. doi: 10.1186/s12864-017-3978-x

Guo, X., Li, Y., Li, C., Luo, H., Wang, L., Qian, J., et al. (2013). Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene* 527, 131–138. doi: 10.1016/j.gene.2013.05.073

He, C., Liu, X., Teixeira da Silva, J. A., Liu, N., Zhang, M., and Duan, J. (2020). Transcriptome sequencing and metabolite profiling analyses provide comprehensive insight into molecular mechanisms of flower development in *Dendrobium officinale* (Orchidaceae). *Plant Mol. Biol.* 104, 529–548. doi: 10.1007/s11103-020-01058-z

He, C., Zhang, J., Liu, X., Zeng, S., Wu, K., Yu, Z., et al. (2015). Identification of genes involved in biosynthesis of mannan polysaccharides in *Dendrobium officinale* by RNA-seq analysis. *Plant Mol. Biol.* 88, 219–231. doi: 10.1007/s11103-015-0316-z

He, L., Fu, S., Xu, Z., Yan, J., Xu, J., Zhou, H., et al. (2017). Hybrid Sequencing of Full-Length cDNA Transcripts of Stems and Leaves in *Dendrobium officinale*. *Genes* 8:257. doi: 10.3390/genes8100257

Hu, J., Huang, W., Zhang, F., Luo, X., Chen, Y., and Xie, J. (2020). Variability of volatile compounds in the medicinal plant *dendrobium officinale* from different regions. *Molecules* 25:5046. doi: 10.3390/molecules25215046

Jiang, W., Wu, Z., Wang, T., Mantri, N., Huang, H., Li, H., et al. (2020). Physiological and transcriptomic analyses of cadmium stress response in *Dendrobium officinale* seedling. *Plant Physiol. Biochem.* 148, 152–165. doi: 10.1016/j.plaphy.2020.01.010

Jiao, C., Song, C., Zheng, S., Zhu, Y., Jin, Q., Cai, Y., et al. (2018). Metabolic profiling of *dendrobium officinale* in response to precursors and methyl jasmonate. *Int. J. Mol. Sci.* 19:728. doi: 10.3390/ijms19030728

Jin, Q., Jiao, C., Sun, S., Song, C., Cai, Y., Lin, Y., et al. (2016). Metabolic analysis of medicinal *dendrobium officinale* and *dendrobium huoshanense* during different growth years. *PLoS One.* 11:e0146607. doi: 10.1371/journal.pone.0146607

Lei, Z., Zhou, C., Ji, X., Wei, G., Huang, Y., Yu, W., et al. (2018). Transcriptome Analysis Reveals genes involved in flavonoid biosynthesis and accumulation in *Dendrobium catenatum* From Different Locations. *Sci. Rep.* 8:6373. doi: 10.1038/s41598-018-24751-y

Li, H., Ye, W., Wang, Y., Chen, X., Fang, Y., and Sun, G. (2021). RNA sequencing-based exploration of the effects of far-red light on lncRNAs involved in the shade-avoidance response of *D. officinale*. *PeerJ* 9:e10769. doi: 10.7717/peerj.10769

Li, N., Dong, Y., Lv, M., Qian, L., Sun, X., Liu, L., et al. (2021). Combined analysis of volatile terpenoid metabolism and transcriptome reveals transcription factors related to terpene synthase in two cultivars of *dendrobium officinale* flowers. *Front. Genet.* 12:661296. doi: 10.3389/fgene.2021.661296

Li, Y., Zhang, B., and Yu, H. (2021). Kilobase-scale genomic deletion of DOTFL1 in dendrobium orchids. *J. Genet. Genomics* 2021:8. doi: 10.1016/j.jgg.2021.07.008

Meng, Y., Yu, D., Xue, J., Lu, J., Feng, S., Shen, C., et al. (2016). A transcriptome-wide, organ-specific regulatory map of *Dendrobium officinale*, an important traditional Chinese orchid herb. *Sci. Rep.* 6:18864. doi: 10.1038/srep18864

Ng, T. B., Liu, J., Wong, J. H., Ye, X., WingSze, S. C., Tong, Y., et al. (2012). Review of research on *Dendrobium*, a prized folk medicine. *Appl. Microbiol. Biotechnol.* 93, 1795–1803. doi: 10.1007/s00253-011-3829-7

Niu, Z., Zhu, F., Fan, Y., Li, C., Zhang, B., Zhu, S., et al. (2021). The chromosome-level reference genome assembly for *Dendrobium officinale* and its utility of functional genomics research and molecular breeding study. *Acta Pharmaceutica Sinica B.* 2021:19. doi: 10.1016/j.apsb.2021.01.019

Ren, Z., Qiu, F., Wang, Y., Yu, W., Liu, C., Sun, Y., et al. (2020). Network analysis of transcriptome and lc-ms reveals a possible biosynthesis pathway of anthocyanins in *dendrobium officinale*. *Biomed. Res. Int.* 2020:6512895. doi: 10.1155/2020/6512895

Shan, T., Zhou, L., Li, B., Chen, X., Guo, S., Wang, A., et al. (2021). The plant growth-promoting fungus MF23 (Mycena sp.) increases production of *dendrobium officinale* (orchidaceae) by affecting nitrogen uptake and NH_4^+ Assimilation. *Front. Plant Sci.* 12:693561. doi: 10.3389/fpls.2021.693561

Shen, C., Guo, H., Chen, H., Shi, Y., Meng, Y., Lu, J., et al. (2017). Identification and analysis of genes associated with the synthesis of bioactive constituents in *Dendrobium officinale* using RNA-Seq. *Sci. Rep.* 7:187. doi: 10.1038/s41598-017-00292-8

Song, C., Jiao, C., Jin, Q., Chen, C., Cai, Y., and Lin, Y. (2020). Metabolomics analysis of nitrogen-containing metabolites between two *Dendrobium* plants. *Physiol. Mol. Biol. Plants* 26, 1425–1435. doi: 10.1007/s12298-020-00822-1

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

Wan, X., Zou, L. H., Zheng, B. Q., Tian, Y. Q., and Wang, Y. (2018). Transcriptomic profiling for prolonged drought in *Dendrobium catenatum*. *Sci. Data* 5:180233. doi: 10.1038/sdata.2018.233

Wang, F., and Shi, H. (2019). Research on the development of *dendrobium candidum* industry in china. *China Foresty Econ.* 3, 88–90.

Wang, T., Song, Z., Wang, X., Xu, L., Sun, Q., and Li, L. (2018). Functional insights into the roles of hormones in the *dendrobium officinale*-tulasnella sp. germinated seed symbiotic association. *Int. J. Mol. Sci.* 19:3484. doi: 10.3390/ijms19113484

Wang, Y. H. (2021). Traditional uses, chemical constituents, pharmacological activities, and toxicological effects of *Dendrobium* leaves: a review. *J. Ethnopharmacol.* 270:113851. doi: 10.1016/j.jep.2021.113851

Wang, Z., Jiang, W., Liu, Y., Meng, X., Su, X., Cao, M., et al. (2021). Putative genes in alkaloid biosynthesis identified in *Dendrobium officinale* by correlating the contents of major bioactive metabolites with genes expression between Protocorm-like bodies and leaves. *BMC Genomics* 22:579. doi: 10.1186/s12864-021-07887-6

Wang, Z., Zhao, M., Cui, H., Li, J., and Wang, M. (2020). Transcriptomic landscape of medicinal *dendrobium* reveals genes associated with the biosynthesis of bioactive components. *Front. Plant Sci.* 11:391. doi: 10.3389/fpls.2020.00391

Wu, Z. G., Jiang, W., Chen, S. L., Mantri, N., Tao, Z. M., and Jiang, C. X. (2016). Insights from the cold transcriptome and metabolome of *dendrobium officinale*: global reprogramming of metabolic and gene regulation networks during cold acclimation. *Front. Plant Sci.* 7:1653. doi: 10.3389/fpls.2016.01653

Xu, X. B., Ma, X. Y., Lei, H. H., Song, H. M., Ying, Q. C., Xu, M. J., et al. (2015). Proteomic analysis reveals the mechanisms of *Mycena dendrobii* promoting transplantation survival and growth of tissue culture seedlings of *Dendrobium officinale*. *J. Appl. Microbiol.* 118, 1444–1455. doi: 10.1111/jam.12781

Yan, L., Wang, X., Liu, H., Tian, Y., Lian, J., Yang, R., et al. (2015). The genome of *dendrobium officinale* illuminates the biology of the important traditional chinese orchid herb. *Mol. Plant.* 8, 922–934. doi: 10.1016/j.molp.2014.12.011

Yuan, Y., Zhang, B., Tang, X., Zhang, J., and Lin, J. (2020a). Comparative transcriptome analysis of different *dendrobium* species reveals active ingredients-related genes and pathways. *Int. J. Mol. Sci.* 21:861. doi: 10.3390/ijms21030861

Yuan, Y., Zhang, J., Liu, X., Meng, M., Wang, J., and Lin, J. (2020b). Tissue-specific transcriptome for *Dendrobium officinale* reveals genes involved in flavonoid biosynthesis. *Genomics* 112, 1781–1794. doi: 10.1016/j.ygeno.2019.10.010

Yuan, Z. Q., Zhang, J. Y., and Liu, T. (2017). Enhancement of polysaccharides accumulation in *Dendrobium officinale* by exogenously applied methyl jasmonate. *Biological. Plantarum.* 61, 438–444. doi: 10.1007/s10535-016-0702-7

Zhan, X., Qi, J., Zhou, B., and Mao, B. (2020). Metabolomic and transcriptomic analyses reveal the regulation of pigmentation in the purple variety of *Dendrobium officinale*. *Sci. Rep.* 10:17700. doi: 10.1038/s41598-020-74789-0

Zhang, G. N., Bi, Z. M., Wang, Z. T., Xu, L. S., and Xu, G. J. (2003). Advances in studies on chemical constituents from plants of *Dendrobium* Sw. *Chinese Trad. Herb. Drugs* 34, S5–S8.

Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R., Hsiao, Y. Y., Niu, S. C., et al. (2017). The Apostasia genome and the evolution of orchids. *Nature* 381:549. doi: 10.1038/nature23897

Zhang, G. Q., Xu, Q., Bian, C., Tsai, W. C., Yeh, C. M., Liu, K. W., et al. (2016). The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* 6:19029. doi: 10.1038/srep19029

Zhang, J., He, C., Wu, K., Teixeira da Silva, J. A., Zeng, S., Zhang, X., et al. (2016). Transcriptome analysis of *dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis. *Front. Plant Sci.* 7:5. doi: 10.3389/fpls.2016.00005

Zhang, M., Yu, Z., Zeng, D., Si, C., Zhao, C., Wang, H., et al. (2021). Transcriptome and metabolome reveal salt-stress responses of leaf tissues from *dendrobium officinale*. *Biomolecules* 5, 736–736. doi: 10.3390/BIOM11050736

Zhang, S., Yang, Y., Li, J., Qin, J., Zhang, W., Huang, W., et al. (2018). Physiological diversity of orchids. *Plant Divers* 40, 196–208. doi: 10.1016/j.pld.2018.06.003

Zou, L. H., Wan, X., Deng, H., Zheng, B. Q., Li, B. J., and Wang, Y. (2018). RNA-seq transcriptomic profiling of crassulacean acid metabolism pathway in *Dendrobium catenatum*. *Sci. Data* 5:180252. doi: 10.1038/sdata.2018.252

Zuo, S. M., Yu, H. D., Zhang, W., Zhong, Q., Chen, W., Chen, W., et al. (2020). Comparative metabolomic analysis of *dendrobium officinale* under different cultivation substrates. *Metabolites* 10:325. doi: 10.3390/metabo10080325

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fpls.2021.830343

Unraveling the Glucosylation of Astringency Compounds of Horse Chestnut *via* Integrative Sensory Evaluation, Flavonoid Metabolism, Differential Transcriptome, and Phylogenetic Analysis

Qinggang Yin[1]*, *Yiding Wei*[1], *Xiaoyan Han*[2], *Jingwang Chen*[3], *Han Gao*[1] and *Wei Sun*[1]*

[1] *Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China,* [2] *Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing, China,* [3] *Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, Beijing, China*

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Zheyong Xue,
Northeast Forestry University, China
Xiaolan Jiang,
Anhui Agricultural University, China
Peng Di,
Jilin Agriculture University, China

***Correspondence:**
Qinggang Yin
qgyin@icmm.ac.cn
Wei Sun
wsun@icmm.ac.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *07 December 2021*
Accepted: *27 December 2021*
Published: *03 February 2022*

Citation:
Yin Q, Wei Y, Han X, Chen J, Gao H and Sun W (2022) Unraveling the Glucosylation of Astringency Compounds of Horse Chestnut via Integrative Sensory Evaluation, Flavonoid Metabolism, Differential Transcriptome, and Phylogenetic Analysis. Front. Plant Sci. 12:830343. doi: 10.3389/fpls.2021.830343

The seeds of Chinese horse chestnut are used as a source of starch and escin, whereas the potential use of whole plant has been ignored. The astringency and bitterness of tea produced from the leaves and flowers were found to be significantly better than those of green tea, suggesting that the enriched flavonoids maybe sensory determinates. During 47 flavonoids identified in leaves and flowers, seven flavonol glycosides in the top 10 including astragalin and isoquercitrin were significantly higher content in flowers than in leaves. The crude proteins of flowers could catalyze flavonol glucosides' formation, in which three glycosyltransferases contributed to the flavonol glucosylation were screened out by multi-dimensional integration of transcriptome, evolutionary analyses, recombinant enzymatic analysis and molecular docking. The deep exploration for flavonol profile and glycosylation provides theoretical and experimental basis for utilization of flowers and leaves of *Aesculus chinensis* as additives and dietary supplements.

Keywords: ***Aesculus chinensis*, kaempferol, astragalin, glycosides, glycosylation**

INTRODUCTION

The Chinese horse chestnut (*Aesculus chinensis*) is a deciduous tree native to Qinling Mountains of China. The seeds of *A. chinensis* are used as a source of starch via alkali treatment or high-temperature detoxification, as well as a source of escin, which is a pentacyclic triterpenoid saponin (Zlatanov et al., 2013; Cheng et al., 2018; Zhang et al., 2020a). The seeds of the Japanese horse chestnut (*A. turbinata*) have been an important food resource since ancient times and are a good source of flavonol O-glycosides, which possess considerable antioxidative capacities for use as food additives and dietary supplements (Kimura et al., 2017). Moreover, some *Aesculus* spp. contain components that are beneficial to human health, especially flavonoids (Wei et al., 2004; Kapusta et al., 2007; Zlatanov et al., 2013; Oszmiański et al., 2014; Zhang et al., 2020b; Jarzebski et al., 2021). Various herbal tea products have been developed from the flowers of the

European horse chestnut (*A. hippocastanum*) and the buds of *A. chinensis*. However, as compared with the widespread use of the seeds of *A. chinensis*, the flowers and leaves have been mostly ignored due to the lack of detailed analysis of key compounds that are potentially beneficial to human health.

Flavonoids, which are abundant in various foods of plant origin, possess antioxidative and anticancer activities, but also contribute to the taste of tea, ginkgo leaves, tartary buckwheat, ginseng leaves, and lotus leaves or seeds (Cui et al., 2016; Su et al., 2017; Cao et al., 2019; Yin et al., 2020; Feng et al., 2021). The sweet aftertaste of tea is due to the contents of (-)-epigallocatechin and (-)-epicatechin, while flavonol glycosides are associated with the astringency and bitterness of tea infusions (Scharbert et al., 2004; Scharbert and Hofmann, 2005; Cao et al., 2019; Dong et al., 2019; Guo et al., 2021). The seeds of *A. chinensis, A. hippocastanum*, *A. carea*, and *A. turbinate* contain various flavonol glycosides (Kapusta et al., 2007; Kimura et al., 2017; Cheng et al., 2018; Zhang et al., 2020a). Glycosides of quercetin and kaempferol have been identified in the leaves of *A. hippocastanum* and *A. carea* (Oszmiański et al., 2014). Kaempferol, which exerts defensive effects against oxidative damage to the brain tissues induced by various types of agents, can cross the blood brain barrier and reduce neuronal damage (Chen and Chen, 2013). Quercetin is most commonly used for conditions of the heart and blood vessels and to prevent cancer (Xu et al., 2018a). The leaves and flowers of *A. chinensis* would been explored as potential food sources if the flavonoid profiles were similar to that of same genus reported.

Flavonoid glycosides are catalyzed by uridine diphosphate glycosyltransferases (UGTs), which transfer glycosyl moieties from UDP sugars to a wide range of acceptor molecules (Li et al., 2021). Different glycosyltransferases catalyze the formation of various flavonoid mono-glycosides. FaGT1 (*Fragaria* x *ananassa*) derived from the strawberry (Griesser et al., 2008), GmUGT88A13 (*Glycine max*) from the soybean (Yin et al., 2017a), and LjUGT72AD1 from *Lotus japonicus* are reported to glucosylate the 3-OH group of flavonols (Yin et al., 2017b). In tartary buckwheat (*Fagopyrum tataricum*), FtUGT73BE5 is the key molecule in the first step glycosylation of rutin (Yin et al., 2020). In ginseng (*Panax gensing*), PgUGT92A10 and PgUGT94Q4 contribute to the formation of kaempferol 3-O-glucoside, as determined by differentiated data-independent acquisition proteomics and phylogenetic analysis (Yin et al., 2021). In ginkgo (*Ginkgo biloba*), GbUGT716A1 possesses unique activities toward flavanol gallates and conveys significant bioactivity beneficial to human health (Su et al., 2017). However, relatively few reports have been identified and characterized as UGT proteins involved in the biosynthesis of flavonoid glycosides in *Aesculus* genus including *A. chinensis*.

In this study, sensory evaluation indicated that herbal tea produced from *A. chinensis* had an excellent taste. Flavonoid profiles of the leaves and flowers revealed that quercetin or kaempferol glycosides may contribute to the strong astringency of herbal tea. Crude protein enzymatic testing indicated that UGTs may contribute to the abundant accumulation of astragalin (kaempferol 3-O-glucoside) in flowers. The sequenced transcriptomes of the flower and seed of *A. chinensis* were used to construct a database of AcUGTs. Screening of 170 AcUGTs identified three that participate in the formation of isoquercitrin (quercetin 3-O-glucoside) and astragalin, which are abundant in the flowers of *A. chinensis*. These results demonstrated that these AcUGT genes, which are responsible for the accumulation of flavonol glycosides in *A. chinensis*, are important to the flavor of herbal tea, and should prove helpful for the development of healthy foods derived from *A. chinensis*.

MATERIALS AND METHODS

Plant Materials and Chemicals

A. chinensis leaves, flowers, and seeds were obtained from the Institute of Medicinal Plants of the Chinese Academy of Medical Sciences (Beijing, China), then immediately washed, frozen in liquid nitrogen, and stored at −80°C for later use. Some leaves and flowers were naturally dried to produce an herbal tea. Green tea made in spring of 2021 was purchased in the market.

The pMAL-c2x expression vector was preserved by the Biotechnology Center of the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences (Beijing, China). UDP-glucose was purchased from Sigma-Aldrich Corporation (St. Louis, MO, USA). Standard flavonols (kaempferol, quercetin, quercitrin, isoquercitrin, and astragalin) were purchased from Shanghai Yuanye Biotechnology Co., Ltd. (Shanghai, China). Chemicals were either analytically pure or high-performance liquid chromatography grade.

Sensory Evaluation of Green Tea and Herbal Tea Infusions

According to Cao et al. method (Cao et al., 2019), green tea or herbal tea (2g) was extracted with distilled boiled water (1:50, w/w), standing for 5 min. Then, infusions of green and herbal tea (about 40°C) were evaluated by a trained team of nine panelists (three men and six women, 23–38 years old). Each member of the team graded the bitterness, astringency, and sweet aftertaste of the tea using a modified nine-point scale, where 8–10 "extremely strong," 6–8 "strong," 4–6 "neutral," 2–4 "weak," and 0–2 "very weak." For the sensory evaluation, a 50 mL aliquot of the sample solution was served in a cup and the intensity of bitterness was recorded by tasting the sample solution while swirling in the mouth for 7–8 s. Then, the solution was spat out and the intensity of astringency was recorded within 3–4 s. After recording the astringency, the examiners were instructed to rinse their mouths with pure water and record the intensity of the sweet aftertaste within 4–5 s. There was a 5 min interval between tests. The results were analyzed statistically to determine significant differences between the mean scores of the different samples. Each evaluation was repeated three times with the order of the samples randomized in each test.

Analysis of Flavonoid Profiles and Flavonol Metabolites by Ultra-High-Performance Liquid Chromatography–Tandem Mass Spectrometry (UPLC/MS/MS)

The freeze-dried samples were crushed with a mixer mill for 240 s at 45 Hz. Then, 10 mg of each powdered sample were transferred

into a 5 mL Eppendorf tube and extracted with 3 mL of 75% methanol/1% acetic acid. After vortexing for 30 s, the samples were homogenized at 40 Hz for 4 min and sonicated for 10 min in an ice-water bath. The homogenization and sonication steps were repeated three times. Following centrifugation at 13, 800 × *g* for 15 min at 4°C, a 2.5 mL aliquot of the supernatant was dried under a gentle flow of nitrogen gas. The residue of each sample was reconstituted in 1.0 mL of 50% methanol/0.1% formic acid, then vortexed for 30 s, ultra-sonicated for 15 min in an ice bath, and centrifuged at 13, 800 × *g* for 15 min at 4°C. The resulting supernatants were filtered through a membrane with a pore size of 0.22 μm, transferred into 2 mL glass vials, and stored at −80°C until analysis by UPLC-MS/MS. A quality control sample was prepared by mixing equal volumes of the supernatants of all samples.

Herbal teas made from flowers and leaves were analyzed using an Acquity UPLC system (Waters Corporation, Milford, MA, USA) equipped with a BEH C18 column (1.7 μm, 2.1 × 150 mm). The samples were separated with a gradient composed of acetonitrile (A) and 0.1% formic acid water (B)

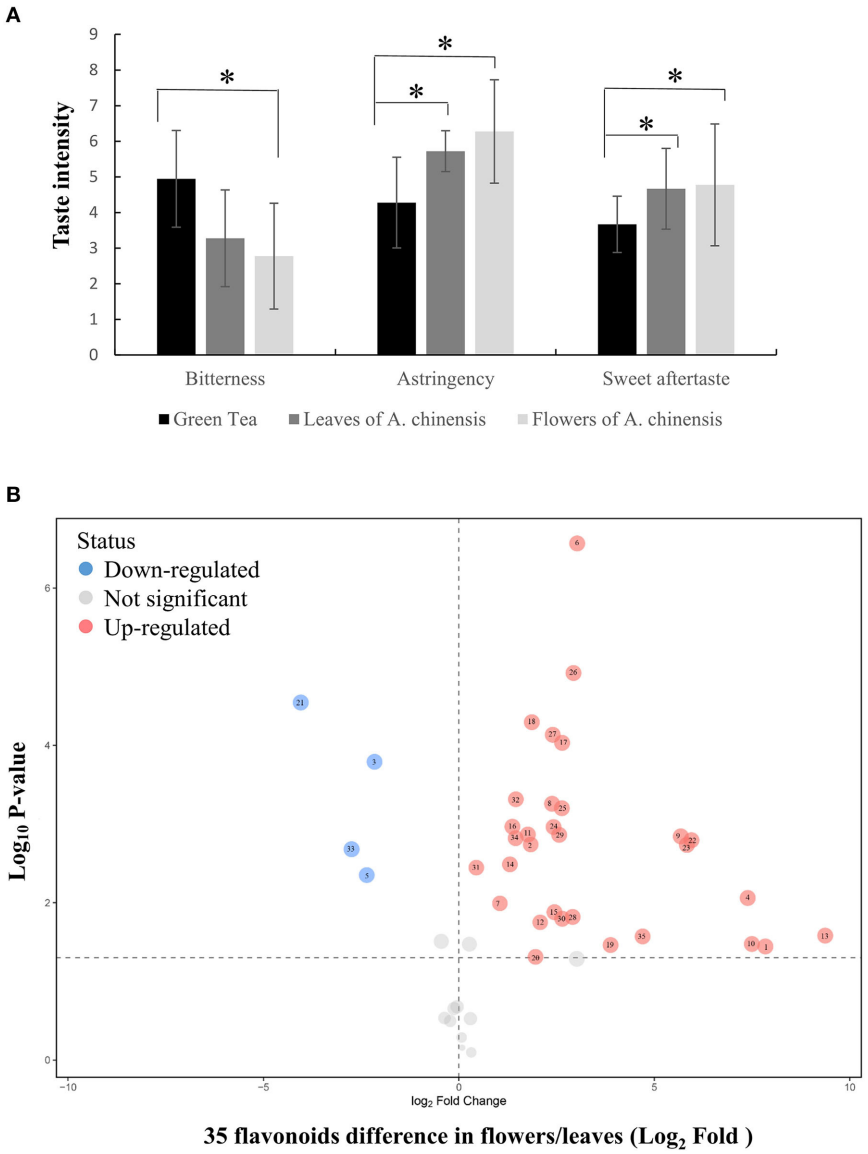


FIGURE 1 | Sensory evaluation and flavonoid profile of *A. chinensis* herbal tea. **(A)** The taste intensity of green tea, herbal tea madding from flowers and leaves of *A. chinensis*. **(B)** The flavonoid differential analysis between flowers and leaves. 1, (-)-Epigallocatechin; 2, Apigenin; 3, Astilbin; 4, Astragalin; 5, Avicularin; 6, Chalconaringenin; 7, Chrysin; 8, Cianidanol; 9, Cynaroside; 10, Dihydromyricetin; 11, Engeletin; 12, Galangin; 13, Gallocatechin; 14, Genkwanin; 15, Hyperoside; 16, isoliquiritigenin; 17, Isoquercitrin; 18, Isorhamnetin; 19, Isorhamnetin-3-O-nehesperidine; 20, Isosakuranetin; 21, Kaempferitrin; 22, Kaempferol; 23, Kaempferol-3-O-rutinoside; 24, Myricitrin; 25, Narcissoside; 26, Naringenin; 27, Narirutin; 28, Phloretin; 29, Pinocembrin; 30, Procyanidin B1; 31, Procyanidin B2; 32, Quercetin; 33, Quercitrin; 34, Rutin; 35, Tiliroside. Data were statistically evaluated using Student's *t* test (*$P < 0.05$).

under the following parameters: injection volume, 2 μL; column temperature, 40°C, and flow rate, 0.3 mL·min^{-1}. The first equilibrium at 90% B was maintained for 0.5 min. Then, the analytes were eluted using a linear gradient of 90%→ 40% B for 0.5–15 min and washed for 3 min with 2% B. The final equilibrium at 90% B was maintained for 2 min. The total running time was about 20 min with the use of a QTRAP® 6500+ LC-MS/MS System (AB SCIEX, Concord, ON, Canada). Typical ion source parameters were: positive (+) and negative (–) mode; ion spray voltage, +5000/−4500 V; curtain gas, 35 psi, temperature, 500°C; ion source gas 1, 55 psi; and ion source gas 2, 60 psi.

The *A. chinensis* leaves, flowers, and seeds stored in −80°C were extracted as described above to quantify the target flavonols. Astragalin, isoquercitroside, quercitrin, quercetin, and kaempferol were detected using an Agilent 1290 Infinity UPLC system coupled an Agilent 6470 Triple Quadruple LC/MS system (Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with an Agilent ZORBAX RRHD Eclipse Plus C18 column. The samples were separated with a gradient composed of acetonitrile (A) and 0.1% formic acid water (B), using eluted program, 0–7 min, 95%→ 5% B; 2 min for wash with 100% A, and 1 min for equilibrium with 95% B. Under the following parameters: injection volume, 2 μL; wavelength, 254 nm; column temperature, 35°C; and flow rate, 0.3 mL·min^{-1}. The mass spectrometer was operated in negative ion mode with the following parameters: sheath gas temperature, 300°C; gas flow, 5.0 L min^{-1}; nebulizer gas, 45 psi; capillary voltage, 3500 V; nozzle voltage, 500 V, and delta electron multiplier voltage, 200 V. The ion conditions for detection by UPLC-MS were the same as described above and metabolites were detected in multiple reactions monitoring mode. Data analyses were conducted with MassHunter qualitative analysis software (version B.07.00; Agilent Technologies).

Crude Protein Extraction and Enzymatic Analysis of *A. chinensis* Flowers

Freshly harvested flowers were ground twice using a freeze grinder (JX-2016, 50 Hz) for 30 s and total protein was extracted

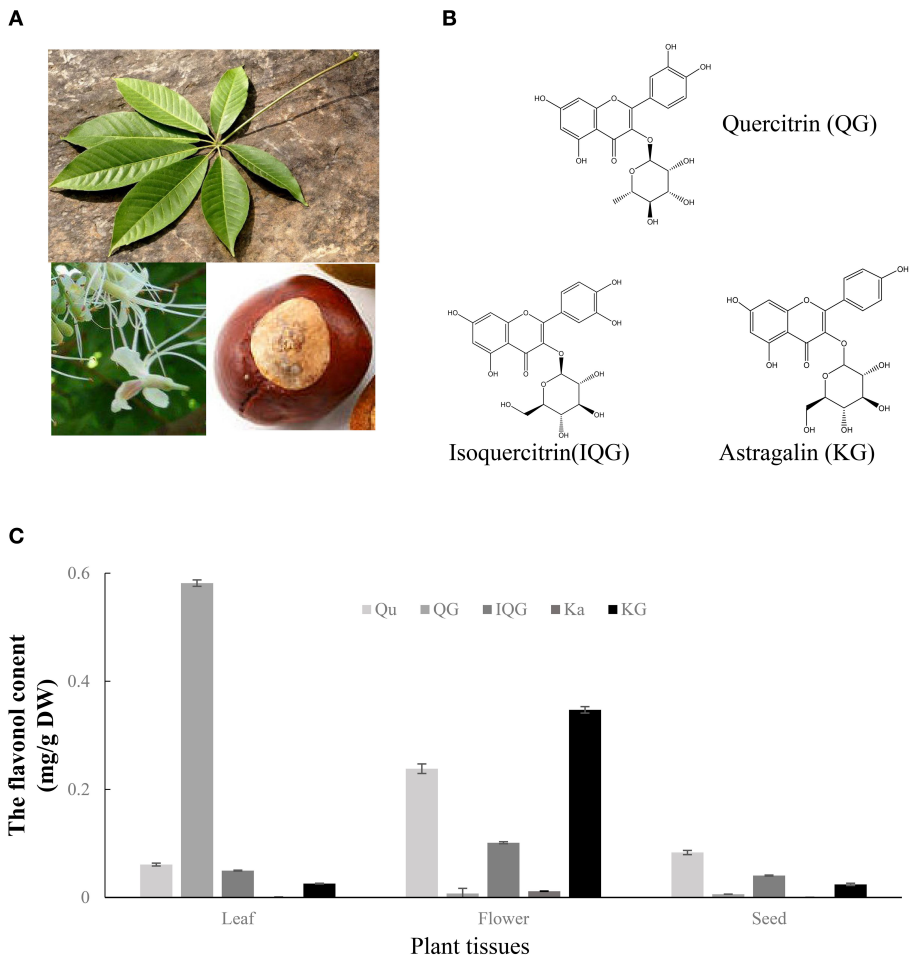


FIGURE 2 | Flavonol glycosides in different tissues of *A. chinensis*. **(A)** Images of the seeds, flowers, and leaves of *A. chinensis*. **(B)** The molecular structures of three main flavonol glycosides. **(C)** The contents of flavonols in the three tissues. Qu, quercetin; Ka, kaempferol; IQG, isoquercitrin; KG, astragalin; QG, quercitrin.

(0.1 g ml^{-1}) in ice-cold extraction buffer (100 mM Tris at pH 7.0, 10 mM dithiothreitol, 0.1 mM phenylmethylsulphonylfluoride, 0.5% polyvinyl pyrrolidone). The extract was centrifuged at 13, 800 × *g* for 10 min at 4°C, and the supernatant was concentrated with an ultrafiltration device (30 KDa) and substitution was conducted with an enzymatic reaction buffer at 4°C. The crude extract was then assayed for enzyme activity in 100 μL reaction volumes consisting of 10 mM dithiothreitol, 50 mM Tris-HCl (pH 7.0), 2 mM UDP-glucose, and 100 μM sugar acceptor. After 30 min of incubation at 37°C, the reactions were terminated with ethyl acetate (400 μL), then, the organic phase was collected after centrifuging at 13, 800 × g for 5 min, passed through a 0.22 μm filter membrane, evaporated to dryness, and re-dissolved with methanol (100 μL). A 2 μL aliquot was analyzed using a UPLC-MS/MS system as described above.

Differential Transcriptome and Phylogenetic Analysis

The expression patterns of AcUGT genes associated with the flavonol glycoside biosynthetic pathway were examined, ∼33.3 Gb of clean data were downloaded from the National Center for Biotechnology Information (NCBI) database (SRR8073719–SRR8073722). RNA-Seq reads from each tissue were assembled using Trinity software with no reference genome. The expression level of each gene was evaluated with the Fragments Per Kilobase per Million mapped fragments method and calculated using Tophat and Cufflinks (version 2.1.1) with default parameters.

Multiple sequence alignments of the deduced amino acid sequences were conducted using DNASTAR bioinformatics software (https://www.dnastar.com/). The predicted amino acid sequences of the UGTs were aligned using the Clustal X2 multiple sequence alignment algorithm (http://www.clustal.org/) before phylogenetic analysis. A neighbor-joining phylogenetic tree was constructed with 1000 boot-strap replicates using Molecular Evolutionary Genetics Analysis 4.0 software (https://www.megasoftware.net/).

Enzymatic Assay of AcUGT Recombinant Proteins and Product Identification

Primers for the 30 AcUGT genes were designed based on the AcGTs database of transcriptomes assembled without a reference genome. Mixed cDNAs from the leaves, flowers, and seeds of *A. chinensis* were used for gene amplification. The PCR products were purified and digested using the corresponding restriction enzymes, then ligated to the pMAL-c2x vector (New England BioLabs, Ipswich, MA, USA) and digested with the same restriction enzymes for expression of recombinant proteins in *Escherichia coli*.

Recombinant UGT proteins in *E. coli* were obtained as Yin et al. (2021). The crude recombinant proteins were used for enzymatic activity testing. The reaction system and conditions were same as for the crude proteins of *A. chinensis*, while the concentration of the tested acceptor substrates (0–400 μM) and purified enzymes (5 μg) was used for kinetic analysis. Further UPLC- MS/MS analysis was conducted as above mentioned. The kinetic parameters *K*m and *K*cat were calculated with the Hyper 32 program (http://hyper32.software.informer.com/).

Homology Modeling and Docking Statistic

Homology models of the AcUGT1, AcUGT22 and AcUGT26 were constructed using the three-dimensional structure of UGT74AC2 (PDB ID: 7BV3; https://www.rcsb.org/, *Siraitia grosvenorii*), UGT85H2 (2PQ6, *Medicago truncatula*) and PaGT3 (6LZX, *Phytolacca Americana*) as a template respectively, by the SWISS-MODEL server (http://swissmodel.expasy.org) (Li et al., 2007, 2021; Maharjan et al., 2020). UDP-glucose or UDP-rhamnose and quercetin were docked with the model structure of AcUGTs using the Igemdock 2.1 docking program (http://gemdock.life.nctu.edu.tw/dock/igemdock.php). The docking result with UDP-glucose or UDP-rhamnose and quercetin was visualized with the Pymol molecular graphics system (http://www.pymol.org).

Statistical Analysis

Statistical analyses were performed using Excel software (Microsoft Corporation, Redmond, WA, USA). The unpaired, two-legged Student's *t*-test was used to identify significant differences. A probability (p) value of <0.05 was considered statistically significant. Data are presented as the mean ± standard deviation ($n \geq 3$).

RESULTS

The Taste of *A. chinensis* Herbal Tea and Metabolic Profiles of Flavonoids

To explore the possibility of utilizing the whole *A. chinensis* plant, the leaves and flowers were naturally dried to produce herbal tea. Sensory evaluation of green tea and *A. chinensis* herbal tea infusions indicated that *A. chinensis* flowers produced an infusion that was obviously less bitter with a superior sweet aftertaste and stronger astringency (>6 points) than the green tea. Notably, the bitterness of the flower infusion was about half of that of green tea (**Figure 1A**). The astringency of the leaf infusion was also significantly stronger than the green tea, although all scores were between 4 and 6 (neutral level). The astringency results suggested that flavonoids may be abundant in the leaves and flowers of *A. chinensis*, which present potential materials for the production of foods to maintain health.

The flavonoid profiles of herbal tea made from *A. chinensis* leaves and flowers were investigated to identify flavonoids that determine the degree of astringency. As shown in **Figure 1**, **Supplementary Figure 1**, there were significant changes in the expression patterns of 35 of 47 flavonoids in *A. chinensis* flowers vs. leaves (31 flavonoids in flowers were higher than those in leaves, only four flavonoids were lower). In flowers, seven of the top 10 peak areas were flavonols other than (-)-epigallocatechin, procyanidin B2, and cynaroside (luteolin-7-glucoside). The peak area of astragalin was the highest at over 160-fold as compared to that in leaves. Similarly, eight of the top 10 peak areas of leaves were flavonols other than procyanidin B2 and cynaroside, with quercitrin (quercetin 3-O-rhamnoside) having the highest

value. Gallocatechin (-)-epigallocatechin were also accumulated in flowers, especially the relative amount of (-)-epigallocatechin was over 229-fold greater than that of leaves, however, we did not detected epigallocatechin-3-gallate and epicatechin-3-gallate in flowers or leaves, which are the key compounds contributed to overall astringent taste of tea infusions (Scharbert and Hofmann, 2005). Collectively, these results show that flavonol glycosides are abundant in *A. chinensis* leaves and flowers, especially flavonol mono-glucosides, which are key to the astringent taste compounds of tea (Scharbert et al., 2004).

To investigate the formation process of flavonol mono-glucosides *in planta*, the flavonols (kaempferol and quercetin) and three flavonol glucosides (i.e., astragalin, isoquercitrin, and quercitrin) were quantified in fresh seeds, flowers, and leaves of *A. chinensis* (**Figures 2A,B**). As shown in **Figure 2C**, Similar to results from herbal tea, tested flavonol or flavonol glycosides in the flowers higher than that in leaves, unless quercitrin which mainly accumulate in leaves reaching 0.59 mg/g. The astragalin content reached to 0.352 mg/g in flowers, which was 14- and 16-fold greater than that of the leaves and seeds, 0.025 and 0.022 mg/g respectively. The isoquercitrin content of the flowers was 0.103 mg/g, which was 2.1- and 2.7-fold greater than that of the leaves and seeds, respectively. The quercetin content of the flowers was 0.229 mg/g, which was 3.8- and 3.1-fold greater than that of the leaves and seeds, respectively. Interestingly, the content of quercetin in flowers was higher than isoquercitrin and quercitrin, and more than half of the content of astragalin. Flavonol glucosides which may play important roles in the enhanced taste, lower bitterness, and higher astringency of *A. chinensis* herbal tea were highly accumulated in the flowers, suggesting that some enzymes contributed to the formation of flavonol glucosides.

Crude Proteins in Fresh Flowers Catalyzed the Formation of Flavonol Glucosides

Enzymes contained in the flowers involved in the glycosylation of flavonol in *A. chinensis* were investigated. Although the

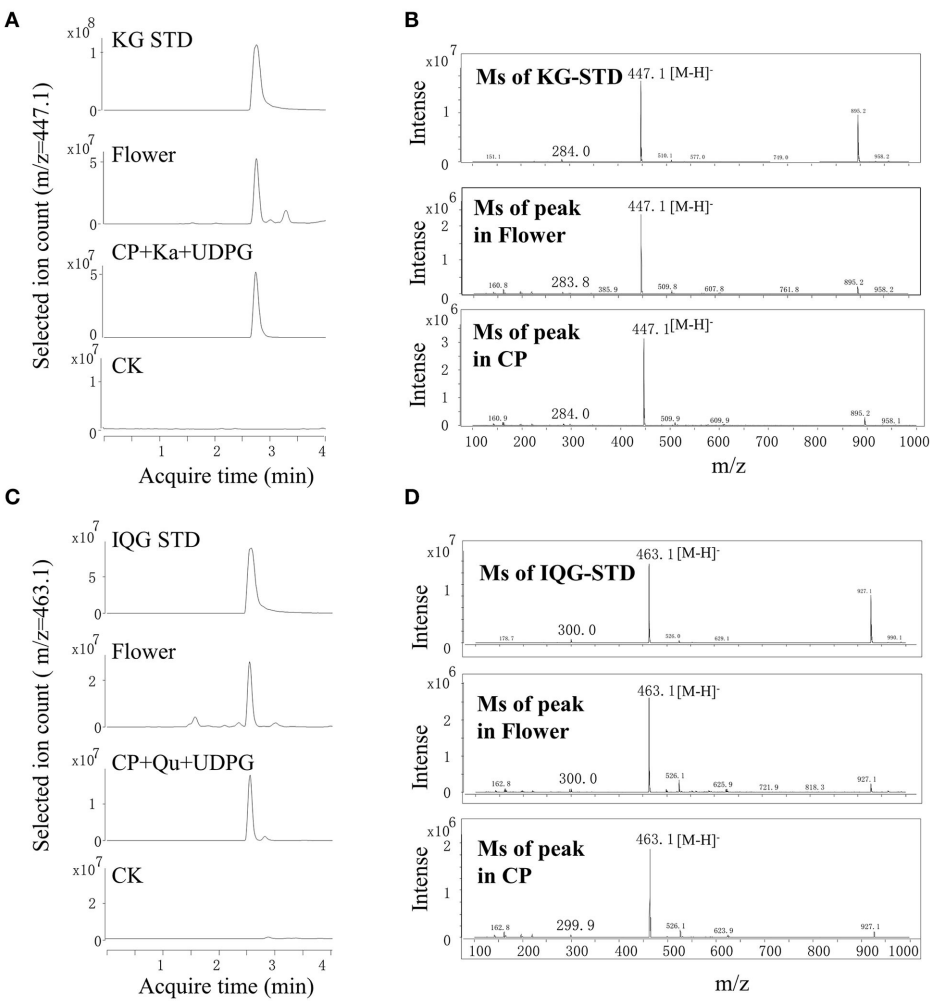


FIGURE 3 | Crude proteins from flowers catalyzed the formation of astragalin and isoquercitrin *in vitro*. **(A,B)** UPLC-MS chromatograms of extracts from flowers and crude proteins (CP) with kaempferol (Ka). **(C,D)** UPLC-MS spectra of extracts from flowers and CP with isoquercitrin (IQG).

leaves contained quercetin 3-O-rhamnose, the lack of UDP-rhamnose (a sugar donor) limited the further investigation. Crude proteins extracted from fresh flowers were used, UDP-glucose acted as a sugar donor and quercetin or kaempferol as a sugar acceptor (**Figure 3**). As compared with the standard of kaempferol 3-O-glucoside, the enzymatic product obtained with UDPG and kaempferol contained a common fragment ([M-H]$^-$, 447.1). The MS spectrum and acquisition time of the new product were identical to those of the *A. chinensis* flower (**Figures 3A,B**). With quercetin as the sugar acceptor and UDPG as the sugar donor, the result was similar, as the crude proteins isolated from the flowers could catalyze the formation of quercetin 3-O-glucoside ([M-H]$^-$, 463.1, **Figures 3C,D**). These enzymatic activities were compatible with the accumulation of flavonol mono-glucosides in flowers. The crude proteins of the *A. chinensis* flower involved in flavonol glycosylation suggested that some UGT genes should highly expressed in these tissues.

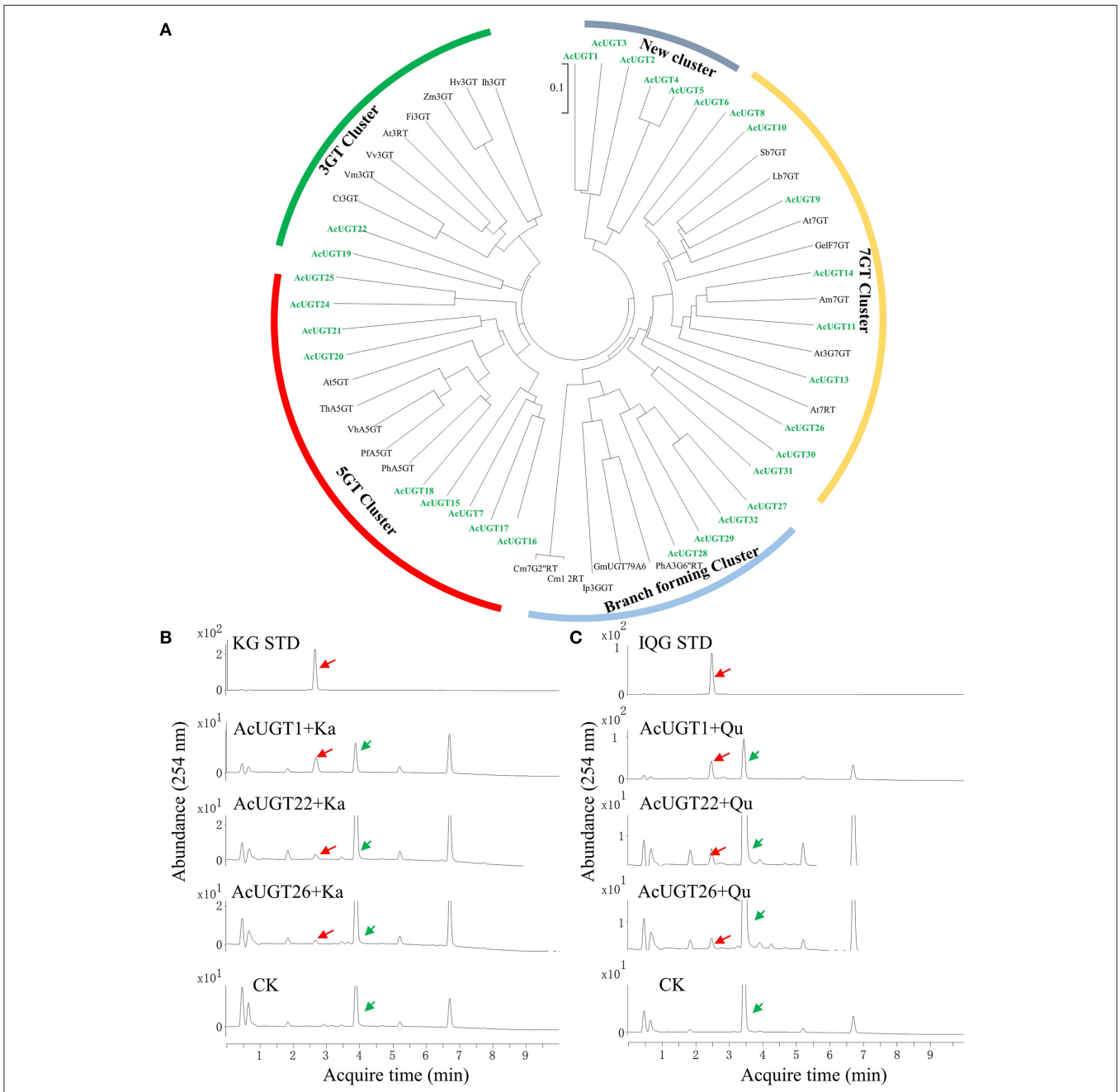


FIGURE 4 | A phylogenetic tree and enzymatic analysis to screen for flavonol glucosyltransferases. **(A)** A phylogenetic tree of 30 AcUGTs and reported UGTs, sequence of reported UGTs were from Yin et al. (2021). **(B)** The UPLC spectra of AcUGT recombinant proteins with UDPG and kaempferol (Ka). **(C)** The UPLC spectra of AcUGT recombinant proteins with UDPG and quercetin (Qu). The green arrows indicate substrates, while the red arrows indicate products. KG STD, astragalin standard; IQG STD, isoquercitrin standard. CK means control, the enzymatic activity of the boiled protein encoded by empty-vector, substrates and UDP-glucose.

A

UGT	Substrates	V_{max} (nmol.S^{-1})	K_m (μM)	k_{cat} (S^{-1})	k_{cat}/K_m (S^{-1}.mM^{-1})
AcUGT1	Ka	59.67±2.17(10^{-3})	22.62±6.29	1.15±0.03	50.63±1.32
	Qu	0.287±0.001(10^{-3})	1.02±0.23	5.34±0.01(10^{-3})	5.24±0.01
AcUGT22	Ka	0.4±0.03(10^{-3})	4.17±0.46	7.66±0.42(10^{-3})	1.84±0.10
AcUGT26	Qu	0.034±0.002(10^{-3})	2.23±0.06	0.65±0.02(10^{-3})	0.29±0.01

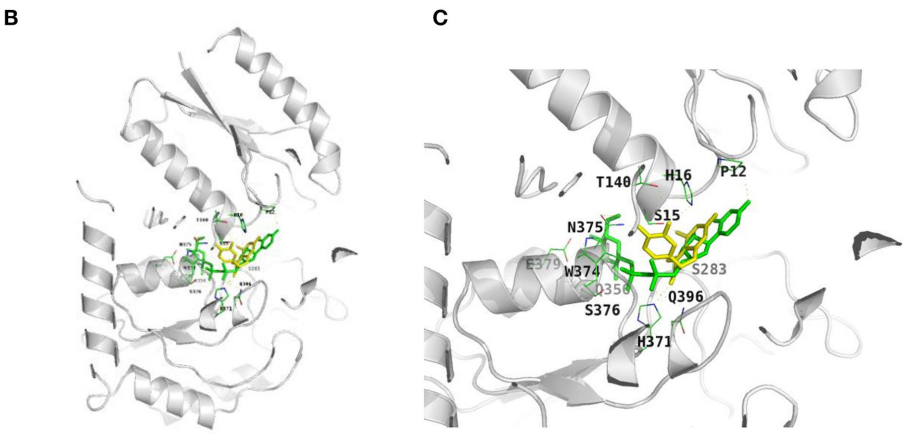


FIGURE 5 | Characterization of AcUGTs. **(A)** Kinetic characteristics of three AcUGTs. **(B)** A ribbon diagram of UDPG and quercetin with AcUGT1. **(C)** An enlarged ribbon diagram of the active docking domain. UDPG is indicated in green and quercetin in yellow.

Differential Transcriptome, Phylogenetic Analysis, and Enzymatic Testing to Screen UGTs

Since the genome of *A. chinensis* has not yet been sequenced, six transcriptome sequences were retrieved from the NCBI database to assess candidate UGT genes. Based on the transcriptome data, 170 genes annotated as GTs were used to construct an AcGT database for expression profile analyses (**Supplementary Figure 2**; **Supplementary Data 1**). Hierarchical clustering analysis showed that 105 and 65 GT genes were highly expressed in the flowers and seeds of *A. chinensis*, respectively. Some GTs were predicted to have other functions, such as the formation of pectin in the cell wall. Unigene37810, Unigene38108, Unigene15907, and Unigene16155 were predicted to code for galacturonosyltransferases, which contribute to pectin formation in the cell wall. Unigene13904, which was only expressed in the flowers, was annotated as O-fucosyltransferase 28. Unigene29217 and Unigene17314 were predicted to code for cellulose synthase and N-acetylglucosaminyltransferase, respectively. These GTs greatly interfered with the selection of candidate genes.

To precisely screen the AcUGTs involved in the formation of flavonoids, the PSPG box was used to manually annotate the AcGT database. Finally, 30 candidate AcUGTs were obtained for further study. Twenty-nine candidate AcUGTs consisted of 44 amino acids with glutamine as the last amino acid of the PSPG box, only the last amino acid of AcUGT29' PSPG box is aspartic acid (**Supplementary Figure 3A**). Glutamine is the key determinant of whether UGTs use UDP-glucose as a sugar donor (Yang et al., 2018). Phylogenetic analysis showed that 55 UGTs, including candidate AcUGTs and 25 previously identified UGTs, clustered into five major branches, which may provide important clues about the enzymatic functions of candidate UGTs (**Figure 4A**). The first cluster, represented by ZmF3GT, consisted of 10 UGTs, including AcUGT19 and AcUGT22, that catalyze the glycosylation of the 3-OH sites of flavonoids. The second cluster contained five UGTs that glycosylate the 5-OH sites of flavonoids, such as PfA5GT, and including nine AcUGTs. This finding suggests that AcUGTs in cluster two may catalyze the glycosylation of the 5-OH sites of flavonoids. In the third branch, seven AcUGTs were clustered with seven UGTs that glycosylate the 7-OH sites of flavonoids, while AcUGT26, AcUGT30, and AcUGT31 are located in the side branch in the cluster. Another

branch formed a cluster of nine UGTs that catalyze multistep glycosylation of flavonoids, such as Ip3GGT. AcUGT1-6 formed a new branch, suggesting that this protein may be involved in different functions. To explore the glycosylation of the 3-OH sites of flavonols, the members of the 3GT cluster and some AcUGTs wth unique positions in phylogenetic tree, such as AcUGT26 or AcUGT1-6, were selected for further enzymatic experimentation.

To investigate the glycosylation of flavonoids, especially kaempferol and quercetin, the 30 AcUGTs were tried to clone, finally three AcUGTs were cloned and further identified using a prokaryotic system (**Supplementary Figure 3B**; **Supplementary Data 1**). UDP-glucose was used as a glycosyl donor for the enzymatic activity test with kaempferol or quercetin as the substrate. AcUGT1, AcUGT22, and AcUGT26 could catalyze the hydroxyl glucosylation of quercetin or kaempferol with the use of UDP-glucose as the glycosyl donor (**Figures 4B,C**). Analysis of the enzymatic products by UPLC-MS showed that recombinant AcUGT1, AcUGT22, and AcUGT26 produced new products with kaempferol and quercetin (**Supplementary Figures 3C,D**; **Supplementary Figure 4**). The m/z value of the glycosylation products was increased by 162 (the $[M\text{-}H]^-$ m/z value of kaempferol and quercetin is 447.1 and 463.1 respectively). Based on the standard products of quercetin 3-*O*-glucoside and kaemperol 3-*O*-glucoside, the glycosylation product of both AcUGTs was flavonol 3-*O*-glucoside.

Characterization of AcUGTs

Enzyme kinetics analysis revealed that the catalytic efficiency of AcUGT1 (k_{cat}/K_m value of kaempferol and quercetin was 50.63 and 5.24 $S^{-1} \cdot mM^{-1}$, respectively) was much higher than that of AcUGT22 with kaempferol or AcUGT26 with quercetin. Moreover, AcUGT1 was more efficient with kaempferol than quercetin over 9.6-fold (**Figure 5A**), contrastly, the affinity of AcUGT1 to kaempferol was just about 4.5% of that to quercetin. AcUGT22 possessed a higher affinity toward kaempferol than AcUGT1 (K_m value of AcUGT22 and AcUGT1, 4.17 vs. 22.62 μM, respectively), while the affinity of AcUGT26 toward quercetin is lower than AcUGT1 (K_m value of AcUGT26 and AcUGT1, 1.02 vs. 2.23 μM, respectively). Although the three AcUGTs could act as a common substrate of the 3-OH sites of flavonols, differences in the affinity or efficiency toward different flavonols were mainly due to structural differences.

As shown in **Figures 5B,C**, the substrate quercetin (Qu) and the glycosyl donor UDP-glucose (UDPG) are both in the pocket created for binding with substrates in the model of AcUGT1. Several residues of AcUGT1, including Pro12, Ser15, His16, Thr140, Ser283, His371, and Asn375, form hydrogen bonds with quercetin, while Pro12, Ser15, Ser283, Gln356, His371, Gly373, Trp374, Asn375, Ser376, Glu379, and Gln396 form hydrogen bonds with UDPG. These residues could be important for the catalysis activity of AcUGT1. Based on the total energy for docking of kaempferol by the three AcUGTs (**Supplementary Table 3**), AcUGT1 required the lowest amount of energy and, thus, might be the most efficient, which is in agreement with the enzymatic activity results (k_{cat}/K_m value of AcUGT1 for kaempferol was the highest at 50.63 $S^{-1} \cdot mM^{-1}$).

To identify the putative glycosyltransferases involved in the formation of quercetin 3-O-rhamnose, the docking result of AcUGT26 showed that quercetin and UDP-rhamnose (UDPR) were located in the cleft of the AcUGT26 model, suggesting that AcUGT26 may contribute to the rhamnosylation of quercetin (**Supplementary Figure 5C**). As compared with AcUGT26 docking with UDPG or quercetin, fewer residues of the PSPG box of AcUGT26 formed hydrogen bonds with UDPR or quercetin, only Asn361, Glu381, and Gln382. However, these active sites in docking AcUGT26-UDPR-quercetin are similar to AcUGT1-UDPG-quercetin (Asn375, Glu379, and Gln396) and AcUGT26-UDPG-quercetin (Asn361, Glu381, and Gln382), indicated that AcUGT1 and AcUGT26 possess the same model for interacting substrates and sugar donor in spatial structure.

DISCUSSION

Metabolic Profiles of *A. chinensis* Flavonoids Revealed Special Spatial Distributions

There are various metabolic patterns of flavonoids in plants. For example, the leaves of *Ginkgo biloba* are rich in flavonol polyglycosides and acylated flavonoids (Su et al., 2017). In tea leaves, flavonols mainly exist in the form of O-glycosides with a glycoside moiety at the C-3 position of aglycones (e.g., quercetin, kaempferol and myricetin), which account for ∼13% of total tea polyphenols in fresh tea leaves following the most abundant flavonoid subclass flavanols (primarily catechins) (Xu et al., 2018b; Zhuang et al., 2020). Soybean seeds contain an abundance of isoflavones, while flavonol di- or multi-glycosides are the main form in leaves (Yin et al., 2017a). The leaves of *Andrographis paniculata* contain andrographolides, while the roots are enriched with flavonoids (Sun et al., 2019). Ginseng leaves contain high concentrations of flavonoids, including kaempferol 3-O-glucoside and panasenoside, which are kaempferol di-glycosides, while the roots contain trace amounts of flavonoids (Yin et al., 2021). The seeds of *A. chinensis* contain flavonol monoglycosides, polyglycosides, and acyl flavonoids (Zhu et al., 2006). Quercitrin was reported to mainly accumulate in leaves of *A. hippocastanum* and *A. carea* (Oszmiański et al., 2014), similarly, the flavonoid profile of *A. chinensis* also reveals quercitrin abundant in its leaves. Furthermore, the flowers of *A. chinensis* was firstly reported to contain high concentrations of flavonol mono-glucosides which contribute to the low bitterness, strong astringency of infusion (Scharbert and Hofmann, 2005).

Flavonols Exploration Herbal Tea Promotes the Whole Plant Utilization of *A. chinensis*

The seeds of *A. chinensis* after alkali treatment or high-temperature detoxification treatment were intake as strarch food, and *Aesculus hippocastanum* extract could used as stabilizer in Hemp seed oil Nanoemulsions for biomedical and food applications (Jarzebski et al., 2021). The seeds of *A. chinensis* were used as the source of escin with a clinically significant activity in chronic venous insufficiency, hemorrhoids and post-operative edema (Zhang et al., 2006). However, the leaves and

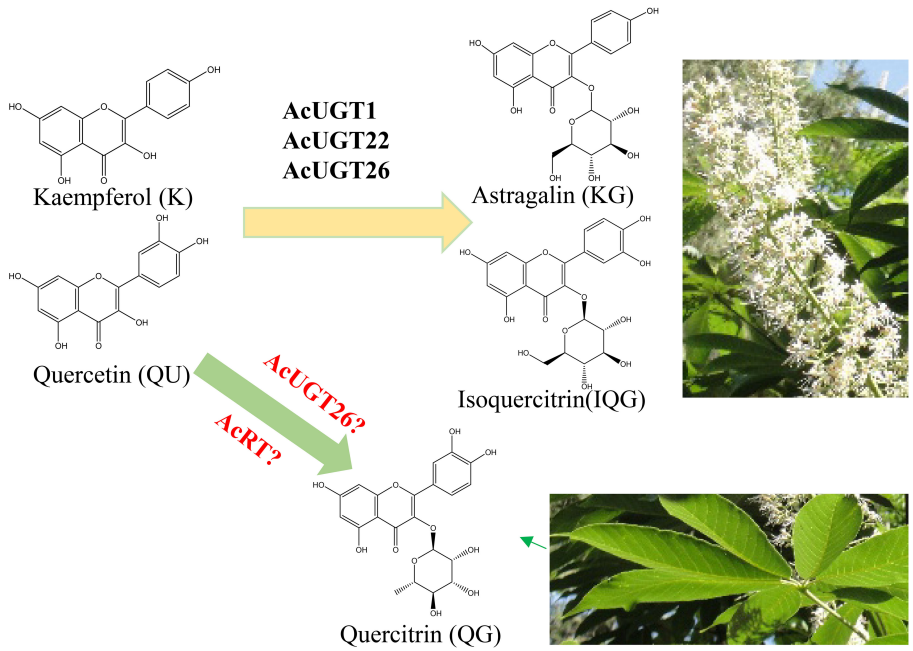


FIGURE 6 | The proposed formation process of astragalin and isoquercitrin in *A. chinensis*.

flowers of *A. chinensis* almost be a virgin land for metabolite investigation.

Many plants with high flavonol contents, such as ginkgo leaves, tartary buckwheat, ginseng, soybean, and lotus, are cultivated for the production of healthy foods or herbal teas. Herein, we found 47 flavnoid accumulated in leaves and flowers, especially flavonol glycosides are the main components. Flavonol glycosides are important astringent and bitter substances in teas, due to their extremely low taste thresholds (Sasaki et al., 2014; Guo et al., 2021). Sensory evaluation indicated that the leaves and flowers of *A. chinensis* match the key taste quality of herbal tea.

Flavonol 3-O-glycosides were found to induce a velvety and mouth-coating sensation at very low threshold concentrations, which were far below those of catechins or theaflavins (Scharbert et al., 2004). Consistently, flavonoid profile of *A. chinensis* herbal tea suggested that quercetin 3-O-rhamnose and kaempferol 3-O-glucoside contributed to the taste parameters of the leaves and flowers, respectively. Similar to flavonol 3-O-glycosides in the seeds of *A. turbinate* (Kimura et al., 2017), the flowers and leaves of *A. chinensis* also could be used as food additives, companing with huge inflorescence and lots of leaves as deciduous tree (**Figure 6**).

Multi-Dimension Strategy to Reveal the Glucosyltransferases Involved in the Formation of Flavonol Glucosides

Multi-dimension strategy had been developed to reveal the formation of key compounds in food or *planta*, such as oolong tea, *Ziziphora bungeana* and ginseng (Yonekura-Sakakibara et al., 2008; Liu et al., 2018; He et al., 2020; Wu et al., 2020; Yin et al., 2021). The glycosyltransferase that catalyzes flavonol mono-glucosides was inferred by analysis of metabolites and crude proteins activities. The transcriptome data and *in vitro* enzyme activity verification of the recombinant protein showed that AcUGT1 may play an important role in the accumulation of these glycosides in flowers. However, the expression levels of AcUGT22 and AcUGT26 in seeds were about 5-fold greater than that in flowers. Hence, flavonoids may be involved in the catalysis of monoglycosides to polyglycosides in seeds, or their products may be further modified and not reflected in flavonoid profiles, as flavonols reportedly undergo acylation in seeds (Sasaki et al., 2014).

Glycosyltransferase, which catalyzes the formation of glycosides, dictates substrate diversity and specificity (Li et al., 2021). Amino acid alignments and phylogenetic tree analysis showed that UGTs tend to be more regioselective, combined with the enzymatic function *in vitro*. AcUGT22, which was classified to the 3-OGT cluster, catalyzes the formation of flavonol mono-glucosides. Interestingly, AcUGT1 and several other UGTs formed a new family. In addition to 3-glycosylation activities, AcUGT1 may have other catalytic activities, such as a triterpenoid substrate, and thus is a potential candidate to explore the mechanism of escin formation. Similarly, the characteristics of AcUGT26 enzyme activity *in vitro* were not consistent with the evolutionary tree. Although close to the 7-OGT cluster, AcUGT26 and other two AcUGTs formed a unique branch, suggesting that this enzyme may have different catalytic activities.

The elucidation of several plant UGT crystals is the basis for exploring the reaction mechanism of these proteins and secondary metabolites in plants (Li et al., 2007; Yang et al., 2018; Maharjan et al., 2020). The simulation results of the three AcUGTs were all >35% and two were 50%, indicating very high feasibility. Molecular docking showed that sugar donor (UDP-glucose) and sugar acceptor (quercetin) were docked to the pocket of AcUGT1 and AcUGT26, and their location were very close, while the energy of the three mimetic proteins to dock quercetin was the lowest. In the simulation, the hydrogen bonds formed between UDP-glucose and quercetin were mainly located in the N terminal, the PSPG box was close to the C terminal, and several amino acids were in the middle loop. Unlike other reports, the PSPG box was responsible for the interactions between the C terminal and UDPG (Yang et al., 2018). Similarly, Pro12 and Ser15 of AcUGT1 and Ser18 and His21 of AcUGT26 can form hydrogen bonds with UDP-glucose. The docking prediction also found that residues of the PSPG box participated in the interactions of quercetin. Especially, the residues Trp360, Asn361, Glu381, and Gln382 of the PSPG box of AcUGT26 can form hydrogen bonds with the substrate. Due to the lack of UDP-rhamnose standards, molecular docking was used to identify putative UGTs. The docking results of UDP-rhamnose, quercetin, and three AcUGTs showed that the rhamnosyl donor and acceptor were docked to the pocket of AcUGT26, suggesting that the protein might catalyze the rhamnosylation of quercetin (**Figure 6**).

CONCLUSION

Sensory evaluation indicated that flavonol glycosides, which are abundant in the flowers and leaves of *A. chinensis* contribute to the taste of herbal tea. The crude proteins of the flowers and recombinant proteins were screened by differential transcriptome and phylogenetic methods. Enzymatic testing demonstrated that three glycosyltransferases that catalyze the formation of flavonol glucosides were mainly present in the flowers. These findings unraveling the glucosylation of astringency compounds of *A. chinensis* via integrative sensory evaluation, metabolite profiling and enzymatic analysis, could push the utilization of the whole plant in the development of healthy functional products or additives.

DATA AVAILABILITY STATEMENT

AcUGT1, AcUGT22, and AcUGT26 were named on behalf of the UGT nomenclature committee as UGT71A46, UGT85A121, and UGT90A25 respectively. 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

QY, XH, and WS conceived and designed the study. QY, YW, and JC performed the experiments and analyzed the data. QY, JC, and XH wrote the paper. HG analyzed the UGTs information from transcriptome. All authors read and approved the manuscript.

FUNDING

This research was supported by Scientific and technological innovation project of China Academy of Chinese Medical Sciences (CACMS Innovation Fund, CI2021A04117, CI2021A05108), Beijing Natural Science Foundation of China (7192138), the National Key R&D Program of China (2019YFC1711100), and the Fundamental Research Funds for the Central Public Welfare Research Institute of China (ZZ13-YQ-097).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.830343/full#supplementary-material

REFERENCES

Cao, Q. Q., Zou, C., Zhang, Y. H., Du, Q. Z., Yin, J. F., Shi, J., et al. (2019). Improving the taste of autumn green tea with tannase. *Food Chem.* 277, 432–437. doi: 10.1016/j.foodchem.2018.10.146

Chen, A. Y., and Chen, Y. C. (2013). A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem.* 138, 2099–2107. doi: 10.1016/j.foodchem.2012.11.139

Cheng, J. T., Chen, S. T., Guo, C., Jiao, M. J., Cui, W. J., Wang, S. H., et al. (2018). Triterpenoid saponins from the seeds of *Aesculus chinensis* and their cytotoxicities. *Nat. Prod. Bioprospect.* 8, 47–56. doi: 10.1007/s13659-017-0148-4

Cui, L., Yao, S., Dai, X., Yin, Q., Liu, Y., Jiang, X., et al. (2016). Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). *J. Exp. Bot.* 67, 2285–2297. doi: 10.1093/jxb/erw053

Dong, F., Hu, J., Shi, Y., Liu, M., Zhang, Q., and Ruan, J. (2019). Effects of nitrogen supply on flavonol glycoside biosynthesis and accumulation in tea leaves (*Camellia sinensis*). *Plant Physiol. Biochem.* 138, 48–57. doi: 10.1016/j.plaphy.2019.02.017

Feng, C. Y., Li, S. S., Taguchi, G., Wu, Q., Yin, D. D., Gu, Z. Y., et al. (2021). Enzymatic basis for stepwise C-glycosylation in the formation of flavonoid di-C-glycosides in sacred lotus (*Nelumbo nucifera* Gaertn.). *Plant J.* 106, 351–365. doi: 10.1111/tpj.15168

Griesser, M., Vitzthum, F., Fink, B., Bellido, M. L., Raasch, C., Munoz-Blanco, J., et al. (2008). Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria x ananassa*) achene and receptacle. *J. Exp. Bot.* 59, 2611–2625. doi: 10.1093/jxb/ern117

Guo, X. Y., Lv, Y. Q., Ye, Y., Liu, Z. Y., Zheng, X. Q., Lu, J. L., et al. (2021). Polyphenol oxidase dominates the conversions of flavonol glycosides in tea leaves. *Food Chem.* 339:128088. doi: 10.1016/j.foodchem.2020.128088

He, J., Yang, W., Cheng, B., Ma, L., Tursunjiang, D., Ding, Z., et al. (2020). Integrated metabolomic and transcriptomic profiling reveals the tissue-specific flavonoid compositions and their biosynthesis pathways in *Ziziphora bungeana*. *Chinese Med.* 15:73. doi: 10.1186/s13020-020-00354-6

Jarzebski, M., Smułek, W., Siejak, P., Rezler, R., Pawlicz, J., Trzeciak, T., et al. (2021). Aesculus hippocastanum L. as a stabilizer in hemp seed oil Nnanoemulsions for potential biomedical and food applications. *Int. J. Mol. Sci.* 22:887. doi: 10.3390/ijms22020887

Kapusta, I., Janda, B., Szajwaj, B., Stochmal, A., Piacente, S., Pizza, C., et al. (2007). Flavonoids in horse chestnut (*Aesculus hippocastanum*) seeds and powdered waste water byproducts. *J. Agric. Food Chem.* 55, 8485–8490. doi: 10.1021/jf071709t

Kimura, H., Ogawa, S., Ishihara, T., Maruoka, M., Tokuyama-Nakai, S., Jisaka, M., et al. (2017). Antioxidant activities and structural characterization of flavonol O-glycosides from seeds of Japanese horse chestnut (*Aesculus turbinata* BLUME). *Food Chem.* 228, 348–355. doi: 10.1016/j.foodchem.2017.01.084

Li, J., Qu, G., Shang, N., Chen, P., Men, Y., Liu, W., et al. (2021). Near-perfect control of the regioselective glucosylation enabled by rational design of glycosyltransferases. *Green Synth. Catal.* 2, 45–53. doi: 10.1016/j.gresc.2021.01.005

Li, L., Modolo, L. V., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A., and Wang, X. (2007). Crystal structure of *Medicago truncatula* UGT85H2–insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *J. Mol. Biol.* 370, 951–963. doi: 10.1016/j.jmb.2007.05.036

Liu, P. P., Yin, J. F., Chen, G. S., Wang, F., and Xu, Y. Q. (2018). Flavor characteristics and chemical compositions of oolong tea processed using different semi-fermentation times. *J. Food Sci. Technol.* 55, 1185–1195. doi: 10.1007/s13197-018-3034-0

Maharjan, R., Fukuda, Y., Nakayama, T., Nakayama, T., Hamada, H., Ozaki, S. I., et al. (2020). Crown-ether-mediated crystal structures of the glycosyltransferase PaGT3 from *Phytolacca americana*. *Acta Crystallogr. D Struct. Biol.* 76(Pt 6), 521–530. doi: 10.1107/S2059798320005306

Oszmiański, J., Kalisz, S., and Aneta, W. (2014). The content of phenolic compounds in leaf tissues of white (*Aesculus hippocastanum* L.) and red horse chestnut (*Aesculus carea* H.) colonized by the horse chestnut leaf miner (*Cameraria ohridella* Deschka and Dimić). *Molecules* 19, 14625–14636. doi: 10.3390/molecules190914625

Sasaki, N., Nishizaki, Y., Ozeki, Y., and Miyahara, T. (2014). The role of acyl-glucose in anthocyanin modifications. *Molecules* 19, 18747–18766. doi: 10.3390/molecules191118747

Scharbert, S., and Hofmann, T. (2005). Molecular definition of black tea taste by means of quantitative studies, taste reconstitution, and omission experiments. *J. Agric. Food Chem.* 53, 5377–5384. doi: 10.1021/jf050294d

Scharbert, S., Holzmann, N., and Hofmann, T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *J. Agric. Food Chem.* 52, 3498–3508. doi: 10.1021/jf049802u

Su, X., Shen, G., Di, S., Dixon, R. A., and Pang, Y. (2017). Characterization of UGT716A1 as a Multi-substrate UDP:Flavonoid Glucosyltransferase Gene in *Ginkgo biloba*. *Front. Plant Sci.* 8:2085. doi: 10.3389/fpls.2017.02085

Sun, W., Leng, L., Yin, Q., Xu, M., Huang, M., Xu, Z., et al. (2019). The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J.* 97, 841–857. doi: 10.1111/tpj.14162

Wei, F., Ma, S. C., Ma, L. Y., But, P. P., Lin, R. C., and Khan, I. A. (2004). Antiviral flavonoids from the seeds of *Aesculus chinensis*. *J. Nat. Prod.* 67, 650–653. doi: 10.1021/np030470h

Wu, L., Huang, X., Liu, S., Liu, J., Guo, Y., Sun, Y., et al. (2020). Understanding the formation mechanism of oolong tea characteristic non-volatile chemical constitutes during manufacturing processes by using integrated widely-targeted metabolome and DIA proteome analysis. *Food Chem.* 310:125941. doi: 10.1016/j.foodchem.2019.125941

Xu, M., Pirtskhalava, T., Farr, J. N., Weigand, B. M., Palmer, A. K., Weivoda, M. M., et al. (2018a). Senolytics improve physical function and increase lifespan in old age. *Nat. Med.* 24, 1246–1256. doi: 10.1038/s41591-018-0092-9

Xu, Y. Q., Zhang, Y. N., Chen, J. X., Wang, F., Du, Q. Z., and Yin, J. F. (2018b). Quantitative analyses of the bitterness and astringency of catechins from green tea. *Food Chem.* 258, 16–24. doi: 10.1016/j.foodchem.2018.03.042

Yang, M., Fehl, C., Lees, K. V., Lim, E. K., Offen, W. A., Davies, G. J., et al. (2018). Functional and informatics analysis enables glycosyltransferase activity prediction. *Nat. Chem. Biol.* 14, 1109–1117. doi: 10.1038/s41589-018-0154-9

Yin, Q., Han, X., Chen, J., Han, Z., Shen, L., Sun, W., et al. (2021). Identification of specific glycosyltransferases involved in flavonol glucoside biosynthesis in ginseng using integrative metabolite profiles, DIA proteomics, and phylogenetic analysis. *J. Agri. Food Chem.* 69, 1714–1726. doi: 10.1021/acs.jafc.0c06989

Yin, Q., Han, X., Han, Z., Chen, Q., Shi, Y., Gao, H., et al. (2020). Genome-wide analyses reveals a glucosyltransferase involved in rutin and emodin glucoside biosynthesis in tartary buckwheat. *Food Chem.* 318:126478. doi: 10.1016/j.foodchem.2020.126478

Yin, Q., Shen, G., Chang, Z., Tang, Y., Gao, H., and Pang, Y. (2017a). Involvement of three putative glucosyltransferases from the UGT72 family in flavonol glucoside/rhamnoside biosynthesis in *Lotus japonicus* seeds. *J. Exp. Bot.* 68, 597–612. doi: 10.1093/jxb/erw420

Yin, Q., Shen, G., Di, S., Fan, C., Chang, Z., and Pang, Y. (2017b). Genome-wide identification and functional characterization of UDP-glucosyltransferase genes involved in flavonoid biosynthesis in *Glycine max*. *Plant Cell Physiol.* 58, 1558–1572. doi: 10.1093/pcp/pcx081

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., et al. (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* 20, 2160–2176. doi: 10.1105/tpc.108.058040

Zhang, N., Liu, D., Wei, S., Cao, S., Feng, X., Wang, K., et al. (2020a). Phenylethanol glycosides from the seeds of *Aesculus chinensis* var. *chekiangensis*. *BMC Chem.* 14:31. doi: 10.1186/s13065-020-00685-3

Zhang, N., Wei, S., Cao, S., Zhang, Q., Kang, N., Ding, L., et al. (2020b). Bioactive triterpenoid saponins from the seeds of *Aesculus chinensis* Bge. *var. chekiangensis*. *Front. Chem.* 7:908. doi: 10.3389/fchem.2019.00908

Zhang, Z., Li, S., Zhang, S., and Gorenstein, D. (2006). Triterpenoid saponins from the fruits of *Aesculus pavia*. *Phytochemistry* 67, 784–794. doi: 10.1016/j.phytochem.2006.01.017

Zhu, C., Peng, W., Li, Y., Han, X., and Yu, B. (2006). Synthesis of 3-O-(beta-D-xylopyranosyl-(1–>2)-beta-D-glucopyranosyl)-3'-O-(beta-D-glucopyranosyl)tamarixetin, the putative structure of aescuflavoside A from the seeds of *Aesculus chinensis*. *Carbohydr Res.* 341, 1047–1051. doi: 10.1016/j.carres.2006.02.036

Zhuang, J., Dai, X., Zhu, M., Zhang, S., Dai, Q., Jiang, X., et al. (2020). Evaluation of astringent taste of green tea through mass spectrometry-based targeted metabolic profiling of polyphenols. *Food Chem.* 305:125507. doi: 10.1016/j.foodchem.2019.125507

Zlatanov, M. D., Antova, G. A., Angelova-Romova, M. J., and Teneva, O. T. (2013). Lipid composition of *Castanea sativa* Mill. and Aesculus hippocastanum fruit oils. *J. Sci. Food Agri.* 93, 661–666. doi: 10.1002/jsfa.5917

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 February 2022
doi: 10.3389/fpls.2022.829447

Genome-Wide Analysis of U-box E3 Ubiquitin Ligase Family in Response to ABA Treatment in *Salvia miltiorrhiza*

Chengan Chen[1†], Can Wang[1†], Junbo Li[1], Xiankui Gao[1], Qikai Huang[1], Yifu Gong[1], Xiaolong Hao[1], Itay Maoz[2], Guoyin Kai[1] and Wei Zhou[1*]*

[1] *Laboratory for Core Technology of TCM Quality Improvement and Transformation, School of Pharmaceutical Sciences, The Third Affiliated Hospital, School of Pharmacy and Academy of Chinese Medical Science, Zhejiang Chinese Medical University, Hangzhou, China,* [2] *Department of Postharvest Science, Agricultural Research Organization, Volcani Center, Rishon LeZion, Israel*

OPEN ACCESS

Edited by:
Xueqing Fu,
Shanghai Jiao Tong University, China

Reviewed by:
Chen Junfeng,
Shanghai University of Traditional Chinese Medicine, China
Zhichao Xu,
Northeast Forestry University, China

****Correspondence:***
Guoyin Kai
guoyinkai1@126.com
Wei Zhou
zhouwei19810501@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *05 December 2021*
Accepted: *04 January 2022*
Published: *09 February 2022*

Citation:
Chen C, Wang C, Li J, Gao X, Huang Q, Gong Y, Hao X, Maoz I, Kai G and Zhou W (2022) Genome-Wide Analysis of U-box E3 Ubiquitin Ligase Family in Response to ABA Treatment in Salvia miltiorrhiza. Front. Plant Sci. 13:829447. doi: 10.3389/fpls.2022.829447

Plant U-box (PUB) proteins are ubiquitin ligases (E3) involved in multiple biological processes and in response to plant stress. However, the various aspects of the genome and the differences in functions between the U-box E3 (UBE3) ubiquitin ligases remain quite obscure in *Salvia miltiorrhiza*. The 60 *UBE3* genes in the *S. miltiorrhiza* genome were recognized in the present study. The phylogenetic analysis, gene structure, motifs, promoters, and physical and chemical properties of the genes were also examined. Based on the phylogenetic relationship, the 60 *UBE3* genes were categorized under six different groups. The U-box domain was highly conserved across the family of *UBE3* genes. Analysis of the *cis*-acting element revealed that the *UBE3* genes might play an important role in a variety of biological processes, including a reaction to the abscisic acid (ABA) treatment. To investigate this hypothesis, an ABA treatment was developed for the hairy roots of *S. miltiorrhiza*. Thirteen out of the *UBE3* genes significantly increased after the ABA treatment. The co-expression network revealed that nine *UBE3* genes might be associated with phenolic acids or tanshinone biosynthesis. The findings of the present study brought fresh and new understanding to the participation of the *UBE3* gene family in plants, specifically in their biological responses mediated by the ABA. In *S. miltiorrhiza*, this gene family may be crucial during the ABA treatment. Significantly, the results of this study contribute novel information to the understanding of the ubiquitin ligase gene and its role in plant growth.

Keywords: *Salvia miltiorrhiza*, genome-wide analysis, ABA treatment, expression pattern, U-box E3 ubiquitin ligases

INTRODUCTION

Salvia miltiorrhiza is a well-known Chinese medicinal plant, being widely used for over 1,000 years in medicine and health foods. Many countries have used it in the treatment of cardiovascular and cerebrovascular diseases. Thus far, *S. miltiorrhiza* has served as a model for Chinese herbal medicine due to its comprehensive investigation of its medicinal value. Specifically, the *S. miltiorrhiza* produces two biologically active constituents: phenolic acids (water-soluble) and tanshinones (lipid-soluble) according to Deng et al. (2020); these show a variety of biological activities, such as anti-tumor, anti-inflammatory, and antibacterial outcomes. Consequently, DNA level gene

identification (genome analysis, functional gene mining, verification of metabolic pathways, etc.) and RNA transcription level regulation (transcription factor, small RNA mining, and function verification) of *S. miltiorrhiza* were systematic researched (Xu et al., 2016; Wu et al., 2018; Shi et al., 2021). However, only some studies reported the influence exerted by the abiotic/biotic stress or hormone treatment on the post-translational protein levels (Hirayama and Shinozaki, 2010). This indicates that the process of the modification of the ubiquitin protein, as well as the degradation of the functional proteins that manage the metabolic synthesis of the active components, continues to remain unclear in *S. miltiorrhiza*.

Both in eukaryotes and plants, the ubiquitin/26S proteasome system (UPS) pathway is one of the well-investigated mechanisms involved in the post-translational regulation of gene expression (Wang et al., 2018). The UPS is crucial in signal transduction, metabolic regulation, differentiation, cell cycle transition, and stress response regulation, resulting in specific proteins becoming degraded (Vandereyken et al., 2018). The UPS includes three enzymes that exhibit synergistic catalytic action, namely, ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3) (Trujillo and Shirasu, 2010). The free ubiquitin in the ATP hydrolysis is activated by the E1 enzyme, after which it is transferred to the E2 enzyme. In the presence of E3, ubiquitin is transferred from the E2-ubiquitin intermediate to the substrate protein (Richburg et al., 2014).

The E3 ubiquitin ligase family is the largest and shows the most diversity of the three enzymes catalyzing the ubiquitination cascade. It is a family of proteins that can recognize the target proteins that will be modified, which is the determining factor for specificity of ubiquitination (Serrano et al., 2018). The E3 ubiquitin ligases include several different families according to their action mode and conserved domains, such as the homology to the E6-associated protein C-terminus (HECT), Really Interesting New Gene (RING)/U-box E3 (UBE3) domain, and Cullin-RING ubiquitin ligases (CRLs). An HECT domain, characterized by 350 amino acids, is found on HECT type E3, and its functions to receive ubiquitin from an E2 enzyme and transfer it to a substrate protein (Rotin and Kumar, 2009). So far, *Arabidopsis thaliana* has been analyzed only for HECT type E3s, which contain seven *HECT* genes named UPL1-UPL7 (Downes et al., 2003). The RING/U-box proteins which are dissimilar to the HECT-type E3s, interact non-covalently with the E2 bearing the thioester-linked ubiquitin *via* the conserved RING/U-box domain to enable the ubiquitin to be transferred to the substrate (Callis, 2014). It is possible to use the RING/UBE3 ligase as a single unit for the direct transfer of the ubiquitin to a target protein. Conversely, the multi-component CRLs can function along with the Skp1-Cullin-F-box (SCF) complex (Harper and Tan, 2012; Serrano et al., 2018; Yu et al., 2020). The E3 ubiquitin ligases can either mediate the direct transfer of the ubiquitin to the protein or produce an intermediate complex (E3 and ubiquitin) prior to transfer to the substrate (Stone et al., 2005; Hua and Vierstra, 2011).

Abbreviations: PUB, Plant U-box; U-box E3, UBE3; ABA, abscisic acid; SA, salicylic acid; GA, gibberellin; MeJA, methyl jasmonate.

UBE3 ubiquitin ligases are a recently discovered subfamily of RING ubiquitinases, containing 70 amino acids residues in the U-box domain (Aravind and Koonin, 2000; Cyr et al., 2002). In comparison with the RING domain, the histidine and cysteine zinc ion chelating residues are absent in the U-box domain. Using the intramolecular interactions, the UBE3-ubiquitin ligases stabilize the U-box scaffolding employing the hydrogen bonds (Kraft et al., 2005; Yee and Goring, 2009). Besides, a domain of the U-box acts as the binding site where the E3-ubiquitin ligase binds with the E2 enzyme. A variety of UBE3-ubiquitin ligases have been identified in several plant genomes. In *Arabidopsis thaliana*, predictions of above 60 *UBE3* genes were made (Cho et al., 2008); other predictions include 77 *UBE3* genes in *Oryza sativa* (Zeng et al., 2008), 56 in Grapevine (*Vitis vinifera L.*) (Yu et al., 2016), 62 in *Solanum lycopersicum* (Sharma and Taganna, 2020), 91 in *Musa acuminate* (Hu et al., 2018), and 125 in Soybean [*Glycine max (L.) Merr.*] (Wang et al., 2016). When the *Arabidopsis PUB*13 was inactivated, it resulted in spontaneous cell death, escalation in salicylic acid (SA) the defense hormone, and early flowering (Trujillo, 2018). The *PUB* gene regulated the accumulation of resistance proteins in Grapevine both in response to biotic and abiotic stressors (Jiao et al., 2017). In Soybean, researchers produced two different *GmPUB8*-overexpressing lines from transgenic *Arabidopsis* plants and found that the heterogeneous over-expression of the *GmPUB8* in *Arabidopsis* resulted in the inhibition of ABA-mediated stomatal closure. The genes connected with drought stress were induced to a lower degree in the *GmPUB8*-overexpression in *Arabidopsis* post the treatment involving exposure to drought (Wang et al., 2016). Hence, from all these findings it appears that the UBE3 proteins may play a fundamental part in the regulation of a variety of biological processes, like hormone-signaling regulations, self-incompatibility, seed germination, and flowering time, as well as in abiotic/biotic stress (Kong et al., 2015; Ma et al., 2015; Trujillo, 2018).

To date, the dissection of the *S. miltiorrhiza* genome in its entirety offers a good platform for molecular biology to analyze its gene family, functional gene mining, and genome evolution. However, the *UBE3* gene family in *S. miltiorrhiza* has not yet been thoroughly examined. As these *UBE3* genes may be crucial in the regulation of the growth and development, it is imperative to fully examine the UBE3-ubiquitin ligase family in *S. miltiorrhiza*. In the present work, the *UBE3* gene number, gene structure, conserved domains, subgroup classification and co-expression analysis were systematically investigated in the entire genome of *S. miltiorrhiza*. Further, the expression profiles of genes in different tissues were studied undergoing the ABA treatment, thus giving a valuable reference for the functional recognition of the *UBE3* genes.

MATERIALS AND METHODS

Sequence Retrieval and Characterization

To look for potential UBE3 ubiquitin ligases, a search was conducted of the database on the *S. miltiorrhiza* genome

(Xu et al., 2016)[1]. From the Pfam database, it was possible to find the seed file of the U-box domain (PF04564). Utilizing the HMMER program, the potential UBE3 ubiquitin ligase members in *S. miltiorrhiza* were identified (Finn et al., 2011). All of the UBE3 protein candidates, drawn from the HMM search, were then submitted to the SMART website[2] in order to ascertain the conserved domain of the U-box. The physical locations of *UBE3* genes on scaffold were obtained from *S. miltiorrhiza* genome database and visualized by TBtools software[3]. Besides, the number of amino acids, molecular weight, theoretical pI, instability index, aliphatic index, and GRAVY (Grand average of hydropathicity) were calculated using the ExPASy-Compute pI/Mw tool (Gasteiger et al., 2005). Cell PLoc 2.0 was used to predict the subcellular localization of the *UBE3* candidate genes[4] (Emanuelsson et al., 2000).

Phylogenetic Tree Construction

In the present work, the UBE3 protein sequences were collected for *A. thaliana*[5], *S. lycopersicum*[6], and *S. miltiorrhiza*. Multiple alignments of 183 UBE3 protein sequences were performed using MEGA 6.0 software (Tamura et al., 2013). The neighbor-joining (NJ) method was employed to construct the phylogenetic tree, with 1,000 replications as the bootstrap value. The phylogenetic tree of the UBE3 proteins was edited by introducing an Interactive Tree of Life (iTOL) as shown by Letunic and Bork (2021).

Gene Structure and Conserved Motif Analysis

The *UBE3* genes in *S. miltiorrhiza* were identified and visualized in terms of the structural organization (coding domain sequences and untranslated regions) using the GSDS (gene structure display system) TOOL[7] (Hu et al., 2015). The conserved motifs of the *UBE3* genes were obtained utilizing the MEME suite[8], as shown by Bailey et al. (2009). Ten MEME motifs, ranging from 6 to 50 amino acids, were used in this study. All the results of the gene structure and conserved motifs were visualized using the TBtools software (Chen et al., 2020).

Promoter and Gene Ontology Analysis

The putative promoter sequences refer to 3,000 bp long upstream of the initiation codon. Sixty promoter sequences of *UBE3* genes were extracted using TBtools. PlanCare predicted the *cis*-acting regulatory elements of the promoter sequences (Rombauts et al., 1999). Based on the functional annotations of the *cis*-acting elements, the candidate elements were identified for more investigation. Those *cis*-acting elements possessing similar functional annotations were classified under the same class. Employing the WordArt tool[9], the word art image of the *cis*-acting elements present in the promoters was generated. From the PANTHER database, the ontology information was drawn and visualized (Mi et al., 2005).

Hairy Root Treated With ABA Treatment, Illumina Sequencing and *de novo* Transcriptome Assembly

The first step was the cultivation of the sterile *S. miltiorrhiza* plants on the Murashige and Skoog (MS) media keeping the temperature at 25°C, and under a photoperiod of 16 h light/8 h darkness (Zhou et al., 2021). Next, cultures of the *A. tumefaciens* strain C58C1 (pRiA4) were used to infect the *S. miltiorrhiza* sterile stems and/or leaves to produce hairy roots (Shi et al., 2016). The various treatments with ABA were performed on the well-developed hairy roots of *S. miltiorrhiza*. After 0, 0.5, 1, 2, 4, and 8 h of treatment, respectively, these hairy roots were selected to undergo RNA isolation and cDNA synthesis.

Employing the cDNA Synthesis Kit (Clontech, United States) and adopting the protocols prescribed, reverse transcription was done. Once separation was done of the double-stranded cDNAs on agarose gel, they were examined for the RNA-seq. The cDNA library was constructed using the Majorbio Biopharm Technology (Shanghai, China) and sequencing was done by Illumina HiSeq TM 2500 with PE100. In fact, all the reads were uploaded in the public database of the National Center for Biotechnology Information (NCBI) under the SRA access number SRP307198. As cited prior, the *de novo* assembly of the Illumina-sequenced short-length reads was accomplished (Zhou et al., 2017). A co-expression network of all *UBE3* genes between important phenolic acids or tanshinone biosynthetic genes was performed by Cytoscape software (Pearson correlation coefficient $r > 0.8$ and p-value < 0.05) (Shannon et al., 2003). Our analysis of RNA-seq data yielded FPKM values for quantification of gene expression, and we visualized our results using TBtools.

Gene Expression Profiles Detected by Quantitative Real-Time PCR

An RNA prep Pure Plant kit (Tiangen Biotech Co., Ltd., Beijing, China) was employed as a simple and inexpensive method of purifying the total RNA resulting from the treatment of different plant tissues (root, stem, leaf, and flower) and hairy roots with ABA, then subjected to RNA isolation and the RNA samples were reverse transcribed to cDNA (Zhou et al., 2021). After synthesizing the cDNA of each sample, a quantitative real-time PCR (qRT-PCR) assay was performed using a SuperReal PreMix Plus (SYBR Green) kit (Tiangen, China), on the ABI StepOnePlus Real-Time PCR System (Applied Biosystems, United States) as explained earlier (Mi et al., 2005). For an internal control in this study, the *SmActin* gene was used. **Supplementary Table 1** shows a summary of the primer pairs for qRT-PCR. The $2^{-\Delta\Delta Ct}$ quantification technique of gene expression was utilized (Shi et al., 2016). Three independent experiments were used to determine every data point.

[1] ftp://danshen.ndctcm.org:10402/Salvia_genome.fas

[2] http://smart.embl-heidelberg.de/

[3] https://github.com/CJ-Chen/TBtools

[4] http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/

[5] https://www.arabidopsis.org/

[6] ftp://ftp.solgenomics.net/tomato_genome/

[7] http://gsds.gao-lab.org/

[8] http://meme-suite.org/tools/meme

[9] https://wordart.com

RESULTS

Identification and Characterization of *UBE3* Gene Family

The HMMER tool with default parameters enabled the identification of 66 *UBE3* candidate genes in the whole genome of *S. miltiorrhiza*. The SMART tool was used to confirm that U-box domains existed on 66 *UBE3* candidate genes. Among them, 6 *UBE3* genes lacking the U-box domain were removed. Finally, a selection of 60 *UBE3* genes possessing the complete U-box domains were done and were categorized as *SmU-box 1-60*, respectively, depending upon the scores of *UBE3* genes presented in the HMM search. The *SmU-box* genes ranged from 555 to 9,332 bp in length, and encoded the polypeptides of 184 to 1,376 amino acids (aa), having a calculated molecular mass of 20 to 152.9 kDa, with 5.19 to 10.04 as the predicted isoelectric point (**Supplementary Table 2**). The sixty *SmU-box* genes were randomly and unequally distributed in the scaffold of the *S. miltiorrhiza* genome. Among them, scaffold 1254, 15043, and 7427 were the most distributed, with two *UBE3* genes (**Supplementary Figure 1**). For example, *SmU-box19* and *SmU-box31* on scaffold 1254 were 7,363 bp apart. *SmU-box43* was 37,433 bp away from *SmU-box49*, and *SmU-box43* shared 91.81% identity with *SmU-box49*. *SmU-box2* was 2,119 bp away from *SmU-box36* (**Supplementary Figure 1**). Interestingly, *SmU-box43* from *SmU-box49* was involved in a tandem duplication event. A majority of the proteins were predicted to be unstable and hydrophilic. From the subcellular localization, it was evident that most of the UBE3 proteins might be found in the nucleus, while the remaining were predicted to be localized either in the plasma membrane or cytosol (**Supplementary Table 2**).

Phylogenetic Relationships of the *UBE3* Gene Family Members

On the basis of the UBE3 proteins identified from *S. miltiorrhiza* (60 members), *S. lycopersicum* (62 members), and *A. thaliana* (61 members), a phylogenetic tree was drawn to examine the evolutionary history of the *UBE3* genes (**Figure 1**). From the three species cited earlier, the 183 *UBE3* genes were generally classified under six subgroups (I-VI). Group I included the *SmU-box52*, *53*, and *56*. It was clear that *SmU-box53* and *S. lycopersicum* ubiquitin fusion degradation 39 (*SlU-box39*) had the highest homology, with the amino acid homology of 49.64%. In Group II, each of the seven members contained a serine/threonine kinase domain in the region of the N-terminal. The biggest groups of the *SmU-box* proteins family included Groups III and IV, of which Group III possessed the ARMADILLO (ARM) repetitions at the C-termini. However, Group IV members all possessed around 100 amino acids in the GKL domain (leucine-rich feature) composed of conserved glycine (G), lysine (K), or arginine (R) residues, situated near the C-terminus of the protein (Zeng et al., 2008). Group VI, on the other hand, had the minor *SmU-box* group, which contained only one gene (*SmU-box*1). *SmU-box1* was highly homologous to the *S. lycopersicum* ubiquitin conjugation factor 18 (*SlU-box18*), reaching to the amino acid homology of 42.42%. According to the NCBI annotation, the *SlU-box18* may be a new ubiquitin conjugation factor. In Group V, seven members were presented, containing six WD40 repeat sequences. Generally, the *UBE3* genes of both *S. miltiorrhiza* and *S. lycopersicum* appeared to be clustered as a single subclade in the phylogenetic tree built, indicative of the relatively closer relationship between *S. miltiorrhiza* and *S. lycopersicum* when compared with *Arabidopsis*.

In all three species, the total number of *UBE3* gene families was relatively stable, revealing the conservative traits of *UBE3* genes. In order to evaluate the degree of gene expansion or loss during evolution, the *UBE3* genes in each group were counted. For the *S. miltiorrhiza*, it was found that Groups I–VI contained 3, 7, 21, 21, 7, and 1 *SmUBE3* genes, respectively (**Supplementary Table 3**). For the *S. lycopersicum*, Groups I, II, III, IV, V, and VI were noted to include 1, 9, 24, 24, 3, and 1 *SlU-box* genes, respectively. However, for *Arabidopsis*, Groups I, II, III, IV, V, and VI revealed 35, 12, 8, 3, 3, and 0 *AtU-box* genes, respectively. In both of the *S. miltiorrhiza* and *S. lycopersicum*, the comparatively consistent number of genes presented in each clade implies the absence of either gene expansion or loss, in these gene families. By contrast, in *S. miltiorrhiza*, Group I revealed rapid gene expansion while Groups III and IV showed two rapid gene losses in comparison to *Arabidopsis*.

Gene Structure and Motif Analysis of *UBE3* Genes

In order to examine the structure of the exons and introns, we aligned all the full-length cDNA sequences of the *UBE3* genes with the corresponding genome DNA sequences (**Figures 2A,B**). The exons in the *SmUBE3* genes ranged in number from 1 to 16. In *S. miltiorrhiza* among the 60 *UBE3* genes, the *SmU-box1* and *SmU-box41* revealed the most striking numbers, showing 16 exons, while 15 of the *UBE3* genes (25.0%) possessed only a single exon (**Supplementary Table 2**). The structural organization was indicative of the wide variations in the *UBE3* genes. This variation in the number of exons may suggest diverse functionality presented within the *UBE3* gene family (*Zhou et al., 2021*).

Furthermore, to identify the conserved motifs, including the UBE3 proteins that were assigned the title motifs 1–10, respectively, the Multiple Em for Motif Elicitation (MEME) was employed (**Supplementary Figure 2**). The core conservative motifs of the UBE3 proteins were Motifs 1, 2, and 3, while the ARM conventional motifs were the 4, 5, and 6 ones. To date, the characteristics of the motifs 7, 8, 9, and 10 remained unknown. The motifs of the *UBE3* genes that were still unclassified may offer crucial evidence for the biological function that has been recently acquired. The common motifs, however, implied a conserved evolutionary relationship and reveal similar enzyme functions (Nystrom and McKay, 2021). On detailed examination, most of the Group IV members contained only one exon. From this finding, it was evident that the members belonging to the same groups revealed similar gene structures and conserved motifs. It is clear that all of the above solid results validated the conclusions from the phylogenetic tree classifications (Wang et al., 2021).

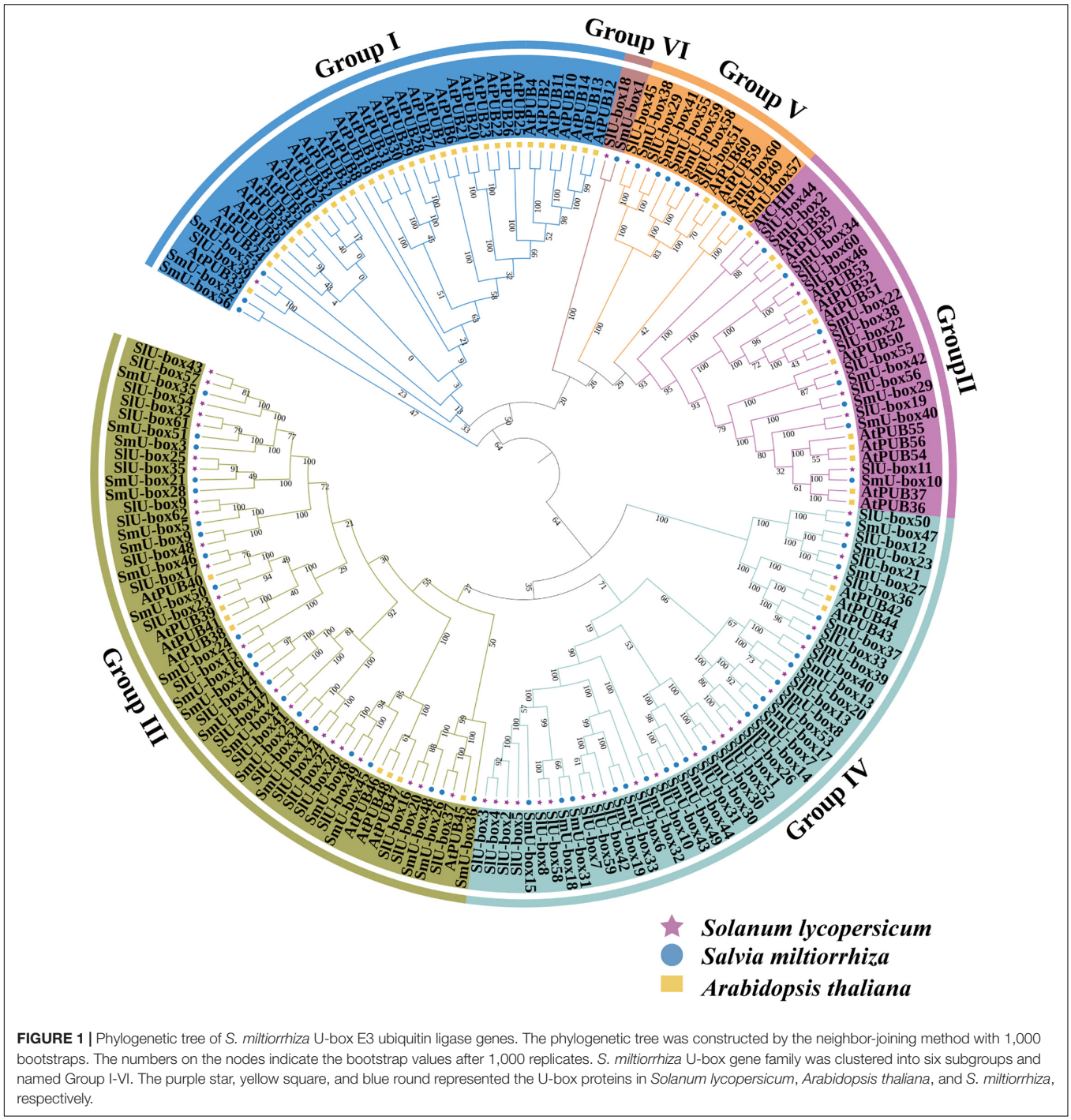


FIGURE 1 | Phylogenetic tree of *S. miltiorrhiza* U-box E3 ubiquitin ligase genes. The phylogenetic tree was constructed by the neighbor-joining method with 1,000 bootstraps. The numbers on the nodes indicate the bootstrap values after 1,000 replicates. *S. miltiorrhiza* U-box gene family was clustered into six subgroups and named Group I-VI. The purple star, yellow square, and blue round represented the U-box proteins in *Solanum lycopersicum*, *Arabidopsis thaliana*, and *S. miltiorrhiza*, respectively.

Prediction of *Cis*-Acting Elements in Promoter Regions and Gene Ontology Analysis

For greater investigation of the function of the *SmUBE3* genes, the PlantCARE database was introduced to predict the *cis*-acting elements. Finally, 14 *cis*-acting elements linked to stress, hormones, plant growth, and development present in the promoter regions of the 60 *UBE3* genes were selected for greater scrutiny. A diverse distribution pattern was observed in the promoter region of the *SmUBE3* genes, as revealed in **Figure 3**, the *S. miltiorrhiza UBE3* genes possessed a variety of biological functions. Several common hormone-related *cis*-acting elements were identified in *SmUBE3* genes including abscisic acid (ABA), salicylic acid (SA), gibberellin (GA), auxin, and methyl jasmonate (MeJA) (**Supplementary Figure 3**). For instance, the presence of an essential *cis*-acting element ABRE observed in response to ABA treatment (Osakabe et al., 2014) was noted in 53 *SmUBE3*

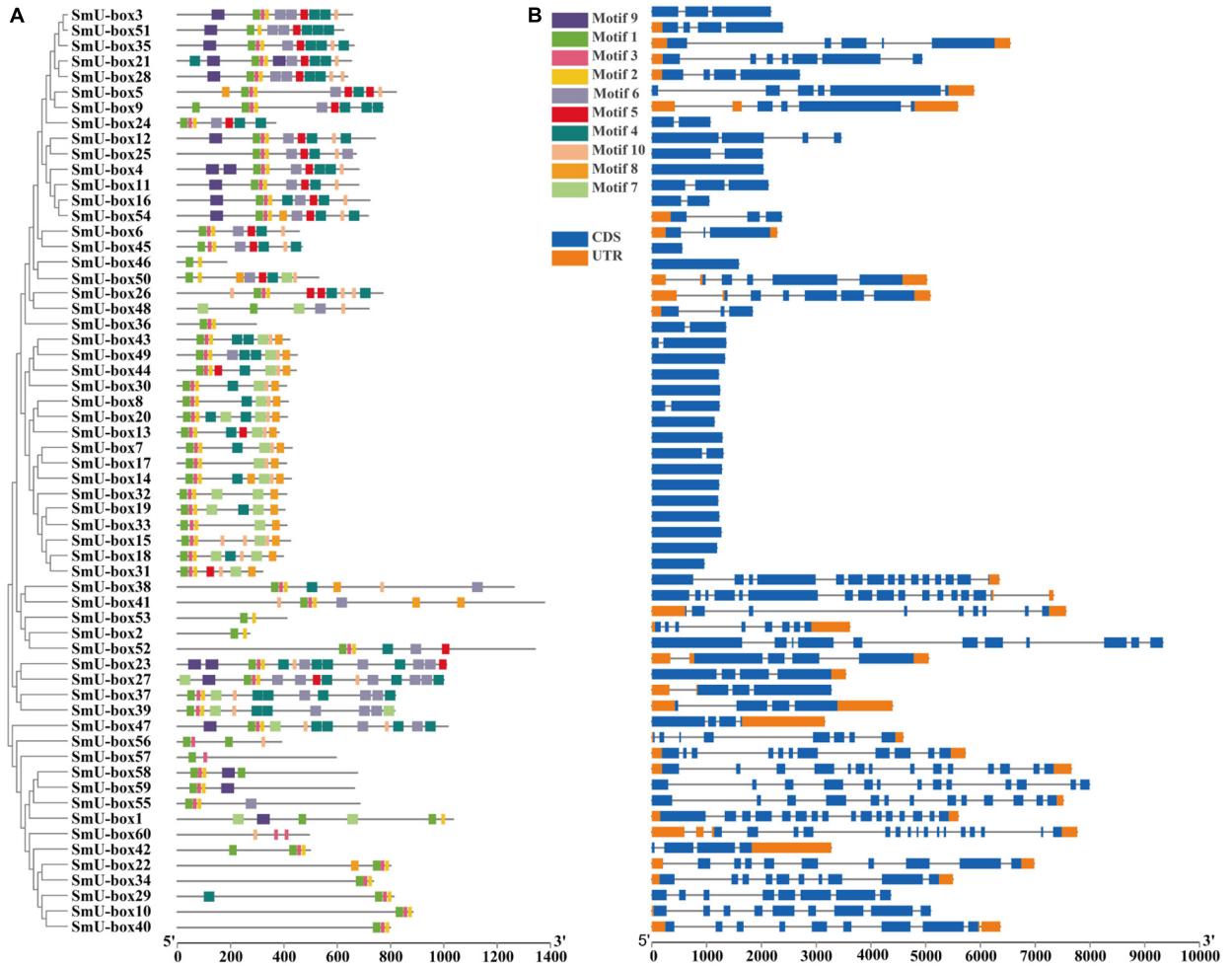


FIGURE 2 | The conserved motifs and gene structure analysis of *U-box* gene family in *S. miltiorrhiza*. **(A)** Distribution of ten motifs in all of the U-box proteins in *S. miltiorrhiza*. A total of 10 motifs were predicated and named motif 1–10. A neighbor-joining (NJ) phylogenetic tree of 60 *S. miltiorrhiza U-box* genes was constructed by Mega 6.0 with 1,000 bootstraps. **(B)** The gene structure of *U-box* genes in *S. miltiorrhiza*. The blue rectangles represented CDS, and the orange rectangles represented UTR.

genes. This implied that a large number of the *UBE3* genes in *S. miltiorrhiza* may reveal sensitivity to the ABA treatment. In fact, MeJA-responsive elements were recognized in 51 *SmUBE3* genes, and 42 *SmUBE3* genes held *cis*-acting elements related to cold stress, indicating that these genes may possess extraordinary resistance to treatments involving low temperatures. Among the essential plant defense mechanisms, flavonoid biosynthesis is very significant (Falcone Ferreyra et al., 2012). In flavonoid biosynthesis pathways, many R2R3-MYBs serve as activators to regulate the expression of structural genes (Schaart et al., 2013; Xu et al., 2015; Zhai et al., 2016; Zhang et al., 2019). For instance, by promoting the expression of flavonoid structural genes, researchers found that overexpression of *RrMYB5* and *RrMYB10* increased procyanidin accumulation in *Rosa rugosa* and tobacco (Shen et al., 2019). In this work, the MYB binding sites were identified, which participated in flavonoid biosynthesis regulation in 8 promoters of the *SmUBE3* genes (*SmU-box6*, *8*, *24*, *25*, *38*, *45*, *46*, and *52*).

A word cloud image was representative of the frequency of occurrence of the promoter element. In the class of *cis*-acting elements, the following components were recognized: abscisic acid (ABA) responsive elements (ACGTG), anaerobic induction (AAACCA), MeJA-responsive elements (TGACG and CGTCA), low-temperature responsive elements (CCGAAA), light-responsive elements (GGTTAA, CACGTG, and TACGTG), MYB binding site which participates in drought induction (CAACTG), auxin-responsive element (AACGAC), SA responsive element (CCATCTTTTT), gibberellin responsive element (TATCCCA), and stress-responsive elements (ATTCTCTAAC). These elements occurred in abundance in the promoters of the family of the UBE3 ubiquitin ligases. These were related to the abscisic acid response, anaerobic and MeJA induction; especially, the ABA induction had the highest frequency (**Figure 3** and **Supplementary Figure 4**). The large quantities of these regulatory elements in the promoter regions of the *SmUBE3* genes suggested that the *UBE3* gene family had strong

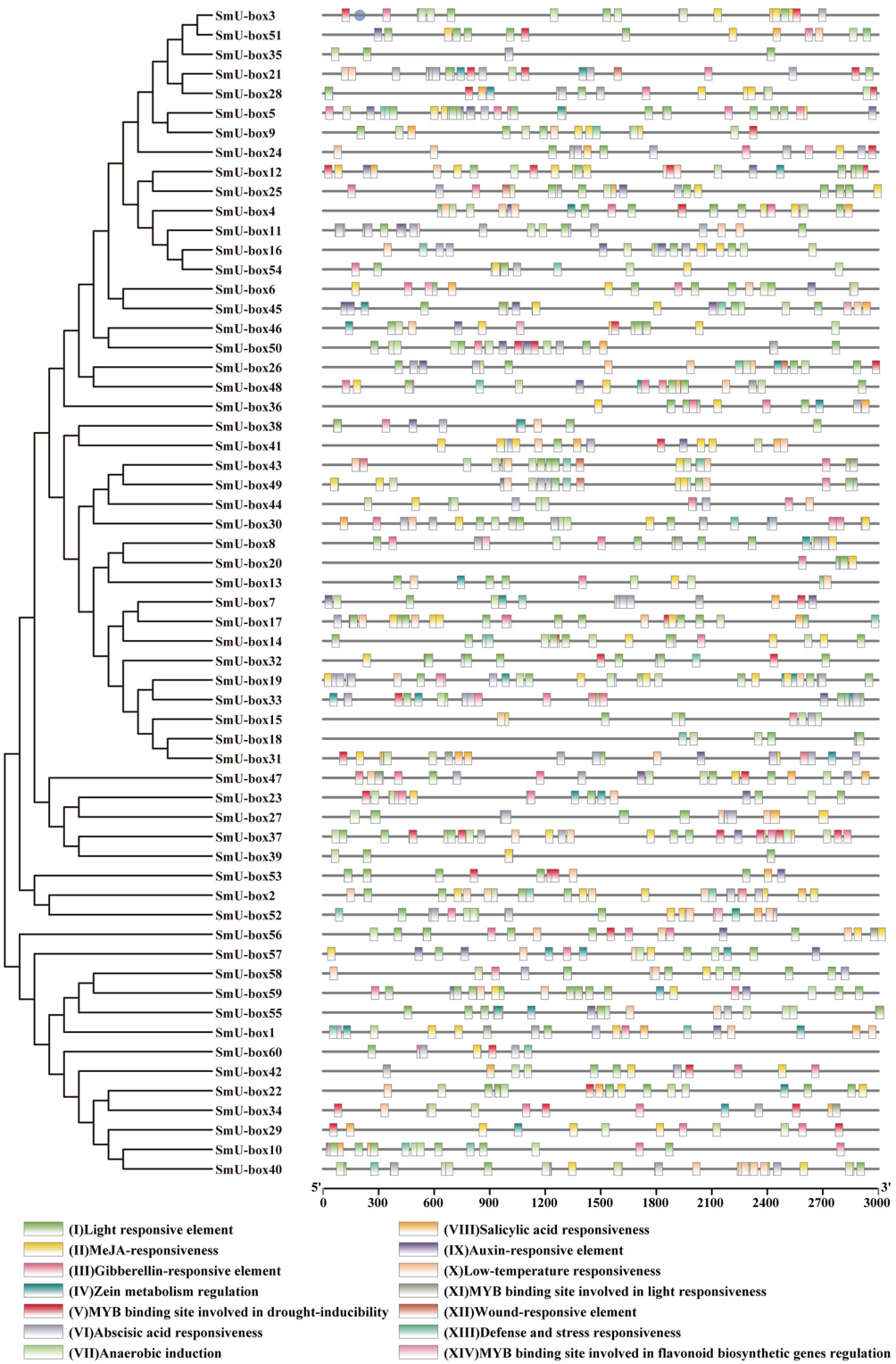


FIGURE 3 | The *cis*-acting elements draft of the putative promoters of 60 *U-box* genes in *S. miltiorrhiza*. The distribution pattern of 14 *cis*-acting elements of the putative promoters of *U-box* genes in *S. miltiorrhiza*. The phylogenetic tree is the same as in **Figure 1**. Fourteen *cis*-acting elements including (I) Light responsive element; (II) MeJA-responsiveness; (III) Gibberellin-responsive element; (IV) Zein metabolism regulation; (V) MYB binding site involved in drought-inducibility; (VI) Abscisic acid responsiveness; (VII) Anaerobic induction; (VIII) Salicylic acid responsiveness; (IX) Auxin-responsive element; (X) Low-temperature responsiveness; (XI) MYB binding site involved in light responsiveness; (XII)Wound-responsive element; (XIII) Defense and stress responsiveness; (XIV) MYB binding site involved in flavonoid biosynthetic genes regulation.

associations with the development of the plant and its response to hormone induction.

On analyzing the Gene Ontology (GO) of the 60 *UBE3* genes identified, the Blast2GO software was used to predict the biological process, cellular component, and molecular function (**Supplementary Figure 5**). A large part of the genes (71.42%) were assigned the GO category of "molecular function," while the classes of "biological process" and "cellular component" were observed to possess the same lower ratio (14.29%). The *UBE3* genes within the "molecular function" category, were mapped largely to the GO term "ubiquitin-protein ligase activity" (GO:0061630), showing a ratio of 71.42%. The prediction was that all the UBE3 proteins would be present in the nucleus (GO: 0005634) and cytoplasm (GO: 0005737). Further, within the "biological process," all the proteins were distinguished into protein ubiquitination (GO:0016567), ubiquitin-dependent protein catabolic process (GO:000651), peptidyl-amino acid modification (GO: 0018193), cellular response to misfolded protein (GO:0071218), protein quality control for misfolded or incompletely synthesized proteins (GO:0006515), positive regulation of proteolysis (GO:0045862), proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161), protein polyubiquitination (GO:0000209), and ubiquitin-dependent ERAD pathway (GO:0030433). From the gene ontology and promoter analysis, strong evidence was observed to support the relationship of the *UBE3* gene family in *S. miltiorrhiza* with a variety of metabolic pathways.

Expression Analysis of UBE3 Ubiquitin Ligases Under ABA Treatment

The ABA was justified as an abiotic inducer to encourage the biosynthesis of the tanshinones and phenolic acids (medicinal metabolites) in *S. miltiorrhiza* (Deng et al., 2020). The highest quantity of the *cis*-acting elements in the promoter regions of the *UBE3* genes was estimated to be ABA-responsive. In fact, six RNA-seq samples undergoing ABA treatment were collected, and after analysis by RNA sequencing, the expression pattern of the *UBE3* genes was studied in response to this treatment. The value of the Fragments Per Kilobases per Million reads (FPKM) for each of the *UBE3* genes was done to group the expression patterns using the k-means quantization method, it was possible to assemble all of them into six clusters, and those with > 4-fold change were considered as the upregulated *UBE3* genes (**Figure 4** and **Supplementary Table 4**). While Cluster II included three *UBE3* genes, Cluster III contained ten *UBE3* genes which were upregulated by the ABA treatment, with a short inducement time span of 0.5 and 1 h, respectively. Clusters of V and VI showed little or no variation in expression. Ultimately, the qRT-PCR method was used to validate the differential gene expression acquired by the RNA-seq. Under the ABA treatment, 13 *UBE3* genes revealed the highest expression level in comparison to the control; also, *SmU-box4*, *12*, *14*, *15*, *17*, *18*, *25*, *32*, *36*, *45*, *55*, *58*, and *59* were picked up for qRT-PCR testing. After applying the qRT-PCR technique, the expression patterns of the 13 genes acquired showed consistency with the trend of the patterns of expression drawn from RNA-seq data (**Figures 4**, **5**).

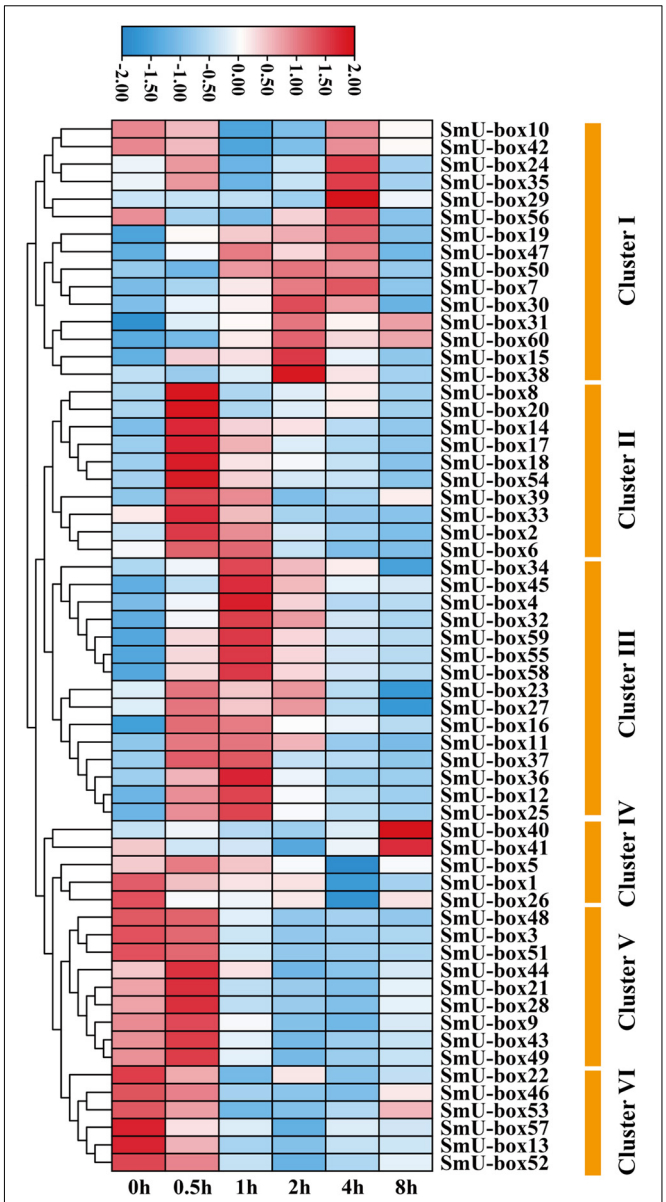


FIGURE 4 | The heat map of *U-box* genes in *S. miltiorrhiza* under ABA treatments based on RNA-seq. Genes with similar expression patterns were clustered into the same group according to the hierarchical clustering method. The top color scale indicated the Fragments Per Kilobases per Million reads (FPKM) values of each gene.

The correlation between the two data sets was observed to be significant, with the correlation coefficient in the range of 0.9091–0.9916 (**Supplementary Figure 6**). This established the reliability of the data obtained by RNA-seq for further investigating the *UBE3* candidate genes in response to the ABA treatments.

Differential Expression of UBE3 Ubiquitin Ligases in Various Tissues

As a means of studying the part they play in plant growth, development, and secondary metabolism biosynthesis of the

S. miltiorrhiza, 13 profiles of the ABA-responsive *UBE3* genes expressions in the vegetative and reproductive tissues (leaf, stem, root, and flower) were investigated (**Figure 6**). It was observed that the *SmU-box14*, *18*, and *25* were hard to detect in the reproductive tissues (flower). The *SmU-box4* and 18 uniformly revealed the highest expression in the leaf. Likewise, the *SmU-box32*, *58*, and 59 showed the highest expression in the stem. The *SmU-box12*, *14*, *15*, *17*, *25*, *36*, and *55* revealed the highest gene

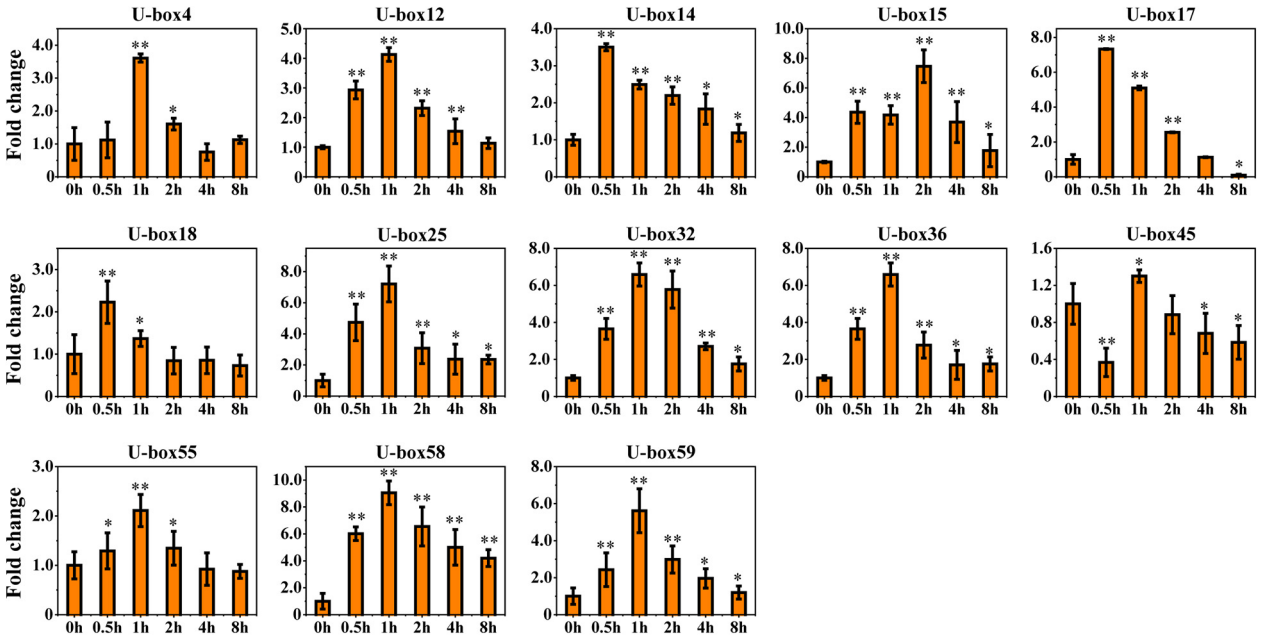


FIGURE 5 | qRT-PCR validation of 13 upregulated genes in response to ABA treatment in RNA-seq dataset. The relative gene expression level changes were normalized to the control without ABA induction at 0 h. Student's *t*-test was used with two significance levels (**$P < 0.01$; *$P < 0.05$), and the data presented the average of three replicates.

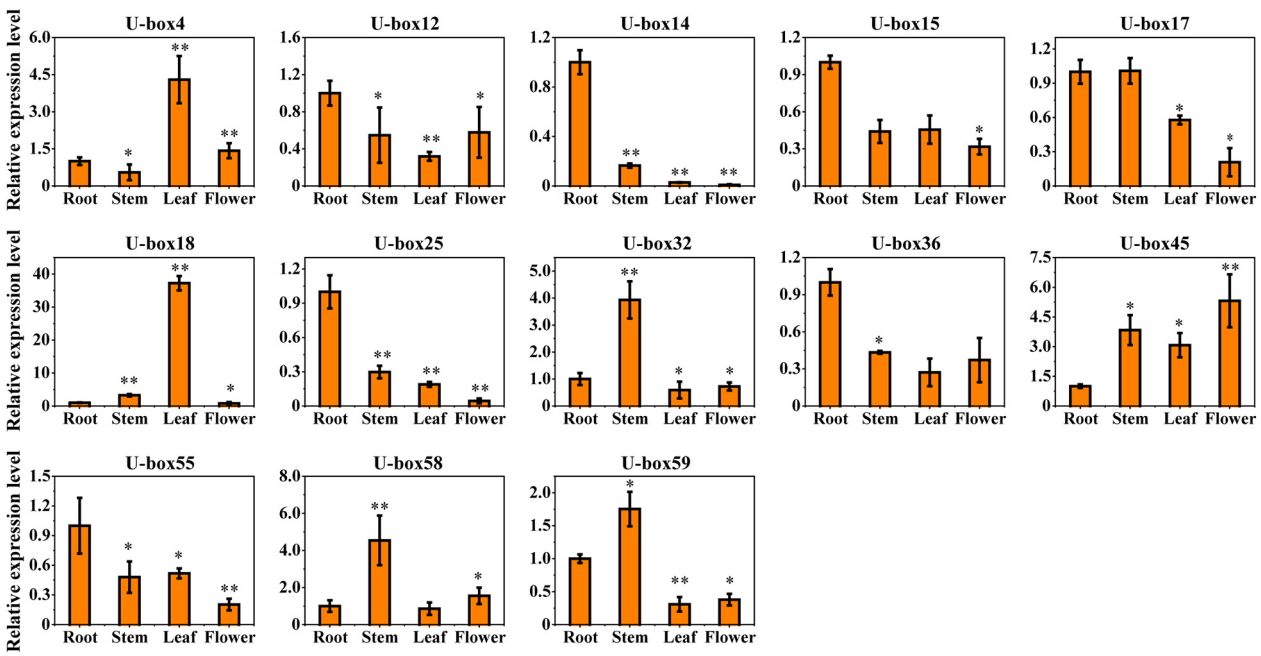


FIGURE 6 | The expression profiles of 13 selected *U-box* genes in different vegetative and reproductive tissues. Asterisks indicate significant differences in the root vs. other tissues by *t*-test with two significant levels (**$P < 0.01$; *$P < 0.05$), and the data presented the average of three replicates.

expression in the root, which in traditional Chinese medicine was utilized as a medicinal material (Zhou et al., 2005, 2021).

Co-expression Network of *UBE3* Genes With Phenolic Acids and Tanshinone Biosynthetic Genes

Previous studies showed that ABA could promote the phenolic acids and tanshinone accumulation in hairy roots by activating the expression of phenylalanine ammonia-lyase (PAL), tyrosine amino transferase (TAT), and geranylgeranyl diphosphate synthase (GGPPS), respectively (Shi et al., 2021). The co-expression network of 60 *UBE3* genes with phenolic acids and tanshinone biosynthetic genes was constructed, and these results indicated that 9 out of 60 *UBE3* genes showed a correlated relationship with phenolic acids and tanshinone biosynthetic genes (**Supplementary Figures 7, 8**). Co-expression analysis revealed that six *UBE3* genes (*SmU-box7*, *15*, *19*, *30*, *32*, *50*) showed a positive correlation with *GGPPS* gene in tanshinone biosynthetic pathway with Pearson correlation coefficient (r) $>$ 0.8 and *P*-value $<$ 0.05 as a cutoff (**Supplementary Table 5**). *SmU-box31* and *SmU-box60* genes were positively correlated with *PAL* gene in phenolic acids biosynthetic pathway ($r > 0.8$). Additionally, five *UBE3* genes (*SmU-box7*, *15*, *30*, *38*, *50*) co-expressed with the *TAT* gene in phenolic acid biosynthetic pathway ($r > 0.8$). Overall, these results suggested that 9 out of 60 *UBE3* genes might participate in phenolic acids and tanshinone biosynthesis.

DISCUSSION

The U-box E3 (UBE3) ubiquitin ligases show the wide distribution in plants and are well-known as participants in several biological functions (Santner and Estelle, 2010; Tian and Xie, 2013; Duplan and Rivas, 2014). As the *UBE3* genes play a crucial part in plant development, their presence has been recognized in many plant species, such as *A. thaliana* (Cho et al., 2008), rice (Zeng et al., 2008), tomato (Sharma and Taganna, 2020), and banana (Hu et al., 2018). Historically, *S. miltiorrhiza* (family Lamiaceae), one of the most valuable herbal plants extensively used in traditional Chinese medicine, is popular in the treatment of cardiovascular and cerebrovascular diseases (Zhou et al., 2005, 2016; Shi et al., 2016). To further characterize the function of the *UBE3* genes in *S. miltiorrhiza*, identifying these genes in the entire genome would be critical to *S. miltiorrhiza*. However, this field has been poorly studied to date. In this work, the HMMER analysis was done using the Pfam databases and 60 *UBE3* genes were recognized in the *S. miltiorrhiza* genome; it is interesting that the number of the *UBE3* genes in *S. miltiorrhiza* beard similarity to that of *A. thaliana* (61), tomato (62), and pear (62).

From phylogenetic analysis, it was evident that 183 UBE3 protein members in these three species (60 in *S. miltiorrhiza*, 62 in *S. lycopersicum*, and 61 in *A. thaliana*) were classified under six groups I-VI, (**Figure 1**). The phylogenetic analysis of *UBE3* genes revealed that all the six groups showed good similarity among themselves because the core U-box domain presented in all the genes. The grouping in the phylogenetic tree beard similarity to some other species, like soybean (Wang et al., 2016) and pear (Wang et al., 2021). Thus, the classification based on the phylogenetic relationship implied that when compared with the *PUBs* in *A. thaliana*, the *UBE3* family in *S. miltiorrhiza* demonstrated greater genetic closeness with the *S. lycopersicum*. This result clearly corresponded to the fact that *S. miltiorrhiza* and tomato got a closer relationship than *A. thaliana*. Besides the U-box domain, sixty *SmU-box* proteins were observed to bind to different domains such as the armadillo (ARM) repeats, protein kinase domain, WD40 repeats, GKL-box and the Ufd2P domain. The UBE3 proteins were derived largely from the U-box proteins possessing the ARM repeats (Wang et al., 2020). It is clear that the ARM repeats mostly mediated the interaction between the U-box proteins and substrate, implying that the interaction enabled the substrate to be available for the ubiquitination leading to degradation (Shu and Yang, 2017). The U-box with ARM domain in *S. miltiorrhiza* presented in twenty-one *UBE3* genes. Eight genes possessed the protein kinase domain, while seven genes included the WD40 repeats. In this work, the Ufd2P domain was identified as a single member of the UBE3 ubiquitin ligases family, namely the *SmU-box1*. Due to the differential distribution of these domains in the UBE3 ubiquitin ligases family, diverse biological functions may be inferred (Duplan and Rivas, 2014).

Further, from the structure and organization of the *UBE3* genes, the diversity existent in a gene family within the species is evident, which is related to the evolution and functional differentiation this gene family has experienced (Xu et al., 2012). In fact, from this study, it was evident that several of the *UBE3* genes in *S. miltiorrhiza* possessed either no introns at all or else contained upward of three introns (**Figure 2**). The general understanding was that a substantial number of introns in the *UBE3* genes behaved like a mutational buffer to ensure that the coding sequence was protected from mutations that are functionally deficient. However, the structural integrity of the gene families was evident from the *UBE3* genes which lacked the introns. All the *UBE3* genes identified were observed to be distributed in 10 classes of motifs. Motifs 1–3 were conserved and revealed good homology with the U-box domain. On the other hand, Motifs 4–6, along with the armadillo-like fold structure, were found to be clustered solely in group III. Hence, in the *S. miltiorrhiza UBE3* genes, the gene structure and motif analysis will be very helpful in predicting both gene evolution and functions for experimental authentication (Bailey et al., 2009).

On analysis of the *cis*-acting elements presented in the promoters, it is understood that the *UBE3* gene families participated in stress-related mechanisms, hormone regulation, growth, and development (**Figure 3**). In earlier studies, the *UBE3* genes were reported to have the ability to respond to ABA signals. Also, the AtPUB44 ubiquitinated abscission aldehyde oxidase 3 (AAO3) was recognized as a mediator in the ABA biosynthesis through 26 proteasomes (Raab et al., 2009). Besides, in *Arabidopsis*, the regulation of the ABI3 transcription factor was accomplished by the AtPUB9, resulting in heightened sensitivity to the ABA at the time of seedling germination (Samuel et al., 2008). In the case of *Pohlia nutans* (PnSAG1), where the U-box domain included the ARM repeat, the expression was seen

to occur rapidly in response to the exogenous ABA (Wang et al., 2019). The AtPUB18, AtPUB19, and AtPUB44 were identified as capable of causing direct interruption of the ABA biosynthesis (Sharma and Taganna, 2020). This study showed that in the putative promoter regions of fifty-three genes, the ABA-responsive elements were presented (**Supplementary Figure 3**). Specifically, in all the candidate genes, 13 ABA-responsive elements within the promoter region of the *SmU-box17* were observed to be at the maximum. From these results, it can be implied that the *UBE3* genes in *S. miltiorrhiza* may have a vital part to play in the ABA signal transduction process. In *Arabidopsis* and *Nicotiana*, it was the abiotic and biotic stresses which controlled the expression of the *PUB* gene (Raab et al., 2009). Earlier studies indicated that the expression of the resistant genes were regulated by the MYB transcription factor. For example, the PbrMYB21 was able to bind specifically to the MYB recognition sites in the *PbrADC* promoters and positively aid in drought tolerance (Li et al., 2017). In this work, many MYB transcription factor binding sites were recognized, limited to the promoter regions of the *UBE3* genes, and which were linked to drought induction resistance. Therefore, we reasonably concluded, that *UBE3* genes might most likely be regulated by the MYB genes concerned, which mediate the signaling of drought stress. In *S. miltiorrhiza*, the promoter region of *UBE3* genes gave supporting evidence for the presence of the defense and stress-responsive elements, low-temperature responsive elements, and wound-responsive elements. Therefore, it is understood that the *UBE3* genes could play a vital role in a variety of biological processes during the growth and development of *S. miltiorrhiza*.

It is assumed that the *UBE3* genes could take part in several events of abiotic stress and hormone induction (Liu et al., 2011; Seo et al., 2012; Song et al., 2017). Due to the frequent presence of the ABA-responsive elements in the promoter regions of the *SmUBE3* genes (**Supplementary Figure 4**), this study was conducted to examine the expression pattern of the *UBE3* genes which exhibited a response to the ABA treatment. From the expression profiles, it was possible to gain a clear picture of the potential participation of the *UBE3* genes in response to the ABA. In 13 of the fifty-three *UBE3* genes, it was possible to identify the ABA-responsive elements and validate their expression. Our data concurred with the earlier studies which reported the response of the *UBE3* genes to ABA induction (Seo et al., 2012). Further, the analysis of the gene expression corresponded to the prediction by the *cis*-elements of the *UBE3* gene family. Manifold elements were related to hormone regulation, not restricted to the ABA-responsive elements only, but were inclusive of the SA, GA, MeJA, and auxins (**Figure 3**). *SnRK2.6* identified as a phosphorylated protein, is considered to be a key gene involved in ABA signaling (Fujii et al., 2009; Melcher et al., 2009; Park et al., 2009; Umezawa et al., 2010). In *S. miltiorrhiza*, a recent study has revealed that *SmSnRK2.6* and *SmAREB1* genes were strongly induced by exogenous ABA. Meanwhile, SmSnRK2.6 protein interacted with SmAREB1 transcription factor. Overexpressing of *SmSnRK2.6* and *SmAREB1* coregulated the rosmarinic acid (RA) and salvianolic acid B (Sal B) accumulation in the transgenic *S. miltiorrhiza* hairy roots. This provided a novel insight on how exogenous ABA promoted the secondary metabolism at the post-translational level in *S. miltiorrhiza* (Jia et al., 2017).

Reports from earlier studies described the expression of the *UBE3* genes in a variety of tissues (Sharma and Taganna, 2020; Wang et al., 2021). In the present study, an investigation was done on the differential levels of expression of 13 *UBE3* genes in several vegetative and reproductive tissues employing the qRT-PCR technique (**Figures 5**, **6**). As the expression of the *UBE3* genes in the *S. miltiorrhiza* is higher in the roots than in the stem, flowers, and leaf, this implied that these genes could have a part to play in the root system formation and thus influence the synthesis of biologically active constituents. On the other hand, the poorest expression was noted in the flowers, indicative that some of these *UBE3* genes could be involved in the growth and development of *S. miltiorrhiza* during exposure to the ABA stress. Additionally, the analysis of co-expression network between *UBE3* genes and phenolic acids or tanshinone biosynthetic genes indicated that 9 *UBE3* genes might be associated with biosynthesis. Our findings show that the *UBE3* genes have a significant spatial association with development and plant growth in *S. miltiorrhiza*.

By way of conclusion, in this study, analyses had been done of the *UBE3* gene family in *S. miltiorrhiza* taking into account the whole genome and transcriptome level. This research presented firstly a systematic and comprehensive analysis of *UBE3* genes in *S. miltiorrhiza*. However, a proper understanding of the functional mechanism of the *UBE3* genes continues to be very meager. Still, this work offers valuable information that will encourage a more detailed understanding of the relationship between the *UBE3* genes and the ABA responses in the future. This paper may assist in a better comprehension of the biological functions of the *UBE3* genes in other plant species, under conditions of abiotic and biotic stress.

DATA AVAILABILITY STATEMENT

All reads of RNA-seq are publicly available in the National Center for Biotechnology Information (NCBI) database under the SRA accession number SRP307198.

AUTHOR CONTRIBUTIONS

WZ and GK formulated the study concept. CC collected the public data available on *S. miltiorrhiza* and the other species in this research. CC, CW, JL, XG, and QH accomplished the analyses of data and bioinformatics and preparation of the manuscript. WZ, GK, YG, and IM contributed to analysis planning and manuscript revision. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (2018YFC1706203); Zhejiang Provincial Natural Science Foundation of China (LY20H280008); the Major

Science and Technology Projects of Breeding New Varieties of Agriculture in Zhejiang Province (2021C02074); Opening Fund of Zhejiang Provincial First-Class Discipline for TCM (ZYAOXYB2019002); Pre-research Projects of Zhejiang Chinese Medical University (2021JKZKTS011A); and "Sannongliufang" Research Joint Project of Zhejiang Province (2021SNLF019).

ACKNOWLEDGMENTS

We would like to acknowledge the great experimental support of the Public Platform of Medical Research Center, Academy of Chinese Medical Science, Zhejiang Chinese Medical University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.829447/full#supplementary-material

Supplementary Figure 1 | The location of *UBE3* genes in *S. miltiorrhiza* genome.

Supplementary Figure 2 | Ten motifs identified using MEME suit for *U-box* gene family.

Supplementary Figure 3 | The number of 14 *cis*-acting elements in each putative promoter of *UBE3* genes. The color scale at the bottom indicated the number of *cis*-acting elements.

Supplementary Figure 4 | The word cloud image of *cis*-acting elements in the promoter of 60 *UBE3* genes.

Supplementary Figure 5 | Gene ontology analysis of 60 *UBE3* genes in *S. miltiorrhiza*. **(A)** Gene ontology analysis of the identified *UBE3* genes in *S. miltiorrhiza*. **(B)** The molecular function of *UBE3* genes in *S. miltiorrhiza*. **(C)** The biological process of *UBE3* genes in *S. miltiorrhiza*. **(D)** The cellular component of *UBE3* genes in *S. miltiorrhiza*.

Supplementary Figure 6 | Correlation of gene expression results obtained from qRT-PCR and RNA-seq.

Supplementary Figure 7 | The co-expression network of *UBE3* genes and phenolic acids biosynthetic genes.

Supplementary Figure 8 | The co-expression network of *UBE3* genes and tanshinone biosynthetic genes.

REFERENCES

Aravind, L., and Koonin, E. V. (2000). The U box is a modified RING finger-a common domain in ubiquitination. *Curr. Biol.* 10, R132–R134. doi: 10.1016/s0960-9822(00)00398-5

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

Callis, J. (2014). The ubiquitination machinery of the ubiquitin system. *Arabidopsis Book/Am. Soc. Plant Biol.* 12:e0174. doi: 10.1199/tab.0174

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

Cho, S. K., Ryu, M. Y., Song, C., Kwak, J. M., and Kim, W. T. (2008). *Arabidopsis* PUB22 and PUB23 are homologous U-Box E3 ubiquitin ligases that play combinatory roles in response to drought stress. *Plant Cell* 20, 1899–1914. doi: 10.1105/tpc.108.060699

Cyr, D. M., Höhfeld, J., and Patterson, C. (2002). Protein quality control: u-box-containing E3 ubiquitin ligases join the fold. *Trends Biochem. Sci.* 27, 368–375. doi: 10.1016/s0968-0004(02)02125-4

Deng, C., Shi, M., Fu, R., Zhang, Y., Wang, Q., Zhou, Y., et al. (2020). ABA-responsive transcription factor bZIP1 is involved in modulating biosynthesis of phenolic acids and tanshinones in *Salvia miltiorrhiza*. *J. Exp. Bot.* 71, 5948–5962. doi: 10.1093/jxb/eraa295

Downes, B. P., Stupar, R. M., Gingerich, D. J., and Vierstra, R. D. (2003). The HECT ubiquitin-protein ligase (UPL) family in *Arabidopsis*: UPL3 has a specific role in trichome development. *Plant J.* 35, 729–742. doi: 10.1046/j.1365-313x.2003.01844.x

Duplan, V., and Rivas, S. (2014). E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. *Front. Plant Sci.* 5:42. doi: 10.3389/fpls.2014.00042

Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903

Falcone Ferreyra, M. L., Rius, S., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S. Y., et al. (2009). In vitro reconstitution of an abscisic acid signalling pathway. *Nature* 462, 660–664. doi: 10.1038/nature08599

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, (Berlin: Springer), 571–607. doi: 10.1385/1-59259-890-0:571

Harper, J. W., and Tan, M. K. M. (2012). Understanding cullin-RING E3 biology through proteomics-based substrate identification. *Mol. Cell. Proteomics* 11, 1541–1550. doi: 10.1074/mcp.R112.021154

Hirayama, T., and Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.* 61, 1041–1052. doi: 10.1111/j.1365-313X.2010.04124.x

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hu, H., Dong, C., Sun, D., Hu, Y., and Xie, J. (2018). Genome-wide identification and analysis of U-box E3 ubiquitin-protein ligase gene family in banana. *Int. J. Mol. Sci.* 19:3874. doi: 10.3390/ijms19123874

Hua, Z., and Vierstra, R. D. (2011). The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* 62, 299–334. doi: 10.1146/annurev-arplant-042809-112256

Jia, Y., Bai, Z., Pei, T., Ding, K., Liang, Z., and Gong, Y. (2017). The protein kinase *SmSnRK2.6* positively regulates phenolic acid biosynthesis in *Salvia miltiorrhiza* by interacting with *SmAREB1*. *Front. Plant Sci.* 8:1384. doi: 10.3389/fpls.2017.01384

Jiao, L., Zhang, Y., and Lu, J. (2017). Overexpression of a stress-responsive U-box protein gene *VaPUB* affects the accumulation of resistance related proteins in *Vitis vinifera* 'Thompson Seedless'. *Plant Physiol. Biochem.* 112, 53–63. doi: 10.1016/j.plaphy.2016.12.019

Kong, L., Cheng, J., Zhu, Y., Ding, Y., Meng, J., Chen, Z., et al. (2015). Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases. *Nat. Commun.* 6, 1–13. doi: 10.1038/ncomms9630

Kraft, E., Stone, S. L., Ma, L., Su, N., Gao, Y., Lau, O. S., et al. (2005). Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of *Arabidopsis*. *Plant Physiol.* 139, 1597–1611. doi: 10.1104/pp.105.067983

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, K., Xing, C., Yao, Z., and Huang, X. (2017). *PbrMYB21*, a novel MYB protein of *Pyrus betulaefolia*, functions in drought tolerance and modulates polyamine levels by regulating arginine decarboxylase gene. *Plant Biotechnol. J.* 15, 1186–1203. doi: 10.1111/pbi.12708

Liu, Y. C., Wu, Y. R., Huang, X. H., Sun, J., and Xie, Q. (2011). AtPUB19, a U-box E3 ubiquitin ligase, negatively regulates abscisic acid and drought

responses in *Arabidopsis thaliana*. *Mol. Plant* 4, 938–946. doi: 10.1093/mp/ssr030

Ma, X., Mo, B., and Cao, X. (2015). New players in ABA signaling: identification of PUB12/13 involved in degradation of ABA co-receptor ABI1. *Sci. China Life Sci.* 58:1173. doi: 10.1007/s11427-015-4947-8

Melcher, K., Ng, L. M., Zhou, X. E., Soon, F. F., Xu, Y., Suino-Powell, K. M., et al. (2009). A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* 462, 602–608. doi: 10.1038/nature08613

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res*. 33, D284–D288. doi: 10.1093/nar/gki078

Nystrom, S. L., and McKay, D. J. (2021). Memes: a motif analysis environment in R using tools from the MEME Suite. *PLoS Computat. Biol.* 17:e1008991. doi: 10.1371/journal.pcbi.1008991

Osakabe, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L. S. P. (2014). ABA control of plant macroelement membrane transport systems in response to water deficit and high salinity. *New Phytol.* 202, 35–49. doi: 10.1111/nph.12613

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* 324, 1068–1071. doi: 10.1126/science.1173041

Raab, S., Drechsel, G., Zarepour, M., Hartung, W., Koshiba, T., Bittner, F., et al. (2009). Identification of a novel E3 ubiquitin ligase that is required for suppression of premature senescence in *Arabidopsis*. *Plant J.* 59, 39–51. doi: 10.1111/j.1365-313X.2009.03846.x

Richburg, J. H., Myers, J. L., and Bratton, S. B. (2014). The role of E3 ligases in the ubiquitin-dependent regulation of spermatogenesis. *Semin. Cell Dev. Biol*. 30, 27–35. doi: 10.1016/j.semcdb.2014.03.001

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

Rotin, D., and Kumar, S. (2009). Physiological functions of the HECT family of ubiquitin ligases. *Nat. Rev. Mol. Cell Biol*. 10, 398–409. doi: 10.1038/nrm2690

Samuel, M. A., Mudgil, Y., Salt, J. N., Delmas, F., Ramachandran, S., Chilelli, A., et al. (2008). Interactions between the S-domain receptor kinases and AtPUB-ARM E3 ubiquitin ligases suggest a conserved signaling pathway in *Arabidopsis*. *Plant Physiol*. 147, 2084–2095. doi: 10.1104/pp.108.123380

Santner, A., and Estelle, M. (2010). The ubiquitin-proteasome system regulates plant hormone signaling. *Plant J.* 61, 1029–1040. doi: 10.1111/j.1365-313X.2010.04112.x

Schaart, J. G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A. M., de Vos, R. C., Jonker, H. H., et al. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits. *New Phytol.* 197, 454–467. doi: 10.1111/nph.12017

Seo, D. H., Ryu, M. Y., Jammes, F., Hwang, J. H., Turek, M., Kang, B. G., et al. (2012). Roles of four *Arabidopsis* U-box E3 ubiquitin ligases in negative regulation of abscisic acid-mediated drought stress responses. *Plant Physiol.* 160, 556–568. doi: 10.1104/pp.112.202143

Serrano, I., Campos, L., and Rivas, S. (2018). Roles of E3 ubiquitin-ligases in nuclear protein homeostasis during plant stress responses. *Front. Plant Sci.* 9:139. doi: 10.3389/fpls.2018.00139

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

Sharma, B., and Taganna, J. (2020). Genome-wide analysis of the U-box E3 ubiquitin ligase enzyme gene family in tomato. *Sci. Rep*. 10:9581. doi: 10.1038/s41598-020-66553-1

Shen, Y., Sun, T., Pan, Q., Anupol, N., Chen, H., Shi, J., et al. (2019). RrMYB5- and RrMYB10- regulated flavonoid biosynthesis plays a pivotal role in feedback loop responding to wounding and oxidation in *Rosa rugosa*. *Plant Biotechnol. J.* 17, 2078–2095. doi: 10.1111/pbi.13123

Shi, M., Hua, Q., and Kai, G. (2021). Comprehensive transcriptomic analysis in response to abscisic acid in *Salvia miltiorrhiza*. *Plant Cell Tissue Organ. Culture (PCTOC)* 147, 389–404. doi: 10.1007/s11240-021-02135-x

Shi, M., Luo, X., Ju, G., Li, L., Huang, S., Zhang, T., et al. (2016). Enhanced diterpene tanshinone accumulation and bioactivity of transgenic *Salvia miltiorrhiza* hairy roots by pathway engineering. *J. Agric. Food Chem*. 64, 2523–2530. doi: 10.1021/acs.jafc.5b04697

Shu, K., and Yang, W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol*. 58, 1461–1476. doi: 10.1093/pcp/pcx071

Song, J., Mo, X., Yang, H., Yue, L., Song, J., and Mo, B. (2017). The U-box family genes in *Medicago truncatula*: key elements in response to salt, cold, and drought stresses. *PLoS One* 12:e0182402. doi: 10.1371/journal.pone.0182402

Stone, S. L., Hauksdóttir, H., Troy, A., Herschleb, J., Kraft, E., and Callis, J. (2005). Functional analysis of the RING-Type ubiquitin ligase family of *Arabidopsis*. *Plant Physiol.* 137, 13–30. doi: 10.1104/pp.104.052423

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tian, M., and Xie, Q. (2013). Non-26S proteasome proteolytic role of ubiquitin in plant endocytosis and endosomal trafficking. *J. Integrat. Plant Biol*. 55, 54–63. doi: 10.1111/jipb.12007

Trujillo, M. (2018). News from the PUB: plant U-box type E3 ubiquitin ligases. *J. Exp. Bot*. 69, 371–384. doi: 10.1093/jxb/erx411

Trujillo, M., and Shirasu, K. (2010). Ubiquitination in plant immunity. *Curr. Opin. Plant Biol.* 13, 402–408. doi: 10.1016/j.pbi.2010.04.002

Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., et al. (2010). Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant Cell Physiol.* 51, 1821–1839. doi: 10.1093/pcp/pcq156

Vandereyken, K., Van Leene, J., De Coninck, B., and Cammue, B. (2018). Hub protein controversy: taking a closer look at plant stress response hubs. *Front. Plant Sci*. 9:694. doi: 10.3389/fpls.2018.00694

Wang, C., Song, B., Dai, Y., Zhang, S., and Huang, X. (2021). Genome-wide identification and functional analysis of U-box E3 ubiquitin ligases gene family related to drought stress response in Chinese white pear (*Pyrus bretschneideri*). *BMC Plant Biol*. 21:235. doi: 10.1186/s12870-021-03024-3

Wang, J., Liu, S., Liu, H., Chen, K., and Zhang, P. (2019). PnSAG1, an E3 ubiquitin ligase of the Antarctic moss *Pohlia nutans*, enhanced sensitivity to salt stress and ABA. *Plant Physiol. Biochem*. 141, 343–352. doi: 10.1016/j.plaphy.2019.06.002

Wang, K., Yang, Q., Lanhuang, B., Lin, H., Shi, Y., Dhanasekaran, S., et al. (2020). Genome-wide investigation and analysis of U-box Ubiquitin-Protein ligase gene family in apple: expression profiles during *Penicillium expansum* infection process. *Physiol. Mol. Plant Pathol*. 111:101487. doi: 10.1016/j.pmpp.2020.101487

Wang, N., Liu, Y., Cong, Y., Wang, T., Zhong, X., Yang, S., et al. (2016). Genome-wide identification of soybean U-box E3 ubiquitin ligases and roles of GmPUB8 in negative regulation of drought stress response in Arabidopsis. *Plant Cell Physiol*. 57, 1189–1209. doi: 10.1093/pcp/pcw068

Wang, Z., Tian, X., Zhao, Q., Liu, Z., Li, X., Ren, Y., et al. (2018). The E3 ligase DROUGHT HYPERSENSITIVE negatively regulates cuticular wax biosynthesis by promoting the degradation of transcription factor ROC4 in rice. *Plant Cell* 30, 228–244. doi: 10.1105/tpc.17.00823

Wu, Y., Zhang, Y., Li, L., Guo, X., Wang, B., Cao, X., et al. (2018). AtPAP1 interacts with and activates SmbHLH51, a positive regulator to phenolic acids biosynthesis in *Salvia miltiorrhiza*. *Front. Plant Sci.* 9:1687. doi: 10.3389/fpls.2018.01687

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. U.S.A*. 109, 1187–1192. doi: 10.1073/pnas.1109047109

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9:949. doi: 10.1016/j.molp.2016.03.010

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Yee, D., and Goring, D. R. (2009). The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *J. Exp. Bot.* 60, 1109–1121. doi: 10.1093/jxb/ern369

Yu, H., Jiang, M., Xing, B., Liang, L., Zhang, B., and Liang, Z. (2020). Systematic analysis of Kelch Repeat F-box (KFB) protein family and identification of phenolic acid regulation members in *Salvia miltiorrhiza* Bunge. *Genes* 11:557. doi: 10.3390/genes11050557

Yu, Y. H., Li, X. Z., Guo, D. L., Zhang, H. L., Li, G. R., Li, X. Q., et al. (2016). Genome-wide identification and analysis of the U-box family of E3 ligases in grapevine. *Russian J. Plant Physiol.* 63, 835–848. doi: 10.1134/S1021443716050186

Zeng, L. R., Park, C. H., Venu, R. C., Gough, J., and Wang, G. L. (2008). Classification, expression pattern, and E3 ligase activity assay of rice U-box-containing proteins. *Mol. Plant* 1, 800–815. doi: 10.1093/mp/ssn044

Zhai, R., Wang, Z., Zhang, S., Meng, G., Song, L., Wang, Z., et al. (2016). Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.). *J. Exp. Bot.* 67, 1275–1284. doi: 10.1093/jxb/erv524

Zhang, X., Xu, Z., Yu, X., Zhao, L., Zhao, M., Han, X., et al. (2019). Identification of two novel R2R3-MYB transcription factors, *PsMYB114L* and *PsMYB12L*, related to anthocyanin biosynthesis in *Paeonia suffruticosa*. *Int. J. Mol. Sci.* 20:1055. doi: 10.3390/ijms20051055

Zhou, L., Zuo, Z., and Chow, M. S. S. (2005). Danshen: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use. *J. Clin. Pharmacol.* 45, 1345–1359. doi: 10.1177/0091270005282630

Zhou, W., Huang, F., Li, S., Wang, Y., Zhou, C., Shi, M., et al. (2016). Molecular cloning and characterization of two 1-deoxy-d-xylulose-5-phosphate synthase genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Mol. Breed.* 36, 1–12. doi: 10.1007/s11032-016-0550-3

Zhou, W., Huang, Q., Wu, X., Zhou, Z., Ding, M., Shi, M., et al. (2017). Comprehensive transcriptome profiling of *Salvia miltiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci. Rep.* 7:10554. doi: 10.1038/s41598-017-10215-2

Zhou, W., Shi, M., Deng, C., Lu, S., Huang, F., Wang, Y., et al. (2021). The methyl jasmonate-responsive transcription factor SmMYB1 promotes phenolic acid biosynthesis in *Salvia miltiorrhiza*. *Hortic. Res.* 8, 1–13. doi: 10.1038/s41438-020-00443-5

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fpls.2022.822374

Qualitative Proteome-Wide Analysis Reveals the Diverse Functions of Lysine Crotonylation in *Dendrobium huoshanense*

Jing Wu[1,2], Xiaoxi Meng[3], Weimin Jiang[4], Zhaojian Wang[1], Jing Zhang[1], Fei Meng[1], Xiaoyan Yao[1], Mengjuan Ye[1], Liang Yao[1], Longhai Wang[5], Nianjun Yu[1,2], Daiyin Peng[1,2,6] and Shihai Xing[1,2,7]**

[1] College of Pharmacy, Anhui University of Chinese Medicine, Hefei, China, [2] Institute of Traditional Chinese Medicine Resources Protection and Development, Anhui Academy of Chinese Medicine, Hefei, China, [3] Department of Horticultural Science, University of Minnesota, Saint Paul, MN, United States, [4] Hunan Key Laboratory for Conservation and Utilization of Biological Resources in the Nanyue Mountainous Region, College of Life Sciences and Environment, Hengyang Normal University, Hengyang, China, [5] School of Integrated Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, China, [6] Synergetic Innovation Center of Anhui Authentic Chinese Medicine Quality Improvement, Hefei, China, [7] Anhui Province Key Laboratory of Research and Development of Chinese Medicine, Hefei, China

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Jian Zhang,
China National Rice Research Institute, Chinese Academy of Agricultural Sciences (CAAS), China
Pingfang Yang,
Hubei University, China

***Correspondence:**
Daiyin Peng
pengdy@ahtcm.edu.cn
Shihai Xing
xshshihai@163.com

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 25 November 2021
Accepted: 18 January 2022
Published: 16 February 2022

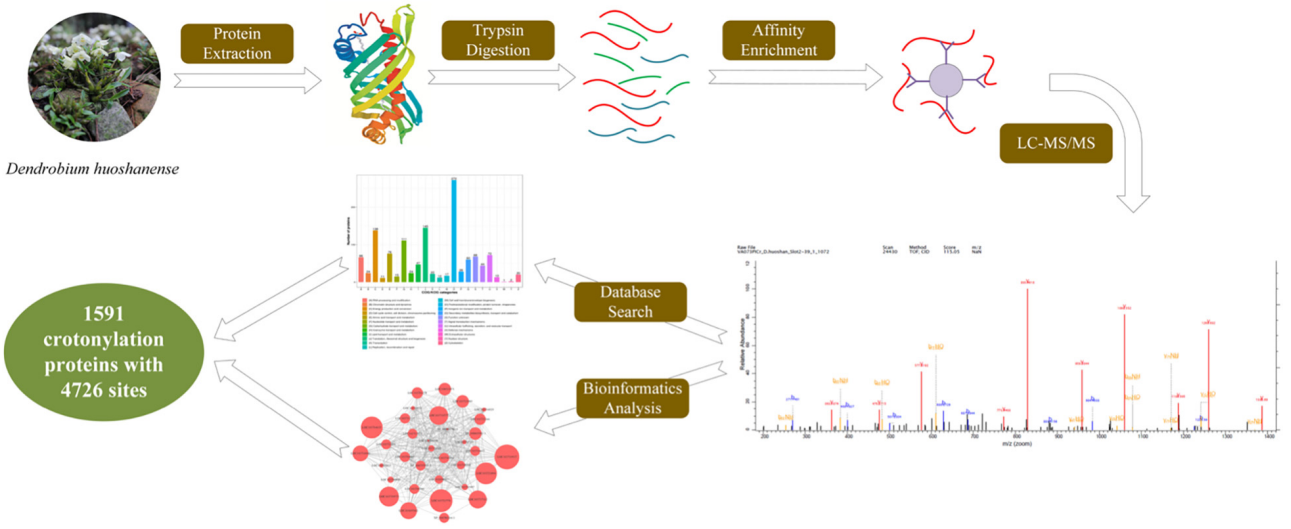
Citation:
Wu J, Meng X, Jiang W, Wang Z, Zhang J, Meng F, Yao X, Ye M, Yao L, Wang L, Yu N, Peng D and Xing S (2022) Qualitative Proteome-Wide Analysis Reveals the Diverse Functions of Lysine Crotonylation in Dendrobium huoshanense. Front. Plant Sci. 13:822374. doi: 10.3389/fpls.2022.822374

The lysine crotonylation of histone proteins is a newly identified posttranslational modification with diversified cellular functions. However, there are few reports on lysine crotonylation of non-histone proteins in medicinal plant cells. By using high-resolution liquid chromatography–mass spectrometry (LC-MS) coupled with highly sensitive-specific immune-affinity antibody analysis, a whole crotonylation proteome analysis of *Dendrobium huoshanense* was performed. In total, 1,591 proteins with 4,726 lysine crotonylation sites were identified; among them, 11 conserved motifs were identified. Bioinformatic analyses linked crotonylated proteins to the drought stress response and multiple metabolic pathways, including secondary metabolite biosynthesis, transport and catabolism, energy production and conversion, carbohydrate transport and metabolism, translation, and ribosomal structure and biogenesis. This study contributes toward understanding the regulatory mechanism of polysaccharide biosynthesis at the crotonylation level even under abiotic stress.

Keywords: *Dendrobium huoshanense*, lysine crotonylation, polysaccharide biosynthesis, heavily-crotonylated, photosynthesis

INTRODUCTION

Dendrobium huoshanense, an edible and medicinal species of family Orchidaceae, was traditionally and popularly called "mihu" and has been claimed to maintain health and prolong life in the Supplement to Compendium of Materia Medica (Zhao, 2007). Famous as a tea material, soup ingredient, and Chinese herbal medicine, *D. huoshanense* can relieve immune deficiencies, improve eyesight, nourish the stomach, and produce body fluid (Bao et al., 2001). A study on plants of the genus *Dendrobium* showed that alkaloids and polysaccharides were their main bioactive compounds (Chen and Guo, 2001). Polysaccharides of *D. huoshanense* exhibit antiglycation and



GRAPHICAL ABSTRACT | Outline of contents in the study.

immunoregulatory activities (Pan et al., 2013; Zha et al., 2014), while its alkaloids are anticataract and neuroprotective (Ng et al., 2012). However, harsh growth habits and overexploitation resulted in *D. huoshanense* becoming endangered and its low yield does not meet its demand. The genome of *D. huoshanense* has recently been sequenced to promote molecular genetics and functional genomics studies (Han et al., 2020). Besides, several studies focusing on the transcriptome of *D. huoshanense* have shown differences in the main active ingredients across tissues (Yuan et al., 2018; Zhou et al., 2020). However, there had been no studies reporting effects in the proteome and protein posttranslational modifications (PTMs) of *D. huoshanense*.

Posttranslational modification, a crucial mechanism aiming at cellular regulation, is attributed to the proteolytic cleavage or addition of a specific group to one or more amino acids (Allfrey et al., 1964; Walsh et al., 2005). As an efficient strategy to modify and expand the properties of 20 natural amino acids by introducing the new groups, there is no denying that PTM plays a vital role in accurate protein folding, molecular structure, as well as physical or chemical properties (Millar et al., 2019). PTMs occur in any stage of a life cycle of a protein, entailing the increase of variance from a genome to proteome level (Doyle and Mamula, 2001). In fact, the vast majority of proteins in cells require modifications to perform their corresponding functions. The PTM procedure is precisely modulated, so that the protein shows a definite and stable or dynamic function in the cell at specific moments (Olsen and Mann, 2013). An increasing number of PTM types have attracted research attention, evolving from small chemical modifications, such as lysine acetylation (Kac) and phosphorylation, to intricate modifications, such as ubiquitylation (Huang et al., 2014). Xue et al. (2018) reported a comprehensive analysis of lysine acetylation and its various functions in rice. In rice, the lysine of histone H3 would be acetylated in response to cold stress stimuli. Kac can regulate biological processes (BPs) to protect plants growing normally. After that, succinylation was detected to be the most enriched modification in strawberry stigma, being involved in regulating AP2/clathrin-mediated vesicular transport (Fang et al., 2018). Another PTM, phosphorylation, was predicted in milk fat globule membrane in human colostrum and mature milk, with a key role in infant development (Yang et al., 2020). Meng et al. (2019) showed that N-glycoproteome proteins in duck egg white can have antibacterial effects. Altogether, many cellular BPs are known to be regulated by PTMs through their impact on signaling, gene expression, protein stability and interactions, and enzyme kinetics (Friso and van Wijk, 2015).

Besides the PTMs mentioned above, there is a newly discovered type, lysine crotonylation (Kcr), which is the addition of a crotonyl group to a lysine residue (**Figure 1A**). In 2011, Kcr was reported to generate a "histone code" conserved from yeast to humans, mainly enriched in active promoters or potential enhancers (Tan et al., 2011). Kcr takes place at the ε-amino group in lysine and can be catalyzed by Kac enzymes, such as P300 (Sabari et al., 2015). The sizes of the four carbons and planar orientation of Kcr render it distinguishable from acetylation (Baumann, 2015). Liu et al. (2018b) analyzed histone crotonylation of rice whole-genome DNA and found a positive correlation between the degree of protein modification and gene expression. Li et al. (2016) proposed that H3 can also upregulate gene expression in the AF9 YEATS domain. Based on evidence that histone crotonylation contributes to chromatin and gene expression regulation, Kcr was revealed to widely exist in non-histones, with other biological functions besides epigenetic regulation (Xu et al., 2017). From then on, crotonylation sites were detected in myofilament and ribosomal proteins in zebrafish embryos, the human lung adenocarcinoma H1299 cells, and cervical cancer HeLa cells and had an impact on many metabolic pathways (Wei et al., 2017; Xu et al., 2017; Kwon et al., 2018). There have been abundant studies of Kcr in animals and microorganisms. Plants can quickly change their habits to

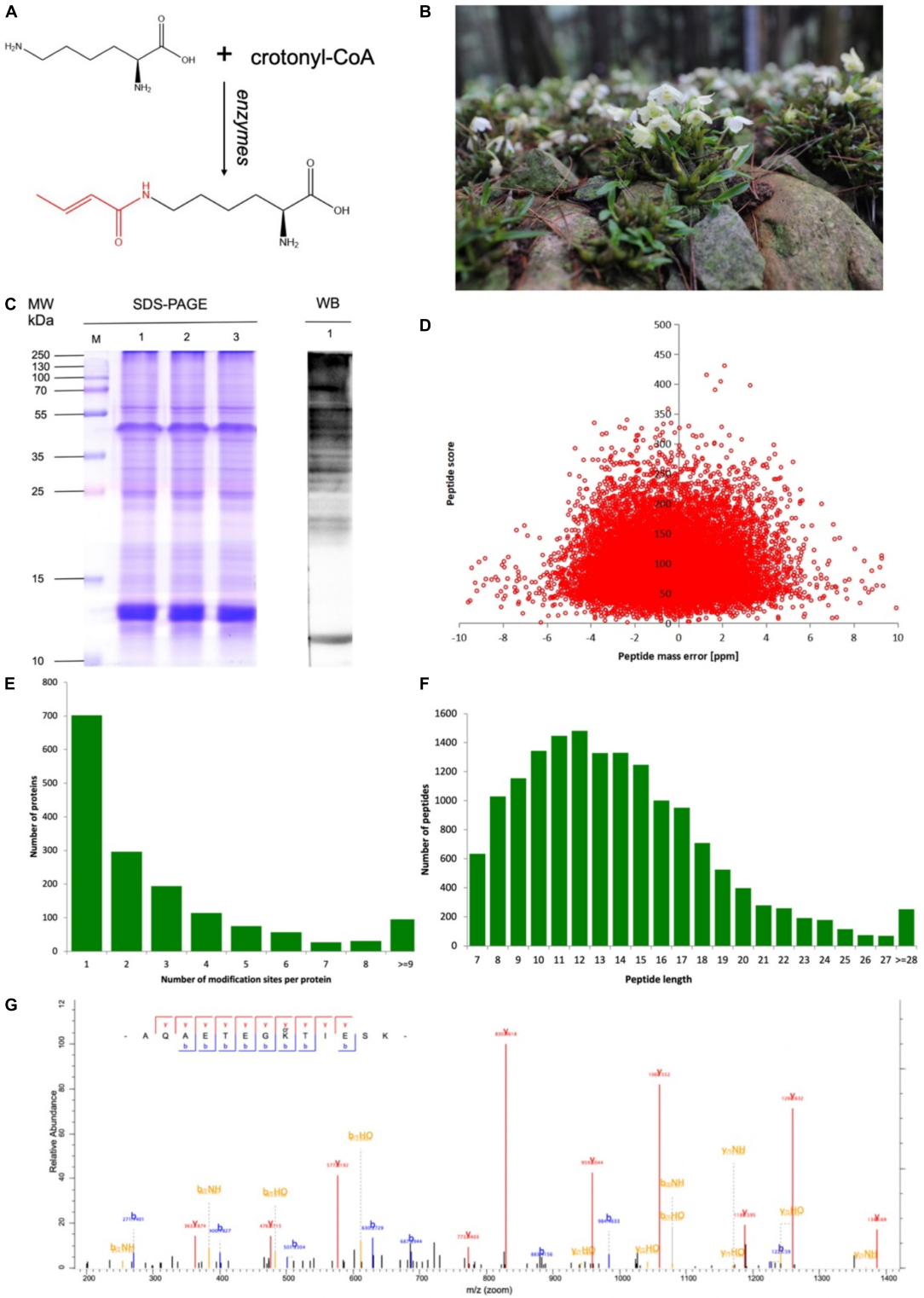


FIGURE 1 | Overview of lysine crotonylation in developing *Dendrobium huoshanense*. **(A)** Structure of lysine crotonylation. **(B)** The picture of an adult *D. huoshanense*. **(C)** Western blotting screening of *D. huoshanense* lysine crotonylome. Molecular weight is labeled on the left. Samples are labeled on the top. 1, 2, and 3 on the left are images of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie brilliant blue. 1 on the right is Western blot image. Same amounts of proteins (20 μg per lane) were loaded for samples. **(D)** The mass quality precision distributions of crotonylation profiles. **(E)** Distribution of lysine crotonylation sites in one protein. **(F)** The peptide length distributions of crotonylation profiles. **(G)** The mass spectrometry (MS/MS) crotonylated protein spectrum of 60S ribosomal protein L7a.

adapt to the changes in their ecological environment. PTM is a more rapid and effective strategy than gene transcription and translation (He et al., 2016). However, Kcr in plants had only been analyzed in rice (Liu et al., 2018b), tobacco (Sun et al., 2017), papaya (Liu et al., 2018a), tea plants (Sun et al., 2019), *Chrysanthemum* (Lin et al., 2021), and peanut (Xu et al., 2021) and was found to contribute to a range of processes, including photosynthesis, glycolysis, and amino acid synthesis.

As a plant with high edible and medicinal value, Kcr in *D. huoshanense* has not been investigated. Therefore, we explored the comprehensive Kcr proteome of *D. huoshanense* by using high-resolution liquid chromatography–mass spectrometry (LC-MS/MS) and high-sensitive immune-affinity purification. Conservative motifs, the Gene Ontology (GO) enrichment, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were also implemented to predict the modified proteins. Our research is the first report on qualitative proteome-wide analysis of Kcr profiling in *D. huoshanense* and compares crotonylated proteins and motifs with those in tea, rice, etc., to obtain orthologous proteins and conserved motifs. Several crotonylated proteins were found to be associated with active compound biosynthesis and response to stress. This offers not only a fundamental basis for exploring the physiological role of Kcr in *D. huoshanense* and PTMs in non-histones, but also a significant resource for future studies on the molecular mechanism of plant regulation by PTMs.

MATERIALS AND METHODS

Chemicals and Plant Materials

Plants originating from Huoshan County, Anhui Province, China were identified as *D. huoshanense* C. Z. Tang et S. J. Cheng by Professor Nianjun Yu from Anhui University of Traditional Chinese Medicine (Tang and Cheng, 1984). The whole plant of *D. huoshanense* consisted of roots, stems, leaves, and flowers (**Figure 1B**). Three independent 3-year-old whole plants had been collected for the samples, which were washed with ddH_2O for 30 min to eliminate potential microbes.

Protein Extraction

Three replicates of whole plants were fully grinded by liquid nitrogen into powders in precooled mortar, mixed, and transferred to a new centrifuge tube. Three volumes of phenolic extraction buffer [10 mM dithiothreitol (DTT), 1% protease inhibitor cocktail, 3 μM trichostatin A, together with 50 mM nicotinamide] were added to the centrifuge tube, followed by ultrasonication. Then, Tris-saturated phenol (pH 8.0) was added in drops with an equal volume into the centrifuge tube for centrifugation (4°C, 10 min, 5,000 *g*). Then, the upper phenol phase was removed to a new centrifuge tube for protein precipitation (at least five volumes of ammonium acetate-saturated methanol were added), followed by protein incubation at −20°C overnight. Centrifugation at 4°C for 10 min was carried out to discard the supernatant. Ice-cold methanol was added to polish the remaining pellet once, accompanied by ice-cold acetone three times. Last, the protein was redissolved in 8 M urea and the concentration was determined by using the bicinchoninic acid (BCA) kit by following the methods of the manufacturer.

Protein Trypsin Digestion

For digestion, an appropriate amount of standard protein and 20% trichloroacetic acid (TCA) was dropped slowly into the protein and mixed thoroughly. The supernatant was discarded after precipitation for 2 h (4°C) and centrifugation for 5 min (4,500 *g*). Dry precipitation was diluted by adding 200 mM triethylammonium bicarbonate (TEAB) to urea, to reach a concentration less than 2 M, which was reduced with 5 mM DTT for 30 min at 56°C and alkylated with 11 mM iodoacetamide (IAA) for 15 min at room temperature in the dark. Finally, trypsin was added at a ratio of 1:50 trypsin-to-protein for digestion overnight.

Western Blot Analysis

Phenol-extracted proteins (20 μg) from *D. huoshanense* were separated by using 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and electroblotted onto a polyvinylidene fluoride (PVDF) membrane for Western blot analysis. Crotonylated proteins were detected by primary anti-crotonyllysine antibody (PTM-502, Cat. No. 1037267K306, PTM Biolabs, Hangzhou, China) in a 1:1,000 (v/v) dilution and secondary antibody goat antimouse immunoglobulin G (IgG) (31430, Thermo Fisher Scientific, Waltham, MA, United States) in a 1:10,000 dilution.

High-Performance Liquid Chromatography Fractionation and Affinity Enrichment

Through reverse-phase high-performance liquid chromatography (HPLC) with high pH by using a Thermo Betasil C18 column (5 μm particles, 10 mm ID, and 250 mm length), the tryptic peptides were first separated with a gradient of 8–32% acetonitrile (pH 9.0) over 60 min into 60 fractions. Then, peptides were dried by vacuum centrifuging.

To enrich crotonylated peptides, trypsin digested peptides were dissolved in NETN buffer (100 mM NaCl,1 mM EDTA, 50 mM Tris–HCl, 0.5% NP-40, pH 8.0) and incubated with prewashed anti-crotonyl lysine antibodies (PTM-503, Cat. No. #TAK901B06, PTM Biolabs, Hangzhou, China) at 4°C overnight with gentle shaking. Then, beads were carefully washed four times with NETN buffer and twice with distilled H_2O. The bound peptides were eluted from the beads with 0.1% trifluoroacetic acid. Finally, the eluted fractions were combined and vacuum dried. The resulting peptides were desalted with C18 ZipTips (Millipore, Burlington, MA, United States) according to the instructions of the manufacturer and then subjected to LC-MS/MS.

Quantitative Proteomic Analysis by Liquid Chromatography–Mass Spectrometry

Followed by loading onto a reversed-phase analytical column (15 cm length, 75 μm id), the tryptic peptides were dissolved in

solvent A, comprised of 0.1% formic acid and 2% acetonitrile. The gradient was comprised of an increase from 7 to 24% solvent B (0.1% formic acid in 98% acetonitrile) over 42 min, 24–32% in 12 min and climbing to 80% in 3 min, and then holding at 80% for the last 3 min, all at a constant flow rate of 450 nl/min on an nanoElute Ultra Performance Liquid Chromatography (UPLC) system. The intact peptides and fragments were subjected to NSI source followed by the timsTOF Pro MS coupled online to the UPLC, whose electrospray voltage was set at 1.75 kV. The first mass m/z scan range was 100–1,700 for a full scan, in which intact peptides were detected in the Orbitrap at a resolution of 70,000. Peptides were then selected for MS/MS by using the NCE setting at 28 and the fragments were detected in the Orbitrap at a resolution of 17,500. The scanning range of secondary MS was set to 100–1,700 m/z by using a parallel accumulation serial fragmentation (PASEF) mode to acquire data. A data-dependent procedure that alternated between one MS scan followed by 20 MS/MS scans with 24 s dynamic exclusion was used. Automatic gain control (AGC) was set at 5E4 and 100 m/z was set for fixed first mass.

Mass-Spectrum Quality Control Analysis

The first-order mass error of most spectra was under 10 ppm, which was in line with the high-precision characteristics of Orbitrap MS. It showed that the qualitative and quantitative analysis of proteins would not be affected due to excessive mass deviation. The score of the matched peptides in the spectrum (indicating the credibility of peptide identification) had a negative correlation with the distribution of quality deviation (the higher the score, the smaller the quality deviation).

Database Search

MaxQuant (version 1.6.15.0) was accessed to process raw data files acquired by the mass spectrometer and tandem mass spectra data were searched against the transcriptome database[1] of *D. huoshanense* uploaded by us that is concatenated with the reverse decoy database and a common pollution database. With four missing cleavages, trypsin/P was used as the cleavage enzyme. For precursor ions, the mass tolerance was set to 20 and 4.5 ppm first and final stages of the search, respectively. As for fragment ions, the mass error was set to 10 ppm. Carbamidomethylation on cysteine was specified as a fixed modification, while crotonylation on lysine was specified as variable modification. False discovery rate (FDR) thresholds for proteins, peptides, and modification sites were set at 1%.

Protein Annotation and Enrichment Analysis

The GO, provided by the UniProt-GO annotation (GOA) database,[2] is a widespread group of evidence-based combinations between terms from the GO data and UniProtKB proteins (Dimmer et al., 2012). Converting identified protein IDs to UniProt IDs is the first step, followed by mapping to the GO ID. If not successful, the InterProScan software would be used to annotate the GO function of a protein based on a protein sequence alignment method. Molecular function (MF), cellular components (CCs), and BPs are three categories of the GO analysis.

The KEGG is a database to annotate the pathways in which proteins participate. All the identified proteins were mapped to a pathway in the KEGG database[3] by using Blastx/Blastp. First, the KEGG Automatic Annotation Server (KAAS) (KEGG online service tool) was used to annotate the KEGG database description of a protein (Moriya et al., 2007). Then, the KEGG mapper was used to map the annotation to the KEGG pathway database.

Protein domain functions were annotated by the InterProScan (a sequence analysis application) based on a protein sequence alignment method (Zdobnov and Apweiler, 2001). InterPro[4] database was used for integrating various information about protein families, domains, and functional sites.

We also used WoLF PSORT, a subcellular localization prediction software (Horton et al., 2007). As for functional enrichment of the GO/pathways/domains of the Kcr protein clusters, the two-tailed Fisher's exact test was employed against the National Center for Biotechnology Information (NCBI) non-redundant protein database. The GO/pathways/domains with an adjusted corrected p-value < 0.05 were considered to be significantly enriched (Huang et al., 2009).

Motif Analysis

To specify the sequence model proximal to the crotonylated residues, we examined the relative frequencies of amino acids 10 positions upstream and downstream of Kcr sites by MoMo (motif-x algorithm), modification motifs prediction software. All the protein sequences in the database were treated as background controls. The minimum number of occurrences was set to 20, p-value ≤ 0.0000001. The option "Emulate original motif-x" was selected and other parameters were set to default.

Protein–Protein Interaction Analysis

The interaction network of proteins can help us to clarify the functions and mechanism of Kcr proteins. Based on the STRING database version 10.5, all the differentially expressed modified protein database accessions or sequences in this study were searched for Protein–protein interactions (PPIs). The STRING treats confidence score as a metric to define whether the interaction is real. Only interactions with a confidence score >0.7 (high confidence) were visualized by using R package "networkD3" for a network of PPIs.

Conservative Analysis

To determine the degree of evolutionary conservation of crotonylation, BLASTP was applied for comparing crotonylated protein sequences of *D. huoshanense* against specific protein sequences from four species: *Dendranthema grandiforum* (PXD010297), *Camellia sinensis* (PXD011610), *Oryza sativa* (PXD008716), and *Nicotiana tabacum* (IPX0000889000). By

[1] https://pan.baidu.com/s/1bdA4oE_LjkhC41eePsv0pg, access code: 2290.

[2] http://www.ebi.ac.uk/GOA/

[3] https://www.genome.jp/kegg/genes.html

[4] http://www.ebi.ac.uk/interpro/

applying a reciprocal best BLAST hit approach, we determined orthologous proteins among these plants.

RESULTS

Proteome-Wide Identification of Lysine Crotonylation Sites in *Dendrobium huoshanense*

To detect the global distribution of Kcr in *D. huoshanense*, the trypsin-hydrolyzed protein was identified by a combination of high-specific Kcr pan antibody and high-resolution LC-MS/MS. The pan anti-crotonylated antibody helped to detect proteins modified by using SDS-PAGE and Western blot (**Figure 1C**). Here, 4,726 crotonylation sites derived from 1,591 proteins were quantifiable with high confidence (**Supplementary Table 1**). A total of 13,765 peptides were identified by trypsin digestion, among which, 4,695 peptides were crotonylated. Original data of all the identified crotonylated peptides are shown in **Supplementary Table 1**. The MS proteomics data have been deposited into the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD028293.[5] The MS data were validated by referring to the mass error of all the identified peptides; most of them had a less than 5 ppm mass error (**Figure 1D**). With a diverse distribution of crotonylation sites in *D. huoshanense* proteins, 702 were crotonylated on one site, 296 were crotonylated on two sites, and 153 proteins contained 7 or more crotonylation sites (**Figure 1E**). Consistent with the properties of tryptic peptides, the lengths of most identified peptides varied from 7 to 28 amino acid residues (**Figure 1F**). In the study of 69,323 spectra generated from the mass spectrometer, 4,341 matched with protein alignments. A representative mass spectrum of a crotonylated protein [60S ribosomal protein L7a (RPL 7a)] is shown in **Figure 1G**.

Functional Annotation and Subcellular Localization of Lysine Crotonylated Proteins

The crotonylated proteins identified were classified based on the GO annotation (**Figure 2A**, **Supplementary Figure 1**, and **Supplementary Table 2**). Metabolic processes are essential life activities in plants, whose products are important active elements for growth. Just as the importance of polysaccharides and alkaloids in *D. huoshanense*. Therefore, the top "biological process" of the GO terms was metabolic process. For example, the number of proteins in cellular metabolic processes, organic substance metabolic processes, and primary metabolic processes was up to 504, 482, and 423, respectively. In the category of "molecular function," the majority of crotonylated proteins were involved in "binding," which mainly consisted of "organic cyclic compound binding" (205 proteins), "heterocyclic compound binding" (199 proteins), and "ion binding" (165 proteins). Under the "cellular components" category, "cytoplasm" and "organelle" were predicted to be the most relevant GO terms in modified proteins, 909 and 831 proteins in each, respectively. The analysis of subcellular localization clarified that a majority of crotonylated proteins were mainly located in the chloroplast (40.29%), cytoplasm (28.41%), and nucleus (13.51%), suggesting that crotonylated proteins were broadly distributed in leaves and stems of *D. huoshanense* (**Supplementary Table 2**). Moreover, 5.66% modified proteins were mitochondria located. Among crotonylated proteins, 3.9 and 2.26% modified proteins were predicted to be localized in the plasma membrane and extracellular space, respectively (**Figure 2B**). We also analyzed function classifications of proteins corresponding to crotonylated sites based on the Clusters of Orthologous Groups (COGs) the COG database. A total of 272 proteins were predicted to have the function of "postmodification, protein turnover, chaperones." "Translation, ribosomal structure, and biogenesis" possessed the second highest amount of identified proteins (145), while 138 modified proteins were predicted to be associated to "energy production and conversion" (**Figure 2C** and **Supplementary Table 2**).

Functional Enrichment Analysis of Lysine Crotonylated Proteins

To explore the roles that crotonylated proteins play, the functional enrichment analysis of identified proteins was operated based on the GO, the KEGG pathways, and protein domain annotations (**Supplementary Table 3**). Regarding the "biological process" category of the GO enrichment, the most well-represented terms were "monocarboxylic acid metabolic process," "starch metabolic process," and "nicotinamide nucleotide metabolic process." In the "cellular components," crotonylated proteins related to "plastid," "chloroplast stroma," and "plastid stroma" accounted for the largest portion. Based on the GO annotation, proteins with a "binding" function occupied a large portion, which included "copper ion binding," "metal ion binding," and "cation binding." There were also many proteins predicted to be enriched in "transition metal ion binding" and "oxidoreductase activity" (**Figure 3A**). The KEGG pathways enrichment analysis showed "carbon fixation in photosynthetic organisms," "proteasome," "pentose phosphate pathway," "photosynthesis," and "glyoxylate and dicarboxylate metabolism" as having highest enrichment score (**Figure 3B**). As it is verified in protein domain enrichment analysis, "proteasome subunit," "alcohol dehydrogenase GroES-like domain," and "thioredoxin" were predicted to be significantly enriched among modified proteins (**Figure 3C**).

Motifs and Secondary Structures of Lysine Crotonylated Peptides

To identify whether specific sequence motifs exist proximal to the K residues, we examined the relative frequencies of amino acids centered around each K residue (**Supplementary Table 4**). By using Motif X, we identified ten motifs flanking the K residues from unique Kcr-containing sites. The 10 consensus sequence motifs were KcrE, K. . .E. . .Kcr, AKcr, KcrD, KcrA, K. . .Kcr, KcrV, KcrG, VKcr, Kcr. . .E, and Ykcr, among which, the KcrE motif was present in the largest number of crotonylated

[5]http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD028293

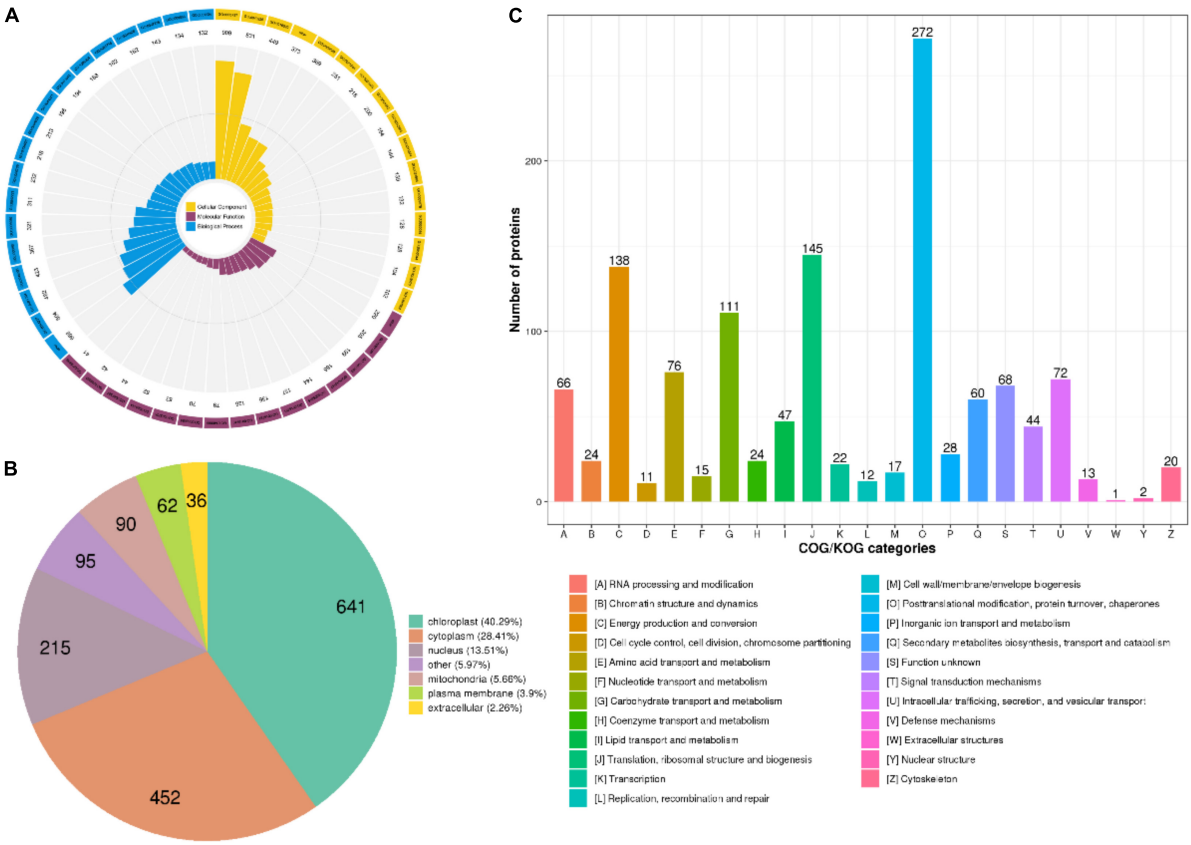


FIGURE 2 | Classification of proteins in *D. huoshanense* corresponding to identified sites. **(A)** The Gene Ontology (GO) classification of the crotonylated proteins in *D. huoshanense* based on biology process, cellular components, and molecular functional. **(B)** Subcellular localization chart of proteins corresponding to modification sites. **(C)** COG/KOG functional classification of crotonylated proteins.

peptides (555 peptides), whereas the YKcr motif was the least abundant in crotonylated peptides (80 peptides) (**Figure 4A**). Kcr indicates the crotonylated lysine, while E, A, D, V, G, and Y indicate glutamic acid, alanine, aspartic acid, valine, glycine, and tyrosine, respectively. These motifs are likely to represent a sequence preference for crotonylation in *D. huoshanense*. Hence, the first six motifs that were highly scored were selected for bioinformatic analysis (**Figure 4B**). Specifically, the GO, the KEGG, and domain function enrichment analyses on the preference of the first six significant motifs were performed. Remarkably, different motifs preferred to harboring diverse functions, as reflected on the various GO categories, pathways, and protein domains (**Supplementary Figures 2–6**). KcrE and KcrD were both distributed to an ATP-dependent peptidase activity (MF category), while glycogen catabolism (BP category) and apoplast and extracellular region (CC category) tended to harbor KcrD and KcrA motifs (**Supplementary Figures 2–4**). According to the KEGG enrichment, the first six motifs belonged to carbon fixation in photosynthetic organisms, but KcrD was enriched in protein processing in the endoplasmic reticulum (**Supplementary Figure 5**). Some proteins had various protein domains, illustrated by peptidase inhibitor I9 showing AKcr and K. . .Kcr motifs and core histone H2A/H2B/H3/H4 showing K. . .E. . .Kcr and KcrD motifs (**Supplementary Figure 6**).

A hierarchical cluster analysis of the frequencies of various residues related to Kcr was also performed on these motifs. As shown in the heat map (**Figure 4C**), the enrichment of A and E residues was mainly observed in the −5 to +5 positions and residue D was markedly enriched in the −3 to +3 position. While the K residue was abundant in the −10 to −5 and +5 to +10 positions, the C residue was not observed. We found that the occurrence of KcrE, K. . .E. . .Kcr, AKcr, and KcrD was comparatively higher than that of other conserved motifs. In accordance with the discovery, it can be inferred that crotonylation in *D. huoshanense* was preferred on K residues adjacent to glutamic acid, alanine, and aspartic acid.

Proteins are composed of linear sequences of amino acids, but only acquire their corresponding activity and biological function when folded into a specific spatial conformation. Thus, understanding the distribution of Kcr in the secondary structure of proteins is helpful to clarify the roles of crotonylated proteins in *D huoshanense*. The structural analysis of crotonylated proteins is shown in **Figure 4D**. Kcr sites preferentially occurred in disordered coils (66.5%), while only approximately 27.07% of sites occurred in α-helices and 6.42% in β-strands (**Figure 4D**). They also tended to occur more frequently on the protein surface, with a high relative surface accessibility of 94% (**Figure 4E**). Therefore, Kcr likely does not affect the surface properties of

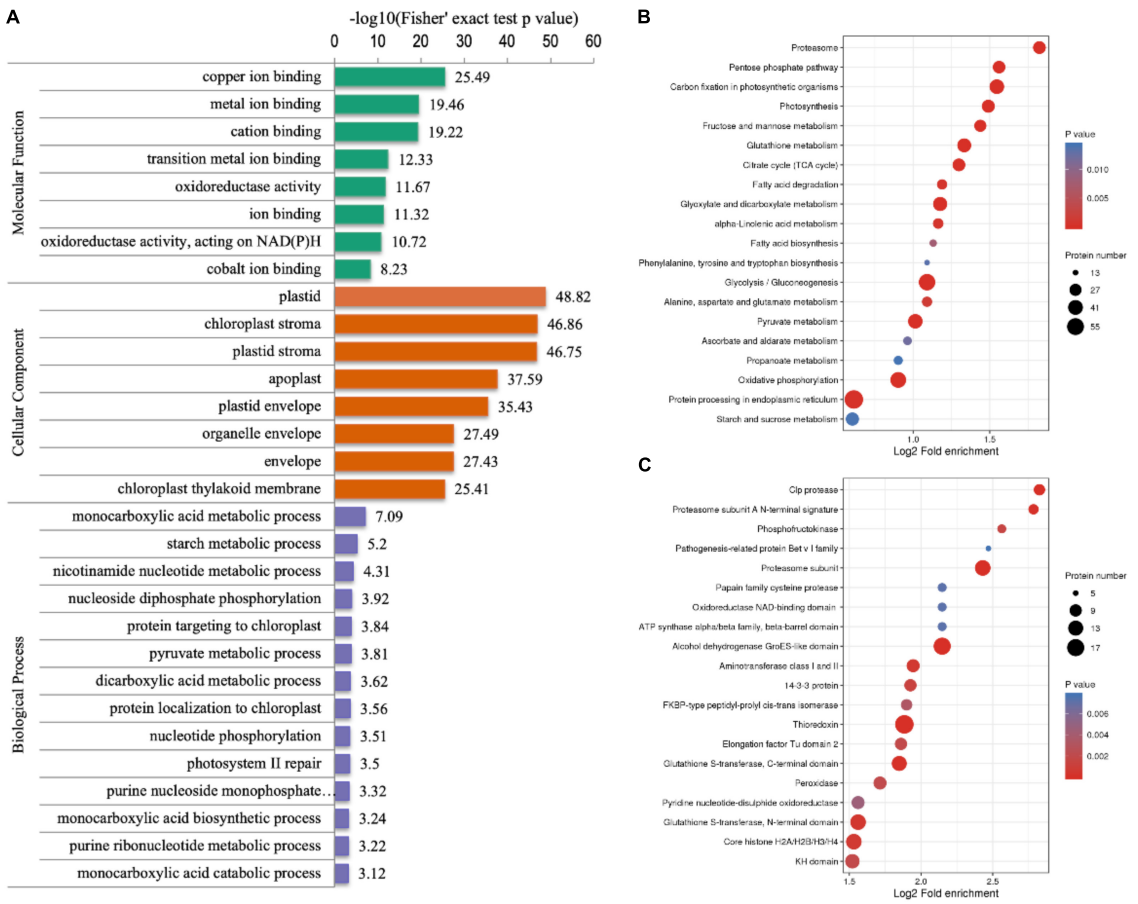


FIGURE 3 | The GO classification, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and protein domain analyses of crotonylated proteins of *D. huoshanense*. The *p*-value obtained by the Fisher's exact test showed the functional classification and pathway of enriched proteins. The results of the first 20 classifications most enriched were given in the bubble plot. In the bubble plot, the vertical axis is the functional classification or pathway and the horizontal axis value is the Log2 conversion value of the proportion of modified proteins in this functional type compared with the proportion of identified proteins. The circle color indicates enrichment significance and the circle size indicates the number of modified proteins in the functional class or pathway. **(A)** The GO enrichment bubble plot of proteins corresponding to modification sites in three categories: biological process, molecular function, and cellular component. **(B)** The KEGG pathway enrichment of proteins corresponding to modification sites. **(C)** Protein domain enrichment bubble plot.

modified proteins. These results imply that crotonylation could selectively modify K residues in proteins and change their structure and activity indirectly, which may affect their functions in *D. huoshanense*.

Protein Interaction Network of the Crotonylated Proteins in *Dendrobium huoshanense*

We generated protein interaction networks of all the identified proteins by using the PPI database to clarify their function. Based on this network, crotonylated proteins associated with various interacting pathways were characterized and a large subinteraction network was further constructed and visualized by Cytoscape. According to the results of the above KEGG pathway enriched analysis, we constructed an interaction network composed of crotonylated proteins originating in three highly scored enrichment pathways, including carbon fixation in photosynthetic organisms, the proteome, and pentose phosphate pathway (**Supplementary Table 5**; **Figures 5A–C**). Proteins were detected to interact with various neighbors in these networks. All the proteins in **Figure 5** are annotated in **Supplementary Table 5**.

Given the lack of functional information of interactions above, further research should be carried out to verify potential interactions. More bioinformatic analyses may benefit to exploring the roles of Kcr in *D. huoshanense*. Furthermore, they will be useful in selecting key proteins and their possible mechanisms.

DISCUSSION

Comparison of Lysine Crotonylation Among *Dendrobium huoshanense* and Other Plants

Although great progress has been made in the study of crotonylation in microorganisms (Li et al., 2016;

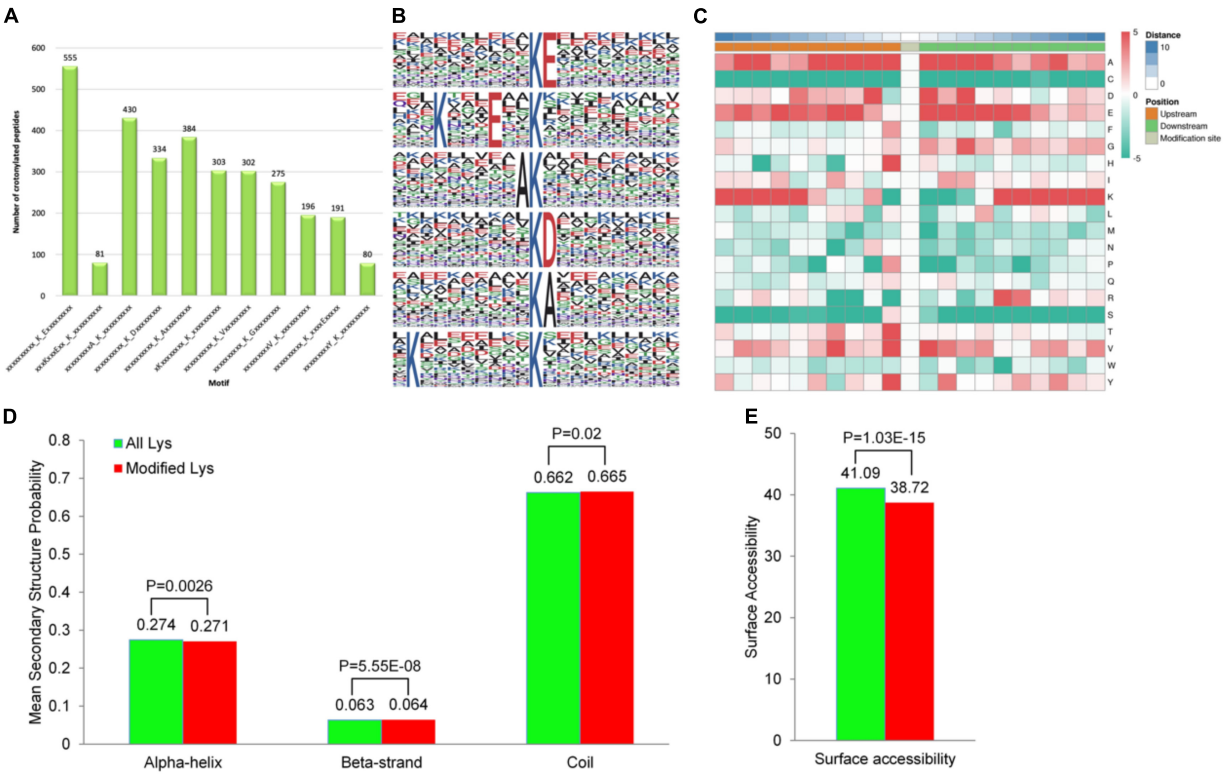


FIGURE 4 | Properties of the lysine crotonylation sites in *D. huoshanense*. **(A)** The peptides length distribution in 10 consensus sequence motifs that were significantly enriched in K residues from unique Kcr-containing sites in all. **(B)** Sequence probability logos of the first six enriched crotonylation site motifs around the lysine crotonylation sites. **(C)** Heat map of the amino acid compositions around the Kcr sites showing the frequency of different types of amino acids around this residue. Red indicates enrichment and green indicates depletion. **(D)** Probabilities of lysine crotonylation in different protein secondary structures (alpha helix, beta-strand, and disordered coil). **(E)** Predicted surface accessibility of crotonylation sites.

Yang et al., 2018), animals (Ruiz-Andres et al., 2016), and humans (Kelly et al., 2018; Wan et al., 2019), Kcr research in plants had been in its infancy. However, there was evidence showing that crotonylation possesses a certain association to plant growth and development. With the innovation of high-resolution MS and highly specific antibodies, research on plant crotonylation has been greatly promoted, such as in peanut (Xu et al., 2021), papaya (Liu et al., 2018a), *Chrysanthemum* (Lin et al., 2021), tea (Sun et al., 2019), rice (Liu et al., 2018b), and tobacco (Sun et al., 2017; **Table 1**). However, there were no reports for crotonylation in *D. huoshanense*. This study identified 4,726 Kcr sites in 1,591 modified proteins that were involved in diverse growth activities. Crotonylated proteins and sites in different plants lead to different consequences. Peanut (*Arachis hypogaea*) showed 6,051 sites in 2,580 proteins (the largest in amount), while the minimum number of proteins (637) containing 2,044 sites was identified in tobacco. On the contrary, what all the plants had in common was their first three categories of subcellular localization. Crotonylated proteins in peanut, papaya, *Chrysanthemum*, tea, rice, tobacco, as well as in *D. huoshanense*, according to ratio order of location, from high to low, were all located to chloroplast, cytosol, followed by nuclear. It was worth noting that crotonylated proteins in the seven species had an extensive abundance in chloroplasts, which is characteristic in plants. Thus, we have considerable reasons to conclude that crotonylation is conserved in the same locations of different plant species.

Crotonylated protein sequences of *D. huoshanense* were compared with specific protein sequences from tea, *Chrysanthemum*, rice, and tobacco. With a reciprocal best BLAST hit approach, we determined the orthologous proteins modified among these species. Consequently, *D. huoshanense* had 774, 728, 515, and 496 orthologous crotonylated proteins with tea, *Chrysanthemum*, rice, and tobacco, respectively (**Figures 6A–D** and **Supplementary Table 6**). We also analyzed conserved motifs identified in *D. huoshanense* with those in other plants, finding a large number of conserved motifs (**Figure 6E** and **Supplementary Table 6**). It is worth noting that KcrD and KcrE were generally conserved in these five plants. It can be inferred that crotonylation was preferred on K residues adjacent to aspartic acid and glutamic acid in plants.

Heavily-Crotonylated Proteins in *Dendrobium huoshanense*

Among the 1,591 proteins identified in *D. huoshanense*, heat shock protein 70 (Hsp70) displayed the most abundant crotonylated sites (25 sites). The second most heavily

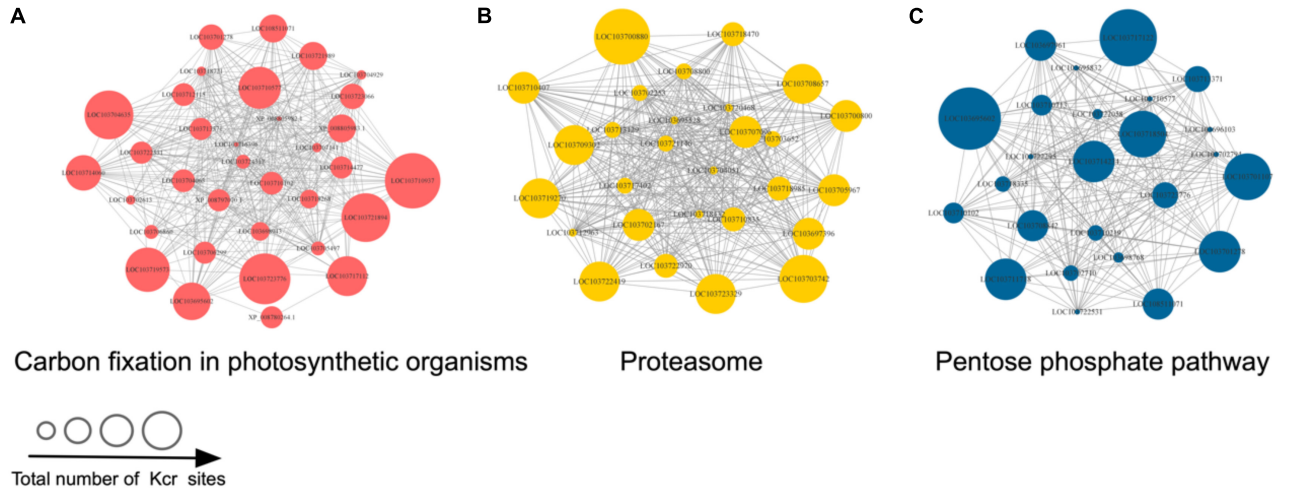


FIGURE 5 | The three KEGG pathways with the highest score in the protein–protein interaction (PPI) networks of Kcr proteins in *D. huoshanense*. The network of Kcr protein interactions was visualized by using Cytoscape (listed with protein ID names). **(A)** Carbon fixation in photosynthetic organisms. **(B)** Proteasome. **(C)** Pentose phosphate pathway. The size of the dots represents the number of Kcr sites in each network.

crotonylated protein was chaperone protein ClpB3 (21 sites), followed by elongation factor 2 (EF2) and ribulose-1,5-bisphosphate carboxylase (rbcl), both with 20 sites. Their MS/MS counts (representing the abundance of crotonylation) were 30, 25, 23, and 47, respectively.

Plants are usually exposed to various abiotic stresses, such as extreme temperatures, drought, heavy metals, and high salinity. These can bring about the imbalance of cellular homeostasis associated with a series of pathological processes (Zhu, 2002). With a disturb to plant growth and gene potential, these stresses were treated as the most challengeable problem in agriculture, which can further entail a loss of grain yield at more than 50% (Boyer, 1982; Parry and Hawkesford, 2012). It was surprising that crotonylated proteins had been proved to reduce the accumulation of reactive oxygen species (ROS) under low temperature stress and improve cold tolerance in *Chrysanthemum* (Huang et al., 2021). Hsp70, a highly conserved protein, had been reported to play a vital role in maintaining cellular homeostasis and protecting living organism against stress, such as through viral suppression (Vega et al., 2006). It is a key molecular and biochemical indicator that can be used to evaluate physiological responses to stresses (Sánchez et al., 2001). Working as a molecular chaperone and stress tolerance response protein, it assists the refolding of unfolded proteins after partial denaturation to keep protein homeostasis (Lindquist, 1986). In addition, it had also been reported that GRPs may positively enhance germination and seedling growth by regulating stomata movement when under heat and drought stresses in *Arabidopsis* (Kim et al., 2008; Koussevitzky et al., 2008). In this study, GRP-2 was shown to be crotonylated in two sites. Gamma-aminobutyric acid (GABA), with seven modified sites, had been reported to play a regulatory role in the diurnal rhythm of drought-tolerant *N. tabacum* (Pelvan et al., 2021).

When exposed to drought stress, *D. huoshanense* seedlings were often compensated by increased levels of enzymes related to adversity resistance and stress response processes. Monodehydroascorbate reductase (MDHAR), catalase (CAT), glutathione peroxidase (GPX), glutathione S-transferase (GST), superoxide dismutase (SOD), glycine-rich RNA-binding protein (GRP), ascorbate peroxidase (APX), and DHAR are ROS-scavenging enzymes that can deal with drought stress (e.g., protein damage and degradation, including ROS damage) (Koh et al., 2015). MDHAR, CAT, GPX, GST, SOD, GRP, APX, and DHAR were identified to have more than 2 Kcr sites in *D. huoshanense*.

Based on conserved protein analysis (**Figures 6A–D**), APX, CAT, GPX, GRP2, and GST were conserved in tea, *Chrysanthemum*, rice, and tobacco, implying that all the five enzymes were extensively crotonylated in plants. Crotonylation had been shown to positively regulate APX activity under abiotic stress, further reduce the oxidative damage caused by it Lin et al. (2021). Therefore, the crotonylation of proteins may modulate ROS stress. The decreased activities of enzymes caused by drought stress (e.g., protein damage and degradation, including ROS damage) could be compensated by

TABLE 1 | The number of crotonylated proteins and sites identified of *Dendrobium huoshanense* compared with previous studies of plants.

Species	Tissues	Kcr proteins	Kcr sites	References
Arachis hypogaea	Leaves	2,580	6,051	Xu et al., 2021
Carica papaya	Fruits	2,120	5,995	Liu et al., 2018a
Dendranthema grandiforum	Leaves	1,199	2,017	Lin et al., 2021
Camellia sinensis	Leaves	971	2,288	Sun et al., 2019
Oryza sativa	Leaves	690	1,265	Liu et al., 2018b
Nicotiana tabacum	Leaves	637	2,044	Sun et al., 2017
Dendrobium huoshanense	Whole plants	1,591	4,726	

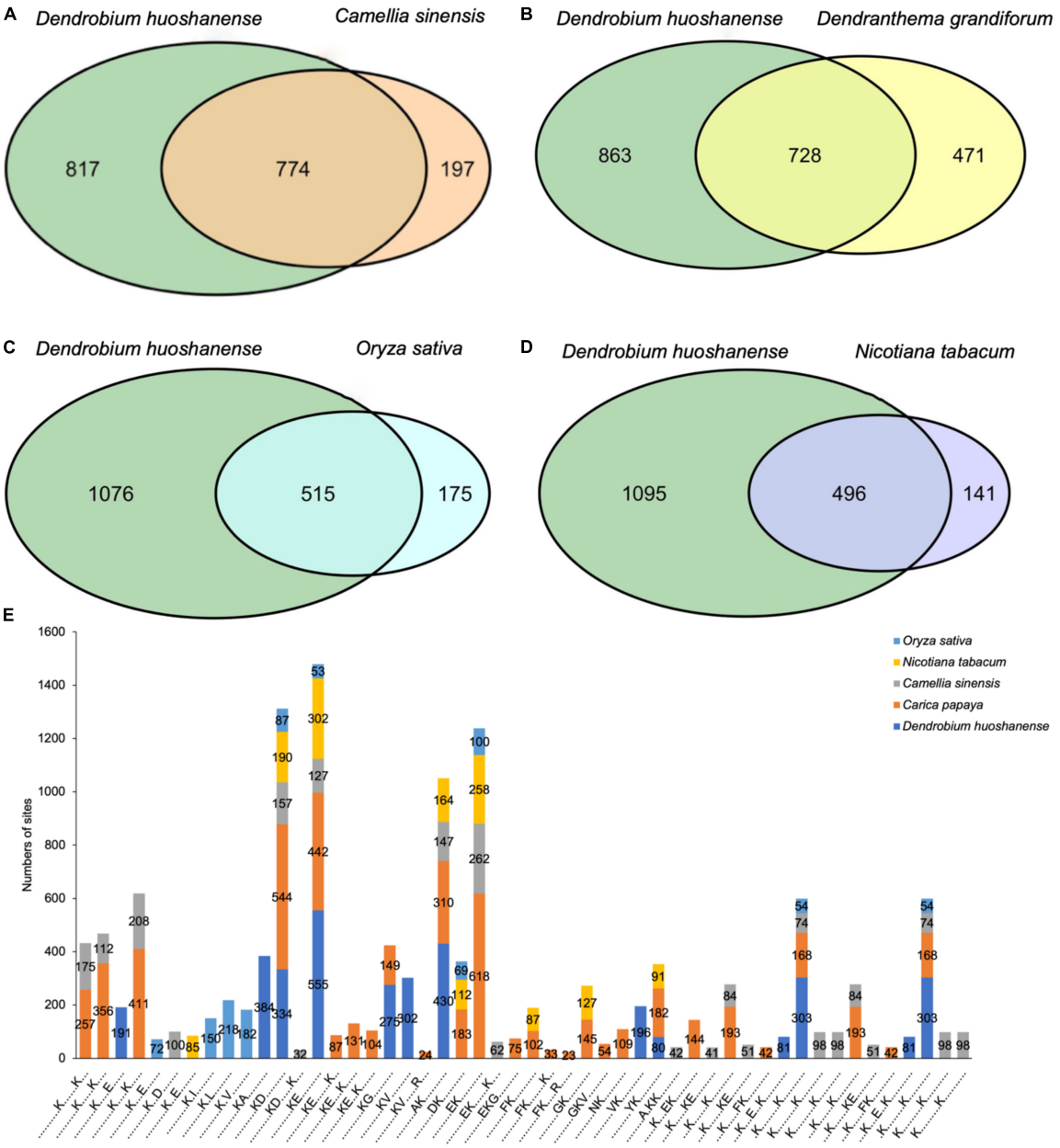


FIGURE 6 | Venn diagrams of the orthologous crotonylated proteins between *D. huoshanense* and tea **(A)**, *Chrysanthemum* **(B)**, rice **(C)**, and tobacco **(D)**. **(E)** Stacked histogram of crotonylation motifs of *D. huoshanense* compared with other species.

increased levels of crotonylated enzymes in *D. huoshanense*. This deduction contributed to further analyzing the regulatory role of crotonylation in plants under environmental stress.

Crotonylated Proteins Are Involved in the Calvin Cycle, Photosynthesis, and Alkaloid and Polysaccharide Biosynthesis Metabolic Pathways

During the growth period, plants synthesize a series of physiological metabolites, which accumulate to form the material basis to allow normal physiological processes to occur (Dyson et al., 2015). In *D. huoshanense*, photosynthesis metabolites and alkaloids are dominant bioactive metabolites supporting its development and medical efficacy (Li et al., 2021; Shang et al., 2021). A large proportion of Kcr proteins were predicted to be found in leaf chloroplasts. Furthermore, carbon fixation in photosynthetic organisms and photosynthesis were two of the most enriched pathways. Polysaccharides are transported from leaves to whole plants, of which bulk accumulated in carbon metabolism of chloroplast (Rennie and Turgeon, 2009; Knoblauch and Peters, 2010).

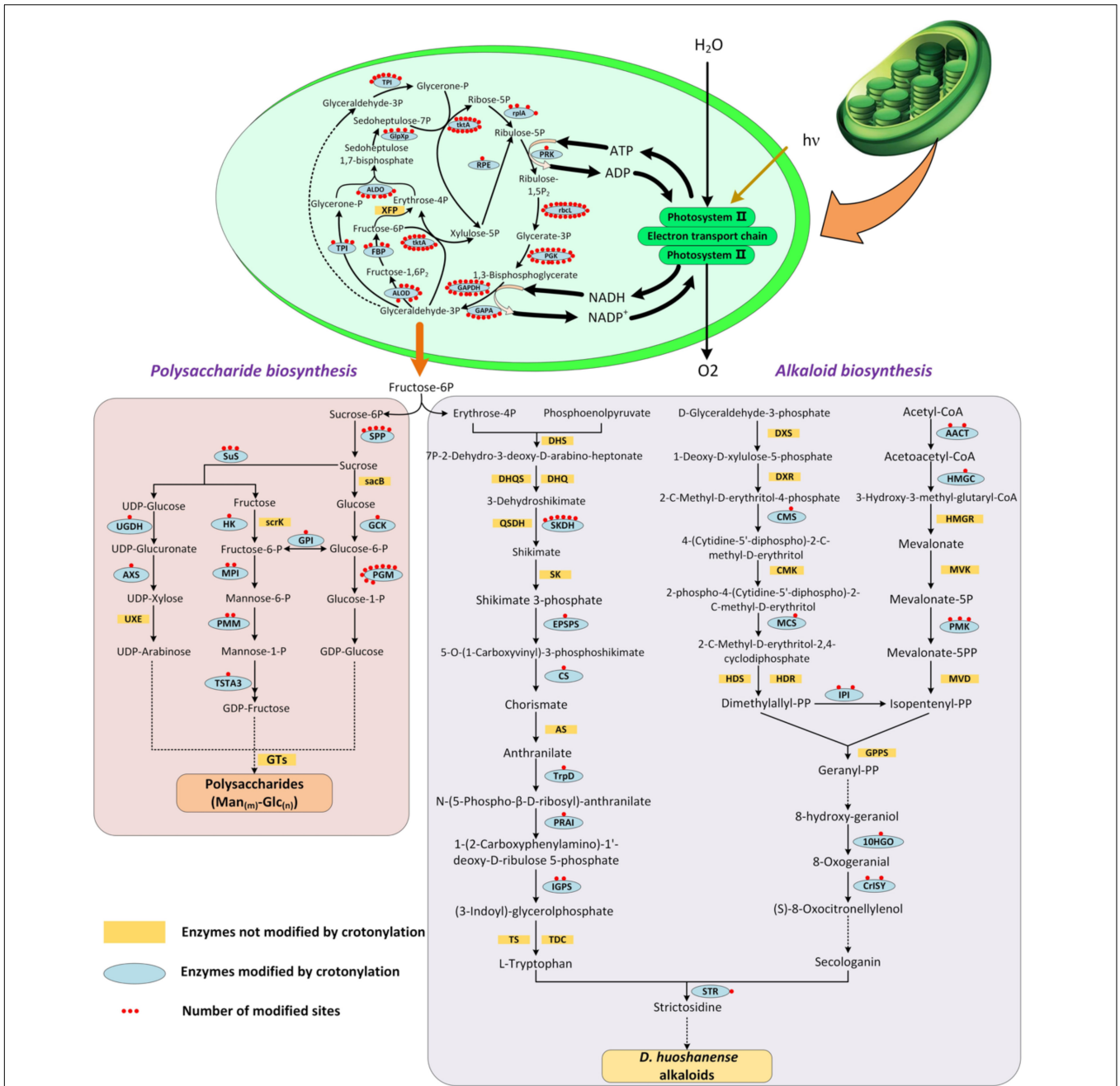


FIGURE 7 | Crotonylated enzymes were involved in Calvin cycle, polysaccharide biosynthesis, and alkaloid biosynthesis pathways of *D. huoshanense*. Yellow box referred to enzymes not modified by crotonylation. Blue oval referred to enzymes modified by crotonylation. Red dots indicated crotonylated sites in each enzyme.

Photosynthesis aims at converting solar energy into energy that can be absorbed by plants in the terms of ATP, sugars, and alkaloids (Gest, 2002). Thus, to shed light on the roles that crotonylation plays in main compound synthesis during photosynthesis, we focused on proteins related to the Calvin cycle, photosynthesis, and alkaloid and polysaccharide biosynthesis metabolic pathways (**Figure 7**). Surprisingly, an enormous number of proteins associated with these processes were modified with more than one site, including chloroplast-encoded large subunits (rbcl, 20 sites) and fructose-1,6-bisphosphatase (FBP) (4 sites). Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) fixes CO_2 from the atmosphere into the biosphere as the rate-limiting enzyme in the Calvin cycle forming carbohydrates (Field et al., 1998; Spreitzer and Salvucci, 2002). Rbcl is a subunit of Rubisco, encoded by the chloroplast genome (Lundqvist and Schneider, 1989). Notably, rbcl owned the most abundant counts of MS/MS among all the proteins crotonylated in this manuscript. It

has been described that the more rbcl is acetylated, the less active Rubisco is (Gao et al., 2016). This negative regulation to Rubisco indicated that CO_2 fixation efficiency can be improved through a lysine modification of rbcl (crotonylation included). FBP was detected to own four sites, being another crucial enzyme of the Calvin cycle and contributing to generating ribulose 1,5-bisphosphate (RuBP) (Serrato et al., 2009). FBP catalyzed fructose-1,6-bisphosphate (F1,6BP) into fructose-6-phosphate (F6P), which is used for synthesizing soluble sugars and starch in chloroplasts and the cytosol (Zimmermann et al., 1976). The deficiency of FBP entailed the impairment of plenty of physiological processes, such as photosynthesis and CO_2 fixation, as revealed in *Arabidopsis thaliana* (Rojas-González et al., 2015). There were also many other enzymes tightly linked to alkaloid and polysaccharide biosynthesis, such as phosphoglycerate kinase (PGK), phosphoglucomutase (PGM), phosphomannose isomerase (PMM), and shikimate dehydrogenase (SKDH) (1, 18, 2, and 6 sites, respectively). PMM, with two sites, was involved in mannose biosynthesis, which will be further explored by our next study concerning key enzymes in *D. huoshanense* polysaccharide biosynthesis. Protein crotonylation in *D. huoshanense* plays an important role in modulating the accumulation of main active ingredients. This role may be exerted through the regulation of enzymes mediating carbon fixation and photosynthetic efficiency. *D. huoshanense* is a precious, but endangered traditional Chinese herb and it is unfortunate that the mechanisms responsible for its synthesis of active compounds are poorly understood. Here, we identified several crotonylated enzymes involved in the biosynthesis of two dominant substances, polysaccharides and alkaloids, which were also involved in photosynthesis. These two compounds were promising for medicinal development due to their wide pharmacological activities. The candidate protein information was provided not only for protein structure analysis and functional verification, but also as the basis for optimizing the breeding quality of *D. huoshanense* and the optimization of germplasm resources.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: the mass spectrometry proteomics dataset has been deposited into the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD028293 (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD028293).

AUTHOR CONTRIBUTIONS

SX, JW, and DP conceived, designed, and implemented the study. JW and ZW contributed to the statistics analysis. JW, SX, DP, NY, FM, XY, and JZ contributed to the reagents, materials, and analysis tools. SX, JW, XM, and WJ drafted the manuscript. All authors edited the manuscript and approved the final version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. U19A2009), the University Synergy Innovation Program of Anhui Province (Grant No. GXXT-2019-043), the NSF of Anhui Province (Grant No. 1908085MH268), the Key Natural Science Research Projects in Anhui Universities (Grant Nos. KJ2019A0453 and KJ2018A0275), the Returnee Program of Anhui People's Society Office (Grant No. DT18100035), and the Foundation of Hunan Key Laboratory for Conservation and Utilization of Biological Resources in the Nanyue Mountainous Region (Grant No. NY20K04).

ACKNOWLEDGMENTS

We would like to thank Qian Shen from Shanghai Jiao Tong University for helping to revise the manuscript. We also thank Zhang Zhang from Jingjie PTM BioLab (Hangzhou) Corporation Ltd. in China for MS analysis and some bioinformatics of proteome.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.822374/full#supplementary-material

REFERENCES

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 51, 86–94. doi: 10.1073/pnas.51.5.786

Bao, X. S., Shun, Q. S., and Chen, L. Z. (2001). *The Medicinal Plants of Dendrobium (Shi-Hu) in China, A Coloured Atlas*. Shanghai: Press of Fudan University and Press of Shanghai Medical University.

Baumann, K. (2015). Post-translational modifications: crotonylation versus acetylation. *Nat. Rev. Mol. Cell. Biol.* 16:265. doi: 10.1038/nrm3992

Boyer, J. S. (1982). Plant productivity and environment. *Science* 218, 443–448. doi: 10.1126/science.218.4571.443

Chen, X. M., and Guo, S. X. (2001). Advances in the research of constituents and pharmacology of *Dendrobium*. *Nat. Prod. Res. Dev.* 1, 70–75. doi: 10.16333/j.1001-6880.2001.01.020

Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M. J., et al. (2012). The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* 40, D565–D570. doi: 10.1093/nar/gkr1048

Doyle, H. A., and Mamula, M. J. (2001). Post-translational protein modifications in antigen recognition and autoimmunity. *Trends Immunol.* 22, 443–449. doi: 10.1016/s1471-4906(01)01976-7

Dyson, B. C., Allwood, J. W., Feil, R., Xu, Y., Miller, M., Bowsher, C. G., et al. (2015). Acclimation of metabolism to light in *Arabidopsis thaliana*: the glucose

6-phosphate/phosphate translocator GPT2 directs metabolic acclimation. *Plant Cell Environ.* 38, 1404–1417. doi: 10.1111/pce.12495

Fang, X., Xin, Y., Sheng, Z., Liu, H., Jiang, A., Wang, F., et al. (2018). Systematic identification and analysis of lysine succinylation in strawberry stigmata. *J. Agric. Food Chem.* 66, 13310–13320. doi: 10.1021/acs.jafc.8b02708

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Friso, G., and van Wijk, K. J. (2015). Posttranslational protein modifications in plant metabolism. *Plant Physiol.* 169, 1469–1487. doi: 10.1104/pp.15.01378

Gao, X., Hong, H., Li, W. C., Yang, L., Huang, J., Xiao, Y. L., et al. (2016). Downregulation of Rubisco activity by non-enzymatic acetylation of RbcL. *Mol. Plant* 9, 1018–1027. doi: 10.1016/j.molp.2016.03.012

Gest, H. (2002). History of the word photosynthesis and evolution of its definition. *Photosynth. Res.* 73, 7–10. doi: 10.1023/A:1020419417954

Han, B., Jing, Y., Dai, J., Zheng, T., Gu, F., Zhao, Q., et al. (2020). A chromosome-level genome assembly of *Dendrobium huoshanense* using long reads and Hi-C data. *Genome Biol. Evol.* 12, 2486–2490. doi: 10.1093/gbe/evaa215

He, D., Wang, Q., Li, M., Damaris, R. N., Yi, X., Cheng, Z., et al. (2016). Global proteome analyses of lysine acetylation and succinylation reveal the widespread involvement of both modification in metabolism in the embryo of germinating Rice Seed. *J. Proteome Res.* 15, 879–890. doi: 10.1021/acs.jproteome.5b00805

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D., and Zhao, Y. (2014). SnapShot: histone modifications. *Cell* 159, 458–458. doi: 10.1016/j.cell.2014.09.037

Huang, Q. X., Liao, X. Q., Yang, X. H., Luo, Y. C., Lin, P., Zeng, Q. H., et al. (2021). Lysine crotonylation of DgTIL1 at K72 modulates cold tolerance by enhancing DgnsLTP stability in chrysanthemum. *Plant Biotechnol. J.* 19, 1125–1140. doi: 10.1111/pbi.13533

Kelly, R. D. W., Chandru, A., Watson, P. J., Song, Y., Blades, M., Robertson, N. S., et al. (2018). Histone deacetylase (HDAC) 1 and 2 complexes regulate both histone acetylation and crotonylation *in vivo*. *Sci. Rep.* 8:14690. doi: 10.1038/s41598-018-32927-9

Kim, J. S., Jung, H. J., Lee, H. J., Kim, K. A., Goh, C. H., Woo, Y., et al. (2008). Glycine-rich RNA-binding protein 7 affects abiotic stress responses by regulating stomata opening and closing in *Arabidopsis thaliana*. *Plant J.* 55, 455–466. doi: 10.1111/j.1365-313X.2008.03518.x

Knoblauch, M., and Peters, W. S. (2010). Münch, morphology, microfluidics - our structural problem with the phloem. *Plant Cell Environ.* 33, 1439–1452. doi: 10.1111/j.1365-3040.2010.02177.x

Koh, J., Chen, G., Yoo, M. J., Zhu, N., Dufresne, D., Erickson, J. E., et al. (2015). Comparative proteomic analysis of *Brassica napus* in response to drought stress. *J. Proteome Res.* 14, 3068–3081. doi: 10.1021/pr501323d

Koussevitzky, S., Suzuki, N., Huntington, S., Armijo, L., Sha, W., Cortes, D., et al. (2008). Ascorbate peroxidase 1 plays a key role in the response of *Arabidopsis thaliana* to stress combination. *J. Biol. Chem.* 283, 34197–34203. doi: 10.1074/jbc.M806337200

Kwon, O. K., Kim, S. J., and Lee, S. (2018). First profiling of lysine crotonylation of myofilament proteins and ribosomal proteins in zebrafish embryos. *Sci. Rep.* 8:3652. doi: 10.1038/s41598018-22069-3

Li, Y., Sabari, B. R., Panchenko, T., Wen, H., Zhao, D., Guan, H., et al. (2016). Molecular coupling of histone crotonylation and active transcription by AF9 YEATS domain. *Mol. Cell.* 62, 181–193. doi: 10.1016/j.molcel.2016.03.028

Li, Z. Q., Zhou, H. Q., Ouyang, Z., Dai, J., Yue, Q., Wei, Y., et al. (2021). Comparison of active ingredients and protective effects of *Dendrobium huoshanense* of different growth years on acute liver injury. *Zhongguo Zhong Yao Za Zhi.* 46, 298–305. doi: 10.19540/j.cnki.cjcmm.20201023.102

Lin, P., Bai, H. R., He, L., Huang, Q. X., Zeng, Q. H., Pan, Y. Z., et al. (2021). Proteome-wide and lysine crotonylation profiling reveals the importance of crotonylation in chrysanthemum (*Dendranthema grandiforum*) under low-temperature. *BMC Genomics* 22:51. doi: 10.1186/s12864-020-07365-5

Lindquist, S. (1986). The heat-shock response. *Annu. Rev. Biochem.* 55, 1151–1191. doi: 10.1146/annurev.bi.55.070186.005443

Liu, S., Xue, C., Fang, Y., Chen, G., Peng, X. J., Zhou, Y., et al. (2018b). Global involvement of lysine crotonylation in protein modification and transcription regulation in Rice. *Mol. Cell. Proteomics* 17, 1922–1936. doi: 10.1074/mcp.RA118.000640

Liu, K., Yuan, C., Li, H., Chen, K., Lu, L., Shen, C., et al. (2018a). A qualitative proteome-wide lysine crotonylation profiling of papaya (*Carica papaya* L.). *Sci. Rep.* 8:8230. doi: 10.1038/s41598018-26676-y

Lundqvist, T., and Schneider, G. (1989). Crystal structure of the binary complex of ribulose-1,5-bisphosphate carboxylase and its product, 3-phospho-D-glycerate. *J. Biol. Chem.* 264, 3643–3646. doi: 10.2210/pdb1rus/pdb

Meng, Y. Q., Qiu, N., Geng, F., Huo, Y. Q., Sun, H. H., and Keast, R. (2019). Identification of the duck egg white N-Glycoproteome and insight into the course of biological evolution. *J. Agric. Food Chem.* 67, 9950–9957. doi: 10.1021/acs.jafc.9b03059

Millar, A. H., Heazlewood, J. L., Giglione, C., Holdsworth, M. J., Bachmair, A., and Schulze, W. X. (2019). The scope, functions, and dynamics of posttranslational protein modifications. *Annu. Rev. Plant Biol.* 70, 119–151. doi: 10.1146/annurev-arplant-050718-100211

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Ng, T. B., Liu, J., Wong, J. H., Ye, X. J., Wing Sze, S. C., Tong, Y., et al. (2012). Review of research on *Dendrobium*, a prized folk medicine. *Appl. Microbiol. Biotechnol.* 93, 1795–1803. doi: 10.1007/s00253-011-3829-7

Olsen, J. V., and Mann, M. (2013). Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics.* 12, 3444–3452. doi: 10.1074/mcp.O113.034181

Pan, L. H., Feng, B. J., Wang, J. H., Zha, X. Q., and Luo, J. P. (2013). Structural characterization and anti-glycation activityin vitroof a water-soluble polysaccharide from *Dendrobium huoshanense*. *J. Food Biochem.* 37, 313–321. doi: 10.1111/j.1745-4514.2011.00633.x

Parry, M. A., and Hawkesford, M. J. (2012). An integrated approach to crop genetic improvement. *J. Integr. Plant Biol.* 54, 250–259. doi: 10.1111/j.1744-7909.2012.01109.x

Pelvan, A., Bor, M., Yolcu, S., Özdemir, F., and Türkan, I. (2021). Day and night fluctuations in GABA biosynthesis contribute to drought responses in *Nicotiana tabacum* L. *Plant Signal. Behav.* 16:1899672. doi: 10.1080/15592324.2021.1899672

Rennie, E. A., and Turgeon, R. (2009). A comprehensive picture of phloem loading strategies. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14162–14167. doi: 10.1073/pnas.0902279106

Rojas-González, J. A., Soto-Súarez, M., García-Díaz, Á, Romero-Puertas, M. C., Sandalio, L. M., Mérida, Á, et al. (2015). Disruption of both chloroplastic and cytosolic FBPase genes results in a dwarf phenotype and important starch and metabolite changes in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 2673–2689. doi: 10.1093/jxb/erv062

Ruiz-Andres, O., Sanchez-Niño, M. D., Cannata-Ortiz, P., Ruiz-Ortega, M., Egido, J., Ortiz, A., et al. (2016). Histone lysine crotonylation during acute kidney injury in mice. *Dis. Model. Mech.* 9, 633–645. doi: 10.1242/dmm.024455

Sabari, B. R., Tang, Z. Y., Huang, H., Yong-Gonzalez, V., Molina, H., Konget, H. E., et al. (2015). Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol. Cell* 58, 203–215. doi: 10.1016/j.molcel.2015.02.029

Sánchez, A., Pascual, C., Sanchez, A., Vargas-Albores, F., Le Moullac, G., and Rosas, C. (2001). Hemolymph metabolic variables and immune response in *Litopenaeus setiferus* adult males: the effect of acclimation. *Aquaculture* 198, 13–28. doi: 10.1016/S0044-8486(00)00576-7

Serrato, A. J., Yubero-Serrano, E. M., Sandalio, L. M., Muñoz-Blanco, J., Chueca, A., Caballero, J. L., et al. (2009). cpFBPaseII, a novel redox-independent chloroplastic isoform of fructose-1,6-bisphosphatase. *Plant Cell Environ.* 32, 811–827. doi: 10.1111/j.1365-3040.2009.01960.x

Shang, Z. Z., Qin, D. Y., Li, Q. M., Zha, X. Q., Pan, L. H., Peng, D. Y., et al. (2021). *Dendrobium huoshanense* stem polysaccharide ameliorates rheumatoid arthritis in mice via inhibition of inflammatory signaling pathways. *Carbohydr. Polym.* 258:117657. doi: 10.1016/j.carbpol.2021.117657

Spreitzer, R. J., and Salvucci, M. E. (2002). Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* 533, 449–475. doi: 10.1146/annurev.arplant.53.100301.135233

Sun, H. J., Liu, X. W., Li, F. F., Li, W., Zhang, J., Xiao, Z. X., et al. (2017). First comprehensive proteome analysis of lysine crotonylation in seedling leaves of *Nicotiana tabacum*. *Sci. Rep.* 7:3013. doi: 10.1038/s41598-017-03369-6

Sun, J. H., Qiu, C., Qian, W. J., Wang, Y., Sun, L. T., Li, Y. S., et al. (2019). Ammonium triggered the response mechanism of lysine crotonylome in tea plants. *BMC Genomics* 20:340. doi: 10.1186/s12864-019-5716-z

Tan, M. J., Luo, H., Lee, S., Jin, F. L., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028. doi: 10.1016/j.cell.2011.08.008

Tang, Z. Z., and Cheng, S. J. (1984). A study on the raw plants for the Chinese traditional medicine "HUOSHAN SHI-HU". *Bull. Bot. Res.* 3, 141–144.

Vega, E. D. L., Hall, M. R., Degnan, B. M., and Wilson, K. J. (2006). Short-term hyperthermic treatment of *Penaeus monodon* increases expression of heat shock protein 70 (HSP70) and reduces replication of gill associated virus (GAV). *Aquaculture* 253, 82–90. doi: 10.1016/j.aquaculture.2005.07.041

Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. Jr. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* 44, 7342–7372. doi: 10.1002/anie.200501023

Wan, J. H., Liu, H. Y., Chu, J., and Zhang, H. Q. (2019). Functions and mechanisms of lysine crotonylation. *J. Cell. Mol. Med.* 23, 7163–7169. doi: 10.1111/jcmm.14650

Wei, W., Mao, A. Q., Tang, B., Zeng, Q. F., Gao, S. N., Liu, X. G., et al. (2017). Large-scale identification of protein crotonylation reveals its role in multiple cellular functions. *J. Proteome Res.* 16, 1743–1752. doi: 10.1021/acs.jproteome.7b00012

Xu, M. L., Luo, J. D., Li, Y., Shen, L. L., Zhang, X., Yu, J., et al. (2021). First comprehensive proteomics analysis of lysine crotonylation in leaves of peanut (*Arachis hypogaea* L.). *Proteomics* 21:e2000156. doi: 10.1002/pmic.202000156

Xu, W. Z., Wan, J. H., Zhan, J., Li, X. Y., He, H. Y., Shi, Z. M., et al. (2017). Global profiling of crotonylation on non-histone proteins. *Cell Res.* 27, 946–949. doi: 10.1038/cr.2017.60

Xue, C., Liu, S., Chen, C., Zhu, J., Yang, X. B., Zhou, Y., et al. (2018). Global proteome analysis links lysine acetylation to diverse functions in *Oryza Sativa*. *Proteomics* 18, 1–42. doi: 10.1002/pmic.201700036

Yang, M., Deng, W., Cao, X. Y., Wang, L. J., Yu, N., Zheng, Y., et al. (2020). Quantitative phosphoproteomics of milk fat globule membrane in human colostrum and mature milk: new insights into changes in protein phosphorylation during lactation. *J. Agric. Food Chem.* 68, 4546–4556. doi: 10.1021/acs.jafc.9b06850

Yang, Q., Li, Y., Apaliya, M. T., Zheng, X., Serwah, B. N. A., Zhang, X., et al. (2018). The response of *Rhodotorula mucilaginosa* to Patulin based on lysine crotonylation. *Front. Microbiol.* 9:2025. doi: 10.3389/fmicb.2018.02025

Yuan, Y. D., Yu, M. Y., Jia, Z. H., Song, X., 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

Zdobnov, E. M., and Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. doi: 10.1093/bioinformatics/17.9.847

Zha, X. Q., Zhao, H. W., Bansal, V., Pan, L. H., Wang, Z. M., and Luo, J. P. (2014). Immunoregulatory activities of *Dendrobium huoshanense* polysaccharides in mouse intestine, spleen and liver. *Int. J. Biol. Macromol.* 64, 377–382. doi: 10.1016/j.ijbiomac.2013.12.032

Zhao, X. M. (2007). *Supplement to Compendium of Materia Medica*. Beijing: China Press of Chinese Traditional Medicine.

Zhou, P. N., Pu, T. Z., Gui, C., Zhang, X. Q., and Gong, L. (2020). Transcriptome analysis reveals biosynthesis of important bioactive constituents and mechanism of stem formation of *Dendrobium huoshanense*. *Sci. Rep.* 10:2857. doi: 10.1038/s41598-020-59737-2

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

Zimmermann, G., Kelly, G. J., and Latzko, E. (1976). Efficient purification and molecular properties of spinach chloroplast fructose 1,6-bisphosphatase. *Eur. J. Biochem.* 70, 361–367. doi: 10.1111/j.1432-1033.1976.tb11025.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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fpls.2022.804593

TcMYB29a, an ABA-Responsive R2R3-MYB Transcriptional Factor, Upregulates Taxol Biosynthesis in *Taxus chinensis*

Xiaoying Cao†, Lingxia Xu†, Ludan Li, Wen Wan and Jihong Jiang**

Key Laboratory of Biotechnology for Medicinal Plants of Jiangsu Province, School of Life Sciences, Jiangsu Normal University, Xuzhou, China

OPEN ACCESS

Edited by:
Yongliang Liu, University of Kentucky, United States

Reviewed by:
Sumit Ghosh, Central Institute of Medicinal and Aromatic Plants, Council of Scientific and Industrial Research (CSIR), India
Lv Zongyou, Shanghai University of Traditional Chinese Medicine, China

***Correspondence:**
Wen Wan
wanwen85@jsnu.edu.cn
Jihong Jiang
jhjiang@jsnu.edu.cn

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

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *29 October 2021*
Accepted: *21 January 2022*
Published: *04 March 2022*

Citation:
Cao X, Xu L, Li L, Wan W and Jiang J (2022) TcMYB29a, an ABA-Responsive R2R3-MYB Transcriptional Factor, Upregulates Taxol Biosynthesis in Taxus chinensis. Front. Plant Sci. 13:804593. doi: 10.3389/fpls.2022.804593

Paclitaxel (Taxol), a highly modified diterpene agent mainly obtained from *Taxus* species, is the most widely used anticancer drug. Abscisic acid (ABA) is a well-known stress hormone that plays important roles in the secondary metabolism of plants, and it can also induce the accumulation of taxol in *Taxus* cell suspension cultures. However, the mechanism behind the regulation of taxol biosynthesis by ABA remains largely unknown. In previous research, a R2R3 MYB transcription factor (TF) TcMYB29a was observed to show a significant correlation with taxol biosynthesis, indicative of its potential role in the taxol biosynthesis. In this study, the TcMYB29a encoded by its gene was further characterized. An expression pattern analysis revealed that *TcMYB29a* was highly expressed in the needles and roots. Overexpression of *TcMYB29a* in *Taxus chinensis* cell suspension cultures led to an increased accumulation of taxol, and upregulated expression of taxol-biosynthesis-related genes, including the taxadiene synthase (TS) gene, the taxane 5α-hydroxylase (T5OH) gene, and the 3′-N-debenzoyl-2′-deoxytaxol-N-benzoyltransferase (DBTNBT) gene as compared to the controls. Chromatin immunoprecipitation (ChIP) assays, yeast one-hybrid (Y1H) assays, electrophoretic mobility shift assays (EMSAs), and dual-luciferase reporter assays verified that TcMYB29a could bind and activate the promoter of *TcT5OH.* Promoter sequence analysis of *TcMYB29a* revealed that its promoter containing an AERB site from -313 to -319 was a crucial ABA-responsive element. Subsequently, the ABA treatment assay showed that *TcMYB29a* was strongly upregulated at 6 h after ABA pretreatment. Furthermore, *TcMYB29a* was strongly suppressed at 3 h after the methyl jasmonate (MeJA) treatment and was depressed to the platform at 12 h. Taken together, these results reveal that TcMYB29a is an activator that improves the accumulation of taxol in *Taxus chinensis* cells through an ABA-medicated signaling pathway which is different from JA-medicated signaling pathways for the accumulation of taxol. These findings provide new insights into the potential regulatory roles of MYBs on the expression of taxol biosynthetic genes in *Taxus*.

Keywords: transcription factor, R2R3-MYB, taxol biosynthesis, *Taxus chinensis*, taxane 5α-hydroxylase

INTRODUCTION

Taxus chinensis is an endangered and economically valuable medicinal woody species of the genus, *Taxus*. Its bark can produce taxol (generic name: paclitaxel), which is one of the most effective anticancer drugs derived from natural sources and is widely used in the treatment of various solid tumors, such as breast, ovarian, and lung cancer and Kaposi's sarcoma (Wani et al., 1971; Gallego-Jara et al., 2020). Taxol is a taxane diterpene, and its biosynthesis mainly needs two metabolic pathways, the diterpenoid pathway and the phenylpropanoid pathway. The former provides the main taxane carbon skeleton, baccatin III, and the latter offers the phenylisoserine side chain (Croteau et al., 2006; Ssmsa and Mnr, 2020). The highly complex taxol biosynthesis pathway involves more than 20 enzymes, which catalyze at least 19 steps of reactions and convert the universal diterpenoid precursor, geranylgeranyl diphosphate (GGPP), into taxol (Ssmsa and Mnr, 2020; **Figure 1**). The first step of taxol biosynthesis, a cyclization of GGPP into taxa-4(5),11(12)-diene, is catalyzed by the taxadiene synthase (TS), which is a slow-starter and a rate-limiting enzyme for the provision of the key intermediate 10-deacetylbaccatin III (10-DAB) (**Figure 1**). Taxane 5α-hydroxylase (T5OH), a cytochrome P450 enzyme, catalyzes the first oxygenation step of taxol biosynthesis, in which taxa-4(5),11(12)-diene is transformed into taxa-4(5),11(12)-diene-5α-ol (Koepp et al., 1995; **Figure 1**), and 3′-N-debenzoyl-2′-deoxytaxol-N-benzoyltransferase (DBTNBT), an important enzyme, is involved in the formation of the functional taxol molecule, which converts the 3′-N-debenzoyltaxol into taxol (**Figure 1**).

Jasmonic acid (JA) and its derivative, methyl jasmonate (MeJA), among the most important stress hormones, have been evaluated as the most effective inducers of taxol biosynthesis in *Taxus* cell cultures (Li et al., 2013; Lenka et al., 2015; Zhang et al., 2015, 2018). Recently, transcription factors (TFs), including WRKY, MYC, basic helix-loop-helix (bHLH), ethylene responsive factor (ERF), and MYB, have been isolated, characterized, and identified as regulators of taxol biosynthesis in *Taxus* (Li et al., 2013; Lenka et al., 2015; Zhang et al., 2015, 2018), and most of them have been verified to be JA-responsive. TcWRKY1, a JA-responsive WRKY transcription factor isolated from *Taxus chinensis*, participates in the regulation of taxol biosynthesis by directly activating the expression of 10-deacetylbaccatin-III-10-b-O-acetyltransferase (DBAT) gene (Li et al., 2013). Another two JA-responsive WRKY TFs, TcWRKY8 and TcWRKY47 in *T. chinensis*, significantly increase the expression levels of several taxol biosynthetic genes, including *DBAT*, *T5OH*, and *TcERF15* (Zhang et al., 2018). The MYC family members, which are the core TFs in the JA signaling pathway, are also found to act in the regulation of taxol biosynthesis (Lenka et al., 2015). The interaction between jasmonate ZIM-domain (JAZ) proteins and MYC proteins is also detected in *Taxus media* in a similar mode of action as in model plants. In a model, JAZs can bind with MYCs and inhibit their transcriptional activities (Cui et al., 2019; **Figure 1**). In *Taxus cuspidata,* three JA-inducted MYC TFs, TcJAMYC1, TcJAMYC2, and TcJAMYC4, are verified to be involved in the negative regulation of the expression of taxol biosynthetic genes in *Taxus cuspidata* (Lenka et al., 2015). Another study showed that TcMYC2a may regulate the expression of *TS*, *TAT*, *DBTNBT*, *T13OH*, and *T5OH* directly or indirectly through ERF12 and ERF15 that are dependent on the JA signaling pathway in *Taxus chinensis* (Li et al., 2013; Zhang et al., 2018; **Figure 1**).

The MYB protein family, defined by the myb domain, constitutes one of the largest groups of TFs described in the plant kingdom, and it acts as a key factor in the regulatory networks controlling various processes in plant growth and development and in response to the biotic and abiotic stresses (Yanhui et al., 2006; Dubos et al., 2010). The repeat motif (R) of the myb domain usually comprises about 52 amino acids and forms three α-helices, in which the second and third helices of each domain build a helix-turn-helix (HTH) fold and are involved in specifically binding to their target DNA sequences (Ogata et al., 1996; Dubos et al., 2010; Baldoni et al., 2015; Hou et al., 2018). Based on the number of myb repeats, the MYB proteins are classified into four subfamilies, including R1R2R3R1/2-MYBs (4R-MYBs), R1R2R3-MYBs (3R-MYBs), R2R3-MYBs, and MYBs (1R-MYBs and MYB-related) in plants (Dubos et al., 2010). Among them, the R2R3-MYB subfamily contains the most abundant MYB proteins and includes more than 100 members in different species, such as *Arabidopsis thaliana* and *Populus trichocarpa* (Wilkins et al., 2009; Dubos et al., 2010; Hou et al., 2018). The R2R3-MYB TFs have a modular structure with a highly conserved myb domain for DNA-binding at the N-terminal (Baldoni et al., 2015) and a transcription activation or repression region usually located in the C-terminus of the proteins. Amino acid motifs in the C-terminal of R2R3-MYB proteins are rich in acidic amino acid residues that are responsible for the diverse regulatory functions of the TFs (Araki et al., 2004; Daniel et al., 2004; Morse et al., 2009). Numerous R2R3-MYB proteins have been characterized and have been found to play central roles in the secondary metabolism of plants (Martin and Paz-Ares, 1997). For instance, R2R3-MYBs are well known for their regulatory role in the pathways of anthocyanin biosynthesis, a specific branch of the phenylpropanoid pathway, by forming a triad with bHLH and WD40 TFs (Wang et al., 2021b). The overexpression of SmMYB9b, an R2R3-MYB TF isolated from *Salvia miltiorrhiza,* could increase the accumulation of tanshinone in *S. miltiorrhiza* (Zhang J. et al., 2017). Also, SmMYB98, an R2R3-MYB TF, isolated from this medicinal herb, could simultaneously promote the biosynthesis of tanshinone and salvianolic acid in *S. miltiorrhiza* (Hao et al., 2020). Two MYB TFs, BpMYB21 and BpMYB61 in birch, regulate triterpenoid biosynthesis (Yin et al., 2020). Furthermore,

Abbreviations: TF, transcription factor; TS, taxadiene synthase; T5OH, taxane 5a-hydroxylase; ChIP, chromatin immunoprecipitation; Y1H, yeast one-hybrid; EMSA, electrophoretic mobility shift assays; MeJA, methyl jasmonate; GGPP, geranylgeranyl diphosphate; 10-DAB, 10-deacetylbaccatin III; bHLH, basic helix-loop-helix; ERF, ethylene responsive factor; DBAT, 10-deacetylbaccatin-III-10-β-O-acetyltransferase; HTH, helix-turn-helix; DAPI, 4′,6′-diamidino-2-phenylindole; LC-MS, liquid chromatography-mass spectrometry; MREs, MYB recognition elements; SD, synthetic dextrose; CDS, coding sequence; OE, over-expressed; DW, dry weight; qRT-PCR, quantitative real-time PCR; DBTNBT, 3′-N-debenzoyl-2′-deoxytaxol-N-benzoyltransferase.

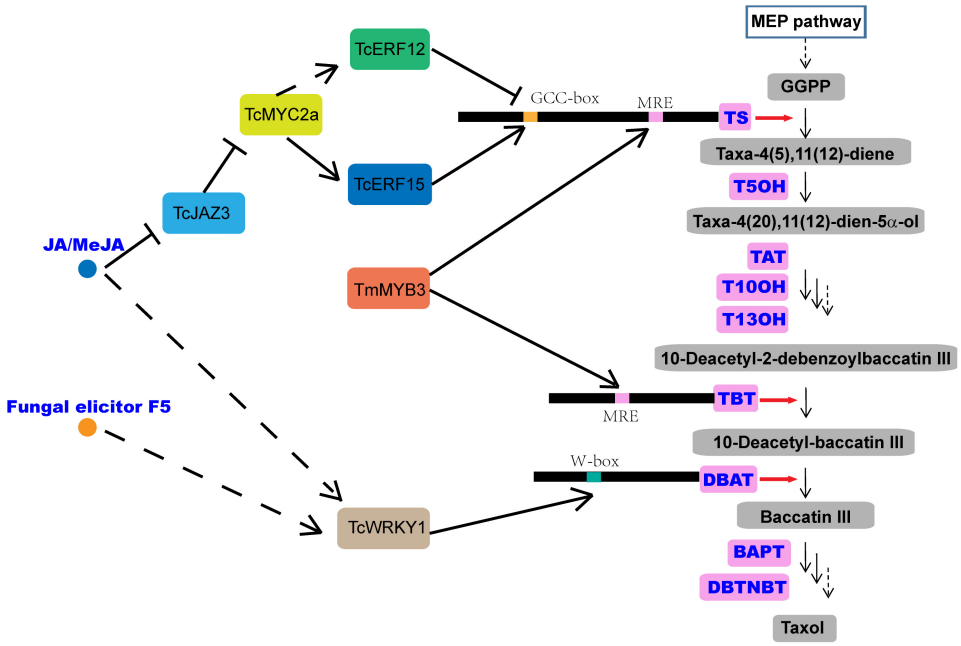


FIGURE 1 | Overview of the taxol biosynthesis pathway. MEP, 2-C-methyl-d-erythritol 4-phosphate; TS, taxadiene synthase; GGPP, geranylgeranyl diphosphate; T5OH, taxadiene-5α-hydroxylase; TAT, taxadien-5α-ol O-acetyltransferase; T10OH, taxane 10β-hydroxylase; T13OH, taxane 13α-hydroxylase; TBT, taxane 2α-O-benzoyltransferase; DBAT, 10-deacetylbaccatin-III-10-β-O-acetyltransferase; BAPT, baccatin III-3-amino-3-phenylpropanoyltransferase; DBTNBT, 3′-N-debenzoyltaxol-N-benzoyltransferase; MRE, MYB recognition element; JAZ, jasmonate ZIM-domain protein.

TmMYB3, an R2R3-MYB TF, has been found to be involved in taxol biosynthesis in *Taxus media,* by activating the expression of *TmTBT* and *TmTS* (Yu et al., 2020). Although previous studies have indicated that the MYBs are involved in the regulation of taxol biosynthesis (Li et al., 2012; Wang et al., 2019; Yu et al., 2020); to date, only a few *MYB* genes have been identified in the *Taxus* spp.

Except for JA, the ABA is considered as another stress hormone for its important roles in response to various biotic and abiotic stresses in plants (Wang et al., 2021a,b). Although ABA has also been evaluated and found to induce taxol biosynthesis in *Taxus* cell cultures (Zhang and Fevereiro, 2010), its function in the regulation of the taxol biosynthesis is largely unknown. In our current research (Cao et al., 2022), the fermentation broth of an endophytic fungus isolated from *T. chinensis* could significantly promote taxol accumulation in *T. chinensis* needles, with a 3.26-fold increase compared to the control. TF-encoding genes, such as MYBs, ERFs, and bHLH, were detected as differentially expressed genes after KL27-FB treatment. In these TFs, an *MYB* gene showed a high expression after treatment with the fungal elicitor, and the overexpression of the *MYB* gene could significantly improve the accumulation of taxol in the cells of *Taxus callus*. Promoter sequence analysis showed that its promoter contains an ABA-responsive element but not a JA-responsive element, suggesting its potential role in the regulation of taxol biosynthesis by ABA. A homolog search and a phylogenic analysis identified the MYB as a typical R2R3-MYB TF that shows high similarity with R2R3-MYB transcription factor 29 (QHG11457.1) (TcMYB29) (Hu et al., 2020) and was thus renamed as TcMYB29a in this study. Although Hu et al. (2020) performed a comprehensive analysis of TcMYB29, the function of TcMYB29 in taxol biosynthesis in *Taxus* was not studied deep. A series of experiments were then performed to determine the function of TcMYB29a in the taxol biosynthesis and to explain the regulation system.

MATERIALS AND METHODS

Plant Materials and Transformation of *Taxus chinensis* Calli

The *Taxus chinensis* plants used were grown in pots containing soil in the greenhouse of the Jiangsu Normal University, Xuzhou, China at 22°C with 50% relative humidity. The needles, stem epidermis, phloem, xylem, and roots were collected from 5-year-old *T. chinensis* seedlings. Specimens of the tissue were collected from three plants. All samples were frozen in liquid nitrogen and stored at −80°C until RNA extraction.

T. chinensis calli used for the genetic transformation were cultured in aseptic bottles containing a solid medium of modified B5 (Lee et al., 2010), which contains 4 g/L of plant gel and 20 g/L of surcose. The calli were cultured at 25°C in the dark, and the wild-type and transgenic calli were subcultured at 2-week intervals on the modified B5 solid medium.

Recombinant plasmids were transformed into *T. chinensis* calli using the *Agrobacterium*-medicated transformation (Zhang et al., 2015). After infection for 30 min, the calli were transferred into the modified B5 solid medium in the dark for 2 days

at 25°C. The cocultured calli were rinsed with a fresh liquid medium containing 50 μg/mL of kanamycin and 300 μg/mL of cefotaxime, then transferred into a fresh solid medium containing 50 μg/mL of kanamycin and 300 μg/mL of cefotaxime, and cultured in the dark at 25°C.

RNA Isolation and Quantitative Real-Time PCR

The RNA isolation was performed by the EASYspin plant RNA extraction kit (Aidlab Bio., China) according to the instructions of the manufacturer. The quality and quantity of the RNA were determined by a NanoDrop2000c spectrophotometer (Thermo Fisher Scientific, United States), and the RNA integrity was identified by electrophoresis on 1.0% of agarose gels. The complementary DNA (cDNA) was synthesized from 1.0 μg of total RNA using the HiScript II Q RT SuperMix by a quantitative PCR (qPCR) kit with gDNase (Vazyme, China) according to the protocols of the manufacturer. To determine the expression levels of *TcMYB29a* in different tissues or treatments, a quantitative real-time PCR (qRT-PCR) analysis was carried out using the ABI StepOnePlus Real-time PCR systems (Thermo Fisher Scientific, United States). One microliter of synthesized cDNA (diluted at 1:5) was used as the template for the qRT-PCR. Specific primers of each gene are listed in **Supplementary Table 1**. The *TcGAPDH* was selected as a reference gene (Zhang et al., 2020). Amplification cycles included 30 s at 95°C, followed by 40 cycles at 95°C for 15 s and 60°C for 30 s. Each measurement was performed with three biological replicates. Data were analyzed by using the $2^{-\Delta\Delta CT}$ method.

Construction of Overexpression and RNAi Vectors

The coding region of the *TcMYB29a* was taken from the transcriptome datasets of *T. chinensis* from previous research. Specific primers (**Supplementary Table 1**) were designed to amplify the *TcMYB29a* DNA segment from the cDNA of *T. chinensis* using the following PCR parameters: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 95°C for 15 s, annealing at 58°C for 30 s, extension at 72°C for 2 min, and final extension at 72°C for 10 min. The PCR products were subcloned into the pHB-GFP vector with *Bam*HI and *Xba*I restriction sites to form pHB-TcMYB29a-OE. The sequence (1,584 bp) of the *TcMYB29a* gene was used to construct an intron-spliced hairpin RNA (RNAi construct) to inhibit the expression of *TcMYB29a*. The amplified fragment was inserted into the pHB vector by reverse orientation to construct the RNAi vector, pHB-TcMYB29a-RNAi. The vectors, including pHB-TcMYB29a-OE, pHB-TcMYB29a-RNAi, and pHB (set as a control), were transformed into the *T. chinensis* calli. Each experiment was conducted using more than three biological replicates.

Sequence Analysis of TcMYB29a

The molecular weight and theoretical isoelectric point of TcMYB29a were computed using the compute pI/Mw tool on the ExPASy server.[1] A BLAST search (BlastX)[2] was used for a homology search from the SWISS-PROT protein database. The TcMYB29a protein was predicted on PlantTFDB.[3] Classification of the myb-like DNA binding domain was predicted using the online software, HMMER3.3.[4] Multiple sequence alignments of the full-length MYB proteins were performed using clustalW in default settings. A phylogenetic tree was described using the MEGA Version 11 adopting the Neighbor-Joining algorithm, and the reliability of the branching pattern was tested with 1,000 bootstrap repetitions.

Subcellular Localization of TcMYB29a in *Taxus* and the Epidermis Cells of Tobacco

The recombinant vector, 35S:GFP-TcMYB29a, was transformed into the *Agrobacterium tumefaciens* strain, GV3101, for plant transformation. *T. chinensis* calli were transformed as described above, and the transgenic *Taxus* calli were then placed on modified B5 plates at 25°C and in the dark for 2 days. Transient expression in tobacco leaves was assessed according to a published method (Zhou et al., 2018). The leaf epidermis of tobacco was peeled to make a temporary squash. After incubation with a phosphate-buffered saline containing 4′,6′-diamidino-2-phenylindole (DAPI), the temporary squash and *Taxus* calli were then observed under a Leica Fluorescence Microscope at 10 × 20 (Leica Microsystems, Wetzlar, GmBH).

Cis-Element Analysis of Promoter Sequences

Genome DNA was extracted by a Plant Genomic DNA extraction Kit (Aidlab Biotech, China) and used as a template. The specific primers for *TcT5OH*, *TcTS*, *TcBAPT*, and *TcDBTNBT* promoters were designed based on the known sequences of the cDNA (accepted from the transcriptome datasets in *T. chinensis* in previous research) (**Supplementary Table 1**) and four general primers provided by the Genome Walking Kit (Takara). Genome walking PCR was performed by applying the Genome Walking Kit, and the PCR products were cloned, sequenced, and aligned with the designed partial sequences of the open reading frame of *TcT5OH*, *TcTS*, *TcBAPT*, and *TcDBTNBT* genes, respectively, to decide the 5′-flanking region. The promoter sequence of *TcMYB29a* was achieved from the *T. chinensis* genome sequence downloaded from the NCBI. Then, the online database PLACE[5] and PlantCARE[6] were used to identify the *cis*-acting elements of the promoters.

Chromatin Immunoprecipitation PCR

Chromatin immunoprecipitation (ChIP) was conducted with the transgenic *Taxus* calli harboring 35S:GFP-TcMYB29a using

[1] http://web.expasy.org

[2] http://www.ncbi.nlm.nih.gov/

[3] http://planttfdb.gao-lab.org/tf.php?sp=Ppe&did=Prupe.I004500.1.p

[4] https://www.ebi.ac.uk/Tools/hmmer

[5] http://www.dna.affrc.go.jp/PLACE/

[6] http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

the method reported in the literature (Meng et al., 2018). The analysis of ChIP DNA products was performed by qRT-PCR using primers that were synthesized to amplify the DNA fragments in the promoter regions of *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT* genes. The primer sequences for R-1 and R-2, R-3 and R-4, R-5 to R-7, and R-8 and R-9 were used for the amplification of DNA fragments in the promoters of *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT*, respectively. Primer sequences TcTSqF and TcTSqR (CDS-1), TcT5OHqF and TcT5OHqR (CDS-2), TcBAPTqF and TcBAPTqR (CDS-3), and TcDBTNBTqF and TcDBTNBTqR (CDS-4) were used for amplifying the regions of coding sequences (CDSs) in *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT*, respectively, and were set as an internal control (**Supplementary Table 1**; Wang and Lindås, 2018). These experiments were repeated more than three times.

Yeast One-Hybrid Assays

To confirm the interaction between TcMYB29a and the promoters of *TcBAPT* and *TcT5OH* genes, yeast one-hybrid (Y1H) assays were conducted using the Matchmaker® Gold Yeast One-Hybrid Library Screening System (Clontech, CA, United States) according to the instructions of the manufacturer. The full-length cDNA segment of *TcMYB29a* was cloned into the *Eco*RI-*Xho*I sites of the GAL4 activation vector (pGADT7-Rec) to form pAD-TcMYB29a. The promoter fragment containing the putative MYB recognition elements (MREs), B1 and B2 of the *TcT5OH* promoter and P5 of *TcBAPT*, were separately amplified from the genome DNA of *T. chinensis* with primers (**Supplementary Table 1**). The DNA segments containing three repeats of B1 binding sites and their mutants were synthesized by Sangon Biotech Co. (Shanghai). Then, these fragments and oligos were cloned into the *Sac*I-*Sal*I sites of the pABAi vector to form pY1-ABAi, pY2-ABAi, p3B1-ABAi, p3mB1-ABAi, and pY3-ABAi using a ClonExpress II One step Cloning Kit (Vazyme, China). Following the protocols of the manufacturer, the vectors were cotransformed into the yeast strain, Y1HGold, according to LiAc conversion protocols. The transformed yeast cells were diluted to an $OD_{600\,nm}$ of 0.005, dropped onto a selective medium containing a synthetic dextrose (SD) without Ura and Leu (SD/-Ura/-Leu), and SD/-Ura/-Leu with 50 ng/mL of ABA (SD/-Ura/-Leu + 50 ng/mL of ABA), 100 ng/mL of ABA (SD/-Ura/-Leu + 100 ng/mL of ABA), and 150 ng/mL of ABA (SD/-Ura/-Leu + 150 ng/mL of ABA), and incubated at 30°C for 48 h.

Electrophoretic Mobility Shift Assays

The coding region of *TcMYB29a* was subcloned into the pGEX-4T-1 vector to form GST-TcMYB29a, in which a GST-tag was fused into the N-terminal of the TcMYB29a. The resulting plasmid was transformed into *Escherichia coli* Rosetta (DE3). The 3′-end biotin B1 probe corresponding to the B1 site was prepared (Sangon Bio, China) (**Supplementary Table 1**). Both the mutant probes containing two mutated nucleotides and the B1 probe were without a biotin label and were set as a competitor probe. The electrophoretic mobility shift assays (EMSAs) were conducted using a LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, United States) according to the instructions of the manufacturer.

Dual-Luciferase Reporter Assays

To further measure the regulation of *TcT5OH* expression by TcMYB29a protein, dual-luciferase reporter assays were performed. For transcription activity analysis, the coding region of *TcMYB29a* was cloned into the pHB-GFP vector with *Hin*dIII and *Pst*I restriction sites under the control of the 35S promoter as an effector (35Spro:TcMYB29a). The promoter sequence of *TcT5OH* and its mutant, *mTcT5OH*, were inserted into a pGreenII 0800-Luc vector and were then cotransformed with 35Spro:TcMYB29a or free pHB vector (35S:pro, set as a negative control) into *Nicotiana benthamiana* leaves, adopting an *Agrobacterium*-mediated method described previously (Zhou et al., 2018). After being cultivated in the dark for 6 h and under long-day conditions (16 h/8 h, day/night) for 36 h, the transformed leaves were sprayed with a D-luciferin sodium salt (Solarbio, Beijing, China) and then examined by adopting a Bio-Rad Gel Doc XR (Bio-Rad, United States). Each assay was carried out with three biological replicates. The sequences of the primers are listed in **Supplementary Table 1**.

Liquid Chromatography-Mass Spectrometry Analysis of Taxanes

The taxanes, including 10-deacetylbaccatin III, baccatin III, and taxol were studied by liquid chromatography-mass spectrometry (LC-MS) as described previously (Kayan et al., 2021). In brief, transgenic *Taxus* calli were frozen in liquid nitrogen, freeze-dried, and ground to powder using a mortar; about 0.1 g powder was mixed with 3 mL of 100% methanol and was then ultrasonicated three times for 60 min each. After centrifugation at 5,000 rpm for 5 min, the supernatant liquor was collected and extracted three times with dichloromethane/water (1:1, v/v). The organic fraction was collected, dried in vacuum, resuspended in 1 mL of methanol, and filtered through a 0.45-μm of organic-phase filter. The column used in all the experiments was a Poroshell 120 EC-C18 (4.6 × 150 mm, 4 μm) column (Agilent Technologies, Cheadle, United Kingdom) and the temperature for the chromatographic separation was set to 45°C. About 10 μl of the sample was injected in each chromatographic run. Mobile phases, delivered at 1.0 mL/min, consisted of 0.1% of formic acid either in 2.0 mmol/L of ammonium acetate aqueous solution (mobile phase A) and acetonitrile (mobile phase B). The chromatographic gradient was set as follows: 50% B for 0–4 min followed by a gradient to 95% B for 2 min, an isocratic step at 95% B for 1 min and then a gradient of50% B for 0.1 min, and an isocratic step at 50% B for 2.9 min. Mass spectral data were attained in positive electrospray mode (ESI +) in the multi-reaction monitoring mode. Operating conditions were optimized as follows: spray voltage of 5,500 V, ion source temperature of 550°C, curtain gas pressure at 25 psi (172 kPa), Ion Source Gas 1 at 50 psi (345 kPa), Ion Source Gas 2 at 60 psi (414 kPa), and residence time of 100 ms.

Methyl Jasmonate and Abscisic Acid Treatments

In vitro long-term subcultured *T. chinensis* calli were maintained on the liquid modified B5 medium for 2 days; then, 8 g of cells

were suspended in 50 mL of fresh liquid modified B5 medium, and incubated at 25°C while being shaken at 100 rpm for 48 h in the dark. Next, for MeJA and ABA treatments, the final concentrations of 50 μmol/L of MeJA or 20 μmol/L of ABA and the same volume of ethanol (set as the control) were added to the liquid medium, respectively, incubated at 25°C, and shaken at 100 rpm on a rotary shaker. These samples were harvested and frozen in liquid nitrogen after MeJA treatment at 0, 0.5, 1, 3, 6, 12, and 24 h, or after the ABA treatment at 0, 6, 24, 48, and 72 h for gene expression analysis, respectively.

RESULTS

Cloning and Basic Analysis of TcMYB29a

Based on the transcriptomes of *T. chinensis*, the full-length of the CDS of TcMYB29a was cloned. Sequence analysis indicated that *TcMYB29a* encodes a protein containing 527 amino acids with a predicted molecular weight of 60.18 kDa and a theoretical pI of 7.95. Multiple sequence alignments revealed that TcMYB29a contains two R motifs in the N-terminal of the amino acid sequences and belongs to the R2R3-MYB protein; there is no common repressor domain, such as EAR or TLLLFR motifs, identified in its C-terminus (**Figure 2**). BlastX hits analysis indicated that TcMYB29a shares the highest similarity with R2R3-MYB transcription factor 29 (QHG11457.1) in *T. chinensis* (98.87%) (Hu et al., 2020). Phylogenetic analysis indicated a greater similarity among TcMYB29a, TcMYB29, and *Larix gmelinii var. olgensis* MYB3 (**Figure 2**). Hu et al. (2020) reported that TcMYB29, AtMYB88, and AtMYB124 were clustered in the S26 subgroup with counterparts in *Arabidopsis thaliana* and *T. chinensis*.

TcMYB29a Localized in Nuclei

The subcellular localization of TcMYB29a was performed *in vivo*. The location of TcMYB29a in plant cells was examined using GFP as a marker. In the epidemics cells of the tobacco leaf that transiently express GFP-TcMYB29a, the GFP signals were located in the nuclei. In the cells of *Taxus* callus that stably express GFP-TcMYB29a, the GFP signals were also observed in the nuclei and colocated with the nuclei marker dye, DAPI (**Figure 3**). However, when tobacco leaf epidemics cells and *Taxus* callus cells harboring the control vector carrying GFP alone were used as controls, the GFP was located in the cytoplasm and the nuclei in the tobacco leaf epidemics cells and *Taxus* callus cells (**Figure 3**). These results indicated that the GFP-TcMYB29a was localized in the nuclei (**Figure 3**), which is consistent with the prediction that TcMYB29a acts as a TF.

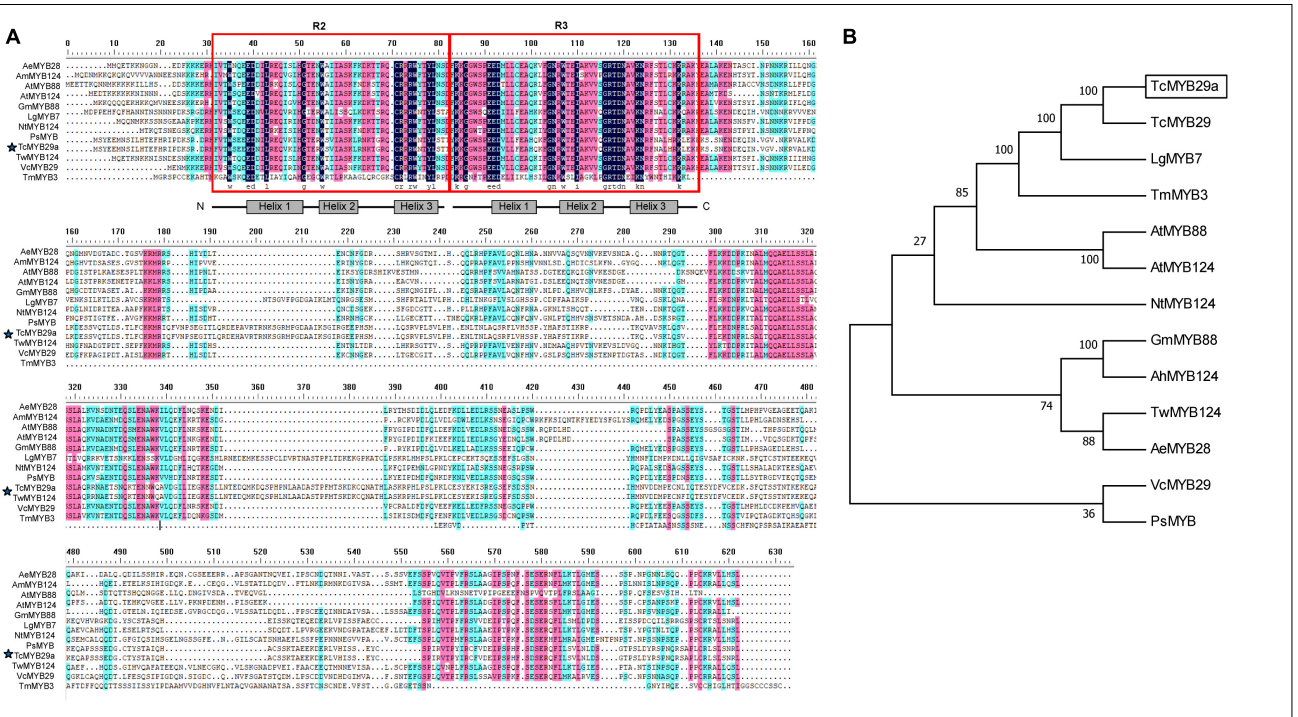


FIGURE 2 | Amino acid sequence alignment and phylogenetic analysis of TcMYB29a. **(A)** Amino acid sequence alignment of TcMYB29a and other known R2R3-MYBs in plants. Conserved residues are highlighted in blue and partial conservation is shown in pink and green. The R2 and R3 domains are indicated with red boxes. The TcMYB29a studied in this study is indicated by blue stars. **(B)** The phylogenetic tree of TcMYB29a. Phylogenetic analysis of full-length TcMYB29a; TcMYB29a is framed by a solid box. The following GenBank accession numbers are used: *Arabidopsis thaliana* AtMYB88 (NP_565291.2), AtMYB124 (NP_001077534.1); *Taxus chinensis* TcMYB29 (QHG11457.1), *Larix gmelinii var. olgensis* LgMYB7 (QFG01315.1), *Nicotiana tomentosiformis* NtMYB124 (XP_009593376.1), *Glycine max* GmMYB88 (XP_003519765.1), *Arachis hypogaea* AhMYB124 (XP_025699147.1), *Vaccinium corymbosum* VcMYB29 (AYC35407.1), *Paeonia suffruticosa* PsMYB (QIG55701.1), *Tripterygium wilfordii* TwMYB124 (XP_038693483.1), and *Abelmoschus esculentus* AeMYB28 (QST87265.1). The amino acid sequence of *Taxus media,* TcMYB3, has been taken from previous research (Yu et al., 2020).

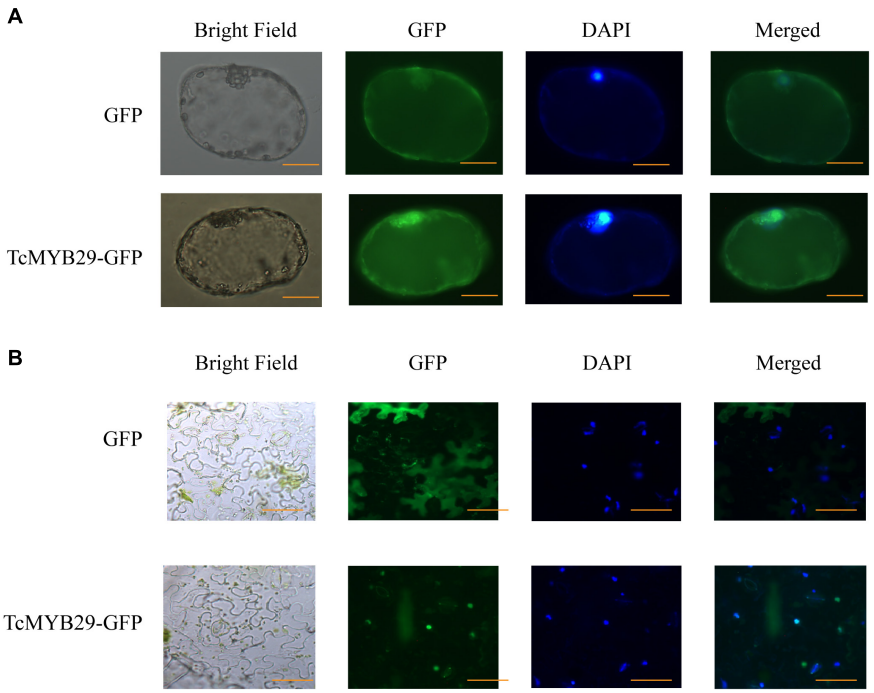


FIGURE 3 | Subcellular localization of TcMYB29a in the **(A)** *Taxus* cells and **(B)** *Nicotiana tabacum*. The fused protein of TcMYB29a and GFP were transformed into *Taxus* cells. GFP fluorescence was observed 2 days after infection by laser scanning with 10 × 20 magnification. Photographs were taken in bright light and dark field for examining the GFP and diamidino-2-phenylindole (DAPI), respectively), and in combination (merged). Scale bar represents 40 μm.

Expression Pattern of *TcMYB29a*

To probe the expression pattern of *TcMYB29a* during the growth and development of *T. chinensis*, the expression levels of *TcMYB29* in the needles, stem epidermis, phloem, xylem, and roots were tested using the qRT-PCR. The transcription of *TcMYB29a* was high in both the needles and roots, while low in both the phloem and xylem (**Figure 4**). Except in roots, the expression patterns of *TcMYB29a* in the needles, the phloem, and the xylem are consistent with the results of its most similar TcMYB29 in the findings of Hu et al. (2020).

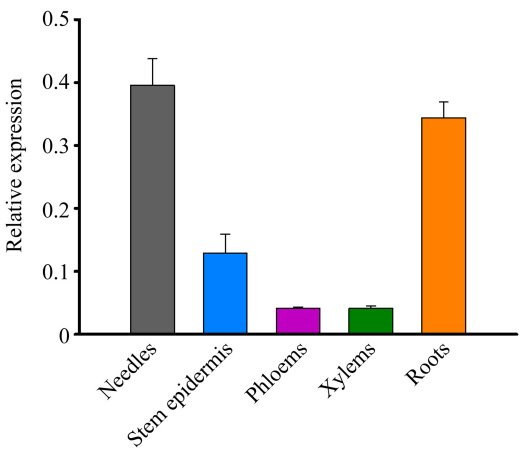


FIGURE 4 | The expression patterns of *TcMYB29a* in the different tissues of *T. chinensis*. The expression of *TcMYB29a* was compared with that of *TcGAPDH*.

TcMYB29a Binds to the Promoter of *TcT5OH*

The role of *TcMYB29a* in the transcription of taxol biosynthesis-related genes was evaluated. Based on the information about the *T. chinensis* genome and genome walking approach, most promoter sequences of taxol biosynthesis-related genes were isolated. Results showed twelve putative MREs (5′-CNGTTR-3′): two MREs in the promoter of *TcT5OH* (-286 to-291 bp and from −488 to −493 bp), three MREs in the promoter of *TcTS* (from −1,346 to −1,341 bp, from −1,312 to −1,307 bp, and from −817 bp to −812 bp), five MREs in the promoter of *TcBAPT* (from −813 to −808 bp, from −787 to −782 bp, from −431 to −426, from −409 to −404, and from −188 to −183 bp), and two MREs in the promoter of *TcDBTNBT* (from −1,097 to −1,092 bp and from −681 to −676 bp) (**Figure 5A**), indicating that TcMYB29a may regulate the expression of *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT* by directly binding with their promoters. To test this possibility, the TcMYB29a over-expressed (OE)-*Taxus* calli were collected and used for ChIP analysis. As shown in **Figure 5**, the region R-3 of *TcT5OH* promoter and R-7 of *TcBAPT* promoter resulted in 3.23-fold and 14.41-fold enrichment compared to the CDS regions of *TcT5OH* and *TcBAPT* (CDS-2 and CDS-3), respectively. While R-4 in *TcT5OH* promoter, R-1 and R-2 in *TcTS* promoter, R-5 and R-6

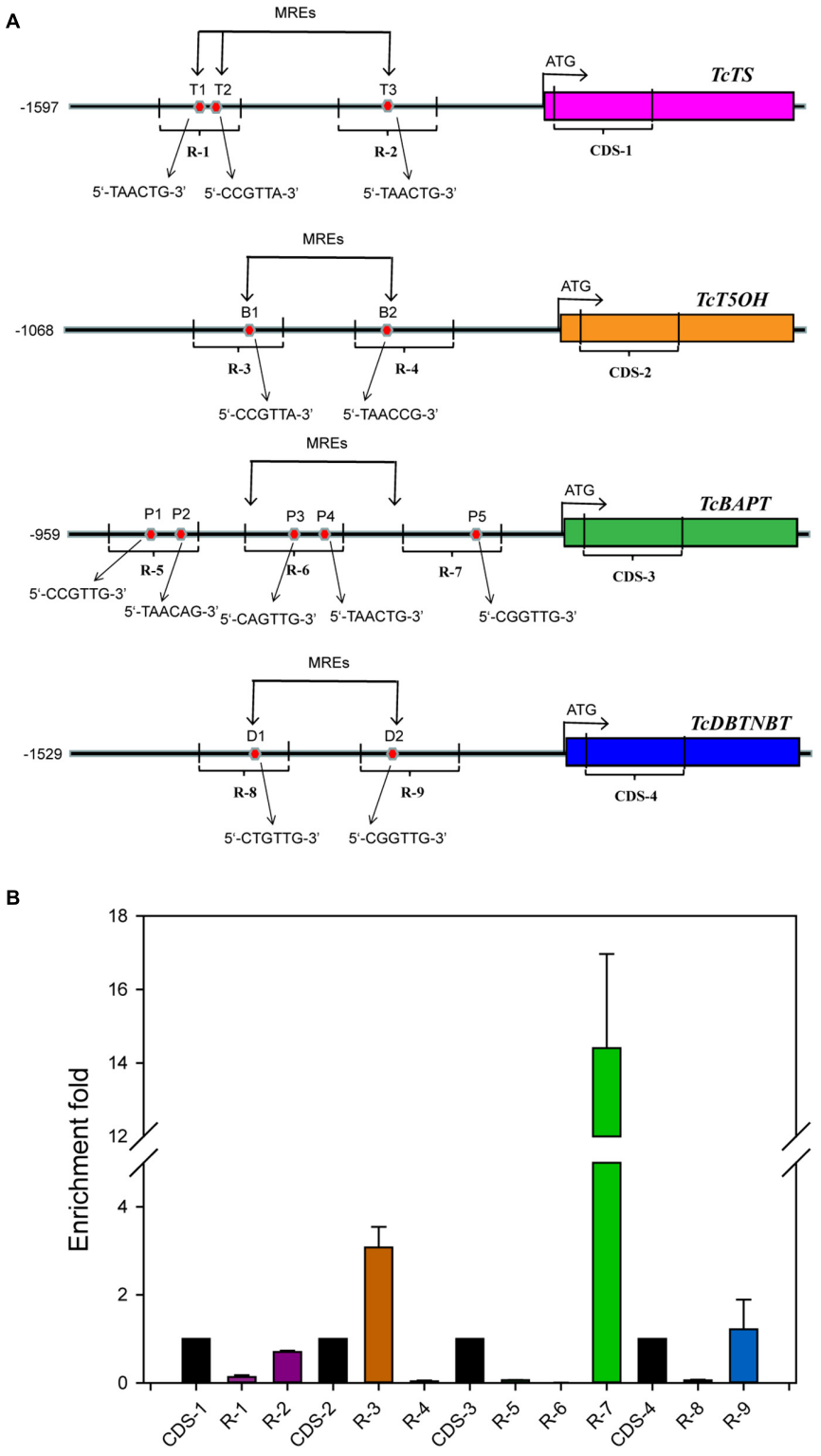


FIGURE 5 | ChIP analysis. **(A)** Schematic of ChIP analysis on *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT* promoter locus and nine amplicons. **(B)** Enrichment of particles, *TcTS*, *TcT5OH*, *TcBAPT*, and *TcDBTNBT* promoter chromatin regions with the anti-GFP antibody in TcMYB29a-OE *Taxus* cells as detected by qPCR, respectively. Error bars represent SD for three independent experiments. R-1 to R-9 represent the promoter regions including T1 to T3, B1and B2, P1 to P5, and D1 and D2, respectively.

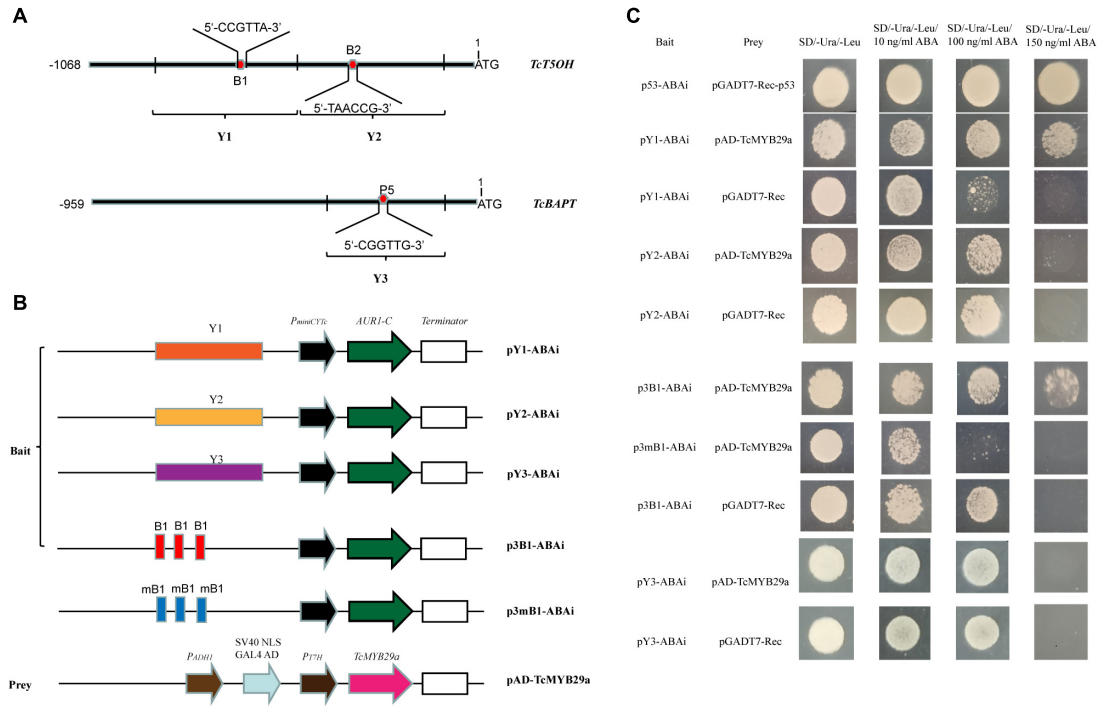


FIGURE 6 | TcMYB29a could bind with MRE in yeast. **(A)** The sketch map of the promoters of *TcT5OH* and *TcBAPT* genes, the red and blue boxes indicate MRE, and B1, B2, and P5 show the putative MREs, respectively. Y1, Y2, and Y3 represent the promoter regions, including B1, B2, and P5, respectively. **(B)** Sketch map of the prey vector and bait vectors. mB1 indicates that the MRE was mutated as shown in **Supplementary Table 1**. **(C)** The binding capability tests of TcMYB29a and *cis*-acting elements in yeast. Related bait vectors with each *cis*-acting element were cotransformed with vacant pGADT7-Rec into the Y1HGold as the control.

in *TcBAPT* promoter, and R-8 and R-9 in *TcDBTNBT* promoter did not show enrichment compared with their controls. These results suggested that TcMYB29a may bind with the promoters of *TcT5OH* and *TcBAPT*.

Binding of TcMYB29a to *TcT5OH* and *TcBAPT* promoters was further validated by the Y1H assay. As shown in **Figure 6**, the yeast Y1HGold cell with pAD-TcMYB29a and bait vector pY1-ABAi grew well on an SD/-Ura/-Leu + 150 ng/mL of ABA. While the Y1HGold cells with pGADT7-Rec and bait vector, pY1-ABAi, with pAD-TcMYB29a and bait vector, pY2-ABAi, and with pGADT7-Rec and bait vector, pY2-ABAi were unable to grow when the concentration of ABA separately reached 150 ng/mL, indicating that TcMYB29a could bind with the Y1 sequence region of the promoter of the *TcT5OH* gene. However, the YIH analysis suggested that TcMYB29a did not bind to the *TcBAPT* promoter (**Figure 6**).

Then, to further identify the direct binding site of the TcMYB29a with the *TcT5OH* promoter, treble MRE B1 was synthesized and used as the bait: Y1HGold cells containing the bait vector, p3B1-ABAi, and pAD-TcMYB29a grew well on an SD/-Ura/-Leu/150 ng/mL of ABA, while the Y1HGold cells containing the mutant B1 bait vector, p3mB1-ABAi and pAD-TcMYB29a, and p3B1-ABAi and pGADT7-Rec were unable to grow on the plates when the ABA concentration separately reached 100 ng/mL and 150 ng/mL, implying that TcMYB29a may directly bind with the B1 site in the promoter of the *TcT5OH* gene. Furthermore, EMSA was also used to assess the binding between TcMYB29a and the TcT5OH promoter *in vitro*. Based on the EMSA, the TcMYB29a binds to the B1 site in the *TcT5OH* promoter (**Figure 7**). Both *in vivo* and *in vitro* results expounded that TcMYB29a could bind with the MRE B1 in the promoter of *TcT5OH*, indicating that *TcT5OH* might be a downstream target of TcMYB2a.

TcMYB29a Activates the Expression of *TcT5OH*

Dual-luciferase reporter assays were performed to validate the effect of TcMYB29a on the promoter activity of *TcT5OH*. As shown in **Figure 8**, the cotransformation of TcT5OHpro:Luc and 35Spro:TcMYB29a exhibited significantly higher luciferase activities than the control. However, the cotransformation of 35Spro:TcMYB29a and mTcT5OHpro:Luc showed significantly lower luciferase activities compared to that of 35Spro:TcMYB29a and TcT5OHpro:Luc. These results suggested that TcMYB29a significantly influences the promoter activity of *TcT5OH*, and the MEB B1 is an important *cis*-acting element for the activity of the *TcT5OH* promoter by way of TcMYB29a (**Figure 8**).

Overexpression of TcMYB29a Promotes Taxol Accumulation in *Taxus* Calli

To corroborate the function of *TcMYB29a* in the process of taxol biosynthesis, 35S:GFP-TcMYB29a and its related empty plasmid 35S:GFP (set as a control) were introduced into *Taxus* calli. Three

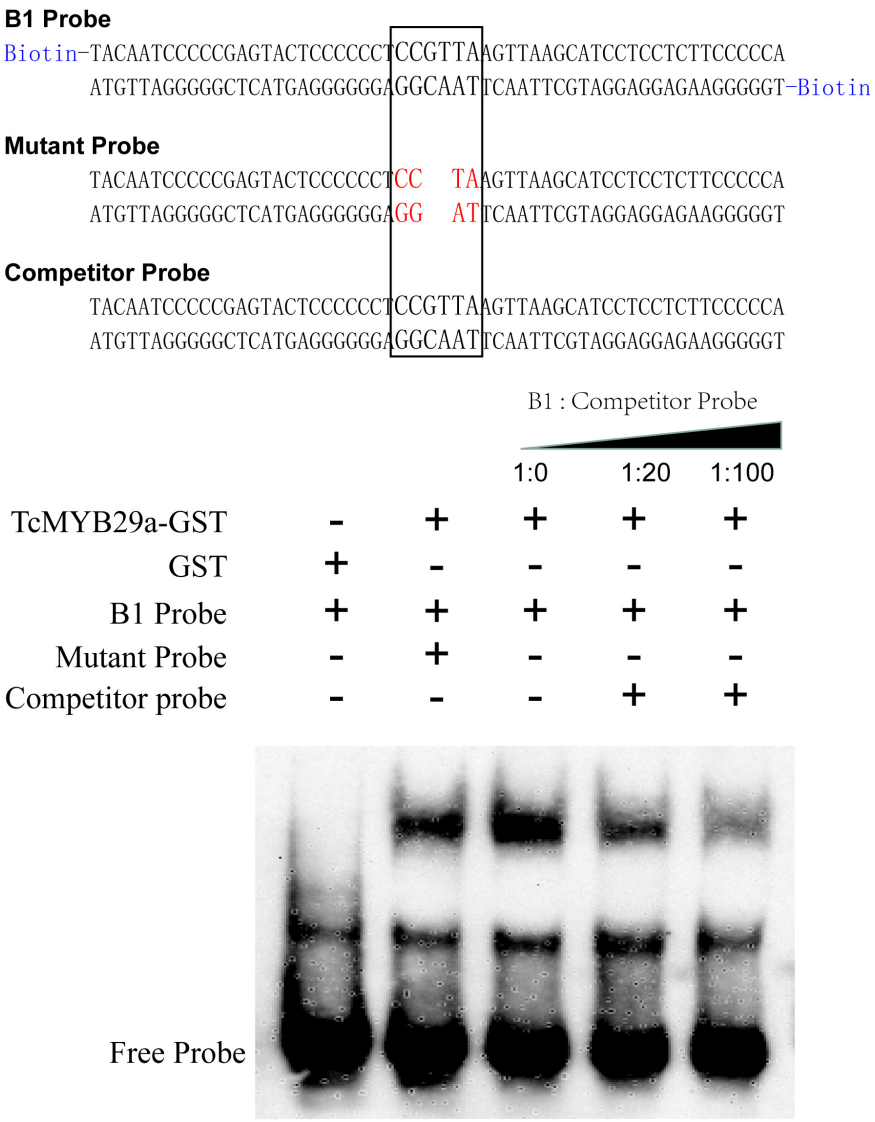


FIGURE 7 | The electrophoretic mobility shift assay **(**EMSA) showing the interaction between TcMYB29a and the *TcT5OH* promoter.

independent transgenic lines of both *TcMYB29a* and the control were chosen for taxane analysis using LC-MS. The contents of taxol and its main precursors (10-DAB and baccatin III) in the OE-calli were higher than those in the control samples. The contents of 10-DAB, baccatin III, and taxol separately rose by 238% from 0.06 ± 0.006 to 0.143 ± 0.31 μg/g of dry weight (DW), 400% from 0.04 ± 0.017 to 0.16 ± 0.006 μg/g of DW, and 419% from 0.313 ± 0.111 to 1.31 ± 0.111 μg/g of DW compared to the control group ($p < 0.01$) (**Figure 9A** and **Supplementary Figure 2**). However, RNAi interference showed no significant effect on the taxol biosynthesis in *T. chinensis* calli (**Supplementary Figure 3**) compared to the control, which may be due to the complex regulatory network for taxol biosynthesis.

Additionally, qRT-PCR analysis of taxol biosynthesis pathway genes showed that the expression of *TS*, *T50H*, *T7OH*, *DBTNBT*, *T10OH*, *T2OH*, and *PAM* were significantly upregulated in TcMYB29a-OE calli, while the expression of *BAPT* had no significant change (**Figure 9B**). These results suggested that TcMYB29a could activate most genes of the taxol biosynthesis pathway and significantly improved the taxol biosynthesis in *T. chinensis* calli.

Abscisic Acid Improved *TcMYB29a* Gene Expression

Both ABA alone and ABA pretreatment of *Trollius yunnanensis* cell suspension cultures before hot stress-induction could significantly enhance taxol yield (Zhang and Fevereiro, 2010). The sequence analysis revealed that the *TcMYB29a* promoter contains a potential ABA regulatory element. To assess whether the expression of *TcMYB29a* is a response to ABA in *T. chinensis* cells, an RT-qPCR was conducted (**Figure 10A**). Six hours

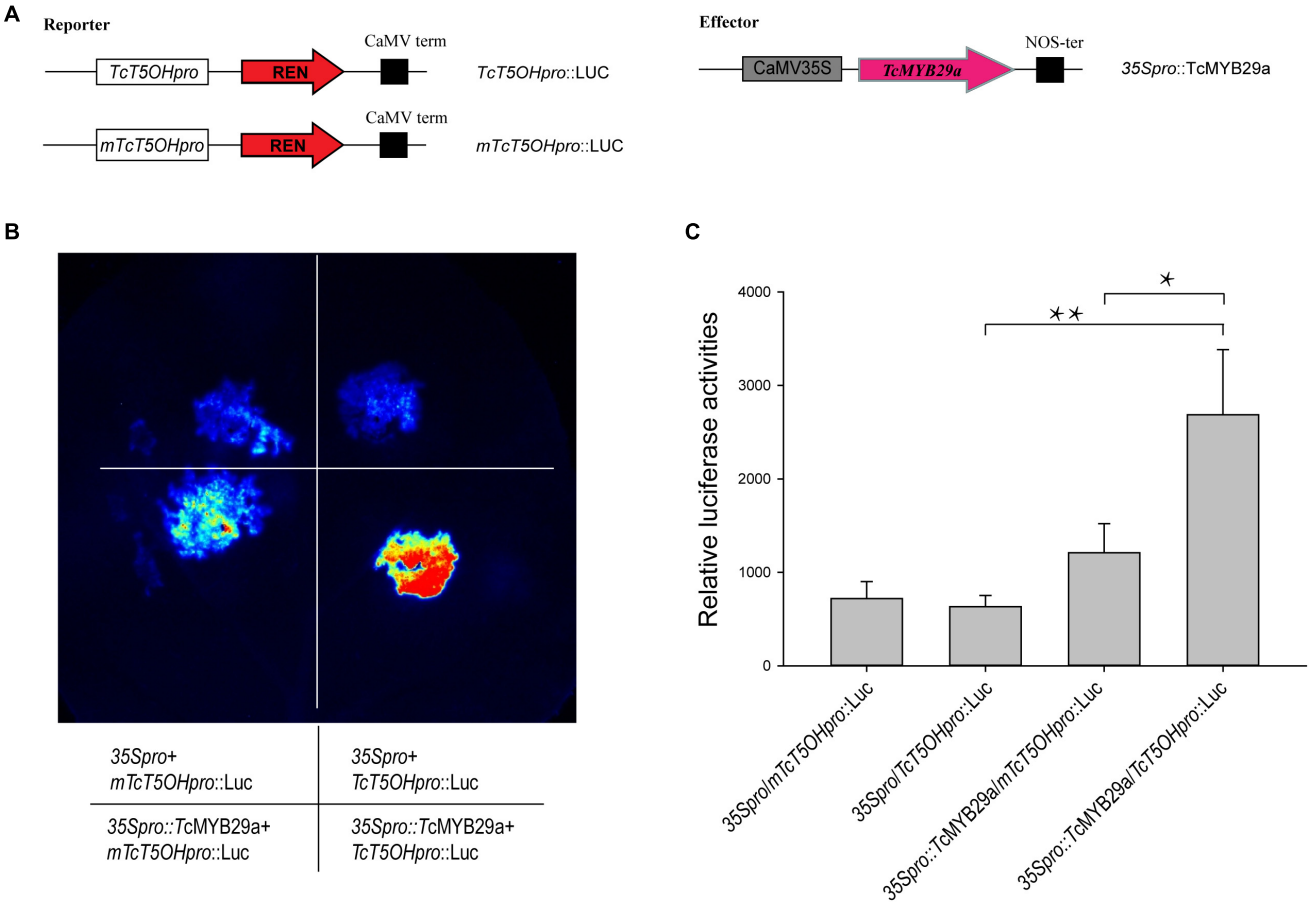


FIGURE 8 | Transcriptional activation ability of TcMYB29a. **(A)** Schematic view of the plasmid combinations of dual-LUC reporters and effector. The promoter fragments of *TcT5OH* and its mutant, *mTcT5OHpro*, were cloned into the pGreenII 0800-LUC vector to generate the reporter constructs. The effector was generated by recombining the *TcMYB29a* gene into the pHB-GFP vector. LUC, firefly luciferase; REN, Renilla luciferase. **(B)** Effects of TcMYB29a on the promoter activity of *TcT5OH* as demonstrated by luciferase reporter assay. TcMYB29a significantly affected the promoter activity of *TcT5OH*. Deletion of two nucleotides in the sequence of the *mTcT5OH* promoter (the MRE B1 site in *TcT5OH* promoter CCGTTA was present in the mutant form as TACC). **(C)** Quantitative analysis of luminescence intensity. Three biological replicates were performed. The *p*-values were evaluated using Student's *t*-test. Stars indicate the level of significance, $*0.01 < p < 0.05$, and $**p < 0.01$.

after the ABA treatment, the expression of *TcMYB29a* was significantly upregulated (8.75-fold compared to the control). However, the expression of *TcMYB29a* was downregulated at 24 h and returned to its initial level some 72 h after the ABA-treatment. In conclusion, *TcMYB29a* was an early response evincing the expression, having been improved by ABA signaling, thereby suggesting that TcMYB29a may participate in the regulation of taxol biosynthesis *via* an ABA-mediated pathway for improving the taxol biosynthesis (**Figure 10A**).

Methyl Jasmonate Suppressed *TcMYB29a* Gene Expression

Methyl jasmonate is the most effective elicitor for taxol biosynthesis in *Taxus* suspended cells (Zhang M. et al., 2017). Most of the reported TFs involved in the taxol biosynthesis are JA-induced (Li et al., 2012, 2013; Lenka et al., 2015; Zhang et al., 2015, 2018). To explore the expression pattern of *TcMYB29a* in response to MeJA in *T. chinensis*, an RT-qPCR was performed (**Figure 10B**). In the initial stage of the MeJA treatment, the expression of *TcMYB29a* had only a very weak upregulation; however, after 3 h of the MeJA treatment, the expression of *TcMYB29a* was decreased significantly, remaining at 8.97-fold downregulation after 12 h. These results indicated that *TcMYB29a* had a later response to MeJA elicitor, and its expression was suppressed by MeJA signaling, thereby suggesting that TcMYB29a may participate in the regulation of taxol biosynthesis *via* a pathway that is different from the JA-mediated signaling pathway for improving the taxol biosynthesis.

DISCUSSION

Taxol, a microtubule-stabilizing drug widely used for treating various cancers, is a diterpenoid mainly isolated from the *Taxus* spp. However, the content of taxol in *Taxus* spp. is especially low

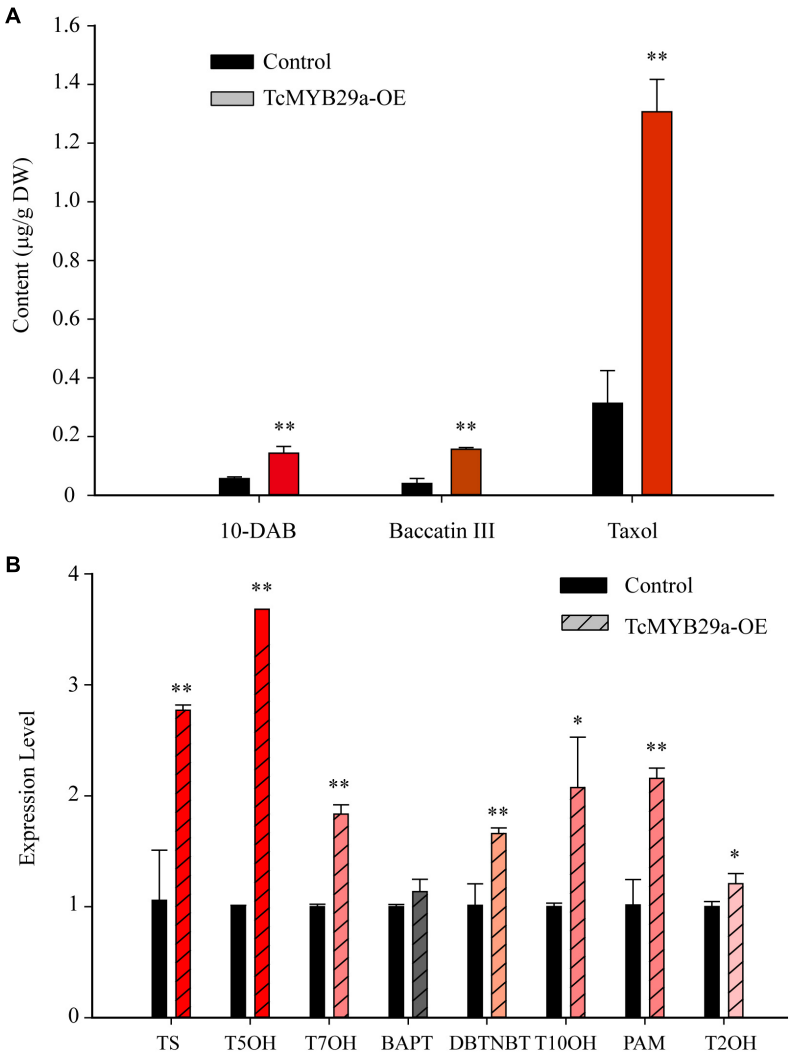


FIGURE 9 | Taxane biosynthesis and taxol biosynthesis-related genes were upregulated in TcMYB29a-OE cells. **(A)** The liquid chromatography-mass spectrometry **(**LC-MS) quantified the contents of taxanes in the control and TcMYB29a-OE cells. **(B)** The qRT-PCR analysis of the expression of taxol biosynthesis-related genes in the control and in the TcMYB29a-OE cells. *TcGAPDH* was set as a control gene, and each experiment was conducted with three biological replicates. The *p*-values were evaluated using Student's *t*-test. Stars indicate the level of significance, $^{**}p < 0.01$, and $^{*}0.01 < p < 0.05$. *TS*, taxadiene synthase; *T5OH,* taxadiene 5-alpha hydroxylase; *T7OH,* taxane 7-beta hydroxylase; *TcBAPT,* phenylpropanoyltransferase; *DBTNBT,* 3′-N-debenzoyl-2′-deoxytaxol-N-benzoyltransferase; *T10OH*, 5-alpha-taxadienol-10-beta-hydroxylase; *PAM,* phenylalanine aminomutase.

and expensive to synthesize (Cragg et al., 1993; Wickremesinhe and Arteca, 1994). A deep understanding of the regulatory mechanism of the taxol biosynthesis pathway is helpful to improve the yield of taxol in *T. chinensis* by metabolic engineering (Wilson and Roberts, 2014).

Transcription factors play important roles in the biosynthesis of secondary metabolites. In the *Taxus* species, several TFs, including WRKYs and ERFs in *T. chinensis*, MYCs in *T. cuspidata* and *T. chinensis*, and MYB in *T. media*, were proved as being involved in the taxol biosynthesis pathway (Li et al., 2013; Lenka et al., 2015; Zhang et al., 2015, 2018; Cui et al., 2019; Yu et al., 2021). Although MYB family genes have been characterized from different plants, little is known about its specific functions and mechanisms on the regulation of taxol biosynthesis in *Taxus* (yew). Several previous studies based on transcriptome sequencing and comparative analysis suggest that MYBs may be involved in regulating the taxol biosynthesis in *Taxus* (Li et al., 2012; Wang et al., 2019). However, the regulatory mechanism of those MYBs in the synthesis of taxol is not further explored. Recently, a phloem-specific R2R3-MYB TF, *TmMYB3*, was isolated from *T. media* and was confirmed to act as a transcriptional activator in the taxol biosynthesis, through binding directly to the promoters of *TmTBT* and *TmTS* genes (Yu et al., 2021). Also, according to previous research, an R2R3-MYB TF (named TcMYB29a) was isolated. The expression of *TcMYB29a* in the needles of *T. chinensis* was upregulated after an endophytic fungus elicitor treatment (Cao et al., 2022). This elicitor treatment improved the accumulation of taxol in

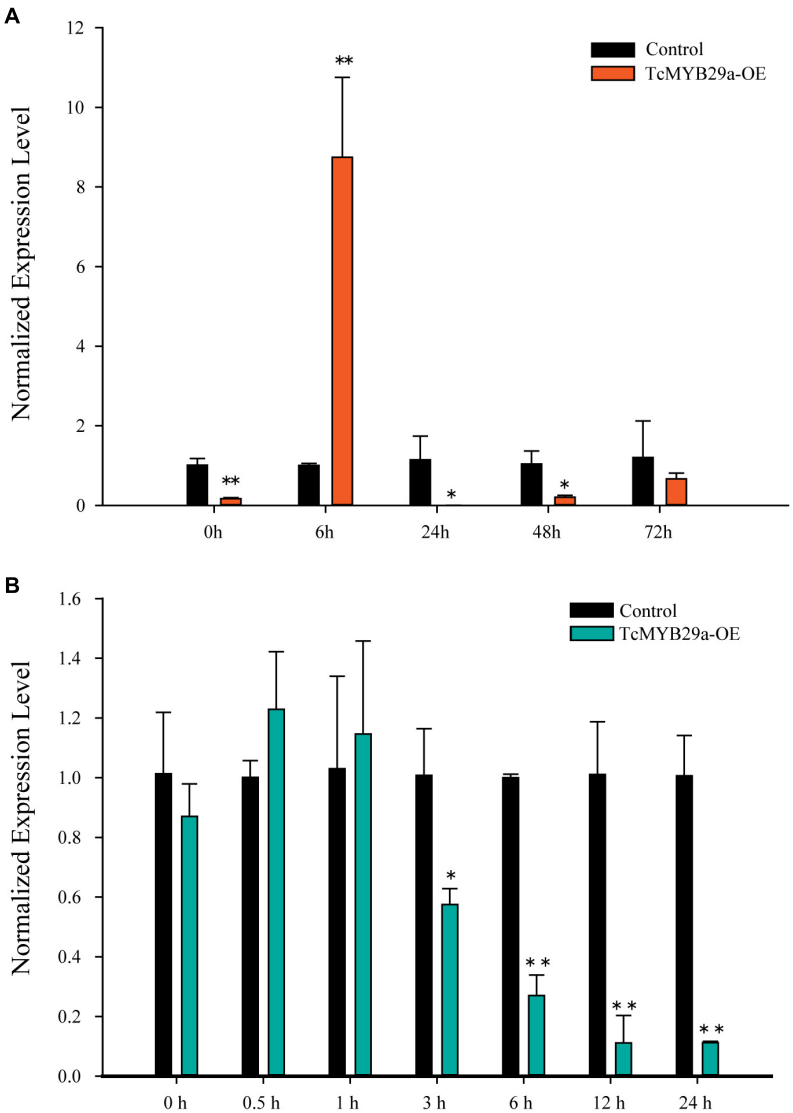


FIGURE 10 | Expression patterns of *TcMYB29a* after ABA and MeJA treatments. The suspended *T. chinensis* cells were treated with 20 μmol/L of ABA **(A)** and 50 μM of MeJA **(B)**, respectively, with the solvent ethanol set as a control; the expression levels of *TcMYB29a* were detected at 0, 6, 24, 48, and 72 h after the ABA treatment, and at 0, 0.5, 1, 3, 6, 12, and 24 h after the MeJA treatment, respectively. The ordinate indicates the normalized expression levels compared to the sample with the added solvent. The *TcGAPDH* was set as a control gene; each experiment was conducted with three biological replicates. The *p*-values were evaluated using the Student's *t*-test. Stars indicate the level of significance, $^{*}0.01 < p < 0.05$, and $^{**}p < 0.01$.

T. chinensis needles, implying that the *TcMYB29a* gene may be involved in the regulation of taxol biosynthesis. Overexpression of *TcMYB29a* in *Taxus* cells could upregulate the expression of most taxol biosynthesis-related genes, especially the *TS* and *T5OH*. An LC-MS analysis also expounded that the contents of taxol and its main precursors, 10-DAB and baccatin III, increased in the cells of TcMYB29a-OE. Subsequently, Y1H, ChIP, EMSA, and LUC-assays further demonstrated that TcMYB29a can regulate the downstream gene expression by binding with the B1 site within the promoter region of *TcT5OH*. All these results indicated that TcMYB29a can directly and indirectly activate the expression of taxol biosynthesis-related genes and improved the taxol biosynthesis in the *Taxus* cells (**Figure 11**).

Overexpression of *TcMYB29a* has been found to be able to significantly upregulate the expression of *TS* and *T5OH* genes during the taxol biosynthesis in the *Taxus* cells. However, ChIP results showed that TcMYB29a was not bound with the predicted MREs in the *TcTS* promoter *in vivo*, implying that there may be unknown factors mediating the indirect regulation of TcMYB29a on the *TS* gene. TS, rather than T5OH, is a rate-limiting enzyme in the taxol biosynthesis, especially in the *Taxus* culture calli in which the accumulation of baccatin III and the *TS* gene expression were very low compared to their original levels in plants (Koksal et al., 2011; Ssmsa and Mnr, 2020); so, the increasing contents of taxol and its precursors (baccatin III and 10-DAB) in the cells of TcMYB29a-OE may be mainly due to

the high expression of the *TS* gene. However, since *T5OH* was identified as a direct target of TcMYB29a, it is possible that the TcMYB29a regulator controls the taxol biosynthesis pathway at least partly by directly regulating the *T5OH* gene expression (**Figure 11**). Therefore, the factors that mediate the activation of TcMYB29a in the *TS* gene require further research.

Methyl jasmonate is confirmed to be one of the most effective inducers of taxol biosynthesis in *Taxus* suspended cell cultures (Zhang M. et al., 2017), except for the TmMYB3, which is not verified as to whether it was induced by MeJA; almost all reported TFs involved in the regulation of taxol biosynthesis are identified as having been induced by MeJA, such as TcWRKY (Li et al., 2012), MYC transcription factors, TcJAMYC1, TcJAMYC2, TcJAMYC4 (Lenka et al., 2015), TcMYC2a (Li et al., 2013; Zhang et al., 2018), TcERF12, and TcERF15 (Zhang et al., 2015). The MYC TFs are key regulators of the JA signaling pathway. This suggests that different members of the MYC family might perform different roles in the regulation of the taxol biosynthesis pathway (Lenka et al., 2015). For example, with MeJA elicitation, the transcript of *T5OH* was significantly repressed in the TcJAMYC2-OE *Taxus* cells but was not significantly regulated in the TcJAMYC1-OE and TcJAMYC4-OE *Taxus* cells (Lenka et al., 2015). In this study, the expression of *TcMYB29a* in *Taxus* callus cells was suppressed by MeJA. However, TcMYB29a is a positive regulator of *T5OH* gene in the taxol biosynthesis. Thus, further studies are needed to uncover the regulatory rules between MYC TFs and TcMYB29a in the taxol biosynthesis.

The R2R3-MYB proteins in *T. chinensis* and *A. thaliana* are classified into 36 subgroups according to their phylogenetic relationships and functions (Hu et al., 2020), of which 24 subgroups include members from *T. chinensis* and *A. thaliana*, three subgroups are specific to *T. chinensis*, while nine subgroups are specific to *A. thaliana*. The TcMYB29a, AtMYB88, and AtMYB124 were clustered in the S26 subgroup with counterparts in *A. thaliana* (Hu et al., 2020). In *A. thaliana,* AtMYB88 (FLP) and AtMYB124 proteins are extensively described as playing functions in the epidermal patterning (Lai et al., 2005), and are required for tolerating the abiotic stress (Xie et al., 2010). The loss of FLP/MYB88 function was found to make *Arabidopsis* plants more susceptible to abiotic stress, and an ABA signal may be involved in this increased sensitivity and it probably acted in the upstream of FLP/MYB88 (Xie et al., 2010). Phloem-specific TmMYB3, a recently reported R2R3-MYB TF isolated from *Taxus*, plays a role in the transcriptional regulation of taxol biosynthesis and may be important for the phloem-specific accumulation of taxol. Although both TmMYB3 and TcMYB29a function to regulate the taxol biosynthesis, they show several differences: first, they show low similarity with each other in terms of amino acid sequences (the similarity was 11.01%); second, TmMYB3 belongs to another subgroup of TcMYB29a, as TmMYB3 and TcMYB29a were separately clustered in the S4 and S26 subgroups with counterparts in *A. thaliana* (**Supplementary Figure 4**); third, the expression patterns of TmMYB3 and TcMYB29a were different, as TmMYB3 exhibited phloem-specific expression while TcMYB29a was highly expressed in the needles and the roots. Those results indicated that MYBs may evolve

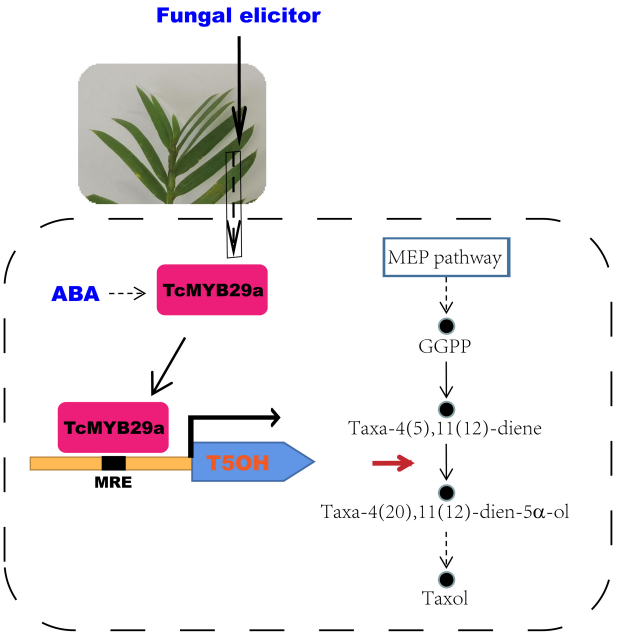


FIGURE 11 | A model for the role of TcMYB29a in the taxol biosynthesis. GGPP, geranylgeranyl diphosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; MRE. MYB recognition element.

independently in the different tissues of *Taxus* to meet the need for the regulation of spatially differential taxol biosynthesis. *TcMYB29a* was highly expressed in the needles and the roots, which show a similar expression pattern with FLP/MYB88 in *Arabidopsis* plants, and its expression was highly induced after the fungal elicitor treatment based on the previous study (Cao et al., 2022). Furthermore, the PLACE website predicted that *TcMYB29a* was induced by ABA, which was verified by qRT-PCR (**Figure 10** and **Supplementary Figure 1**). The similarities in the sequences and expression patterns between TcMYB29a and the well-characterized AtMYB88 (FLP) and AtMYB124 suggest a conserved response to ABA, although there are significant divergences between the lineages of the angiosperm and the gymnosperm. These results presented here suggest that TcMYB29a acts as a positive regulator of ABA-medicated expression of taxol biosynthesis-related genes in the *Taxus* cell cultures, and it may play its role in response to the biotic and abiotic stresses of *T. chinensis* needles and roots.

In summary, TcMYB29a was involved in the regulation of taxol biosynthesis in *T. chinensis* partly by activating the expression of the *TcT5OH* gene, and it is highly expressed in the needles and roots compared to the stem epidermis, the phloem, and the xylem. These results provided a potential explanation for the accumulation of taxol in *Taxus* needles after the fungal elicitor treatment; the expression levels of *TcMYB29a* were improved after ABA treatments but suppressed after the MeJA treatment, indicating a new ABA-mediated pathway different from the JA-mediated pathway for regulating the taxol biosynthesis. Further

studies are needed to elucidate the complex signaling network of the MYB involved in the taxol biosynthesis.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JJ, XC, and WW planned and designed the research. WW and XC wrote the manuscript. LX, XC, LL, and WW conducted the research and analyzed the data. All authors read and approved the final manuscript.

FUNDING

The work was funded by the Natural Science Foundation of China (No. 31770613) and the Basic Research Program of Xuzhou (No. KC21028).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.804593/full#supplementary-material

REFERENCES

Araki, S., Ito, M., Soyano, T., Nishihama, R., and Machida, Y. (2004). Mitotic cyclins stimulate the activity of c-Myb-like factors for transactivation of G2/M phase-specific genes in tobacco. *J. Biol. Chen.* 279, 32979–32988. doi: 10.1074/jbc.M403171200

Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: their role in drought response mechanisms. *IJMS* 16, 15811–15851. doi: 10.3390/ijms160715811

Cao, X. Y., Xu, L. X., Wang, J. Y., Dong, M. M., Xu, C. Y., Kai, G. Y., et al. (2022). Endophytic fungus *Pseudodidymocyrtis lobariellae* KL27 promotes taxol biosynthesis and accumulation in *Taxus chinensis*. *BMC Plant Biol.* 22:12. doi: 10.1186/s12870-021-03396-6

Cragg, G. M., Schepartz, S. A., Suffness, M., and Grever, M. R. (1993). The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *J. Nat. Prod.* 56, 1657–1668. doi: 10.1021/np50100a001

Croteau, R., Ketchum, R. E. B., Long, R. M., Kaspera, R., and Wildung, M. R. (2006). Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* 5, 75–97. doi: 10.1007/s11101-005-3748-2

Cui, Y., Mao, R., Chen, J., and Guo, Z. (2019). Regulation mechanism of MYC family transcription factors in jasmonic acid signalling pathway on taxol biosynthesis. *Int. J. Mol. Sci.* 20, 1843. doi: 10.3390/ijms20081843

Daniel, X., Sugano, S., and Tobin, E. M. (2004). CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in *Arabidopsis*. *PNAS* 101, 3292–3297. doi: 10.1073/pnas.0400163101

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

Gallego-Jara, J., Lozano-Terol, G., Sola-Martínez, R. A., Cánovas-Díaz, M., and de Diego Puente, T. (2020). A compressive review about Taxol®: history and future challenges. *Molecules* 25:5986. doi: 10.3390/molecules25245986

Hao, X., Pu, Z., Cao, G., You, D., and Kai, G. (2020). Tanshinone and salvianolic acid biosynthesis are regulated by SmMYB98 in *Salvia miltiorrhiza* hairy roots. *J. Adv. Res.* 23, 1–12. doi: 10.1016/j.jare.2020.01.012

Hou, D., Cheng, Z., Xie, L., Li, X., Li, J., Mu, S., et al. (2018). The R2R3 MYB gene family in *Phyllostachys edulis*: genome-wide analysis and identification of stress or development related R2R3 MYBs. *Front. Plant Sci.* 9:738. doi: 10.3389/fpls.2018.00738

Hu, X., Zhang, L., Wilson, I., Shao, F., and Qiu, D. (2020). The R2R3-MYB transcription factor family in *Taxus chinensis*: identification, characterization, expression profiling and posttranscriptional regulation analysis. *Peer J.* 8, e8473. doi: 10.7717/peerj.8473

Kayan, B., Gizir, A. M., and Kalderis, D. (2021). Ultrasonic-assisted extraction of 10-deacetylbaccatin III from *Taxus baccata* L.: optimization using response surface methodology. *J. Iran. Chem. Soc.* 18, 37–45. doi: 10.1007/s13738-020-02003-z

Koepp, A. E., Hezari, M., Zajicek, J., Vogel, B. S., LaFever, R. E., Lewis, N. G., et al. (1995). Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of taxol biosynthesis in Pacific yew. *J. Biol. Chem.* 270, 8686–8690. doi: 10.1074/jbc.270.15.8686

Koksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–120. doi: 10.1038/nature09628

Lai, L. B., Nadeau, J. A., Lucas, J., Lee, E. K., Nakagawa, T., Zhao, L., et al. (2005). The *Arabidopsis* R2R3 MYB proteins FOUR LIPS and MYB88 restrict divisions late in the stomatal cell lineage. *Plant Cell.* 17, 2754–2767. doi: 10.1105/tpc.105.034116

Lee, E. K., Jin, Y. W., Park, J. H., Yoo, Y. M., Hong, S. M., Amir, R., et al. (2010). Cultured cambial meristematic cells as a source of plant natural products. *Nat. Biotechnol.* 28, 1213–1217. doi: 10.1038/nbt.1693

Lenka, S. K., Nims, N. E., Vongpaseuth, K., Boshar, R. A., Roberts, S. C., and Walker, E. L. (2015). Jasmonate-responsive expression of paclitaxel biosynthesis genes in *Taxus cuspidata* cultured cells is negatively regulated by the bHLH transcription factors TcJAMYC1, TcJAMYC2, and TcJAMYC4. *Front. Plant Sci.* 6:115. doi: 10.3389/fpls.2015.00115

Li, S., Zhang, P., Zhang, M., Fu, C., and Yu, L. (2013). Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in Taxus chinensis. *Plant Biol.* 15, 19–26. doi: 10.1111/j.1438-8677.2012.00611.x

Li, S. T., Zhang, P., Zhang, M., Fu, C. H., Zhao, C. F., Dong, Y. S., et al. (2012). Transcriptional profile of *Taxus chinensis* cells in response to methyl jasmonate. *BMC Genomics* 13:295. doi: 10.1186/1471-2164-13-295

Martin, C., and Paz-Ares, J. (1997). MYB transcription factors in plants. *Trends Genet.* 13, 67–73. doi: 10.1016/S0168-9525(96)10049-4

Meng, L. S., Xu, M. K., Wan, W., Yu, F., Li, C., Wang, J. Y., et al. (2018). Sucrose signaling regulates anthocyanin biosynthesis through a MAPK cascade in *Arabidopsis thaliana*. *Genetics* 210, 607–619. doi: 10.1534/genetics.118.301470

Morse, A. M., Whetten, R. W., Dubos, C., and Campbell, M. M. (2009). Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytol.* 183, 1001–1013. doi: 10.1111/j.1469-8137.2009.02900.x

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., et al. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Mol. Biol.* 3, 178–187. doi: 10.1038/nsb0296-178

Ssmsa, B., and Mnr, C. (2020). Sites of biosynthesis and storage of taxol in *Taxus media* (rehder) plants: mechanism of accumulation. *Phytochemistry* 175, 112369. doi: 10.1016/j.phytochem.2020.112369

Wang, C., Ji, W., Liu, Y., Zhou, P., Meng, Y., Zhang, P., et al. (2021b). The antagonistic MYB paralogs RH1 and RH2 govern anthocyanin leaf markings in *Medicago truncatula*. *New Phytol.* 229, 3330–3344. doi: 10.1111/nph.17097

Wang, C., Deng, Y., Liu, Z., and Liao, W. (2021a). Hydrogen sulfide in plants: crosstalk with other signal molecules in response to abiotic stresses. *Int. J. Mol. Sci.* 22:12068. doi: 10.3390/ijms222112068

Wang, K., and Lindås, A. C. (2018). Chromatin immunoprecipitation assay in the hyperthermoacidophilic crenarchaeon, *Sulfolobus acidocaldarius*. *Methods Mol. Biol.* 1689, 139–146. doi: 10.1007/978-1-4939-7380-4_12

Wang, T., Chen, Y., Zhuang, W., Zhang, F., Shu, X., Wang, Z., et al. (2019). Transcriptome sequencing reveals regulatory mechanisms of taxol synthesis in *Taxus wallichiana* var. *Mairei*. *Int. J. Genomics* 2019:1596895. doi: 10.1155/2019/1596895

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., and McPhail, A. T. (1971). Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chen. Soc.* 93, 2325–2327. doi: 10.1021/ja00738a045

Wickremesinhe, E. R., and Arteca, R. N. (1994). Roots of hydroponically grown *Taxus* plants as a source of taxol and related taxanes. *Plant Sci.* 101, 125–135. doi: 10.1016/0168-9452(94)90248-8

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. (2009). Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol.* 149, 981–993. doi: 10.1104/pp.108.132795

Wilson, S. A., and Roberts, S. C. (2014). Metabolic engineering approaches for production of biochemicals in food and medicinal plants. *Curr. Opin. Biotech.* 26, 174–182. doi: 10.1016/j.copbio.2014.01.006

Xie, Z., Li, D., Wang, L., Sack, F. D., and Grotewold, E. (2010). Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. *Plant J.* 64, 731–739. doi: 10.1111/j.1365-313X.2010.04364.x

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., et al. (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol.* 60, 107–124. doi: 10.1007/s11103-005-2910-y

Yin, J., Sun, L., Li, Y., Xiao, J., and Zhan, Y. (2020). Functional identification of BpMYB21 and BpMYB61 transcription factors responding to MeJA and SA in birch triterpenoid synthesis. *BMC Plant Biol.* 20:374. doi: 10.1186/s12870-020-02521-1

Yu, C., Luo, X., Zhang, C., Xu, X., Huang, J., Chen, Y., et al. (2020). Tissue-specific study across the stem of *Taxus* media identifies a phloem-specific TmMYB3 involved in the transcriptional regulation of paclitaxel biosynthesis. *Plant J.* 103, 1.

Yu, C., Zhang, C., Xu, X., Huang, J., Chen, Y., Luo, X., et al. (2021). Omic analysis of the endangered Taxaceae species *Pseudotaxus chienii* revealed the differences in taxol biosynthesis pathway between *Pseudotaxus* and *Taxus yunnanensis* trees. *BMC Plant Biol.* 21:104. doi: 10.1186/s12870-021-02883-0

Zhang, C., and Fevereiro, P. S. (2010). The effect of heat shock on paclitaxel production in *Taxus yunnanensis* cell suspension cultures: role of abscisic acid pretreatment. *Biotechnol. Bioeng.* 96, 506–514. doi: 10.1002/bit.21122

Zhang, J., Zhou, L., Zheng, X., Zhang, J., Yang, L., Tan, R., et al. (2017). Overexpression of SmMYB9b enhances tanshinone concentration in *Salvia miltiorrhiza* hairy roots. *Plant Cell Rep.* 36, 1297–1309. doi: 10.1007/s00299-017-2154-8

Zhang, K., Fan, W., Chen, D., Jiang, L., Li, Y., Yao, Z., et al. (2020). Selection and validation of reference genes for quantitative gene expression normalization in *Taxus* spp. *Sci. Rep.* 10:22205. doi: 10.1038/s41598-020-79213-1

Zhang, M., Chen, Y., Nie, L., Jin, X., Fu, C., and Yu, L. (2017). Molecular, structural, and phylogenetic analyses of *Taxus chinensis* JAZs. *Gene* 620, 66–74. doi: 10.1016/j.gene.2017.04.005

Zhang, M., Chen, Y., Nie, L., Jin, X., Liao, W., Zhao, S., et al. (2018). Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis*. *Sci. Rep.* 8:5197. doi: 10.1038/s41598-018-23558-1

Zhang, M., Li, S., Nie, L., Chen, Q., Xu, X., Yu, L., et al. (2015). Two jasmonate-responsive factors, TcERF12 and TcERF15, respectively act as repressor and activator of tasy gene of taxol biosynthesis in *Taxus chinensis*. *Plant Mol. Biol.* 89, 463–473. doi: 10.1007/s11103-015-0382-2

Zhou, A., Sun, H., Feng, S., Zhou, M., Gong, S., Wang, J., et al. (2018). A novel cold-regulated gene from *Phlox subulata*, PsCor413im1, enhances low temperature tolerance in *Arabidopsis*. *Biochem. Bioph. Res. Commun.* 495, 1688–1694. doi: 10.1016/j.bbrc.2017.12.042

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fpls.2022.860033

Overexpression of *SmSCR1* Promotes Tanshinone Accumulation and Hairy Root Growth in *Salvia miltiorrhiza*

*Wei Zhou[1†], Shuai Wang[1†], Yafang Shen[1†], Yunhui Liu[1], Itay Maoz[2], Xiankui Gao[1], Chengan Chen[1], Tingyao Liu[1], Can Wang[1] and Guoyin Kai[1]**

[1] Laboratory for Core Technology of Traditional Chinese Medicine (TCM) Quality Improvement and Transformation, School of Pharmaceutical Sciences, The Third Affiliated Hospital, School of Pharmacy and Academy of Chinese Medical Science, Zhejiang Chinese Medical University, Hangzhou, China, [2] Department of Postharvest Science, Agricultural Research Organization, Volcani Center, Rishon LeZion, Israel

Lipid-soluble tanshinone is one of the main bioactive substances in the medicinal plant *Salvia miltiorrhiza*, and its medicinal demand is growing rapidly. Yeast extract (YE) modulates the tanshinone biosynthesis, but the underlying regulatory network remains obscure. In this study, a YE-responsive transcriptional factor Scarecrow1 (*SCR1*) was identified in *S. miltiorrhiza* from the YE-induced transcriptome dataset. *SmSCR1* is located in the nucleus. Overexpression of *SmSCR1* in *S. miltiorrhiza* roots resulted in a significantly higher accumulation of tanshinone than the control, with the highest 1.49-fold increase. We also detected upregulation of tanshinone biosynthetic genes, *SmSCR1* and *SmHMGR1*, and distinct alteration of growth and development of the hairy roots in the overexpression lines compared to the control. An inverse phenotype was observed in *SmSCR1*-SRDX suppression expression lines. We found that *SmSCR1* can bind to the promoter of *SmCPS1* to induce its expression. This study provides new insight into the regulatory mechanism on the growth and development of hairy roots, tanshinone accumulation, and the metabolic engineering of bioactive compounds in *S. miltiorrhiza*.

Keywords: *Salvia miltiorrhiza*, GRAS transcription factor, tanshinone accumulation, hairy root growth, regulatory mechanism

OPEN ACCESS

Edited by:
Fangyuan Zhang, Southwest University, China

Reviewed by:
Xiaojian Yin, China Pharmaceutical University, China
Qinlong Zhu, South China Agricultural University, China

***Correspondence:**
Guoyin Kai
guoyinkai1@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *22 January 2022*
Accepted: *08 February 2022*
Published: *08 March 2022*

Citation:
Zhou W, Wang S, Shen Y, Liu Y, Maoz I, Gao X, Chen C, Liu T, Wang C and Kai G (2022) Overexpression of SmSCR1 Promotes Tanshinone Accumulation and Hairy Root Growth in Salvia miltiorrhiza. Front. Plant Sci. 13:860033. doi: 10.3389/fpls.2022.860033

INTRODUCTION

Salvia miltiorrhiza Bunge is a Chinese herbal medicine that belongs to the *Salvia* genus in the Lamiaceae family (Jia et al., 2019; Jung et al., 2020). In Asian countries, dried roots and stems of *S. miltiorrhiza* are a common treatment for cardiovascular system-related diseases (Hao et al., 2015). The bioactivity of *S. miltiorrhiza* is associated with lipid-soluble tanshinones, including tanshinone I (TA-I), cryptotanshinone (CT), dihydrotanshinone (DT), and tanshinone IIA (TA-II) (Fu et al., 2020). Increasing attention has been given to the abovementioned compounds in recent years, mainly focusing on improving the yield of tanshinone in *S. miltiorrhiza*.

Abbreviations: CPS1, copalyl diphosphate synthase I; HMGR1, 3-hydroxy-3-methylglutaryl-CoA-reductase I; GGPPS1, geranylgeranyl diphosphate synthase I; HMGS, hydroxymethylglutaryl-CoA synthase; IPPI, isopentenyl diphosphate isomerase; KSL, ent-kaurene synthase like; CPS1, copalyl diphosphate synthase I; CYP76AH1, miltiradiene oxidase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS, deoxy-D-xylulose 5-phosphate synthase II; Dual-LUC, dual-luciferase; YE, yeast extract; GRAS1, GRAS transcriptional factor 1; DT, dihydrotanshinone; CT, cryptotanshinone; TA-I, tanshinone I; TA-II, tanshinone IIA; TTA, total tanshinone.

Tanshinones are a type of diterpenoids produced in two phases in *S. miltiorrhiza* roots **(Supplementary Figure 1)**. First, common terpenoid precursors [e.g., isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)] are generated by two distinct processes (e.g., mevalonate (MVA) pathway in cytosol and 2-C-Methyl-D-erythritol-4-phosphate (MEP) pathway in plastids) (Ma et al., 2012). The universal diterpenoid precursor, geranylgeranyl diphosphate (GGPP), is then biosynthesized GGPP synthase (GGPPS) (Shi et al., 2016a,b). Then, three known synthases, namely, copalyl diphosphate synthase 1 (CPS1), kaurene synthase-like 1 (KSL1), and miltiradiene oxidase (CYP76AH1), and a still-unidentified enzyme(s) produce tanshinones (Gao et al., 2009; Xing et al., 2018). Overexpression or downregulation of one or two synthases significantly impacts tanshinone synthesis (Kai et al., 2011; Wang et al., 2015; Ma et al., 2016). Plant transcription factors (TFs) play important regulatory roles in resistance to stress and metabolic engineering. TFs can regulate one or multiple biosynthetic genes from a single or multiple pathways (Xu et al., 2016; Wang et al., 2017; Yang et al., 2017). TFs are a diverse group of genes including many families and subfamilies, such as bHLH, MYB, ERF, bZIP, and GRAS (Deng et al., 2020a,b; Hao et al., 2020; Shi et al., 2021a,b; Zhou et al., 2021a,b). Less is known about the regulatory actions or target genes of the GRAS TFs, which play a critical role in the tanshinone biosynthetic pathway in *S. miltiorrhiza*.

The GRAS TF family has been discovered to be unique in plants. GRAS TFs consist of three gene members, namely, gibberellic acid insensitive (GAI), repressor of GA1-3 mutant (RGA), and Scarecrow (*SCR*). A whole-genome analysis from *Arabidopsis thaliana*, rice (*Oryza sativa*), and musk lily (*Lilium longiflorum*) grouped the GRAS families into ten distinctive subfamilies, namely, *SCR*, short-root (*SHR*), *L. longiflorum* SCR-like (*LISCL*), *DELLA*, SCL-like 3 (*SCL3*), SCL-like 4 and SCL-like 7 (*SCL4/7*), *A. thaliana* Lateral suppressor (*AtLAS*), hairy meristem (*HAM*), *A. thaliana* phytochrome A signal transduction 1 (*AtPAT1*), and dwarf and low tillering (*DLT*) (Sun et al., 2011). The SCR and SHR subfamilies were validated to act as positive regulators to promote the formation of root radial morphology and root monolayer cells in *Arabidopsis* (Koizumi et al., 2012; Rich et al., 2017). SCR is mainly expressed in root epidermal cells regulating the division of daughter cells differentiation in root, whereas SHR not only affects the polar differentiation of root endothelial cells but also activates the SCR promoter in specific tissues to regulate its function (Slewinski et al., 2012). In addition, SHR proteins are expressed explicitly in the mid-column sheath cells. They can transfer from external monolayers to endothelial cells to form a dimer with SCR, which jointly activates the transcription of the *SCR* gene and other downstream target genes. Therefore, it can be inferred that SCR and SHR are transcriptionally dependent on each other, and the resulting SCR-SHR complex positively modulates the transcription of SCR (Koizumi et al., 2012; Slewinski et al., 2012). Yeast extract (YE), also known as yeast flavoring, plays a vital role in promoting the accumulation of secondary metabolites in plants. In *S. miltiorrhiza*, after treating hairy root culture with YE, the content of tanshinone was significantly increased (Chen et al., 2001). However, the transcriptional mechanism involved in YE-triggered tanshinone biosynthesis remains obscure.

In this study, we isolated and identified a new GRAS TF named *SmSCR1*. Our studies uncover the regulatory role of *SmSCR1* gene in response to YE-induced regulation of tanshinone biosynthesis and hairy root growth and development. Our data discover a novel TF family, GRAS, as a candidate for improving the biosynthesis potential of tanshinone through metabolic engineering strategies.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Salvia miltiorrhiza plants were grown in solid Murashige and Skoog (MS) medium at 25°C, in 60% humidity, and in a light rhythm of 16-h light and 8-h dark (Shi et al., 2021a,b). *S. miltiorrhiza* hairy roots were subcultured in 250 ml Erlenmeyer flasks with 100 ml 1/2 MS liquid medium and placed in a shaker at 110 rpm min^{-1}, 25°C in the dark. The hairy roots were collected after 60 days for gene expression analysis (Shi et al., 2016a, 2020). *Nicotiana benthamiana* was directly sown in soil and cultivated in pots for further experimental needs (Huang et al., 2019). RNA was extracted from different tissues of annual plants of *S. miltiorrhiza* and stored at −80°C using liquid nitrogen flash freezing.

Gene Isolation and Sequence Analysis

All the GRAS families in *S. miltiorrihiza* were identified against local datasets. *SmSCR1*, which belongs to the GRAS family and intensively responds to YE induction, was cloned using a homology-based cloning method, as described previously (Zhou et al., 2016, 2017). The cloned *SmSCR1* gene was analyzed using the BlastP tool in the NCBI database. The reported GRAS members, i.e., SmGRAS1 (accession number: KY435886) and SmGRAS2 (KY435887) in *S. miltiorrihiza*, together with *O. sativa* OsGRAS32 (Os06g0127800), *O. sativa* OsSCR1 (XP_015615402.1), *O. sativa* OsSCR2 (XP_015620600.1), *A. thaliana* AtSCR (Q9M384.1), and *A. thaliana* AtSCL23 (NM_123557), which are highly homologous to *SmSCR1* (OM032820), were subjected to amino acid sequence alignment using Vector NTI software (Invitrogen, Carlsbad, CA, United States). The full-length amino acid sequences of the GRAS were aligned using the neighbor-joining (NJ) method in the Clustal X computer, and the MEGA 6.0 program was used to generate a phylogenetic tree. Using Clustal X computing tools, the amino acid sequences of candidate GRAS family members were aligned based on the NJ method, and MEGA 6.0 software was used to generate phylogenetic trees. Subfamily members of SCR and SCL from *Arabidopsis* and rice were compared with *SmSCR1* using Vector NTI 10.0 software.

Subcellular Localization Analysis of *SmSCR1*

To verify the subcellular localization profile of *SmSCR1*, pHB-*SmSCR1-YFP*, and pHB-*YFP* recombinant plasmids were

introduced into *Agrobacterium rhizogenes* strain EHA105 for transient transformation, respectively. pHB-YFP was used as the positive control. After 48 h of infection by *Agrobacterium tumefaciens*, YFP signals from infected *N. benthamiana* leaves were visualized using a high-resolution microscope observation system. The nuclei of epidermal cells of infected *N. benthamiana* leaves were stained with 4′, 6-diamidino-2-phenylindole dihydrochloride (DAPI) solution for 3 h before observation (Zhou et al., 2020).

Quantitative Real-Time PCR

According to the procedure reported before (Shi et al., 2016b), the quantitative real-time PCR (qRT-PCR) analysis was conducted. Different tissues (e.g., taproot, lateral root, fibrous root, phloem, xylem, stem, petiole, young leaves, and mature leaves) were taken from an adult plant, as well as hairy roots with the treatment of YE, and then frozen in liquid nitrogen for RNA isolation (Zhang et al., 2011). *Actin* gene of *S. miltiorrhiza* was used as the internal reference. The sequences of all primer pairs used in qRT-PCR studies are summarized in **Supplementary Table 1**. The gene expression was quantified by the comparative *Ct* method (Hao et al., 2015).

Transformation of *SmSCR1* in *Salvia miltiorrhiza*

The open reading frame (ORF) of *SmSCR1* was introduced into the double restriction insertion site of *Bam*HI and *Spe*I of the pHB vector under the CaMV 35S promoter and NOS terminator (**Supplementary Figure 2A**). The DNA sequence encoding the SRDX structural domain (LDLDLELRLGFA) was inserted into the ORF of *SmSCR1* just before the stop codon (TAA) using the method described previously (Hiratsu et al., 2003; Deng et al., 2020a,b). The SmSCR1-SRDX was then subcloned into the pHB to create the pHB-SmSCR1-SRDX suppressive expression construct (**Supplementary Figure 2B**). To obtain the transgenic hairy roots, all plasmids were transformed into

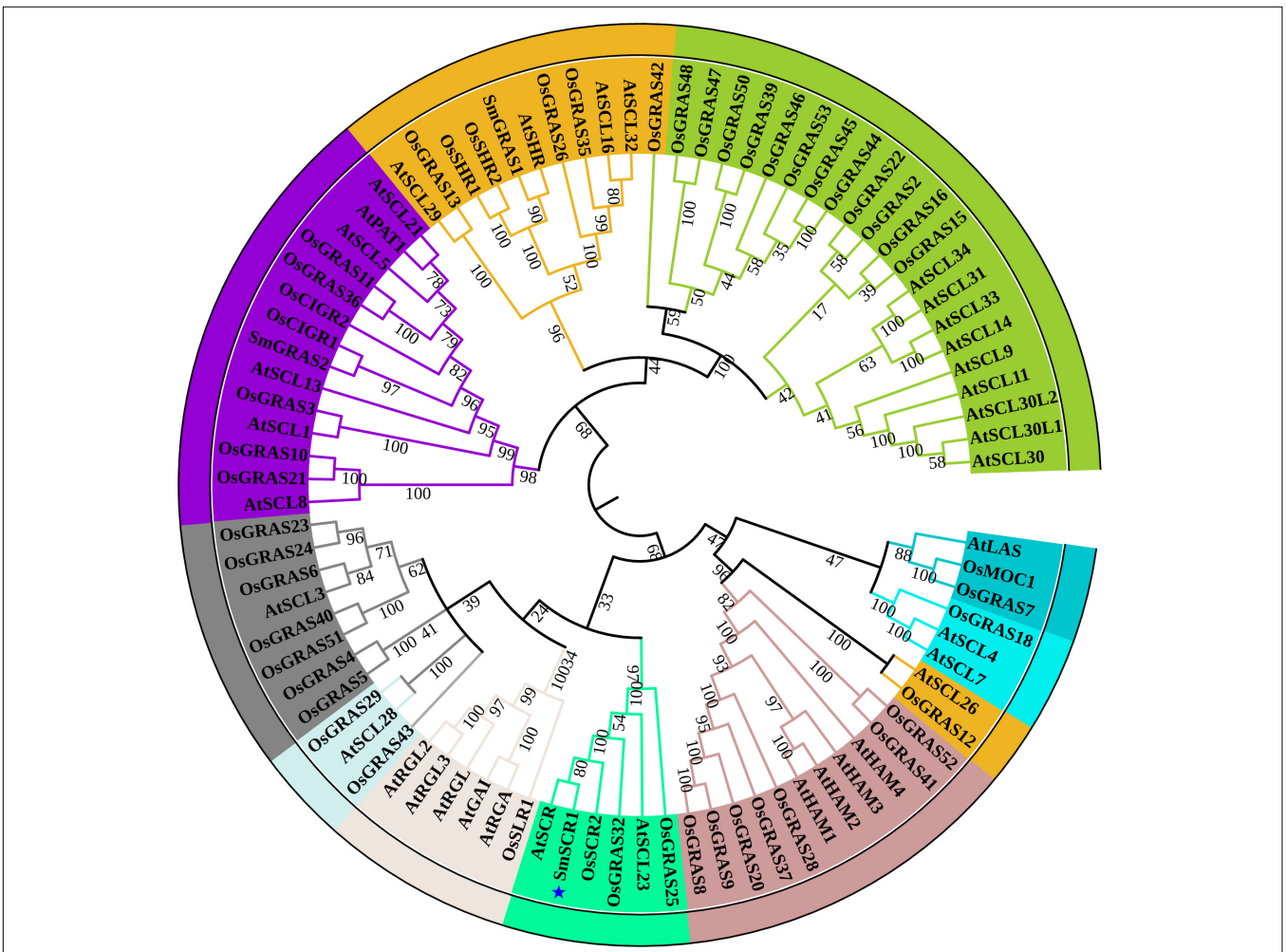


FIGURE 1 | Bioinformatics analysis of *SmSCR1*. Phylogenic tree analysis of *SmSCR1* and GRAS transcription factors from different species. The MEGA 6.05 software was employed to construct the phylogenetic tree with the amino acid sequences of *SmSCR1* in *Salvia miltiorrhiza* and 89 GRAS family members in *Arabidopsis* and rice using the neighbor-joining approach. The bootstrap values after 1,000 replicates are indicated by the numbers on the nodes.

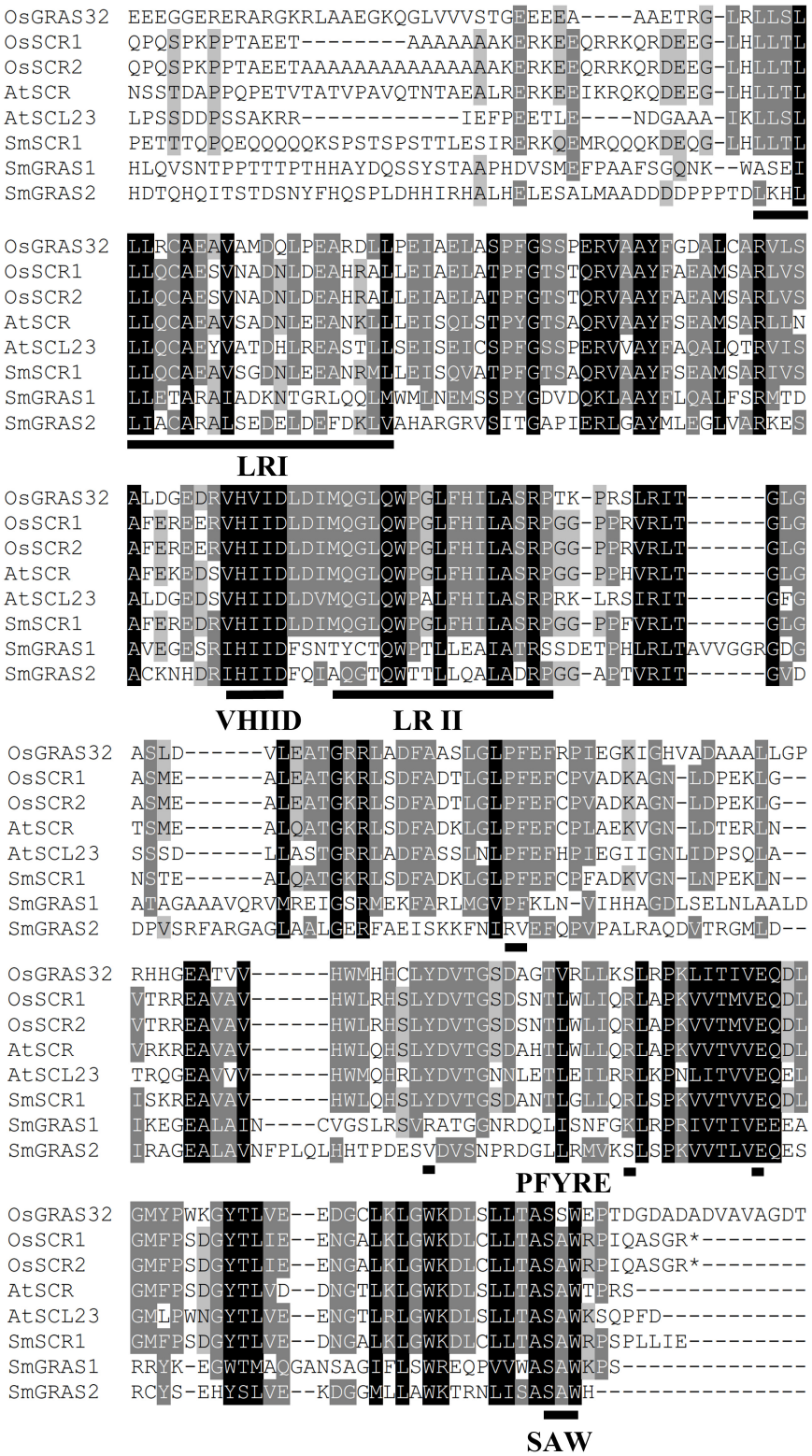


FIGURE 2 | Multiple alignments of *SmSCR1* with related GRAS proteins from other plants. SmGRAS1 (accession number: KY435886), SmGRAS2 (KY435887), OsGRAS32 (Os06g0127800), OsSCR1 (XP_015615402.1), OsSCR2 (XP_015620600.1), AtSCR (Q9M384.1), AtSCL23 (NM_123557), and *SmSCR1* (OM032820). Black line indicates the conserved amino acid sites. ILR I, leucine-rich region I; LR II, leucine-rich region II; VHIID, the conserved region consisting of valine, histidine, isoleucine, and D-aspartic acid; PFYRE, the conserved amino acid sites consisting of proline, phenylalanine, tyrosine, arginine, and glutamic acid; SAW, the conserved region consisting of serine, alanine, and tryptophan.

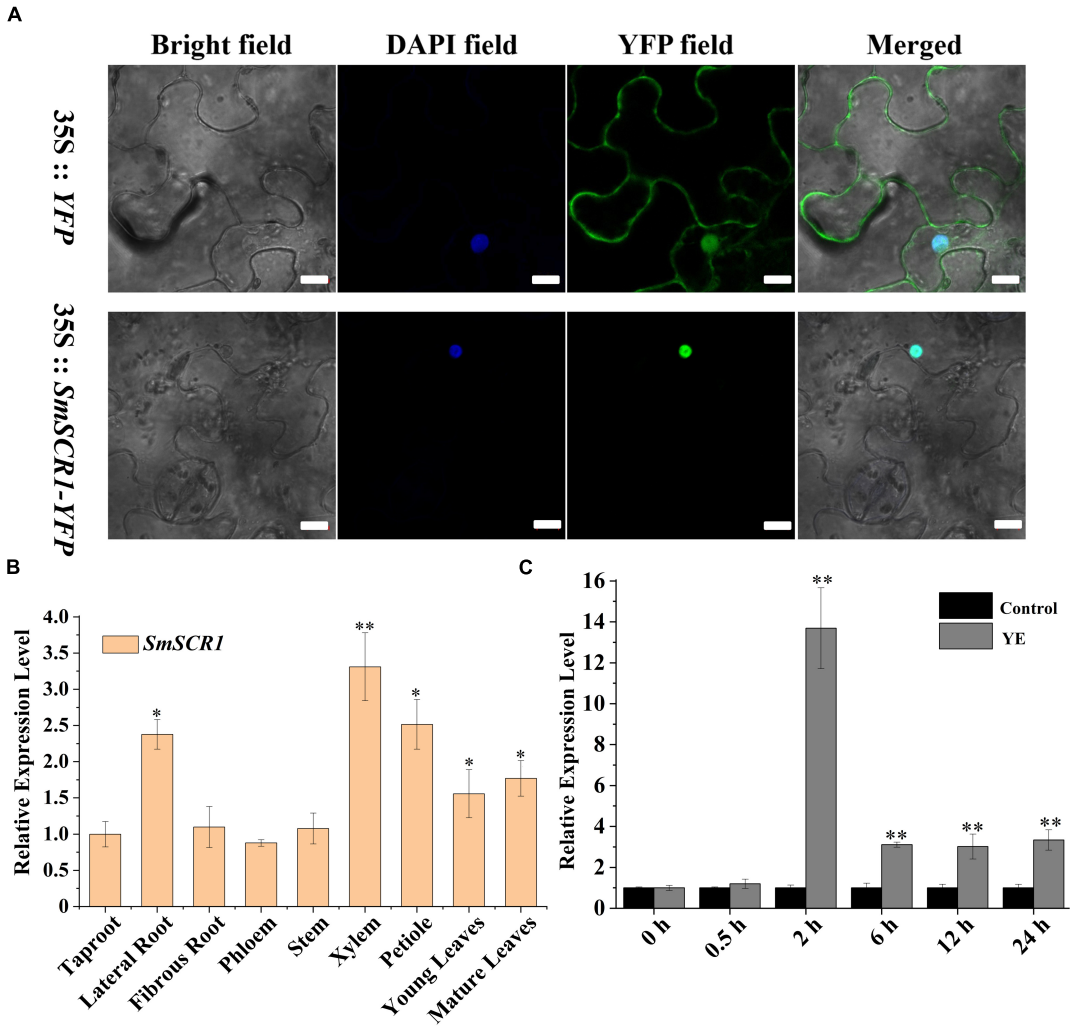


FIGURE 3 | Expression profile of *SmSCR1*. **(A)** Transient expression of the *SmSCR1* in *Nicotiana benthamiana* leaf epidermal cells. Scale bars represent 10 μm. **(B)** The expression patterns of *SmSCR1* in different tissues. Fold changes of the relative expression level of *SmSCR1* gene in other tissues are all normalized to the taproot. **(C)** Induced effect of YE on the expression of *SmSCR1*. Fold changes of the gene relative expression levels are all normalized to the control without induction by YE at 0 h. Asterisks indicate significant differences between the taproot and other tissues in **(B)** and between treatment and control in **(C)** at one significant level of *t*-test (*$p < 0.05$; **$p < 0.01$). Data are means of three replicates with SDs.

A. tumefaciens C58C1 and then transinfected into *S. miltiorrhiza* (Zhou et al., 2016).

Determination of Tanshinones by High-Performance Liquid Chromatography

After 60 days of continuous culture in 1/2 MS liquid medium, each hairy root line was harvested and dried in a freeze dryer for 24 h, and then extracted with 16 ml methanol/dichloromethane (3:1, v/v). Notably, 100 mg of dried powder of *S. miltiorrhiza*, sonicated for 1 h, then soaked overnight in the dark and centrifuged for 10 min at 12,000 × *g*. After centrifugation, the supernatant was poured into a distillation flask. After centrifugation, the supernatant was poured into a distillation flask and spun dry at 50°C. We then added 2 ml of methanol to dissolve the material in the distillation flask. After further centrifugation at 6,500 × *g* for 10 min, the samples were filtered separately through 0.22 μM organic membranes (Jinten, China) using a 1 ml sterile syringe (Huang et al., 2019). All metabolites were quantified by high-performance liquid chromatography (HPLC), as described previously (Agilent Technologies, Santa Clara, CA, United States) (Zhou et al., 2016; Sun et al., 2019; Yang et al., 2021).

Dual-Luciferase Assay

The dual-luciferase (Dual-LUC) assays were carried out as previously reported (Deng et al., 2019). The recombinants of pHB-YFP and pHB-SmSCR1-YFP were introduced into *A. tumefaciens* GV3301,respectively. Gene promoters from the tanshinone biosynthetic pathway were cloned and inserted into

pGREEN0800 vector, respectively, which were co-transformed into the GV3301 strain with pSoup19. The test was carried out as previously reported (Cao et al., 2018). The pHB-YFP construct was employed as the negative control, and the *Renilla* gene was used as the internal control.

Yeast One-Hybrid Assay

Similar to a previous study (Deng et al., 2019), the ORF of *SmSCR1* was inserted into the pB42AD vector. Three fragments of SmCPS1 promoter sequence (i.e., from −1 to −700, −701 to −1,400, and −1,401 to −2,100 bp, relative to translation start site, respectively) were inserted into a pLacZ plasmid. The resulting recombinants were co-transformed into yeast cell EGY48a. After cultivation on SD/-Ura/-Leu medium for 48 h, the positive binding activity was examined on SD/-Ura/-Leu medium with X-gal for 24 h. PB42AD and pLacZ empty carriers were set as negative controls.

Morphological Observations of *SmSCR1*

To investigate the effect of the *SmSCR1* gene on plant growth and development, we examined different phenotypes of the transgenic hairy roots on 1/2 MS medium. In addition, three different transgenic hairy roots of *SmSCR1* with good growth and similar developmental time were selected for developmental microstructure observation, as reported previously (Wu et al., 2020). The pith (Pi) and posterior xylem (Mx) of the hairy roots were observed under a light microscope (200 ×).

Statistical Analysis

Each data represent the average of three independent experimental calculations, and the data are expressed as the mean ± SD. SPSS 16.0 software was used to perform a single-sample *t*-test and one-way ANOVA, and *p*-values < 0.05 were considered statistically significant.

RESULTS

Isolation and Sequence Analysis of *SmSCR1* Gene

First, we identified a *GRAS* TF (unigene number: SmiContig9465) and *SmSCR1* (NCBI accession number:

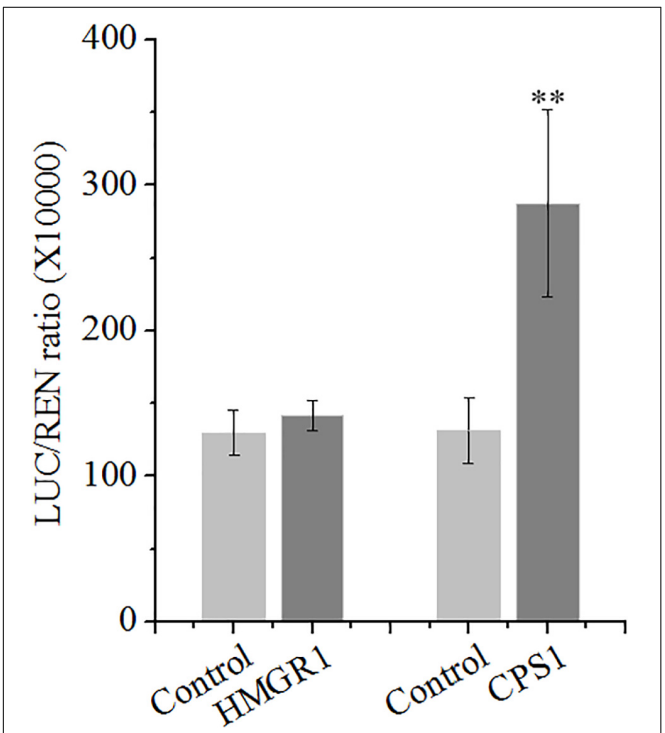


FIGURE 4 | *SmSCR1* activates the transcription of *SmCPS1*. *SmSCR1* activates the promoter of *SmCPS1* by the dual-luciferase (Dual-LUC) assay. The SmCPS1 promoter was fused to the LUC reporter, and its activity was determined by a transient Dual-LUC assay in *Nicotiana benthamiana*. The value of LUC/REN was recorded as the activating activity. Asterisk indicates significant differences at one significant level of *t*-test (**$p < 0.01$). Data are means of three replicates with SDs.

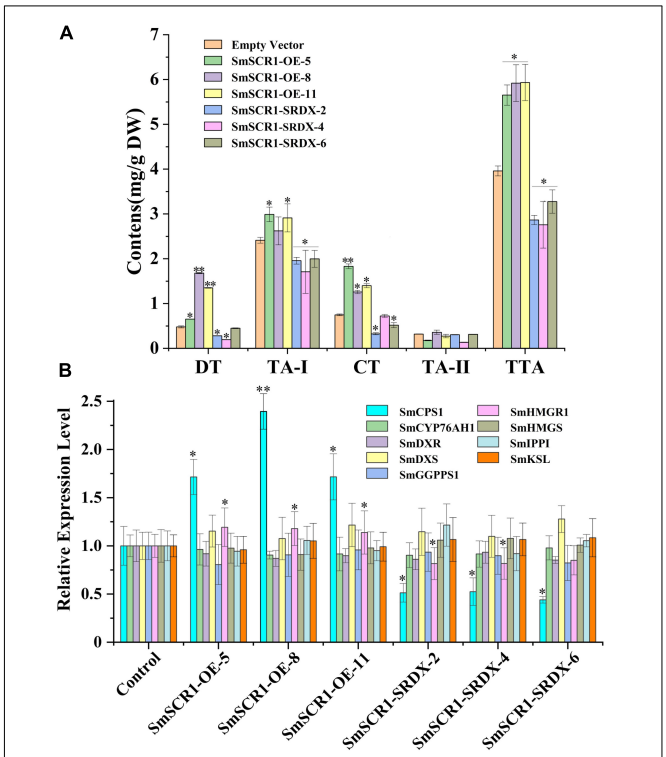


FIGURE 5 | Tanshinone content and gene expression in the tanshinone biosynthetic pathway in *Salvia miltiorrhiza* transgenic hairy root lines. **(A)** High-performance liquid chromatography analysis of tanshinone content in hairy root lines of SmSCR1-OE (overexpression) and SmSCR1-SRDX (suppression). DT, dihydrotanshinone; CT, cryptotanshinone; TA-I, tanshinone I; TA-II, tanshinone IIA; TTA, total tanshinone. **(B)** Gene expression in the SmSCR1-OE and SmSCR1-SRDX lines for the underlying tanshinone biosynthesis. The blank vector without *SmSCR1* gene is used as the control to standardize fold differences in relative gene expression levels in transgenic hairy roots. The standard errors of the mean are represented by error bars. The asterisks on the bar indicate significant differences by *t*-test compared to the control at two significant levels (*$p < 0.05$; **$p < 0.01$).

OM032820), responding to YE-induction (fold change > 2) from the *S. miltiorrhiza* transcriptome (Zhou et al., 2017). *SmSCR1* length is 1,944 bp long and encodes 647 amino acids, with a molecular weight size of 70.515 kDa coupled with a theoretical isoelectric point of 5.78. GRAS family members from *Arabidopsis* and rice were analyzed with *SmSCR1* to construct a phylogenetic tree. As shown in **Figure 1**, *SmSCR1* has the highest relationship with *AtSCR* in *Arabidopsis* and *OsSCR2* in rice, implying that *SmSCR1* belongs to the *SCR* subgroup of the *GRAS* family. *SmSCR1* consists of several conserved motifs, including VHIID, PFYRE, SAW motifs, and the leucine-rich regions (LRs), i.e., LR I and LR II (**Figure 2**).

Expression Analysis of *SmSCR1*

Transient expression of *SmSCR1* in epidermal cells from 45-day-old *N. benthamiana* leaves exhibited that *SmSCR1* located in the nucleus (**Figure 3A**). The gene expression of *SmSCR1* was characterized from nine different tissues of *S. miltiorrhiza* by qRT-PCR. *SmSCR1* was expressed in all tissues, namely, taproot, lateral root, fibrous root, phloem, xylem, stem, petiole, young leaves, and mature leaves, with the highest expression in xylem and petiole, and the lowest expression in taproot (**Figure 3B**). The expression of *SmSCR1* after YE-induction was examined, reaching the highest levels for 2 h after post-induction with 13.6-fold compared to the control (**Figure 3C**).

SmSCR1 Activates the Transcription of *Copalyl Diphosphate Synthase 1* Gene

The role of *SmSCR1* in regulating tanshinone biosynthetic gene expressions was examined. The Dual-LUC assay showed that *SmSCR1* uniquely activates the transcription of the *SmCPS1* promoter, leading to a 2.19-fold increase compared with the control, but not to other promoters (e.g., *HMGR1* promoter) (**Figure 4**). Y1H assay showed that self-activating of pB42AD vector binding to the promoter of the *CPS1* gene is visible (**Supplementary Figure 3**). Those results indicated that *SmSCR1* activated the transcription of *SmCPS1* to increase the production of tanshinone in *S. miltiorrhiza*.

SmSCR1 Promotes Tanshinone Biosynthesis in the Transgenic Hairy Root of *Salvia miltiorrhiza*

The OE and SRDX suppressive hairy root lines were first examined by genomic PCR detection (**Supplementary Figure 4**). Then, the transcript profiles of *SmSCR1* in transgenic hairy roots were quantified by qRT-PCR. Three independent OE lines (i.e., OE-5, OE-8, and OE-11) and three SRDX lines (i.e., SRDX-2, SRDX-4, and SRDX-6) of *SmSCR1* with the highest expression levels in transgenic lines were subsequently chosen for further analysis (**Supplementary Figure 5**). After 60-day-old subculture, the hairy root lines were collected independently and were subjected to examine the content of tanshinones by HPLC (**Figure 5A**). The concentrations of DT, TA-1, CT, and TTA were significantly increased compared to the control in the OE lines, with the highest total tanshinone (TTA) content in SmSCR1-OE-11, reaching 1.49-fold than that of the control. In contrast, in the SRDX repression lines, DT, TA-1, CT, and TTA were significantly decreased compared to the control, with the lowest TTA content of 2.76 mg/g DW. The above results suggest that the *SmSCR1* TF can promote the tanshinone accumulation in *S. miltiorrhiza*.

SmSCR1 Upregulates *Copalyl Diphosphate Synthase 1* and *HMGR1* in Transgenic Hairy Roots of *Salvia miltiorrhiza*

Copalyl diphosphate synthase 1 and *HMGR1* were significantly upregulated in OE lines (e.g., OE-5, OE-8, and OE-11), with the SmSCR1-OE-11 line, showing the highest expression. In contrast, the two genes were drastically downregulated in the three SRDX lines (i.e., 2, 4, and 6), with the SmSCR1-SRDX-6 line exhibiting the lowest decrease compared to the control (**Figure 5B**). These findings showed that *SmSCR1* stimulated *CPS1* and *HMGR1* together, resulting in increased production in transgenic hairy roots of *S. miltiorrhiza*.

SmSCR1 Promotes the Growth of Transgenic Hairy Roots of *Salvia miltiorrhiza*

The biomass of the transgenic hairy root lines significantly increased in the *SmSCR1* OE lines and significantly reduced in the SRDX lines compared to the control (**Figure 6A**). Moreover, the *SmSCR1* overexpression hairy root lines had a better growth status than the control, with more branches in the *SmSCR1* OE lines and fewer in the SRDX suppressive hair roots (**Figure 6B**). Roots are formed by a series of asymmetrical divisions of early differentiated cells to produce the inner and outer cortex of the root, which then develops into xylem and phloem. The development of transgenic hairy roots by paraffin sectioning demonstrated that the *SmSCR1* gene promotes the formation of hairy root radiation morphology of the Mx cell and Pi cell differentiation in *SmSCR1*-OE hairy roots when compared with the control and the SRDX suppressive hairy roots (**Figure 6**). Those results suggest that *SmSCR1* can modulate the growth and development of hairy roots in *S. miltiorrhiza*.

DISCUSSION

The GRAS TF family is an important TF class involved in regulating plant growth, development, environmental stress response, and growth signal transduction (Sun et al., 2011). With the development of genome sequencing technology, a few GRAS family members have been identified in plants, such as plum (*Armeniaca mume*), lotus (*Nelumbo nucifera*), and pine blue (*Isatis indigotica*) (Kamiya et al., 2003; Hou et al., 2010). However, only a few reports systematically characterized GRAS TFs in medicinal plants. We identified a novel YE-responsive GRAS TF gene, *SmSCR1* gene, encoding 647 amino acids with several key GRAS conserved structural domains (e.g., VHIID, LR I, LR II, PFYRE, and SAW motifs) (Koizumi et al., 2011; Zhou et al., 2017). In a previous study, SCR and SHR families

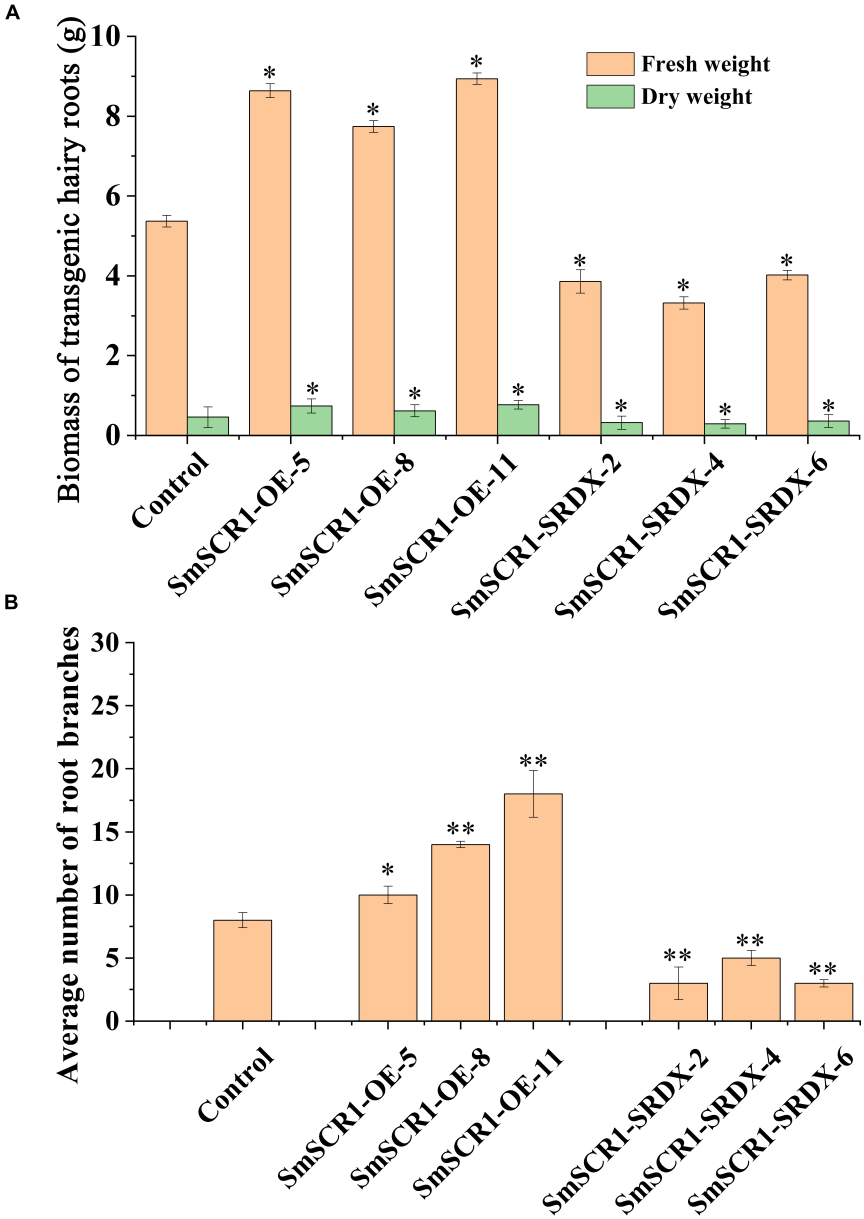


FIGURE 6 | Phenotype of the *SmSCR1* transgenic hairy root. **(A)** Biomass of transgenic hairy root. **(B)** An average number of root branches. Asterisks indicate significant differences between the transgenic hairy root lines and control at one significant level of *t*-test (*$p < 0.05$; **$p < 0.01$). Data are means of three replicates with SDs.

were thought to co-regulate the plant growth, development, and radial structures in roots (Koizumi et al., 2011, 2012; Fan et al., 2017). We also identified that *SmSCR1* was very sensitive to YE-induction (**Figure 3C**), regulating tanshinone accumulation in *S. miltiorrhiza* (Zhou et al., 2017, 2021a).

We revealed that *SmSCR1* upregulates tanshinone biosynthesis in *SmSCR1* suppressive hairy root lines (**Figure 5A**). Recently, two GRAS families, including *SmGRAS1* (GenBank accession number: KY435886) and *SmGRAS2* (GenBank accession number: KY435887), clustered into SHR and PAT1 subgroups, respectively (**Figure 1**), were characterized in *S. miltiorrhiza*. SmGRAS1 and SmGRAS2 were verified by transgenic validation to upregulate tanshinone accumulation in hairy roots. Therefore, many GRAS families might play a more prominent role in regulating tanshinone accumulation in *S. miltiorrhiza* and require further investigation.

Both *CPS1* and *HMGR1* positively correlated to the expression of *SmSCR1* in the transgenic lines, while *SmSCR1* can only bind to the *CPS1* promoter to activate its expression as examined. Due to the limited data about the specific *cis*-element in promoter binding by SCR1 in other plants (Li et al., 2019), further examining the direct binding of *SmSCR1* to the specific

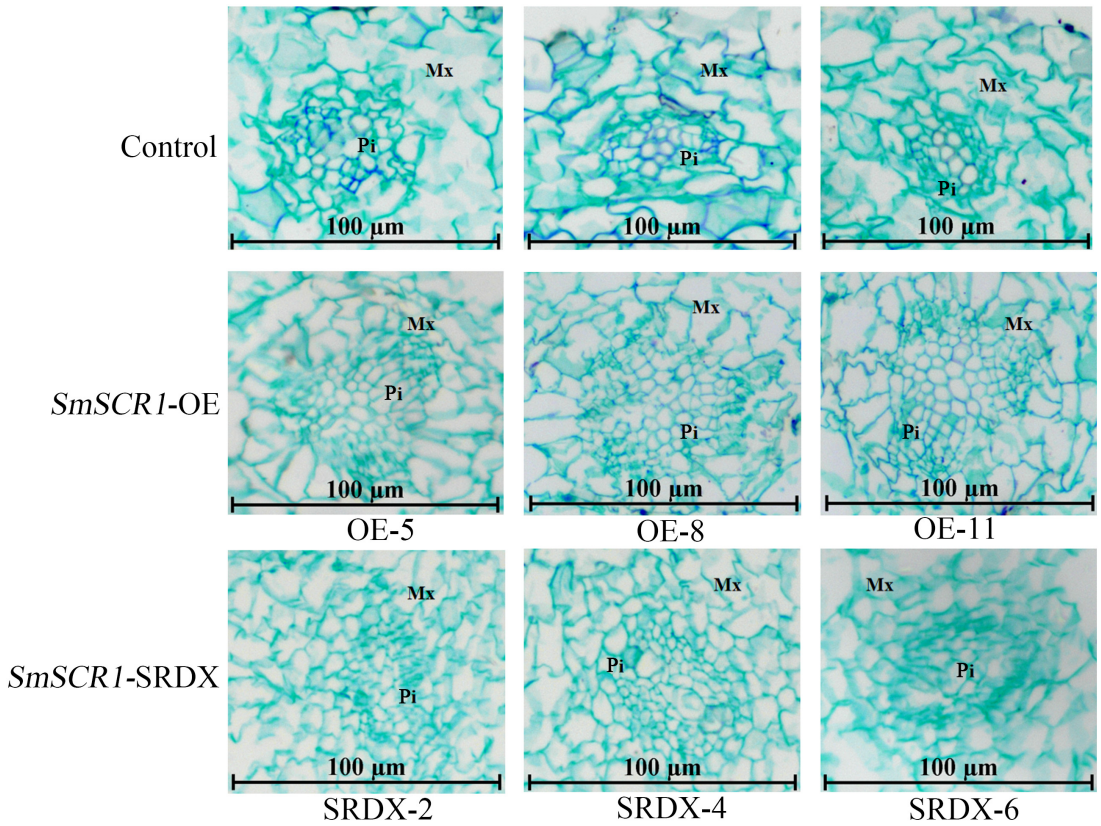


FIGURE 7 | Paraffin sectioning analysis of the transgenic hairy roots. The control means the transgenic hairy roots with the transformation of pHB-X-YFP blank vector. pHB-SmSCR1 represents the overexpression lines, and pHB-SmSCR1-SRDX represents the suppression lines. Pi represents the pith. Mx represents the posterior xylem.

promoter elements by EMSA assay *in vitro* would be unfeasible. In this study, overexpressing *SmSCR1* upregulated not only the expression of its target gene *CPS1* but also the *HMGR1* gene in the tanshinone biosynthetic pathway. Overexpressing *SmERF1L1* promotes the expression of five genes in the tanshinone biosynthetic pathway, but only *DXR* is a target (Huang et al., 2019). In SmERF128 transgenic hairy roots, the expression of six genes in the tanshinone biosynthetic pathway was activated, and only three of the six genes were validated to be targets of SmERF128 (Zhang et al., 2019). Thus, it can be inferred that the ectopic expression of certain TF genes upregulates the expression of not only their target genes but also other genes in the same biosynthetic pathway.

The root formation is a series of asymmetric divisions of stem cells to generate the inner and outer cortex of the root and then differentiate into different forms of tissue. In *A. thaliana*, SCR is mainly expressed in root epidermal cells. As a positive regulator, SCR regulates root radiation morphology and plays a vital role in root monolayer cell formation (Koizumi et al., 2012; Rich et al., 2017). When SCR function is lost and mutated, the normal development of the cell layer will be affected. In addition, SCR can also regulate the division of progenitor cells and promote the growth of biomass (Koizumi et al., 2011; Slewinski et al., 2012). In rice, *OsSCR* was thought to affect the development of plant leaf guard cells (Hou et al., 2010). In the three *SmSCR1* OE lines, we found that *SmSCR1* significantly promoted the biomass increase of the three transgenic hairy root lines, whereas a visible decrease was observed in *SmSCR1* SRDX lines.

Moreover, *SmSCR1* can regulate the development of Pi and the Mx. In *SmSCR1* OE lines, the cell division rate of progeny

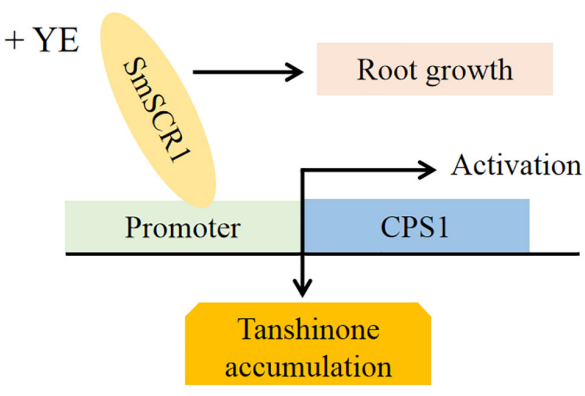


FIGURE 8 | A proposed model for the role of *SmSCR1* in modulating tanshinone biosynthesis and root growth in *Salvia miltiorrhiza*.

and the root radial structure growth rate were significantly higher than that of control. In contrast, the three SRDX suppressive hairy root lines have an opposite phenotype (**Figure 7**). The function of *SmSCR1* to regulate the growth and development of hairy roots in *S. miltiorrhiza* is similar to the model plants, e.g., *Arabidopsis* and rice.

Based on the present findings, a deductive model of *SmSCR1* as an activator in response to YE-induction promotes tanshinone accumulation, Pi cell division, radial structure formation, and biomass increase of hairy roots in *S. miltiorrhiza* was outlined (**Figure 8**).

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

GK and WZ conceived and designed the project. SW, YS, YL, XG, CC, and TL prepared materials and performed the experiments. WZ, SW, YS, IM, YL, and CW performed the bioinformatics analysis and analyzed the data. GK, WZ, SW, and YS wrote the manuscript. IM, WZ, and GK revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

We acknowledge the support extended toward this research by the bodies listed: Zhejiang Provincial Natural Science Foundation of China (LY20H280008), National Key R&D Program of China (2018YFC1706203), the Major Science and Technology Projects of Breeding New Varieties of Agriculture in Zhejiang Province (2021C02074-3-4), Pre-research Projects of Zhejiang Chinese Medical University (2021JKZKTS011A), and Sannongliufang Research Joint Project of Zhejiang Province (2021SNLF019), and the Opening Project of Zhejiang Provincial Preponderant and Characteristic Subject of Key University (Traditional Chinese Pharmacology), Zhejiang Chinese Medical University (ZYAOX2018003 and ZYAOX2018009).

ACKNOWLEDGMENTS

We gratefully acknowledge the great experimental support of the Public Platform of Medical Research Center, Academy of Chinese Medical Science, and Zhejiang Chinese Medical University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.860033/full#supplementary-material

Supplementary Figure 1 | The skeleton of tanshinone biosynthetic pathway in *Salvia miltiorrhiza*.

Supplementary Figure 2 | Recombinants of *SmSCR1*-overexpressive and suppressed vectors. **(A)** Recombinant vector of pHB -*SmSCR1*-YFP. **(B)** Recombinant vector of pHB-*SmSCR1*-SRDX.

Supplementary Figure 3 | Yeast one-hybrid assay of *SmSCR1* binding to the SmCPS1 promoter. P1, P2, and P3 represent the three fragments of *SmCPS1* promoter sequence from −1 to −700, −701 to −1,400, and −1,401 to −2,100 bp, relative to translation start site, respectively.

Supplementary Figure 4 | Detection of the expression level of *SmSCR1* gene in transgenic hairy root lines by qRT-PCR.

Supplementary Figure 5 | Identification of transgenic hairy root lines by PCR. The asterisks on the bar indicate significant differences by *t*-test compared to the control (**$P < 0.01$).

REFERENCES

Cao, W., Wang, Y., Shi, M., Hao, X., Zhao, W., Wang, Y., et al. (2018). Transcription factor SmWRKY1 positively promotes the biosynthesis of tanshinones in *Salvia miltiorrhiza*. *Front. Plant Sci.* 9:554. doi: 10.3389/fpls.2018.00554

Chen, H., Chena, F., Chiu, F. C., and Lo, C. M. (2001). The effect of yeast elicitor on the growth and secondary metabolism of hairy root cultures of *Salvia miltiorrhiza*. *Enzyme Microb. Tech.* 28, 100–105. doi: 10.1016/s0141-0229(00)00284-2

Deng, C., Hao, X., Shi, M., Fu, R., Wang, Y., Zhang, Y., et al. (2019). Tanshinone production could be increased by the expression of SmWRKY2 in *Salvia miltiorrhiza* hairy roots. *Plant Sci.* 284, 1–8. doi: 10.1016/j.plantsci.2019.03.007

Deng, C., Shi, M., Fu, R., Zhang, Y., Wang, Q., Zhou, Y., et al. (2020a). ABA-responsive transcription factor bZIP1 is involved in modulating biosynthesis of phenolic acids and tanshinones in *Salvia miltiorrhiza*. *J. Exp. Bot.* 71, 5948–5962. doi: 10.1093/jxb/eraa295

Deng, C., Wang, Y., Huang, F., Lu, S., Zhao, L., Ma, X., et al. (2020b). SmMYB2 promotes salvianolic acid biosynthesis in the medicinal herb *Salvia miltiorrhiza*. *J. Integr. Plant Biol.* 62, 1688–1702. doi: 10.1111/jipb.12943

Fan, S., Zhang, D., Gao, C., Zhao, M., Wu, H., Li, Y., et al. (2017). Identification, classification, and expression analysis of GRAS gene family in *Malus domestica*. *Front. Physiol.* 8:253. doi: 10.3389/fphys.2017.00253

Fu, L., Han, B., Zhou, Y., Ren, J., Cao, W., Patel, G., et al. (2020). The anticancer properties of tanshinones and the pharmacological effects of their active ingredients. *Front. Pharmacol.* 11:193. doi: 10.3389/fphar.2020.00193

Gao, W., Hillwig, M. L., Huang, L., Cui, G., Wang, X., Kong, J., et al. (2009). A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org. Lett.* 11, 5170–5173. doi: 10.1021/ol902051v

Hao, X., Pu, Z., Cao, G., You, D., Zhou, Y., Deng, C., et al. (2020). Tanshinone and salvianolic acid biosynthesis are regulated by SmMYB98 in *Salvia miltiorrhiza* hairy roots. *J. Adv. Res.* 23, 1–12. doi: 10.1016/j.jare.2020.01.012

Hao, X., Shi, M., Cui, L., Xu, C., Zhang, Y., and Kai, G. (2015). Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic *Salvia miltiorrhiza* hairy roots. *Biotechnol. Appl. Bioc.* 62, 24–31. doi: 10.1002/bab.1236

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J.* 34, 733–739. doi: 10.1046/j.1365-313x.2003.01759.x

Hou, X., Lee, L., Xia, K., Yan, Y., and Yu, H. (2010). DELLAs modulate jasmonate signaling via competitive binding to JAZs. *Dec. Cell* 19, 884–894. doi: 10.1016/j.devcel.2010.10.024

Huang, Q., Sun, M., Yuan, T., Wang, Y., Shi, M., Lu, S., et al. (2019). The AP2/ERF transcription factor SmERF1L1 regulates the biosynthesis of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Food Chem.* 274, 368–375. doi: 10.1016/j.foodchem.2018.08.119

Jia, Q., Zhu, R., Tian, Y., Chen, B., Li, R., Li, L., et al. (2019). *Salvia miltiorrhiza* in diabetes: a review of its pharmacology, phytochemistry, and safety. *Phytomedicine* 58:152871. doi: 10.1016/j.phymed.2019.152871

Jung, I., Kim, H., Moon, S., Lee, H., and Kim, B. (2020). Overview of *Salvia miltiorrhiza* as a potential therapeutic agent for various diseases: an update on efficacy and mechanisms of action. *Antioxidants* 9:857. doi: 10.3390/antiox9090857

Kai, G., Xu, H., Zhou, C., Liao, P., Xiao, J., Luo, X., et al. (2011). Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metab. Eng.* 13, 319–327. doi: 10.1016/j.ymben.2011.02.003

Kamiya, N., Itoh, J., Morikami, A., Nagato, Y., and Matsuoka, M. (2003). The SCARECROW gene's role in asymmetric cell divisions in rice plants. *Plant J.* 36, 45–54. doi: 10.1046/j.1365-313x.2003.01856.x

Koizumi, K., Hayashi, T., Wu, S., and Gallagher, K. L. (2012). The SHORT-ROOT protein acts as a mobile, dose-dependent signal in patterning the ground tissue. *P. Natl. Acad. Sci. U S A.* 109, 13010–13015. doi: 10.1073/pnas.1205579109

Koizumi, K., Wu, S., MacRae-Crerar, A., and Gallagher, K. L. (2011). An essential protein that interacts with endosomes and promotes movement of the SHORT-ROOT transcription factor. *Curr. Biol.* 21, 1559–1564. doi: 10.1016/j.cub.2011.08.013

Li, W., Bai, Z., Pei, T., Yang, D., Mao, R., Zhang, B., et al. (2019). SmGRAS1 and SmGRAS2 regulate the biosynthesis of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Front. Plant Sci.* 10:1367. doi: 10.3389/fpls.2019.01367

Ma, Y., Ma, X. H., Meng, F. Y., Zhan, Z. L., Guo, J., and Huang, L. Q. (2016). RNA interference targeting CYP76AH1 in hairy roots of *Salvia miltiorrhiza* reveals its key role in the biosynthetic pathway of tanshinones. *Biochem. Bioph. Res. Co* 477, 155–160. doi: 10.1016/j.bbrc.2016.06.036

Ma, Y., Yuan, L., Wu, B., Li, X., Chen, S., and Lu, S. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exp. Bot.* 63, 2809–2823. doi: 10.1093/jxb/err466

Rich, M. K., Courty, P. E., Roux, C., and Reinhardt, D. (2017). Role of the GRAS transcription factor ATA/RAM1 in the transcriptional reprogramming of arbuscular mycorrhiza in *Petunia hybrida*. *BMC Genomics* 18:589. doi: 10.1186/s12864-017-3988-8

Shi, M., Gong, H., Cui, L., Wang, Q., Wang, C., Wang, Y., et al. (2020). Targeted metabolic engineering of committed steps improves anti-cancer drug camptothecin production in *Ophiorrhiza pumila* hairy roots. *Ind. Crop. Prod.* 148:112277. doi: 10.1016/j.indcrop.2020.112277

Shi, M., Du, Z., Hua, Q., and Kai, G. (2021a). CRISPR/Cas9-mediated targeted mutagenesis of bZIP2 in Salvia miltiorrhiza leads to promoted phenolic acid biosynthesis. *Ind. Crop. Prod.* 167:113560. doi: 10.1016/j.indcrop.2021.113560

Shi, M., Liao, P., Nile, S. H., Georgiev, M. I., and Kai, G. (2021b). Biotechnological exploration of transformed root culture for value-added products. *Trends Biotechnol.* 39, 137–149. doi: 10.1016/j.tibtech.2020.06.012

Shi, M., Luo, X., Ju, G., Li, L., Huang, S., Zhang, T., et al. (2016a). Enhanced diterpene tanshinone accumulation and bioactivity of transgenic *Salvia miltiorrhiza* hairy roots by pathway engineering. *J. Agr. Food Chem.* 64, 2523–2530. doi: 10.1021/acs.jafc.5b04697

Shi, M., Zhou, W., Zhang, J., Huang, S., Wang, H., and Kai, G. (2016b). Methyl jasmonate induction of tanshinone biosynthesis in *Salvia miltiorrhiza* hairy roots is mediated by JASMONATE ZIM-DOMAIN repressor proteins. *Sci. Rep.* 6:20919. doi: 10.1038/srep20919

Slewinski, T. L., Anderson, A. A., Zhang, C., and Turgeon, R. (2012). Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* 53, 2030–2037. doi: 10.1093/pcp/pcs147

Sun, M., Shi, M., Wang, Y., Huang, Q., Yuan, T., Wang, Q., et al. (2019). The biosynthesis of phenolic acids is positively regulated by the JA-responsive transcription factor ERF115 in *Salvia miltiorrhiza*. *J. Exp. Bot.* 70, 243–254. doi: 10.1093/jxb/ery349

Sun, X., Xue, B., Jones, W. T., Rikkerink, E., Dunker, A. K., and Uversky, V. N. (2011). A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.* 77, 205–223. doi: 10.1007/s11103-011-9803-z

Wang, B., Sun, W., Li, Q., Li, Y., Luo, H., Song, J., et al. (2015). Genome-wide identification of phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. *Planta* 241, 711–725. doi: 10.1007/s00425-014-2212-1

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., et al. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii f. niedzwetzkyana*). *Plant J.* 90, 276–292. doi: 10.1111/tpj.13487

Wu, X., Riaz, M., Yan, L., Zhang, Z., and Jiang, C. (2020). How the cells were injured and the secondary metabolites in the shikimate pathway were changed by boron deficiency in trifoliate orange root. *Plant Physiol. Biochem.* 151, 630–639. doi: 10.1016/j.plaphy.2020.04.009

Xing, B., Liang, L., Liu, L., Hou, Z., Yang, D., Yan, K., et al. (2018). Overexpression of SmbHLH148 induced biosynthesis of tanshinones as well as phenolic acids in *Salvia miltiorrhiza* hairy roots. *Plant Cell Rep.* 37, 1681–1692. doi: 10.1007/s00299-018-2339-9

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant.* 9, 949–952. doi: 10.1016/j.molp.2016.03.010

Yang, M., Wang, Q., Liu, Y., Hao, X., Wang, C., Liang, Y., et al. (2021). Divergent camptothecin biosynthetic pathway in *Ophiorrhiza pumila*. *BMC Biol.* 19:122. doi: 10.1186/s12915-021-01051-y

Yang, N., Zhou, W., Su, J., Wang, X., Li, L., Wang, L., et al. (2017). Overexpression of SmMYC2 increases the production of phenolic acids in *Salvia miltiorrhiza*. *Front. Plant Sci.* 8:1804. doi: 10.3389/fpls.2017.01804

Zhang, L., Yan, X. M., Wang, J., Li, S. S., Liao, P., and Kai, G. Y. (2011). Molecular cloning and expression analysis of a new putative gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase from *Salvia miltiorrhiza*. *Acta Physiol. Plant.* 33, 953–961. doi: 10.1007/s11738-010-0627-2

Zhang, Y., Ji, A., Xu, Z., Luo, H., and Song, J. (2019). The AP2/ERF transcription factor SmERF128 positively regulates diterpenoid biosynthesis in *Salvia miltiorrhiza*. *Plant Mol. Biol.* 100, 83–93. doi: 10.1007/s11103-019-00845-7

Zhou, W., Huang, F. F., Li, S., Wang, Y., Zhou, C. C., Shi, M., et al. (2016). Molecular cloning and characterization of two 1-deoxy-D-xylulose-5-phosphate synthase genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Mol. Breed.* 36:124. doi: 10.1007/s11032-016-0550-3

Zhou, W., Huang, Q., Wu, X., Zhou, Z., Ding, M., Shi, M., et al. (2017). Comprehensive transcriptome profiling of Salvia miltiorrhiza for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci. Rep.* 7:10554. doi: 10.1038/s41598-017-10215-2

Zhou, W., Li, S., Maoz, I., Wang, Q., Xu, M., Feng, Y., et al. (2021a). SmJRB1 positively regulates the accumulation of phenolic acid in *Salvia miltiorrhiza*. *Ind. Crop. Prod.* 164:113417. doi: 10.1016/j.indcrop.2021.113417

Zhou, W., Shi, M., Deng, C., Lu, S., Huang, F., Wang, Y., et al. (2021b). The methyl jasmonate-responsive transcription factor SmMYB1 promotes phenolic acid biosynthesis in *Salvia miltiorrhiza*. *Hortic. Res.* 8:10. doi: 10.1038/s41438-020-00443-5

Zhou, Y., Sun, W., Chen, J., Tan, H., Xiao, Y., Li, Q., et al. (2020). Author correction: SmMYC2a and SmMYC2b played similar but irreplaceable roles in regulating the biosynthesis of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Sci. Rep.* 10:7201. doi: 10.1038/s41598-020-62994-w

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 March 2022
doi: 10.3389/fpls.2022.860152

Genome-Wide Identification of BAHD Superfamily and Functional Characterization of Bornyl Acetyltransferases Involved in the Bornyl Acetate Biosynthesis in *Wurfbainia villosa*

*Huilin Liang[1,2], Xiaojing Lin[1,2], Peng Yang[1,2,3], Yewen Sun[1,2], Qingwen Wu[1,2], Shamukaer Alimujiang[2], Haiying Zhao[1], Dongming Ma[1,2], Ruoting Zhan[1,2] and Jinfen Yang[1,2]**

[1] Key Laboratory of Chinese Medicinal Resource from Lingnan (Ministry of Education), Guangzhou University of Chinese Medicine, Guangzhou, China, [2] School of Pharmaceutical Science, Guangzhou University of Chinese Medicine, Guangzhou, China, [3] Hunan Provincial Key Laboratory for Synthetic Biology of Traditional Chinese Medicine, School of Pharmaceutical Sciences, Hunan University of Medicine, Huaihua, China

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Xu Lu,
China Pharmaceutical University, China
Shihai Xing,
Anhui University of Chinese Medicine, China

***Correspondence:**
Jinfen Yang
yangjf@gzucm.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *22 January 2022*
Accepted: *25 February 2022*
Published: *28 March 2022*

Citation:
Liang H, Lin X, Yang P, Sun Y, Wu Q, Alimujiang S, Zhao H, Ma D, Zhan R and Yang J (2022) Genome-Wide Identification of BAHD Superfamily and Functional Characterization of Bornyl Acetyltransferases Involved in the Bornyl Acetate Biosynthesis in Wurfbainia villosa. Front. Plant Sci. 13:860152. doi: 10.3389/fpls.2022.860152

Bornyl acetate (BA) is known as a natural aromatic monoterpene ester with a wide range of pharmacological and biological activities. Borneol acetyltransferase (BAT), catalyzing borneol and acetyl-CoA to synthesize BA, is alcohol acetyltransferase, which belongs to the BAHD super acyltransferase family, however, BAT, responsible for the biosynthesis of BA, has not yet been characterized. The seeds of *Wurfbainia villosa* (homotypic synonym: *Amomum villosum*) are rich in BA. Here we identified 64 members of the BAHD gene family from the genome of *W. villosa* using both PF02458 (transferase) and PF07247 (AATase) as Hidden Markov Model (HMM) to screen the BAHD genes. A total of sixty-four WvBAHDs are distributed on 14 chromosomes and nine unanchored contigs, clustering into six clades; three WvBAHDs with PF07247 have formed a separated and novel clade: clade VI. Twelve candidate genes belonging to clade I-a, I-b, and VI were selected to clone and characterize *in vitro*, among which eight genes have been identified to encode BATs acetylating at least one type of borneol to synthesize BA. All eight WvBATs can utilize (−)-borneol as substrates, but only five WvBATs can catalyze (+)-borneol, which is the endogenous borneol substrate in the seeds of *W. villosa*; WvBAT3 and WvBAT4 present the better catalytic efficiency on (+)-borneol than the others. The temporal and spatial expression patterns of *WvBATs* indicate that *WvBAT3* and *WvBAT4* are seed-specific expression genes, and their expression levels are correlated with the accumulation of BA, suggesting WvBAT3 and WvBAT4 might be the two key BATs for BA synthesis in the seeds of *W. villosa*. This is the first report on BAT responsible for the last biosynthetic step of BA, which will contribute to further studies on BA biosynthesis and metabolism engineering of BA in other plants or heterologous hosts.

Keywords: bornyl acetate, borneol acetyltransferase, *Wurfbainia villosa*, BAHD gene family, alcohol acetyltransferase

INTRODUCTION

Natural products have been a prominent source of pharmacologically active molecules in medicines for years, with advantages of effectiveness and low occurrence of side effects. Bornyl acetate (BA) has been reported for its rich pharmacological effects. BA has shown a high lipoxygenase inhibition, leading to the reduction of the inflammatory/allergic response, tumoral and Alzheimer's disorders, and reducing 5-fluorouracil-induced intestinal mucositis (Zhang et al., 2017; Cutillas et al., 2018). BA has also been demonstrated to be the potential proliferation inhibitor against human MCF-7, HT-29, and H-1299 cells (Sajjadi et al., 2015). Moreover, it has been reported that BA has analgesic, antioxidant, whitening, anticancer, antitumor, antiabortion, and anti-anxiety effects, and it has antibacterial, insecticidal, and anesthetic effects symbiotically with other aromatic compounds in the volatile oil (Wang et al., 2011; Asghari et al., 2012; Ohtsubo et al., 2015; Perestrelo et al., 2016; Zhang et al., 2017; Ao et al., 2019). However, BA is only distributed in a few families of plants, such as *Zingiberaceae, Pinaceae*, *Cupressaceae*, *Rutaceae*, *Umbelliferae*, *Lamiaceae*, and *Anacardiaceae*. *Wurfbainia villosa* (homotypic synonym: *Amomum villosum*) is rich in BA; BA content in the seeds essential oil is 10.53% (50.35% of the total monoterpene content), higher than the contents of BA in other plants, such as *Laurus nobilis* (fruits, 4.40%), *Illicium pachyphyllum* (fruits, 4.01%), and *Valeriana Jatamansi* (root, 0.6–1.5%) (Liu et al., 2012; Raina and Negi, 2015; Wang et al., 2018; Fidan et al., 2019; Chen et al., 2020b). Therefore, *W. villosa* is a significant material for illuminating the BA biosynthetic pathway. *W. villosa* is a plant of the ginger family, and its fragrant dried fruit is known as Fructus Amomi (Chinese medicine name: Sharen). It is a well-known traditional Chinese medicine and is widely used in the treatment of digestive system diseases and daily cooking in China, especially in southern China. BA is the medicinal substance and quality index of *W. villosa*, and its content determines the quality of *W. villosa* (Commission of Chinese Materia Medica, 1999; Commission of Chinese Pharmacopoeia, 2015). In addition to high levels of BA, *W. villosa* contains a variety of terpene acetates with lower or trace levels, such as isobornyl acetate (IBA), nerolidyl acetate, and santalyl acetate, which are the unique aroma and flavor components of *W. villosa* relating to its pharmacological activities (Zhao et al., 2021b; **Supplementary Table 1**).

In plants, all terpenoid skeletons are generated by the (MVA) pathway and the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway, which have been well documented (Vranová et al., 2013). Bornyl diphosphate synthase (BPPS) is the first key enzyme in the downstream pathway of BA biosynthesis, which catalyzes the monoterpene precursor, geranyl diphosphate (GPP), to generate bornyl diphosphate, the precursor of borneol. *BPPS* has been cloned and characterized from a few plants, such as WvBPPS from *W. villosa*, CbBPPS from *Cinnamomum burmannii* (Wang et al., 2018; Ma et al., 2021). Diphosphate diphosphatase is converted to borneol by dephosphorylation, and borneol is acetylated by alcohol acetyltransferase to synthesize BA. Alcohol acetyltransferase (AAT, EC 2.3.1.84) can catalyze terpene alcohol, aromatic alcohol, or aliphatic alcohol to synthesize aromatic volatile acetate in plants (**Figure 1**; Beekwilder et al., 2004; Souleyre et al., 2005). *AATs* have been cloned and identified from the fruits of strawberry, banana, apple, peach, and flowers of rose and lavender, and traditional herbs, such as *Celastrus angulatus*, *Ocimum basilicum,* and other plants (**Supplementary Table 2**; Beekwilder et al., 2004; Souleyre et al., 2005; Zhang et al., 2010; Sarker and Mahmoud, 2015; Dhar et al., 2020; Yan et al., 2020). BA is synthesized by acetylation of borneol, which has three optical isomers, including levorotatory, dextrorotatory, and racemate (Drienovská et al., 2020; Chánique et al., 2021). In contrast, chemically synthesized borneol contains four different stereoisomers, (+)-borneol, (−)-borneol, (+)-isoborneol, and (−)-isoborneol (Khine et al., 2020). However, the AAT responsible for the synthesis of BA and other monoterpene acetates in *W. villosa* has not yet been characterized; furthermore, the gene encoding borneol acetyltransferase (BAT) catalyzing different types of borneol to synthesize BA has not yet been reported to date.

Alcohol acetyltransferase is one important member of the BAHD acyltransferase family, which was named according to the first alphabet of the first four biochemically characterized enzymes of this family (BEAT, AHCT, HCBT, and DAT) and features with two highly conserved motifs: "HXXXD" and "DFGWG" (St-Pierre and De Luca, 2000). The BAHD members identified to date are all monomeric enzymes with a molecular mass ranging from 48 to 55 kDa (D'Auria, 2006). AATs are mainly high expressed in plant tissues which accumulate significant amounts of volatile esters, such as glandular trichomes, flowers (petals/stigmas), or fruits (pulp/receptacle), and some are expressed in leaves or stems (Beekwilder et al., 2004; Souleyre et al., 2005; Sarker and Mahmoud, 2015; Dhar et al., 2020; Yan et al., 2020). AATs expressed in fruits are maturation-induced. The substrates including alcohol precursors and acyl-CoA donors of AAT are broad, and one AAT might be involved in the synthesis and accumulation of multiple esters in plants at the same time (Ma et al., 2005; Balbontín et al., 2010; Nimitkeatkai et al., 2011; Cumplido-Laso et al., 2012; Souleyre et al., 2014). AAT can utilize more kinds of alcohol substrates *in vitro* other than the endogenous substrates, however, the biosynthesis of ester compounds mainly depends on the available substrates in the plant. There might be multiple AATs in the seeds of *W. villosa*, which is the major organ enriched with BA and other monoterpene acetates.

In this study, as part of the ongoing efforts to complete the BA biosynthesis pathway in *W. villosa*, we identified 64 BAHD members from the *W. villosa* genome database and selected 12 WvAAT candidate genes to clone and characterize. Eight genes were identified by biochemical assay to encode for BAT, converting at least one type of borneol to BA, and two of them were further proved to be the key BATs responsible for the BA synthesis in the seeds of *W. villosa*.

MATERIALS AND METHODS

Plant Material

Wurfbainia villosa is from Yangchun City, Guangdong Province, China. The leaves, rhizome, flowers, and fruits from healthy

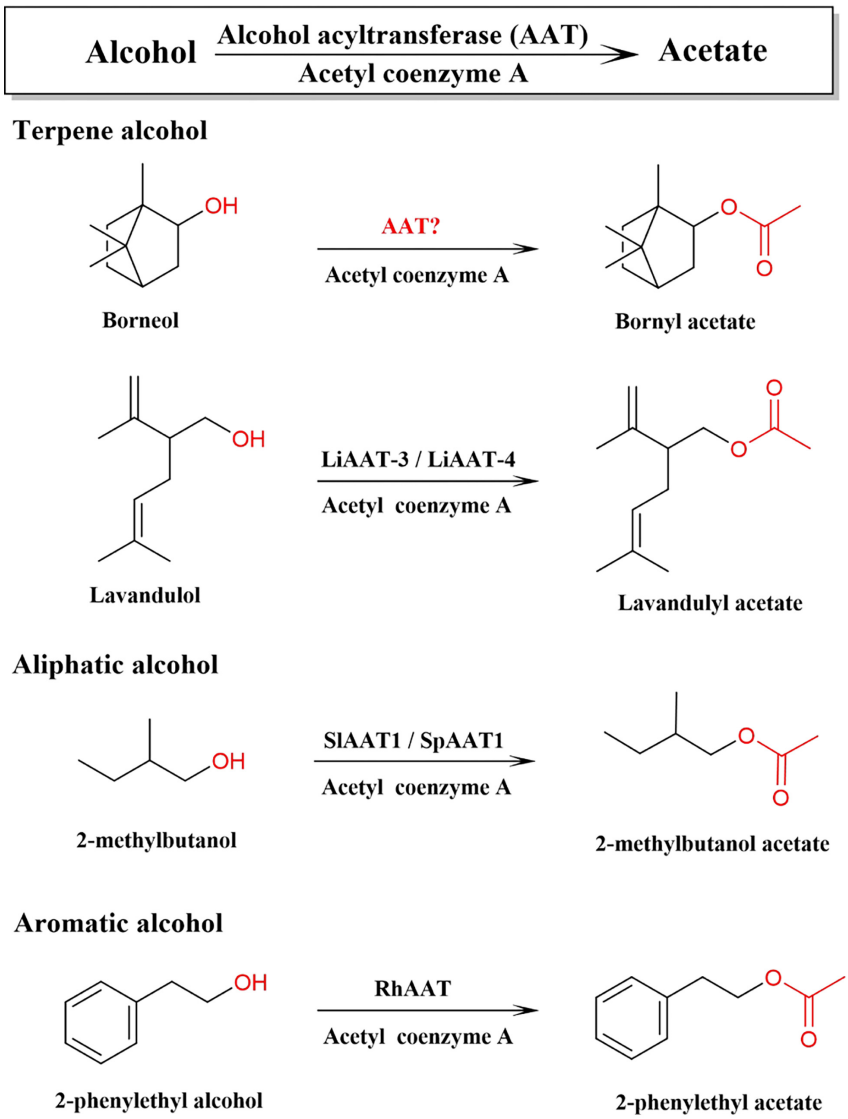


FIGURE 1 | Alcohol acetyltransferase (AATs) involved in the biosynthetic pathway of acetate with different precursor alcohols.

plants were collected and frozen at −80°C. The fruits at different developmental stages: 30-days after flowering (DAF), 45-DAF, 60-DAF, 75-DAF, and 90-DAF were separated into pericarp and seeds.

Genome-Wide Identification of BAHD Superfamily

To identify putative BAHD genes in *W. villosa*, we used the Protein Basic Logical Alignment Search Tool (BLASTP) (*E*-value cutoff of 1e-5) to compare *Arabidopsis thaliana* and *Malus domestica* BAHD protein sequences (Liu et al., 2020) with the *W. villosa* genomic data (unpublished). The amino-acid sequences of *A. thaliana* and *M. domestica* were obtained from phytozome.[1] Then, based on the hidden Markov model (HMM) of the BAHD domain (PF02458) obtained from the Pfam database,[2] we also identified candidate WvBAHDs using Hmmsearch (3.2.1) (Potter et al., 2018). Then, we used the Web CD-Search Tool[3] and Web SMART[4] to confirm the domains of the candidate WvBAHDs identified by the two methods mentioned above, and then their HXXXD and DFGWG motifs were inspected. Furthermore, PF07247 (alcohol acetyltransferase, AATase) was used to screen AAT as the supplement of BAHD genes. Collinear blocks of the WvBAHD were identified using MCSanX and the results of circular plots were generated using Circos (Wang et al., 2012). The amino acid sequences alignments and the maximum-likelihood (ML) phylogenetic tree

[1]https://phytozome.jgi.doe.gov/pz/portal.html#

[2]http://pfam.xfam.org/

[3]https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi

[4]http://smart.embl-heidelberg.de/

with bootstraps of 1,000 were obtained by using the MUSCLE and IQ-TREE of TBtools (Chen et al., 2020a), and the clade of clustering adapted from Liu et al. (2020).

Selection of Candidate WvAAT Genes Involved in Bornyl Acetate Biosynthesis

The WvAAT candidate genes were selected directly by their clade clustered into and their transcripts per million (TPM) expression value (from unpublished RNA-seq data) combined with the upstream gene *WvBPPS* in different tissues and the fruits at different developmental stages. To predict the substrate preference of WvAAT, the multi-sequence alignments and phylogenetic analysis with the reported AATs were performed using the ML method with the tool MEGA X (Kumar et al., 2018). The information of the AATs from other plants used for the phylogenetic analysis was shown in **Supplementary Table 2**. Candidates were further analyzed for a multiple sequence comparison analysis of the nucleotide and amino acid sequences using the software Jalview (Waterhouse et al., 2009). The conserved and active motifs predictions were used by the online tool MEME[5] (Bailey and Elkan, 1994); the protein structure, signal peptide, and transit peptide of the WvAAT candidates were predicted and analyzed by online platform TMHMM 2.0,[6] SignalP,[7] and ChloroP 1.1 Server[8] (Emanuelsson et al., 1999; Krogh et al., 2001; Petersen et al., 2011). According to the results of the sequence alignment phylogenetic tree, the evolutionary tree was beautified by the online tool iTOL[9] (Letunic and Bork, 2021).

Amplification of Full-Length WvAAT Candidate Gene

The extraction method of total RNA and complementary DNA (cDNA) was the same as the previous reports (Wang et al., 2018). The coding regions of AAT candidate genes were amplified from *W. villosa* cDNA using the Primer STAR Max DNA Polymerase (Takara, China) with appropriate primers (**Supplementary Table 3**). The PCR conditions used were the following: 98°C, 1 min; 98°C 10 s, 50–60°C, 15 s, 72°C, 15 s, 30 cycles; 72°C, 5 min. The purified PCR products were then ligated into pLB (Tiangen, China) or 007 vs. (TSINGKE, China) cloning vector, which were consequently transformed into *E. coli* DH5α cells and sequencing.

Prokaryotic Expression and Purification of WvAAT Recombinant Protein

The full-length ORFs of the candidate WvAATs were ligated into pET32a (+) expression vector, using the In-Fusion Cloning Kit (Takara, China), and then transformed into the *Escherichia coli* Rosetta (DE3). The primers were described in **Supplementary Table 3**. Cells were grown at 37°C until the OD_{600} reaches.4–0.6, and induced at 16°C with isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 10 μM for 16–20 h in Luria–Bertani (LB) media supplemented with 50 μg/μL carbenicillin and 25 μg/μL chloramphenicol. The recombinant protein was purified using NI-NTA resin (Qiagen, Hilden, Germany). The purified protein was dialyzed on a PD-10 desalting column (GE Healthcare).

Enzyme Assay and Product Analysis of WvAAT Candidates

The *in vitro* enzymatic reactions were performed at 250–350 μL total volume [20 mM sodium hydrogen phosphate buffer (pH 9.0), 10% glycerol] containing 20–80 μg purified protein, 0.2 mM acetyl-CoA, and 0.02 mM terpene alcohol substrate [(+)-borneol, (−)-borneol, isoborneol, geraniol, nerol, α-terpineol, (−)-menthol, linalool and nerolidol]. According to the previous reports on AAT enzymatic assays (Croteau and Hooper, 1978; Sarker and Mahmoud, 2015) and the results of our preliminary experiments, the optimum pH and temperature for WvAATs were determined to be 9.0 and 32°C, respectively. All reaction mixture was incubated at 32°C for 6 h, then immediately overlaid with 250 μL hexane. The mixture was then centrifuged at 12,000 rpm for 5 min to separate the phases. The hexane extraction was analyzed using Agilent 7890B Gas Chromatograph with 5977A inert Mass Selective Detector (Agilent, United States). Helium was used as the carrier gas (1 mL/min) and then separated on the HP-5MS column (30 m × 250 μm × 0.25 μm film thickness). The gas chromatography (GC) oven temperature was programmed at an initial temperature of 35°C for 2 min with an increase of 12°C/min to 300°C. The temperature was then kept at 300°C for 5 min. For chiral compounds, CycloSil-B column (30 m × 0.25 mm id, 0.25 mm film thickness) was used for separation, initially at 50°C for 2 min, and then at 5°C/min from 50 to 180°C, increase to 230°C at 10°C/min, and hold at 230°C for 2 min. NIST14/Wiley275 Mass Spectral Library was used for metabolite identification. Meanwhile, the standards were also utilized for further identification. Each WvAAT was tested in triplicate.

Volatile Terpenoid Extraction and Analysis

Approximately, 0.1 g of seeds at different developmental stages were ground frizzed in liquid nitrogen and extracted with 1 mL hexane using an ultrasonic cleaner for 30 min, and then incubated at 40°C for 1 h. The samples were then centrifuged at 10,000 rpm for 15 min and the resulting supernatants were pipetted into novel 2 mL tubes. One milliliter of hexane extract was pipetted into 1 mL vial for GC- mass spectrometry (MS) analysis. The extraction was analyzed using Agilent 7890B Gas Chromatograph with 5977A inert Mass Selective Detector (Agilent, United States). Helium was used as the carrier gas (1 mL/min) and then separated on the CycloSil-B column (30 m × 0.25 mm id, 0.25 mm film thickness). The GC oven temperature was programmed at an initial temperature of 35°C for 2 min, and then at 5°C/min from 35to 200°C, increase to 230°C at 10°C/min, and hold at 230°C for 2 min. The temperature was then kept at 240°C for 5 min. NIST14/Wiley275 Mass Spectral Library was used for metabolite identification. The terpene compounds were identified by the

[5]https://meme-suite.org/meme/tools/meme

[6]https://services.healthtech.dtu.dk/service.php?TMHMM-2.0

[7]https://services.healthtech.dtu.dk/service.php?SignalP-4.1

[8]https://services.healthtech.dtu.dk/service.php?ChloroP-1.1

[9]https://itol.embl.de/

mass spectral library. The predominant terpene acetates and their precursor terpene alcohols in this research were further identified using their authentic standards. There were three biological replicates and three technical replicates for each organ.

Quantitative Real-Time PCR of WvBATs

Based on the correlation analysis between transcriptome and ester data of *W. villosa*, eight identified WvBATs were selected for quantitative real-time PCR (qRT-PCR). The primers for real-time PCR of all the WvBAT genes were designed manually (Fluorescence-specific primers shown in **Supplementary Table 3**). The quantitative PCR (qPCR) of WvBATs was performed using 2xTSINGKE®Master qRT-PCR Mix-SYBR (+ UDG, TSINGKE, China) in the CFX96 real-time PCR detection system (Bio-Rad, United States). The WvBAT transcript levels were monitored using the internal reference gene TUA, and calculated using the $2^{-\Delta\Delta Ct}$ method. There were three biological replicates and three technical replicates for all experiments.

Analysis of the WvBATs Promoters

Preliminary element analysis was performed on the approximate area of the promoter. The 2,000 bp non-coding region upstream of the *WvBATs* coding region was extracted through TBtools (Chen et al., 2020a) and then analyzed on the online website PlantCare[10] (Lescot et al., 2002) to find out whether the element was related to seed-specific expression. The classification of the *cis*-acting elements of the promoters of functional genes was referred to Abdullah et al. (2018).

RESULTS

Genome-Wide Identification of WvBAHD Gene Family and WvAAT Candidates

Sixty-one putative *BAHD* genes with complete ORF and two conserved motifs were identified from the genome database of *W. villosa* using the BAHD HMM configuration file (PF02458, transferase). In addition, three genes (*WvBAHD3*, *52,* and *53*) were screened out using PF07247 (AATase) as HMM. AATase (PF07247) comes from CL0149, a protein clan named CoA-acyltrans with a characteristic HXXXD motif, same as the transferase family (PF02458). The AATase family contains a number of AATs from bacteria and metazoa, catalyzing the esterification of isoamyl alcohol by acetyl-CoA, similar to the AATs from the plant (Minetoki et al., 1993; Zhu et al., 2015; Lin et al., 2016; Reyes-Sánchez et al., 2019). Accordingly, we speculated that these three genes screened by PF07247 (AATase) should be members of the *BAHD* superfamily. In total, 64 *WvBAHDs* have been identified as BAHD super acyltransferase family members (**Supplementary Table 4**), which are distributed on 14 chromosomes and nine unanchored contigs (**Figure 2A**). The genome synteny analysis showed that 15 segmental duplications (23.4%) and four tandem duplication (6.3%) events occurred, suggesting that segmental duplication events have an important contribution to the expansion of the *WvBAHD* gene family (**Figure 2B**). The prediction results of the *trans*-membrane function indicated that most *WvBAHDs* were located in the cytoplasm (**Supplementary Table 5**).

To determine the classification of the *WvBAHD* gene family, we constructed a phylogenetic tree using amino acid sequences encoded by 235 *BAHD* genes from *W. villosa*, *A. thaliana*, and *M. domestica*, and 14 identified AATs from other plants. The 64 putative WvBAHDs were classified into five clades: I-a (twenty-two genes), I-b (fifteen genes), II-a (nine genes), II-b (fourteen genes), III-a (one gene), and VI (three genes) (**Figure 3A**). The genes belonging to clade I-a are involved in modifying aromatic and terpenoid alcohols in *A. thaliana* and *M. domestica*, therefore, we speculated that the WvBAHDs clustered into clade I-a in *W. villosa* might encode proteins with similar functions. However, the identified AAT were also clustered into clade I-b and clade IV, which is inconsistent with the reports that the members of clade I-b had functions related to the biosynthesis of lignin monomeric intermediates (Hoffmann et al., 2003; Petrik et al., 2014). This might be due to the phylogenetic tree algorithm and these enzymes have not been comprehensively identified to date. Furthermore, three AATase-WvBAHDs are clustered separately as a novel clade and named clade VI. Therefore, the range of WvAAT candidate genes we inferred here was expanded from clade I-a to clade I-b and clade VI. To confirm the correlation between the above clustering and conserved motifs, we detected eight conserved motifs in the WvBAHDsof *W. villosa* by MEME. Motif1 and Motif2 correspond to conserved motifs HXXXD and DFGWG, respectively (**Figure 3B** and **Supplementary Figures 1, 2**). The similarity in the type and distribution of conserved motifs in the same clade further supports the classification of evolutionary trees. Among the members of clade VI, the histidine residues from the HXXXD motif were substituted by threonine or serine, and four residues from the DFGWG motif were substituted, causing the clade VI members distant relation with other members of BAHD and other identified AATs (**Figures 3A**, **4**).

Then, we compared the expression profiles of the genes from clade I and VI with *WvBPPS*, the key gene for borneol biosynthesis, and identified four genes (*WvBAHD14*, *WvBAHD40*, *WvBAHD39*, *WvBAHD50*; TPM > 10) co-expressed with *WvBPPS* in 60-DAF seeds, suggesting that these genes might be the key genes involved in the borneol acetylation (**Figure 3C**). Furthermore, in clade I-a, WvBAHD8, WvBAHD38, and WvBAHD31 share high identity (>45%) with the reported AATs (CaAT20, LiAAT3, VpAAT1, and FaAAT2); WvBAHD9, WvBAHD14, WvBAHD39, WvBAHD40, and WvBAHD50 are clustered with MsBanAAT into a close branch; in clade I-b, WvBAHD18 and WvBAHD28 share approximately 40% identity with ClLAAT1, an AAT from *Citrus limon* (**Figure 4**). These reported AATs to show enzymatic affinity to geraniol and nerol *in vitro* (Aharoni et al., 2000; **Supplementary Table 2**), therefore, we speculated that these WvBAHDs might have AAT activity similar to these reported AATs. Thence, ten genes (*WvBAHD8*, *WvBAHD9*, *WvBAHD14*, *WvBAHD18*, *WvBAHD28*, *WvBAHD31*, *WvBAHD38*, *WvBAHD39*, *WvBAHD40*, and *WvBAHD50*) from clade I and two genes

[10]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

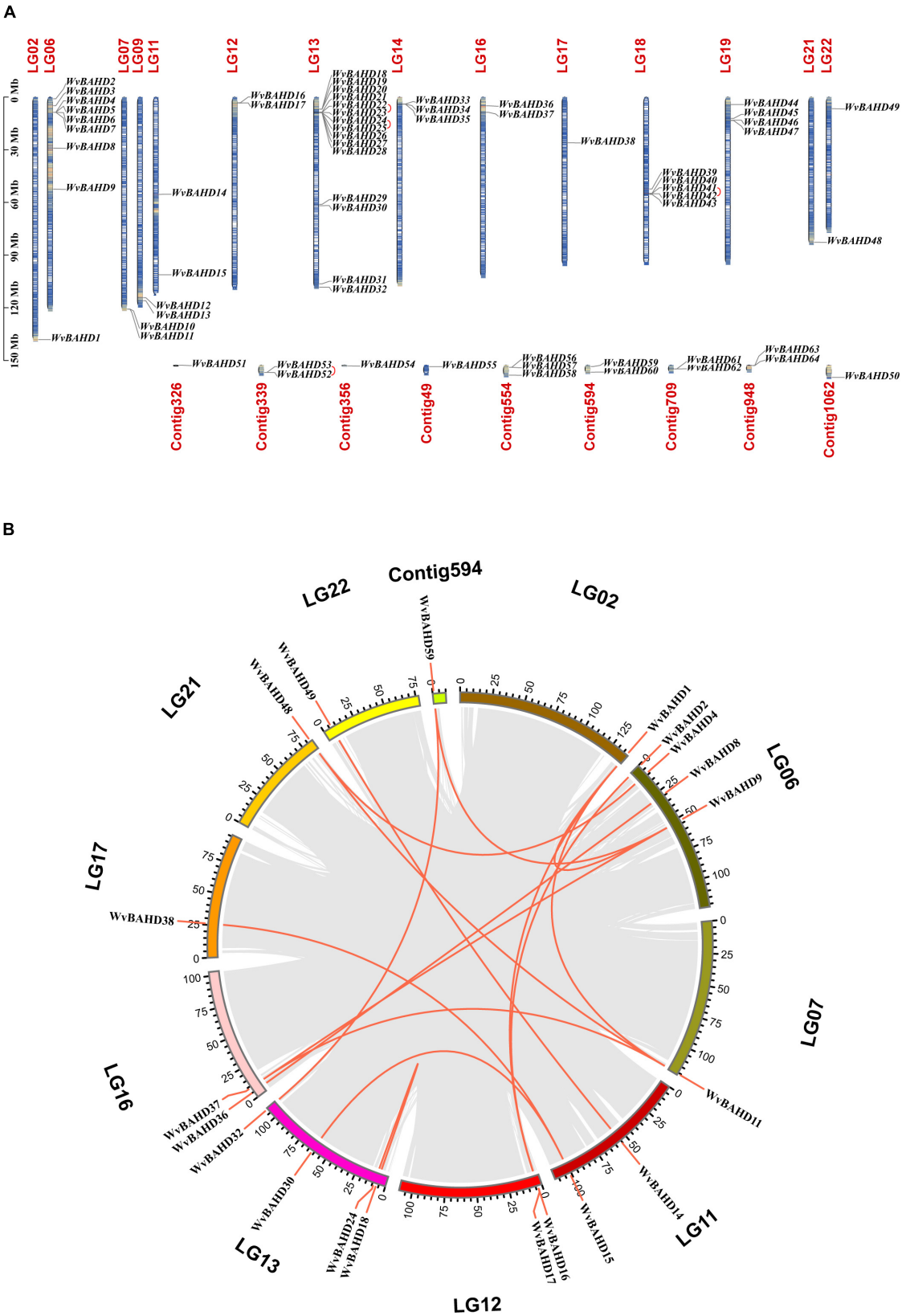


FIGURE 2 | Gene location and collinearity analysis of the WvBAHD gene family. **(A)** Chromosome location and distribution analysis of WvBAHD genes. Tandem duplicated genes are linked by a red line. **(B)** Chromosomal duplication analysis of WvBAHD genes in circos graph. Red lines represent the syntenic gene pairs.

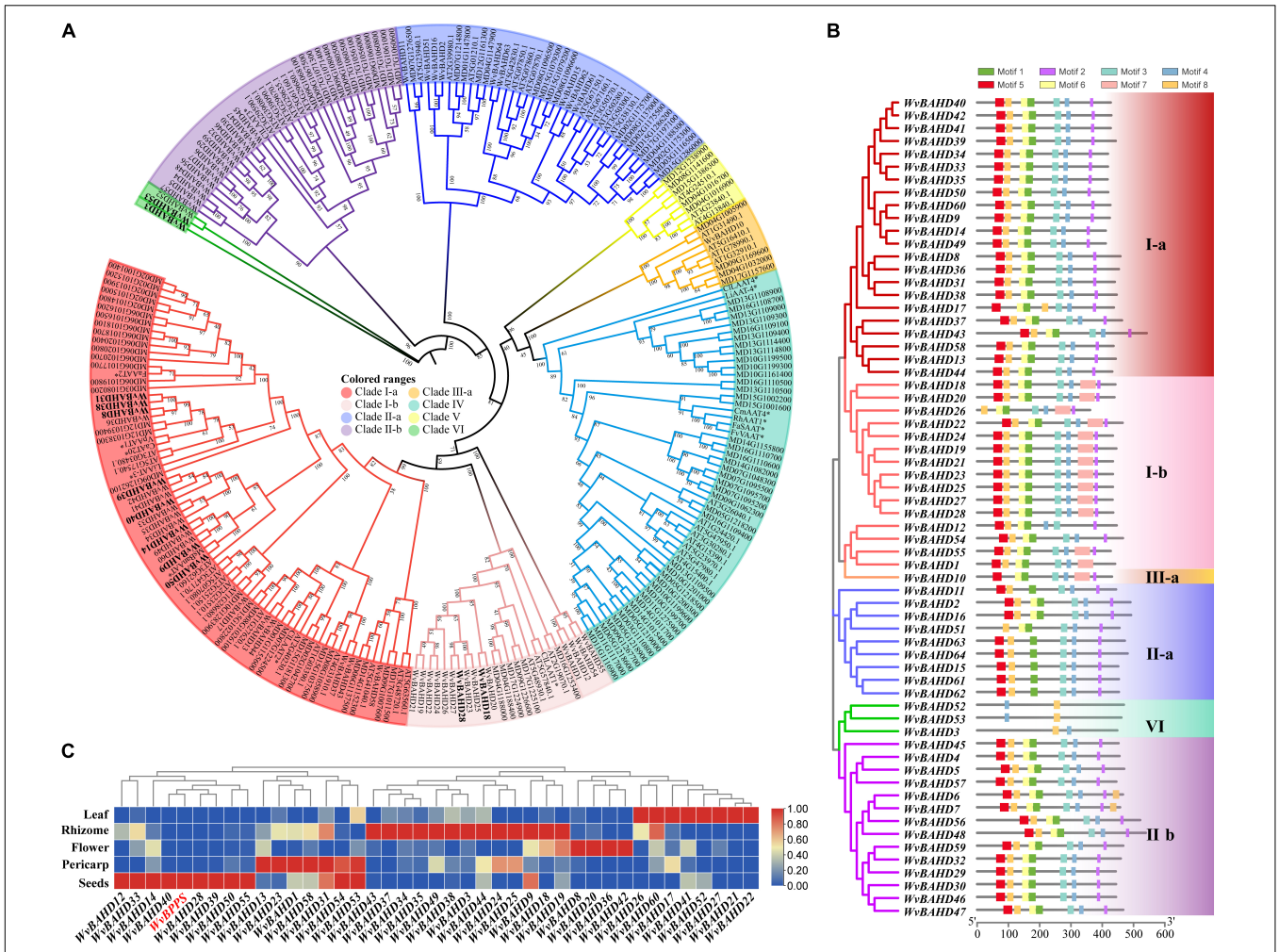


FIGURE 3 | (A) Phylogenetic analysis of WvBAHDs from *Wurfbainia villosa*, AtBAHDs from *Arabidopsis thaliana,* and MdBAHDs from *Malus domestica*, and 14 identified AATs marked with asterisk. The phylogenetic tree was constructed by maximum-likelihood with 1,000 replications, and the clade of clustering was adapted from Liu et al. (2020). The WvAAT candidates were indicated in bold black font. **(B)** Conserved motifs analysis of WvBAHDs. Legend depicting the amino acid sequence of the conserved motifs were shown in **Supplementary Figure 1**. **(C)** Transcriptome differential expression analysis of *WvBAHDs* from clade om clare 1. **(C)** nd *WvBPPS*. Pericarp and seeds were from 60-DAF (Days after flowering) fruits.

(*WvBAHD53* and *WvBAHD3*) from clade VI, which are all expressed in the seeds (**Supplementary Table 6**), were selected as candidate *WvAATs* for cloning and functional characterization.

Eight WvAATs Were Characterized as Borneol Acetyltransferase

Enzymatic assays were conducted *in vitro* using recombinant proteins extracted and purified from *E. coli* expression strains (**Supplementary Figure 3**). The enzymatic assays showed that eight WvAAT recombinant proteins (WvBAHD8, WvBAHD14, WvBAHD28, WvBAHD39, WvBAHD40, WvBAHD50, WvBAHD3, and WvBAHD53) can acetylate geraniol and nerol, which are primary alcohols easily to be catalyzed, to form geranyl acetate and neryl acetate, respectively, even with the small amount of soluble proteins, verifying that these eight candidates have AAT enzymatic activity, and this method is effective for the AAT characterization *in vitro*, according to previous research (Shalit et al., 2003; Sarker and Mahmoud, 2015; **Supplementary Figure 4**). In the case of the other four candidate WvAATs, no geranyl acetate or neryl acetate was detected from the enzymatic reaction, and they could not catalyze bornyl-type substrate either.

To determine the BAT activity of these eight WvAAT recombinant proteins mentioned above, the catalytic products using borneol-type substrates, including (−)-borneol, (+)-borneol, and isoborneol, were analyzed. The results reveal that all these WvAATs can catalyze the substrate (−)-borneol to produce (−)-BA, however, (+)-BA was only detected from the catalytic products of five WvAATs (WvBAHD8, WvBAHD14, WvBAHD39, WvBAHD40, and WvBAHD3) when (+)-borneol was used as substrate (**Figures 5A,B**); we thus concluded that these five WvAATs capable to catalyze both (−)-borneol and (+)-borneol have no strict substrate specificity for different

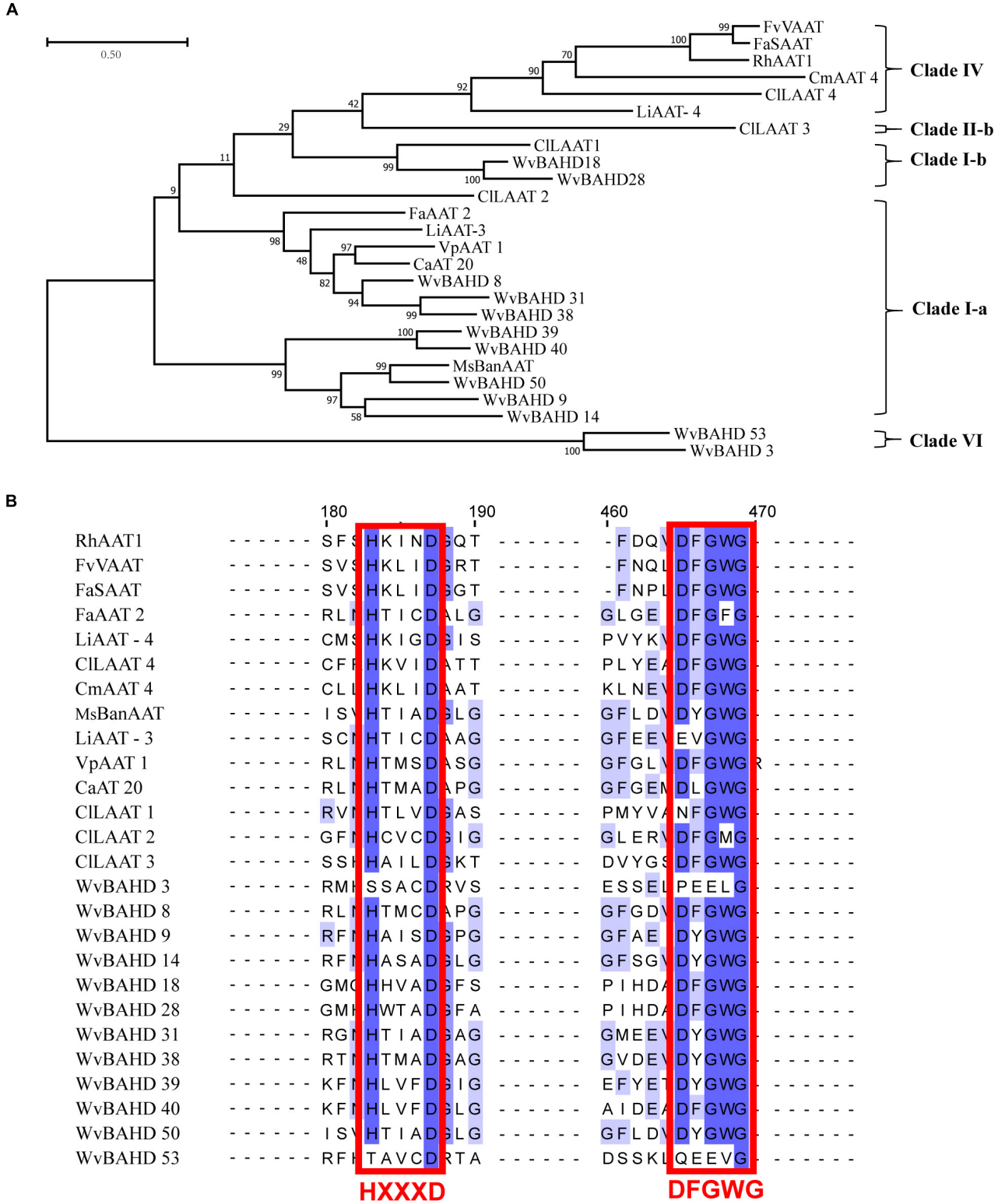


FIGURE 4 | (A) Phylogenetic analysis of 12 WvAAT candidate genes and 14 identified AATs from other plants. The phylogenetic tree was constructed by maximum likelihood with 1,000 replications, and the clade of clustering was adapted from Liu et al. (2020). **(B)** Sequence comparison of HXXXD and DFGWG motifs in WvAATs and identified AATs proteins. The percentage identify threshold was 30% or above. Conserved motifs are highlighted in red. **Supplementary Table 2** contains the GenBank accession numbers of each AAT sequence used.

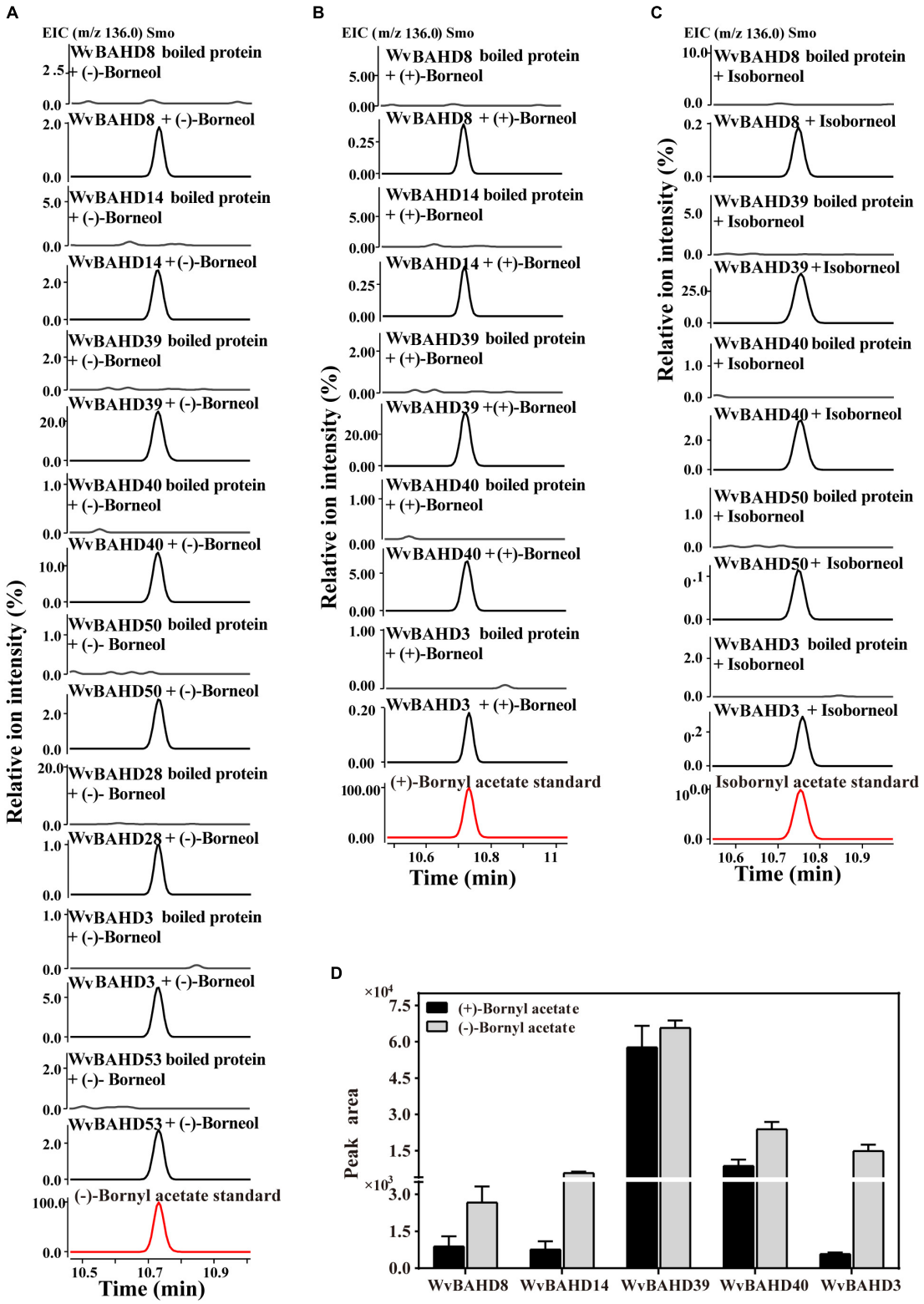


FIGURE 5 | Functional characterization of WvBAHDs. **(A–C)** The gas chromatography (GC)–mass spectrometry (MS) chromatogram of the *in vitro* reaction products yielded by each WvBAHD using (−)-borneol, (+)-borneol, and isoborneol as the substrate, respectively. **(D)** The GC-MS chromatograms peak area (EIC, m/z 136.0) of products yielded by each WvBAHD at the same substrate concentration (enzyme concentrations vary among WvBAHDs), using (−)-borneol, (+)-borneol, and isoborneol as the substrate, respectively.

optical isomers of borneol, while WvBAHD28, WvBAHD50, and WvBAHD53 have substrate specificity for (−)-borneol. Among these five WvAATs capable to catalyze both (−)-borneol and (+)-borneol, WvBAHD8, WvBAHD39, WvBAHD40, and WvBAHD3 can also catalyze isoborneol to produce IBA (**Figure 5C**). In addition to borneol-type substrate, more diverse substrates were used to perform the enzyme assay, and none of these WvAATs could react with tertiary alcohols, such as (−)-menthol, linalool, or nerolidol, to form relative acetate products, similar to that of LiAAT3 and LiAAT4 (Sarker and Mahmoud, 2015). Since all these WvAATs can convert at least one type of borneol to BA, demonstrating that they have the BAT activity, therefore they are renamed as WvBAT1–8, respectively (**Table 1** and **Supplementary Table 7**). To compare the catalytic activity for different optical isomers of borneol substrate, the product peak areas of the five WvBATs which can catalyze both (−)-borneol and (+)-borneol, reacting with the same amount of each substrate, were analyzed. The results indicated that these WvBATs had the better catalytic ability for (−)-borneol than for (+)-borneol; notably, WvBAT3 (WvBAHD39) exhibited the best catalytic efficiency for (+)-borneol than the other WvBATs, followed by WvBAT4 (WvBAHD40) (**Figure 5D**). In addition, WvBAT3 is the only enzyme capable to acetylate the α-terpineol to synthesize α-terpinyl acetate (**Table 1** and **Supplementary Figure 5**).

The Optical Configuration of Borneol and Bornyl Acetate in the Seeds of *Wurfbainia villosa*

As we mentioned above, most of the WvBATs present a better catalytic ability for (−)-borneol *in vitro*. (−)-Borneol and (+)-borneol are optical isomers and both of them are natural metabolites in plants, however, the optical configuration of endogenous borneol in the seeds of *W. villosa* and the WvBPPS-catalyzed product hasn't been identified with the chiral column to date (Wang et al., 2018). Therefore, we used GC-MS with CycloSil-B column to identify the chiral configuration of borneol in the seeds of *W. villosa*, and that of the product in the enzymatic reaction of WvBPPS. The results showed that no (−)-borneol was detected but a significant amount of (+)-borneol and a small amount of isoborneol were detected in the mature (90-DAF) seeds of *W. villosa* (**Figure 6A**); this result was also observed in the volatile extract from seeds at different developmental stages (30-DAF to 75-DAF), and consistent with the observation that (+)-borneol was the WvBPPS-catalyzed product with phosphatase treatment (**Figure 6B**). Furthermore, the content of (+)-BA rises with the increasing of (+)-borneol throughout the seed development, peaking at 90 DAF; (+)-borneol has a substantially larger content than isoborneol, which matches the data that (+)-BA has a significantly higher amount than IBA (**Figure 6C**). These results suggest that (+)-borneol is the major endogenous borneol-type substrate in the seeds of *W. villosa*.

Correlation Between Gene Expression Levels of WvBATs With Acetates Accumulation in *Wurfbainia villosa*

To analyze the temporal and spatial expression patterns of *WvBATs*, qRT-PCR was performed using the leaf, rhizome, flower, and 60-DAF fruit (separated into pericarp and seeds), and the seeds at five developmental stages. The result demonstrated that the relative expression levels of *WvBATs* were basically consistent with the TMP (transcripts per million) expression value from the RNA-sequencing (RNA-seq) data, except for *WvBAT5*; *WvBAT5* was actually expressed higher in rhizome than in the seeds. The expressional patterns of *WvBAT3*, *WvBAT4*, and *WvBAT6* are seed-specific, similar to that of *WvBPPS*. Although *WvBAT6* was expressed specifically in the seeds, the enzyme it encodes cannot catalyze (+)-borneol and isoborneol, indicating that it might participate in the biosynthesis of other acetates in the seeds. *WvBAT1* was expressed specifically in the flower, suggesting that it might be involved in the acetates biosynthesis in the flower. Furthermore, *WvBAT7* was expressed in all the organs with the highest level in the rhizome,

TABLE 1 | The information of the characterized WvBATs in *Wurfbainia Villosa*.

Family in Pfam (HMMs)	BAHD family	Original ID	Renamed ID	Protein size (aa)	Protein molecular weight (kDa)	Catalytic substrate					
						(+)-Borneol	(−)-Borneol	Isoborneol	Geraniol	Nerol	α- Terpineol
Transferase (PF02458)	Clade I-a	WvBAHD8	WvBAT1	458	49.76	✓	✓	✓	✓	✓	n.d.
		WvBAHD14	WvBAT2	411	45.93	✓	✓	n.d.	✓	✓	n.d.
		WvBAHD39	WvBAT3	443	49.65	✓	✓	✓	✓	✓	✓
		WvBAHD40	WvBAT4	427	47.35	✓	✓	✓	✓	✓	n.d.
		WvBAHD50	WvBAT5	419	46.05	n.d.	✓	✓	✓	✓	n.d.
	Clade I-b	WvBAHD28	WvBAT6	435	47.21	n.d.	✓	n.d.	✓	✓	n.d.
AATase (PF07247)	Clade VI	WvBAHD3	WvBAT7	488	48.64	✓	✓	✓	✓	✓	n.d.
		WvBAHD53	WvBAT8	461	50.75	n.d.	✓	n.d.	✓	✓	n.d.

"✓" indicated that the substrate can be converted to the corresponding acetate; "n.d." indicated that no corresponding acetate was detected in the enzymatic reaction.

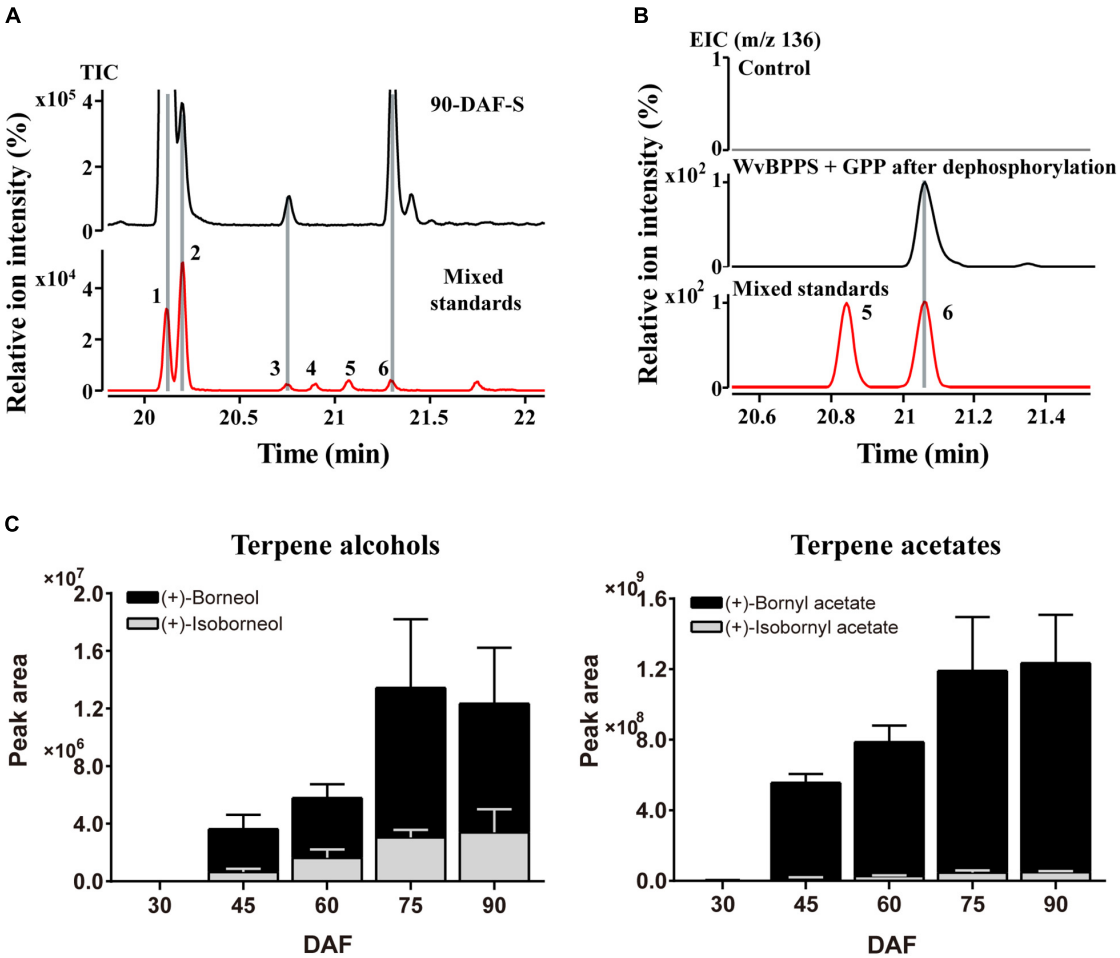


FIGURE 6 | (A) The GC-MS chromatogram of the volatile terpene in the 90-DAF seeds of *Wurfbainia villosa*, and the mixed standards. **(B)** The GC-MS chromatogram of the products generated by WvBPPS protein after dephosphorylation and the mixed-borneol standard and its negative control was treated with boiled protein. The compounds of peaks: 1. bornyl acetate; 2. isobornyl acetate; 3. (+)-isoborneol; 4. (−)-isoborneol; 5. (−)-borneol; 6. (+)-borneol. **(C)** The GC-MS chromatographic peak (EIC, m/z 136.0) areas of (+)-borneol, (+)-isoborneol and their corresponding acetates (+)-bornyl acetate and (+)-isobornyl acetate, in seeds at different developmental stages of *W. villosa*.

indicating that it might be involved in the BA synthesis in the rhizome, which contains low levels of BA (Wang et al., 2018; **Figure 7A**).

As we mentioned above, (+)-borneol is the major endogenous substrate in the seeds, and BA accumulates in the seeds with an obvious occurrence at 45-DAF (**Figure 6C**); only four *WvBATs* (*WvBAT2*, *WvBAT3*, *WvBAT4*, and *WvBAT7*) are capable to catalyze (+)-borneol and expressed in the seeds as well. Therefore, to determine the major BAT(s) responsible for the BA biosynthesis, we investigated their expression patterns in the seeds at different developmental stages and compared their expression levels in 45-DAF and 60-DAF seeds. Of which, the expression levels of *WvBAT2* and *WvBAT7* peaked at 45 DAF, while their expression levels were much lower than *WvBAT4* (**Figures 7B,C**). Notably, both *WvBAT3* and *WvBAT4*, the seed-specific expressed genes, were not expressed at 30-DAF, but started to be highly expressed at the 45-DAF stage, similar to that the highest expression level of *WvBPPS* presented in 45-DAF, in accord with the initial accumulation levels of (+)-borneol and (+)-BA in the seeds (Wang et al., 2018; **Figure 7B**). In addition, *WvBAT4* was the highest-expressed gene in the seeds at 45-DAF and 60-DAF, and the expression level of *WvBAT3* in 60-DAF seeds is also higher than the other two *WvBATs* (**Figure 7C**). Considering WvBAT3 (WvBAHD39) is the enzyme presenting the best catalytic efficiency on (+)-borneol (**Figure 5D**), we speculated that both WvBAT3 and WvBAT4 are the key enzymes responsible for the synthesis of BA. In addition, IBA initial accumulation in the seeds also increased from 45-DAF (**Figure 6C**), implying that the WvBATs (WvBAT3, WvBAT4, and WvBAT7) enable to catalyze (+)-isoborneol, were also involved in the IBA synthesis in the seeds of *W. villosa*.

Promoter Analysis of the WvBATs

To ascertain the potential biological roles of WvBATs in *W. villosa*, the element analysis was performed on the approximate area of the promoter of *WvBAT1*–8 genes. In

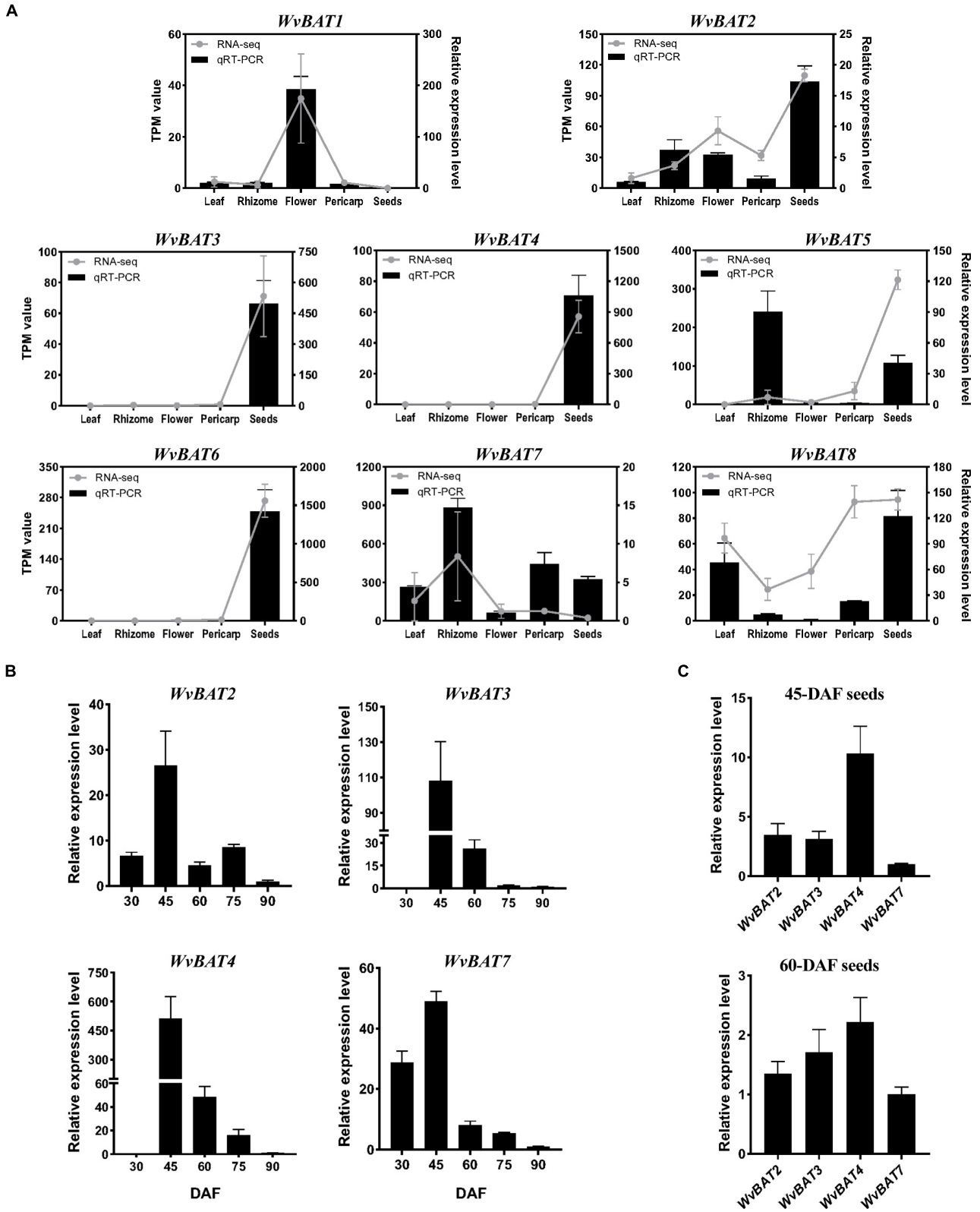


FIGURE 7 | Expression pattern of *WvBATs*. **(A)** The relative expression levels of *WvBATs* in different tissues. Pericarp and seeds were from 60-days after flowering (DAF) fruits. Transcripts per million reads (TPM) value was from the RNA-sequencing (RNA-seq) data. Comparison of the expression levels of the *WvBATs*, capable to catalyze (+)-borneol, in seeds at different developmental stages **(B)** or at the same stage, 45-DAF or 60-DAF **(C)**. Data represent the means ± SDs (n = 3).

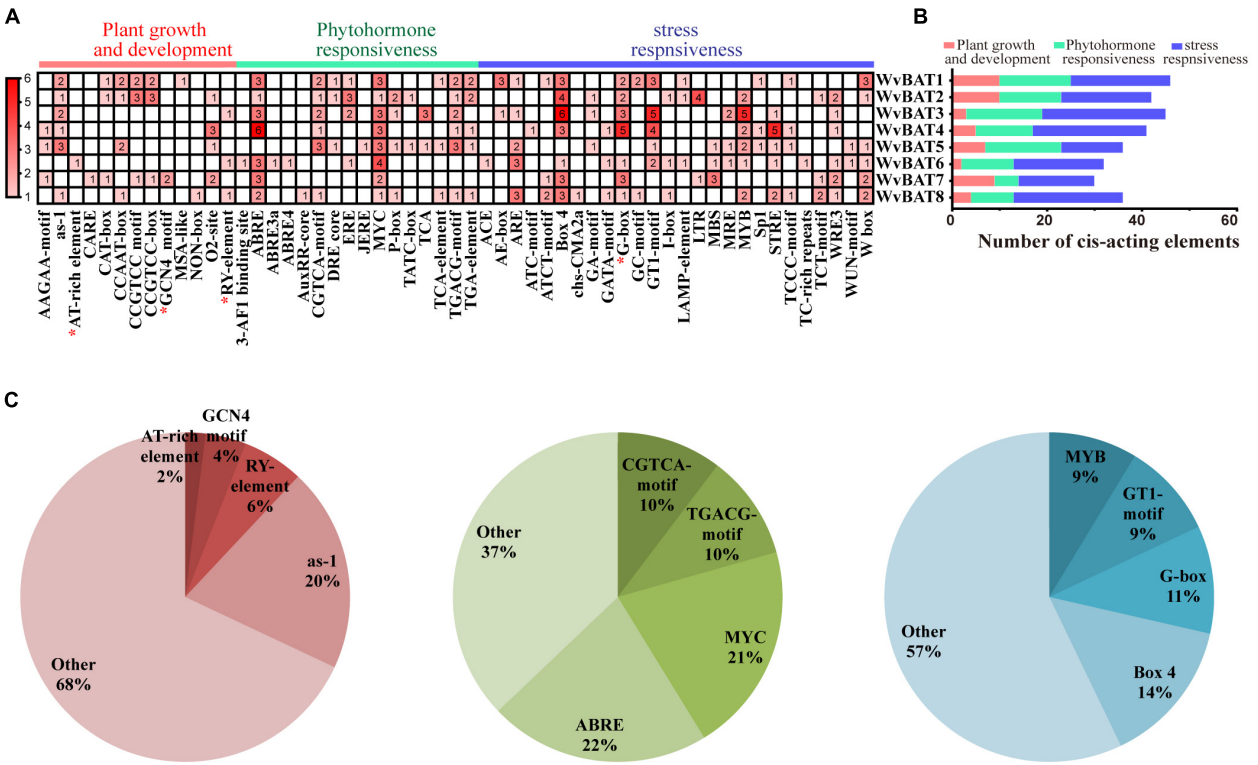


FIGURE 8 | (A) The number of *cis*-acting elements in *WvBATs*, indicated by the intensity of the red color and numbers in the grid. The cis-acting elements with seed-special expression were indicated by a red asterisk. **(B)** Histograms indicate the number of *cis*-acting elements involved in plant growth and development, phytohormone responsiveness, or stress responsiveness. **(C)** Pie charts showing the ratio of different *cis*-acting elements in each structural category.

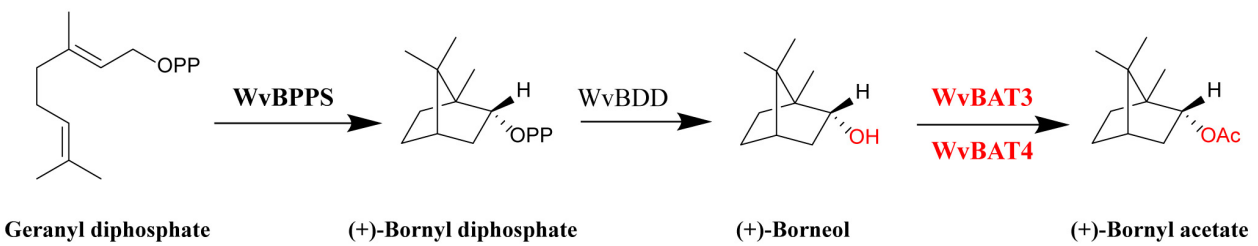


FIGURE 9 | Biosynthetic pathway of bornyl acetate (BA) in *Wurfbainia villosa*. BPPS: bornyl diphosphate synthase; BDD, bornyl diphosphate diphosphatase; BAT, bornyl acyltransferase. The WvBAT in red was identified to be the key enzyme involved in the synthesis of bornyl acetate (BA) in seeds.

total, 308 *cis*-acting elements were identified, and they were grouped into three categories, phytohormone responsiveness, stress responsiveness, and plant growth and development (**Figures 8A,B**). The greatest proportion of *cis*-acting regulatory elements related to phytohormone response was ABRE elements that were associated with abscisic acid (ABA, 22%), followed by MYC (21%), TGACG-motif (10%), and CGTCA-motif (10%); the latter three elements were related to methyl jasmonate (MeJA). Nearly half of all the *cis*-acting elements were associated with stress responsiveness (161/304), of which the most common 4 *cis*-acting elements were the box 4 (14%), G-box (11%), GT1-motif (9%), and MYB (9%), the first three were associated with responsiveness to light, and the last one was related to water stress (**Figure 8C**). These results suggested that *WvBATs* might be induced or suppressed by MeJA and ABA, and involved in plant responses to a variety of abiotic stressors.

We previously observed that promoters of *WvBPPS* featured a seed-specific expression element GCN4-motif, which was consistent with the main accumulation of borneol in the seeds (Zhao et al., 2021a). The seed-expression *cis*-acting elements include the GCN4 motif involved in the endosperm *cis*-elements (Onodera et al., 2001), RY-element, and AT-rich element involved in seed-specific regulation (Bäumlein et al., 1992; Ellerström et al., 1996), and G-box (Izawa et al., 1993, 1994). Expectedly, excluding *WvBAT5*, these *WvBAT* promoters all contained *cis*-acting elements related to seed expression; for example,

the *WvBAT7* promoter contained two GCN4 motifs, and the promoters of *WvBAT1*, *WvBAT2*, *WvBAT4*, *WvBAT6*, and *WvBAT7* contained G-box.

DISCUSSION

A Novel Pfam Hidden Markov Model PF07247 for Identifying BAHD Acyltransferase Genes in Plant

There are relatively few studies devoted to the HMM of the BAHD super acyltransferase family for screening BAHD members in plants. PF02458, the HMM of the transferase family, which belong to CL0149 with a characteristic HXXXD motif, was obtained by comparing the HMM in the Pfam database of reported members of the BAHD family, which was consistent with the results of previous studies (Liu et al., 2020; Kumar et al., 2021). We identified 61 WvBAHDs by PF02458 from *W. villosa* genomic data, featured "HXXXD" and "DFGWG," which is consistent with the characteristic of the BAHD family.

In our investigation, besides PF02458, the other Pfam HMM PF07247 of the AATase family, which belongs to CL0149 as well, was also used to identify the BAHD family members of *W. villosa*. Accordingly, three WvBAHDs (WvBAHD3, WvBAHD52, and WvBAHD53) annotated as AATase had been screened out. However, there was no reported HMM profile of BAHD members similar to PF07247. Actually, "AATase" usually describes the AAT of bacteria, metazoan, and fungi, such as yeast *Saccharomyces* and *Kluyveromyces* (Zhu et al., 2015; Reyes-Sánchez et al., 2019). To further confirm whether these three AATase-WvBAHDs belong to the BAHD acyltransferase family, we constructed another phylogenetic tree using 64 WvBAHDs and the identified BAHDs from other species (**Supplementary Table 8**) based on the cluster system from D'Auria, which clusters BAHD acyltransferases into six major groups (D'Auria, 2006; Sarker and Mahmoud, 2015). Unexpectedly, these three AATase-WvBAHDs were all clustered into the clade V, which consists mostly of AATs involved in volatile ester biosynthesis (**Supplementary Figure 6**); however, they are not clustered closely with the other identified AATs from plants, which might be due to amino acid residue substitutions in their DFGWG motif, one of the conserved motifs of BAHD acyltransferase family. Notably, WvBAHD3 and WvBAHD53 do have the activity of AAT, and both can acetylate borneol, geraniol, and nerol (**Table 1**). Therefore, this work has verified "PF07247/AATase" is feasible for screening BAHD superfamily members in the plant, thus, the BAHD family would be enriched by adding new AAT members.

The First Report on the Borneol Acetyltransferases Responsible for the Last Biosynthetic Step of Bornyl Acetate

Bornyl acetate is the aromatic monoterpene ester with a wide range of pharmacological and biological activities, but only a few plants accumulate a significant amount of BA. BPPS, the enzyme catalyzing the synthesis of borneol precursor, the first step on BA downstream biosynthesis pathway of BA, has been identified from other plants, including *Salvia officinalis* (SoBPPS), *Lippia dulcis* (LdBPPS), *Lavandula angustifolia* (LaBPPS), and *Cinnamomum burmannii* (CbBPPS) (Wise et al., 1998; Despinasse et al., 2017; Hurd et al., 2017; Ma et al., 2021); except for *Lippia dulcis*, these plants all contain BA (Al-Dhubiab, 2012; Arceusz et al., 2013), however, the gene encoding BAT, catalyzing the acetylation of borneol, has not yet been reported. In this study, based on the identification of the BAHD family, we cloned and characterized eight *WvBATs* from *W. villosa*. This is the first report on the BAT responsible for the last biosynthetic step of BA. These WvBATs can acetylate at least one type of borneol to synthesize BA, and they are substrate-promiscuous enzymes, similar to other AATs. AAT is the substrate-promiscuous enzyme, which can acetylate different types of alcohols and acyl-CoA to synthesize aromatic volatile ester compounds in plants (Beekwilder et al., 2004; Souleyre et al., 2005). However, the type and content of acylation products depend on the presence and content of endogenous substrates in plants. For example, the type of acetate ester synthesized in mint depends not on the specificity of AATs, but on the availability of the terpene alcohol in mint oil (Croteau and Hooper, 1978). With regard to *W. villosa*, the volatile oil of the seeds contains high levels of (+)-borneol, much lower levels of isoborneol, and only trace levels of α-terpineol, accordingly, (+)-borneol is the predominant endogenous substrate for WvBATs in the seeds. Two WvBATs, WvBAT3 and WvBAT4, capable of acetylating (+)-borneol with higher efficiency, and expressed in the seeds with positive correlation to the BA accumulation, might be the key enzymes collaborate to synthesize BA in the seeds of *W. villosa*. This speculation is consistent with the previous reports that the plants contain several AATs which might collaborate in the synthesis of a single ester (Cumplido-Laso et al., 2012; Sarker and Mahmoud, 2015).

In summary, based on the biochemical characterization of WvBATs in this study and previous literature (Wise et al., 1998; Despinasse et al., 2017; Wang et al., 2018), we illuminated the downstream biosynthetic pathway of BA in *W. villosa* (**Figure 9**). This work will contribute to a better understanding of the biosynthesis of BA and other acetates in *W. villosa*, and lay the foundation for further studies on BA biosynthesis and metabolism engineering of BA in other plants or heterologous hosts.

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

JY and HL designed the experiments and wrote the manuscript. HL, XL, PY, YS, QW, SA, and HZ performed the experiments

and analyzed the data. RZ and DM revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (No. 81872954) and the Key-Area Research and Development Program of Guangdong Province (No. 2020B20221001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.860152/full#supplementary-material

REFERENCES

Abdullah, M., Cheng, X., Cao, Y., Su, X., Manzoor, M. A., Gao, J., et al. (2018). Zinc finger-homeodomain transcriptional factors (ZHDs) in upland cotton (*Gossypium hirsutum*): genome-wide identification and expression analysis in fiber development. *Front. Genet.* 9:357. doi: 10.3389/fgene.2018.00357

Aharoni, A., Luecker, J., Verhoeven, H. A., Van, T., and O'Connell, A. P. (2000). *Fruit Flavour Related Genes and Use Thereof*. [Accessed on Nov 30, 2021]

Al-Dhubiab, B. (2012). Pharmaceutical applications and phytochemical profile of *Cinnamomum burmannii*. *Phcog. Rev.* 6:125. doi: 10.4103/0973-7847.99946

Ao, H., Wang, J., Chen, L., Li, S., and Dai, C. (2019). Comparison of volatile oil between the fruits of *Amomum villosum* Lour and *Amomum villosum* Lour var xanthioides T. L. Wu et Senjen based on GC-MS and chemometric techniques. *Molecules* 24:1663. doi: 10.3390/molecules24091663

Arceusz, A., Occhipinti, A., Capuzzo, A., and Maffei, M. E. (2013). Comparison of different extraction methods for the determination of α- and β-thujone in sage (*Salvia officinalis* L.) herbal tea. *J. Sep. Sci.* 36, 3130–3134. doi: 10.1002/jssc.201300206

Asghari, G., Jalali, M., and Sadoughi, E. (2012). Antimicrobial activity and chemical composition of essential oil from the seeds of *Artemisia aucheri* Boiss. *Jundishapur J. Nat. Pharm. Prod.* 7, 11–15. doi: 10.5812/kowsar.17357780.3530

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Balbontín, C., Gaete-Eastman, C., Fuentes, L., Figueroa, C. R., Herrera, R., Manriquez, D., et al. (2010). VpAAT1, a gene encoding an alcohol acyltransferase, is involved in ester biosynthesis during ripening of mountain papaya fruit. *J. Agric. Food Chem.* 58, 5114–5121. doi: 10.1021/jf904296c

Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D., and Wobus, U. (1992). Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2, 233–239. doi: 10.1046/j.1365-313x.1992.t01-45-00999.x

Beekwilder, J., Alvarez-Huerta, M., Neef, E., Verstappen, F. W. A., Bouwmeester, H. J., and Aharoni, A. (2004). Functional characterization of enzymes forming volatile esters from strawberry and banana. *Plant Physiol.* 135, 1865–1878. doi: 10.1104/pp.104.042580

Chánique, A. M., Dimos, N., Drienovská, I., Calderini, E., Pantín, M. P., Helmer, C. P. O., et al. (2021). A structural view on the stereospecificity of plant borneol-type dehydrogenases. *ChemCatChem* 13, 2262–2277. doi: 10.1002/cctc.202100110

Chen, L. X., Lai, Y.-F., Zhang, W.-X., Cai, J., Hu, H., Wang, Y., et al. (2020b). Comparison of volatile compounds in different parts of fresh *Amomum villosum Lour.* from different geographical areas using cryogenic grinding combined HS-SPME-GC-MS. *Chin. Med.* 15:97. doi: 10.1186/s13020-020-00377-z

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

Commission of Chinese Materia Medica. (1999). *Chinese Materia Medica, Vol. 8*. Shanghai: Shanghai Science and Technology Publishing House.

Commission of Chinese Pharmacopoeia. (2015). *Pharmacopoeia of the People's Republic of China*. Beijing: China Medico-Pharmaceutical Science.

Croteau, R., and Hooper, C. L. (1978). Metabolism of monoterpenes: acetylation of (-)-menthol by a soluble enzyme preparation from peppermint (*Mentha piperita*) leaves. *Plant Physiol.* 61, 737–742. doi: 10.1104/pp.61.5.737

Cumplido-Laso, G., Medina-Puche, L., Moyano, E., Hoffmann, T., Sinz, Q., Ring, L., et al. (2012). The fruit ripening-related gene FaAAT2 encodes an acyl transferase involved in strawberry aroma biogenesis. *J. Exp. Bot.* 63, 4275–4290. doi: 10.1093/jxb/ers120

Cutillas, A.-B., Carrasco, A., Martinez-Gutierrez, R., Tomas, V., and Tudela, J. (2018). Thymus mastichina L. essential oils from Murcia (Spain): composition and antioxidant, antienzymatic and antimicrobial bioactivities. *PLos One* 13:e0190790. doi: 10.1371/journal.pone.0190790

D'Auria, J. C. (2006). Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* 9, 331–340. doi: 10.1016/j.pbi.2006.03.016

Despinasse, Y., Fiorucci, S., Antonczak, S., Moja, S., Bony, A., Nicolè, F., et al. (2017). Bornyl-diphosphate synthase from *Lavandula angustifolia*: a major monoterpene synthase involved in essential oil quality. *Phytochemistry* 137, 24–33. doi: 10.1016/j.phytochem.2017.01.015

Dhar, N., Sarangapani, S., Reddy, V. A., Kumar, N., Panicker, D., Jin, J., et al. (2020). Characterization of a sweet basil acyltransferase involved in eugenol biosynthesis. *J. Exp. Bot.* 71, 3638–3652. doi: 10.1093/jxb/eraa142

Drienovská, I., Kolanoviæ, D., Chánique, A., Sieber, V., Hofer, M., and Kourist, R. (2020). Molecular cloning and functional characterization of a two highly stereoselective borneol dehydrogenases from *Salvia officinalis* L. *Phytochemistry* 172:112227. doi: 10.1016/j.phytochem.2019.112227

Ellerström, M., Stålberg, K., Ezcurra, I., and Rask, L. (1996). Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. *Plant Mol. Biol.* 32, 1019–1027. doi: 10.1007/BF00041385

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978

Fidan, H., Stefanova, G., Kostova, I., Stankov, S., Damyanova, S., Stoyanova, A., et al. (2019). Chemical composition and antimicrobial activity of *Laurus nobilis* L essential oils from Bulgaria. *Molecules* 24:804. doi: 10.3390/molecules24040804

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., and Legrand, M. (2003). Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.* 278, 95–103. doi: 10.1074/jbc.M209362200

Hurd, M. C., Kwon, M., and Ro, D.-K. (2017). Functional identification of a *Lippia dulcis* bornyl diphosphate synthase that contains a duplicated, inhibitory arginine-rich motif. *Biochem. Biophys. Res. Commun.* 490, 963–968. doi: 10.1016/j.bbrc.2017.06.147

Izawa, T., Foster, R., and Chua, N. H. (1993). Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* 230, 1131–1144. doi: 10.1006/jmbi.1993.1230

Izawa, T., Foster, R., Nakajima, M., Shimamoto, K., and Chua, N. H. (1994). The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. *Plant Cell* 6, 1277–1287. doi: 10.1105/tpc.6.9.1277

Khine, A. A., Yang, M.-Y., Hu, A., Lin, G.-H., Toh, Y.-H., and Chen, H.-P. (2020). Production of optically pure (–)-borneol by *Pseudomonas monteilii*

TCU-CK1 and characterization of borneol dehydrogenase involved. *Enzyme Microb. Technol.* 139:109586. doi: 10.1016/j.enzmictec.2020.109586

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11 Edited by F. *Cohen. J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Kumar, G., Kumar, P., Kapoor, R., Lore, J. S., Bhatia, D., and Kumar, A. (2021). Characterization of evolutionarily distinct rice BAHD-Acyltransferases provides insight into their plausible role in rice susceptibility to *Rhizoctonia solani*. *Plant Genome* 14:e20140. doi: 10.1002/tpg2.20140

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Lin, J.-L., Zhu, J., and Wheeldon, I. (2016). Rapid ester biosynthesis screening reveals a high activity alcohol-O-acyltransferase (AATase) from tomato fruit. *Biotechnol. J.* 11, 700–707. doi: 10.1002/biot.201500406

Liu, C., Qiao, X., Li, Q., Zeng, W., Wei, S., Wang, X., et al. (2020). Genome-wide comparative analysis of the BAHD superfamily in seven *Rosaceae* species and expression analysis in pear (*Pyrus bretschneideri*). *BMC Plant Biol.* 20:14. doi: 10.1186/s12870-019-2230-z

Liu, P., Liu, X.-C., Dong, H.-W., Liu, Z.-L., Du, S.-S., and Deng, Z.-W. (2012). Chemical composition and insecticidal activity of the essential oil of *Illicium pachyphyllum* fruits against two grain storage insects. *Molecules* 17, 14870–14881. doi: 10.3390/molecules171214870

Ma, R., Su, P., Guo, J., Jin, B., Ma, Q., Zhang, H., et al. (2021). Bornyl diphosphate synthase from *Cinnamomum burmanni* and its application for (+)-borneol biosynthesis in Yeast. *Front. Bioeng. Biotechnol.* 9:631863. doi: 10.3389/fbioe.2021.631863

Ma, X., Koepke, J., Panjikar, S., Fritzsch, G., and Stöckigt, J. (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.* 280, 13576–13583. doi: 10.1074/jbc.M414508200

Minetoki, T., Bogaki, T., Iwamatsu, A., Fujii, T., and Hamachi, M. (1993). The purification, properties and internal peptide sequences of alcohol acetyltransferase isolated from *Saccharomyces cerevisiae* Kyokai No. 7. *Biosci. Biotechnol. Biochem.* 57, 2094–2098. doi: 10.1271/bbb.57.2094

Nimitkeatkai, H., Shishido, M., Okawa, K., Ohara, H., Ban, Y., Kita, M., et al. (2011). Effect of jasmonates on ethylene biosynthesis and aroma volatile emission in Japanese apricot infected by a pathogen (*Colletotrichum gloeosporioides*). *J. Agric. Food Chem.* 59, 6423–6429. doi: 10.1021/jf2010996

Ohtsubo, S., Fujita, T., Matsushita, A., and Kumamoto, E. (2015). Inhibition of the compound action potentials of frog sciatic nerves by aroma oil compounds having various chemical structures. *Pharmacol. Res. Perspect.* 3:e00127. doi: 10.1002/prp2.127

Onodera, Y., Suzuki, A., Wu, C. Y., Washida, H., and Takaiwa, F. (2001). A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *J. Biol. Chem.* 276, 14139–14152. doi: 10.1074/jbc.M007405200

Perestrelo, R., Silva, C. L., Rodrigues, F., Caldeira, M., and Câmara, J. S. (2016). A powerful approach to explore the potential of medicinal plants as a natural source of odor and antioxidant compounds. *J. Food Sci. Technol.* 53, 132–144. doi: 10.1007/s13197-015-2022-x

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Petrik, D. L., Karlen, S. D., Cass, C. L., Padmakshan, D., Lu, F., Liu, S., et al. (2014). p-Coumaroyl-CoA: monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in *Brachypodium distachyon*. *Plant J.* 77, 713–726. doi: 10.1111/tpj.12420

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Raina, A. P., and Negi, K. S. (2015). Essential oil composition of *Valeriana Jatamansi* Jones from himalayan regions of India. *Indian J. Pharm. Sci.* 77, 218–222. doi: 10.4103/0250-474x.156614

Reyes-Sánchez, F. J., Páez-Lerma, J. B., Rojas-Contreras, J. A., López-Miranda, J., Soto-Cruz, Ó, and Reinhart-Kirchmayr, M. (2019). Study of the enzymatic capacity of *Kluyveromyces marxianus* for the synthesis of esters. *J. Mol. Microbiol. Biotechnol.* 29, 1–9. doi: 10.1159/000507551

Sajjadi, S.-E., Jamali, M., Shokoohinia, Y., Abdi, G., Shahbazi, B., and Fattahi, A. (2015). Antiproliferative evaluation of terpenoids and terpenoid coumarins from *Ferulago macrocarpa* (Fenzl) Boiss. fruits. *Pharmacognosy Res.* 7, 322–328. doi: 10.4103/0974-8490.158437

Sarker, L. S., and Mahmoud, S. S. (2015). Cloning and functional characterization of two monoterpene acetyltransferases from glandular trichomes of *L. x intermedia*. *Planta* 242, 709–719. doi: 10.1007/s00425-015-2325-1

Shalit, M., Guterman, I., Volpin, H., Bar, E., Tamari, T., Menda, N., et al. (2003). Volatile ester formation in roses. Identification of an acetyl-coenzyme A. Geraniol/Citronellol acetyltransferase in developing rose petals. *Plant Physiol.* 131, 1868–1876. doi: 10.1104/pp.102.018572

Souleyre, E. J. F., Chagné, D., Chen, X., Tomes, S., Turner, R. M., Wang, M. Y., et al. (2014). The AAT1 locus is critical for the biosynthesis of esters contributing to 'ripe apple' flavour in 'Royal Gala' and 'Granny Smith' apples. *Plant J.* 78, 903–915. doi: 10.1111/tpj.12518

Souleyre, E. J. F., Greenwood, D. R., Friel, E. N., Karunairetnam, S., and Newcomb, R. D. (2005). An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor: characterization of apple alcohol acyl transferase. *FEBS J.* 272, 3132–3144. doi: 10.1111/j.1742-4658.2005.04732.x

St-Pierre, B., and De Luca, V. (2000). "Chapter Nine Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism," in *Recent Advance Phytochem Evolution of Metabolic Pathways*, eds J. T. Romeo, R. Ibrahim, L. Varin, and V. De Luca (Netherlands: Elsevier), 285–315. doi: 10.1016/S0079-9920(00)80010-6

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116

Wang, H., Ma, D., Yang, J., Deng, K., Li, M., Ji, X., et al. (2018). An integrative volatile terpenoid profiling and transcriptomics analysis for gene mining and functional characterization of AvBPPS and AvPS involved in the monoterpenoid biosynthesis in *Amomum villosum*. *Front. Plant Sci.* 9:846. doi: 10.3389/fpls.2018.00846

Wang, X., Ma, A., Shi, W., Geng, M., Zhong, X., and Zhao, Y. (2011). Quercetin and bornyl acetate regulate T-lymphocyte subsets and INF- γ /IL-4 ratio I in pregnant mice. *Evid. Based Complement. Alternat. Med.* 2011:745262. doi: 10.1155/2011/745262

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wise, M. L., Savage, T. J., Katahira, E., and Croteau, R. (1998). Monoterpene synthases from common sage (*Salvia officinalis*). *J. Biol. Chem.* 273, 14891–14899. doi: 10.1074/jbc.273.24.14891

Yan, X., Qin, X., Li, W., Liang, D., Qiao, J., and Li, Y. (2020). Functional characterization and catalytic activity improvement of BAHD acyltransferase from *Celastrus angulatus* Maxim. *Planta* 252:6. doi: 10.1007/s00425-020-03413-2

Zhang, B., Shen, J., Wei, W., Xi, W., Xu, C.-J., Ferguson, I., et al. (2010). Expression of genes associated with aroma formation derived from the fatty acid pathway during peach fruit ripening. *J. Agric. Food Chem.* 58, 6157–6165. doi: 10.1021/jf100172e

Zhang, T., Lu, S. H., Bi, Q., Liang, L., Wang, Y. F., Yang, X. X., et al. (2017). Volatile oil from Amomi Fructus attenuates 5-fluorouracil-induced intestinal mucositis. *Front. Pharmacol.* 8:786. doi: 10.3389/fphar.2017.00786

Zhao, H., Li, M., Zhao, Y., Lin, X., Liang, H., Wei, J., et al. (2021b). A comparison of two monoterpenoid synthases reveals molecular mechanisms associated with the difference of bioactive monoterpenoids between *Amomum villosum*

and *Amomum longiligulare*. *Front. Plant Sci.* 12:695551. doi: 10.3389/fpls.2021.695551

Zhao, H., Li, M., Lin, X., Zhao Yuan, Y., Liang, H., Ye, Z., et al. (2021a). Cloning and bioinformatics analysis of promoters of three terpene synthase genes and their terpenoid regulation in *Amomum villosum*. *Chinese Tradit. Herb Drugs* 52, 1117–1127.

Zhu, J., Lin, J.-L., Palomec, L., and Wheeldon, I. (2015). Microbial host selection affects intracellular localization and activity of alcohol-O-acetyltransferase. *Microb. Cell Fact.* 14:35. doi: 10.1186/s12934-015-0221-9

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

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

frontiers | Frontiers in Plant Science

REVIEW
published: 29 April 2022
doi: 10.3389/fpls.2022.881856

Research Progress and Trends in Metabolomics of Fruit Trees

Jing Li[1,2,3,4†], Guohua Yan[2,3,4†], Xuwei Duan[2,3,4†], Kaichun Zhang[2,3,4], Xiaoming Zhang[2,3,4], Yu Zhou[2,3,4], Chuanbao Wu[2,3,4], Xin Zhang[2,3,4], Shengnan Tan[1,5], Xin Hua[1*] and Jing Wang[2,3,4*]*

[1] Key Laboratory of Saline-Alkali Vegetation Ecology Restoration, Ministry of Education, Northeast Forestry University, Harbin, China, [2] Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [3] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Ministry of Agriculture and Rural Affairs, Beijing, China, [4] Beijing Engineering Research Center for Deciduous Fruit Trees, Beijing, China, [5] Analysis and Test Center, Northeast Forestry University, Harbin, China

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Ge Bai,
Yunnan Academy of Tobacco Agricultural Sciences, China
Aijia Ji,
Guangzhou University of Chinese Medicine, China

***Correspondence:**
Shengnan Tan
shengnantan@126.com
Xin Hua
huaxin@nefu.edu.cn
Jing Wang
sduwj@126.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 23 February 2022
Accepted: 01 April 2022
Published: 29 April 2022

Citation:
Li J, Yan G, Duan X, Zhang K, Zhang X, Zhou Y, Wu C, Zhang X, Tan S, Hua X and Wang J (2022) Research Progress and Trends in Metabolomics of Fruit Trees. Front. Plant Sci. 13:881856. doi: 10.3389/fpls.2022.881856

Metabolomics is an indispensable part of modern systems biotechnology, applied in the diseases' diagnosis, pharmacological mechanism, and quality monitoring of crops, vegetables, fruits, etc. Metabolomics of fruit trees has developed rapidly in recent years, and many important research results have been achieved in combination with transcriptomics, genomics, proteomics, quantitative trait locus (QTL), and genome-wide association study (GWAS). These research results mainly focus on the mechanism of fruit quality formation, metabolite markers of special quality or physiological period, the mechanism of fruit tree's response to biotic/abiotic stress and environment, and the genetics mechanism of fruit trait. According to different experimental purposes, different metabolomic strategies could be selected, such as targeted metabolomics, non-targeted metabolomics, pseudo-targeted metabolomics, and widely targeted metabolomics. This article presents metabolomics strategies, key techniques in metabolomics, main applications in fruit trees, and prospects for the future. With the improvement of instruments, analysis platforms, and metabolite databases and decrease in the cost of the experiment, metabolomics will prompt the fruit tree research to achieve more breakthrough results.

Keywords: fruit tree, metabolomics, quality, resistance, mQTL, mGWAS

INTRODUCTION

To comprehensively and quantitatively study key metabolites in a biological system, "metabolomics," which refers to the qualitative and quantitative analysis of all metabolites with molecular weight < 1,000 Da in organisms or cells, was proposed by Nicholson in 1999 (Nicholson et al., 1999).

Unlike DNA, RNA, and proteins, which consist of several fixed and known structural units, metabolites are reactive, are structurally diverse, and have a wide range of concentrations. Metabolites do not follow a fixed structural module and have different physical properties (Sumner et al., 2003). The Human Metabolome Database (HMDB) has identified more than 40,000 human metabolites. The environment in which plants grow and reproduce is far more complex than those of humans and animals and often disturbed by a variety of biotic and abiotic factors. Considering their own factors, plants can only adjust the whole process of their life cycle in a more flexible and

diverse way, with the total number of metabolites produced by plants ranging from 200,000 to 1 million (Dixon and Strack, 2003).

Fruits provide irreplaceable nutrition to human and have an increasingly important impact on human health. Compared with animals and main crops, metabolomics of fruit trees starts later. But fruits are beneficial to human health due to their diverse secondary metabolites, including organic acids, flavonoids, anthocyanins, terpenes, etc., which are very suitable for the study by metabolomics technology. Thus, a rapid increase of research articles appears from 2017 to 2021 (**Figure 1**) according to the Web of Science[1]. Many important research results have been achieved in metabolomics of a fruit tree in recent years. In this article, main achievements in metabolomics of fruit tree will be introduced.

METABOLOMICS STRATEGIES AND KEY TECHNIQUES

Metabolomics Strategies

Metabolites in plants can generally be divided into primary and secondary metabolites. Primary metabolites are the basic material conditions necessary for the growth and development of organisms, while secondary metabolites are metabolites produced only at a certain stage or period of growth of an organism (Koch, 2004; Carreno-Quintero et al., 2013). Depending on the purpose of the experiment, researchers can choose individual metabolomics *strategies*, including targeted metabolomics, non-targeted metabolomics, pseudo-targeted metabolomics, and widely targeted metabolomics.

Targeted metabolomics assays are performed on known compounds with targeted extraction methods and high-purity standards to quantify the target metabolism. Generally, targeted metabolomics focuses on fewer compounds and can be searched directly in classified databases, for example, databases for glycans and lipids such as LipidMaps and LipidBank (Cajka and Fiehn, 2016). Non-targeted metabolomics assays are performed on unknown compounds. It covers as many compounds of different properties and classes as possible in the extraction and detection, with a relatively broad coverage of substances. However, both the qualitative and quantitative properties of compounds are compromised by the lack of standards. Generally, non-targeted metabolomics studies are used to screen for differential metabolites in different treatments, and targeted metabolomics studies are used for the next more precise study.

To overcome the disadvantages of non-targeted metabolomics, widely targeted metabolomics, also known as second-generation targeted metabolomics, allows for the simultaneous detection of thousands of substances by the establishment of a database of thousands of metabolite specimens (Wei et al., 2013). Pseudo-targeted metabolomics is a quantitative ion selection algorithm based on untargeted metabolomics. It is used to perform quantitative ion selection for all detected metabolites. The ion-pair information of metabolites is obtained by high-resolution mass spectrometry (HR-MS), and the abundance of metabolites is measured by targeted multiple reaction monitoring (MRM). The method has high coverage, good linearity, and reproducibility and does not require standards to limit the metabolites' detection. Both known and unknown metabolites in the sample can be measured (Zheng et al., 2020).

Spatial metabolomics integrates traditional metabolomics technology and MS imaging technology, which can not only identify what substances are contained in a sample and their contents as traditional metabolomics but also detect the spatial distribution information of compounds in a single experiment to achieve qualitative and quantitative positioning at the same time (Li et al., 2018).

The comparison of advantages and disadvantages of different metabolomic strategies is shown in **Table 1**.

Key Techniques in Metabolomics

The exploration of the diversity of fruit tree metabolism and potential molecular mechanisms by which fruit tree cells control their own chemical composition depends on the advances of analytical methods.

Sample Processing

Samples should be collected representatively, with priority given to normally developing plants. Furthermore, 3–6 individuals should be selected as a source of biological replicates to avoid differences due to excessive individual differences. Ice boxes or ice packs should be used to collect samples in the field to reduce the degradation of metabolites by enzymes that are still active after sample collection. At the same time, samples should be collected with certain accuracy, and the samples of the experimental and control groups should be consistent in terms of time (Urbanczyk-Wochniak et al., 2005), site, and processing conditions, as far as the experiment allows. If the kernels are not studied, it should be removed with a scalpel in a low-temperature metal bath on an RNAase reagent-treated bench. For the study of the peel, the pulp should be taken as less as possible (Montefiori et al., 2009; de Godoy et al., 2013; Rudell et al., 2017). Freezing of the whole fruit should be avoided if possible during the sampling process, which will make the subsequent freeze-drying and grinding work difficult, and once the whole fruit is freeze-thawed, the whole experimental results will be affected. Then, the sample will be stored in a suitable volume of a centrifuge tube, marked and quickly put into liquid nitrogen for precooling, and then put into −80°C refrigerator for storage. Note that the samples should not be stored in tin foil, self-sealing bags, Ziploc bags, plastic bags, etc.; such packaging is easy to break under low temperature or in the process of transportation and easy to make the markings blurred, resulting in sample confusion.

The metabolites are usually extracted separately with water or organic solvents (e.g., methanol, chloroform, etc.) to obtain aqueous and organic solvent extracts, thus separating the non-polar lipophilic phase from the polar phase. To enrich the concentration of metabolites during the study, the process of vacuum concentration and spin-drying of the extracts before redissolution is often added. In the process of sample analysis

[1]https://www.webofscience.com/

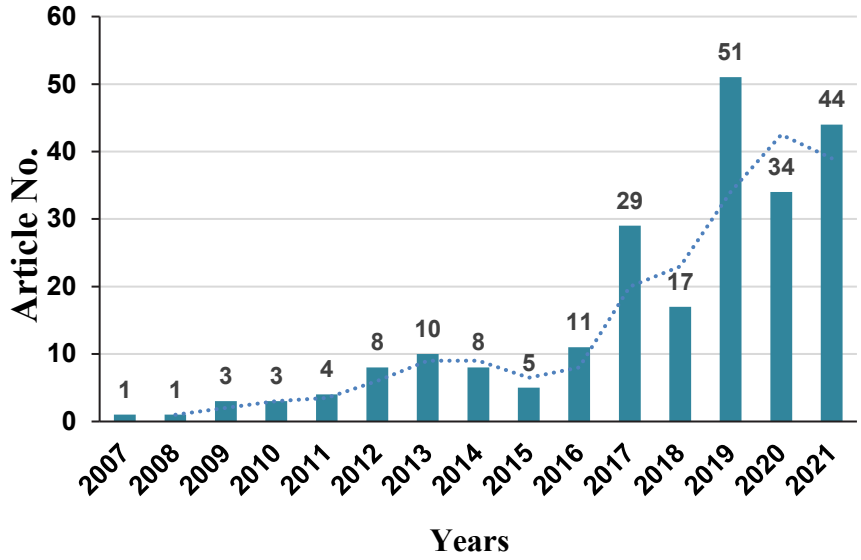


FIGURE 1 | Published research articles about metabolomics of fruit tree (data from Web of Science).

TABLE 1 | Different strategies of metabolomics.

Metabolomic	Advantages	Disadvantages	Application
Targeted metabolomics	Qualitative and precise quantitative	Limited to detection of compounds of known molecular weight, low throughput	Qualitative and quantitative detection of a small number of known compounds
Non-targeted metabolomics	High throughput, wide detection range	Qualitative difficulty, quantitative accuracy is low	High throughput detection of known and unknown compounds
Widely-targeted metabolomics	High throughput, accurate qualitative and quantitative	Limited to library-building compounds	Qualitative and quantitative detection of a large number of known compounds
Pseudo-targeted metabolomics	Repeatability, high throughput, wide detection range	/	High throughput detection of known and unknown compounds
Spatial metabolomics	Qualitative, quantitative, positioning, high throughput	/	Qualitative, quantitative, and localized analysis of a large number of known or unknown compounds directly

method establishment, the optimal method can be figured out by trying different combination of extraction solvents, extraction temperatures, ratios of sample and solvent, redissolved solvent, mobile phase gradients, and mobile phases and by the controlled variable method (Kim and Verpoorte, 2010). The detection of non-targeted metabolites should include as many compound peaks as possible, and the detection of targeted metabolites should ensure that all concerned metabolites can be detected. The sample sequence requires the insertion of quality control samples at intervals of 10 or 20 during the detection by liquid chromatography-MS (LC-MS), which generally includes all compounds of the sample being tested or a mixture of a representative batch of samples. In addition, compounds not contained in the fruit trees are added as internal standards during the extraction of the samples to calibrate the loss of sample compounds during the extraction and detection process and thus to correct the peak areas of other compounds. *In vivo* direct-immersion solid-phase microextraction followed by gas chromatography-time-of-flight mass spectrometry method (SPME GC × GC-TOF-MS) can detect volatile compounds directly without extracting the compounds (Risticevic et al., 2016, 2020).

The flowchart of metabolomics sample preparation, analysis, and data processing is shown in **Figure 2**.

Analysis

Chromatography is used for the purpose of separating metabolites according to the different affinities of compounds and columns and is generally divided into LC and gas chromatography (GC). LC is used to detect non-volatile compounds, polar compounds, thermally unstable compounds, and high-molecular-weight compounds (including proteins, peptides, and polymers). The mobile phase generally uses methanol, acetonitrile, water, or isopropanol; the added acid is generally acetic acid or formic acid, inorganic acid cannot be used (inorganic acid may corrode the pipeline); the added base is generally ammonium hydroxide or ammonia, generally does not use metal bases (will corrode the pipeline). Trifluoroacetic acid (TFA) helps to improve the separation of the liquid phase but will produce ion suppression in the positive and negative

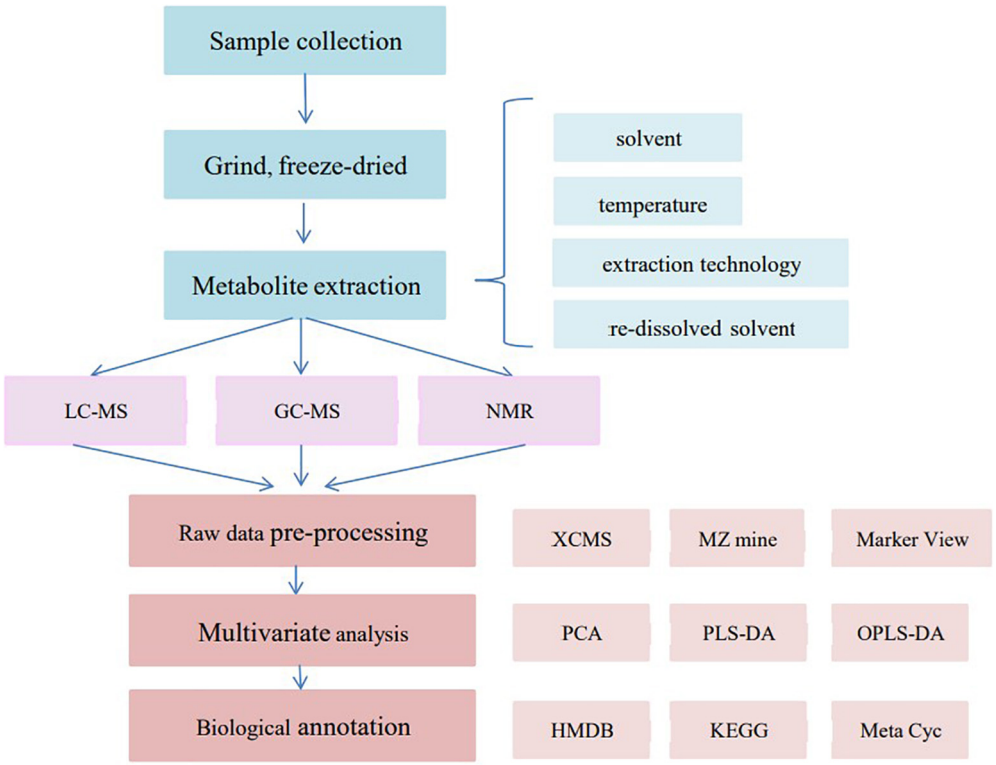


FIGURE 2 | Metabolomics sample preparation, analysis, and data processing flowchart.

ion modes of MS. Triethylamine/trimethylamine (TEA/TMA) contributes to the formation of negative ions. GC can be used for the detection of volatile compounds, the mobile phase is an inert gas, and samples with a certain vapor pressure and good stability at column temperature can be detected. Substances with low volatility and easily decomposed by heat can be transformed by derivatization into derivatives with high volatility and good thermal stability for detection. *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (MSTFA) and methoxamine hydrochloride are often used as derivatization reagents, as preliminary studies have shown that these compounds are the most suitable for the derivatization of plant metabolites (Fiehn et al., 2000).

Mass spectrometry separates some of the ions based on the mass to charge ratio. The ions pass through the analyzer and are separated according to different mass to charge ratios. Ions with the same mass to charge ratio are clustered together to form a mass spectrogram. Commonly used mass spectrometers are triple quadrupole mass analyzer (Yost and Enke, 1978), time-of-flight mass analyzer, ion trap mass analyzer, and electrostatic field orbital trap. The triple quadrupole mass analyzer consists of four straight rod electrodes suspended equidistantly parallel to the axis. Under the action of a certain DC/VC, ions with m/z meeting certain requirements can pass through the quadrupole to the detector, while other ions are filtered out. The advantages of the triple quadrupole are light weight, small size, and low cost. The time-of-flight mass analyzer uses pulses to extract ions from an ion source. The ions are accelerated by an accelerating voltage, have the same kinetic energy, and enter the drift tube. Ions with a small mass to charge ratio are the fastest and reach the detector first. Larger ions reach the detector last. The advantages of TOF MS are a wide range of mass to charge ratio of the detected ions, high sensitivity, and suitability for secondary tandem MS. A fast scanning speed is suitable for studying very fast processes. The principle of an ion trap mass analyzer is similar to that of quadrupole mass analyzer. When the HF voltage amplitude and HF voltage frequency are fixed to a certain value, only ions with a certain mass to charge ratio can be stabilized on a certain track in the trap. By rotating and changing the electrode voltage, different m/z ions fly out of the trap and reach the detector. The advantage of ion trap mass spectrometer is that a single ion trap can realize multistage "time" tandem MS, simple structure, low price, high cost performance, high sensitivity, and large mass range. The electrostatic field ion orbitrap is where analytes are ionized in an ion source and then sequentially enter a quadrupole, a C-trap, and an Orbitrap. The electrostatic field ion orbitrap mass spectrometer can perform high precision mass scans, and it is more stable than other mass analyzers.

Mass spectrometry characterizes compounds by different characteristic ions of different compounds. Chromatographic methods enhance the coverage of metabolites and improve the quantitative accuracy of MS. The most commonly used separation and analysis methods are LC-MS (Vos et al., 2007), GC-MS (Lisec et al., 2015). For LC-MS and GC-MS methods,

internal standards are added before sample processing to correct errors caused by sample pretreatment and instrument response.

Data Processing

Data processing in metabolomics generally consists of two major steps, namely, data preprocessing and statistical analysis (Ma and Qi, 2021). Data preprocessing includes steps such as removing system noise signals, removing disturbing signals caused by system instability, and removing operational errors to provide a more reliable dataset for the next statistical analysis. Prior to formal data processing, tools such as XCMS, MZmine, and MarkerView are used to process the original peak area and other data for non-targeted metabolomics data analysis. The first step of the statistical analysis is unsupervised multivariate statistical analysis, usually using the principal component analysis (PCA). The second step may be the univariate statistical analysis to screen for variables with statistically significant differences in different groups. The third step is a supervised multivariate statistical analysis such as partial least squares discriminant analysis (PLS-DA) to select the variables that contribute more to the classification of the sample, i.e., screen for markers.

The purpose of the PCA analysis (Shaffer, 2002) is to find specific values that produce differences in a large amount of sample data. PLS-DA analysis will classify samples into groups, which is good for finding the similarities and differences between sample groups; orthogonal partial least squares discriminant analysis (OPLS-DA) can also distinguish the group differences between samples, but this analysis can better focus on some positive correlation data. The results of data analysis are also subject to the *t*-test and variable importance in projection (VIP) ranking values to screen for differential metabolites. It is generally accepted that variables meeting both $P < 0.05$ and VIP > 1.0 are differential metabolites. Metabolic pathways can be resolved using databases such as HMDB, KEGG, Reactome, BioCyc, and MetaCyc.

APPLICATION OF METABOLOMICS IN FRUIT TREES

Fruit tree metabolomics has developed rapidly in recent years, and many important research results have been achieved in combination with transcriptomics, genomics, proteomics, quantitative trait locus (QTL), and genome-wide association study (GWAS) (**Tables 2**, **3**). These research results mainly focus on the mechanism of fruit quality formation, metabolite markers of special quality or physiological period, key genes and metabolites of fruit tree resistance to biotic/abiotic stress, the influence of environment on the fruit tree, and fruit tree population genetic basis by combining with QTL and GWAS to narrow down QTL regions, predict candidate genes, construct the regulatory network, etc. (**Figure 3**).

Fruit Quality Formation

Fruit quality, such as fruit color, sweet/acid ratio, size, firmness, aroma, and special nutrients, is very important trait for consumers and is the main factor determining market competitiveness. Research on the mechanisms of fruit quality formation can reveal new information about fruit physiology, provide new biomarkers for variety or producing area identification, and accelerate the improvement of the fruit quality.

From a flower to a mature fruit, undergoing complex changes, metabolomics reveals the composition and changes of the compounds that form the fruit quality. A widely targeted metabolomic analysis in pineapple (*Ananas comosus* L.) young fruits (YF), mature fruits (MF), and fully mature fruits (FMF) identified 466 metabolites. In comparison, lysophospholipids (LPLs) were found to be the highest in YF, and their content decreased gradually with fruit development and remained almost constant in FMF (Hong et al., 2021). During the ripening of ginseng berry berries, the content of malate gradually increases, and most amino acids and organic acids gradually decrease. Compounds that could distinguish different cultivation ages and maturity stages of ginseng were obtained (Mamat et al., 2020). Content variation in amino acid and isoprenoid in *ASR1*-antisense transgenic tomato (*Solanum lycopersicum*) fruits and normal tomato was examined by targeted metabolomics. Amino acids and isoprenoids were found to be significantly different in these two species. The combined analysis with transcriptome data demonstrates that the Asr1 [i.e., abscisic acid (ABA), stress, and ripening 1] transcription factor is involved in the cascade regulation of red tomato fruit maturation (Dominguez et al., 2021). Special metabolites also accumulate in mangosteen (Mamat et al., 2020), ginseng berry (Park et al., 2019), starfruit (Ramadan et al., 2020), avocado (Hurtado-Fernández et al., 2015), strawberry (Zhang et al., 2011; Wang et al., 2021e), apple (Rudell and Mattheis, 2009; Risticevic et al., 2016, 2020; Jing and Malladi, 2020; Xu et al., 2020), grape berry (Dai et al., 2013), mulberry (Lee et al., 2016), cupressaceae (Falasca et al., 2014), olea (Olmo-García et al., 2018), walnuts (Wang et al., 2021b), and African baobab (Baky et al., 2021) during fruit maturing.

Fruit color not only affects appetite but also affects nutrients. By targeting carotenoids in white fleshed apricots (*Prunus armeniaca* L.) "Kuchebaixing" and orange fleshed apricots "Shushangganxing" at different developmental stages (i.e., S1–S4), 14 carotenoids and 27 carotenoid lipids were identified in apricot flesh. The comparison revealed significant differences in the carotenoid content between these two different apricot cultivars during S3 and S4 development and finally identified β-carotene and (E/Z)-octahydro lycopene as the key metabolites responsible for the differences in total carotenoid content in the different cultivars. Combination of metabolome and transcriptome reveals *PSY*, *NCED1*, and *CCD4* as key genes responsible for differential accumulation of the total carotenoid content in apricot fruit (Zhou et al., 2021). The red pericarp longan (*Dimocarpus longan* Lour.) (RP) and white longan (SX) were analyzed using widely targeted metabolomics; a total of 597 substances were identified in both white and red longan. Among them, 33 compounds, mainly flavonoids, were present in red-skinned longan and 23 compounds, mainly phenolic acids, were present in white-skinned longan. Cyanidin 3-*O*-glucoside, cyanidin 3-*O*-6″-malonyl-glucoside, and cyanidin

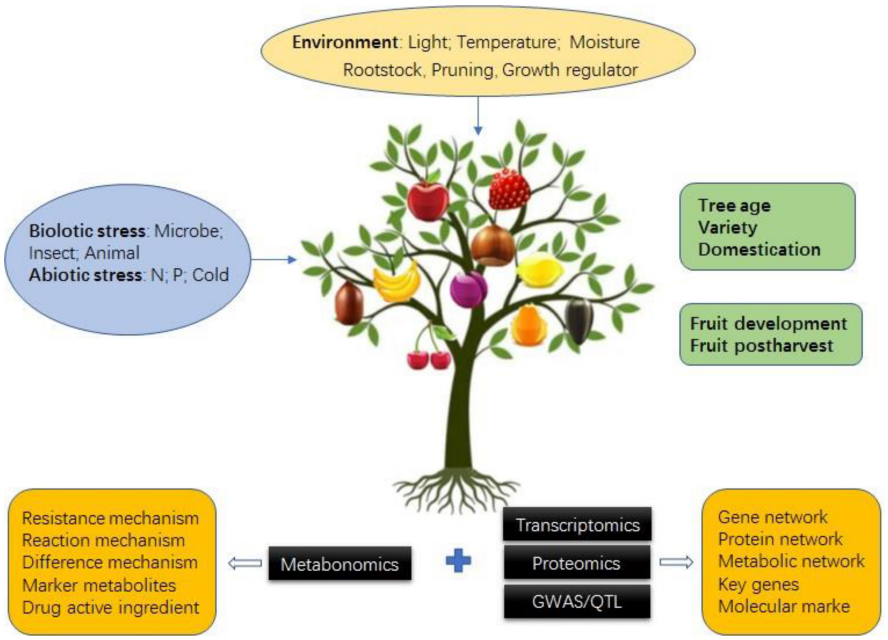


FIGURE 3 | Application of metabolomics with/without multi-omics in fruit tree.

O-syringic acid specially existed in RP longan. Delphinin 3-*O*-glucoside specially existed in SX longan. Accordingly, genes in the biosynthetic pathway of anthocyanin, such as *F3H*, *F3'H, UFGT, GST, MYB, bHLH, NAC, and MADS*, were detected significantly upregulated in RP longan using the transcriptome analysis (Yi et al., 2021). The analysis of six Tuscan sweet cherries (*Prunus avium* L.) endemic to Italy by non-targeted metabolomics successfully identified 15 metabolites in the positive ion mode and 14 metabolites in the anion mode. Most of these metabolites belong to flavonoids, including a highly potent antioxidant, cinchonain, while the highest flavonoid content was found in two varieties, namely, Crognola and Morellona (Berni et al., 2021). Anthocyanins also have important effects on fruit coloration in jujube (Zhang et al., 2020), fig (Wang et al., 2017), and *Cerasus humilis* (Ji et al., 2020).

The taste and aroma of the fruit are important intrinsic qualities of the fruit. A variation of metabolites in kiwifruit (*Actinidia chinensis*) flavor formation was revealed by widely targeted metabolomics to detect the content of sugar, organic matter, and volatile compounds in the flesh of kiwifruit during six developmental and six postharvest periods. A total of 34 flavor-related metabolites were identified. Further studies showed that the accumulation of metabolites in kiwifruit differed between the immature and mature stages. In the immature stage, the accumulated metabolites will tend to accumulate esters and sugars, and in the mature stage, they will tend to accumulate aldehydes, alcohols, ketones, and organic acids. In addition, a regulatory network that can regulate the production and accumulation of soluble sugars, organic acids, and volatiles was constructed by combining metabolomic and transcriptomic data analysis. This regulatory network can identify key structural and regulatory genes associated with flavor metabolism, key transcription factors that regulate the metabolism of soluble sugars and esters, and moreover, reveal their role in regulating the transcription of key structural genes involved in these metabolic pathways (Wang et al., 2022). Lipid compounds and phenylpropane derivatives were identified as the main components of the dates' aroma by detecting 13 volatile compounds of dates (*Phoenix dactylifera* L.). In addition, 2,3-butanediol, hexaldehyde, hexol, and cinnamaldehyde can be used to distinguish between the different dates varieties (Khalil et al., 2017). Methyl hexatate, which was not detected in two other producing regions, is a volatile unique to *Averrhoa carambola* L. in Egypt (Ramadan et al., 2020). Special metabolites were differentially accumulated in different cultivar and wild persimmon varieties (Ramírez-Briones et al., 2019; Ryu et al., 2019), *Citrus* L. (Wang et al., 2016; Fayek et al., 2019; Villa-Ruano et al., 2019), and peach varieties (Karagiannis et al., 2021).

After the fruit is picked, the quality may change depending on the storage condition before entering the consumer's table (Karagiannis, 2018; Wu et al., 2019, 2021; Lwin and Lee, 2020; Fuentealba, 2021). Comparing the changes of volatile compounds before and after 7 days of storage of 3 hairy peaches (*Prunus persica* L.) and 3 nectarines, it was found that the distribution of volatile compounds in fruits tended to be similar after storage. Using random forest analysis, it was possible to identify 15 volatile organic compounds, namely, 6 terpenes, 6 esters, 2 lactones, and 1 aromatic VOC, which differed between varieties before and after storage. Of which, 4 terpenes and 4 esters were positively correlated with nectarine. Another 16 VOCs highly correlated with 11 key VOC pathway genes (Muto et al., 2020).

TABLE 2 | Application of metabolomics and multi-omics in fruit tree.

Species	Tissue	Muti-omics[a]	Strategies of metabolomics	Technology of metabolomics	Metabolic	References
Apricot (*Prunus armeniaca* L.)	Pulp	T + M	Targeted	UHPLC-APCI-MS/MS	Carotenoid	Zhou et al., 2021
Pineapple (*Ananas comosus* L.)	Pulp	T + M	Widely targeted	UPLC-MS/MS GC-MS	Fatty acid	Hong et al., 2021
Kiwi fruit (*Actinidia chinensis*)	Pulp	T + M	Widely targeted	GC-MS HS-SPME-GC-MS	Sugars, organic acids, volatiles	Wang et al., 2022
Longan (*Dimocarpus longan* Lour.)	Pericarp	T + M	Widely targeted	HPLC UPLC-MS/MS	Anthocyanins	Yi et al., 2021
Tomato (*Solanum lycopersicum*)	Pulp	T + M	Targeted	GC-MS	Amino acid and isoprenoid amino acid	Dominguez et al., 2021
Sweet Cherry (*Prunus avium* L.)	Pulp	T + M	Untargeted	UHPLC-MS/MS	Flavonoids	Berni et al., 2021
Peach (*Prunus persica* L.Batsch)	Pulp	T + M	Untargeted	GC × GC-ToF-MS	Volatile organic compounds	Muto et al., 2020
Pear (*Pyrus bretschneideri* × *Pyrus pyrifolia*)	Pulp	T + M	Untargeted	LC-MS	Gibberellin GA3	Wang et al., 2021f
Apple (*Malus* × *domestica* Borkh)	Pulp	T + M + P	Widely Targeted	UPLC-QQQ-MS	Carbohydrates, flavonoids, amino acids	Wang et al., 2021a
Strawberry (*Fragaria* × *ananassa*)	Pulp	T + M	Untargeted	GC-TOF-MS	Organic acids and sugars	Vallarino et al., 2019
Apple (*Malus* × *domestica* Borkh)	Pulp and pericarp	T + M	Untargeted	LC-Q-TOF-MS	Phenol	Khan et al., 2012b
Jujube (*Zizyphus jujuba* Mill.)	Pulp	T + M	Widely targeted	LC-MS/MS	Triterpenoids, alkaloids, and flavonoids	Zhang et al., 2021
Jujube (*Zizyphus jujuba* Mill.)	Pericarp	T + M	Targeted		Flavonoid and anthocyanins	Zhang et al., 2020
Citrus (*Citrus reticulata*)	Pericarp	T + M	Untargeted	GC-MS	Sugars, organic acids, amino acid	Lu et al., 2017
Fig (*Ficus carica* L.)	Pericarp	T + M	Untargeted	LC-MS	Flavonoid and anthocyanins	Wang et al., 2017
Apple (*Malus* × *domestica* Borkh)	Pulp and pericarp	T + M	Untargeted	UPLC-TOF-MS	Flavonoid	Wang et al., 2021c
Apple (*Malus* × *domestica* Borkh)	Root and leaf	T + M	Untargeted	LC-ESI-MS/MS	Flavonoid	Sun et al., 2021
Citrus (*Citrus reticulata*)	Pericarp	T + M	Untargeted	HS-SPME-GC–MS GC-MS	Sugars, organic acids, volatiles	Tang et al., 2018
Cerasus humilis (Bge.) Sok	Pericarp	T + M	Untargeted	UPLC-MS/MS	Anthocyanins	Ji et al., 2020
Apple (*Malus* × *domestica* Borkh)	Pulp and pericarp	T + M	Widely targeted	LC-MS/MS GC-MS	Sugars, organic acids, amino acid	Xu et al., 2020
Citrus (*Citrus reticulata*)	Root	T + M + P	Untargeted	NMR	Polar component	Padhi et al., 2019
Walnut (*Juglans regia* L.)	Pulp and pericarp	T + M + P	Untargeted	LC-MS	Raffinose	Wang et al., 2021b

[a] *T is transcriptomics; M is metabolomics; P is proteomics.*

In ethephon-treated sweet cherries, fruit ripening and ethylene-related metabolites, such as malate and monosaccharides, have changed (Smith et al., 2011).

The development of metabolomics has made it possible to identify the active components of special fruit trees with medicinal value. Fifteen compounds were identified from the *Kigelia africana* fruit using UHPLC-TOF-MS and GC-MS. In which, physostigmine was demonstrated with an excellent anticancer activity (Fagbohun et al., 2021). In addition, nine antiaging compounds in *Nephelium lappaceum* (rambutan) seeds were identified (Lee et al., 2020). Genipap (*Genipa americana* L.) is a known fruit with medicinal value. Biomarkers after the administration of genipap in human urine is identified by LC-MS (Dickson et al., 2020).

Biotic/Abiotic Stress on Fruit Tree

Biotic stress and abiotic stress are great challenges to the fruit tree with impact on production, quality, and survival. The biotic/abiotic stress causes physiological changes, which eventually causes corresponding metabolites changes. By metabolomics analysis and in comparison with normal fruit tree, differential metabolites will be identified as new biomarkers of the biotic/abiotic stress and serve for disease

TABLE 3 | Application of metabolomics in fruit tree.

Species	Tissue	Strategies of metabolomics	Technology of metabolomics	Application	References
Pear (*Pyrus bretschneideri* × *Pyrus pyrifolia*)	Flower buds	Untargeted	GC-TOF-MS	Environment	Horikoshi et al., 2018
Blueberry (*Vaccinium* spp.)	Flower buds	Untargeted	GC-TOFMS	Abiotic stress	Wang et al., 2021d
Peach (*Prunus persica* L.Batsch)	Flower buds	Widely targeted	LC-MS	Biotic stress	Zhang et al., 2019
Pear (*Pyrus bretschneideri* × *Pyrus pyrifolia*)	Stem	Untargeted	LC-MS-MS	Environment	Cui et al., 2021
Banana (*Musa nana* Lour.)	Pulp	Untargeted	GC–MS	Environment	Nascimento et al., 2019
Juniperus communis L.	Pulp	Untargeted	^{1}H NMR	Biotic stress	Falasca et al., 2014
African baobab fruit (*Adansonia digitata* L.)	Pulp	Untargeted	UHPLC-HRMS/MS + HS-SPME/GC-MS GC-MS	Quality	Baky et al., 2021
Ginseng (*Panax ginseng Meyer*)	Pulp	Untargeted	GC-MS	Quality	Park et al., 2019
Olive (*Olea europaea* L.)	Pulp	Untargeted	LC-MS + GC-MS	Quality	Olmo-García et al., 2018
Dates (*Phoenix dactylifera* L.)	Pulp and pericarp	Untargeted	SPME GC–MS	Quality	Khalil et al., 2017
Apple (*Malus* × *domestica* Borkh.)	Pulp and pericarp	Untargeted	DI-SPME-GCxGC-ToFMS	Quality	Risticevic et al., 2016, 2020
Strawberry (*Fragaria* × *ananassa*)	Pulp and pericarp	Untargeted	(MALDI-TOF IMS)	Quality	Wang et al., 2021e
Citrus (*Citrus reticulata*)	Pulp and pericarp	Untargeted	LC-MS	Quality	Wang et al., 2016
Citrus (*Citrus reticulata*)	Pericarp	Untargeted	UPLC-QQQ-MS	Biotic stress	Wang et al., 2020
Citrus (*Citrus reticulata*)	Leaf	Targeted	LC-MS/MS	Biotic stress	Suh et al., 2021
Grapevine (*Vitis vinifera* L.)	Pulp and pericarp	Targeted	UHPLC	Genetic mechanism	Arrizabalaga-Arriazu et al., 2021
Grapevine (*Vitis vinifera* L.)	Stem	Widely targeted	LC-HRMS	Genetic mechanism	Teh et al., 2019
Mulberry (*Morus alba* L.)	Leaf and phloem sap	Untargeted	GC-MS	Biotic stress	Gai et al., 2014
Citrus (*Citrus reticulata*)	Pericarp	Untargeted	HPLC-PDA-QTOF-MS	Biotic stress	Ballester et al., 2013
Apple (*Malus* × *domestica* Borkh)	Immature fruit and shoot	Untargeted	LC-MS	Biotic stress	Klee et al., 2019
Genipap (Genipa americana L.)	Fruit	Untargeted	UHPLC-MS	Quality	Dickson et al., 2020
Rambutan (*Nephelium lappaceum* L.)	Seed	Untargeted	UPLC-qTOF-MS/MS	Quality	Lee et al., 2020
Crescentia cujete (*Bignoniaceae*)	Fruit	Untargeted	UPLC-MS/MS + NMR	Quality	Rivera-Mondragón et al., 2020
Kigelia africana (Lam.) Benth.	Fruit	Untargeted	UHPLC/GC-TOF-MS + GC–MS	Quality	Fagbohun et al., 2021
Butia spp. (Arecaceae)	Fruit	Untargeted	LC-MS	Environment	Hoffmann et al., 2017
Pear (*Pyrus bretschneideri* × *Pyrus pyrifolia*)	Pulp and pericarp	Untargeted	GC-MS + LC-MS	Environment	Rudell et al., 2017
Peach (*Prunus persica* L. Batsch)	Pulp and pericarp	Untargeted	GC-MS	Environment	Anthony et al., 2020

(Continued)

TABLE 3 | (Continued)

Species	Tissue	Strategies of metabolomics	Technology of metabolomics	Application	References
Apple *(Malus × domestica* Borkh*)*	Pulp and pericarp	Untargeted	NMR	Environment	D'Abrosca et al., 2017
Goji (*Lycium barbarum* L.)	Fruit	Untargeted	UPLC–ESI–MS/MS	Abiotic stress	Wei et al., 2020
Sweet cherry (*Prunus avium* L.)	Pulp and pericarp	Untargeted	GC–MS	Abiotic stress	Smith et al., 2011
Pomegranate (*Punica granatum* L.)	Stem and buds	Untargeted	HPLC-GC	Abiotic stress	Yazdanpanah et al., 2021
Sweet cherry (*Prunus avium* L.)	Fruit and stem	Untargeted	GC–MS	Quality	Karagiannis, 2018
Sweet cherry (*Prunus avium* L.)	Fruit	Targeted	HPLC	Quality	Fuentealba, 2021
Noni (*Morinda citrifolia* Linn.)	Pulp and pericarp	Untargeted	LC–MS	Quality	Wu et al., 2019
Noni (*Morinda citrifolia* Linn.)	Fruit	Untargeted	LC-MS	Quality	Wu et al., 2021
Mangosteen (*Garcinia mangostana* L.)	Pulp	Untargeted	GC–MS	Quality	Mamat et al., 2020
Pear *(Pyrus bretschneideri × Pyrus pyrifolia)*	Fruit	Targeted	HPLC	Quality	Lwin and Lee, 2020
Apple *(Malus × domestica* Borkh*)*	Pulp and pericarp	Targeted	GC	Quality	Rudell and Mattheis, 2009
Citrus (*Citrus reticulata*)	Pericarp	Untargeted	GC-TOF-MS	Quality	Fayek et al., 2019
Strawberry (*Fragaria × ananassa*)	Fruit	Untargeted	GC-MS + HPLC	Quality	Zhang et al., 2011
Starfruit (*Averrhoa Carambola* L.)	Fruit	Untargeted	HS-SPME-GC-MS	Quality	Ramadan et al., 2020
Mangosteen (*Garcinia mangostana* Linn.)	Fruit	Untargeted	GC-MS + LC-MS	Quality	R. Parijadi et al., 2019
Japanese persimmon (*Diospyros kaki* Thunb.)	Fruit	Untargeted	UPLC-MS	Quality	Ramírez-Briones et al., 2019
Japanese persimmon (*Diospyros kaki* Thunb.)	Fruit	Untargeted	NMR	Quality	Ryu et al., 2019
Mulberry fruit (*Morus alba Linnaeus*)	Fruit	Untargeted	GC-MS + HPLC	Quality	Lee et al., 2016
Grape berry (*Vitis vinifera* L.)	Fruit	Untargeted	LC-MS	Quality	Dai et al., 2013
Avocado (*Persea americana*)	Fruit	Untargeted	GC-APCI-TOF MS	Quality	Hurtado-Fernández et al., 2015

diagnosis. Key metabolites pathways responding to biotic/abiotic stress can reveal molecular mechanism about resistance to biotic/abiotic stress.

Several researches have been carried out to find biomarkers and pathology of biotic stress, especially in microbe infection. Huanglongbing (HLB) (*Candidatus Liberibacter* sp.) is a common disease on citrus. Infested sugar orange (*Citrus reticulata*) is divided into one normal type and three abnormal types based on fruit peel color. Non-targeted metabolomics was used, and 215 significant differences metabolites were found between abnormal and normal fruit. Two unique metabolites (i.e., *O*-caffeoyl maltotriose and prunetin) were detected only in normal pericarp. Besides, it is also revealed that "phenylpropanoid biosynthesis" pathway exhibited obvious enrichment in all comparison groups, according to the metabolic pathway enrichment analysis of differential metabolites (Wang et al., 2020). *Penicillium* spp. is a citrus fruits susceptible pathogen. The contents of sugars, organic acids, vitamin C, and D-citmonene decreased in the citrus peel after infection, and the amount of ethanol and beta-terpene alcohol increased (Tang et al., 2018). Amino acids and their derivatives were significantly higher in the citrus peel of roughing disorder (Lu et al., 2017). In mulberry and apple, pathogen infection altered the levels of abscisic acid and cytokinin (Gai et al., 2014; Klee et al., 2019). Bird foraging flower buds will affect the survival of fruit trees. Salvianolic acid A is more abundant in peach buds where birds are prone to forage and may be a potential metabolite that attracts birds to forage (Zhang et al., 2019).

The study on mechanism of resistance to disease is important for breed improvement. Scanning electron microscopy (SEM), transcriptome, and metabolomics analysis revealed that exogenous application of $NaH_2PO_42H_2O$ (P) and abscisic acid (ABA) increased the expression of "Huangguan" pear (*Pyrus bretschneideri* × *Pyrus pyrifolia*) wax-related and calcium-regulated genes, as well as increased "Huangguan" pear resistance to brown spot disease during harvest and storage (Wang et al., 2021f). Widely targeted metabolomics analysis was used to analyze leaves of HLB-tolerant and -sensitive citrus (*Citrus reticulata*) varieties, covering primary and secondary metabolic pathways including carbohydrate metabolism, nucleotide metabolism, amino acid metabolism, energy metabolism, and biosynthesis of secondary metabolites. In which, aspartate and glutamate metabolism, purine metabolism, biosynthesis of plant hormones, and catabolic pathways were upregulated in the tolerant group. A total of 50 metabolites related to HLB tolerance were identified, and these metabolites were considered as potential markers for the identification of citrus HLB (Suh et al., 2021). It was found that phenylpropanes and their derivatives play an important role in citrus fruits to resistant *Penicillium* infection (Ballester et al., 2013). In addition, fruit trees also produce special metabolites to affect the microbial populations in the rhizosphere to help resist the invasion of pathogens (Padhi et al., 2019).

Improper fertilization and abnormal temperature are common abiotic stresses on fruit trees. Nitrogen fertilizer is often over-applied in apple (*Malus* × *domestica* Borkh) cultivation in China (Approx. 600–800 kg N ha^{-1}). Although high nitrogen fertilizer application promoted apple yield and fruit weight per unit, at the same time, the carbon-to-nitrogen ratio, soluble sugars, flavonoids, and other fruit quality indicators decreased. The combined analysis of transcriptome, proteomics, and metabolomics revealed the global pattern of high nitrogen effect on apple fruits. High nitrogen significantly inhibited the accumulation of carbohydrates (sucrose, glucose, and alginose) and flavonoids (rutin, rhamnolipid-3-*O*-rutinoside, and trihydroxyisoflavone-7-*O*-galactoside) in the fruits. More carbon groups are used in the synthesis of amino acids (especially arginine) and their derivatives under high nitrogen conditions to enhance the nitrogen metabolism process (Wang et al., 2021a). Under low phosphorus stress, the content of flavonoids and anthocyanin in leaves increased, and the amino acid and their derivatives, organic acids, and flavonoids in roots increased. In contrast, under high phosphorus stress, a higher flavone content in the leaves and lower anthocyanins in the roots were found. Thus, apple seedlings respond to phosphorus stress by regulating the flavonoid pathway (Sun et al., 2021). Different concentrations of phosphate fertilizer were also found leading to changes in flavonoids in Chinese wolfberry (*Lycium barbarum* L.) (Wei et al., 2020).

Influence of Environment on Fruit Tree

The growth and development of fruit trees need light, temperature, air, water, and other non-nature environment. The change of environment will affect the physiology of fruit tree and quality of fruit. Metabolomics provides new method and new insight to reveal the mechanism of environmental regulation on fruit tree.

Light plays an important role in fruit growth. Bagging apple fruit reduces insect infestation, but it also reduces light exposure and increases the risk of browning. As found by Wang et al. (2021c), the bagging resulted in the production of 50 differential metabolites, in which a significant decrease in flavonoids was positively correlated with peel browning. In general, the inner fruit of turmeric trees received less light. The LC-MS and GC-MS analysis showed that sucrose and sorbitol contents were higher in the outer fruit, while glucose and malic acid contents were higher in the inner fruit (Rudell et al., 2017).

Persistent low temperature favors fruit trees to break dormancy, and changing temperature has an adverse effect on germination. Research of flower buds of pears treated at different temperatures showed that metabolites associated with the pentose phosphate pathway, energy production, and the tricarboxylic acid cycle (TCA) cycle may have reduced the germination rate of pear buds (Horikoshi et al., 2018). In contrast, [H_2CN_2 (HC)] breaking of blueberry dormancy is associated with increases in soluble sugars, organic acids, and amino acids (aspartate, glutamate, and phenylalanine) (Wang et al., 2021d). Treated with low temperature after harvesting leads to decrease of xylose, galactose, galacturonic acid, glucuronate, glycine, and rhamnose (R. Parijadi et al., 2019). Studies on *Butia* spp. (Arecaceae) growing at different latitude, longitude, and altitude in Brazil have found that fruits at lower altitude are sweeter, and fruits at higher altitude contain more carotenoids and phenols (Hoffmann et al., 2017).

Metabolomics was used to analyze the variation in fruit quality of four main Spanish grape cultivars RJ43, CL306, VN31, and 1084 under simulated climatic conditions [i.e., expected temperature, CO_2 content, and relative humidity (RH)] and different irrigation patterns (i.e., water sufficiency/water deficit) for 2,100 years. The results indicate that CO_2 levels in 2100 will increase sugar accumulation and reduce acidity in the grapes, but this effect is partially mitigated by insufficient water. In addition, climate change and moisture conditions significantly affected the concentration and intensity of amino acid accumulation in grapes, but these effects were different in different varieties. Besides, CO_2 content and water deficiency combined to reduce anthocyanin and anthocyanin/total soluble solids ratio in grapes. In general, although global climate change and moisture conditions can have a significant impact on grape quality, that is limited by the grape varieties. For example, RJ43 and CL306 are more affected by global changes, while 1,084 is relatively less affected (Arrizabalaga-Arriazu et al., 2021).

In addition, the biological environment can also affect the physiological and biochemical characteristics of fruit trees. Banana fruits near natural forests (near-NF) are very different from that distant from natural forests (distant-NF) in color, ripening characteristics, and susceptibility. Mature banana fruits harvested in near-NF contain more organic acids, GBAB (a four-carbon non-proteinogenic amino acid) and unsaturated fatty acids, and these metabolites increase the flavor and nutritional composition of the banana, while also enhancing the ability of the banana fruit to cope with biotic and abiotic stresses. Fruits harvested in distant-NF contain more glutamate and putrescine,

which makes bananas have an unpleasant smell after harvest and is also bad for storage (Nascimento et al., 2019).

Rootstocks and interstocks are generally used to dwarf fruit trees to achieve light and simplified management of the orchard. Comparison of the metabolome of pears grafted to different stocks revealed a significant increase in the concentrations of flavonoids and phenolic acids in scions grafted to dwarf stocks, along with significantly decreased concentrations of D-sorbitol and D-mannitol in the roots. It indicates that flavonoids and phenolic acids are key compounds involved in reducing scion growth and dwarf rootstock may control tree growth by regulating carbohydrate partitioning from shoot to roots (Cui et al., 2021). Training can also have effects on apple tree metabolism (D'Abrosca et al., 2017; Anthony et al., 2020). Girdling improves the quality of the sweet cherry fruit. The anthocyanin content in the peel varied significantly between pruning systems, and the slender spindle training system was the most favorable for fruit growth (Michailidis et al., 2020).

Genetic Mechanism of Fruit Tree

While the inheritance and variation of metabolites are controlled by smaller loci, traditional phenotypic traits are often located at larger loci. The addition of metabolomics can accelerate the research progress and candidate genes screening in fruit trees. By considering the metabolite levels as quantitative traits, the metabolic traits can be analyzed by QTL, i.e., metabolic QTL (mQTL) analysis (Carreno-Quintero et al., 2012). Combining metabolomics and GWAS, named mGWAS (Luo, 2015). Most of the mQTL work has been focused on *Arabidopsis thaliana*, *S. Lycopersicum* (Toubiana et al., 2015; Knoch et al., 2017), *Triticum aestivum*, *Oryza sativa,* and *Zea mays* (Gong et al., 2013; Jiang et al., 2013; Chen et al., 2021), and recently mQTL have also been applied in study the mechanism of fruit genetic basis.

To obtain hybrid plants with high-soluble solids content (SSC)/titratable acidity (TA) ratios in strawberries (*Fragaria* × *ananassa*), Vallarino et al. (2019) used GC-MS to detect primary metabolites in "232," "1,392," and their hybrid population F1 for 2 years. A total of 50 compounds were detected and successfully identified, including soluble sugars, organic acids, amino acids, soluble alcohols, etc. Pearson correlation analysis revealed possible synergistic regulation of fructose and glucose, glucose and fructose 6-phosphate, succinic acid and fumaric acid, and raffinose and succinic acid. The mQTL analysis was performed for these 50 major metabolites, SSC, TA, pH, and L-AA data using the constructed "232" × "1,392" linkage maps. The final results of 133 QTL localizations were obtained, and QTL for organic acids and sugars (glucuronide and succinate, raffinose and sucrose) were found in LG V-4. It was speculated that a common gene with pleiotropy for sucrose, raffinose, and succinate might cause these changes. The expression level of FvH4_5g03890 (glucose-6 -β 1-episomerase) gene was found to be negatively correlated with sucrose and raffinose contents by glucose biosynthesis, metabolism-related annotation and quantitative RT-PCR (Vallarino et al., 2019).

A hotspot of QTLs of phenolic compounds, including procyanidins and flavan-3-ols, in ripe apple fruits were mapped by mQTLs at the top of LG16 (Khan et al., 2012a). Higher expression of leucoanthocyanidin reductase gene (MdLAR1), which located on the LG16, was associated with a fourfold increase of procyanidin. Combining with expression analysis of structural and regulation genes of the phenylpropanoid and flavonoid pathways, MdLAR1 was identified the candidate gene for procyanidins and flavan-3-ols (Khan et al., 2012b).

A total of metabolic profiling of 101 F_2-generation grape stems were analyzed by LC-HR MS, and 1,317 characteristic ions were detected. Among them, 19 were related to stilbene compounds, and 5 mQTLs were finally identified by QTL localization. Two large-effect mQTLs, corresponding to a stilbenoid dimer and a trimer, were jointly localized in the 25.0–27.8 Mbps of LG18. This range also included 48 genes of bio-resistance, 15 genes in flavonoid biosynthesis pathway, and 6 genes of triterpenoid biosynthesis. In addition, haplotype dosage effects of 3 mQTLs were statistically significant. Fruit trees with high quality and disease resistance can be screened simultaneously (Teh et al., 2019).

Two hundred and seven jujube varieties, including 25 wild and 182 cultivated jujubes (*Ziziphus jujuba* Mill.), were used to study domestication-driven changes of fruit bioactive molecules by a combination of genome-sequencing, widely targeted metabolomics, and mGWAS. Genome-sequencing revealed 109 domestication-relative sweeps, which mainly distributed on chromosomes 1, 4, and 10. A total of 2,985 metabolites were annotated by a metabolome library with various bioactive metabolites. In which, 187 metabolites were associated with domestication. Totally, 1,080 associated loci for 986 metabolites were identified by mGWAS. Among them, 15 triterpenoids, which were significantly reduced in cultivated jujube, existed at six major loci. Mutations in the promoter of an OSC gene associated with triterpenes content reduction during domestication. In addition, by comparing the metabolites of 24 dry and 31 fresh types of jujube, 203 significantly different metabolites were identified, of which 7 cyclopeptide alkaloids (CPA) were highly accumulated in the dry variety. An *N*-methyltransferase were identified the CPA candidate gene by divergence sweep identification and its genomic sequence variations were responsible for content of CPAs in 39 jujube accessions (Zhang et al., 2021).

PERSPECTIVES

In recent years, fruit tree metabolomics-related research has increased rapidly, revealing the mechanism of fruit quality formation, biotic/abiotic stress, and environmental influence on fruit tree. Alternatively, the obtained differential metabolites can be used as novel biomarkers for the identification of variety, origin, and pathogen infection. Metabolomics-transcriptomic combined analysis, further revealed the gene/metabolic regulatory network and candidate genes leading to trait. Metabolomics, combining with GWAS/QTL, narrows the range of candidate genes and increases the odds of acquiring key genes.

However, the development of metabolomics in fruit trees is still in its initial stage, and most studies are the method

exploration and simple repetition of metabolomics application to fruit trees, obtaining differential metabolites but lacking in-depth mechanistic research. If the changes in the metabolome are combined with the changes in the organ, tissue, cellular, and molecular levels, the mechanism of trait changes should be even more systematically revealed. In addition, non-targeted metabolome is widely used in fruit tree metabolomics research, but only a few dozen compounds in thousands obtained can be identified, which not only leads to low metabolomics efficiency, but also the differential metabolites obtained may not real represent differences. The current metabolic library is mainly prepared for human, animal, and drug experiments and rarely covers the unique metabolites of fruit trees. The construction of metabolomics database of fruit tree will greatly promote the application of metabolomics in fruit tree. Widely targeted and pseudo-targeted metabolomes allow the quantitative and qualitative acquisition of thousands of compounds and are alternative solutions if the cost is reduced. Then, extreme metabolite diversity not only makes the fruit tree metabolomics study particularly required but also brings challenges to the various stages of metabolomics experiments, analysis, and library building. To further promote the application of metabolomics to fruit trees, the following problems or bottlenecks should be solved. Firstly, a more broad-spectrum and universal detection method should to be constructed. Secondly, a really unbiased, high-sensitive, and high-throughput metabolism analysis platform with the ability to parallel analyze huge amounts of data is needed. Thirdly, accurate and efficient identification or annotation of metabolites database or tools should be developed.

With the improvement of instruments and metabolite database and decrease of cost of experiment, metabolomics will prompt the fruit tree research to achieve more breakthrough results.

AUTHOR CONTRIBUTIONS

JL, ST, GY, XD, XH, and JW provided suggestions for the frame structure and content of the manuscript. JL wrote the first draft. ST assisted in the translation of the manuscript. GY and XD provided help in the arrangement of the table. KZ, GY, XD, XMZ, YZ, CW, and XZ contributed to the revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Promotion and Innovation of Beijing Academy of Agriculture and Forestry Sciences (KJCX201910), the technological innovation ability project of Beijing Academy of Agriculture and Forestry Sciences (KJCX20210403), Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team), the National Natural Science Foundation of China (32072536), and National Key R&D Program of China from the Ministry of Science and Technology of China (Grant No. 2019YFC1711100), Youth Research Fund of the Beijing Academy of Forestry and Fruit Sciences (LGYJJ202006 and LGYJJ202005).

REFERENCES

Anthony, B. M., Chaparro, J. M., Prenni, J. E., and Minas, I. S. (2020). Early metabolic priming under differing carbon sufficiency conditions influences 2 peach fruit quality development. *Plant Physiol. Biochem.* 157, 416–431. doi: 10.1016/j.plaphy.2020.11.004

Arrizabalaga-Arriazu, M., Gomès, E., Morales, F., Irigoyen, J. J., Pascual, I., and Hilbert, G. (2021). Impact of 2100-projected air temperature, carbon dioxide, and water scarcity on grape primary and secondary metabolites of different *Vitis vinifera* cv. Tempranillo Clones. *J. Agric. Food Chem.* 69, 6172–6185. doi: 10.1021/acs.jafc.1c01412

Baky, M. H., Badawy, M. T., Bakr, A. F., Hegazi, N. M., Abdellatif, A., and Farag, M. A. (2021). Metabolome-based profiling of African baobab fruit (*Adansonia digitata* L.) using a multiplex approach of MS and NMR techniques in relation to its biological activity. *RSC Adv.* 11, 39680–39695. doi: 10.1039/D1RA08277A

Ballester, A.-R., Lafuente, M. T., de Vos, R. C. H., Bovy, A. G., and González-Candelas, L. (2013). Citrus phenylpropanoids and defence against pathogens. Part I: metabolic profiling in elicited fruits. *Food Chem.* 136, 178–185. doi: 10.1016/j.foodchem.2012.07.114

Berni, R., Charton, S., Planchon, S., Legay, S., Romi, M., Cantini, C., et al. (2021). Molecular investigation of Tuscan sweet cherries sampled over three years: gene expression analysis coupled to metabolomics and proteomics. *Hortic. Res.* 8:12. doi: 10.1038/s41438-020-00445-3

Cajka, T., and Fiehn, O. (2016). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* 88, 524–545. doi: 10.1021/acs.analchem.5b04491

Carreno-Quintero, N., Acharjee, A., Maliepaard, C., Bachern, C., Mumm, R., Bouwmeester, H., et al. (2012). Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiol.* 158, 1306–1318. doi: 10.1104/pp.111.188441

Carreno-Quintero, N., Bouwmeester, H. J., and Keurentjes, J. J. B. (2013). Genetic analysis of metabolome–phenotype interactions: from model to crop species. *Trends Genet.* 29, 41–50. doi: 10.1016/j.tig.2012.09.006

Chen, J., Xue, M., Liu, H., Fernie, A. R., and Chen, W. (2021). Exploring the genic resources underlying metabolites through mGWAS and mQTL in wheat: from large-scale gene identification and pathway elucidation to crop improvement. *Plant Commun.* 2:100216. doi: 10.1016/j.xplc.2021.100216

Cui, Z., Zhang, H., Galarneau, E. R. -A., Yang, Y., Li, D., Song, J., et al. (2021). Metabolome analysis reveals important compounds related to dwarfing effect of interstock on scion in pear. *Ann. Appl. Biol.* 179, 108–122. doi: 10.1111/aab.12684

D'Abrosca, B., Scognamiglio, M., Corrado, L., Chiocchio, I., Zampella, L., Mastrobuoni, F., et al. (2017). Evaluation of different training systems on Annurca apple fruits revealed by agronomical, qualitative and NMR-based metabolomic approaches. *Food Chem.* 222, 18–27. doi: 10.1016/j.foodchem.2016.11.144

Dai, Z. W., Léon, C., Feil, R., Lunn, J. E., Delrot, S., and Gomès, E. (2013). Metabolic profiling reveals coordinated switches in primary carbohydrate metabolism in grape berry (*Vitis vinifera* L.), a non-climacteric fleshy fruit. *J. Exp. Bot.* 64, 1345–1355. doi: 10.1093/jxb/ers396

de Godoy, F., Bermúdez, L., Lira, B. S., de Souza, A. P., Elbl, P., Demarco, D., et al. (2013). Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. *J. Exp. Bot.* 64, 2449–2466. doi: 10.1093/jxb/ert106

Dickson, L., Tenon, M., Svilar, L., Fança-Berthon, P., Martin, J. C., Rogez, H., et al. (2020). Genipap (*Genipa americana* L.) juice intake biomarkers after medium-term consumption. *Food Res. Int.* 137:109375. doi: 10.1016/j.foodres.2020.109375

Dixon, R. A., and Strack, D. (2003). Phytochemistry meets genome analysis, and beyond. *Phytochemistry* 62, 815–816. doi: 10.1016/S0031-9422(02)00712-4

Dominguez, P. G., Conti, G., Duffy, T., Insani, M., Alseekh, S., Asurmendi, S., et al. (2021). Multiomics analyses reveal the roles of the ASR1 transcription factor in tomato fruits. *J. Exp. Bot.* 72, 6490–6509. doi: 10.1093/jxb/erab269

Fagbohun, O. F., Olawoye, B., Ademakinwa, A. N., Oriyomi, O. V., Fagbohun, O. S., Fadare, O. A., et al. (2021). UHPLC/GC-TOF-MS metabolomics, MTT assay, and molecular docking studies reveal physostigmine as a new anticancer agent from the ethyl acetate and butanol fractions of *Kigelia africana* (Lam.) Benth. fruit extracts. *Biomed. Chromatogr.* 35:e4979. doi: 10.1002/bmc.4979

Falasca, A., Melck, D., Paris, D., Saviano, G., Motta, A., and Iorizzi, M. (2014). Seasonal changes in the metabolic fingerprint of *Juniperus communis* L. berry extracts by ^{1}H NMR-based metabolomics. *Metabolomics* 10, 165–174. doi: 10.1007/s11306-013-0566-1

Fayek, N. M., Farag, M. A., Abdel Monem, A. R., Moussa, M. Y., Abd-Elwahab, S. M., and El-Tanbouly, N. D. (2019). Comparative Metabolite Profiling of Four *Citrus* Peel Cultivars *via* ultra-performance liquid chromatography coupled with quadrupole-time-of-flight-mass spectrometry and multivariate data analyses. *J. Chromatogr. Sci.* 57, 349–360. doi: 10.1093/chromsci/bmz006

Fiehn, O., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000). Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* 72, 3573–3580. doi: 10.1021/ac991142i

Fuentealba, C. (2021). Cell wall and metabolite composition of sweet cherry fruits from two cultivars with contrasting susceptibility to surface pitting during storage. *Food Chem.* 342:128307. doi: 10.1016/j.foodchem.2020.128307

Gai, Y. P., Han, X. J., Li, Y. Q., Yuan, C. Z., Mo, Y. Y., Guo, F. Y., et al. (2014). Metabolomic analysis reveals the potential metabolites and pathogenesis involved in mulberry yellow dwarf disease: phytoplasma invasion impaired mulberry metabolism. *Plant Cell Environ.* 37, 1474–1490. doi: 10.1111/pce.12255

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., et al. (2013). Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20320–20325. doi: 10.1073/pnas.1319681110

Hoffmann, J. F., Carvalho, I. R., Barbieri, R. L., Rombaldi, C. V., and Chaves, F. C. (2017). *Butia* spp. (Arecaceae) LC-MS-based metabolomics for species and geographical origin discrimination. *J. Agric. Food Chem.* 65, 523–532. doi: 10.1021/acs.jafc.6b03203

Hong, K., Chen, L., Gu, H., Zhang, X., Chen, J., Nile, S. H., et al. (2021). Novel insight into the relationship between metabolic profile and fatty acid accumulation altering cellular lipid content in pineapple fruits at different stages of maturity. *J. Agric. Food Chem.* 69, 8578–8589. doi: 10.1021/acs.jafc.1c02658

Horikoshi, H. M., Sekozawa, Y., Kobayashi, M., Saito, K., Kusano, M., and Sugaya, S. (2018). Metabolomics analysis of "Housui" Japanese pear flower buds during endodormancy reveals metabolic suppression by thermal fluctuation. *Plant Physiol. Biochem.* 126, 134–141. doi: 10.1016/j.plaphy.2018.02.028

Hurtado-Fernández, E., Pacchiarotta, T., Mayboroda, O. A., Fernández-Gutiérrez, A., and Carrasco-Pancorbo, A. (2015). Metabolomic analysis of avocado fruits by GC-APCI-TOF MS: effects of ripening degrees and fruit varieties. *Anal. Bioanal. Chem.* 407, 547–555. doi: 10.1007/s00216-014-8283-9

Ji, X., Ren, J., Lang, S., Wang, D., Zhu, L., and Song, X. (2020). Differential regulation of Anthocyanins in *Cerasus humilis* fruit color revealed by combined transcriptome and metabolome analysis. *Forests* 11:1065. doi: 10.3390/f11101065

Jiang, P. S., Zhang, H. X., Lü, X. L., Hao, Z. F., Li, B., Li, M. S., et al. (2013). Analysis of Meta-QTL and candidate genes related to yield components in maize. *Acta Agron. Sin.* 39:969. doi: 10.3724/SP.J.1006.2013.00969

Jing, S., and Malladi, A. (2020). Higher growth of the apple (*Malus* × *domestica* Borkh.) fruit cortex is supported by resource intensive metabolism during early development. *BMC Plant Biol.* 20:75. doi: 10.1186/s12870-020-2280-2

Karagiannis, E. (2018). Postharvest responses of sweet cherry fruit and stem tissues revealed by metabolomic profiling. *Plant Physiol. Biochem.* 127, 478–484. doi: 10.1016/j.plaphy.2018.04.029

Karagiannis, E., Sarrou, E., Michailidis, M., Tanou, G., Ganopoulos, I., Bazakos, C., et al. (2021). Fruit quality trait discovery and metabolic profiling in sweet cherry genebank collection in Greece. *Food Chem.* 342:128315. doi: 10.1016/j.foodchem.2020.128315

Khalil, M. N. A., Fekry, M. I., and Farag, M. A. (2017). Metabolome based volatiles profiling in 13 date palm fruit varieties from Egypt *via* SPME GC–MS and chemometrics. *Food Chem.* 217, 171–181. doi: 10.1016/j.foodchem.2016.08.089

Khan, S. A., Chibon, P. Y., de Vos, R. C. H., Schipper, B. A., Walraven, E., Beekwilder, J., et al. (2012a). Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16. *J. Exp. Bot.* 63, 2895–2908. doi: 10.1093/jxb/err464

Khan, S. A., Schaart, J. G., Beekwilder, J., Allan, A. C., Tikunov, Y. M., Jacobsen, E., et al. (2012b). The mQTL hotspot on linkage group 16 for phenolic compounds in apple fruits is probably the result of a leucoanthocyanidin reductase gene at that locus. *BMC Res. Notes* 5:618. doi: 10.1186/1756-0500-5-618

Kim, H. K., and Verpoorte, R. (2010). Sample preparation for plant metabolomics. *Phytochem. Anal.* 21, 4–13. doi: 10.1002/pca.1188

Klee, S. M., Sinn, J. P., Finley, M., Allman, E. L., Smith, P. B., Aimufua, O., et al. (2019). *Erwinia amylovora* auxotrophic mutant exometabolomics and virulence on apples. *Appl. Environ. Microbiol.* 85:e00935-19. doi: 10.1128/AEM.00935-19

Knoch, D., Riewe, D., Meyer, R. C., Boudichevskaia, A., Schmidt, R., and Altmann, T. (2017). Genetic dissection of metabolite variation in Arabidopsis seeds: evidence for mQTL hotspots and a master regulatory locus of seed metabolism. *J. Exp. Bot.* 68, 1655–1667. doi: 10.1093/jxb/erx049

Koch, K. (2004). Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* 7, 235–246. doi: 10.1016/j.pbi.2004.03.014

Lee, K.-M., Oh, T.-J., Kim, S.-H., Kim, H.-Y., Chung, H., Min, D. S., et al. (2016). Comprehensive metabolic profiles of mulberry fruit (*Morus alba* Linnaeus) according to maturation stage. *Food Sci. Biotechnol.* 25, 1035–1041. doi: 10.1007/s10068-016-0167-7

Lee, Y. R., Cho, H. M., Park, E. J., Zhang, M., Doan, T. P., Lee, B. W., et al. (2020). Metabolite Profiling of Rambutan (*Nephelium lappaceum* L.) seeds using UPLC-qTOF-MS/MS and senomorphic effects in aged human dermal fibroblasts. *Nutrients* 12:1430. doi: 10.3390/nu12051430

Li, B., Neumann, E. K., Ge, J., Gao, W., Hua, Y., Ping, L., et al. (2018). Interrogation of spatial metabolome of ginkgo biloba with High-resolution MALDI and LDI mass spectrometry imaging. *Plant Cell Environ.* 41, 2693–2703. doi: 10.1111/pce.13395

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A. R. (2015). Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc.* 1, 387–396. doi: 10.1038/nprot.2006.59

Lu, X.-P., Li, F. F., Xiong, J., Cao, X. J., Ma, X. C., Zhang, Z. M., et al. (2017). Transcriptome and metabolome analyses provide insights into the occurrence of peel roughing disorder on *Satsuma Mandarin* (Citrus unshiu Marc.) Fruit. *Front. Plant Sci.* 8:1907. doi: 10.3389/fpls.2017.01907

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24, 31–38. doi: 10.1016/j.pbi.2015.01.006

Lwin, H. P., and Lee, J. (2020). Fruit quality and major metabolites in cold-stored 'Wonhwang' Asian pears are differentially affected by fruit size. *J. Sci. Food Agric.* 100, 5117–5125. doi: 10.1002/jsfa.10422

Ma, A., and Qi, X. (2021). Mining plant metabolomes: methods, applications, and perspectives. *Plant Commun.* 2:100238. doi: 10.1016/j.xplc.2021.100238

Mamat, S. F., Azizan, K. A., Baharum, S. N., Noor, N. M., and Aizat, W. M. (2020). GC-MS and LC-MS analyses reveal the distribution of primary and secondary metabolites in mangosteen (*Garcinia mangostana* Linn.) fruit during ripening. *Sci. Hortic.* 262:109004. doi: 10.1016/j.scienta.2019.109004

Michailidis, M., Karagiannis, E., Tanou, G., Samiotaki, M., Sarrou, E., Karamanoli, K., et al. (2020). Proteomic and metabolic analysis reveals novel sweet cherry fruit development regulatory points influenced by girdling. *Plant Physiol. Bioch.* 149, 233–244. doi: 10.1016/j.plaphy.2020.02.017

Montefiori, M., McGhie, T. K., Hallett, I. C., and Costa, G. (2009). Changes in pigments and plastid ultrastructure during ripening of green-fleshed and yellow-fleshed kiwifruit. *Sci. Hortic.* 119, 377–387. doi: 10.1016/j.scienta.2008.08.022

Muto, A., Müller, C. T., Bruno, L., McGregor, L., Ferrante, A., Chiappetta, A. A. C., et al. (2020). Fruit volatilome profiling through GC × GC-ToF-MS and gene expression analyses reveal differences amongst peach cultivars in their response to cold storage. *Sci. Rep.* 10:18333. doi: 10.1038/s41598-020-75322-z

Nascimento, T. P., Castro-Alves, V. C., Castelan, F. P., Calhau, M. F. N. S., Saraiva, L. A., Agopian, R. G., et al. (2019). Metabolomic profiling reveals that natural biodiversity surrounding a banana crop may positively influence the nutritional/sensorial profile of ripe fruits. *Food Res. Int.* 124, 165–174. doi: 10.1016/j.foodres.2018.07.050

Nicholson, J. K., Lindon, J. C., and Holmes, E. (1999). "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli *via* multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29, 1181–1189. doi: 10.1080/004982599238047

Olmo-García, L., Kessler, N., Neuweger, H., Wendt, K., Olmo-Peinado, J., Fernández-Gutiérrez, A., et al. (2018). Unravelling the distribution of secondary metabolites in *Olea europaea* L.: exhaustive characterization of eight olive-tree derived matrices by complementary platforms (LC-ESI/APCI-MS and GC-APCI-MS). *Molecules* 23:2419. doi: 10.3390/molecules23102419

Padhi, E. M. T., Maharaj, N., Lin, S.-Y., Mishchuk, D. O., Chin, E., Godfrey, K., et al. (2019). Metabolome and microbiome signatures in the roots of citrus affected by huanglongbing. *Phytopathology* 109, 2022–2032. doi: 10.1094/PHYTO-03-19-0103-R

Park, S. E., Seo, S. H., Kim, E. J., Park, D. H., Park, K. M., Cho, S. S., et al. (2019). Metabolomic approach for discrimination of cultivation age and ripening stage in ginseng berry using gas chromatography-mass spectrometry. *Molecules* 24:3837. doi: 10.3390/molecules24213837

R. Parijadi, A. A., Ridwani, S., Dwivany, F. M., Putri, S. P., and Fukusaki, E. (2019). A metabolomics-based approach for the evaluation of off-tree ripening conditions and different postharvest treatments in mangosteen (*Garcinia mangostana*). *Metabolomics* 15:73. doi: 10.1007/s11306-019-1526-1

Ramadan, N. S., Wessjohann, L. A., Mocan, A., Vodnar, D. C., El-Sayed, N. H., El-Toumy, S. A., et al. (2020). Nutrient and sensory metabolites profiling of *Averrhoa Carambola* L. (Starfruit) in the context of its origin and ripening stage by GC/MS and chemometric analysis. *Molecules* 25:2423. doi: 10.3390/molecules25102423

Ramírez-Briones, E., Rodríguez Macías, R., Casarrubias Castillo, K., del Río, R. E., Martínez-Gallardo, N., Tiessen, A., et al. (2019). Fruits of wild and semi-domesticated *Diospyros* tree species have contrasting phenological, metabolic, and antioxidant activity profiles. *J. Sci. Food Agric.* 99, 6020–6031. doi: 10.1002/jsfa.9878

Risticevic, S., Souza-Silva, E. A., DeEll, J. R., Cochran, J., and Pawliszyn, J. (2016). Capturing plant metabolome with direct-immersion *in vivo* solid phase microextraction of plant tissues. *Anal. Chem.* 88, 1266–1274. doi: 10.1021/acs.analchem.5b03684

Risticevic, S., Souza-Silva, E. A., Gionfriddo, E., DeEll, J. R., Cochran, J., Hopkins, W. S., et al. (2020). Application of *in vivo* solid phase microextraction (SPME) in capturing metabolome of apple (*Malus* × *domestica* Borkh.) fruit. *Sci. Rep.* 10:6724. doi: 10.1038/s41598-020-63817-8

Rivera-Mondragón, A., Tuenter, E., Ortiz, O., Sakavitsi, M. E., Nikou, T., Halabalaki, M., et al. (2020). UPLC-MS/MS-based molecular networking and NMR structural determination for the untargeted phytochemical characterization of the fruit of *Crescentia cujete* (*Bignoniaceae*). *Phytochemistry* 177:112438. doi: 10.1016/j.phytochem.2020.112438

Rudell, D. R., and Mattheis, J. P. (2009). Superficial scald development and related metabolism is modified by postharvest light irradiation. *Postharvest Biol. Technol.* 51, 174–182. doi: 10.1016/j.postharvbio.2008.07.008

Rudell, D. R., Serra, S., Sullivan, N., Mattheis, J. P., and Musacchi, S. (2017). Survey of 'd'Anjou' pear metabolic profile following harvest from different canopy positions and fruit tissues. *HortScience* 52, 1501–1510. doi: 10.21273/HORTSCI12375-17

Ryu, S., Muramatsu, T., Furihata, K., Wei, F., Koda, M., Miyakawa, T., et al. (2019). NMR-based metabolic profiling and comparison of Japanese persimmon cultivars. *Sci. Rep.* 9:15011. doi: 10.1038/s41598-019-51489-y

Shaffer, R. E. (2002). Multi- and Megavariate Data Analysis. Principles and applications, I. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold, Umetrics Academy, Ume, 2001, ISBN 91-973730-1-X, 533pp. *J. Chemom.* 16, 261–262. doi: 10.1002/cem.713

Smith, E. D., Whiting, M. D., and Rudell, D. R. (2011). Metabolic profiling of ethephon-treated sweet cherry (*Prunus avium* L.). *Metabolomics* 7, 126–133. doi: 10.1007/s11306-010-0238-3

Suh, J. H., Tang, X., Zhang, Y., Gmitter, F. G., and Wang, Y. (2021). Metabolomic analysis provides new insight into tolerance of Huanglongbing in Citrus. *Front. Plant Sci.* 12:710598. doi: 10.3389/fpls.2021.710598

Sumner, L. W., Mendes, P., and Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62, 817–836. doi: 10.1016/S0031-9422(02)00708-2

Sun, T., Zhang, J., Zhang, Q., Li, X., Li, M., Yang, Y., et al. (2021). Transcriptome and metabolome analyses revealed the response mechanism of apple to different phosphorus stresses. *Plant Physiol. Biochem.* 167, 639–650. doi: 10.1016/j.plaphy.2021.08.040

Tang, N., Chen, N., Hu, N., Deng, W., Chen, Z., and Li, Z. (2018). Comparative metabolomics and transcriptomic profiling reveal the mechanism of fruit quality deterioration and the resistance of citrus fruit against *Penicillium digitatum*. *Postharvest Biol. Technol.* 145, 61–73. doi: 10.1016/j.postharvbio.2018.06.007

Teh, S. L., Rostandy, B., Awale, M., Luby, J. J., Fennell, A., and Hegeman, A. D. (2019). Genetic analysis of stilbenoid profiles in grapevine stems reveals a major mQTL hotspot on chromosome 18 associated with disease-resistance motifs. *Hortic. Res.* 6:121. doi: 10.1038/s41438-019-0203-x

Toubiana, D., Batushansky, A., Tzfadia, O., Scossa, F., Khan, A., Barak, S., et al. (2015). Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* 81, 121–133. doi: 10.1111/tpj.12717

Urbanczyk-Wochniak, E., Baxter, C., Kolbe, A., Kopka, J., Sweetlove, L. J., and Fernie, A. R. (2005). Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. *Planta* 221, 891–903. doi: 10.1007/s00425-005-1483-y

Vallarino, J. G., Pott, D. M., Cruz-Rus, E., Miranda, L., Medina-Minguez, J. J., Valpuesta, V., et al. (2019). Identification of quantitative trait loci and candidate genes for primary metabolite content in strawberry fruit. *Hortic. Res.* 6:4. doi: 10.1038/s41438-018-0077-3

Villa-Ruano, N., Pérez-Hernández, N., Zepeda-Vallejo, L. G., Quiroz-Acosta, T., Mendieta-Moctezuma, A., Montoya-García, C., et al. (2019). 1 H-NMR based metabolomics profiling of citrus juices produced in veracruz, México. *Chem. Biodivers.* 16:e1800479. doi: 10.1002/cbdv.201800479

Vos, R., Moco, S., Lommen, A., Keurentjes, J. J., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778–791. doi: 10.1038/nprot.2007.95

Wang, F., Huang, Y., Wu, W., Zhu, C., Zhang, R., Chen, J., et al. (2020). Metabolomics analysis of the peels of different colored citrus fruits (*Citrus reticulata* cv. 'Shatangju') during the maturation period based on UHPLC-QQQ-MS. *Molecules* 25:396. doi: 10.3390/molecules25020396

Wang, F., Ge, S., Xu, X., Xing, Y., Du, X., Zhang, X., et al. (2021a). Multiomics analysis reveals new insights into the apple fruit quality decline under high nitrogen conditions. *J. Agric. Food Chem.* 69, 5559–5572. doi: 10.1021/acs.jafc.1c01548

Wang, H., Asker, K., Zhan, C., and Wang, N. (2021b). Transcriptomic and metabolic analysis of fruit development and identification of genes involved in raffinose and hydrolysable tannin biosynthesis in walnuts. *J. Agric. Food Chem.* 69, 8050–8062. doi: 10.1021/acs.jafc.1c02434

Wang, H., Wang, S., Fan, M. M., Zhang, S. H., Sun, L. L., and Zhao, Z. Y. (2021c). Metabolomic insights into the browning of the peel of bagging 'Rui Xue' apple fruit. *BMC Plant Biol.* 21:209. doi: 10.1186/s12870-021-02974-y

Wang, H., Xia, X., and An, L. (2021d). Metabolomics analysis reveals the mechanism of hydrogen cyanamide in promoting flower bud break in blueberry. *Agronomy* 11:102. doi: 10.3390/agronomy11010102

Wang, J., Yang, E., Chaurand, P., and Raghavan, V. (2021e). Visualizing the distribution of strawberry plant metabolites at different maturity stages by MALDI-TOF imaging mass spectrometry. *Food Chem.* 345:128838. doi: 10.1016/j.foodchem.2020.128838

Wang, Q., Wu, X., Liu, L., Yao, D., Li, J., Fang, J., et al. (2021f). Transcriptome and metabolomic analysis to reveal the browning spot formation of 'Huangguan' pear. *BMC Plant Biol.* 21:321. doi: 10.1186/s12870-021-03049-8

Wang, R., Shu, P., Zhang, C., Zhang, J., Chen, Y., Zhang, Y., et al. (2022). Integrative analyses of metabolome and genome-wide transcriptome reveal the regulatory network governing flavor formation in kiwifruit (*Actinidia chinensis*). *New Phytol.* 233, 373–389. doi: 10.1111/nph.17618

Wang, S., Tu, H., Wan, J., Chen, W., Liu, X., Luo, J., et al. (2016). Spatio-temporal distribution and natural variation of metabolites in citrus fruits. *Food Chem.* 199, 8–17. doi: 10.1016/j.foodchem.2015.11.113

Wang, Z., Cui, Y., Vainstein, A., Chen, S., and Ma, H. (2017). Regulation of Fig (*Ficus carica* L.) Fruit Color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front. Plant Sci.* 8:1990. doi: 10.3389/fpls.2017.01990

Wei, C., Liang, G., 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

Wei, F., Shi, Z., Wan, R., Li, Y., Wang, Y., An, W., et al. (2020). Impact of phosphorus fertilizer level on the yield and metabolome of goji fruit. *Sci. Rep.* 10:14656. doi: 10.1038/s41598-020-71492-y

Wu, T., Hu, D., and Wang, Q. (2021). Noni fruit's water spot appearance on the second day of harvest: a trade-off between resistance and energy. *Chem. Biol. Technol. Agric.* 8:8. doi: 10.1186/s40538-020-00207-2

Wu, T., Li, M., and Lan, Z. (2019). Reveal the variation patterns of chemical composition in the fruit of *Morinda citrifolia* (noni) during postharvest storage through metabolomic characterization. *Trop. Plant Biol.* 12, 85–97. doi: 10.1007/s12042-019-09222-6

Xu, J., Yan, J., Li, W., Wang, Q., Wang, C., Guo, J., et al. (2020). Integrative analyses of widely targeted metabolic profiling and transcriptome data reveals molecular insight into metabolomic variations during apple (*Malus domestica*) fruit development and ripening. *Int. J. Mol. Sci.* 21:4797. doi: 10.3390/ijms21134797

Yazdanpanah, P., Jonoubi, P., Zeinalabedini, M., Rajaei, H., Ghaffari, M. R., Vazifeshenas, M. R., et al. (2021). Seasonal metabolic investigation in pomegranate (*Punica granatum* L.) highlights the role of amino acids in genotype- and organ-specific adaptive responses to freezing stress. *Front. Plant Sci.* 12:699139. doi: 10.3389/fpls.2021.699139

Yi, D., Zhang, H., Lai, B., Liu, L., Pan, X., Ma, Z., et al. (2021). Integrative analysis of the coloring mechanism of red longan pericarp through metabolome and transcriptome analyses. *J. Agric. Food Chem.* 69, 1806–1815. doi: 10.1021/acs.jafc.0c05023

Yost, R. A., and Enke, C. G. (1978). Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J. Am. Chem. Soc.* 100, 2274–2275. doi: 10.1021/ja00475a072

Zhang, J., Wang, X., Yu, O., Tang, J., Gu, X., Wan, X., et al. (2011). Metabolic profiling of strawberry (*Fragaria* × *ananassa* Duch.) during fruit development and maturation. *J. Exp. Bot.* 62, 1103–1118. doi: 10.1093/jxb/erq343

Zhang, Q., Wang, L., Liu, Z., Zhao, Z., Zhao, J., Wang, Z., et al. (2020). Transcriptome and metabolome profiling unveil the mechanisms of *Ziziphus jujuba* Mill. peel coloration. *Food Chem.* 312:125903. doi: 10.1016/j.foodchem.2019.125903

Zhang, S., Ying, H., Pingcuo, G., Wang, S., Zhao, F., Cui, Y., et al. (2019). Identification of potential metabolites mediating bird's selective feeding on *Prunus mira* flowers. *BioMed. Res. Int.* 2019:1395480. doi: 10.1155/2019/1395480

Zhang, Z., Shi, Q., Wang, B., Ma, A., Wang, Y., Xue, Q., et al. (2021). Jujube metabolome selection determined the edible properties acquired during domestication. *Plant J.* 109, 1116–1133. doi: 10.1111/tpj.15617

Zheng, F., Zhao, X., Zeng, Z., Wang, L., Lv, W., Wang, Q., et al. (2020). Development of a plasma pseudotargeted metabolomics method based on ultra-high-performance liquid chromatography–mass spectrometry. *Nat. Protoc.* 15, 2519–2537. doi: 10.1038/s41596-020-0341-5

Zhou, W., Zhao, S., Xu, M., Niu, Y., Nasier, M., Fan, G., et al. (2021). Identification of key genes controlling carotenoid metabolism during apricot fruit development by integrating metabolic phenotypes and gene expression profiles. *J. Agric. Food Chem.* 69, 9472–9483. doi: 10.1021/acs.jafc.1c00496

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 May 2022
doi: 10.3389/fpls.2022.881394

Global Transcriptome Analysis Revealed the Molecular Regulation Mechanism of Pigment and Reactive Oxygen Species Metabolism During the Stigma Development of *Carya cathayensis*

Yulin Xing†, Ketao Wang†, Chunying Huang, Jianqin Huang, Yirui Zhao, Xiaolin Si and Yan Li**

State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, China

OPEN ACCESS

Edited by:
Yongliang Liu,
University of Kentucky, United States

Reviewed by:
Yanqun Xu,
Zhejiang University, China
Jian Gao,
Yangtze Normal University, China

***Correspondence:**
Ketao Wang
wangkt@zafu.edu.cn
Yan Li
20180061@zafu.edu.cn

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

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *22 February 2022*
Accepted: *22 March 2022*
Published: *09 May 2022*

Citation:
Xing Y, Wang K, Huang C, Huang J, Zhao Y, Si X and Li Y (2022) Global Transcriptome Analysis Revealed the Molecular Regulation Mechanism of Pigment and Reactive Oxygen Species Metabolism During the Stigma Development of Carya cathayensis. Front. Plant Sci. 13:881394. doi: 10.3389/fpls.2022.881394

Hickory (*Carya cathayensis* Sarg.) is a monoecious plant of the genus *Carya* of the Juglandaceae family. Its nuts contain a number of nutritional compounds and are deeply loved by consumers. Interestingly, it was observed that the color of hickory stigma changed obviously from blooming to mature. However, the molecular mechanism underlying color formation during stigma development and the biological significance of this phenomenon was mostly unknown. In this work, pigment content, reactive oxygen species (ROS) removal capacity, and transcriptome analysis of developing stigma of hickory at 4 differential sampling time points (S1, S2, S3, and S4) were performed to reveal the dynamic changes of related pigment, antioxidant capacity, and its internal molecular regulatory mechanism. It was found that total chlorophyll content was decreased slightly from S1 to S4, while total carotenoids content was increased from S1 to S3 but decreased gradually from S3 to S4. Total anthocyanin content continued to increase during the four periods of stigma development, reaching the highest level at the S4. Similarly, the antioxidant capacity of stigma was also gradually improved from S1 to S4. Furthermore, transcriptome analysis of developing hickory stigma identified 31,027 genes. Time-series analysis of gene expressions showed that these genes were divided into 12 clusters. Cluster 5 was enriched with some genes responsible for porphyrin and chlorophyll metabolism, carotenoid metabolism, and photosynthesis. Meanwhile, cluster 10 was enriched with genes related to flavonoid metabolism, including anthocyanin involved in ROS scavenging, and its related genes were mainly distributed in cluster 12. Based on the selected threshold values, a total of 10432 differentially expressed genes were screened out and enriched in the chlorophyll, carotenoid, anthocyanin, and ROS metabolism. The expression trends of these genes provided plausible explanations for the dynamic change of color and ROS level of hickory stigma with development. qRT-PCR analyses were basically consistent with the results of RNA-seq. The gene co-regulatory networks of pigment and ROS metabolism were further constructed and *MYB113* (CCA0887S0030) and *WRKY75*

(CCA0573S0068) were predicted to be two core transcriptional regulators. These results provided in-depth evidence for revealing the molecular mechanism of color formation in hickory stigma and its biological significance.

Keywords: *Carya cathayensis*, color change, gene expression, gene co-regulatory networks, pigment metabolism, ROS scavenging, stigma development, transcriptome analysis

INTRODUCTION

The flower stigma is receptive portions of the female tissues that bind pollen and mediate tube migration into the style. It can be divided into dry and wet stigmas based on the structure. Angiosperms that produce trinucleate pollen usually have a dry stigma, while binucleate pollen usually interacts with a wet stigma. Although they have different morphological structures, all stigmas perform similar functions, including pollen capture and hydration, pollen tube guidance, and dispersal, all of which are crucial for successful fertilization and controlling seed yield (Nasrallah, 2000; Edlund et al., 2004).

Flower blossoms are products of sexual selection for traits that enhance mating success. Flowers vary in color, pattern, shape, and scent, which alone or in combination can act as signals for the attraction of animal pollinators (Pannell, 2017). Flower color plays an essential function in plant ecology and evolution by attracting animal pollinators (Grotewold, 2006), being a central feature of plant observation, and having a significant ornamental function (Xue et al., 2020). Due to the importance of color formation in angiosperm, especially in the ornamental plant, the biosynthetic pathways of pigment in color formation have been widely reported (Grotewold, 2006; Chen et al., 2016; Xia et al., 2021). Three chemically distinct groups of pigments, chlorophylls, carotenoids, and anthocyanins are widely distributed in plants and are the major pigments in flower color formation (Xue et al., 2020; Xia et al., 2021).

Chlorophyll is a critical component in almost all plants and makes plants green, which mainly participates in photosynthesis (Hörtensteiner, 2013). When photosynthetic organs are overexcited, such as under light stress conditions, chlorophyll can act as a photosensitizer, leading to cell damage and death to protect them (Apel and Hirt, 2004). After a lot of research, the pathway of chlorophyll metabolism has been very clear, which need the involvement of enzymes and transcription factors encoded by many genes, such as *glutamyl tRNA reductase (Glu-TR), glutamate-1-semialdehyde-2,1-amino mutase (GSA-AM), Porphobilinogen synthase (PBCS), bile chromogen dehydrogenase (PBGD), Uroporphyrinogen III synthase (UROS), Uroporphyrinogen III decarboxylase (UROD), Coproporphyrinogens III oxidase (CPOX), Protoporphyrinogen oxidase (PPOX), Magnesium chelatase H subunit (MgCh), Magnesium proto IX methyltransferase (MgPM), Mg-protoporphyrin IX monomethylester (MgPEC), 3,8-divinyl protochlorophyllide a 8-vinyl reductase (DVR), Protochlorophyllide oxidoreductase (POR), Chlorophyll synthase (CHLG), Chlorophyllide a oxygenase (CAO), hydroxy-Chl a reductasehydroxy-Chl a reductase (HCAR), non-yellow coloring1/NYC1-like (NYC1/NOL), metal-chelating substance (MCS), pheophytinase (PPH), pheophorbide a oxygenase (PAO), primary fluorescent Chl Catabolite (pFCC), and red Chl catabolite reductase (RCCR)* (Beale, 2005; Hörtensteiner, 2013).

Carotenoids are the second most abundant natural pigments on earth and are usually function in photosynthesis and photoprotection as well as growth and development (Tanaka and Ohmiya, 2008). Carotenoids are terpenoids chemicals and contribute to the formation of the yellow, orange, and red hue of most flowers (Tanaka and Ohmiya, 2008). It is the only precursor to vitamin A biosynthesis that is particularly beneficial in humans by promoting antioxidant activity and reducing age-related macular degeneration of the eye (Davies, 2007; Giuliano et al., 2008). Carotenoids can also quench O_2 by a chemical mechanism involving their oxidation (Ramel et al., 2012). Its metabolism has been clarified, and some genes encoding enzymes and transcription factors are involved in the process. These genes included *1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-c-methyl-d-erythritol 4-phosphate cytidyl transferase (MCT), 2-c-methyl-d-erythritol 4-phosphate cytidyl transferase (CMK), 2-C-Methyl-D-erythritol2,4-cyclodiphosphate synthase (MDS), 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate synthase (HDS), 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR), isopentenyl diphosphate isomerase (IPI), geranylgeranyl diphosphate synthase (GGPPS), phytoene synthase (PSY), phytoene desaturase (PDS),* (*-carotene isomerase (ZISO)),* (*-carotene desaturase (ZDS)), lycopene-*(*-cyclase (LCY-E)), lycopene-*(*-cyclase (LCY-B)), LUTEIN DEFICIENT 5 (LUT5), LUTEIN DEFICIENT 1 (LUT1), violaxanthin de-epoxidase (ATVDE), zeaxanthin epoxidase (ZEP), carotene-*(*-hydroxylase (CHY-B)), carotene-*(*-hydroxylase (CHY-E)), violaxanthin de-epoxidase (VDE), Carotenoid Cleavage Dioxygenase (CCD), 9-cis-epoxycarotenoid dioxygenase (NCED),etc.* (Goodwin, 1971; Bramley, 1985; Vallabhaneni et al., 2010; Brandi et al., 2011; Fantini et al., 2013; Gao et al., 2021).

Flavonoids can produce the broadest spectrum of colors, ranging from pale yellow to blue-purple (Zhao and Tao, 2015). Anthocyanin is a type of flavonoid compound and is accumulated in different tissues of plants stimulated by drought, high light, and hormones (Zeng et al., 2010; Qi et al., 2011; Li et al., 2016). Anthocyanin is commonly found in various plants with widely biological functions, including protection against UV radiation, protecting leaf cells from photo-oxidative damage, and attracting pollinators (Winkel-Shirley, 2002; Bradshaw and Schemske, 2003; Falcone et al., 2012). Additionally, it can prevent cardiovascular disease, diabetes, aging, and cancer *via* scavenging ROS, bringing high health value to humans (He and Giusti, 2010; Li et al., 2022). After years of research by scientists, the synthetic pathway of flavonoids is also very clear,

which involves *phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumaroyl-CoA ligase (4CL), catalysis of chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), flavanone-3′-hydroxylase (F3′H), flavanone-3′5′-hydroxylase (F3′5′H), anthocyanin synthase (ANS), Flavonol synthase (FLS)*, *dihydroflavonol reductase (DFR), flavonoid-3-o-glucosyltransferase (UFGT), etc.* (Chen et al., 2016). In *Arabidopsis*, a ternary protein complex (MYB-bHLH-WD40) is considered to regulate the expression of these anthocyanin biosynthetic genes (Zheng et al., 2019), and the R2R3-MYB transcription factors include *production of anthocyanin pigment 1 (PAP1, MYB75)*, *production of anthocyanin pigment 2 (PAP2, MYB90)*, *MYB113*, *MYB114*, *transparent testa 8 (TT8)*, *enhancer of glabra (EGL3)*, and *transparent testa glabra 1 (TTG1)* (Walker et al., 1999; Nesi et al., 2000; Gonzalez et al., 2008; Jaakola, 2013).

Reactive oxygen species plays an essential role in diverse physiological processes. It exists in various forms in the aerobic environment, including superoxide, hydroxyl radicals, and other radical molecules (Mittler, 2017). In plants, ROS is produced mainly in chloroplasts, mitochondria, and peroxisomes continuously (Elstner, 1991; Rogers and Munné-Bosch, 2016). The ROS metabolism in plant cells is an intricate network involving many enzymatic and metabolite elements. It has dual roles in plants, promoting growth and development and playing a defensive role, but also producing toxic by-products of oxygen metabolism that lead to plant senescence (Mittler et al., 2004). In order to suppress the excessive accumulation of ROS, plants have evolved a variety of scavenging mechanisms, including a variety of ROS scavenging enzymes and non-enzymatic antioxidant compounds. For example, it was found that post-harvest UV-C exposure enhanced total anthocyanins and phenolic compounds in stored strawberries to scavenging the ROS, acting as a shield against UV-B (Raven, 2000; Severo et al., 2015; Xu Y. et al., 2017).

Hickory (*Carya cathayensis* Sarg.) is a deciduous tree of the genus *Carya* in the Juglandaceae family, endemic to China, mainly distributed in Tianmu mountainous area at the border of Zhejiang and Anhui provinces, integrating economic value and ecological value. Hickory is monoecious, with male flowers in catkins, inflorescence pendulous, and no tepal, and male flower buds usually begin to differentiate from mid to late May, then enter dormancy until late July (Huang et al., 2007, 2013). The female flowers show short spikes without tepal. Its buds begin to differentiate in early April, forming bracteoles and pistillate primordia, entering bud morphological differentiation in mid-April (Wang et al., 2012; Shen et al., 2014). Interesting, the hickory is a wind-pollinated plant, while the changing pattern of the color of stigma shows the characteristics of insect pollination, different from typical wind-pollinated plants. The color of pistil stigma varies markedly from opening to mature, in turn, presenting green, light red, bright red, purple-red, and, dark purple. During the pollination period, the stigma is bright red, bearing downy glands, and secretes mucus to facilitate the reception of pollen (Xia, 2006). Nevertheless, the reason for this color change in the hickory stigma is unclear. Whether this phenomenon is a protective mechanism of the plant itself against ROS accumulation or a feature of evolution is worth studying, which will be beneficial to solving the problem of the low yield of a hickory nut.

In this study, hickory stigma at different developmental periods offered ideal materials for revealing the molecular regulatory mechanism underlying the color changes of stigma and the biological significance of this phenomenon. By microplate assay, transcriptome sequence, and qRT-PCR, we were to: (1) measure the levels of chlorophylls, anthocyanins and, carotenoids of hickory stigma during the development; (2) determine the 1,1-Diphenyl-2-picrylhydrazyl radical 2,2-Diphenyl-1-(2,4,6-trinitrophenyl) hydrazyl (DPPH) and 2, 2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) radical scavenging capacity of stigma with development; (3) excavate the key differential genes with specific expression required for the color change and ROS metabolism during the hickory stigma development; (4) analyze the expression levels of genes required for pigment formation of stigma; (5) construct the gene co-regulatory networks between pigment and ROS metabolism. With these studies, the molecular mechanism of color formation and change in hickory stigma and its biological significance will be deeply understood, providing important theoretical guidance for the improvement of hickory nut yields in the new future.

MATERIALS AND METHODS

Plant Material

The stigmas of hickory were collected from the experimental base in Linglong Mountain, Lin'an, China (119°38′51″E, 30°12′39″N, elevation: 119 m). A total of six hickory trees with similar growth and good vigor were selected, and female flowers were collected every 2–3 days from late April, according to the developmental state. Also, 4 sampling time points were selected, noted as S1, S2, S3, and S4, respectively. The pistillate flowers were taken back to the laboratory immediately after collection. After the sepals were removed, the stigma and ovary parts were separated, placed in liquid nitrogen for freezing, and then stored at −80°C in the refrigerator for pigment determination and RNA extraction for transcriptome sequencing. Each stage (S1 to S4) had three biological replicates.

Total Chlorophyll and Carotenoid Content Measurement

The extraction and analysis of the total content of chlorophylls and carotenoids were determined according to the method described (Lichtenthaler, 1987; Xue et al., 2020). First, 20 mg of freeze-dried hickory stigma powder was added to 1,000 μl 95%(v/v) ethanol/water solution. After 24 h, it was centrifuged at 5,000 × *g* for 10 min, measured the absorbance at 652, 665, and 470 nm of the supernatants using an enzyme-labeled instrument (Tecan Spark®, Swiss). The following equations were used to calculate total chlorophyll and carotenoid contents:

Chl a (μg/g) = $(16.29A_{665}-8.54A_{652})*10$;

Chl b (μg/g) = $(30.66A_{652}-13.58A_{665})*10$;

Total Chl (μg/g) = Chl a + Chlb;

Total carotenoids (μg/g) = $(1000A_{470}-1.63$Chl a-104.96Chl b$)\div 221*10$

Total Anthocyanin Content Measurement

The extraction and analysis of the total content of anthocyanin were performed using the pH-differential method (Wrolstad and Giusti, 2001; Xue et al., 2020). First, 4 stages of the stigmas were placed in a lyophilizer (Christ Alpha 2–4 LD plus, Germany) for 48 h, and then grind thoroughly to powder. Next, 25 milligrams of freeze-dried hickory stigma powder were added to 250 μL extraction buffer (Trifluoroacetic acid: formic acid: water: ethanol = 1:2:27:70, v/v), and the mixture was stood 24 h at 4°C in the dark. Following centrifugation at 12,000 rpm for 10 min, the supernatant was transferred to a fresh tube. Next, 100 μL supernatant was mixed with 900 μL.025 mol potassium chloride buffer (pH = 1.0), and another 100 μL supernatant was mixed with 900 μL.4 mol sodium acetate buffer (pH = 4.5). The mixture was stood for 1.5 h (complete sedimentation) at room temperature. The absorption of the mixture was recorded at 510 and 700 nm using the enzyme-labeled instrument with 3 repetitions. Finally, total anthocyanin content was calculated based on the molar absorbance of a cyanidin-3-glucoside standard by the equation:

total anthocyanins (μg/g) = (A × MW × DF × 10000)/(ε × L), where $A = (A_{510} - A_{700})_{pH1.0} - (A_{510} - A_{700})_{pH4.5}$, ε = 26900 L·mol^{-1}cm^{-1}(extinction coefficient of cyanidin-3-glucoside at 510 nm), L = path length, MW = molecular weight of cyanidin-3-glucoside, DF = dilution factor.

Antioxidant Capacity Measurement

In this experiment, 300 ul anhydrous ethanol was added to a 30 mg sample, sonicated for 3 h, and then macerated overnight. The extract was centrifuged at 8000 rpm for 10 min, and the supernatant was prepared at a concentration of 100 mg/ml (Burits and Bucar, 2000).

The DPPH radical scavenging capacity of hickory female stigma was determined by referring to the method with minor modifications (Burits and Bucar, 2000). The extracts were diluted to 0.25, 0.5, 1, 1.5, and 2 mg/ml with anhydrous ethanol and prepared for use later. Took 500 ul of the diluted extract in a test tube, added 500 ul of.2 mM DPPH free radical ethanol solution, reacted at room temperature, avoided light for 30 min, and then measured the absorbance at 517 nm using an enzyme marker. The absorbance Ac of 500 ul of DPPH free radical ethanol solution mixed with 500 ul of ethanol was measured simultaneously. The IC50 (concentration of samples scavenging 50% of DPPH free radicals) was used to evaluate the magnitude of antioxidant activity of hickory female flower stigma extracts from four periods. The lower the IC50 value, the higher the antioxidant activity.

Clearance rate (%) = $(A_c - A_s)/A_c \times 100\%$

The ABTS radical scavenging capacity of hickory female flower stigma was determined according to the method (Petretto et al., 2016). First, ABTS radical stock solution was prepared by mixing 7 mmol/L ABTS solution with 2.45 mmol/L potassium persulfate at a 1:1 ratio by volume and left overnight. The prepared ABTS radical solution was diluted with anhydrous ethanol to produce an absorbance value of 0.7 ± 0.02 at 734 nm. Then, 50 μL of the extract dilutions (0.25, 0.5, 1, 1.5, and 2 mg/mL) were mixed with 1 ml of ABTS radical solution, and the absorbance values were measured at 734 nm after a 30-min reaction at room temperature and protected from light.

Clearance rate (%) = $(A_c - A_s)/A_c \times 100\%$

Ac is the absorbance value of anhydrous ethanol mixed with ABTS radical solution, and As is the absorbance value of the extract mixed with the solution with ABTS radical.

Total RNA Extraction, mRNA Library Construction, and Sequencing

Samples were immediately frozen in liquid nitrogen and stored at –80°C for RNA-sequence analysis. Total RNA was extracted from female stigma at different developmental time points using CTAB-PBIOZOL reagent and ethanol precipitation. After extracting, the concentration, 28S/18S and RIN/RQN of total RNA were detected by Agilent 2100, and OD260/280 and OD260/230 were quantified by the NanoDrop. The construction of cDNA libraries and transcriptome sequencing were completed by BGI Technology (Shenzhen, China). First, total RNA was disposed of by mRNA enrichment using Oligo (dT)-attached magnetic beads. The mRNA was further broken using the fragment buffer after ligating sequence adapters. PCR amplification and cyclizing were used to get a single-strand cDNA library. The DNBSEQ platform performed the sequencing after the library was certified. All of the experiments were performed with three replicates. Raw reads were obtained from the sequencing, and the low-quality reads with adaptors were processed to obtain clean reads.

RNA-Seq Data Analysis

Raw reads obtained were filtered with SOAPnuke (version 1.4) (Li et al., 2008) to filter out low-quality reads, contaminated junctions, and high levels of unknown base N. The clean reads were obtained and stored in FASTQ format, then aligned to the reference genome (hickory genome), using Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT2 version 2.1) (Kim et al., 2015), followed by Bowtie2 (version 2.2.5) to align the clean reads to the gene set (Langmead and Salzberg, 2012). RSEM was used to calculate gene expression levels for individual samples (v1.2.8) (Li and Dewey, 2011). Once the new transcripts are obtained, we added the new transcripts with potential protein-coding to the reference gene sequence to form a complete reference sequence and then calculate gene expression levels. Differential gene expression analysis was performed by DESeq2 (Love et al., 2014) with a Q-value < 0.05. Finally, multiple samples are tested for differentially expressed genes as required, and in-depth clustering and functional enrichment analyses of differentially expressed genes (DEGs) are performed.

Identification and Analysis of Differentially Expressed Genes

The assembled unigenes were functionally annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG)[1] and Gene Ontology (GO) databases. The Blast2GO program provided GO

[1]https://www.kegg.jp/

annotation. WEGO then generated GO classification maps. The biological processes and unigenes annotations of the pathway were analyzed using KEGG. The KEGG comprehensive database resource was compared with our data for analysis. DEGs were identified using the NOISeq method (Tarazona et al., 2015) based on the gene expression level of each sample with a fold change of at least 2, *Q*-value < 0.01, then were analyzed using RNAseq (3 biological replicates per group), characterized by gene ontology enrichment analysis.

Transcription Factor Prediction and Co-expression Network Construction

To characterize patterns across samples and identify interesting and highly covariant sets of genes, we constructed a gene co-expression network based on symbolic hybrid network types using the weighted correlation network analysis (WGCNA) package (Langfelder and Horvath, 2008; Linn et al., 2017). The WGCNA package was also used to construct gene co-expression networks from identified DEGs to establish the transcriptional regulatory structure of anthocyanin biosynthesis, carotenoid biosynthesis, chlorophyll metabolism, and ROS metabolism. The network was visualized using Cytoscape software (Smoot et al., 2011).

QRT-PCR Analysis of Genes Involved in Selected Genes

Totals of 21, 18, and 21 unigenes related to chlorophyll, carotenoid, and anthocyanin metabolism were selected for qRT-PCR analysis. Total RNA was isolated from the stigma collected at four different developmental time points as previously described, and then transcribed 1000 ng of total RNA into cDNA for qRT-PCR using PrimeScript ^{TM} RT reagent Kit with gDNA Eraser (removing genomic DNA) (Perfect Real Time) (TaKaRa, Japan) according to the instructions. The templet cDNA was diluted into 100 ng/μL (± 3 ng/μL). NCBI primer blast was used to design PCR primers to quantify gene expression involved in pigments formation, shown in **Supplementary Table 1**. The reaction system for RT-qPCR is as follow: ddH_2O 3.8 μL, TB Green 5 μL, primerF.2 μL, primerR.2 μL, cDNA0.8 μL, with 3 repetitions. All qRT-PCR assays were performed using TB Green® Premix Ex Taq^{TM} (Tli RNaseH Plus) (TaKaRa, Japan) in a CFX96 Touch^{TM} Real-Time PCR Detection System (BIO-RAD, United States) with the following reaction conditions: 95°C for the 30 s and 40 cycles of amplification (95°C for 5 s, 60°C for 30 s). The relative expression levels of target genes were calculated using the $2^{-\Delta\Delta Ct}$ method against the internal control (Livak and Schmittgen, 2001), and the *CcUBC9-5* gene was used as a control to normalize the relative expression levels of target genes. Experiments were performed with three independent biological replicates and three technical replicates.

Analysis of Protein Physicochemical Properties and Prediction of Subcellular Localization

The protein physicochemical properties, including protein length, isoelectric points (PI), molecular weight (MW), and grandaverage of hydropathicity (GRAVE), were predicted using ProtParam[2]. The ProtScale[3], SignalP3[4], and Plant-mPLoc[5] were used to predict protein hydrophobic/hydrophilic, signal peptides, and subcellular localization, respectively.

Statistical Analysis

Statistical analysis was carried out with one-way ANOVA in SPSS17 software (ANCOVA; SPSS17, SPSS Inc., Chicago IL, United States). Significant differences among treatments were obtained according to *P*-values determined by Tukey's honestly significant difference (Tukey's HSD) with HSD ($P < 0.05$). All the experiments were repeated 3 times, and the results were presented as mean ± standard deviation (*SD*).

RESULTS

Dynamic Changes in the Contents of Total Anthocyanin, Chlorophyll, and Carotenoid

During the development of the hickory flower, the stigma color changed obviously from green to dark-purple (S1: green, S2: light-purple red, S3: dark-purple red, S4: dark purple) (**Figure 1A**). To clarify the pigment components and their dynamic change, total chlorophyll, carotenoid, and anthocyanin at four stages of stigma development were further measured. It was shown that total chlorophyll content was relatively high overall and showed a continuous slight downtrend with stigma development (**Figure 1B**), while total carotenoid content was increased from S1 to S3 and reached the highest value of 36.4 (ug/g, while decreased gradually from S3 to S4 (**Figure 1C**). Total anthocyanin content gradually increased during the four periods of stigma development, sharped reaching the highest content (131.4 ug/g, with cyanidin-3-glucoside used as the standard) at the S4 (**Figure 1D**). These results suggest that the color change of the stigma of hickory flowers is due to the joint action of these pigments. The green color at the early stage is mainly caused by chlorophyll. In the middle and late stages, the accumulation of carotenoid and anthocyanin gradually makes the stigma appear red and purple.

Differences in the Scavenging Capacity of Stigma for DPPH and ABTS Radicals

It is well-known that anthocyanins have a strong antioxidant capacity (Raven, 2000; Severo et al., 2015). The sharper changes of anthocyanins contents than the other two pigments from S3 to S4 inspire us to measure the ROS scavenging capacities. To explore the antioxidant capacity of stigma, removal-capacities of DPPH and ABTS radicals were further measured. It was found that the scavenging ability of DPPH radicals gradually increased with the process of stigma development and could reach 84.14%

[2]https://web.expasy.org/protparam/

[3]https://web.expasy.org/protscale/

[4]https://services.healthtech.dtu.dk/service.php?SignalP-3.0

[5]http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/

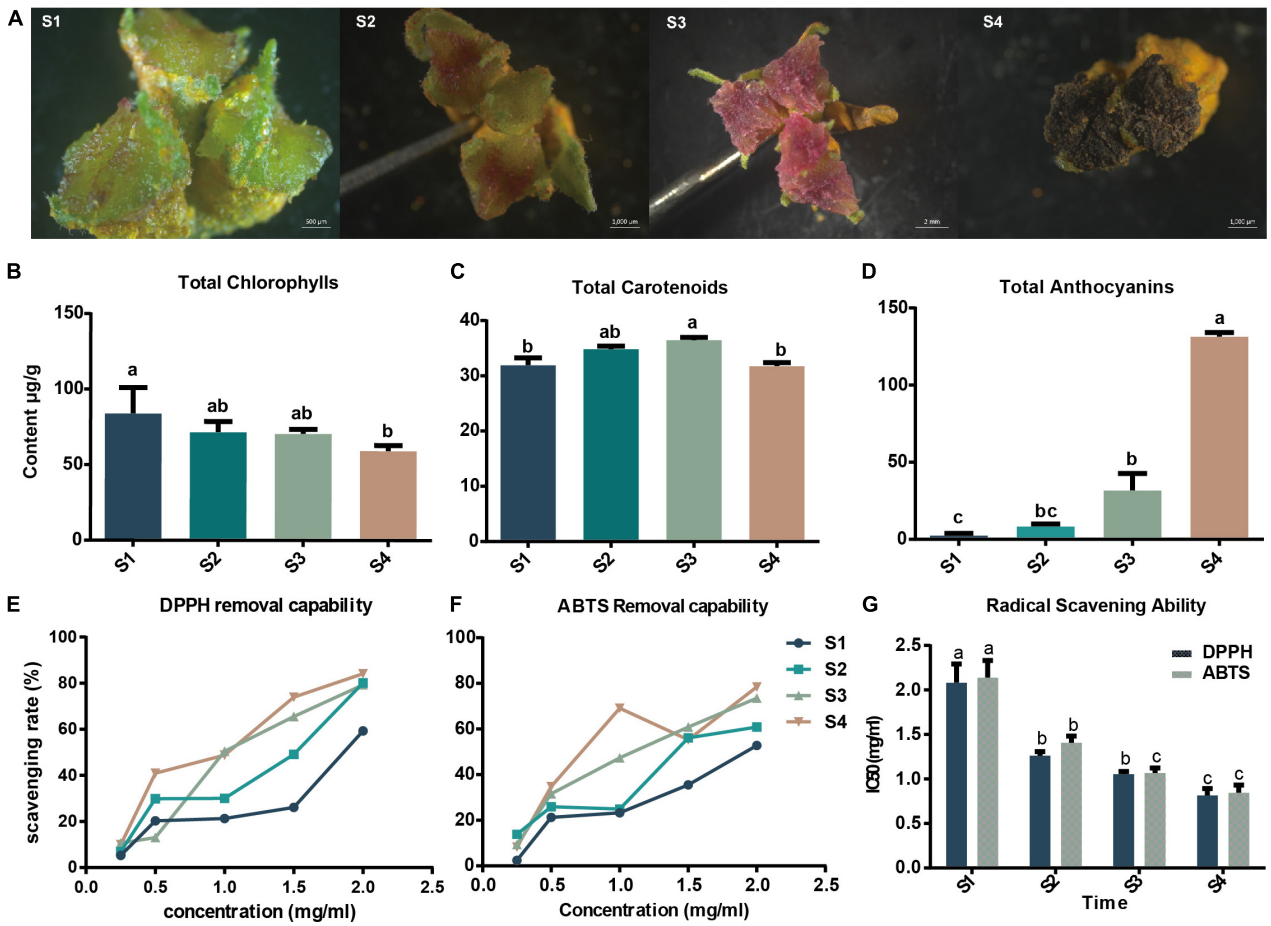


FIGURE 1 | Pigmentation and phenotypes of hickory stigmas at the four flowering stages. **(A)** Morphological characteristics of hickory stigma during flower development (S1-S4). **(B–D)** Dynamic changes in the content of total chlorophylls, carotenoids, and total anthocyanins (cyanidin-3-glucoside used as the standard), expressed as μg/g dry weight (DW). **(E–F)** Differences in DPPH and ABTS radical scavenging ability in different concentrations of sample extracts expressed as mg/ml. **(G)** Differences in DPPH and ABTS scavenging capacity at the 4 different stages. Results are mean ± *SD* of three parallel measurements. Different letters correspond to significant differences ($P < 0.05$, Tukey's HSD) by ANOVA.

at the S4 period (**Figure 1E**). Similarly, the change of ABTS free radical scavenging capacity was close to that of DPPH, with 78.46% scavenging capacity at S4 (**Figure 1F**). It was also observed that DPPH and ABTS radical scavenging rates gradually increased with the increasing concentration of the hickory stigma extracts (**Figures 1E,F**). DPPH and ABTS IC50 could also be used to describe the oxidative scavenging capacity, the extracts from the stigma samples at S4 were more active in scavenging DPPH radicals, and their IC50s were significantly ($P < 0.05$) lower than those of the other three periods (S1, S2, and S3) (**Figure 1G**). The oxidative scavenging capacity obtained by the ABTS IC50 showed similar results. The highest clearance activity was observed during the S4 period and the lowest during the S1 period (**Figure 1G**).

Overview of the Transcriptomic Analysis

To obtain a global transcriptome profile of developing stigma, twelve stigma samples at S1, S2, S3, and S4 were sequenced using the BGIseq500 platform. After removing adapter pollution, poly-N sequences, and low-quality reads, totals of 75.72 Gb clean bases were acquired, including an average of 45.24 Mb raw sequencing reads and 42.06 Mb clean reads, in each sample. The Q30 base percentage was greater than 90.78% and the average ratio of clean reads to raw reads was 92.97%. Using HISAT and Bowtie2, 93.43–94.53% of the clean reads could be mapped to the genome of hickory, 56.9–62.31% could be mapped to the genome of hickory uniquely. 67.88–71.17% (total), 52.83–54.87% (unique), and the clean reads were mapped to the hickory genes per sample, so that 31,027 genes were identified (**Supplementary Table 2**). Principal components analysis (PCA) showed that the sample clusters with high similarity converged, and the 4 sampling time points were separated without an outlier sample (**Supplementary Figure 1**). Based on the expression levels of all genes in every two samples, the Pearson correlation coefficient (PCC) was calculated and the value was at least 0.94, suggesting that the repeatability of the sample is very good (**Figure 2A**). Moreover, Hierarchical cluster analysis (HCA) was

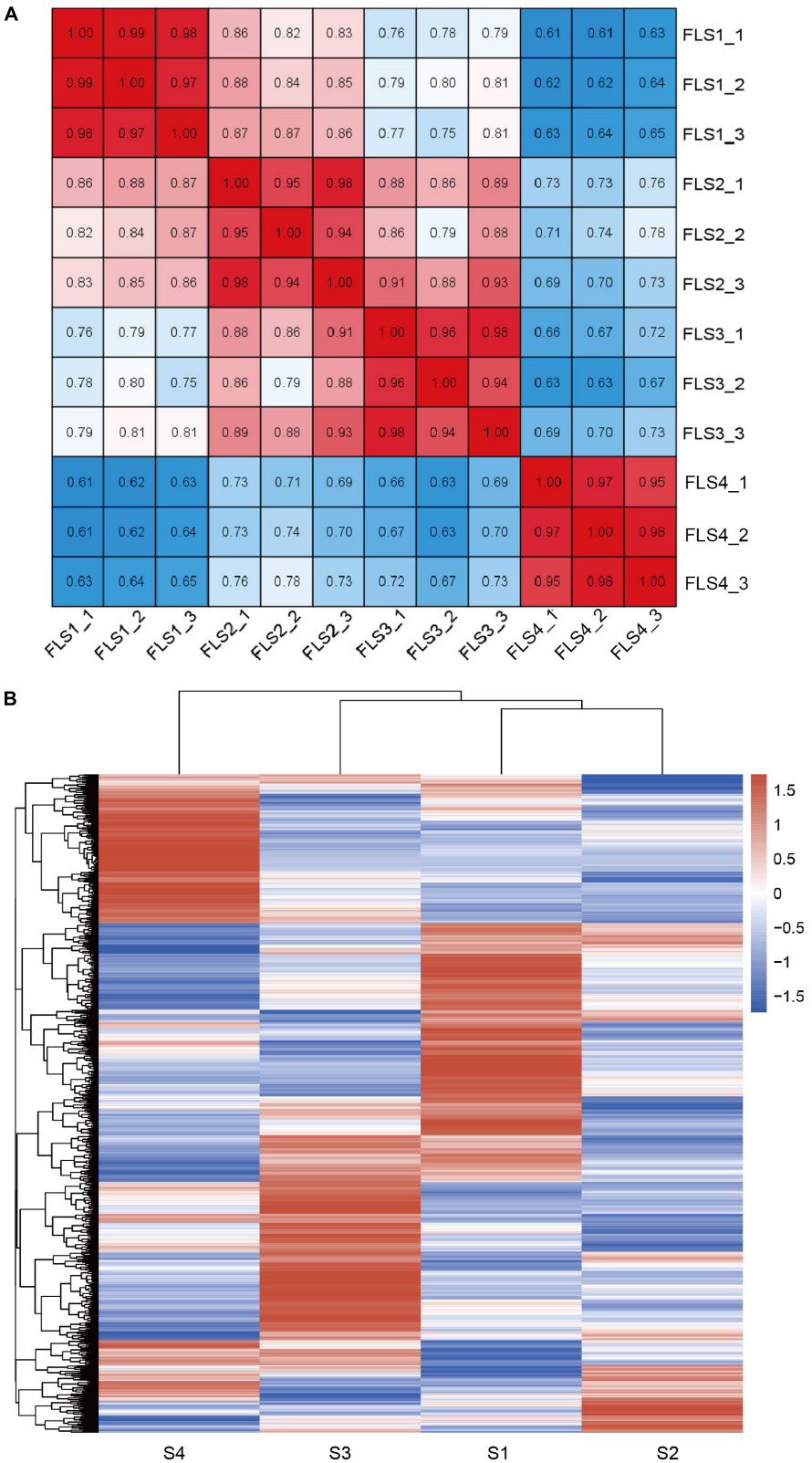


FIGURE 2 | Basic survey of stigma transcriptome during flower development of hickory. **(A)** Pearson correlation results are based on all transcripts identified from RNA-seq. **(B)** Heatmaps of all genes by hierarchical cluster analysis (HCA) from RNA-seq.

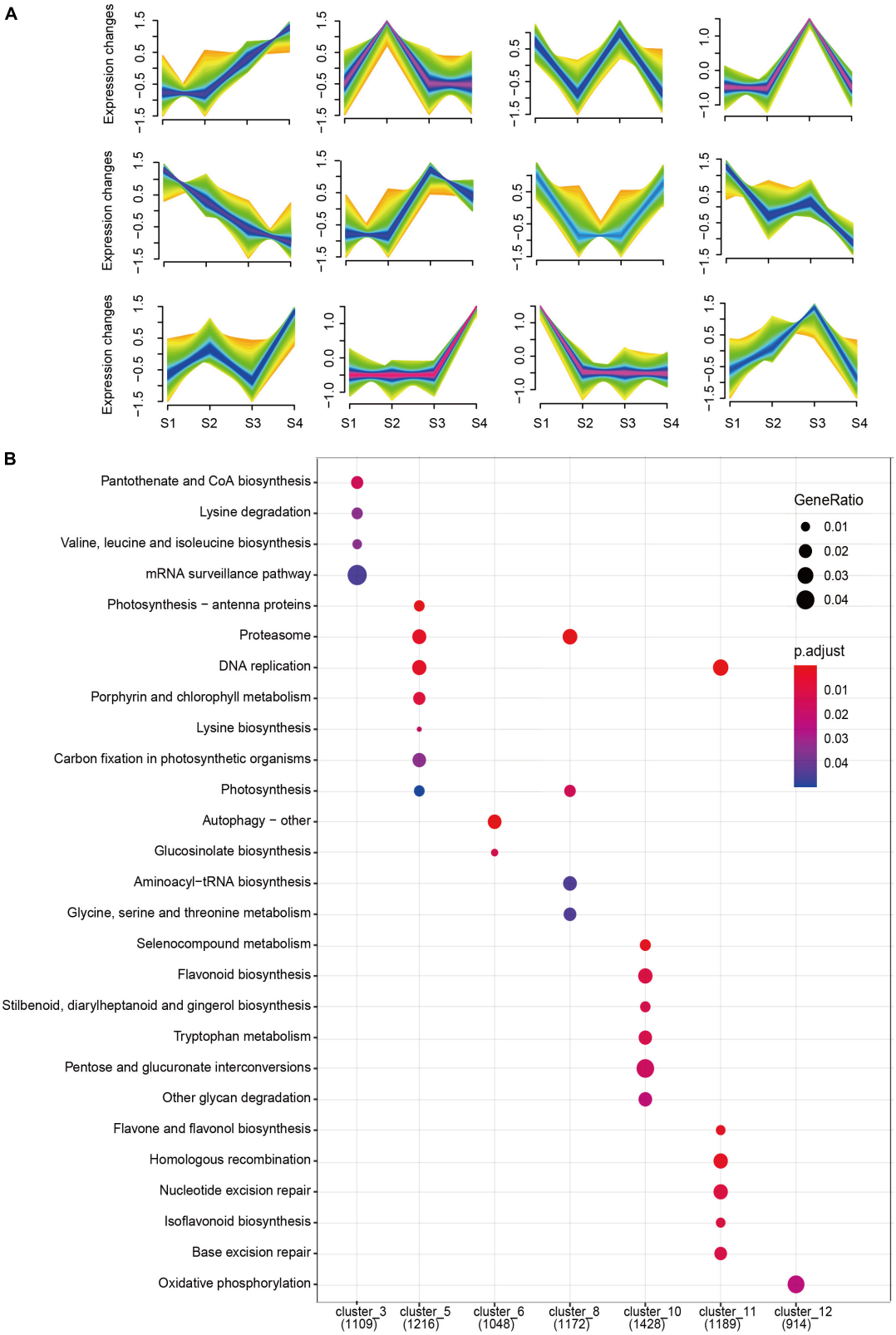


FIGURE 3 | Analysis of 12 time-series clusters during stigma development of hickory. **(A)** Clustering analysis of gene expression patterns at different time points using fuzzy c-means algorithm (Mfuzz). **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of 12 clusters.

accompanied and displayed in the form of heatmaps plotted by a heatmap package in R software (**Figure 2B**), showing that significant differences in the expression levels of genes occurred between the sample groups. Using Mfuzz, twelve time-related gene clusters were obtained based on the similar expression patterns of some genes during stigma development (**Figure 3A** and **Supplementary Table 3**). KEGG analysis showed that some genes in cluster 5 were associated with "photosynthesis-antenna proteins," "porphyrin and chlorophyll metabolism," "carbon fixation in a photosynthetic organism," "photosynthesis," and "carotenoids metabolism," with gradual downtrends from S1 to S4 (**Figure 3B** and **Supplementary Table 4**). These results indicated that photosynthesis mainly dominated in the pre-developmental period. Furthermore, the "Flavonoid biosynthesis" pathway was enriched in cluster 10, with a smooth trend in the pre-period, and increased significantly from S3 to S4 (**Figure 3B** and **Supplementary Table 4**). Other flavonoids, such as flavone, flavonol, and isoflavonoid, were found to be enriched in cluster 11, with a drastic reduction tendency from S1 to S2 but a slight change from S2 to S4 (**Figure 3B** and **Supplementary Table 4**). Notably, it was also found that oxidative phosphorylation-related genes were clustered in cluster 12, which showed a trend of increasing first from S1to S3 and then decreasing from S3 to S4 (**Figure 3B** and **Supplementary Table 4**).

Based on the selected threshold values of fold change with at least 2 (Q-value < 0.01, P-value of the cutoff standard < 0.05), a total of 10,432 DEGs were screened out in all the pair-wise comparison groups. In these DEGs, 2,604 (S2 vs. S1), 2,780 (S3 vs. S2) and 4,955 (S4 vs. S3), 4722(S3 vs. S1), and 6915 (S4 vs. S1) DEGs were counted out, including 1,801, 1026, 3011, 2383 and 3795 upregulated genes and 803, 1,754, 1,944, 2339 and 3120 downregulated genes, respectively (**Figure 4A**). In addition, Venn diagrams were mapped to figure out the relationship between the DEGs in the above comparison groups, indicating that a total of 185 DEGs existed in all five comparison groups (**Figure 4B**). These results suggested that there were dynamic changes in identified genes during stigma development and the active transcription occurred during the late period. KEGG analysis indicated that all identified DEGs were enriched in 15 pathways (**Supplementary Table 5**). It was also found that these DEGs enriched in the "Flavonoid biosynthesis" pathway (S4 vs. S1, S4 vs. S3, ko00941), "Photosynthesis-antenna proteins" (S3 vs. S1, S3 vs. S2, and S4 vs. S1, ko00196) pathways, "Photosynthesis" (S4 vs. S1, ko00195) pathway, "anthocyanin biosynthesis pathway" (S3 vs. S1, and S4 vs. S1, ko00942), "carotenoid biosynthesis pathways" (S4 vs. S1, ko00906) and ABC transporters (S4 vs. S3, ko02010) (**Figure 4C** and **Supplementary Table 5**). Furthermore, volcanic maps in S3 vs. S1, S3 vs. S2, S4 vs. S1, and S4 vs. S3 were drawn, intuitively showing the up-regulated genes (blue dots), down-regulated genes (red genes), and non-differential genes (green dots) (**Figures 4D-G**). These results indicated that the genetic manifestations of these pathways were active at the middle and late stages of female stigma development, offering a clear direction for elucidating the internal molecule mechanisms of the color formation and dynamic changes.

Identification and Verification of Genes in Pigment Metabolism Pathways and Their Expression Analysis

The above results indicated that chlorophyll, carotenoid, and flavonoid, including anthocyanins, may be involved in color formation. According to the whole genomic information of the hickory and KEGG database, the metabolism pathways of chlorophyll, carotenoid, and flavonoid, including anthocyanins, were drawn and genes encoding enzymes and transcription factors were annotated in these pathways (**Figures 5–7** and **Supplementary Table 6**). Flower color-related genes with critical roles were further analyzed and 91 genes were screened out. Among these genes, 27 genes are required for chlorophyll metabolism, 28 genes were related to carotenoid metabolism and 36 genes were associated with anthocyanins metabolism. Among these genes, 21, 19, 22, and 18 from S1, S2, S3, and S4, respectively, had FPKM values of over 100. In the chlorophyll metabolic pathways, some genes, such as *GSA, HEMB, HEMD, HEMF1/2, CLH1/2, FCC,* and *NYE1,* were upregulated, while *HEMA, HEMC, HEME, HEMG, CHLD, CHLH, CHLI, CRD, DVR, POR, CAO, CHLG, NYC1,* and *PAO,* were downregulated (**Figure 5** and **Supplementary Table 6**). In the carotenoid biosynthesis pathway, some genes, including *DXS, MCT, CMK, MDS, HDS, GGPPS*, and the downregulated genes *DXR, PSY, PDS, ZISO, ZDS, LCY-E, LCY-B, LUT5, LUT1, ATVDE,* and *ZEP,* were shown to be upregulated (**Figure 6** and **Supplementary Table 6**).

Based on the results showing that there were greater changes in anthocyanin level and strong antioxidant capacity, a detailed analysis of the expression patterns of anthocyanin-related genes was conducted according to the data in **Supplementary Table 6**. Through the analysis of the synthesis pathway, we found that the expression patterns of most genes, including *PAL, C4H, 4CL1/5, 4CL2, 4CL3, CHS, F3H, F3′H, F3′5 ′H, DFR, ANS*, *UF3GTs, UGT, UGT73B1/B2/B3/B4/5,* and *UGT88A1,* were basically consistent, showing upward trends from S3 to S4, contributing to rapid accumulation of anthocyanins at a late stage of stigma development. It was also observed that the *FLS* gene required for quercetin-Der, Kaempferol-Der, and myricetin-Der biosynthesis showed the opposite trends as the above-described genes, suggesting that more premise substances enter into the leucocyanidin, leucopelargonidin, and leucodelphin for anthocyanins accumulation. After the anthocyanins biosynthesis is finished, they will be transported into the vacuole, involving some transporter proteins. It could be seen that the expression trends of *TT2* coding transporter proteins were similar to that of anthocyanins biosynthesis genes, while others, including *TT19, MRPs,* and *AHA10,* showed the opposite trends from S3 to S4. This is an indication that *TT2* may be the main transporter for anthocyanins in hickory stigma. In addition, previous studies have reported that the anthocyanin metabolism pathway was regulated by a suite of transcription factors that include MYB, bHLH, and WD-repeat proteins (Walker et al., 1999; Nesi et al., 2000; Gonzalez et al., 2008; Jaakola, 2013). It was observed that some TFs, including *TT2*, *TT1,* and *TTG1,* showed upward trends from S3 to S4, while others like *MYB75*, *MYB11*, *MYB12*, *MYC1*,

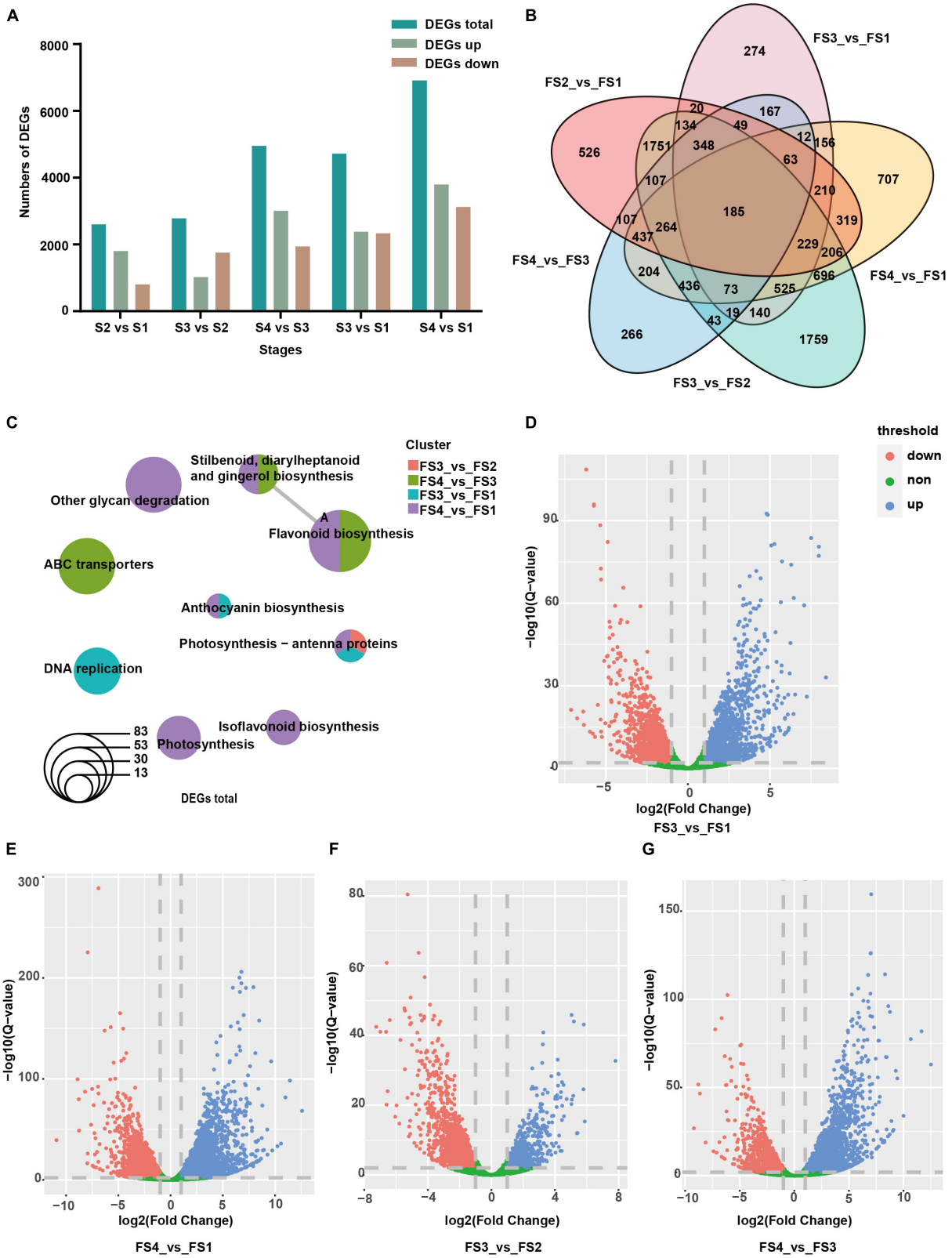


FIGURE 4 | (A) Overview of DEGs and their pathway enrichment based on KEGG analysis. Numbers of differentially expressed genes (DEGs) in five comparisons; **(B)** Venn diagram of the number of DEGs revealed by five comparisons; **(C)** KEGG enrichment of all DEGs; **(D–G)** Volcanic maps for DEGs in four comparisons enriched by KEGG. Red dots, downregulated genes. Blue dots, upregulated genes. Green dots, non-differential genes.

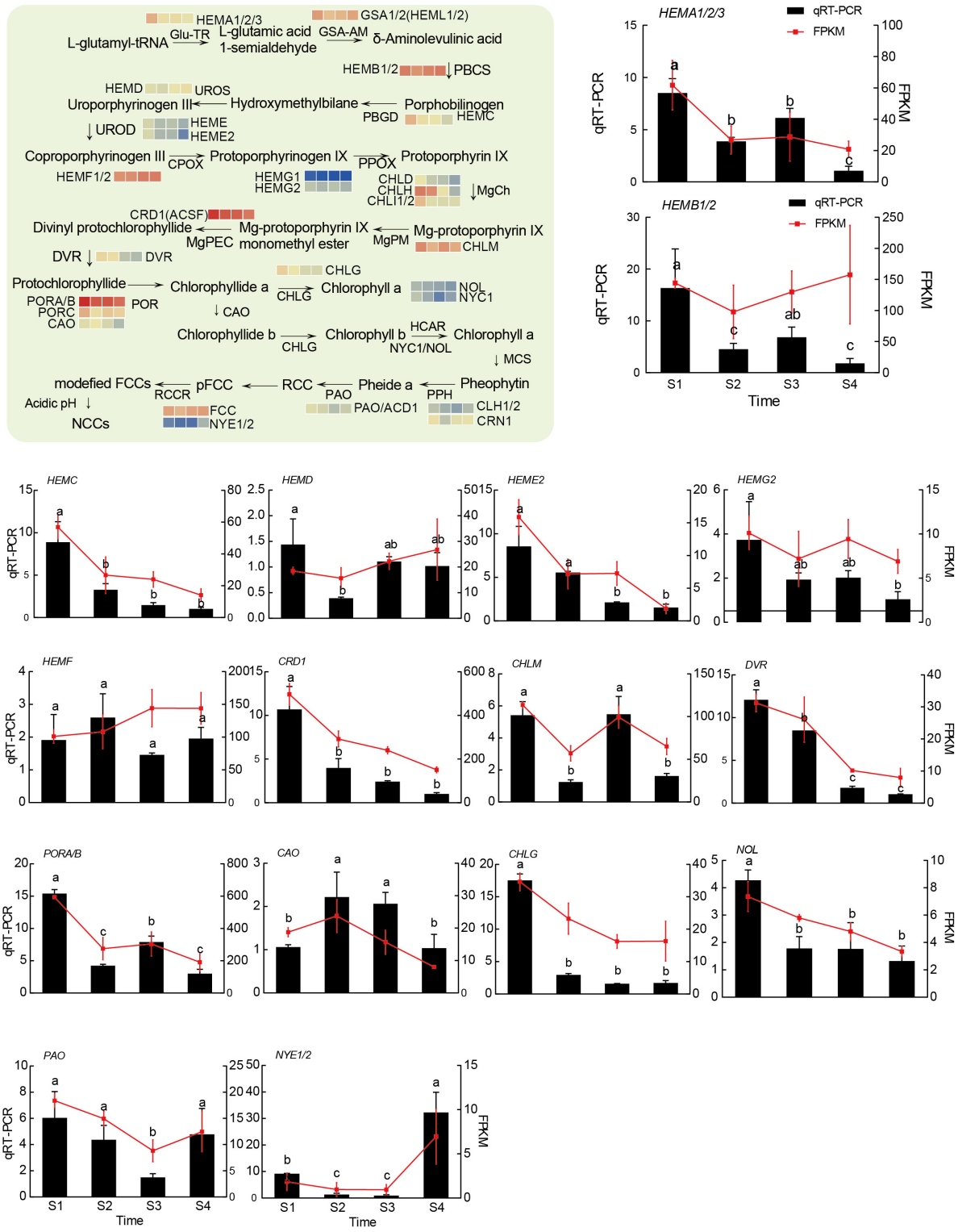


FIGURE 5 | The model diagram of the main chlorophyll biosynthesis pathway, and transcriptional analysis of and qRT-PCR assays of 16 genes related to chlorophyll biosynthesis. R-package heatmaps were used to plot the expression heatmap of these genes. The heatmap data were homogenized by log2 (FPKM). The relative expression levels of target genes were calculated based on the $2^{-\Delta\Delta Ct}$ method against the internal control, and the *CcUBC9-5* gene was used as a control. Experiments were performed with three independent biological replicates and three technical replicates. Different letters showed significant differences ($P < 0.05$, Tukey's HSD) by ANOVA among groups.

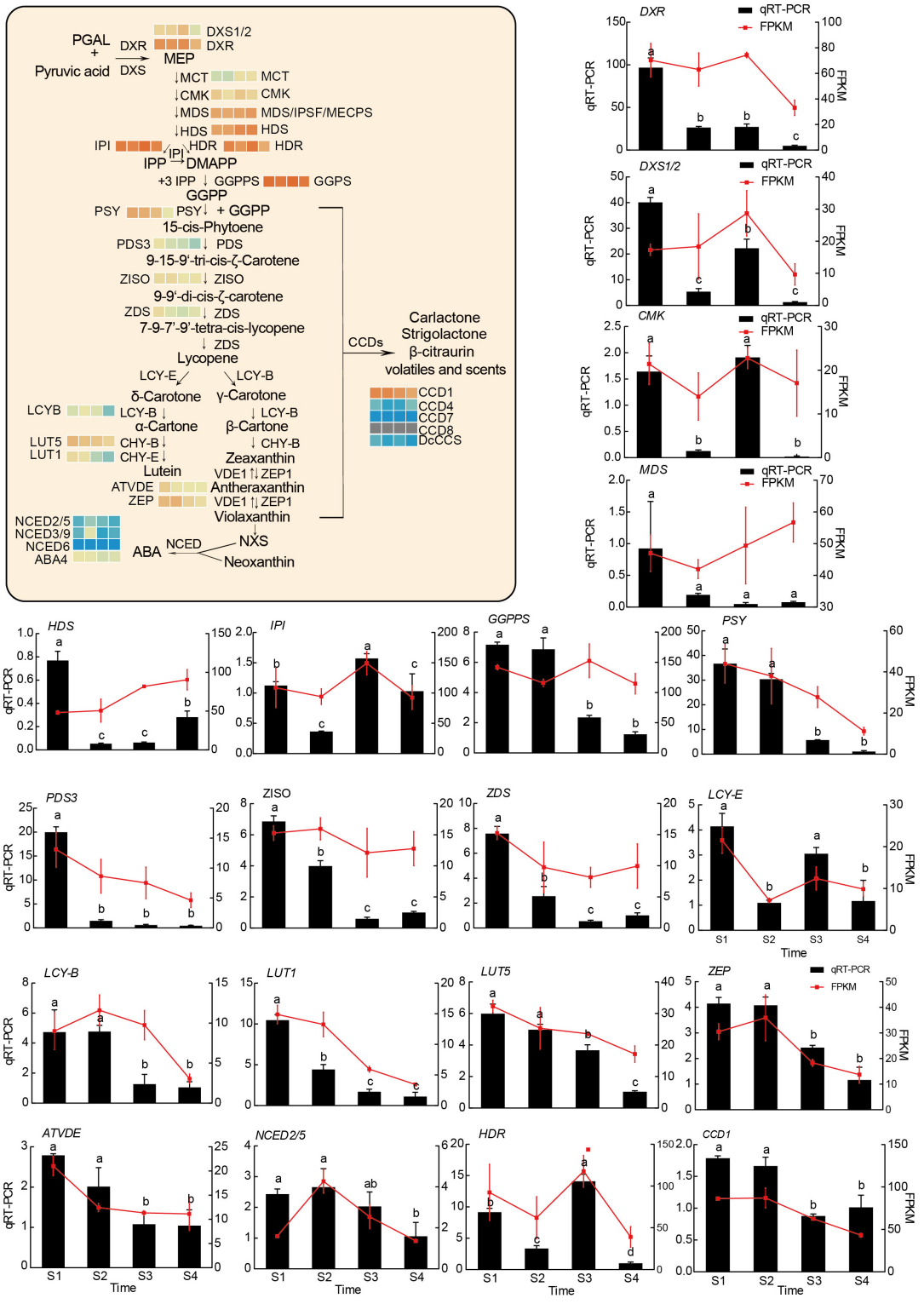


FIGURE 6 | The model diagram of the main carotenoid metabolism pathway, and transcriptional analysis of and qRT-PCR assays of 20 genes associated with carotenoid metabolism. R-package heatmaps were applied to map the expression heatmap of these genes. The heatmap data were homogenized by log2 (FPKM). The relative expression levels of target genes were calculated by the $2^{-\Delta\Delta Ct}$ method against the internal control (*CcUBC9-5*). Experiments were conducted with three independent biological replicates and three technical replicates. Different letters indicated significant differences ($P < 0.05$, Tukey's HSD) by ANOVA among groups.

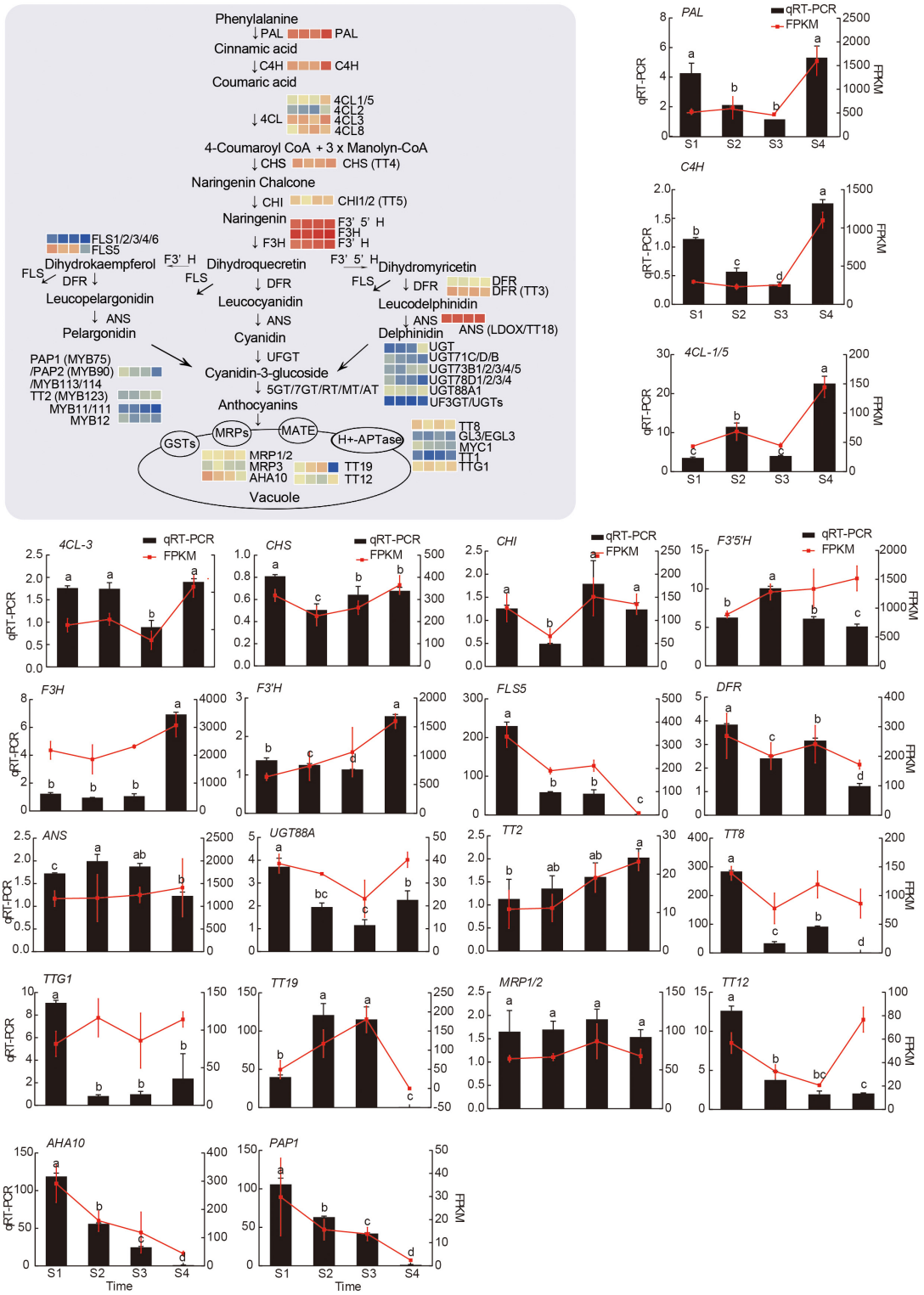


FIGURE 7 | The model diagram of the main anthocyanin biosynthesis pathway, and transcriptional analysis of and qRT-PCR assays of 21 genes associated with anthocyanins metabolism. R-package heatmaps were applied to map the expression heatmap of these genes. The heatmap data were homogenized by log2 (FPKM). The relative expression levels of target genes were calculated using the $2^{-\Delta\Delta Ct}$ method and the *CcUBC9-5* was taken as an internal control. Experiments were carried out with three independent biological replicates and three technical replicates. Different letters represented significant differences ($P < 0.05$, Tukey's HSD) by ANOVA among groups.

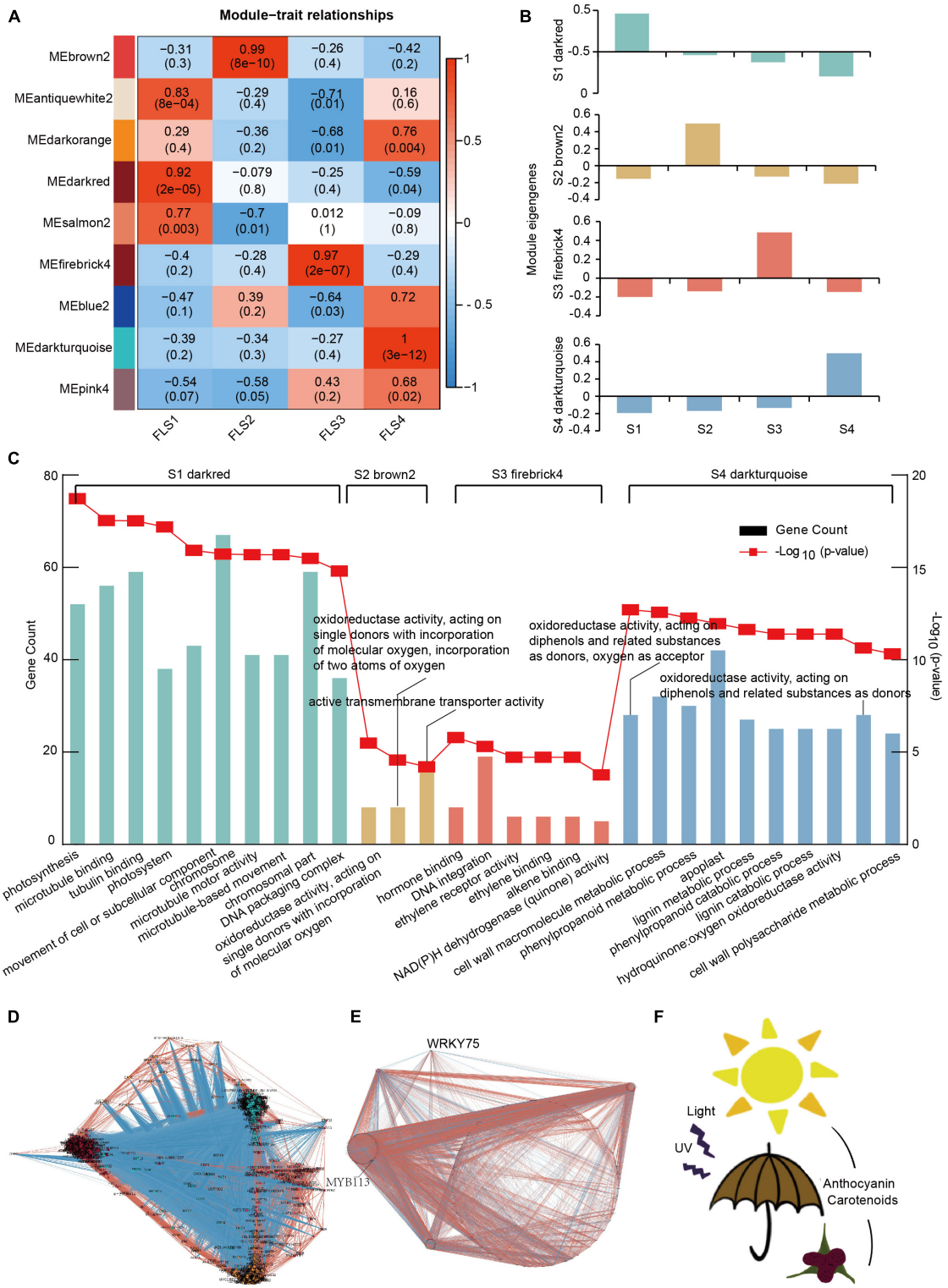


FIGURE 8 | Establishment of WGCNA modules of the differentially expressed genes (DEGs) at the four stages of hickory stigma. **(A)** Module-trait correlations and DEGs network analysis established by WGCNA. Colors indicated the strength and direction of the correlation according to the color legend at the right, and the color scale represented the module-trait correlations from −1 to 1. The numbers in parentheses are the partial Pearson correlation coefficients and the corresponding *P*-values. **(B)** Module eigengenes of four hub modules. **(C)** GO analysis of four hub modules (darkred, brown2, firebrick4, and dark turquoise). **(D)** co-expression networks of anthocyanin, chlorophyll, carotenoid, ROS, and transcription factors in four modules. **(E)** co-expression networks of *MYB113 and WRKY75* and pigments, ROS, TFs with the FPKM more than 10 at any stage. **(F)** The working module of hickory stigmas is photo protected by forming pigments, such as anthocyanins, to scavenge reactive oxygen species.

TABLE 1 | Physicochemical properties of proteins of *WRKY75* and *MYB113* in hickory.

Gene	Gene ID	Length	PI	MW (kDa)	GRAVE	localization	Hydrophobic/hydrophilic	Signal peptides
WRKY75	CCA0573S0068	152	9.64	17.64	−1.156	nucleus	Hydrophilic	No
MYB113	CCA0887S0030	264	9.1	30.21	−0.536	nucleus	Hydrophilic	No

TT8, and *GL3* were on the contrary. This may be attributed to the fact that these TFs usually form protein complexes to function. To regulate the expression of anthocyanin-related genes, these TFs need to be transcribed earlier. When the data were homogenized *via* Z-score normalization, the trends in gene expression became clearer (**Supplementary Figure 3**). Together, the expression trends of these pigment-related genes provided plausible explanations for the dynamic change of hickory stigma color with development.

To validate the results of RNA-seq, 20, 16, and 21 unigenes related to chlorophyll metabolism, carotenoid biosynthesis, and anthocyanin biosynthesis were selected for qRT-PCR analysis. As was shown in **Figures 5–7**, the results of the qRT-PCR were similar to the RNA-seq analysis, showing a high correlation, except for minor differences in some genes. These results further confirmed the reliability of the RNA-seq data in this study.

Establishment and Analysis of Weighted Correlation Network Analysis Modules

To gain insight into the patterns of genetic association between different samples and further study the high covariate gene sets associated with pigment and ROS metabolism, we used WGCNA to construct a co-expressed gene network based on 10,432 DEGs (**Supplementary Table 7**). After removing the low expression genes, nine different gene modules were obtained by HCA and were demonstrated in different colors. Among nine modules, four significant modules were found through analyzing the relationship between modules and traits (**Figure 8A**). The darkred-, brown2-, firebrick4- and dark turquoise-modules were significantly correlated with the trait at S1 (PCC = 0.92, P = $2^{*}e^{-5}$), S2 (PCC = 0.99, P = $8^{*}e^{-10}$), S3 (PCC = 0.97, P = $2^{*}e^{-7}$), and S4 (PCC = 1, P = $3^{*}e^{-12}$), respectively. **Figure 8B** further showed the module eigengene in the four modules, which indicates these four modules were strongly interrelated in the four stages (**Figure 8B**). This result illustrated that the DEGs in the darkred, brown2, firebrick4, and the dark turquoise module were more important to correlate with the 4 stages during hickory stigma development. In the 4 hub modules, 17 chlorophyll-related DEGs and 12 carotenoid-related DEGs were mostly contained in the darkred module, while most anthocyanin biosynthesis-related DEGs were distributed in the dark turquoise module, with 112 DEGs (**Supplementary Table 8**). GO analysis further showed the "photosynthesis" and "photosystem" were enriched in the darkred module, which was closely related to S1 (**Figure 8C** and **Supplementary Table 9**). Notably, "oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, GO:0016701," "oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen, GO:0016702," and "active transmembrane transporter activity, GO:0022804" related pathway were concentrated on the brown2 module (**Figure 8C** and **Supplementary Table 9**). These related DEGs in the firebrick4 module were associated with "hormone-binding, GO:0042562," "ethylene receptor activity, GO: 0038199," "ethylene binding, GO:0051740" and "NAD(P)H dehydrogenase (quinone) activity, GO:0003955" (**Figure 8C** and **Supplementary Table 9**). In the dark turquoise module, these DEGs were involved in "oxidoreductase activity, acting on diphenols and related substances as donors, oxygen as acceptor, GO:0016682," "phenylpropanoid metabolic process, GO:0009698" and hydroquinone: oxygen oxidoreductase, GO:0052716" (**Figure 8C** and **Supplementary Table 9**). These results further suggested that pigment and ROS metabolism were very active during hickory stigma development.

Co-expression Network of Genes Related to Reactive Oxygen Species and Pigments Metabolism

According to pigment and ROS metabolism as well as related transcription factors in 4 key modules, a co-expression network was further constructed (**Figure 8D**). It was found that this network included 23 chlorophyll-related genes, 19 carotenoid-related genes, 141 anthocyanin-related genes, 46 ROS-related genes, and 623 transcription factors (TFs) (**Supplementary Table 10**). These transcription factors belonged to different families, including MYB (99), AP2(52), bHLH (52), NAC(30), WRKY(28), C_2H_2(21), HB-HD-ZIP(20), bZIP9(19), Trihelix(18), and others (284) (**Supplementary Figure 2**). Notably, 7 transcription factors (6 in the MYB family and 1 in the bHLH family) were shown to be related to anthocyanin metabolism. Moreover, the number of connections between genes was also counted, ranging from 49 to 447 (**Figure 8E** and **Supplementary Table 10**). In the chlorophyll metabolism pathway, *HEME2* (CCA1503S0015) and *CAO* (CCA1377S0020) had relatively high connections with more than 400 connections. It was also found that the connections of only the *PSY* (CCA0887S0096) gene in carotenoid biosynthesis-related genes were over 400. In the anthocyanins metabolism pathway, genes with more than 400 connections were *FLS1* (CCA0006S0007), *4CL8* (CCA0881S0026), *F3′H* (CCA0743S0031), *F3′H* (CCA0507S0050), and *TT12* (TRANSPARENT TESTA 12-like, CCA1606S0005). In ROS metabolism, *GSTU18* (glutathione S-transferase TAU 18, CCA0747S0058) was shown to have over 400 connections. Among genes encoding all the transcription factors, there were 21 genes with more than 400 connections, of which, the gene expression difference of *WRKY75* (WRKY transcription factor 75, CCA0573S0068) was the largest during stigma development. Remarkably, *MYB113* (MYB transcription factor 113, CCA0887S0030) was more connected to other genes in 7 anthocyanin-related transcription factors. Based on

this, we further mapped the co-expression network of these two transcription factors and genes related to pigment and ROS metabolism, which more clearly showed the regulatory relationship of transcription factors on them (**Figure 8E**). Using ProtParam, the physicochemical properties of the two transcription factors were further analyzed, as shown in **Table 1**. It was demonstrated that the protein length, pI, MW, and GRAVY of *WRKY75* and *MYB113* were 152 and 264 aa, 9.64 and 9.1, 17.64 and 30.21 kD, and –1.156 and –0.536, respectively. Both proteins were hydrophilic with no signal peptides. Furthermore, the subcellular localization prediction showed that they were both located in the nucleus. Taken together, these results provide in-depth guidance for revealing the transcriptional regulation of pigment and ROS metabolism in the future.

DISCUSSION

Darwin had noted that when the wind fertilizes a flower, it never has a gaily colored corolla. It means that the relationship between floral traits and pollinator behavior has been an important factor influencing the co-evolution of plants and pollinators (Chittka et al., 2001). Hickory is a monoecious plant of the genus *Carya* in the Juglandaceae, and its nuts are one of the most popular nuts among consumers because of their high nutritional value. It is a wind-pollinated plant, while the stigma of female flowers shows different colors with development. From opening to mature, the stigma color presents successively green, light-purple red, dark-purple red, dark purple (**Figure 1A**). Chlorophylls, carotenoids, and anthocyanins are three chemically diverse groups of pigments and the core members in flower color formation (Xue et al., 2020; Xia et al., 2021). Our present work showed that the stigma appeared green at the first stage (S1) when the chlorophyll content was higher than the other two pigments, gradually turning red and purple as the anthocyanin and carotenoid content increased (S2 to S3), and eventually getting dark purple due to the reduction of carotenoid content and sharp increase of anthocyanin content of from S3 to S4 (**Figures 1A–D**). These results suggested that chlorophyll, carotenoid, and anthocyanin were involved in the color formation of hickory stigma, supporting an identified possibility to consider that flavonoids and especially anthocyanins play many essential functions in reproductive tissues, but do not attract pollinators (Gould, 2004; Taylor and Grotewold, 2005). It was also shown that there were sharper changes in anthocyanin than that in other pigments from S3 to S4 (**Figures 1B–D**). Anthocyanins are known to have a strong ROS-removal capacity (Raven, 2000; Severo et al., 2015), stimulating our curiosity about the ROS elimination in hickory stigmas. Notably, the results of the oxidative scavenging ability of the sample extracts were consistent with the pigment change trends, especially the anthocyanin (**Figures 1E–G**). Therefore, it was speculated that this phenomenon was a protective mechanism of the plant itself against ROS accumulation.

In hickory, the whole genome has been sequenced and assembled, and the transcriptomes of several tissues have been sequenced (Huang et al., 2019). However, most studies have focused on male and female flower organ development and fruit formation stages (Huang et al., 2022), but there is no transcriptomic analysis of color changes in different flowering stages. In this research, RNA-seq technology was performed to analyze 4 sample stages with three replicates during hickory stigma development to investigate the model and mechanism of changes in transcriptional patterns associated with flower color.

Through time-series expression analysis, 12 clusters were observed (**Figure 3A**). KEGG analysis demonstrated cluster 5 was associated with porphyrin and chlorophyll metabolism, and the expression patterns were basically consistent with chlorophyll content and flower color (**Figures 1B**, **3B**). In cluster 10, the "flavonoid biosynthesis" pathway was enriched, and the expression trends were similar to that of anthocyanin content, forming red and purple stigmas. In cluster 11, other flavonoids, such as flavone, flavonol, and isoflavonoid, were enriched with a drastic reduction tendency from S1 to S2 but slight changes from S2 to S4, which may be beneficial for more prerequisite substances of phenylpropane synthesis pathway to enter anthocyanin synthesis (**Figures 1D**, **3B**). In cluster 12, oxidative phosphorylation-related genes were clustered and showed a trend of increasing from S1 to S3, which was similar to that of oxidative scavenging ability, proving some molecular evidence for ROS metabolism (**Figures 1E–G**, **3B**). According to all genes, 10432 DEGs were found, including 6549 upregulated genes and 6181 downregulated genes (**Figure 4A**). KEGG analysis of these DEGs showed that 14 pathways were enriched, including "chlorophyll metabolism," "carotenoid biosynthesis," and "anthocyanin biosynthesis." These results suggested that these genes required for the color formation were very active, offering some explanation for the dynamic changes of hickory stigma color.

The chlorophyll metabolism pathway had been well elucidated (Hörtensteiner, 2013). The chlorophyll metabolism pathway had been painted based on the whole genomic information of the hickory and KEGG database, and 27 DEGs were identified in this study (**Figure 5** and **Supplementary Table 6**). According to RNA-seq analysis and qRT-PCR assays, most genes showed high expression at the beginning of development and decreased at later stages. In contrast, chlorophyll degradation genes showed higher expression at later stages. Similarly, this phenomenon has been verified in *Lonicera japonica* Thunb. and *Lilium* (Xu L. et al., 2017; Xue et al., 2020).

Carotenoids can make plants yellow, orange, and red (Nisar et al., 2015). In this study, 28 DEGs were identified, including the critical candidate enzymes and TFs related to carotenoid metabolism (**Supplementary Table 6**). The present study showed that most of the DEGs had higher expression levels at the middle stage (**Figure 6**), consistent with the determination of carotenoid content. Interestingly, based on WGCNA and co-expression network results, 4 of these genes were connected with ABA, one of the carotenoid degradation products (**Figure 8A** and **Supplementary Table 10**) (North et al., 2007). Therefore, it meant that the carotenoids participated in pigment formation in hickory stigma.

Based on the analysis of the three pigment content measurements, the anthocyanin content appeared to be a

significantly sharper change (**Figure 1D**), so we believe that anthocyanin is not only a role pigment in the accumulation and color change of hickory stigma but also one of the reasons that may lead to other biological phenomena during the development process, which is worthy of attention. The anthocyanin synthesis pathway has been thoroughly investigated as a type of flavonoid compound (Ferrer et al., 2008). Using the KEGG analysis and selected criteria, the study had identified 36 DEGs associated with the anthocyanins pathway (**Supplementary Table 6**). According to the RNA-seq results, KEGG database, and previous studies, the anthocyanin synthesis, and transport pathway had been diagrammed, with the heatmaps of several role genes being plotted beside it (**Figure 7**), qRT-PCR assays had been used to qualify the role genes, and the trends of expression of the most consisted with transcriptome results (**Figure 7**). It has been reported that the accumulation of UV-absorbing pigments, especially flavonoids in epidermal tissues, was the main mechanism of photo-protection in plants against UV light (Agati and Tattini, 2010; Agati et al., 2013; Xu Y. et al., 2019). It has also been shown that the quercetin derivatives biosynthesis genes *CHS, CHI, F3H, FLS*, and *F3H* were present in lower terrestrial plants (Markham, 1988; Rausher, 2006), which were the most sensitive genes in the face of oxidative damage in modern terrestrial plants (Lin et al., 2003; Pollastri and Tattini, 2011). Notably, these genes have also been identified in anthocyanin-related DEGs. In May, the hickory flower begins to form, when strong sunlight hits the stigma surface, the young stigma can be directly and easily burnt. Therefore, it is likely that anthocyanin-related genes were activated by strong sunlight to promote the anthocyanin accumulation to counteract high-intensity light-induced oxidative damage by scavenging ROS.

Previous studies have been reported that some transcription factors (TFs) regulate the anthocyanin metabolism, some of which including MYB, bHLH, and WD-40 can be united to form complex functions (Gonzalez et al., 2008; Liu et al., 2016; Wang et al., 2019). So it is very important to find core TFs regulating anthocyanin and ROS metabolism. Based on WGCNA and co-expression network analysis (**Figure 8A** and **Supplementary Table 10**), 7 transcription factors related to anthocyanin synthesis, especially, *MYB113* (*CCA0887S0030*) had the most connectivity, suggesting that it was a core TF in this pathway, which have also been reported as a core regulator in anthocyanin synthesis in *Arabidopsis, Solanum tuberosum*, and *Solanum melongena* (Walker et al., 1999; Liu et al., 2016; Zhou et al., 2020).

Our results have also shown that the DEGs were clustered into nine different color modules by WGCNA, among which, four key modules (darkred, brown2, firebrick4, and dark turquoise) (**Figure 8A**) were found. Moreover, after investigating WGCNA, together with GO analysis (**Figure 8C**), the darkred module could cluster some genes responsible for "photosynthesis" and "photosystem," and these genes were highly expressed at the early stage of female flower stigma development. This is an indication that the early stage stigma mainly performs photosynthesis to provide plant growth and development nutrients. Previous studies have reported that adjustment of growth and development to light conditions is usually established by changes in hormone levels (de Wit et al., 2016). As a classic plant hormone, Ethylene can regulate anthocyanin biosynthesis and ROS scavenging in a variety of plants (Ni et al., 2021). In *Arabidopsis*, *ERF4* and *ERF8* promote the accumulation of light-induced anthocyanins (Koyama and Sato, 2018). It has also been reported that ethylene can also regulate anthocyanin synthesis in ripe apple fruit through antagonism of the R2R3-MYB repressor *MYB17* and activators *MYB1* and *MdEIL1* (Wang et al., 2022). In the middle stage of hickory stigma development, the firebrick4 module was enriched in "hormone-binding," "ethylene receptor activity," and "ethylene binding" pathway, and related DEGs including some TFs have been sorted out, providing a new direction for our future study. At the late stages of flower stigma development, these DEGs were mainly enriched in "oxidoreductase activity," "active transmembrane transporter activity," "hormone-related," and "phenylpropanoid metabolic process," which was contributed to the balance of pigment and ROS accumulation. These results provide some molecular evidence for the scavenging rate of reactive oxygen radicals in the extracts of samples from different periods. So, it is likely that the stigma color change act on the ability to scavenge ROS to protect stigma against various abiotic and biotic stresses. Notably, the co-expression network also showed that *WRKY75* (CCA0573S0068) may be a core transcription factor involved in pigment and ROS metabolism. In *Arabidopsis*, *WRKY75* has been shown to induce H_2O_2 accumulation as a key way to accelerate the leaf senescence process (Li et al., 2012; Guo et al., 2017). In poplar, *WRKY75* was reported to reduce the reactive oxygen scavenging capacity of leaves under stress and negatively regulate salt and osmotic tolerance by modulating multiple physiological processes (Zhao et al., 2019). Previous studies have suggested that the WRKY structural domain can bind to the cis-acting element W-box (Devaiah et al., 2007). Further work will be done to verify the direct regulating relationship of *WRKY75* on the targeted gene with more W-box required for pigment and ROS metabolism, including *ATAF1, GSTL3,* and *GRF7,* using some molecular techniques.

CONCLUSION

In conclusion, it was demonstrated that color changes during hickory female stigma development were due to the dynamic changes in the content of chlorophylls, carotenoids, and anthocyanins. The antioxidant capacities had similar changing trends with carotenoids from S1 to S3 and anthocyanins at all four stages during stigma development. Transcription analysis of developing stigma provided the comprehensive molecular mechanism for the dynamic changes in pigment content and antioxidant capacities. *MYB113* (CCA0887S0030) and *WRKY75* (CCA0573S0068) were further predicted to be two core transcriptional regulators responsible for pigment and ROS metabolism. These results suggested that one of the important biological significance of the color change of female flower stigma was the photo-protection and anti-oxidation, the color variation

makes the appearance of this ability in plant stigmas a possibility, and corresponding models have also been proposed to explain this possibility as well (**Figure 8F**). At present, we have not yet found any link between stigma color and the pollination process in hickory, and some additional analyses are still needed to verify this possibility in the new future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA810757.

AUTHOR CONTRIBUTIONS

YL, YX, and JH conceived and designed the study. KW and YX analyzed the data. YX performed the experiments. YL and YX wrote the manuscript. YL, JH, and KW edited and reviewed the writing. All authors have read and approved this manuscript.

FUNDING

This research was financially supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2022C02009), Zhejiang Key Research and Development Project (2021C02037), the National Natural Science Foundation of China (No. 32101557), the National Key R&D Program of China (2018YFD1000604), Key Research and Development Project of Zhejiang Province, China (2020C02005), and the Zhejiang Agriculture New Variety Breeding Major Science and Technology Special (2021C02066-12).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.881394/full#supplementary-material

REFERENCES

Agati, G., Brunetti, C., Di Ferdinando, M., Ferrini, F., Pollastri, S., Tattini, M., et al. (2013). Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant Physiol. Biochem.* 72, 35–45. doi: 10.1016/j.plaphy.2013.03.014

Agati, G., and Tattini, M. (2010). Multiple functional roles of flavonoids in photoprotection. *New Phytol.* 186, 786–793. doi: 10.1111/j.1469-8137.2010.03269.x

Apel, K., and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701

Beale, S. I. (2005). Green genes gleaned. *Trends Plant Sci.* 10, 309–312. doi: 10.1016/j.tplants.2005.05.005

Bradshaw, H. D., and Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426, 176–178. doi: 10.1038/nature02106

Bramley, P. (1985). The *in Vitro* Biosynthesis of Carotenoids. *Adv. Lipid Res.* 21, 243–279. doi: 10.1016/B978-0-12-024921-3.50013-7

Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., et al. (2011). Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* 11:24. doi: 10.1186/1471-2229-11-24

Burits, M., and Bucar, F. (2000). Antioxidant activity of Nigella sativa essential oil. *Phytother. Res.* 14, 323–328. doi: 10.1002/1099-1573(200008)14:5<323::aid-ptr621<3.0.co;2-q

Chen, Q., Man, C., Li, D., Tan, H., Xie, Y., Huang, J., et al. (2016). Arogenate Dehydratase Isoforms Differentially Regulate Anthocyanin Biosynthesis in Arabidopsis thaliana. *Mol. Plant* 9, 1609–1619. doi: 10.1016/j.molp.2016.09.010

Chittka, L., Spaethe, J., Schmidt, A., and Hickelsberger, A. (2001). "Adaptation, constraint, and chance in the evolution of flower color and pollinator color vision," in *Cognitive Ecology of Pollination: Animal Behaviour and Floral Evolution*, eds L. Chittka and J. Thomson (Cambridge: Cambridge University Press), 106–126. doi: 10.1017/CBO9780511542268.007

Davies, K. M. (2007). Genetic modification of plant metabolism for human health benefits. *Mutat. Res.* 622, 122–137. doi: 10.1016/j.mrfmmm.2007.02.003

de Wit, M., Galvão, V. C., and Fankhauser, C. (2016). Light-Mediated Hormonal Regulation of Plant Growth and Development. *Annu. Rev. Plant Biol.* 67, 513–537. doi: 10.1146/annurev-arplant-043015-112252

Devaiah, B. N., Karthikeyan, A. S., and Raghothama, K. G. (2007). *WRKY75* transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. *Plant Physiol.* 143, 1789–1801. doi: 10.1104/pp.106.093971

Edlund, A. F., Swanson, R., and Preuss, D. (2004). Pollen and stigma structure and function: the role of diversity in pollination. *Plant Cell* 16, S84–S97. doi: 10.1105/tpc.015800

Elstner, E. F. I. O. B. (1991). Mechanisms of oxygen activation in different compartments of plant cells. *Curr. Opin. Solid State Mater. Sci.* 15, 342–343. doi: 10.1046/j.1365-2427.1996.00099.x

Falcone, F. M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222

Fantini, E., Falcone, G., Frusciante, S., Giliberto, L., and Giuliano, G. (2013). Dissection of tomato lycopene biosynthesis through virus-induced gene silencing. *Plant Physiol.* 163, 986–998. doi: 10.1104/pp.113.224733

Ferrer, J. L., Austin, M. B., Stewart, C. Jr, and Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* 46, 356–370. doi: 10.1016/j.plaphy.2007.12.009

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

Giuliano, G., Tavazza, R., Diretto, G., Beyer, P., and Taylor, M. A. (2008). Metabolic engineering of carotenoid biosynthesis in plants. *Trends Biotechnol.* 26, 139–145. doi: 10.1016/j.tibtech.2007.12.003

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

Goodwin, T. W. (1971). Biosynthesis of carotenoids and plant triterpenes. *Biochem. J.* 123, 293–329. doi: 10.1042/bj1230293

Gould, K. S. (2004). Nature's Swiss Army Knife: the Diverse Protective Roles of Anthocyanins in Leaves. *J. Biomed. Biotechnol.* 2004, 314–320. doi: 10.1155/S1110724304406147

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57, 761–780. doi: 10.1146/annurev.arplant.57.032905.105248

Guo, P., Li, Z., Huang, P., Li, B., Fang, S., Chu, J., et al. (2017). A Tripartite Amplification Loop Involving the Transcription Factor *WRKY75*, Salicylic Acid, and Reactive Oxygen Species Accelerates Leaf Senescence. *Plant Cell* 29, 2854–2870. doi: 10.1105/tpc.17.00438

He, J., and Giusti, M. M. (2010). Anthocyanins: natural colorants with health-promoting properties. *Annu. Rev. Food Sci. Technol.* 1, 163–187. doi: 10.1146/annurev.food.080708.100754

Hörtensteiner, S. (2013). Update on the biochemistry of chlorophyll breakdown. *Plant Mol. Biol.* 82, 505–517. doi: 10.1007/s11103-012-9940-z

Huang, C., Li, Y., Wang, K., Xi, J., Xu, Y., Si, X., et al. (2022). Analysis of lipidomics profile of Carya cathayensis nuts and lipid dynamic changes during embryonic development. *Food Chem.* 370:130975. doi: 10.1016/j.foodchem.2021.130975

Huang, Y., Liu, L., Huang, J., Wang, Z., Chen, F., Zhang, Q., et al. (2013). Use of transcriptome sequencing to understand the pistillate flowering in hickory (*Carya cathayensis* Sarg.). *BMC Genom.* 14:691. doi: 10.1186/1471-2164-14-691

Huang, Y., Xia, G., Wang, Z., Zheng, B., Liang, J., Huang, J., et al. (2007). Studies on anatomy of development of female flower in *Carya cathayensis* Sarg. *Acta Agric. Univ. Jiangxien.* 29, 723e–726e.

Huang, Y., Xiao, L., Zhang, Z., Zhang, R., Wang, Z., Huang, C., et al. (2019). The genomes of pecan and Chinese hickory provide insights into Carya evolution and nut nutrition. *GigaScience* 8:giz036. doi: 10.1093/gigascience/giz036

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18, 477–483. doi: 10.1016/j.tplants.2013.06.003

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Meth.* 12, 357–360. doi: 10.1038/nmeth.3317

Koyama, T., and Sato, F. (2018). The function of ETHYLENE RESPONSE FACTOR genes in the light-induced anthocyanin production of *Arabidopsis thaliana* leaves. *Plant Biotechnol. J.* 35, 87–91. doi: 10.5511/plantbiotechnology.18.0122b

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

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9, 357–359. doi: 10.1038/nmeth.1923

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

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713–714. doi: 10.1093/bioinformatics/btn025

Li, S., Wang, W., Gao, J., Yin, K., Wang, R., Wang, C., et al. (2016). MYB75 Phosphorylation by MPK4 Is Required for Light-Induced Anthocyanin Accumulation in *Arabidopsis*. *Plant Cell* 28, 2866–2883. doi: 10.1105/tpc.16.00130

Li, Y., Wang, J., Wang, K., Lyu, S., Ren, L., Huang, C., et al. (2022). Comparison analysis of widely-targeted metabolomics revealed the variation of potential astringent ingredients and their dynamic accumulation in the seed coats of both Carya cathayensis and Carya illinoinensis. *Food Chem.* 374:131688. doi: 10.1016/j.foodchem.2021.131688

Li, Z., Peng, J., Wen, X., and Guo, H. (2012). Gene network analysis and functional studies of senescence-associated genes reveal novel regulators of Arabidopsis leaf senescence. *J. Integr. Plant Biol.* 54, 526–539. doi: 10.1111/j.1744-7909.2012.01136.x

Lichtenthaler, H. (1987). Chlorophylls and carotenoids: pigments of photosynthetic membranes. *Meth. Enzymol.* 148, 350–383. doi: 10.1042/bst0110591

Lin, Y., Irani, N. G., and Grotewold, E. (2003). Sub-cellular trafficking of phytochemicals explored using auto-fluorescent compounds in maize cells. *BMC Plant Biol.* 3:10. doi: 10.1186/1471-2229-3-10

Linn, J., Ren, M., Berkowitz, O., Ding, W., van der Merwe, M. J., Whelan, J., et al. (2017). Root Cell-Specific Regulators of Phosphate-Dependent Growth. *Plant Physiol.* 174, 1969–1989. doi: 10.1104/pp.16.01698

Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., Yu, B., et al. (2016). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *J. Exp. Bot.* 67, 2159–2176. doi: 10.1093/jxb/erw014

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

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

Markham, K. R. (1988). "Distribution of flavonoids in the lower plants and its evolutionary significance," in *The Flavonoids: Advances in Research Since 1980*, ed. J. B. Harborne (London: Chapman and Hall), 427e–468e.

Mittler, R. (2017). ROS Are Good. *Trends Plant Sci.* 22, 11–19. doi: 10.1016/j.tplants.2016.08.002

Mittler, R., Vanderauwera, S., Gollery, M., and Van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends Plant Sci.* 9, 490–498. doi: 10.1016/j.tplants.2004.08.009

Nasrallah, J. B. (2000). Cell-cell signaling in the self-incompatibility response. *Curr. Opin. Plant Biol.* 3, 368–373. doi: 10.1016/s1369-5266(00)00098-4

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

Ni, J., Premathilake, A. T., Gao, Y., Yu, W., Tao, R., Teng, Y., et al. (2021). Ethylene-activated *PpERF105* induces the expression of the repressor-type R2R3-MYB gene *PpMYB140* to inhibit anthocyanin biosynthesis in red pear fruit. *Plant J.* 105, 167–181. doi: 10.1111/tpj.15049

Nisar, N., Li, L., Lu, S., Khin, N. C., and Pogson, B. J. (2015). Carotenoid metabolism in plants. *Mol. Plant* 8, 68–82. doi: 10.1016/j.molp.2014.12.007

North, H. M., De Almeida, A., Boutin, J. P., Frey, A., To, A., Botran, L., et al. (2007). The Arabidopsis ABA-deficient mutant aba4 demonstrates that the major route for stress-induced ABA accumulation is *via* neoxanthin isomers. *Plant J.* 50, 810–824. doi: 10.1111/j.1365-313X.2007.03094.x

Pannell, J. R. (2017). Plant Sex Determination. *Curr. Biol.* 27, R191–R197. doi: 10.1016/j.cub.2017.01.052

Petretto, G. L., Maldini, M., Addis, R., Chessa, M., Foddai, M., Rourke, J. P., et al. (2016). Variability of chemical composition and antioxidant activity of essential oils between Myrtus communis var. Leucocarpa DC and var. Melanocarpa DC. *Food Chem.* 197, 124–131. doi: 10.1016/j.foodchem.2015.10.056

Pollastri, S., and Tattini, M. (2011). Flavonols: old compounds for old roles. *Ann. Bot.* 108, 1225–1233. doi: 10.1093/aob/mcr234

Qi, T., Song, S., Ren, Q., Wu, D., Huang, H., Chen, Y., et al. (2011). The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in Arabidopsis thaliana. *Plant Cell* 23, 1795–1814. doi: 10.1105/tpc.111.083261

Ramel, F., Birtic, S., Cuiné, S., Triantaphylidès, C., Ravanat, J. L., and Havaux, M. (2012). Chemical quenching of singlet oxygen by carotenoids in plants. *Plant Physiol.* 158, 1267–1278. doi: 10.1104/pp.111.182394

Rausher, M. D. (2006). "The evolution of flavonoids and their genes," in *The Science of Flavonoids*, ed. E. Grotewold (New York: Springer), 175e–211e.

Raven, J. A. (2000). Land plant biochemistry. *Phil. Trans. R. Soc. Lond. B* 355, 833–846. doi: 10.1098/rstb.2000.0618

Rogers, H., and Munné-Bosch, S. (2016). Production and Scavenging of Reactive Oxygen Species and Redox Signaling during Leaf and Flower Senescence: similar But Different. *Plant Physiol.* 171, 1560–1568. doi: 10.1104/pp.16.00163

Severo, J., de Oliveira, I. R., Tiecher, A., Chaves, F. C., and Rombaldi, C. V. (2015). Postharvest UV-C treatment increases bioactive, ester volatile compounds and a putative allergenic protein in strawberry. *LWT Food Sci. Technol.* 64, 685–692. doi: 10.1016/j.lwt.2015.06.041

Shen, C., Xu, Y., Huang, J., Wang, Z., Qiu, J., Huang, Y., et al. (2014). Molecular characterization and expression analysis of the critical floral genes in hickory (*Carya cathayensis* Sarg.). *Plant Physiol. Biochem.* 83, 142–150. doi: 10.1016/j.plaphy.2014.07.020

Smoot, M., Ono, K., Ideker, T., and Maere, S. (2011). PiNGO: a Cytoscape plugin to find candidate genes in biological networks. *Bioinformatics* 27, 1030–1031. doi: 10.1093/bioinformatics/btr045

Tanaka, Y., and Ohmiya, A. (2008). Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. *Curr. Opin. Biotechnol.* 19, 190–197. doi: 10.1016/j.copbio.2008.02.015

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., Nueda, M. J., Ferrer, A., et al. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* 43, e140. doi: 10.1093/nar/gkv711

Taylor, L. P., and Grotewold, E. (2005). Flavonoids as developmental regulators. *Curr. Opin. Plant Biol.* 8, 317–323. doi: 10.1016/j.pbi.2005.03.005

Vallabhaneni, R., Bradbury, L. M., and Wurtzel, E. T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. *Arch. Biochem. Biophys.* 504, 104–111. doi: 10.1016/j.abb.2010.07.019

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., et al. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1350. doi: 10.1105/tpc.11.7.1337

Wang, L., Lu, W., Ran, L., Dou, L., Yao, S., Hu, J., et al. (2019). R2R3-MYB transcription factor *MYB6* promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in Populus tomentosa. *Plant J.* 99, 733–751. doi: 10.1111/tpj.14364

Wang, S., Li, L. X., Zhang, Z., Fang, Y., Li, D., Chen, X. S., et al. (2022). Ethylene precisely regulates anthocyanin synthesis in apple *via* a module comprising MdEIL1, MdMYB1, and MdMYB17. *Hort. Res.* uhac034. [Epub online ahead of print]. doi: 10.1093/hr/uhac034

Wang, Z. J., Huang, J. Q., Huang, Y. J., Li, Z., and Zheng, B. S. (2012). Discovery and profiling of novel and conserved microRNAs during flower development in *Carya cathayensis via* deep sequencing. *Planta* 236, 613–621. doi: 10.1007/s00425-012-1634-x

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* 5, 218–223. doi: 10.1016/s1369-5266(02)00256-x

Wrolstad, R., and Giusti, M. (2001). Characterization and Measurement of Anthocyanins by UV-Visible Spectroscopy. *Curr. Protocols Food Anal. Chem.* 1, doi: 10.1002/0471142913.faf0102s00

Xia, G. (2006). *Reproductive Biology of Carya cathayensis Sarg*. [PhD thesis]. Xuanwu: Nanjing Forestry University.

Xia, Y., Chen, W., Xiang, W., Wang, D., Xue, B., Liu, X., et al. (2021). Integrated metabolic profiling and transcriptome analysis of pigment accumulation in *Lonicera japonica* flower petals during colour-transition. *BMC Plant Biol.* 21:98. doi: 10.1186/s12870-021-02877-y

Xu, L., Yang, P., Feng, Y., Xu, H., Cao, Y., Tang, Y., et al. (2017). Spatiotemporal Transcriptome Analysis Provides Insights into Bicolor Tepal Development in Lilium "Tiny Padhye". *Front. Plant Sci.* 8:398. doi: 10.3389/fpls.2017.00398

Xu, Y., Charles, M. T., Luo, Z., Mimee, B., Tong, Z., Véronneau, P. Y., et al. (2019). Ultraviolet-C priming of strawberry leaves against subsequent Mycosphaerella fragariae infection involves the action of reactive oxygen species, plant hormones, and terpenes. *Plant Cell Environ.* 42, 815–831. doi: 10.1111/pce.13491

Xu, Y., Charles, M. T., Luo, Z., Mimee, B., Veronneau, P. Y., Rolland, D., et al. (2017). Preharvest Ultraviolet C Irradiation Increased the Level of Polyphenol Accumulation and Flavonoid Pathway Gene Expression in Strawberry Fruit. *J. Agric. Food Chem.* 65, 9970–9979. doi: 10.1021/acs.jafc.7b04252

Xue, Q., Fan, H., Yao, F., Cao, X., Liu, M., Sun, J., et al. (2020). Transcriptomics and targeted metabolomics profilings for elucidation of pigmentation in *Lonicera japonica* flowers at different developmental stages. *Ind. Crops Prod.* 145:111981. doi: 10.1016/j.indcrop.2019.111981

Zeng, X. Q., Chow, W. S., Su, L. J., Peng, X. X., and Peng, C. L. (2010). Protective effect of supplemental anthocyanins on Arabidopsis leaves under high light. *Physiol. Plant* 138, 215–225. doi: 10.1111/j.1399-3054.2009.01316.x

Zhao, D., and Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Front. Plant Sci.* 6:261. doi: 10.3389/fpls.2015.00261

Zhao, K., Zhang, D., Lv, K., Zhang, X., Cheng, Z., Li, R., et al. (2019). Functional characterization of poplar WRKY75 in salt and osmotic tolerance. *Plant Sci.* 289:110259. doi: 10.1016/j.plantsci.2019.110259

Zheng, T., Tan, W., Yang, H., Zhang, L., Li, T., Liu, B., et al. (2019). Regulation of anthocyanin accumulation *via MYB75/HAT1/TPL*-mediated transcriptional repression. *PLoS Genet.* 15:e1007993. doi: 10.1371/journal.pgen.1007993

Zhou, L., He, Y., Li, J., Liu, Y., and Chen, H. (2020). CBFs Function in Anthocyanin Biosynthesis by Interacting with *MYB113* in Eggplant (*Solanum melongena* L.). *Plant Cell Physiol.* 61, 416–426. doi: 10.1093/pcp/pcz209

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.

GLOSSARY

Glu-TR, glutamyl tRNA reductase; GSA-AM, glutamate-1-semialdehyde-2,1-amino mutase; PBCS, Porphobilinogen synthase; PBGD, bile chromogen dehydrogenase; UROS, Uroporphyrinogen III synthase; UROD, Uroporphyrinogen III decarboxylase; CPOX, Coproporhyrinogen III oxidase; PPOX, Protoporphyrinogen oxidase; MgCh, Magnesium chelatase H subunit; MgPM, Magnesium proto IX methyltransferase; MgPEC, Mg-protoporphyrin IX monomethylester; DVR, 3,8-divinyl protochlorophyllide a 8-vinyl reductase; POR, Protochlorophyllide oxidoreductase; CHLG, Chlorophyll synthase; CAO, Chlorophyllide a oxygenase; HCAR, hydroxy-Chl a reductasehydroxy-Chl a reductase; NYC1/NOL, non-yellow coloring1/NYC1-like; MCS, metal chelating substance; PPH, pheophytinase; PAO, pheophorbide a oxygenase; pFCC, primary fluorescent Chl Catabolite; RCCR, red Chl catabolite reductase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT, 2-c-methyl-d-erythritol 4-phosphate cytidyl transferase; CMK, 2-c-methyl-d-erythritol 4-phosphate cytidyl transferase; MDS, 2-C-Methyl-D-erythritol2,4-cyclodiphosphate synthase; HDS, 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IPI, isopentenyl diphosphate isomerase; GGPPS, geranylgeranyl diphosphate synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZISO, -carotene isomerase; ZDS, -carotene desaturase; LCY-E, lycopene-(-cyclase); LCY-B, lycopene-(-cyclase); LUT5, LUTEIN DEFICIENT 5; LUT1, LUTEIN DEFICIENT 1; ATVDE, violaxanthin de-epoxidase; ZEP, zeaxanthin epoxidase; CHY-B, carotene-(-hydroxylase); CHY-E, carotene-(-hydroxylase); VDE, violaxanthin de-epoxidase; CCD, Carotenoid Cleavage Dioxygenase; NCED, 9-cis-epoxycarotenoid dioxygenase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CHS(catalysis of chalcone synthase; CHI, chalcone isomerase; F3H, flavanone-3-hydroxylase; F3′H, flavanone-3′-hydroxylase; F3′5′H, flavanone-3′5′-hydroxylase; ANS, anthocyanin synthase; FLS, Flavonol synthase; DFR, dihydroxyflavone reductase; UFGT, flavonoid-3-o-glucosyltransferase; MYB113, MYB transcription factor 113; PAP1, production of anthocyanin pigment 1, MYB75; PAP2, production of anthocyanin pigment 2, MYB90; TT8, transparent testa 8; EGL3, enhancer of glabra; and TTG1, transparent testa glabra 1. ABTS (2, 2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid))" and "DPPH (1,1-Diphenyl-2-picrylhydrazyl radical 2,2-Diphenyl-1-(2,4,6-trinitrophenyl)hydrazyl).

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fpls.2022.879390

CRISPR/Cas9-Mediated Targeted Mutagenesis of *FtMYB45* Promotes Flavonoid Biosynthesis in Tartary Buckwheat (*Fagopyrum tataricum*)

Dong Wen[1], Lan Wu[1], Mengyue Wang[1], Wei Yang[1], Xingwen Wang[1], Wei Ma[2], Wei Sun[1], Shilin Chen[1], Li Xiang[1] and Yuhua Shi[1]**

[1]Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [2]College of Pharmaceutical Sciences, Heilongjiang University of Chinese Medicine, Harbin, China

OPEN ACCESS

Edited by:
Yongliang Liu,
University of Kentucky,
United States

Reviewed by:
Chong Ren,
Institute of Botany (CAS), China
Miaoying Tian,
University of Hawaii at Manoa,
United States

***Correspondence:**
Yuhua Shi
yhshi@icmm.ac.cn
Li Xiang
lxiang@icmm.ac.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *19 February 2022*
Accepted: *25 April 2022*
Published: *12 May 2022*

Citation:
Wen D, Wu L, Wang M, Yang W, Wang X, Ma W, Sun W, Chen S, Xiang L and Shi Y (2022) CRISPR/Cas9-Mediated Targeted Mutagenesis of FtMYB45 Promotes Flavonoid Biosynthesis in Tartary Buckwheat (Fagopyrum tataricum). Front. Plant Sci. 13:879390. doi: 10.3389/fpls.2022.879390

The clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9 (CRISPR/Cas9) technology is an efficient genome editing tool used in multiple plant species. However, it has not been applied to Tartary buckwheat (*Fagopyrum tataricum*), which is an important edible and medicinal crop rich in rutin and other flavonoids. *FtMYB45* is an R2R3-type MYB transcription factor that negatively regulates flavonoid biosynthesis in Tartary buckwheat. Here, the CRISPR/Cas9 system polycistronic tRNA-sgRNA (PTG)/Cas9 was employed to knock out the *FtMYB45* gene in Tartary buckwheat. Two single-guide RNAs (sgRNAs) were designed to target the second exon of the *FtMYB45* gene. Twelve transgenic hairy roots were obtained using *Agrobacterium rhizogenes*-mediated transformation. Sequencing data revealed that six lines containing six types of mutations at the predicted double-stranded break site were generated using sgRNA1. The mutation frequency reached 50%. A liquid chromatography coupled with triple quadrupole mass spectrometry (LC-QqQ-MS) based metabolomic analysis revealed that the content of rutin, catechin, and other flavonoids was increased in hairy root mutants compared with that of lines transformed with the empty vector. Thus, CRISPR/Cas9-mediated targeted mutagenesis of *FtMYB45* effectively increased the flavonoids content of Tartary buckwheat. This finding demonstrated that the CRISPR/Cas9 system is an efficient tool for precise genome editing in Tartary buckwheat and lays the foundation for gene function research and quality improvement in Tartary buckwheat.

Keywords: *Fagopyrum tataricum*, PTG/Cas9 system, targeted genome editing, *FtMYB45* gene, flavonoid biosynthesis

INTRODUCTION

Tartary buckwheat [*Fagopyrum tataricum* (L.) Gaertn], also called bitter buckwheat, is an annual dicotyledonous plant belonging to the Polygonaceae family and the Fagopyrum genus. It is a diploid species ($2n = 2x = 16$), originates in the mountains of Western China at 400–3,900m of altitude, and is mainly cultivated in the Himalayas, Southeast Asia, Europe, and South

America, particularly in China, Afghanistan, Bhutan, Kazakhstan, Northern India, and Nepal (Wang and Campbell, 2007; Guo et al., 2011). Tartary buckwheat is an important traditional medicinal and edible plant, which is considered a new plant-based ingredient to enrich corn-based gluten-free formulations (Appiani et al., 2021). It is rich in various flavonoids, high-quality proteins, amino acids, and dietary fiber (Zhao et al., 2012). Flavonoids such as rutin, catechin, and epicatechin are the most important biologically active components of Tartary buckwheat (Md et al., 2013). Studies have shown that flavonoids effectively improve the symptoms of and prevent cardiovascular diseases, hypertension, diabetes, retinal hemorrhage, and acute hemorrhagic nephritis. They also have positive effects on the stomach, promote digestion, and improve immunity (Martínez Conesa et al., 2005; Jiang et al., 2007; Tomotake et al., 2007). Thus, Tartary buckwheat has become an important functional food (Zhou et al., 2015).

MYB transcription factors play important roles in the regulation of flavonoid biosynthesis in plants. In *Arabidopsis thaliana*, *AtMYB3*, *AtMYB4*, *AtMYB7*, and *AtMYB32* inhibit phenylpropanoid biosynthesis (Jin et al., 2000; Preston et al., 2004; Dubos et al., 2010; Fornalé et al., 2014). In *Salvia miltiorrhiza*, *SmMYB36* and *SmMYB39* prevent the accumulation of phenolic acid (Zhang et al., 2013; Ding et al., 2017). In *F. tataricum*, several MYB transcription factors were reported to activate or repress flavonoid biosynthesis. Overexpressing *FtMYB1* and *FtMYB2* enhances the biosynthesis and accumulation of anthocyanins (Bai et al., 2014). *FtMYB116* can be induced by red and blue light and promotes the accumulation of rutin by directly inducing the expression of flavonoid-3′-hydroxylase *(F3'H)*, which is involved in flavonoid biosynthesis (Zhang et al., 2018). The R2R3-MYB transcription factor *FtMYB6* is also induced by light and promotes flavonol biosynthesis by activating the expression of *FtF3H* and *FtFLS1* (Yao et al., 2020). *FtMYB11* represses phenylpropanoid biosynthesis (Zhou et al., 2017). *FtMYB13*, *FtMYB14*, *FtMYB15*, and *FtMYB16* are considered as negative regulators repressing rutin biosynthesis (Zhang et al., 2018).

The clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9 (CRISPR/Cas9) system has been recently developed from the adaptive immune system of *Streptococcus pyogenes* and is a powerful tool for targeted genome editing (Jinek et al., 2012). The CRISPR/Cas9 technology usually consists of two parts, an artificial single-guide RNA (sgRNA) and Cas9 nuclease. It has been successfully used for targeted gene modifications in a wide variety of plants (Feng et al., 2013), such as Arabidopsis (Jiang et al., 2014), rice (Shan et al., 2015; Srivastava et al., 2017), potato (Wang et al., 2015), maize (Shin et al., 2017), soybean (Cai et al., 2018), and rapeseed (Wang et al., 2017). However, the CRISPR/Cas9 system has been seldom used in *F. tataricum*. The polycistronic tRNA-sgRNA (PTG)/Cas9 system has been reported to be more efficient for gene editing in rice (Xie et al., 2015), kiwifruit (Wang et al., 2018), sweet orange (Tang et al., 2021), and grape (Ren et al., 2021). This technology uses the endogenous tRNA processing system to boost CRISPR/Cas9 gene editing capability. It consists of multiple tandemly arrayed tRNA-sgRNA units that form the PTG gene. Studies indicated that the start and end sites of the tRNA in the tandemly arrayed tRNA-sgRNA transcripts can be precisely recognized and cleaved by endogenous RNases (RNase P and RNase Z in plants) to simultaneously produce multiple functional sgRNAs (Xie et al., 2015).

Here, the PTG/Cas9 system was employed for targeted mutagenesis of *FtMYB45* in *F. tataricum*. The *FtMYB45 (MYB15)* gene has been identified as a transcriptional repressor of the flavonoid biosynthetic pathway, particularly of rutin (Zhang et al., 2018). In this study, the PTG/Cas9 system effectively induce mutations of the target gene *FtMYB45* in transgenic hairy roots, and the content of flavonoids such as rutin increased in mutant lines. Thus, the PTG/Cas9 gene editing system was efficacious in *F. tataricum*. To our knowledge, the present work is the first report of the CRISPR/Cas9 technology applied in *F. tataricum*, which provides a good technical foundation for molecular genetic studies in Tartary buckwheat.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The Tartary buckwheat variety Jinqiao No. 2 used in this study was provided by Professor Qingfu Chen from Guizhou Normal University. The peeled seeds were soaked in distilled water for 20 min, then sterilized in 75% ethanol for 45 s and in 1 g L^{-1} mercuric chloride for 8 min, and washed 3–4 times with sterile water. Afterward, the seeds were blotted on filter paper to remove excess water and sown onto Murashige and Skoog (MS) medium in a greenhouse with a 16-h light/8-h dark photoperiod at 25°C.

For UV-B treatment, the seedlings were grown in full darkness for 5 days and then irradiated with UV-B light (300 nm, 2.0×100 μw/cm^2) for 6 h. Seedlings kept in the dark were used as controls. The treatment comprised three biological replicates. Seedlings were frozen in liquid nitrogen and stored at −80°C.

Quantitative Real-Time Reverse Transcription-Polymerase Chain Reaction

Total RNA was isolated using an RNA Extraction Kit (Takara, Dalian, Liaoning, China). First-strand cDNA synthesis was performed using 2 μg of the total RNA and PrimeScript™ RT reagent Kit (Takara), and qRT-PCR was conducted in a total volume of 20 μl on the qTOWER 3 real-time PCR system (ChemStudio SA, Analytik Jena, Germany) using SYBR Premix ExTaq Mix (Takara). The primers for qRT-PCR were designed using Primer Premier 5 (Premier Biosoft, United States) and are listed in **Supplementary Table S1**, and the Tartary buckwheat *Histone 3 (H3)* gene (GenBank accession number. JF769134) was used as an internal control gene (Gocal et al., 2001). PCR cycling began with a denaturing step at 95°C for 2 min, followed by 40 cycles at 95°C for 5 s, 60°C for 10 s, and 72°C for 10 s. Finally, the dissolution curve signals were collected from 60°C to 95°C. Three biological and three technical replicates were

performed. The gene expression levels were analyzed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

sgRNA Design and Vector Construction

Two sgRNAs (sgRNA1 and sgRNA2) targeting *FtMYB45* were designed and their off-target effects were analyzed based on the published genome sequence of *F. tataricum* (GenBank accession number: GCA_002319775.1) and the website (Concordet and Haeussler, 2018).[1] Secondary structure analysis of target-sgRNA sequences was carried out with the program RNA Folding Form (Ma et al., 2015b).[2] The specific PCR primers C45-F and C45-R spanning sgRNA target sites were designed (**Supplementary Table S1**). A 360-bp fragment was amplified by PCR using genomic DNA of Jinqiao No. 2 as template, purified using a PCR purification kit (TransGen Biotech, Beijing, China), and sequenced to verify the sequences of sgRNA1 and sgRNA2.

The CRISPR/Cas9 vector targeting the *FtMYB45* gene was constructed using PTG/Cas9 method according to method of Wang et al. (2018). The sgRNA intermediate vector pHLW-sgRNA-tRNA and the Cas9 binary vector pPTG-sgRNA-Cas9-U6-1 were used. First, the fragment containing the first *Bsa* I site, sgRNA1, sgRNA scaffold, tRNA, sgRNA2, and the second *Bsa* I site was amplified from vector pHLW-sgRNA-tRNA using the target-specific primers 45sg-F and 45sg-R (**Supplementary Table S1**). Then, the PCR fragment was digested with restriction enzymes *Bsa* I (New England Biolabs, United States), and ligated into the BsaI-linearized vector pPTG-gRNA-Cas9-U6-1 with T4 DNA ligase (New England Biolabs, United States) to generate vector PTG/Cas9-FtMYB45. The ligation mixture was transformed into *Escherichia coli* DH5α competent cells and plated on LB-kanamycin agar plate (50 mg/L). Positive clones were confirmed by colony PCR using the primers SP-F and SP-R primers (**Supplementary Table S1**). All primers were synthesized commercially (Sangon Biotech Co., Ltd., Shanghai, China), *E. coli* competent cells were produced in our laboratory.

Agrobacterium rhizogenes-Mediated Hairy Root Transformation in *Fagopyrum tataricum*

Hairy roots transformation of Tartary buckwheat mediated by *A. rhizogenes* was performed as previously reported (Mi et al., 2020). Briefly, the cotyledons and hypocotyls of 7–10 day-old old Tartary buckwheat seedlings were used as explants. The cotyledons were cut into small squares and the hypocotyls were cut into approximately 0.5-cm segments. All explants were precultured on MS solid medium for 1 day. The plasmid PTG/Cas9-FtMYB45 was introduced into *A. rhizogenes ACCC10060* by electrotransformation. *Agrobacterium rhizogenes ACCC10060* strain harboring PTG/Cas9-FtMYB45 was cultured in a shaker at 200 rpm and 28°C until the OD_{600} value reached 0.2. The prepared explants were soaked in the bacteria suspension for 10 min and cocultivated on a cocultivation medium (MS + 100mM acetosyringone) with filter paper at 25°C for 3 days in the dark. Then, the cocultured explants were transferred onto a selection medium (MS + 200 mg/L cefadroxil + 50 mg/L kanamycin) and cultured under a 16-h light/8-h dark cycle at 25°C for about 2 weeks until hairy roots were induced. Afterward, hairy roots were cut into 2–3 cm pieces and transferred into 100 ml glass bottles containing 10 ml of the selection medium and cultured in a shaker at 80 rpm at 25°C in the dark until they overspread to the bottom of the glass bottles (replace medium every 7 days if necessary). After 10–12 days, the hairy roots were collected and frozen at −80°C for identification and subsequent analysis.

Determination of Flavonoid Metabolites by UPLC-QqQ/MS

Ground fresh hairy roots were accurately weighed and 0.1 g were extracted in 500 µl 70% methanol for 2 h at 4°C. The extract was sonicated for 30 min and centrifuged at 12,000 rpm for 10 min at 4°C, then filtered through a 0.22-µm hydrophilic organic nylon microporous membrane (SCAA-104). The extracted samples were analyzed by Agilent UPLC 1290II-G6400 QqQ MS (Agilent Technologies, Santa Clara, CA, United States) following the method published by Yang et al. (2020).

Mutant Analysis

Genomic DNA was extracted from T0 transgenic hairy roots using the DNAsecure Plant Kit (TianGen Biotech Co., Ltd., Beijing, China). Positive transgenic hairy roots were verified by PCR, using primers specific from the kanamycin resistance gene (Kan-F and Kan-R; **Supplementary Table S1**). The primers C45-F and C45-R were used to amplify the sgRNA region. The PCR products were sequenced directly by C45-F and C45-R. The sequencing chromatograms were decoded using the Degenerate Sequence Decoding method (DSDecode) and predicted the mutant types (Liu et al., 2015; Ma et al., 2015a). To accurately identify the mutation types, the PCR fragment was purified and cloned into pEASY-T1 cloning vector (TransGen Biotech, Beijing, China), then the ligated product was identified by PCR and sequenced by Sanger sequencing. For each mutant line, at least 10 positive colonies were randomly selected and sequenced. The sequence alignment and mutation analysis were performed using the DNAMAN software (Version 4.0; Lynnon Corporation, Canada).

Data Statistical Analysis

Student's *t*-test and one-way analysis of the variance (ANOVA) were performed using GraphPad Prism8.0.1. p values <0.01 is considered statistically significant.

RESULTS

FtMYB45 Is Repressed by UV-B and Inhibits Flavonoid Biosynthesis

Ultraviolet-B (UV-B) is an important environmental signal that regulates plant growth and development. Previous studies have shown that UV-B can induce the key genes in the flavonoid

[1]http://crispor.tefor.net/

[2]http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3

biosynthetic pathway and increased the accumulation of flavonoids in *Ginkgo biloba* (Zhao et al., 2020), strawberry (Warner et al., 2021), apple (Hu et al., 2020), blueberry (Li et al., 2021), and other plant species (Suzuki et al., 2005; Huang et al., 2016). To investigate the effect of UV-B on the accumulation of flavonoids in Tartary buckwheat, 5-day-old seedlings were treated with UV-B light. Liquid chromatography–mass spectrometry (LC-MS) analyses showed that the content of rutin, epicatechin, and catechin were significantly increased after UV-B irradiation ($p < 0.01$; **Figure 1A**), indicating that UV-B irradiation promoted the accumulation of these flavonoids in Tartary buckwheat. Previous study revealed that the *FtMYB45* gene was induced by MeJA and repressed rutin biosynthesis (Zhang et al., 2018). Our qRT-PCR result indicated that the *FtMYB45* expression in seedlings was significantly downregulated ($p < 0.01$) after UV-B treatment compared with the control cultured in the dark (**Figure 1B**). It implied that *FtMYB45* was also repressed by UV-B treatment and inhibited flavonoid biosynthesis in Tartary buckwheat. Based on this result, *FtMYB45* was selected as the target gene for the development of a CRISPR/Cas9 workflow in Tartary buckwheat.

SgRNA Design and PTG/Cas9-FtMYB45 Vector Construction

The *FtMYB45* gene is located in chromosome 5 and is 1,145 bp in size, with two exons. Two sgRNAs targeting exon 2 of the *FtMYB45* gene were designed (**Figure 2A**), the GC content of sgRNA1 and sgRNA2 was 52.17 and 47.83%, respectively. On-target and off-target the designed sgRNAs were analyzed by the CRISPROR tool. The cutting frequency determination (CFD) score is widely used to measure sgRNA on-target specificity, and a high CFD specificity score indicates high sgRNA specificity (Doench et al., 2016). The result indicated the CFD score of sgRNA1 was 99 with 0 off-target within four mismatch bases, and the CFD score of sgRNA2 was 98 with only 1 off-target within four mismatch bases. Thus, both sgRNA1 and sgRNA2 were highly specific.

To verify the accuracy of the sgRNA sequences in Jinqiao No.2, the sgRNA region was amplified using the specific primer pair C45-F and C45-R and sequenced. The results showed that the sequences of sgRNA1 and sgRNA2 in Jinqiao No. 2 were 100% matched to the reference sequences (**Supplementary Text S1**). The PTG/Cas9-FtMYB45 vector was constructed according to the method published by Wang et al. (2018). In this vector, the spCas9 expression cassette was driven by the CaMV 35S promoter, and the polycistronic tRNA-sgRNA cassette (PTG) was driven by the AtU6-1 promoter (**Figure 2B**). After validation of the construct sequence by Sanger sequencing, the PTG/Cas9-FtMYB45 was introduced into *A. rhizogenes ACCC10060* cells for the transformation of *F. tataricum*.

Targeted Mutagenesis of *FtMYB45* Gene Using the PTG/Cas9 System

Since the plant regeneration and genetic transformation have not yet been refined, *A. rhizogene*-mediated hairy root transformation is still the main method for genetic transformation in Tartary buckwheat. Our data indicated that *FtMYB45* was expressed in different organs of Tartary buckwheat, and was also expressed in hairy root (**Supplementary Figure S1**).

Therefore, *A. rhizogene* strain *ACCC10060* harboring the PTG/Cas9-FtMYB45 vector was transformed into Tartary buckwheat explants to induce hairy roots. Twelve transgenic hairy roots were obtained according to the PCR detection of the kanamycin resistance gene (neomycin phosphotransferase gene, nptII) using the primer pair Kan-F and Kan-R (**Figure 2C**). The sgRNA target region was amplified from the transgenic hairy roots using the specific primer pair C45-F and C45-R and sequenced to analyze *FtMYB45* mutations. The direct Sanger sequencing chromatograms were decoded by DSDecode (**Supplementary Figure S2**) and the mutation types were further

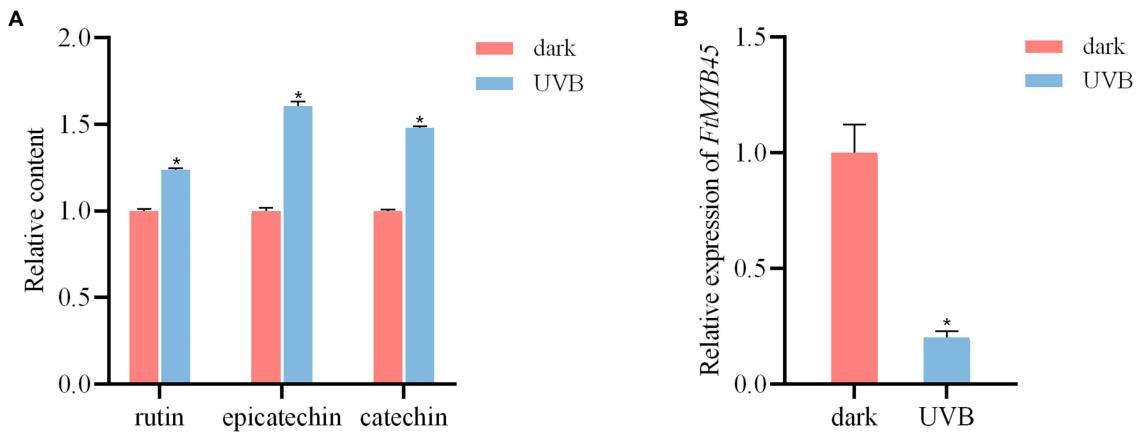


FIGURE 1 | UV-B irradiation enhance the contents of flavonoids and reduce *FtMYB45* gene expression in Tartary buckwheat seedlings. **(A)** Content changes of rutin, epicatechin, and catechin in Tartary buckwheat seedlings after 6 h of UV-B treatment. **(B)** Changes of *FtMYB45* gene expression after 6 h of UV-B treatment. The values represent the means ± standard deviations (SDs) of three biological replicates. Asterisks indicate statistically significant differences compared with control seedlings under dark (*$p < 0.01$, Student's *t*-test).

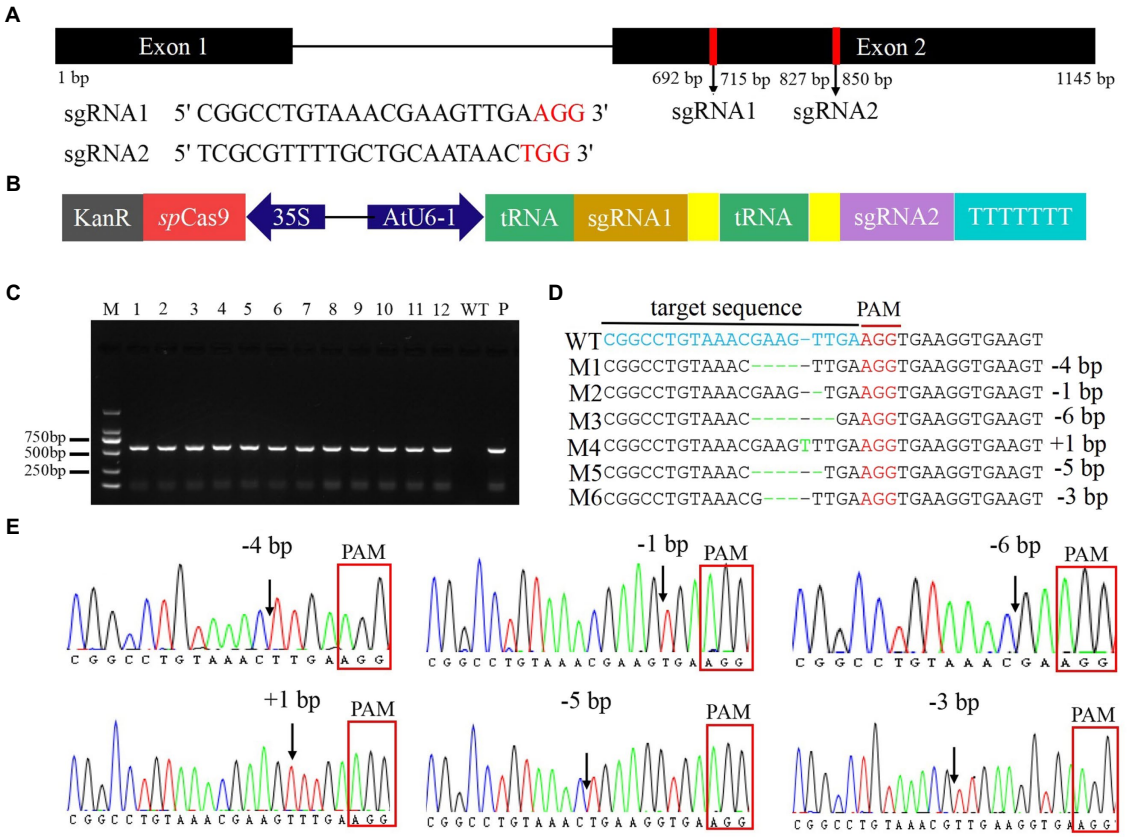


FIGURE 2 | Construction of the PTG/Cas9-FtMYB45 vector and targeted modification of the *FtMYB45* gene. **(A)** Schematic illustration of the two sgRNAs target sites in the *FtMYB45* gene. The black rectangles represent exons, the black line represents the intron, and the numbers below represent the number of bases. The red vertical bars represent the locations of the sgRNA1 and sgRNA2. The red letters represent the protospacer adjacent motif (PAM) of each sgRNA. **(B)** Schematic diagram of the PTG/Cas9-FtMYB45 vector. The spCas9 expression cassette was driven by the CaMV 35S promoter, and the polycistronic tRNA-sgRNA cassette (PTG) was driven by the AtU6-1 promoter. The yellow rectangles represent the sgRNA scaffold. **(C)** Identification of the transgenic hairy root line by PCR amplification of the kanamycin resistance gene. The length of the PCR product was 563 bp. M represents the DNA marker DL2000. Lines 1–12 are individual hairy root lines, WT, wild type; P, positive control. **(D)** Mutation types induced by sgRNA1 in *FtMYB45*. The blue letters represent the sgRNA1 target sequence. The red letters represent the PAM sequence. The green letter represents the nucleotide insertion and the green dashes represent the nucleotide deletions. M1–M6 on the left side represent the mutation types. WT, wild type; +, insertion; –, deletion. **(E)** The sequencing chromatograms of mutation types of *FtMYB45*. The black arrowheads represent the locations of mutations. The red rectangles represent the PAM sequence.

genotyped by cloning and Sanger sequencing. The results showed that six hairy root lines (45–12, 45–13, 45–14, 45–17, 45–18, and 45–19) presented mutations at the target sites of sgRNA1 (**Table 1**; **Supplementary Table S2**), and the editing efficiency reached 50%. Sequence alignment revealed that there were six types of mutations (named M1–M6; **Figure 2D**), including insertion and deletion (**Figure 2E**). Among them, line 45–12 was a chimeric mutant, line 45–13, 45–14, 45–17, and 45–18 were biallelic mutants and line 45–19 was heterozygous mutant (**Table 1**). Unfortunately, no mutations were detected at the target sites of sgRNA2.

Changes in Flavonoids Content in *FtMYB45* Mutants

UPLC-QqQ-MS metabolomics analysis (Yang et al., 2020) was used to determine the changes in flavonoids content in *FtMYB45* mutant hairy roots compared with that in the control line. The main 10 flavonoids in hairy roots including naringenin-7-O-glucoside, kaempferol-3-O-rutinoside, kaempferol-3-O-β-D-glucoside, methylquercetin-O-hexose, methylquercetin-O-rutinoside, rutin, epicatechin, catechin, epicatechin-3-O-glucoside, and catechin-3-O-glucoside were detected (The chromatograms are shown in **Supplementary Figure S3**). The content of six flavonoids (kaempferol-3-O-β-D-glucoside, methylquercetin-O-hexose, rutin, catechin, epicatechin-3-O-glucoside, and catechin-3-O-glucoside) were increased in all *FtMYB45* mutant lines, and most of these increases were significant. However, the content changes of naringenin-7-O-glucoside, kaempferol-3-O-rutinoside, and methylquercetin-O-rutinoside in mutant lines were variable, for example, kaempferol-3-O-rutinoside in 45–12 line, naringenin-7-O-glucoside and methylquercetin-O-rutinoside in 45–13 and 45–14 lines, and methylquercetin-O-rutinoside in 45–17 line were significantly decreased, while they were increased in other mutant lines. Moreover, epicatechin levels were slightly diminished in the

TABLE 1 | Mutant genotypes and mutant type by Sanger sequencing analysis.

Mutant line	No. of clone sequenced	WT	Mutant type						Genotype
			M1 (−4 bp)	M2 (−1 bp)	M3 (−6 bp)	M4 (+1 bp)	M5 (−5 bp)	M6 (−3 bp)	
45–12	18		4	9	2	3			Chimeric
45–13	11			8		3			Biallele
45–14	11			9		2			Biallele
45–17	11					5	6		Biallele
45–18	11					6	5		Biallele
45–19	11	7						4	Heterozygote

+, insertion; −, deletion; WT, wild type: reference sequences of sgRNA1 in the FtMYB45 gene.

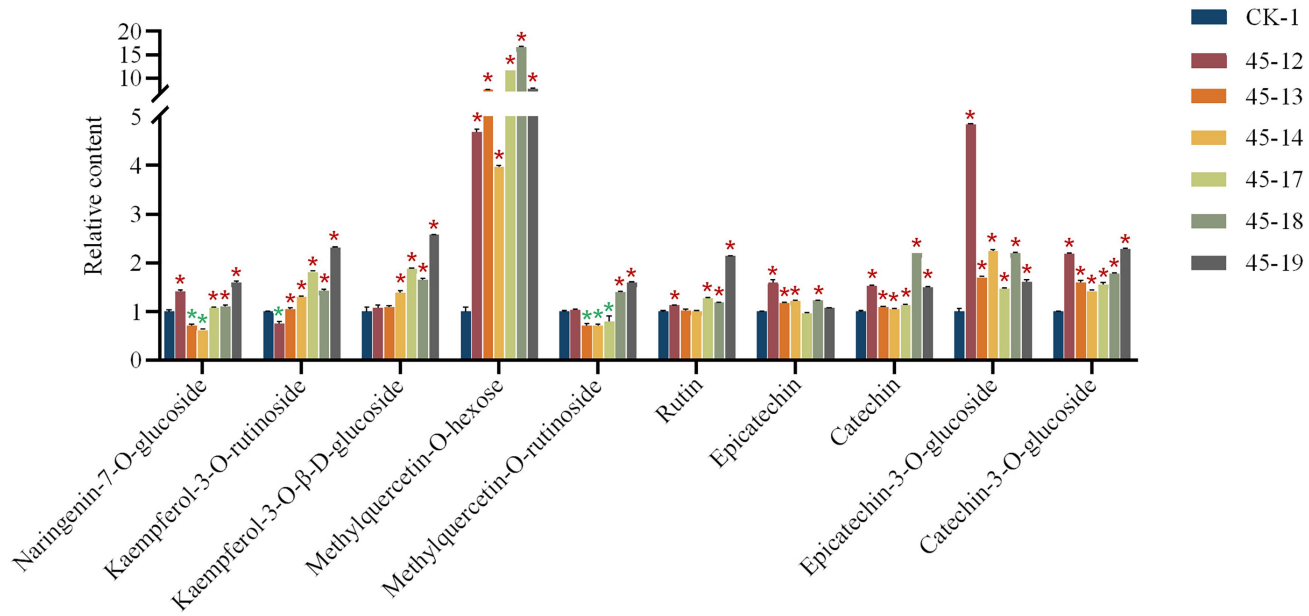


FIGURE 3 | UPLC-QqQ-MS analysis of the changes of flavonoids content in mutant hairy root lines. CK represents hairy root transformed with the empty vector, 45–12, 45–13, 45–14, 45–17, 45–18, and 45–19 are individual mutant hairy root lines. The values represent the means ± standard deviations (SDs) of three biological replicates. Asterisks represent statistically significant differences compared with CK. The red asterisks represent significant increase, and the green asterisks represent significant decrease (*$p < 0.01$, one-way ANOVA).

45–17 and 45–19 (**Figure 3**). Altogether, the data indicated that *FtMYB45* mutation caused an increase of the amount of most flavonoids in hairy roots of Tartary buckwheat, suggesting that *FtMYB45* negatively regulated flavonoid biosynthesis.

DISCUSSION

Tartary buckwheat contains a large amount and variety of flavonoids, making it a popular health food. MYB transcription factors are key regulators of flavonoid biosynthesis in plants (Cao et al., 2020). Here, we focused on the flavonoids present in Tartary buckwheat, which, despite their importance, have not been thoroughly investigated. Previous studies demonstrated that *FtMYB45* was a JA responsive factor that repressed rutin biosynthesis (Zhang et al., 2018). Our study showed that UV-B irradiation significantly decreased the expression level of *FtMYB45* and significantly increased the content of rutin, epicatechin, and catechin in Tartary buckwheat (**Figure 1**), indicating that *FtMYB45* also inhibits flavonoid biosynthesis through UV-B signal transduction. Therefore, *FtMYB45* may be involved in the crosstalk between UV-B and JA signaling pathways, and regulate flavonoid biosynthesis in Tartary buckwheat. This finding provides new insight into the function of *FtMYB45* in Tartary buckwheat and might be of importance for the culture of Tartary buckwheat with high content in flavonoid metabolites.

The CRISPR/Cas9 gene editing is a fast, simple, efficient, and flexible technique for gene function analysis and crop improvement (Gupta et al., 2019; Triozzi et al., 2021). It has been widely used in a variety of plants (Jinek et al., 2012; Feng et al., 2013), and has also been applied in medicinal plants (Li et al., 2017; Feng et al., 2018, 2021). However, the

application of CRISPR/Cas9 technology in Tartary buckwheat has not been reported yet. In this study, the *FtMYB45* was selected as the target gene to test the CRISPR/Cas9 system in Tartary buckwheat. Two sgRNAs of *FtMYB45* were designed to ensure efficient knockout of *FtMYB45*, and the off-target analysis indicated that sgRNA1 and sgRNA2 were highly specific. The PTG/Cas9-FtMYB45 vector was transformed into Tartary buckwheat using *A. rhizogenes* to induce transgenic hairy roots. Twelve transgenic hairy roots were obtained. Sequencing analyses showed successful gene editing in the region targeted by sgRNA1 in six hairy root lines, with the editing efficiency reaching 50%. A total of six types of mutations, including base insertions and deletions, were detected at the target site (**Figure 2**). Therefore, we successfully knocked out the *FtMYB45* gene in Tartary buckwheat using the PTG/Cas9 system, and consequently provide a new tool for gene function research and genetic improvement in Tartary buckwheat.

Unfortunately, no mutation was observed at the target sites of sgRNA2. The GC content of sgRNA has been considered as one of the key factors affecting sgRNA editing efficiency. Previous reports have shown that 97% of sgRNAs which have been experimentally validated in plants have a GC content between 30 and 80% (Liang et al., 2016). In our study, the GC content of sgRNA1 and sgRNA2 was 52.17 and 47.83%, respectively. Thus, GC content may not be the reason why sgRNA2 does not edit. Another main reason affecting sgRNA activity is the secondary structure of sgRNAs (Makarova et al., 2011). Assessment of the secondary structures of the studied sgRNA1 and sgRNA2 (**Supplementary Figure S4**) found that sgRNA2 formed an 8 bp typical stem-loop structure. Ma et al. (2015b) also reported an inactive sgRNA formed a stem-loop structure with a pairing of continuous 14 and 4 bp of the target, and suggested the sgRNA selection should avoid those with pairing to the sgRNA by more than continuous 6 bp. Thus, the continuous 8 bp stem-loop structure might inhibit the binding of the sgRNA2 to the target strand, leading to the failure of gene editing.

Hairy root cultures established by transforming plants with *A. rhizogenes* have been utilized to produce transgenic plants, investigate plant metabolic processes, and increase secondary metabolites. They are genetically and biochemically stable during rapid growth (Guillon et al., 2006a,b). Recently, the hairy root transformation has been widely utilized to validate and optimize induced mutagenesis by the CRISPR/Cas9 system (Li et al., 2019; Le et al., 2020). In addition, biotechnological approaches which used hairy root culture have greatly enhanced the production of rutin by common buckwheat (Lee et al., 2007; Kim et al., 2010). Therefore, hairy root cultures have been used as a useful model system to study the production of flavonoids and a variety of other secondary metabolites. Now, the flavonoid biosynthesis pathway is relatively clear (**Supplementary Figure S5**; Falcone Ferreyra et al., 2012; Dong and Lin, 2021). Dihydroflavonols are precursors used for flavonoid biosynthesis. Flavonol synthase (FLS) links flavonoids and flavonols synthesis pathways and is involved in dihydroflavonol desaturation to form flavonols (Forkmann and Martens, 2001). Dihydroflavonol reductase (DFR) is a key enzyme and an important branch point in the synthesis pathway of anthocyanins and catechins (Landry et al., 1995). In this study, we detected 10 flavonoids including flavonols, flavanols, and their glycosides in the obtained *FtMYB45* mutant hairy root lines. The UPLC-QqQ-MS result showed that the content of most of these flavonoids was significantly increased in mutant lines. In particular, the content of methylquercetin-O-hexose, proanthocyanidins, including catechin, epicatechin-3-O-glucoside, and catechin-3-O-glucoside, were greatly increased in all six mutant lines. Thus, our data demonstrated that *FtMYB45* is a negative regulator of flavonoid biosynthesis. This is consistent with the previous report showing that *FtMYB45* directly represses phenylalanine ammonia-lyase (*FtPAL*) gene expression, and thus affecting the entire flavonoid metabolic pathway (Zhang et al., 2018). Moreover, flavonoids content among the six mutant lines showed different change levels, which may be due to different mutations types of *FtMYB45*. The mutant lines with the same genotype showed similar content changes in detected flavonoids, for example, line 45–13 and 45–14, line 45–17, and 45–18 (**Figure 3**). However, the increase of some flavonoids in heterozygote mutant line 45–19 was greater than that of biallelic or chimeric mutant lines, which does not meet our expectations. The possible reason we suppose is that the Transfer DNA (T-DNA) insertion in line 45–19 may affect the related genes in flavonoid biosynthesis. We also noticed that the content of some flavonoids showed decreased in few mutant lines, for example, kaempferol-3-O-rutinoside in line 45–12, and naringenin-7-O-glucoside in line 45–13 and 45–14. The reason is still not clear and needs to be further studied.

Taken together, our results indicated that the application of the PTG/Cas9 gene editing system effectively knocked out *FtMYB45* and increased the content of flavonoids in mutant hairy roots in Tartary buckwheat. These *FtMYB45* mutant hairy root lines will be good candidate biomaterials for the production of flavonoids.

CONCLUSION

In this study, the PTG/Cas9 genome editing system was successfully utilized for genome editing in Tartary buckwheat, which lays a valuable foundation for the application of CRISPR/Cas9 technology in gene function study and molecular breeding in Tartary buckwheat. Additionally, we performed targeted mutagenesis of the *FtMYB45* gene, which resulted in an increased content of flavonoids in mutant hairy roots of Tartary buckwheat. This finding provides further evidence to support the negative regulatory role of the *FtMYB45* gene in the flavonoid biosynthetic pathway, and the obtained mutant hairy root lines with increased amounts of flavonoids will provide good sources for the production of flavonoids.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

DW performed the experiments, analyzed the data, and wrote the paper. MW and XW performed part of the hairy root transformation experiment. LW and WY analyzed part of the data. WM, WS, and SC revised the paper. YS and LX initiated and supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the CACMS Innovation Fund (CI2021A04107), the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZ13-YQ-101), National Key R&D Program of China from the Ministry of Science and Technology of China (2019YFC1711100), and the Agilent Thought Leader Program and ACT-UR Program.

ACKNOWLEDGMENTS

We thank Qingfu Chen for providing Tartary buckwheat seeds of Jinqiao No. 2 and thank Yifei Liu for providing the PTG/Cas9 vector system.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.879390/full#supplementary-material

REFERENCES

Appiani, M., Rabitti, N. S., Proserpio, C., Pagliarini, E., and Laureati, M. (2021). Tartary buckwheat: a new plant-based ingredient to enrich corn-based gluten-free formulations. *Foods* 10, 2613–2627. doi: 10.3390/foods10112613

Bai, Y. C., Li, C. L., Zhang, J. W., Li, S. J., Luo, X. P., Yao, H. P., et al. (2014). Characterization of two tartary buckwheat R2R3-MYB transcription factors and their regulation of proanthocyanidin biosynthesis. *Physiol. Plant.* 152, 431–440. doi: 10.1111/ppl.12199

Cai, Y. P., Chen, L., Liu, X. J., Guo, C., Sun, S., Wu, C., et al. (2018). CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soyabean. *Plant Biotechnol. J.* 16, 176–185. doi: 10.1111/pbi.12758

Cao, Y. P., Li, K., Li, Y. L., Zhao, X. P., and Wang, L. H. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biology* 9:61. doi: 10.3390/biology9030061

Concordet, J. P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. doi: 10.1093/nar/gky354

Ding, K., Pei, T. L., Bai, Z. P., Jia, Y. Y., Ma, P. D., and Liang, Z. (2017). SmMYB36, a novel R2R3-MYB transcription factor, enhances tanshinone accumulation and decreases phenolic acid content in *Salvia miltiorrhiza* hairy roots. *Sci. Rep.* 7:5104. doi: 10.1038/s41598-017-04909-w

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437

Dong, N. Q., and Lin, H. X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J. Integr. Plant Biol.* 63, 180–209. doi: 10.1111/jipb.13054

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

Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222

Feng, Y., Qinsong, Y., Yuhao, G., Yunjing, M., Ying, X., Yuanwen, T., et al. (2021). Establishment of dual-cut CRISPR/Cas9 gene editing system in pear establishment of dual-cut CRISPR/Cas9 gene editing system in pear calli. *Yuan Yi Xue Bao* 5, 873–882. doi: 10.16420/j.issn.0513-353x.2020-0829

Feng, S., Song, W., Fu, R. R., Zhang, H., Xu, A. R., and Li, J. (2018). Application of the CRISPR/Cas9 system in *Dioscorea zingiberensis*. *Plant Cell Tissue Organ Cult.* 135, 133–141. doi: 10.1007/s11240-018-1450-5

Feng, Z. Y., Zhang, B. T., Ding, W. N., Liu, X. D., Yang, D. L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* 23, 1229–1232. doi: 10.1038/cr.2013.114

Forkmann, G., and Martens, S. F. (2001). Metabolic engineering and applications of flavonoids. *Curr. Opin. Biotechnol.* 12, 155–160. doi: 10.1016/S0958-16690000192-0

Fornalé, S., Lopez, E., Salazar-Henao, J. E., Fernández-Nohales, P., Rigau, J., and Caparros-Ruiz, D. (2014). AtMYB7, a new player in the regulation of UV-sunscreens in *Arabidopsis thaliana*. *Plant Cell Physiol.* 55, 507–516. doi: 10.1093/pcp/pct187

Gocal, G. F. W., Sheldon, C. C., Gubler, F., Moritz, T., Bagnall, D. J., MacMillan, C. P., et al. (2001). GaMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. *Plant Physiol.* 127, 1682–1693. doi: 10.1104/pp.010442

Guillon, S. H., Tremouillaux-Guiller, J., Pati, P. K., Rideau, M., and Gantet, P. (2006a). Hairy root research: recent scenario and exciting prospects. *Curr. Opin. Plant Biol.* 9, 341–346. doi: 10.1016/j.pbi.2006.03.008

Guillon, S. H., Tremouillaux-Guiller, J., Pati, P. K., Rideau, M., and Gantet, P. (2006b). Harnessing the potential of hairy roots: dawn of a new era. *Trends Biotechnol.* 24, 403–409. doi: 10.1016/j.tibtech.2006.07.002

Guo, X. D., Ma, Y. J., Parry, J., Gao, J. M., Yu, L. L., and Wang, M. (2011). Phenolics content and antioxidant activity of Tartary buckwheat from different locations. *Molecules* 16, 9850–9867. doi: 10.3390/molecules16129850

Gupta, D., Bhattacharjee, O., Mandal, D., Sen, M. K., Dey, D., Dasgupta, A., et al. (2019). CRISPR-Cas9 system: a new-fangled dawn in gene editing. *Life Sci.* 232, 116636–116651. doi: 10.1016/j.lfs.2019.116636

Hu, J. F., Fang, H. C., Wang, J., Yue, X. X., Su, M. Y., Mao, Z., et al. (2020). Ultraviolet B-induced MdWRKY72 expression promotes anthocyanin synthesis in apple. *Plant Sci.* 292:110377. doi: 10.1016/j.plantsci.2019.110377

Huang, X., Yao, J. W., Zhao, Y. Y., Xie, D. F., Jiang, X., and Xu, Z. (2016). Efficient rutin and quercetin biosynthesis through flavonoids-related gene expression in *Fagopyrum tataricum* Gaertn. Hairy root cultures with UV-B irradiation. *Front. Plant Sci.* 7:63. doi: 10.3389/fpls.2016.00063

Jiang, P., Burczynski, F., Campbell, C., Pierce, G., Austria, J. A., and Briggs, C. J. (2007). Rutin and flavonoid contents in three buckwheat species *Fagopyrum esculentum*, *F. tataricum*, and *F. homotropicum* and their protective effects against lipid peroxidation. *Food Res. Int.* 40, 356–364. doi: 10.1016/j.foodres.2006.10.009

Jiang, W. Z., Yang, B., and Weeks, D. P. (2014). Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and inheritance of modified genes in the T2 and T3 generations. *PLoS One* 9:e99225. doi: 10.1371/journal.pone.0099225

Jin, H. L., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., et al. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *EMBO J.* 19, 6150–6161. doi: 10.1093/emboj/19.22.6150

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Kim, Y. K., Xu, H., Park, W. T., Park, N. I., Lee, S. Y., and Park, S. U. (2010). Genetic transformation of buckwheat (*Fagopyrum esculentum* M.) with

Agrobacterium rhizogenes and production of rutin in transformed root cultures. *Aust. J. Crop. Sci.* 4, 485–490. doi: 10.1055/s-0028-1084808

Landry, L. G., Chapple, C. C., and Last, R. L. (1995). Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol.* 109, 1159–1166. doi: 10.1104/pp.109.4.1159

Le, H., Nguyen, N. H., Ta, D. T., Le, T. N. T., Bui, T. P., Le, N. T., et al. (2020). CRISPR/Cas9-mediated knockout of Galactinol synthase-encoding genes reduces raffinose family oligosaccharide levels in soybean seeds. *Front. Plant Sci.* 11:612942. doi: 10.3389/fpls.2020.612942

Lee, S. Y., Cho, S. I., Park, M. H., Kim, Y. Y., Choi, J. E., and Park, S. U. (2007). Growth and rutin production in hairy root cultures of buckwheat (*Fagopyrum esculentum* M.). *Prep. Biochem. Biotechnol.* 37, 239–246. doi: 10.1080/10826060701386729

Li, B., Cui, G. H., Shen, G. A., Zhan, Z., Huang, L. Q., Chen, J., et al. (2017). Targeted mutagenesis in the medicinal plant *Salvia miltiorrhiza*. *Sci. Rep.* 7:43320. doi: 10.1038/srep43320

Li, C. L., Nguyen, V., Liu, J., Fu, W. Q., Chen, C., Yu, K. F., et al. (2019). Mutagenesis of seed storage protein genes in soybean using CRISPR/Cas9. *BMC. Res. Notes* 12, 176–182. doi: 10.1186/s13104-019-4207-2

Li, T. S., Yamane, H., and Tao, R. (2021). Preharvest long-term exposure to UV-B radiation promotes fruit ripening and modifies stage-specific anthocyanin metabolism in highbush blueberry. *Hortic. Res.* 8, 67–78. doi: 10.1038/s41438-021-00503-4

Liang, G., Zhang, H. M., Lou, D. J., and Yu, D. Q. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci. Rep.* 6, 21451–21458. doi: 10.1038/srep21451

Liu, W. Z., Xie, X. R., Ma, X. L., Li, J., Chen, J. H., and Liu, Y. G. (2015). DSDecode: a web-based tool for decoding of sequencing chromatograms for genotyping of targeted mutations. *Mol. Plant* 8, 1431–1433. doi: 10.1016/j.molp.2015.05.009

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

Ma, X. L., Chen, L. T., Zhu, Q. L., Chen, Y. L., and Liu, Y. G. (2015a). Rapid decoding of sequence-specific nuclease-induced heterozygous and biallelic mutations by direct sequencing of PCR products. *Mol. Plant* 8, 1285–1287. doi: 10.1016/j.molp.2015.02.012

Ma, X. L., Zhang, Q. Y., Zhu, Q. L., Liu, W., Chen, Y., Qiu, R., et al. (2015b). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas9 systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577

Martínez Conesa, C., Vicente Ortega, V., Yáñez Gascón, M. J., Alcaraz Baños, M., Canteras Jordana, M., Benavente-García, O., et al. (2005). Treatment of metastatic melanoma B16F10 by the flavonoids tangeretin, rutin, and diosmin. *J. Agric. Food Chem.* 53, 6791–6797. doi: 10.1021/jf058050g

Md, R. U., Li, X. H., Kim, Y. B., Chae, S. C., Kim, S. J., and Park, S. U. (2013). Phenolic compounds in different organs of tartary buckwheat (*Fagopyrum tataricum* Gaertn.) cultivars. *Aust. J. Crop. Sci.* 7, 1861–1865. doi: 10.3316/informit.677888683359450

Mi, Y. L., Zhu, Z. H., Qian, G. T., Li, Y., Meng, X. X., Xue, J., et al. (2020). Inducing hairy roots by *Agrobacterium rhizogenes*-mediated transformation in tartary buckwheat *Fagopyrum tataricum*. *J. Vis. Exp.* 157:e60828. doi: 10.3791/60828

Preston, J., Wheeler, J., Heazlewood, J., Li, S. F., and Parish, R. W. (2004). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Plant J.* 40, 979–995. doi: 10.1111/j.1365-313X.2004.02280.x

Ren, C., Liu, Y. F., Guo, Y. C., Duan, W., Fan, P. G., Li, S., et al. (2021). Optimizing the CRISPR/Cas9 system for genome editing in grape by using grape promoters. *Hortic. Res.* 8:52. doi: 10.1038/s41438-021-00489-z

Shan, Q. W., Zhang, Y., Chen, K. L., Zhang, K., and Gao, C. X. (2015). Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol. J.* 13, 791–800. doi: 10.1111/pbi.12312

Shin, H. Y., Wang, C. C., Lee, H. K., Yoo, K. H., Zeng, X. K., Kuhns, T., et al. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.* 8:15464. doi: 10.1038/ncomms15464

Srivastava, V., Underwood, J. L., and Zhao, S. (2017). Dual-targeting by CRISPR/Cas9 for precise excision of transgenes from rice genome. *Plant Cell Tissue Organ Cult.* 129, 153–160. doi: 10.1007/s11240-016-1166-3

Suzuki, T., Honda, Y., and Mukasa, Y. (2005). Effects of UV-B radiation, cold and desiccation stress on rutin concentration and rutin glucosidase activity in tartary buckwheat (*Fagopyrum tataricum*) leaves. *Plant Sci.* 168, 1303–1307. doi: 10.1016/j.plantsci.2005.01.007

Tang, X. M., Chen, S. L., Yu, H. W., Zheng, X. J., Zhang, F., Deng, X., et al. (2021). Development of a gRNA–tRNA array of CRISPR/Cas9 in combination with grafting technique to improve gene-editing efficiency of sweet orange. *Plant Cell Rep.* 40, 2453–2456. doi: 10.1007/s00299-021-02781-7

Tomotake, H., Yamamoto, N., Kitabayashi, H., Kawakami, A., Kayashita, J., Ohinata, H., et al. (2007). Preparation of tartary buckwheat protein product and its improving effect on cholesterol metabolism in rats and mice fed cholesterol-enriched diet. *J. Food Sci.* 72, S528–S533. doi: 10.1111/j.1750-3841.2007.00474.x

Triozzi, P. M., Schmidt, H. W., Dervinis, C., Kirst, M., and Conde, D. (2021). Simple, efficient and open-source CRISPR/Cas9 strategy for multi-site genome editing in *Populus tremula* × *alba*. *Tree Physiol.* 41, 2216–2227. doi: 10.1093/treephys/tpab066

Wang, Y. J., and Campbell, C. G. (2007). Tartary buckwheat breeding *Fagopyrum tataricum* L. Gaertn. through hybridization with its Rice-Tartary type. *Euphytica* 156, 399–405. doi: 10.1007/s10681-007-9389-3

Wang, J. L., Tang, M. Q., Chen, S., Zheng, X. F., Mo, H. X., Li, S. J., et al. (2017). Down-regulation of BnDA1, whose gene locus is associated with the seeds weight, improves the seeds weight and organ size in *Brassica napus*. *Plant Biotechnol. J.* 15, 1024–1033. doi: 10.1111/pbi.12696

Wang, Z. P., Wang, S. B., Li, D. W., Zhang, Q., Li, L., Zhong, C., et al. (2018). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. *Plant Biotechnol. J.* 16, 1424–1433. doi: 10.1111/pbi.12884

Wang, S. H., Zhang, S. B., Wang, W. X., Xiong, X. Y., Meng, F. R., and Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Rep.* 34, 1473–1476. doi: 10.1007/s00299-015-1816-7

Warner, R., Wu, B., MacPherson, S., and Lefsrud, M. (2021). A review of strawberry photobiology and fruit flavonoids in controlled environments. *Front. Plant Sci.* 12:611893. doi: 10.3389/fpls.2021.611893

Xie, K. B., Minkenberg, B., and Yang, Y. N. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3570–3575. doi: 10.1073/pnas.1420294112

Yang, W., Su, Y., Dong, G. Q., Qian, G. T., Shi, Y. H., Mi, Y., et al. (2020). Liquid chromatography-mass spectrometry-based metabolomics analysis of flavonoids and anthraquinones in *Fagopyrum tataricum* L. Gaertn. Tartary buckwheat seeds to trace morphological variations. *Food Chem.* 331:127354. doi: 10.1016/j.foodchem.2020.127354

Yao, P. F., Huang, Y. J., Dong, Q. X., Wan, M., Wang, A. H., Chen, Y., et al. (2020). FtMYB6, a light-induced SG7 R2R3-MYB transcription factor, promotes flavonol biosynthesis in tartary buckwheat (*Fagopyrum tataricum*). *J. Agric. Food Chem.* 68, 13685–13696. doi: 10.1021/acs.jafc.0c03037

Zhang, K. X., Logacheva, M. D., Meng, Y., Hu, J. P., Wan, D. P., Li, L., et al. (2018). Jasmonate-responsive MYB factors spatially repress rutin biosynthesis in *Fagopyrum tataricum*. *J. Exp. Bot.* 69, 1955–1966. doi: 10.1093/jxb/ery032

Zhang, S. C., Ma, P. D., Yang, D. F., Li, W. J., Liang, Z. S., Liu, Y., et al. (2013). Cloning and characterization of a putative R2R3 MYB transcriptional repressor of the rosmarinic acid biosynthetic pathway from *Salvia miltiorrhiza*. *PLoS One* 8:e73259. doi: 10.1371/journal.pone.0073259

Zhao, B. B., Wang, L., Pang, S. Y., Jia, Z. C., Wang, L., Li, W., et al. (2020). UV-B promotes flavonoid synthesis in *Ginkgo biloba* leaves. *Ind. Crop. Prod.* 151:112483. doi: 10.1016/j.indcrop.2020.112483

Zhao, G., Zhao, J. L., Peng, L. X., Zou, L., Wang, J. B., Zhong, L., et al. (2012). Effects of yeast polysaccharide on growth and flavonoid accumulation in *Fagopyrum tataricum* sprout cultures. *Molecules* 17, 11335–11345. doi: 10.3390/molecules171011335

Zhou, M. L., Sun, Z. M., Ding, M. Q., Logacheva, M. D., Kreft, I., Wang, D., et al. (2017). FtSAD2 and FtJAZ1 regulate activity of the *FtMYB11* transcription repressor of the phenylpropanoid pathway in *Fagopyrum tataricum*. *New Phytol.* 216, 814–828. doi: 10.1111/nph.14692

Zhou, M. L., Wang, C. L., Qi, L. P., Yang, X. B., Sun, Z. M., Tang, Y., et al. (2015). Ectopic expression of *Fagopyrum tataricum* FtMYB12 improves cold tolerance in *Arabidopsis thaliana*. *J. Plant Growth Regul.* 34, 362–371. doi: 10.1007/s00344-014-9472-7

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.

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 May 2022
doi: 10.3389/fpls.2022.868472

Transcriptome and Metabonomics Analysis Revealed the Molecular Mechanism of Differential Metabolite Production of *Dendrobium nobile* Under Different Epiphytic Patterns

Qingqing Li[1†], Chaobo Liu[1†], Ceyin Huang[1], Mufei Wang[1], Teng Long[1], Jingyi Liu[1], Junhua Shi[2], Junli Shi[2], Lin Li[1], Yuqi He[3] and De-Lin Xu[1]**

[1] Department of Medical Cell Biology, Zunyi Medical University, Zunyi, China, [2] Affiliated Hospital of Zunyi Medical University, Zunyi, China, [3] School of Medicine, Zunyi Medical University, Zunyi, China

The cultivation medium of *Dendrobium nobile* has an effect on the contents of its main medicinal components, but the specific mechanism is still unclear. In this study, the callus, seedlings, rhizomes, and leaves of *D. nobile* were sequenced for the PacBio SMRT. The 2-year-old stems were selected for the Illumina sequencing and metabolome sequencing to analyze the genetic mechanism of metabolic differences under different epiphytic patterns. As a result, a total of 387 differential genes were obtained, corresponding to 66 differential metabolites. Different epiphytic patterns can induce a series of metabolic changes at the metabolome and transcriptome levels of *D. nobile*, including flavonoid metabolism, purine metabolism, terpenoid backbone biosynthesis, amino acid metabolism, and alpha-linolenic acid metabolic, and related regulatory genes include *ALDH2B7, ADC, EPSPS-1, SHKA, DHAPS-1, GES, ACS1, SAHH, ACS2, CHLP, LOX2, LOX2.3,* and *CYP74B2*. The results showed that the genetic mechanism of *D. nobile* under various epiphytic patterns was different. In theory, the content of metabolites under the epiphytic patterns of Danxia stone is higher, which is more suitable for field cultivation.

Keywords: ***Dendrobium nobile*** **(Shihu), epiphytic patterns, metabonomic, transcriptome, molecular mechanism**

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Lei Zhang,
Second Military Medical University, China
Weidong Zhao,
China Medical University, China

****Correspondence:***
Yuqi He
184860466@qq.com
De-Lin Xu
xudelin2000@163.com
orcid.org/0000-0003-3695-2997

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *02 February 2022*
Accepted: *30 March 2022*
Published: *17 May 2022*

Citation:
Li Q, Liu C, Huang C, Wang M, Long T, Liu J, Shi J, Shi J, Li L, He Y and Xu D-L (2022) Transcriptome and Metabonomics Analysis Revealed the Molecular Mechanism of Differential Metabolite Production of Dendrobium nobile Under Different Epiphytic Patterns. Front. Plant Sci. 13:868472. doi: 10.3389/fpls.2022.868472

INTRODUCTION

As a perennial herb of *Dendrobium* of Orchidaceae, *Dendrobium nobile* Lindl. contains alkaloids, terpenoids, phenols, flavonoids, dibenzyl, and glycosides with a high medicinal value (Ling-Hu et al., 2021). At present, sesquiterpenes (Wang et al., 2020), phenanthrene (Cheng et al., 2020), dibenzyl derivatives (Zhang et al., 2019), phenanthraquinone (Liu Y. H. et al., 2021), and other compounds have been isolated. Among them, *D. nobile* alkaloid (ADNL) was conducive to the expression of liver glucose and lipid metabolism genes, played a regulatory role in metabolic disorders (Xu et al., 2017), had a neuroprotective effect, and can cure Alzheimer's disease (Li et al., 2022). *D. nobile* polysaccharide had an antiviral effect and had broad development prospects (Li Z. et al., 2020). *D. nobile* sesquiterpenes had physiological activities in neuroprotection,

immune regulation, and antitumor (Ma et al., 2019). Among the *Dendrobium* recorded in the Pharmacopoeia of the People's Republic of China, only *D. nobile* has the ability of treating weakness of the five internal organs, strengthening Yin and benefiting essence, holding thick stomach for a long time, and tonifying the kidney and benefiting power, and the alkaloid content in *D. nobile* was the highest among all herbs of *Dendrobium*, which was the main source of medicinal *Dendrobium* (Wang, 2021). The extract of *D. nobile* can also improve constipation (Wang et al., 2021), and the color of the flower was gorgeous, which had an ornamental value. It had moisturizing, whitening, antiaging, and other effects (Nie et al., 2020), and had good application prospects in the cosmetics industry (Yu et al., 2020). It was an economic plant with medicinal and ornamental value.

The quality of Chinese medicinal materials was affected by cultivation methods and environmental conditions, resulting in different synthesis and accumulation of secondary metabolites, and different types and contents of secondary metabolites are responsible for the differences in the quality of Chinese medicinal materials (Guo et al., 2020). The results of the study by Li Y. et al. (2020) showed that the cultivation area, cultivation substrate, and soil conditions had effects on the quality of wendangshen (Jiao et al., 2021), the accumulation of various components in *Phellinus igniarius* (Li, 2020), the yield and quality of *Paeonia lactiflora* Pall (Meng et al., 2020), and so on. Therefore, to ensure the quality of traditional Chinese medicine, the conditions suitable for the growth of traditional Chinese medicine can be selected, and the growth and secondary metabolism of traditional Chinese medicine can be balanced. *D. nobile* was mainly distributed in Guizhou, Shaanxi, Sichuan, and other places in China. The quality of Chishui and Xishui in Guizhou was the best. Among them, the planting technology of Chishui in Guizhou was the most advanced, mainly cultivated on Danxia stones. The cultivation medium of *D. nobile* had an impact on the content of its main medicinal components (Zhang J. Q. et al., 2020). Therefore, it is of great significance to study the effects of different epiphytic patterns on the secondary metabolites of *D. nobile*.

In recent years, with the development of science and technology, transcriptome sequencing technology (RNA-Seq) (Liu H. et al., 2021) and other high-throughput sequencing technologies are more widely used, mainly for functional gene mining (Shovlin and Tropea, 2018), gene network analysis (Pratapa et al., 2020), genetic mechanism analysis (Xu et al., 2020), and so on. The combination of transcriptome and metabolome analysis can more comprehensively analyze the regulatory mechanism of secondary metabolite biosynthesis (Mao et al., 2021). At present, the transcriptome of *D. nobile* has been studied, but there is no study on the determination of secondary metabolites and genetic mechanism of *D. nobile* under different epiphytic patterns (Li et al., 2017; Wen et al., 2017). Using the combined analysis of transcriptomics and metabolomics, this study compared the differential genes and metabolites of *D. nobile* under the four epiphytic patterns of Danxia stone, Crushed stone, Sawdust, and Stump. Then, association analysis was carried out to explore the genetic mechanism of differences. The results were expected to provide reference for the planting of *D. nobile*.

MATERIALS AND METHODS

Materials

Dendrobium nobile with the same genotype cultivated by Chishui Xintian Company (105°74′E, 28°56′N) was sampled for the following study. In 2018, capsules were collected from *D. nobile* plants in Dendrobium resource Park of Chishui Xintian company. After surface disinfection, callus and seedlings were induced by tissue culture, and then, the seedlings were cultured and planted to form full-grown plants. Notably, 1 g callus, 1 g seedlings, and 1 g rhizomes and leaves of *D. nobile* were randomly selected, each with three replicates. Total RNA was extracted by the Trozil method, and mixed in equal amounts for RNA-Seq sequencing to construct the nucleotide library of the whole organ and tissues of *D. nobile* during the whole growth period.

From the Dendrobium resource Park of Chishui Xintian Company, the 2-year-old stems with the same shape and size on Danxia stone (CSXTDS), Crushed stone (CSXTSS), Sawdust (CSXTJM), and Stump (CSXTSZ) were selected, respectively. Each stem was divided into two parts, and after triturating with liquid nitrogen, 200 mg of each stem was used to extract total RNA. One part was used for Illumina sequencing, and the other part was used to separate secondary metabolites for metabolome sequencing. Each sample had three biological replicates.

Construction of cDNA Library and RNA-Seq Sequencing

Advanced molecular biology equipment should be used to detect the purity, concentration, and integrity of RNA samples after mixing. After the samples pass the test, the full-length cDNA of mRNA was synthesized by SMARTer^{TM} PCR cDNA Synthesis Kit, and was then amplified by PCR. The end of the full-length cDNA was repaired, connected with SMRT dumbbell connector, digested by exonuclease, and then the sequencing library was obtained. After being inspected in the library, the full-length transcriptome was sequenced by PacBio instrument according to the target off-line data. The sequence was divided into full-length sequence and non-full-length sequence according to whether 3 "primer, 5" primer, and PolyA (optional) were present or absent. The full-length sequences from the same transcript should be clustered, high-quality sequences should be extracted, and then the transcript sequence after removing redundancy should be obtained. The nucleotide database of the whole growth period, whole tissues and organs, and the full length of the transcript of *D. nobile* was established.

Illumina Sequencing and Differential Gene Screening

After the total RNA of the samples under different epiphytic patterns passed the quality inspection, the eukaryotic mRNA was enriched using the magnetic beads connected with Oligo(dT).

The extracted mRNA is randomly fragmented into short fragments by Fragmentation Buffer, and a single-strand cDNA is synthesized by using the fragmented mRNA as a template with six-base random primers (Random hexamers). Buffer, dNTPs, RNaseH, and DNA polymerase I were then added for double-stranded cDNA synthesis. The second strand of cDNA was amplified by PCR to obtain the final sequencing library. The library was sequenced by Illumina Nova seq^{TM} 6000 after passing the quality inspection, and the sequencing read length was paired-end 2 × 150 bp (PE150). *De novo* assembly of the transcriptome was performed with Trinity 2.4.0. Trinity groups transcript into clusters based on shared sequence content. Such a cluster of transcripts is very commonly referred to as a "gene." The longest transcript in the cluster was chosen as the "gene" sequence (aka Unigene).

Unigene Annotation and Functional Classification

Salmon was used to perform expression level analysis for Unigenes by calculating transcripts per million (TPM). The differentially expressed Unigenes were selected with log2 (fold change) > 1 or log2 (fold change) < -1 and with statistical significance ($p < 0.05$) by R package edgeR. All assembled Unigenes were aligned against the non-redundant (Nr) protein database Gene ontology (GO), SwissProt, Kyoto Encyclopedia of Genes and Genomes (KEGG), and eggnog databases using DIAMOND with a threshold *E* value < 0.00001. Pfam.

Metabolome Sequencing and Secondary Metabolite Isolation

To determine the metabolites that significantly changed in *D. nobile* under different epiphytic patterns, four stem strips under four attachment methods, Danxia stone (CSXTDS), Crushed stone (CSXTSS), Sawdust (CSXTJM), and Stump (CSXTSZ) were analyzed in three replicates per group. The collected samples were thawed on ice, and metabolites were extracted from 20 μl of each sample using 120 μl of precooled 50% methanol buffer. Then, the mixture of metabolites was vortexed for 1 min and incubated for 10 min at room temperature, and stored at −20°C overnight. The mixture was centrifugated at 4,000 *g* for 20 min; subsequently, the supernatant was transferred to 96-well plates. Pooled quality control (QC) sample was also prepared by combining 10 μl of each extraction mixture. All samples were analyzed using a Triple TOF 5600 Plus higher solution tandem mass spectrometer (SCIEX, Warrington, United Kingdom). Chromatographic separation was performed using an ultra-performance liquid chromatography (UPLC) system (SCIEX, United Kingdom). During the entire acquisition period, the mass accuracy was calibrated every 20 samples. Furthermore, a QC sample was analyzed every 10 samples to evaluate the stability of the LC-MS.

The group datasets were normalized before the analysis was performed. Data normalization was performed on all samples using the probabilistic quotient normalization algorithm. Then, QC-robust spline batch correction was performed using QC samples. The *p* value analyzed by Student's *t*-test, which was then adjusted for multiple tests using an FDR (Benjamini–Hochberg), was used for the selection of different metabolites. We also conducted the supervised PLS-DA using metaX for variables that discriminate profiling statistical method to identify more specific differences between the groups. The VIP cut-off value of 1.0 was set to select important features.

Combined Analysis of Differential Genes and Differential Metabolites

Based on the RNA-Seq sequencing, differential analysis of gene expression was performed. Differential genes were screened with $p < 0.05$, FC > 2, or < -2 as criteria, and statistical analyses such as fold-change and *t*-test were used for univariate analysis, and q-values were obtained by BH correction. Combined with multivariate statistical analysis of the VIP (Variable Important for the Projection) value obtained by PLS-DA, the satisfying ratio ≥ 2 or ratio $\leq 1/2$, *p* value ≤ 0.05, and VIP ≥ 1 of differential metabolites were screened. R language was used to standardize the expression values of differential genes and differential metabolites, and correlation analysis was performed. Taking $p < 0.05$, $r > 0.8$ as the standard, the KEGG pathway annotation was used to analyze the intersection in order to obtain key genes and metabolites.

Quantitative Real-Time PCR Validation

To verify the accuracy of transcriptome data, differential genes (DEG) were selected for qRT-PCR verification. The stems of 2-year-old *D. nobile* collected on the Danxia stone (CSXTDS), Crushed stone (CSXTSS), Sawdust (CSXTJM), and Stump (CSXTSZ) were used as materials. Total RNA was extracted by BIOFIT polysaccharide polyphenol biological kit, and cDNA was synthesized by TIANGEN reverse transcription kit. The volume of qRT-PCR reaction system was 10 μl in total, containing 1 μl cDNA template with a concentration of 1–10 ng/μl, 5 μl SYBR Green qPCR Master Mix (Universal), 0.2 μl primer each with a concentration of 10 μmol/L, and 3.6 μl ddH_2O. The reaction was performed using the following conditions: denaturation at 95°C for 30 s, followed by 40 cycles of amplification (95°C for 10 s, 60°C for 30 s, and 72°C for 30 s), and extension at 95°C for 15 s. For each sample, three technical replicates of the qRT-PCR assay were used with three biological replicates. Gene expression was evaluated using the $2^{-\Delta\Delta Ct}$ method.

RESULTS

Analysis of Illumina Sequencing Results

Based on the Illumina sequencing, 370,032 CCS polished sequences were obtained. After clustering the full-length nonchimeric sequences, 118,574 high-quality consistent sequences were obtained. Then, after removing redundancy, 68,401 transcript sequences were obtained (**Figure 1A**). The length of the sequence was less than 6,000 nt, mainly distributed around 2,000 nt. With the increase of dress matching length, the single gene showed a decreasing trend. By analyzing the

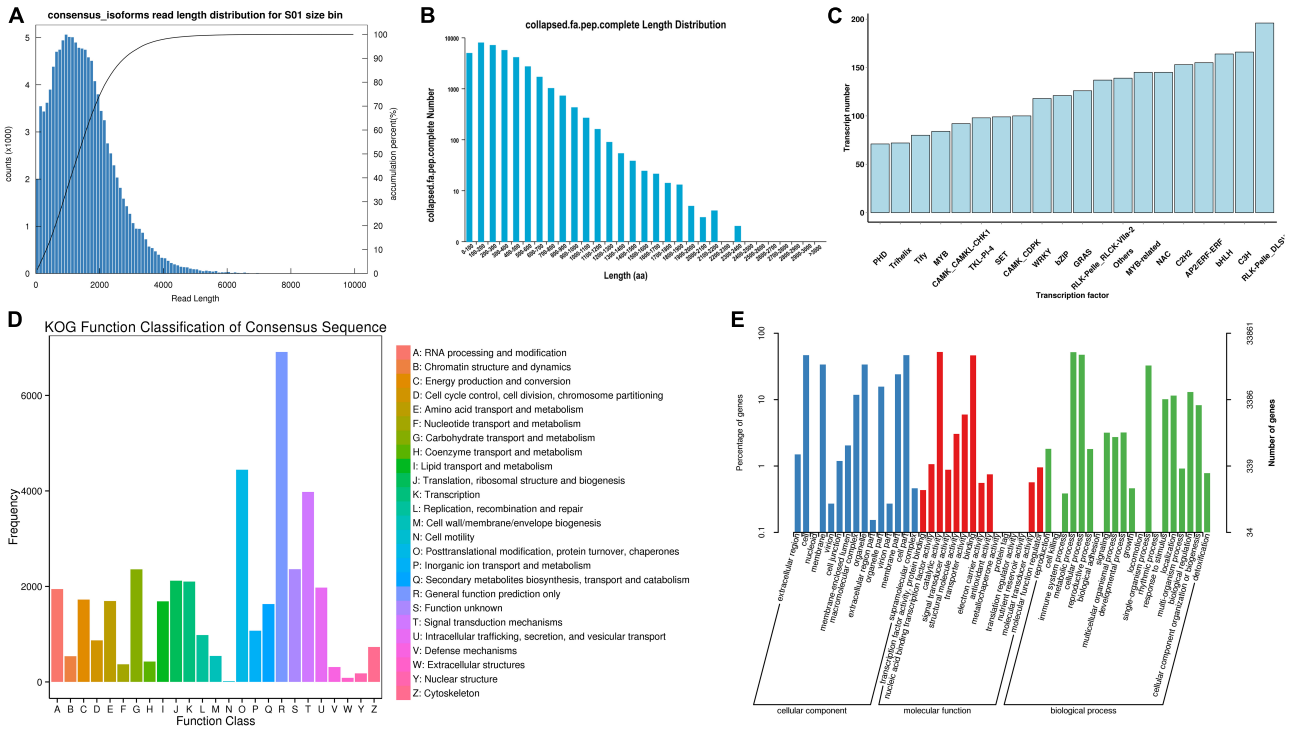


FIGURE 1 | Sequencing results of third-generation transcriptome. **(A)** Length distribution of transcripts after redundancy removal; **(B)** Length distribution of complete ORF; **(C)** Statistics of different types of transcription factors; **(D)** KOG function annotation; and **(E)** GO function annotation.

transcript sequence, 2,034 alternative splices were obtained for each sample, and a total of 25,591 SSRs and 9,427 lncRNA were predicted. In the prediction of coding sequence (CDs), 36,730 regions of CDs were obtained, and the complete open reading frame (ORF) length was mainly distributed in 100–200 aa, accounting for 21.45% (**Figure 1B**). A total of 5,572 transcription factors were predicted (**Figure 1C**), which were distributed in 4,065 gene families. To obtain the annotation information of transcripts, the obtained transcripts were compared with NR, Swissprot, GO, COG, KOG, Pfam, and KEGG databases.

A total of 59,148 transcripts were annotated, and 59,065, 55,554, 40,735, and 44,748, 26,738, and 36,610 transcripts were annotated in NR, eggNOG, Swissprot, and Pfam, KEGG, KOG databases, respectively (**Figure 1D**). All these databases divided transcripts into 25 categories, among which the functional annotation of KOG database was relatively comprehensive. Most life activities were included in the annotation, and the number of genes related to general function prediction was the largest, with 6,021. In contrast, the number of genes associated with cell motility was only 15, and other kinds of transcripts were expressed differently. Among them, the number of genes related to metabolic function was 1,517. GO database annotated 33,861 transcripts (**Figure 1E**), which were mainly divided into three aspects: biological process, molecular function, and cellular component. Among them, Oxidation-reduction process (2,715 genes), ATP binding (4,148 genes), and Integral component of membrane (7,011 genes) were the most abundant subgroups, respectively. This sequencing constructed a nucleotide database of the whole growth period, all tissues and organs, and full-length transcripts of *D. nobile*.

Analysis of RNA-Seq Sequencing Results Under Different Epiphytic Patterns

To obtain the gene expression of *D. nobile*, the RNA of *D. nobile* was extracted from four different epiphytic patterns, and then, a sequencing library was established to sequence the qualified samples. The PCA diagram shows (**Figure 2A**) that the sample has a high resolution and can be used in subsequent quantitative and qualitative studies. The 12 sample libraries constructed were sequenced on a machine and were tested and then qualified (**Table 1**), resulting in a total of 165,609 transcripts being spliced into 80,135 Unigenes. Different authoritative databases, namely, NCBI_nr, GO, KEGG, Pfam, Swiss-Prot, and eggNOG were selected for a comparison of Unigenes.

The GO (Gene Ontology) functional annotation results showed (**Figure 2B**) that a total of 27,761 (34.64%) Unigenes were annotated in three areas, namely, molecular functions, cellular components, and biological processes. In terms of molecular function, it was mainly concentrated in protein binding, molecular function, and ATP binding. In terms of cellular components, it was mainly enriched in three aspects, namely, nucleus, cytoplasm, and plasma membrane. In terms of biological process, it was mainly enriched in three aspects: biological process, transcriptional regulation, and transcription. The KEGG results showed (**Figure 2C**) that a

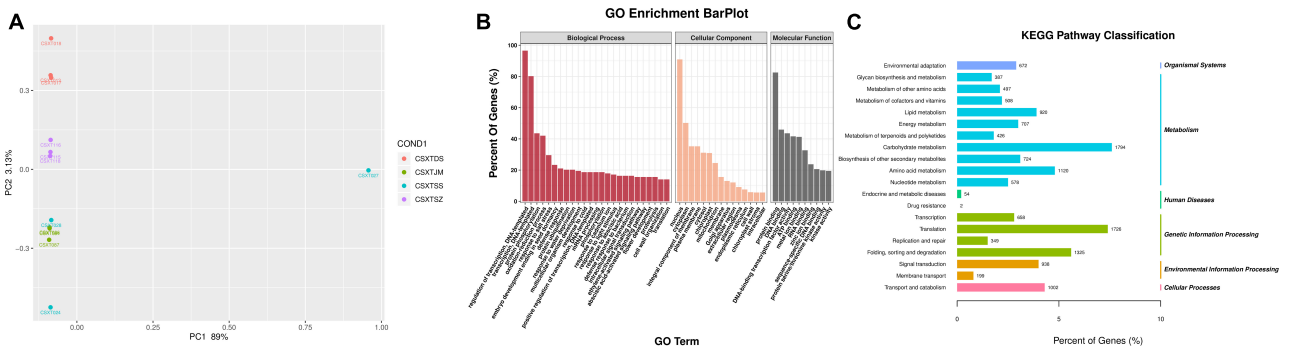


FIGURE 2 | Transcriptome sequencing results. **(A)** PCA map of sample differential genes; **(B)** GO functional annotation map of genes; and **(C)** KEGG transcriptome annotation results.

total of 23,454 Unigenes were annotated, which were related to 141 different metabolic pathways. These pathways were mainly divided into six categories (i.e., Organismal Systems, Metabolism, Human Diseases, Genetic Information processing, Environmental Information Processing, and Cellular Processes). Interestingly, Metabolism, Genetic Information Processing, and Environmental Information Processing were the three main genes. Carbohydrate metabolism (1,794 genes), Translation (1,726 genes) and folding, sorting, and degradation (1,325 genes) were the three main categories. These important metabolic pathways may affect the synthesis of active components in *D. nobile*. In addition, a total of 32, 98, 89, 115, and 19 genes were related to Indole alkaloid biosynthesis (K01593, K21026), Isoquinoline alkaloid biosynthesis (K01593, K00811, K00276, K00422, K14454, K00815, K15849, K21311, K14455), Tropane, piperidine, and pyridine alkaloid biosynthesis (K00811, K00276, K14454, K08081, K00815, K15849, K00817, K14455), Flavonoid biosynthesis (K13081, K01859, K00487, K08695, K13065, K00475, K00660, K05278, K05277, K05280, K13083, K13082, K00588, K09754), and Isoflavonoid biosynthesis (K13258, K13264, K13260), respectively. There remains a sea of unigenes that were found in the plant hormone signal transduction (524 unigenes), plant-pathogen interaction (527 unigenes), MAPK signaling pathway (414 unigenes), and Circadian rhythm (146 unigenes) categories. These pathways not only play a significant role in plant growth and development but were also related to the synthetic processes for primary and secondary products.

TABLE 1 | Transcriptome sequencing results.

Samples	Raw_Reads	Valid_Reads	Valid%	Q20%	Q30%	GC%
CSXT013	49,669,070	48,255,252	97.15	98.25	94.43	47.14
CSXT017	57,695,118	56,023,998	97.1	98.26	94.43	46.89
CSXT018	45,646,752	44,231,148	96.9	98.35	94.67	46.95
CSXT084	46,658,872	45,367,348	97.23	98.24	94.37	45.85
CSXT086	45,166,582	42,998,662	95.2	98.32	94.59	46.5
CSXT087	39,684,332	38,740,132	97.62	98.3	94.54	46.38
CSXT024	49,847,190	48,614,568	97.53	98.34	94.63	46.68
CSXT027	42,795,584	41,971,920	98.08	98.3	94.47	44.17
CSXT028	49,252,856	48,219,690	97.9	98.25	94.38	46.6
CSXT115	44,594,872	43,740,732	98.08	98.25	94.39	46.23
CSXT116	52,407,542	51,185,366	97.67	98.19	94.23	46.15
CSXT118	53,060,766	51,965,238	97.94	98.21	94.31	46.53

Differential Gene Analysis Under Different Epiphytic Patterns

To explore the key differential genes of *D. nobile* under different epiphytic patterns, the genes of Danxia and other epiphytic patterns were compared and analyzed. The change in gene expression was different under each epiphytic pattern (**Figures 3A,C**), and the change in gene expression in Danxia stone was the largest. There were 1,401 differential genes in the DSvsJM comparison group; 2,620 differential genes in the DSvsSS comparison group and 2,244 differential genes in the DSvSZ comparison group. In the three comparison groups, 828, 1,811, and 1,582 genes were upregulated and 573, 809, and 662 genes were downregulated, respectively. Among them, there were 387 differentially shared genes in the three comparison groups (**Figure 3B**), and 289 genes were annotated. These genes encode different genes and proteins, including α-galactosidase-like, chalcone synthase, retrovirus-related pol polyprotein from transposon TNT 1-94, allene oxide synthase 1, mitogen-activated protein kinase kinase kinase ANP1-like, cinammate 4-hydroxylase, 3-ketoacyl-CoA synthase 4-like, E3 ubiquitin-protein ligase PUB23-like, fructokinase-1-type, linoleate 13S-lipoxygenase 2-1, chloroplastic-like, L-ascorbate oxidase homolog, flavonoid 3′-monooxygenase-like isoform X2, and WRKY transcription factor.

The KEGG analysis annotated differential genes in different pathways (**Figure 3D**), including starch and sucrose metabolism (dct00500), MAPK signaling pathway-plant (dct04016), carotenoid biosynthesis (dct00906, zma00906), ubiquitin mediated proteolysis (cre04120), fructose and mannose metabolism (dct00051), pentose and glucuronate interconversions (smo00040), proteasome (dct03050), flavonoid biosynthesis (nnu00941), plant-pathogen interaction (csat04626, han00564, dct04626), pyruvate metabolism (dct00620), glycerophospholipid metabolism (han00564), endocytosis (vvi04144), folate biosynthesis (ghi00790), amino sugar and nucleotide sugar metabolism (dct00520), carbon fixation in

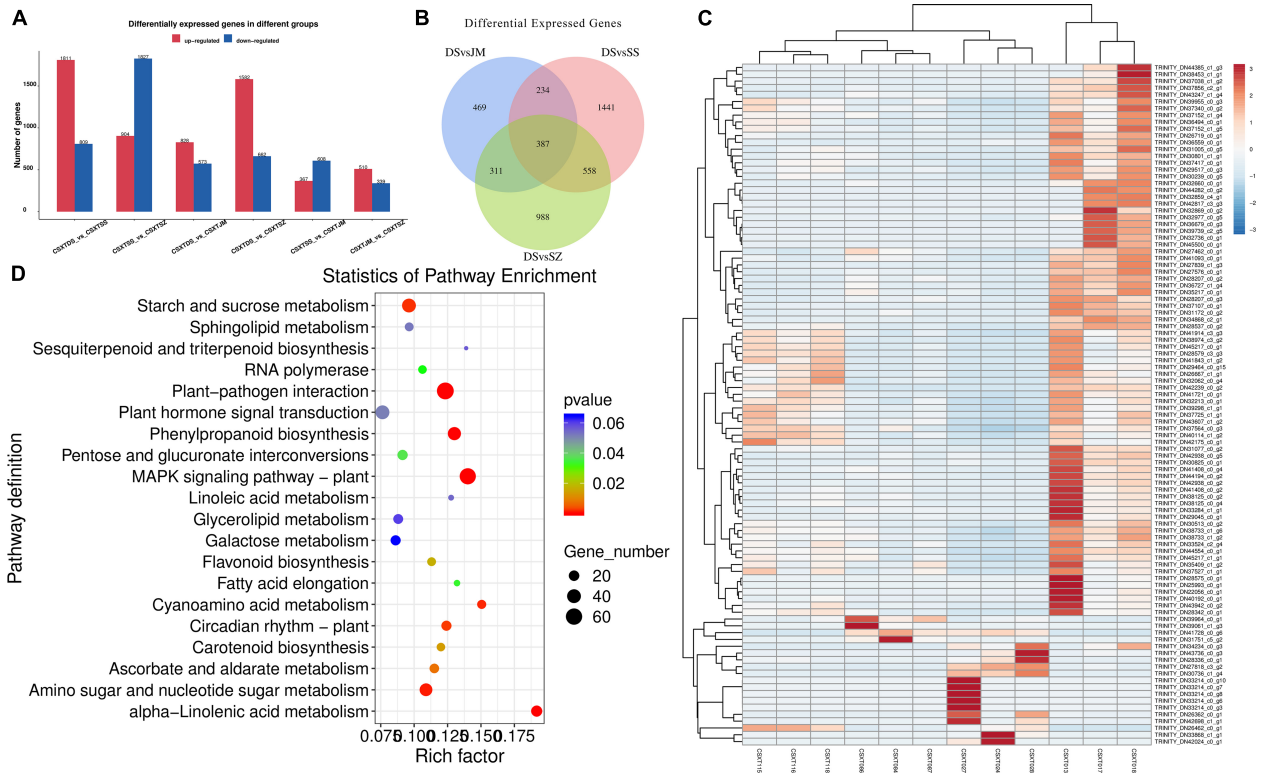


FIGURE 3 | Statistics of differentially expressed genes. **(A)** Number statistics of differentially expressed genes; **(B)** Veen map of Danxia stone; **(C)** Differential gene clustering heat map; and **(D)** Enrichment of KEGG pathway of differential genes.

photosynthetic organisms (dct00710), alpha-Linolenic acid metabolism (han00592), and other pathways. These genes were selected for further correlation analysis.

Analysis of Differential Metabolites Under Different Epiphytic Patterns

To explore the metabolic differences of *D. nobile* under different epiphytic patterns, the samples were analyzed by metabolic analysis. A total of 29,518 metabolites were identified. The results of HMDB superclass classification showed (**Figure 4A**) that lipid molecules were the most abundant metabolites, followed by phenylpropane and polyketones, and then heterocyclic compounds, organic oxygen compounds, organic acids, and their derivatives. Results of KEGG database showed (**Figure 4B**) that differential metabolites were mainly enriched in metabolic process, secondary metabolite biosynthetic process, ABC transporter, cyanoamino acid metabolism, and phenylpropanoid biosynthesis.

Comparing Danxia stone with the other three epiphytic patterns, the results of analysis showed (**Figure 5A**) that there were 1,072 differential metabolites and 39 secondary metabolites in the DSvsJM comparison group. There were 1,478 differential metabolites, and 39 metabolites were annotated in DSvsSS comparison group. There were 945 differential metabolites, and 21 secondary metabolites were annotated in the DSvsSZ comparison group. In the three comparison groups, 407, 535, and 343 metabolites were upregulated, and 665, 943, and 602 were downregulated. There were 66 differential metabolites in common among the three comparison groups, including butadiene-styrene rubber, 2-methyl-4-heptanone, (Z)-3-nonen-1-ol, 3,5,5-trimethylhexanal, *S*-propyl-L-cysteine, 2-pentadecylfuran, and other metabolites. KEGG pathway (**Figure 5B**) showed that the differential metabolites of the DSvsJM, DSvsSS, and DSvsSZ groups were enriched in 14, 31, and 7 pathways, respectively. The common pathways of different metabolites in the three comparison groups were phenylpropanoid biosynthesis, biosynthesis of alkaloids derived from shikimate pathway, biosynthesis of plant secondary metabolites, biosynthesis of secondary metabolites, metabolic pathways, flavonoid biosynthesis, and arginine and proline metabolism. An analysis showed that the Crushed stone and Danxia stone had the most differential metabolites, and Danxia stone had an effect on the synthesis of phenyl propionic acids, alkaloids, flavonoids and other amino acids, and secondary metabolites. These differential metabolites were selected for further experiments.

Combined Analysis of Differential Genes and Metabolites

To understand the connection between genes and metabolites under different epiphytic patterns, the KEGG pathway was analyzed, and the correlation analysis of differential genes

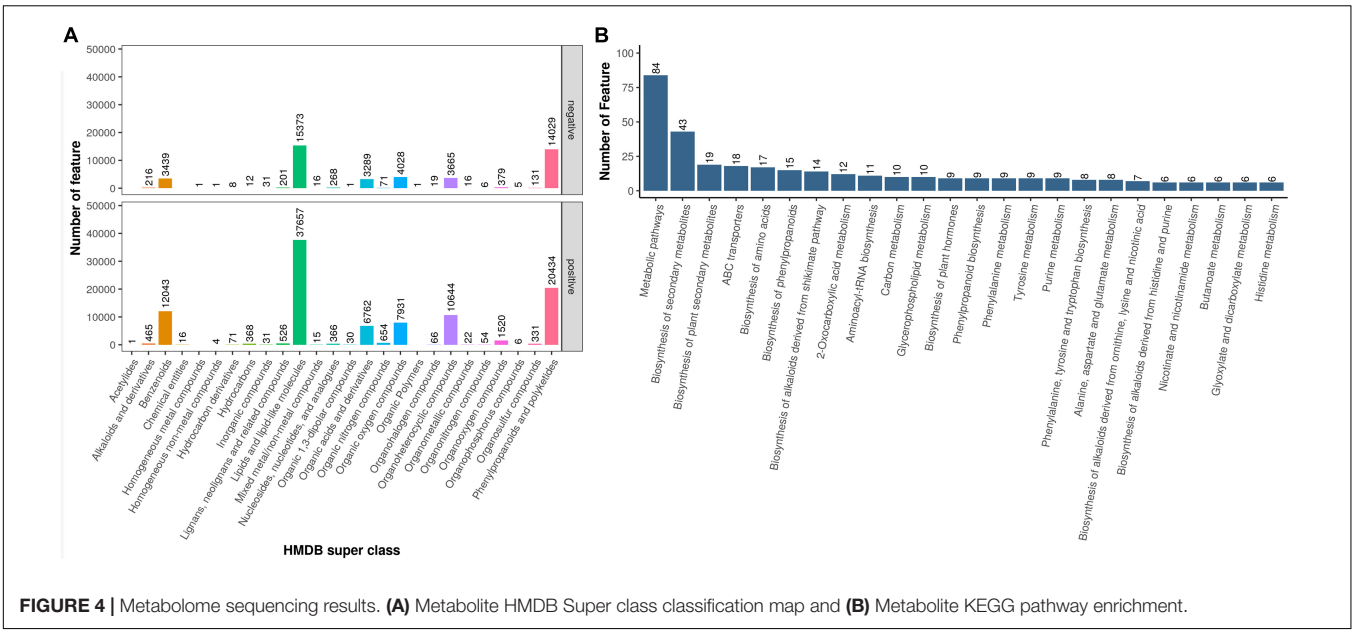


FIGURE 4 | Metabolome sequencing results. **(A)** Metabolite HMDB Super class classification map and **(B)** Metabolite KEGG pathway enrichment.

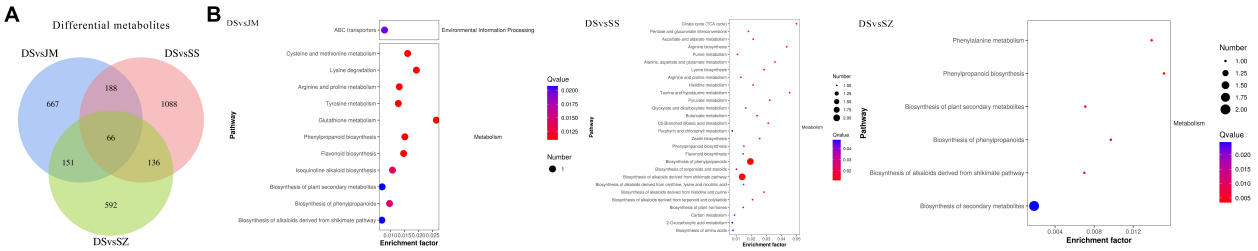


FIGURE 5 | Statistics of differential metabolites. **(A)** Veen plot of differential metabolites and **(B)** KEGG enrichment pathway of differential metabolites.

and differential metabolites was performed. There were 387 differential genes among the three groups of samples, corresponding to 66 differential metabolites. We detected 542 significant associations ($p < 0.01$, $R > 0.8$) between 386 genes and 66 metabolites. Among them, different metabolites are associated with many different genes. Dimethomorph was associated with the maximum number of genes (89), and many metabolites were only associated with *RVE1*, such as Epicatechin-(4beta- > 8)-ent-epicatechin, Procyanidin_B4, Procyanidin_B2, Procyanidin_B5, and dTDP-4-amino-4,6-dideoxy-5-C-methyl-D-allose. Among them, *ACS1* was negatively correlated with hydroxymethylbilane, which indicated that this gene had an inhibitory effect on the synthesis and accumulation of hydroxymethylbilane. There was a positive correlation between *LOX2.3* and 2,3-Dinor-8-iso_prostaglandin_F1alpha, indicating that this gene can promote the synthesis and accumulation of lipid molecules and lipid-like molecules.

Different metabolites were associated with different genes (**Figure 6**), and six common pathways of KEGG pathway were observed. *CYP75B2* and Malonylapiin are coenriched in the Flavone and flavonol biosynthesis (map00944) pathway, and both showed a downward trend. *CYP75B2* encodes flavonoid 3′-monooxygenase, which was also associated with cell membrane (GO:0016020) and oxidoreductase activity (GO:0016709) and the synthesis process of secondary metabolites (GO:0044550), which affected not only the metabolism of flavonoids, but also the metabolism of other secondary products (Zhang X. et al., 2020), such as anthocyanin synthesis (Jin et al., 2020). *ALDH2B7, ADC1*, and N-Acetyl putrescine were enriched in the metabolic pathway of arginine and proline. *EPSPS-1, SHKA, DHAPS-1*, and L-Homophenylalanine were enriched in the metabolic pathway of phenylalanine. These genes regulated the anabolism of amino acid analogs. *TRINITY_DN29493_c2_g1* was related to the synthesis of 5′-Phosphoribosyl-N-formylglycinamide, and *TRINITY_DN36595_c0_g2* was related to the synthesis of Allantoate. They were jointly enriched in the metabolism pathway of purine, and the gene expression was upregulated, whereas the metabolite expression was downregulated, indicating that this gene may inhibit purine metabolism. *GES* and Gibberellin_A8-catabolite were coenriched in the diterpenoid biosynthesis. *ACS1, SAHH, ACS2*, and S-Adenosyl-L-homocysteine were coenriched in the cysteine and methionine metabolism (map00270) pathway. The expression of *ACS1* and *ACS2* was upregulated and the expression of *SAHH* was downregulated, which jointly regulated cystine and methionine metabolism. *CHLP, LOX2, LOX2.3, CYP74B2*, and

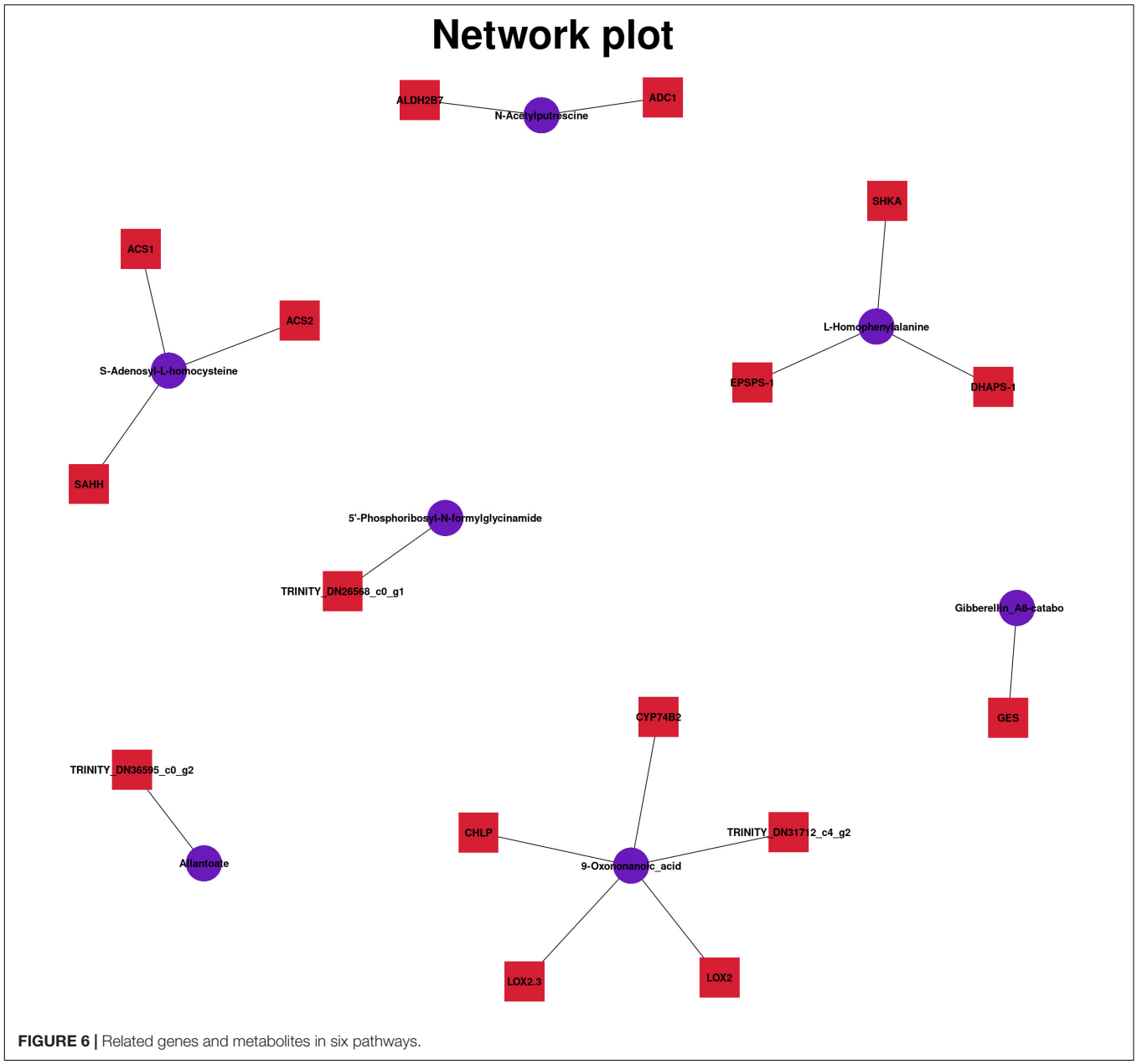


FIGURE 6 | Related genes and metabolites in six pathways.

TRINITY_DN31712_c4_g2 were related to alpha-Linolenic acid metabolism (map00592). These genes were downregulated, indicating that the genes inhibited synthesis and accumulation of alpha-linolenic acid. The results showed that the metabolism of flavonoids, purines, and alpha-linolenic acid in Danxia stone decreased to varying degrees. There were a variety of regulatory mechanisms in organisms, and genes and metabolites had different regulatory modes.

Expression Characteristics of Differential Metabolites and Genes

According to the KEGG pathway enrichment of differential genes and metabolites, 17 differential genes (**Figure 7A**) and eight metabolites (**Figure 7B**) were finally obtained. After the expression and metabolite peak area of genes were standardized in R, their expression characteristics were analyzed. The results showed that the expression of *CYP75B2* and the metabolite Malonylapiin was downregulated in Danxia stone, while the expression levels were higher in other epiphyte modalities, indicating that the biosynthesis of flavonoids and flavonols was reduced in Danxia stone. In the stone epiphytic patterns, *CHLP* was most actively expressed, encoding geranylgeranyl reductase, affecting the synthesis of metabolites Isopentenyl-diphosphate and geranylgeranyl-diphosphate, and regulating the terpenoid backbone biosynthesis. The biosynthesis of terpenoid backbone was not active in the wooden forms, and the most active one was *LOX2*, encoding lipoxygenase, but its expression level was

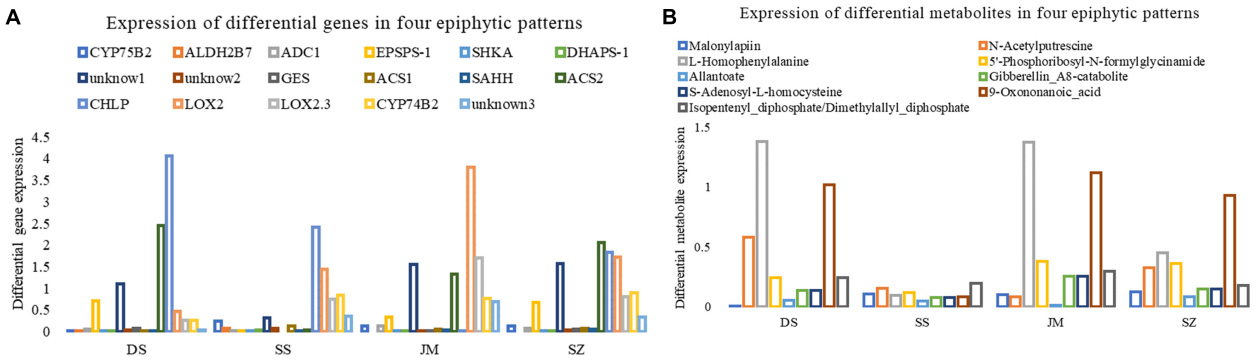


FIGURE 7 | Changes of differential genes and differential metabolites under different epiphytic patterns. The abscissa represents four different epiphytic patterns. Danxia stone (CSXTDS, DS), Crushed stone (CSXTSS, SS), Sawdust (CSXTJM, SS), and Stump (CSXTSZ, SZ). **(A)** Differential gene expression. Unknow1 is *TRINITY_DN29493_c2_g1*, unknow2 is *TRINITY_DN36595_c0_g2*, unknow3 is *TRINITY_DN31712_c4_g2*. **(B)** Peak areas of differential metabolites under different epigenetic patterns.

low in the epilithic forms. The content of the corresponding metabolites in the wooden forms was higher than that in the stone patterns. It indicated that differential regulation of alpha-linolenic acid metabolic pathways occurs in the two epiphytic modalities. *TRINITY_DN31712_c4_g2* also encoded lipoxygenase, but its expression was downregulated, indicating that lipoxygenase was regulated by different genes and was a key enzyme in the alpha-linolenic acid metabolic pathway. L-Homophenylalanine was higher in both Danxia stone and Stump, although almost absent in Crushed stone, which indicated that amino acid metabolism was more active in Danxia and Sawdust. In general, the content of secondary metabolites was higher in Danxia stone than in other epiphytic patterns, and lower in Crushed stone.

qRT-PCR Validation

To validate the reliability of the transcriptome data, we selected five DEGs and analyzed their expression levels by qRT-PCR. The actin gene of *D. nobile* was used as a reference control. Primers were designed using Primers Blast on NCBI website (**Table 2**). The results showed that the expression patterns of the five DEGs were consistent with the transcriptome sequencing data (**Figure 8**). In Danxia stone, the expression of *CYP75B2* was downregulated, and in the Crushed stone, the expression of *CHLP* was the most active, which confirmed the reliability of our transcriptome data.

DISCUSSION

In this study, through transcriptomic and metabolic analysis, the hub genes and pathways generated by *D. nobile* under different epiphytic patterns were identified, and the genetic mechanism of the differences was gained. In this study, we screened the differential genes related to differential metabolites. *CYP75B2*, which encoded flavonoid 3′-monooxygenase, regulated the metabolism of Malonylapiin, *ACS1* and *ACS2* regulated the synthesis of *S*-Adenosyl-L-homocysteine, and indirectly regulated the metabolism of cysteine and methionine, and *CHLP* and *CYP74B2* inhibited the synthesis and accumulation of α-linolenic acid. The amino acid metabolism in Danxia stone was increased, and it was closely related to *EPSPS-1, SHKA, DHAPS-1*, *ACS1*, *SAHH*, *ALDH2B7*, and *ADC1*. Amino acids, as signal molecules, were involved in regulating the synthesis of ethylene and other hormones in plants (Wu et al., 2020), and had a certain impact on the tissue metabolism, growth, maintenance, and repair of *D. nobile* (Heinemann and Hildebrandt, 2021). Danxia stone may induce certain stress response and interfere with the metabolism of amino acids, which was the result of *D. nobile* adaptation to the environment (Joshi et al., 2010). Danxia stone may be different from other epiphytes due to different contents of mineral elements and trace elements, and the response concentration and trend of different plants to various trace elements are different (Liu et al., 2010). Mineral nutrition may indirectly affect the synthesis and accumulation of flavonoids, terpenes, and other components in plants by affecting the levels of endogenous hormones in plants. In general, the content of total metabolites in Danxia

TABLE 2 | The information of qPCR primers.

Gene name	Primer	Tm	Length	Product length
CHLP-F	ACGTCTCACCTGATTTCTACGG	59.84	22	85
CHLP-R	TCTGATTTGTGGGTGACGGTG	60.54	21	
ACS1-F	TGGTTTAGGGTTTGCTTCGC	59.04	20	148
ACS2-R	TCAACTTCAGCCCAGTGTTCT	59.51	21	
CYP75B2-F	GTGGGGAGGAGAGTGTTTGG	60.0	20	116
CYP75B3-R	CCAAGCCCAGGCACAAAATC	60.0	20	
ADC1-F	GGATGTCCTGCGAGCGATG	60.95	19	82
ADC2-R	ACTGTGAGCGTATTCCTCGG	59.54	20	
CYP74B2-F	GCTAGGCTTCAACGCATTCG	60.0	20	92
CYP74B3-R	GTCTCTGCTTCAGCCCTGTT	60.0	20	
ACS2-F	GTGAAGCTTAACGTGTCGCC	59.8	20	115
ACS3-R	TTATTCTCCTCAGCGCCACC	59.8	20	
actin-F	AATCCCAAGGCAAACAGA	51.0	18	–
actin-R	CACCATCACCAGAATCCAG	53.0	19	

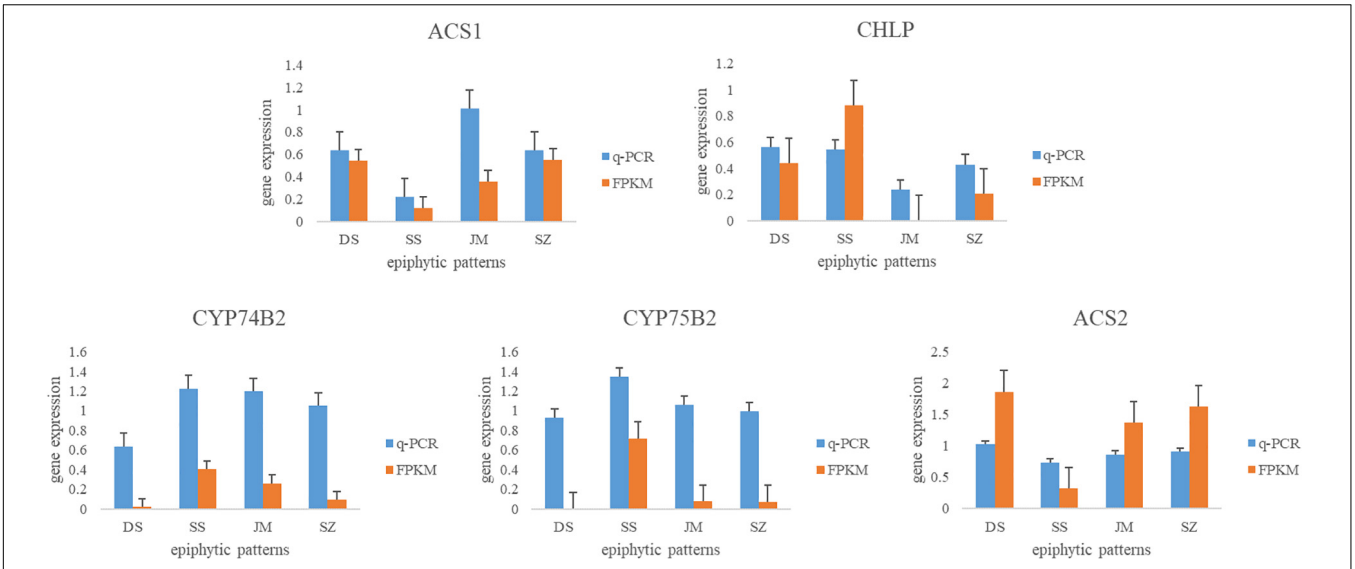


FIGURE 8 | qRT-PCR analysis of differential genes in *Dendrobium nobile*. The ordinate corresponds to the average change of expression value, the abscissa represents four different epiphytic patterns, and the FPKM represents the value of transcriptome sequencing.

stone was higher, which indicated that under the cultivation method of Danxia stone, the content of medicinal active ingredients was higher, and the quality of medicinal materials was good.

In this study, the full-length transcriptome database of the whole organs and tissues of *D. nobile* during the whole growth period was constructed by sequencing, which provides a basis for other studies of this species. Through the integrated omics analysis of transcriptome and metabolome, this study obtained the main metabolic pathways and genes produced by differences, and had a certain understanding of genetic mechanism concerning the epiphytic patterns of orchid plants. Theoretically, the cultivation method of Danxia was more conducive to the accumulation of metabolites in *D. nobile*, which was suitable for field cultivation. However, the accumulation of secondary metabolites is related not only to the cultivation mode, but also to the root fungi, bacteria, light, and harvesting season, which needs further research. Besides, this study could be serving as a high-quality reference for subsequent genomics and bioinformatics research of additional species within this genus.

CONCLUSION

In this study, based on the PacBio SMRT sequencing and Illumina sequencing, a full-length transcriptome database of *D. nobile* during the whole growth period and whole organ tissue was constructed. UPLC-mass spectrometry was used to study the changes in metabolites of *D. nobile* under different epiphytic patterns. Combined transcriptome and metabolome analysis were used to explore the differences and genetic mechanism of *D. nobile* under different epiphytic patterns.

As a result, a total of 80,135 Unigenes and 29,518 metabolites were obtained. Among them, there were 387 differential genes, corresponding to 66 metabolites, of which 542 had significant correlations. The metabolic pathways which have changed were mainly flavone and flavonol biosynthesis, arginine and proline metabolism, metabolic pathway of phenylalanine, purine metabolism, cysteine and methionine metabolism, and alpha-linolenic acid metabolic pathway. Differential metabolites included Malonylapiin, *N*-Acetylputrescine, L-Homophenylalanine, 5′-Phosphoribosyl-N-formylglycinamide, Allantoate, Gibberellin_A8-catabolite, *S*-Adenosyl-L-homocysteine, 9-Oxononanoic_acid, Isopentenyl_diphosphate, and Dimethylallyl_diphosphate. Relevant regulatory genes included *ALDH2B7, ADC*, *EPSPS-1, SHKA*, *DHAPS-1*, *GES, ACS1, SAHH, ACS2*, *CHLP, LOX2, LOX2.3,* and *CYP74B2*. Among them, *ALDH2B7, ADC1* were related to arginine and proline metabolism, *EPSPS-1, SHKA, DHAPS-1* were related to metabolic pathway of phenylalanine, and *ACS1, SAHH, ACS2* were related to cysteine and methionine metabolism. The unknown differential genes *TRINITY_DN29493_c0_g2, TRINITY_DN36595_c0_g2* were related to purine metabolism. *TRINITY_DN31712_c4_g2* were related to alpha-linolenic acid metabolic pathway, which regulated lipoxygenase with *LOX2* and *LOX2.3*. Multiple metabolic pathways included *CYP75B2*, *RVE1*, *ACS1*, *LOX2.3*, which jointly regulated the synthesis of metabolites. Different epiphytic patterns will affect the metabolic regulatory network of *D. nobile*, and the differences were mainly enriched in flavonoid metabolism, purine metabolism, terpenoid backbone biosynthesis, amino acid metabolism, and alpha-linolenic acid metabolism. The synthesis of flavonoids and the metabolism of purines in Danxia stone decreased, while the synthesis of terpenoids and metabolism of amino acids was increased. At the same time, the metabolite content in Danxia stone was higher, and the metabolite content in Crushed stone

was lower, indicating that Danxia stone was conducive to the accumulation of secondary metabolites in *D. nobile*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA806713.

AUTHOR CONTRIBUTIONS

D-LX conceived, supervised, and wrote-reviewed the manuscript. QL and CL originally wrote and reviewed the draft. QL, CH, MW, TL, JL, JHS, and LL performed the experiments and carried out the analysis. JLS and D-LX designed the experiments. D-LX and YH cofounded and co-administrated the project. All authors approved the final version.

FUNDING

This research was supported financially by the National Natural Science Foundation of China (31560079 and 31960074), the Science and Technology Department Foundation of Guizhou Province of China [Nos. (2017)5733-050, (2019)-027, and (2019)5657], the Special Joint Bidding Project of Zunyi Sci & Tech Bureau and Zunyi Medical University (ZSKHHZ-2020-91), and Honghuagang Sci & Tech Project of Zunyi City [ZHKHNZT (2020)04].

REFERENCES

Cheng, L., Guo, D. L., Zhang, M. S., Linghu, L., Fu, S. B., Deng, Y., et al. (2020). Dihydrophenanthrofurans and bisbibenzyl derivatives from the stems of *Dendrobium nobile*. *Fitoterapia* 143:5. doi: 10.1016/j.fitote.2020.104586

Guo, L. P., Zhou, L. Y., Kang, C. Z., Wang, H. Y., Zhang, W. J., Wang, S., et al. (2020). Strategies for medicinal plants adapting environmental stress and "simulative habitat cultivation" of Dao-di herbs. *Chinese J. Tradit. Chinese Med.* 45, 1969–1974. doi: 10.19540/j.cnki.cjcmm.20200302.101

Heinemann, B., and Hildebrandt, T. M. (2021). The role of amino acid metabolism in signaling and metabolic adaptation to stress-induced energy deficiency in plants. *J. Exp. Bot.* 72, 4634–4645. doi: 10.1093/jxb/erab182

Jiao, H. H., Chen, J., Li, B., and Bai, D. T. (2021). Analysis of the factors affecting the quality of wendangshen medicinal materials. *Western J. Tradit. Chinese Med.* 34, 73–74.

Jin, J. F., Li, Z. J., and Zhao, L. Y. (2020). Clonging and Analysis of RrF3'H in *Rosa rugosa*. *Am. J. Mol. Biol.* 10, 51–60.

Joshi, V., Joung, J. G., Fei, Z., and Jander, G. (2010). Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino Acids* 39, 933–947. doi: 10.1007/s00726-010-0505-7

Li, D. D., Zheng, C. Q., Zhang, F., and Shi, J. S. (2022). Potential neuroprotection by *Dendrobium nobile* Lindl. alkaloid in Alzheimer's disease models. *Neural Regen. Res.* 17, 972–977. doi: 10.4103/1673-5374.324824

Li, F. S. (2020). *Effects Of Different Substrates On The Accumlation Of The Main Components And The Assessment Of Anti-Inflammatory Activity In Phellinus Igniarius*. Changchun: Changchun Normal University.

Li, Q., Ding, G., Li, B., and Guo, S. X. (2017). Transcriptome analysis of genes involved in dendrobine biosynthesis in *Dendrobium nobile* Lindl. Infected with mycorrhizal fungus MF23 (Mycena sp.). *Sci. Rep.* 7:316. doi: 10.1038/s41598-017-00445-9

Li, Y., Kong, D., Fu, Y., Sussman, M. R., and Wu, H. (2020). The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol. Biochem.* 148, 80–89. doi: 10.1016/j.plaphy.2020.01.006

Li, Z., Xiang, J., Hu, D., and Song, B. (2020). Naturally potential antiviral agent polysaccharide from *Dendrobium nobile* Lindl. *Pest. Biochem. Physiol.* 167, 1–9. doi: 10.1016/j.pestbp.2020.104598

Ling-Hu, C., Gu, R. H., and Qin, L. K. (2021). Research progress on chemical consitituents and pharmacological effects of *Dendrobium nobile*. *Chinese Herb. Med.* 40, 44–51.

Liu, D., Guo, L., Huang, L., Jin, H., Liu, W., and Zhu, D. (2010). Effects of mineral nutrition on metabolism of flavonoids in medicinal plants. *Chinese J. Tradit. Chinese Med.* 35, 2367–2371.

Liu, H., Shi, J. H., Wu, M., and Xu, D. L. (2021). The application and future prospect of RNA-Seq technology in Chinese medicinal plants. *J. Appl. Res. Med. Aromatic Plants* 24:100318. doi: 10.1016/j.jarmap.2021.100318

Liu, Y. H., Lin, F. X., Tan, Y. F., Yang, J. Y., Zhang, B., Zhou, X. M., et al. (2021). Three new phenanthraquinones from the Root of *Dendrobium nobile*. *Chinese J. Organic Chem.* 41, 2112–2115. doi: 10.6023/cjoc202012005

Ma, C., Meng, C. W., Zhou, Q. M., Peng, C., Liu, F., Zhang, J. W., et al. (2019). New sesquiterpenoids from the stems of *Dendrobium nobile* and their neuroprotective activities. *Fitoterapia*. 138:104351. doi: 10.1016/j.fitote.2019.104351

Mao, X., Ge, M., Wang, X., Yu, J., Li, X., Liu, B., et al. (2021). Transcriptomics and metabolomics analyses provide novel insights into glucose-induced trophic transition of the marine diatom *Nitzschia laevis*. *Mar. Drugs* 19:426. doi: 10.3390/md19080426

Meng, X., Li, J. J., Yao, J., Chen, N., and Yu, J. (2020). Research Progress on the effects of different factors on the yield and quality of *Paeonia lactiflora* Pall. *Modern Agric. Sci. Technol.* 20, 59–62.

Nie, X., Chen, Y., Li, W., and Lu, Y. (2020). Anti-aging properties of *Dendrobium nobile* Lindl.: from molecular mechanisms to potential treatments. *J. Ethnopharmacol.* 257:112839. doi: 10.1016/j.jep.2020.112839

Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17, 147–154. doi: 10.1038/s41592-019-0690-6

Shovlin, S., and Tropea, D. (2018). Transcriptome level analysis in Rett syndrome using human samples from different tissues. *Orphanet J. Rare Dis.* 13:113. doi: 10.1186/s13023-018-0857-8

Wang, P., Chen, X., Cai, C. H., Kong, F. D., Huang, S. Z., Yuan, J. Z., et al. (2020). A new picrotoxane-type sesquiterpene from *Dendrobium nobile* Lindl. *Nat. Product Res.* 17, 1–6. doi: 10.1080/14786419.2020.1851224

Wang, W. W., Fan, H. J., Yang, W. H., Xia, K. Y., Li, H. G., Li, X. L., et al. (2021). Aperient effect of *Dendrobium nobile* Lindl extract on constipation model mice. *Sci. Technol. Food Ind.* 42, 1–6. doi: 10.13386/j.issn1002-0306.2020120264

Wang, Y. H. (2021). Traditional uses, chemical constituents, pharmacological activities, and toxicological effects of *Dendrobium* leaves: a review. *J. Ethnopharmacol.* 270:113851. doi: 10.1016/j.jep.2021.113851

Wen, Z., Guo, W., Li, J., Lin, H., He, C., Liu, Y., et al. (2017). Comparative transcriptomic analysis of vernalization- and cytokinin-induced floral transition in *Dendrobium nobile*. *Sci. Rep.* 7:45748. doi: 10.1038/srep45748

Wu, Y. Y., Liu, X. F., Fu, B. L., Zhang, Q. Y., Tong, Y., Wang, J., et al. (2020). Methyl jasmonate enhances ethylene synthesis in kiwifruit by inducing NAC genes that activate ACS1. *J. Agric. Food Chem.* 68, 3267–3276. doi: 10.1021/acs.jafc.9b07379

Xu, L., Mei, Y., Liu, R., Chen, X., Li, D., and Wang, C. (2020). Transcriptome analysis of *Spodoptera litura* reveals the molecular mechanism to pyrethroids resistance. *Pest. Biochem. Physiol.* 169:104649. doi: 10.1016/j.pestbp.2020.104649

Xu, Y. Y., Xu, Y. S., Wang, Y., Wu, Q., Lu, Y. F., Liu, J., et al. (2017). *Dendrobium nobile* Lindl. alkaloids regulate metabolism gene expression in livers of mice. *J. Pharm. Pharmacol.* 69, 1409–1417. doi: 10.1111/jphp.12778

Yu, Y. Z., Lu, Y. L., Chen, W. G., and Wen, X. Y. (2020). Study on different cultivation techniques and ornamental value of 12 species of *Dendrobium*. *J. Anhui Agric. Sci.* 48, 156–157.

Zhang, M. S., LingHu, L., Zhang, J. Y., and Nie, X. Q. (2019). Bibenzyl derivatives from *Dendrobium nobile*. *Chinese J. Organic Chem.* 39, 3289–3293. doi: 10.6023/cjoc201903035

Zhang, J. Q., Zhou, T., Xiao, C. H., Jiang, W. K., Guo, L. P., Kang, C. Z., et al. (2020). Technical evaluation and principle analysis of simulative habitat cultivation of *Dendrobium nobile*. *Chinese J. Tradit. Chinese Med.* 45, 2042–2045. doi: 10.19540/j.cnki.cjcmm.20200302.110

Zhang, X., Wu, F., Gu, N., Yan, X., and Zhang, H. (2020). Postharvest biological control of Rhizopus rot and the mechanisms involved in induced disease resistance of peaches by *Pichia membranefaciens*. *Postharvest Biol. Technol.* 163:111146. doi: 10.1016/j.postharvbio.2020.111146

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/fpls.2022.905275

Comparative Transcriptomics and Metabolites Analysis of Two Closely Related *Euphorbia* Species Reveal Environmental Adaptation Mechanism and Active Ingredients Difference

Han Zheng[1†], Mu-Yao Yu[1†], Yang Han[1†], Badalahu Tai[1,2], Sheng-Fa Ni[3], Rui-Feng Ji[1], Chun-Juan Pu[1], Kang Chen[1], Fu-Quan Li[4], Hua Xiao[4], Ye Shen[1], Xiu-Teng Zhou[1*] and Lu-Qi Huang[1*]*

[1] State Key Laboratory of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [2] Mongolian Medicine College, Inner Mongolia Minzu University, Tongliao, China, [3] Anhui University of Science and Technology, Huainan Xinhua Hospital, Huainan, China, [4] Hulunbeier Mongolian Medical Hospital, Hulunbeier, China

Roots of *Euphorbia fischeriana* and *Euphorbia ebracteolata* are recorded as the source plant of traditional Chinese medicine "Langdu," containing active ingredients with anticancer and anti-AIDS activity. However, the two species have specific patterns in the graphic distribution. Compared with *E. ehracteolata*, *E. fischeriana* distributes in higher latitude and lower temperature areas and might have experienced cold stress adaptation. To reveal the molecular mechanism of environmental adaptation, RNA-seq was performed toward the roots, stems, and leaves of *E. fischeriana* and *E. ehracteolata*. A total of 6,830 pairs of putative orthologs between the two species were identified. Estimations of non-synonymous or synonymous substitution rate ratios for these orthologs indicated that 533 of the pairs may be under positive selection (Ka/Ks $>$ 0.5). Functional enrichment analysis revealed that significant proportions of the orthologs were in the TCA cycle, fructose and mannose metabolism, starch and sucrose metabolism, fatty acid biosynthesis, and terpenoid biosynthesis providing insights into how the two closely related *Euphorbia* species adapted differentially to extreme environments. Consistent with the transcriptome, a higher content of soluble sugars and proline was obtained in *E. fischeriana*, reflecting the adaptation of plants to different environments. Additionally, 5 primary or secondary metabolites were screened as the biomarkers to distinguish the two species. Determination of 4 diterpenoids was established and performed, showing jolkinolide B as a representative component in *E. fischeriana*, whereas ingenol endemic to *E. ebracteolate*. To better study population genetics, EST-SSR markers were generated and tested in 9 species of *Euphorbia*. A total of 33 of the 68 pairs were screened out for producing clear fragments in at least four species, which will furthermore facilitate the studies on the genetic improvement and phylogenetics of this rapidly adapting taxon. In this study, transcriptome and metabolome analyses revealed the evolution of genes related to cold stress tolerance,

OPEN ACCESS

Edited by:
Xueqing Fu,
Shanghai Jiao Tong University, China

Reviewed by:
Ling Li,
Shanghai Jiao Tong University, China
Xiaojian Yin,
China Pharmaceutical University, China
Xu Lu,
China Pharmaceutical University, China

***Correspondence:**
Lu-Qi Huang
huangluqi01@126.com
Xiu-Teng Zhou
zxt_0508@163.com
Ye Shen
shenye@nrc.ac.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 26 March 2022
Accepted: 02 May 2022
Published: 31 May 2022

Citation:
Zheng H, Yu M-Y, Han Y, Tai B, Ni S-F, Ji R-F, Pu C-J, Chen K, Li F-Q, Xiao H, Shen Y, Zhou X-T and Huang L-Q (2022) Comparative Transcriptomics and Metabolites Analysis of Two Closely Related Euphorbia Species Reveal Environmental Adaptation Mechanism and Active Ingredients Difference. Front. Plant Sci. 13:905275. doi: 10.3389/fpls.2022.905275

biosynthesis of TCA cycle, soluble sugars, fatty acids, and amino acids, consistent with the molecular strategy that genotypes adapting to environment. The key active ingredients of the two species were quantitatively analyzed to reveal the difference in pharmacodynamic substance basis and molecular mechanism, providing insights into rational crude drug use.

Keywords: *Euphorbia*, comparative transcriptomics, metabolome, environmental adaptation, diterpenoid, EST-SSR

INTRODUCTION

Euphorbia is the largest genus in the Euphorbiaceae family, consisting of more than 2,000 species, with 77 species distributed in China (Li et al., 2008). Roots of *Euphorbia fischeriana* Steud (Li et al., 2021b) and *Euphorbia ebracteolata* Hayata (Yang et al., 2021) were used as the same kind of traditional Chinese medicine "Lvru" (now known as Langdu) for more than 2,000 years of treating swelling and ulcer of scabies (Zhao et al., 1996). As main active ingredients of Langdu, jolkinolide B (abietane type, multicyclic diterpenoid) and its derivatives have been proven to display potent anticancer activity (Wang et al., 2009, 2017a); ingenol's esters (ingenane type, bicyclic diterpenoid) have a great potential in treating human immunodeficiency virus (HIV) (Miana et al., 2015; Huang et al., 2019) and actinic keratosis (Parker et al., 2017), such as prostratin, Picato.

Although *E. fischeriana* and *E. ehracteolata* have been used interchangeably, the graphic distribution pattern of these two Langdu species is different (Li et al., 2008). Compared with *E. ehracteolata* growing in eastern and central China, *E. fischeriana* growing in northern and northeast China (higher latitude) has the habitat characteristics of lower annual average temperature (data from www.nmc.cn). In particular, *E. fischeriana* will be subjected to a strong cold stress in winter. Under this long-term environmental factor, the two species exhibit typical patterns of adaptive evolution and explosive speciation. From a genetic perspective, during speciation among closely related species, genes evolving rapidly are more differentiated than the rest, which is thought to be responsible for habitat differentiation and adaptation (Zhang et al., 2013; Mao et al., 2016; Zhao et al., 2016). However, few genomic resources are available for *E. fischeriana* and *E. ebracteolata*, leading to the unavailability of positive selection gene detection and the study of important loci. A few studies have been reported toward mechanism of ontogenesis (Prenner and Rudall, 2007), evolution of major structural characters (Horn et al., 2012), and diversity of species (Frajman et al., 2019) in *Euphorbia* (Euphorbiaceae), but neither of the two Langdu species was involved. Although the transcriptome of *E. fischeriana* has been sequenced (Barrero et al., 2011), it is not enough to explain the adaptive growth. With paucity of genetic data such as genome sequences and associated molecular markers, stress resistance or evolutionary analysis toward Langdu remains a challenge. In addition, studies have reported that the accumulation of compounds including soluble sugar, amino acids, and fatty acids contributes to adaptation toward abiotic stress (Wanner and Junttila, 1999; Duan et al., 2012; Wei et al., 2018). So far, chemical studies on Langdu have only been conducted on the active components of the drug, and few studies have been conducted on the compounds coping with stress resistance.

Transcriptome, characterized by its low cost and high efficiency, can not only provide additional genome resources and information about the process of speciation or adaptive evolution, such as time estimation of divergence, or detection of adaptive genes (Koenig et al., 2013; Zhao et al., 2018), but also provide expressed sequence tag—simple sequence repeat (EST-SSR) markers for species identification and germplasm evaluation (Varshney et al., 2005; Zhao et al., 2019; Li et al., 2021a). Furthermore, using metabolomics to study the metabolic changes under stress can reveal the response mechanism of plants to the changes in external environment or genes (Fiehn, 2002; De Vos et al., 2007). Therefore, we should not only reveal the molecular mechanism of adaptation to abiotic stress, but also study the effect of metabolite accumulation on adaptation to abiotic stress and the difference in active ingredients between the two plants. As the first step toward answering these questions, we obtained transcripts and metabolites for *E. fischeriana* and *E. ebracteolata*, furthermore carrying out a comprehensive analysis. 1) Transcripts of two Langdu species were identified, and their genetic differences were compared, providing additional genetic resources for Langdu breeding or evolutionary analysis; 2) evolutionary dynamics of two species were determined, obtaining an estimated time of differentiation, as well as the characteristics of adaptive evolution between the two species; 3) the differential metabolites of the two species were identified through metabolomics, and the chemical mechanism of adaptation toward stress was analyzed; 4) A UPLC-MS/MS method determining 4 diterpenoids simultaneously was established to detect diterpenoid contents in root and leaf of the two Langdu species; 5) and genus-specific EST-SSR markers based on the two species of Langdu were developed in preliminarily.

MATERIALS AND METHODS

Plant Materials

We collected two Langdu species at fructescence. *E. fischeriana* is from hulun buir (48°34′57.18″N, 119°54′07.56″E, alt.746 m, Inner Mongolia); *E. ebracteolata* is from Jiyuan (35°12′46.56″N, 112°25′55.58″E, alt.708 m, Henan). Fresh leaves, stems, and roots from twelve individual plants for both species were stored in liquid nitrogen until total RNA extraction and metabolites analysis.

Microscope Observation

Roots of *E. fischeriana* and *E. ebracteolata* were first embedded with paraffin and then cut into slices of 5 μm. Slices were dewaxed using xylene for 20 min two times, ethanol for 5 min two times, and 75% ethanol for 5 min and then washed with water. Dewaxed slices were soaked in saffron dye for 2 h and decolored with 50, 70, and 80% ethanol solution for 5 s. Fast green and ethanol were then used to dye and decolor, respectively. Slices were permeabilized using xylene and mounted with neutral gum. Microscopic features were captured using an Olympus BX51 microscope.

RNA-Seq and Data Analysis

To provide a preliminary indication of genetic variation within species, the high-quality mRNA from leaves, stems, and roots from three individual plants for each species was isolated using Plant RNA Purification Reagent (Invitrogen, Life Technologies, USA) following the manufacturer's protocol. Both RNA-seq library preparation and paired-end sequencing were performed using an Illumina HiSeq 6000 platform (Zheng et al., 2021). All unigenes were first subjected to BLASTX (Altschul et al., 1997) against the National Center for Biotechnology Information (NCBI), including the non-redundant protein (Nr) database and non-redundant nucleotide sequence (Nt) databases with an *E*-value threshold of 10^{-5}. The predicted gene name for each contig was assigned according to the best BLASTX hit. Protein structure and function were annotated with Swiss-Prot and Protein family (Pfam) database (Finn et al., 2010). Based on the annotations in NR, BLAST2GO v2.5 (Conesa et al., 2005) was used to obtain GO annotations for the aligned unigene sequences with an *E*-value threshold of 10^{-6}, and the Web Gene Ontology Annotation Plot (WEGO) software (Ye et al., 2006) was used to establish GO functional classifications for all unigenes. The unigenes were aligned to the euKaryotic Ortholog Groups/Cluster of Orthologous Groups (KOG/COG) database to predict and classify possible functions, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database (Kanehisa et al., 2008) was used to obtain pathway annotations (*E*-value threshold 10^{-10}).

Identification of Gene Orthologous Groups and Calculation of Ka/Ks

The Coding sequence (CDS) of each putative unigene were extracted according to the BLASTX results, and the orientation of the unaligned sequences was determined using ESTScan software. The CDSs extracted from the respected unigene were translated into amino acid sequences using the standard codon table. Self-to-self BLASTP was conducted for all amino acid sequences with a cutoff *E*-value of 10^{-5}. Based on the predicted CDS regions of both *E. ehracteolata* and *E. fischeriana* transcriptomes, ORTHOMCL version 2.0.9 (Li et al., 2003) with default settings was used to reconstruct the clusters of orthologous groups (COGs). Pairs of putative orthologous genes were identified based on the reciprocal best matches with an *E*-value threshold of 10^{-20}. Only pairs of sequences that mapped unambiguously to the same protein in Swiss-Prot database were selected as orthologous genes. The protein-coding sequences with unexpected stop codons in the BLAST hit region and/or shorter than 150 bp in length were removed.

Then, the Ks value of the obtained orthologous genes and the formula $T = K/2r$ were used to estimate divergence time (T) between *E. fischeriana* and *E. ehracteolata* (Graur et al., 2000). "K" is a genetic divergence expressed in terms of mean number of synonymous substitutions between orthologs; "r" is the mean rate of synonymous substitution and is considered to be $1.5\ e^{-8}$ substitutions/synonymous site/year for all dicots (Koch et al., 2000). Ka/Ks calculation was performed with PAML package (Yang, 2007) using default settings. Based on the Ka/Ks value and a threshold at 0.5, the orthologs were sub-categorized into two datasets: a test set with Ka/Ks above 0.5, and a reference dataset with Ka/Ks value < 0.5. The significance of the difference in GO term abundance between the two datasets was tested using the Fisher's exact test with the GOSSIP package (Bluthgen et al., 2005) implemented in BLAST2GO V.2.6 (Conesa et al., 2005).

Metabolome Analysis

About 50 mg of fresh pulverized roots and leaves of *E. fischeriana* and *E. ehracteolata* was accurately weighed and extracted with 1.0 ml 80% methanol for LC-MS. Extraction was done at room temperature (RT) for 3 h in an orbital shaker set at 220 rpm. Resulting extracts were spun down at 3,000 *g* for 15 min to sediment tissue material. About 200 μl of supernatant was transferred to vial for LC-MS analysis using an established LC-MS-based approach (Su et al., 2016; Kong et al., 2017). The LC-MS detection data were extracted and preprocessed using SIEVE software (Thermo), and the data were normalized and later edited in Excel 2013. Finally, it is organized into a two-dimensional data matrix form, which contains information such as retention time (RT), compound molecular weight (CompMW), observations (sample name), amount of material extracted (ID), and peak intensity. A total of 1,220 features at (ESI+) ion mode and 1,506 features at (ESI-) ion mode after editing were performed multivariate analysis (MVA) using SIMCA-P software (version 13.0, Umetrics AB, Umea, Sweden). Principal component analysis (PCA) and partial least squares-discriminate analysis (PLS-DA) were employed to identify biochemical patterns. The metabolites that differed between the two classes were quantified using a combination of VIP statistics (threshold > 1) of the OPLS-DA model and *t*-tests ($p < 0.05$). Compounds were identified by a comparison of m/z or precise molecular mass at http://metlin.scripps.edu.

Determination of Secondary Metabolites

Roots and leaves of *E. fischeriana* and *E. ehracteolata* were first freeze-dried then pulverized into powder. About 0.15 g of the fine powder was accurately weighed and extracted with 15 ml ethanol in an ultrasonic water bath at 30°C for 30 min. After cooling to room temperature, the extraction solution was filtered through a 0.22-μm millipore filter and injected directly into the UPLC system for analysis. Jolkinolide A (ChemFaces, CAS:37905-07-0, lot no: CFS201902), jolkinolide B (Nature Standard, CAS:37905-08-1, lot no:4069), jolkinolide E (ChemFaces, CAS:54494-34-7, lot no: CFS201901), and ingenol (Nature Standard, CAS:30220-46-3, lot no:3925) were used as the standard substances and were

FIGURE 1 | Two species of Langdu characteristic diagrams. **(A)** Fruit of *E. fischeriana*; **(B)** longitudinal section of *E. fischeriana* root; **(C)** cross-section micrograph of *E. fischeriana* root; **(D)** whole *E. fischeriana* plant; **(E)** whole *E. ehracteolata* plant; **(F)** fruit of *E. ehracteolata*; **(G)** longitudinal section of *E. ehracteolata* root; **(H)** cross-section micrograph of *E. ehracteolata* root.

dissolved in ethanol to obtain stock solutions of 1 mg/ml^{-1}. A number of 4 reference solutions were mixed and diluted with ethanol to obtain a series of mixture solutions. All the solutions were stored at −20°C before use.

Quantitation of active components was performed on an ultra-high-performance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS, LCMS 8045, Shimadzu, Japan). An ACQUITY BEH C_{18} column (2.1 mm × 100 mm, 1.7 μm) was used to separate diterpenoids. The mobile phase was composed of formic acid in water (0.05%, V/V) as solvent A and formic acid in ACN (0.05%, V/V) as solvent B, with a flow rate of 0.4 ml/min^{-1} at 40°C. The optimized gradient elution was as follows: 0–6 min, 30–93%B; 6–9 min, 93%B; 9–9.2 min, 93–30%B. About 1 μl of the solution was injected. The mass system was equipped with an electrospray ionization (ESI) source operating in both positive and negative ion modes, using multiple reaction monitoring (MRM) mode. Optimized parameters are listed in **Supplementary Table S1**. The nebulizer and drying gas were 99.95% nitrogen, and their flow rates were 3.0 and 10.0 L/min^{-1}, respectively. The heating gas was 99.95% air with a flow rate of 10.0 L/min^{-1}. The collision gas was 99.99% argon with a pressure of 270 kPa. Other parameters were as follows: interface voltage 4.0 kV, interface temperature 300°C, desolvation line temperature 250°C, and heat block temperature 400°C. The precision [relative standard deviation (RSD) < 5.77%] and accuracy (recovery was from 95.9 to 105.1%) of this method met the requirements for quantitative determination. Linearity was verified using coefficients of determination (R2), which were all > 0.999 within the adopted linear range (**Supplementary Material**).

Development and Detection of EST-SSR Markers

The B MISA (http://pgrc.ipk-gatersleben.de/misa/) (Dieringer and Schlötterer, 2003) was used to identify and localize microsatellite motifs in the two Langdu species and SSRs were considered to contain motifs of two to six nucleotides and a minimum of five contiguous repeat units. The alignments of 6,153 pairs of orthologs were extracted as the input file for the MISA program. Using the detailed information on SSR loci obtained from the output of the MISA program, primers for each SSR-containing sequence with a repetitive at least 15 bp in length were designed with Program Primer Premier 5 (PREMIER Biosoft Int., Palo Alto, CA). To validate the SSRs identified *in silico* identified SSRs, 68 primer pairs shared between the two Langdu species were synthesized [Sangon Biotech (Shanghai) Co., Ltd., Shanghai, China] and validated by polymerase chain reaction (PCR) in 9 species, including *Euphorbia fischeriana* (Hailar, Inner Mongolia), *Euphorbia ebracteolata* (Jiyuan, Henan province), *Euphorbia lathyris* (Lijiang, Yunnan province), *Euphorbia pekinensis* (Chuzhou, Anhui province), *Euphorbia sieboldiana* (Wang qing, Jilin province), *Euphorbia kansui*

TABLE 1 | Summary of assembly results for *E. fischeriana* and *E. ehracteolata* using trinity.

Sequence	*E. fischeriana*			*E. ehracteolata*		
	Root	Stem	Leaf	Root	Stem	Leaf
Raw reads	50845208	46010830	49087484	61659938	56456010	51479006
Clean reads	49966652	45210234	48124262	60048398	54973026	50043212
Q20 (%)	97.09	97.07	96.92	97.17	97.2	97.13
GC (%)	43.54	42.53	42.83	42.62	42.62	42.51
Total number of contigs/unigenes		149,207/112,487			143,857/101,211	
Length range of contigs/unigenes		200–15,623			200–16,340	
N50 value of contigs /unigenes		1,559/1,051			1,635/1,155	
Mean length of contigs/unigenes		813/631			885/686	
Median length of contigs/unigenes		398/325			459/366	

(Yongji, Shanxi Province), *Euphorbia humifusa* (Yongji, Shanxi Province), *Euphorbia helioscopia* (Jiyuan, Henan province), and *Euphorbia esula* (Ewenke, Inner Mongolia autonomous region). According to the Flora of China (Li et al., 2008), samples are all commonly used Chinese medicine. Fresh leaves were dried in silica gel at the time of collection. Voucher specimens were deposited in National Resource Center for Chinese Materia Medica (Beijing, China).

Genomic DNA was extracted from the dry leaves using Plant Genomic DNA Kit (TIANGEN). PCRs were performed in a 25 μl volume containing 50 ng of template genomic DNA. The PCRs were carried out under the following conditions: initial denaturation at 94°C for 2 min, 35 cycles at 94°C for 30 s, 55°C for 30 s, 72°C for 40 s, and a final extension at 72°C for 10 min. The separation of alleles was performed on a 3% polyacrylamide gel with a 500-bp DNA marker (TaKaRa) to calculate the length of the EST-SSR amplicons.

RESULTS

Differences in Habitats and Phenotypes Between the Two Langdu Species

To study these two closely related species, we collected wild *E. fischeriana* in Hailar, Inner Mongolia and wild *E. ehracteolata* in Jiyuan, Henan province, which are the main producing areas of two species, respectively. The average annual temperature in Hailar is −2.0°C, and that in Jiyuan is 14.6°C (data from www.nmc.cn). Therefore, in terms of habitat distribution, *E. fischeriana* is more tolerant of coldness than *E. ehracteolata*. This adaptive distribution might also result in phenotypic differences between the two Langdu species. Classification of the two species majorly depends on whether the ovary and capsule have trichome (**Figures 1A,F**) (Li et al., 2008), but the identification and comparison of the experimental materials showed that their roots (medicinal part) are also significantly different. The root of *E. fischeriana* is elongated and enlarged with multiple layers of semi-exfoliated epidermis, whereas *E. ehracteolata* is fusiform and the outer epidermis is tight (**Figures 1D,E**). In root of *E. fischeriana*, latex is white (**Figure 1B**) with concentric ring vascular bundle (**Figure 1C**), whereas yellow latex (**Figure 1G**) and abnormal vascular bundle (**Figure 1H**) are obtained in *E. ebracteolata*. The medicinal materials from the two species are similar in appearance, mostly in transverse slices or oblique slices, collectively referred to as "white Langdu," which is easy to be confused. Therefore, we further developed biomarkers to assist the identification of medicinal materials.

De novo Assembly and Functional Annotation

To reveal the molecular mechanism of adaptive distribution, cDNA libraries of the root, stem, and leaf of *E. fischeriana and E. ehracteolata* were sequenced, with 169594954 and 145943522 raw reads obtained (**Table 1**). The average Q20 percentage (percentage of sequences with sequencing error rate lower than 1%) was 97.17 and 97.03%. The average GC percentage was 42.58 and 42.97%. Using the trinity *de novo* assembly program, we assembled the short read sequences from the three tissues into 149027 transcripts, with a mean length of 813 bp and an N50 length (the contig size such that 50% of the entire assembly is contained in contigs equal to or longer than this value) of 1,559 bp for the *E. fischeriana*. For *E. ehracteolata*, 143,857 contigs longer than 200 bp, with a mean length of 885 bp and an N50 length of 1,635 bp, were generated. In total, all contigs were connected into 112,487 unigenes with a mean length of 631 bp and an N50 of 1,051 bp for *E. fischeriana* and 101,211 unigenes with a mean length of 686 and an N50 value of 1,115 for *E. ehracteolata*.

To obtain comprehensive gene function information, we carried out gene function annotation of seven databases, including Nr, Nt, Pfam, KOG/COG, Swiss-Prot, KEGG, and GO (**Supplementary Table S2**). The all unigenes were assigned putative gene descriptions based on the BLAST (*E*-value $\leq 1 \times 10e^{-5}$) search against the NCBI non-redundant (Nr) protein database. A total of 36,204 (32.18%) unigenes for *E. fischeriana* and 38,018 (37.56%) unigenes for *E. ehracteolata* were shown significant similarity with proteins in the Nr database. There were 48,033 (42.70%) unigenes for *E. fischeriana* and 44,919 (44.38%) for *E. ehracteolata*

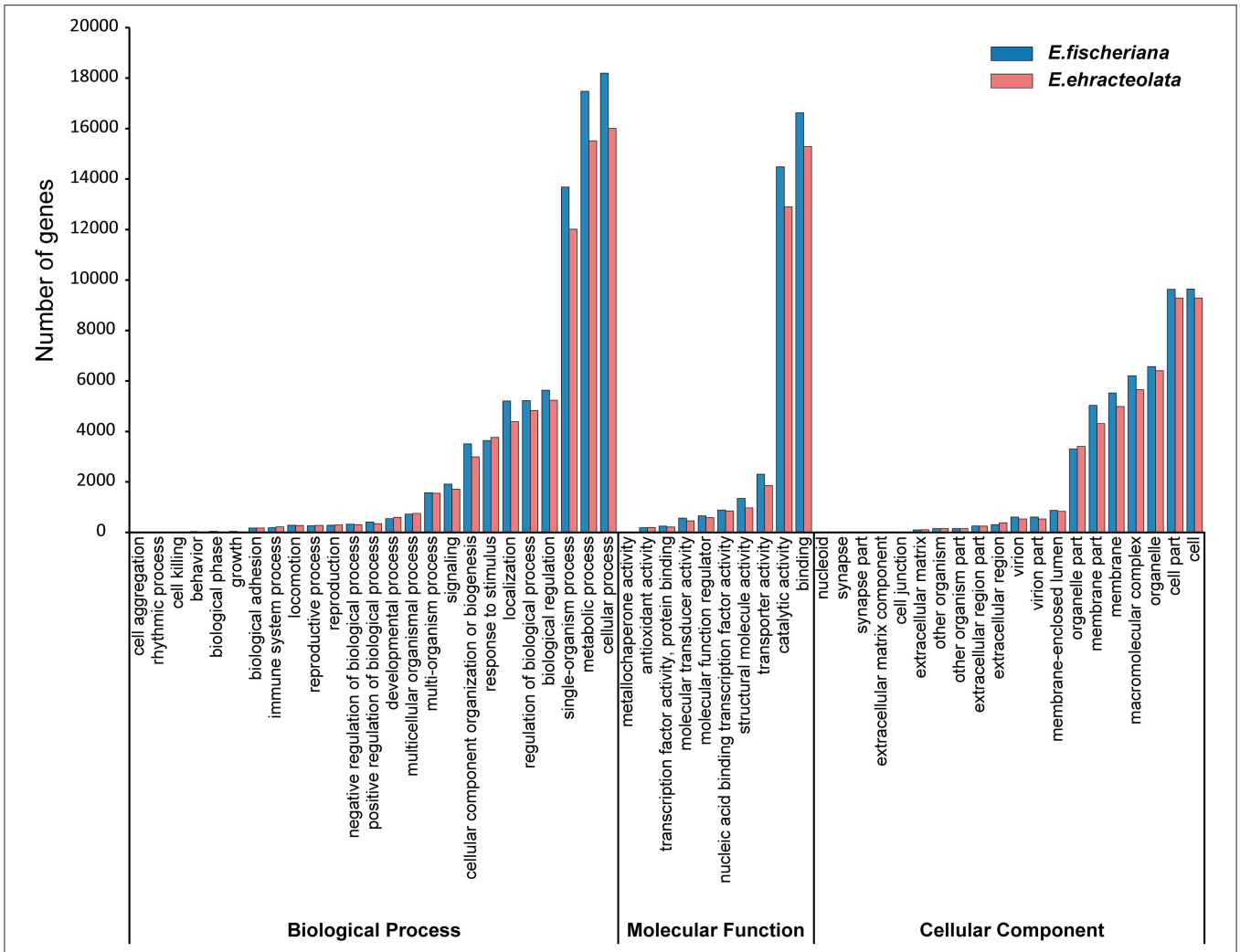


FIGURE 2 | Functional annotation of assembled sequences based on gene ontology (GO) categorization.

with at least one significant match to the above databases. For both species, a BLASTX top-hit species distribution of gene annotations showed highest homology to *Jatropha curcas* [*E. fischeriana* (38.49%) vs. *E. ehracteolata* (39.02%)]. Based on Nr annotations, we used the GO classification system to functionally categorize unigenes. A total of 28,472 (28.13%) unigenes for *E. ehracteolata* and 32,598 (28.97%) unigenes for *E. fischeriana* were assigned to at least one GO term annotation. The unigenes were assigned to three main GO categories (**Figure 2**): biological process (*E. ehracteolata*: 71,495, 47.30%; *E. fischeriana*: 79,428, 47.91%), molecular function (33,335, 22.06 vs. 37,307, 22.51%), and cellular component (46,314, 30.64 vs. 49,036, 29.58%). These categories were similarly distributed in both species.

For the biological process category, the two mostly highly represented terms among the 25 level-2 categories were cellular process, metabolic process, and single-organism process; for the molecular function category, among the 10 level-2 categories, binding and catalytic activity were overrepresented; for the 20 level-2 categories in the cellular component category, cell, cell part, and organelle were the most abundant terms (**Figure 2**). These categories were similarly distributed in both species. Specifically, compared with *E. ehracteolata*, more genes in *E. fischeriana* were involved in biological process and molecular function, including metabolic process (15,517, 10.27 vs. 17,479, 10.54%), localization (4,390, 2.91 vs. 5,209, 3.14%), cellular component organization or biogenesis (2,994, 1.98 vs. 3,508, 2.12%), positive regulation of biological process (351, 0.23 vs. 416, 0.25%), molecular transducer activity (451, 0.30 vs. 571, 0.34%), structural molecule activity (975, 0.65 vs. 1,347, 0.81%), transporter activity (1,853, 1.23 vs. 2,305, 1.39%), and so on. However, genes involved in biological phase (34, 0.02 vs. 42, 0.03%), growth (36, 0.02 vs. 46, 0.03%), immune system process (183, 0.11 vs. 226, 0.15%), and developmental process (544, 0.33 vs. 598, 0.40%) showed a conversed trend, comparing *E. fischeriana* with *E. ehracteolata*.

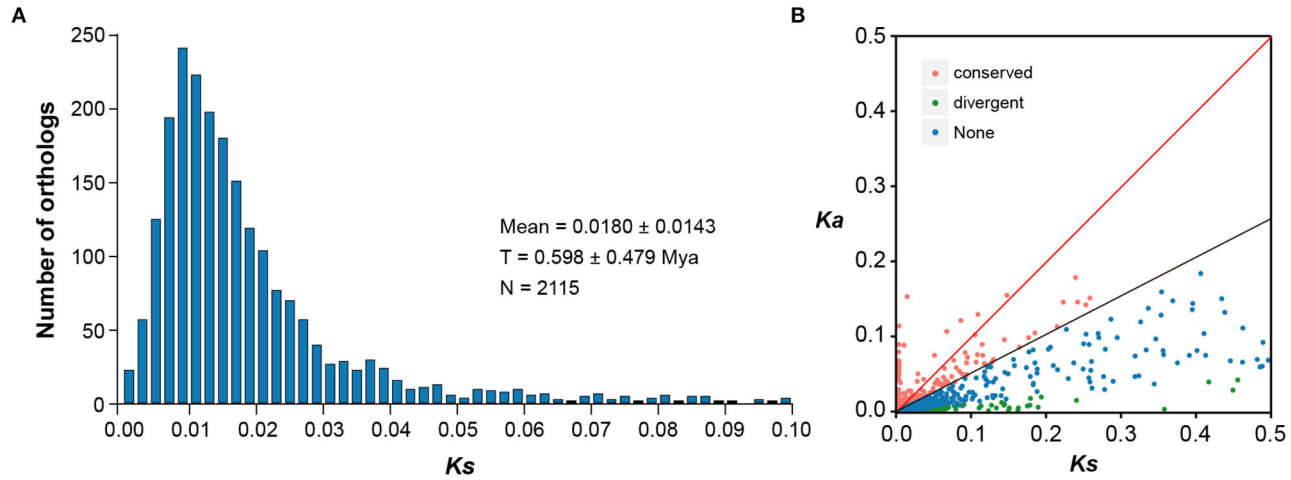


FIGURE 3 | Distribution of Ka and Ks for 6,830 pairs of the putative orthologous genes between *E. fischeriana* and *E. ehracteolata*. **(A)** The Ks distribution of orthologs between *E. fischeriana* and *E. ehracteolata* and their divergence is shown by the peak of Ks at 0.018 ± 0.014. **(B)** The orthologs with *Ka/Ks*>1 fall above the red line whereas those with Ka/Ks = 0.5–1.0 fall between the black and red lines.

Orthologous Genes and Substitution Rates Between Two Langdu Species

Based on the predicted CDS regions of both *E. fischeriana* and *E. ehracteolata* transcriptomes, we identified initial putative orthologous pairs. After removing the pairs with unexpected stop codons in the BLAST hit region and/or shorter than 150 bp in length, 6,830 ortholog pairs were retained. Out of the 6,830 ortholog pairs, 3,193 pairs had both non-synonymous (Ka) and synonymous (Ks) substitutions and thus were allowed for the calculation of Ka/Ks ratios. A peak of Ks value distribution between *E. fischeriana* and *E. ehracteolata* was observed at 0.0180 ± 0.0143 (**Figure 3A**), and the low level of Ks value indicates close relationship. It was estimated by the formula T = K/2r that the age of the speciation event between *E. fischeriana* and *E. ehracteolata* is ∼0.598 Mya, which falls in the Middle Pleistocene. Among the 3,193 ortholog pairs, 533 pairs showed high Ka/Ks values (**Figure 3B**), in which 191 pairs have a Ka/Ks > 1 ($p < 0.05$), indicating positive selection, whereas 342 pairs have a Ka/Ks between 0.5 and 1, indicating weak positive selection. By contrast, most of the orthologous pairs (2,658) showed a Ka/Ks ratio < 0.5, indicating that these gene pairs are likely under purifying selection.

Gene Functions Under Positive Selection and Implications for Adaptive Evolution Between Two Langdu Species

In enrichment analysis, we used the orthologs into one dataset with Ka/Ks > 0.5 (Swanson et al., 2004). In KEGG database, the positive selection orthologous genes were enriched to 57 pathways, including (1) citrate cycle (TCA cycle); (2) amino acid biosynthesis and metabolism including arginine biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, cysteine and methionine metabolism, glycine, serine and threonine metabolism, as well as valine, leucine, and isoleucine degradation; (3) fructose and mannose metabolism; (4) starch and sucrose metabolism; (5) fatty acid biosynthesis; and (6) terpenoid backbone biosynthesis (**Figure 4A**). The expression levels of orthologous genes in pathway are also different in two Langdu species (**Figure 4B**).

In an analysis of GO terms with at least five hits, 31 GO terms annotated to 76 pairs of orthologs were found to be over-represented (p-value < 0.05) in the test dataset. For the selected genes, we used BLASTX search to find their orthologous genes, and genes with function in stress tolerance, development, TCA-related, and so on were filtered (**Supplementary Table S3**). In particular, half of terms involved genes were related to stress tolerance, including cold stress, heat stress, and drought stress. Terms containing cold stress tolerance gene accounted for 30%, including *AtGCN1* (Wang et al., 2017b), *AtYLS9* (Griebel et al., 2022), *BrRZFP* (Jung et al., 2013), *AtPTP1* (Liu et al., 2016), and *AtPFD* (Perea-Resa et al., 2017). In terms of growth and development, we enriched genes related to trichome and root hair development, such as *AtBLT* (Kasili et al., 2011; Mazie and Baum, 2016), *AtVTI13* (Larson et al., 2014), *AtNRP2* (Zhu et al., 2017), and *AtWER1* (Wang et al., 2019). TCA-related genes were also enriched, including *AtSAM1* (Sekula et al., 2020), *AtLIP1* (Wang et al., 2022), *AtPES1* (Lippold et al., 2012), *AtCCDC* (Lohmeier-Vogel et al., 2008), and *AtBASS2* (Furumoto et al., 2011; Mueller et al., 2014; Lee et al., 2017), involving amino acid biosynthesis, catabolism of triacylglycerols, fatty acid, starch metabolism, pyruvate transport, etc.

Identification of Metabolites Differentially Accumulated in Two Langdu Species

Due to the above genetic differences, we further studied the metabolic differences between the two species, to reveal the effects of genotypes on chemical phenotypes. The chemical

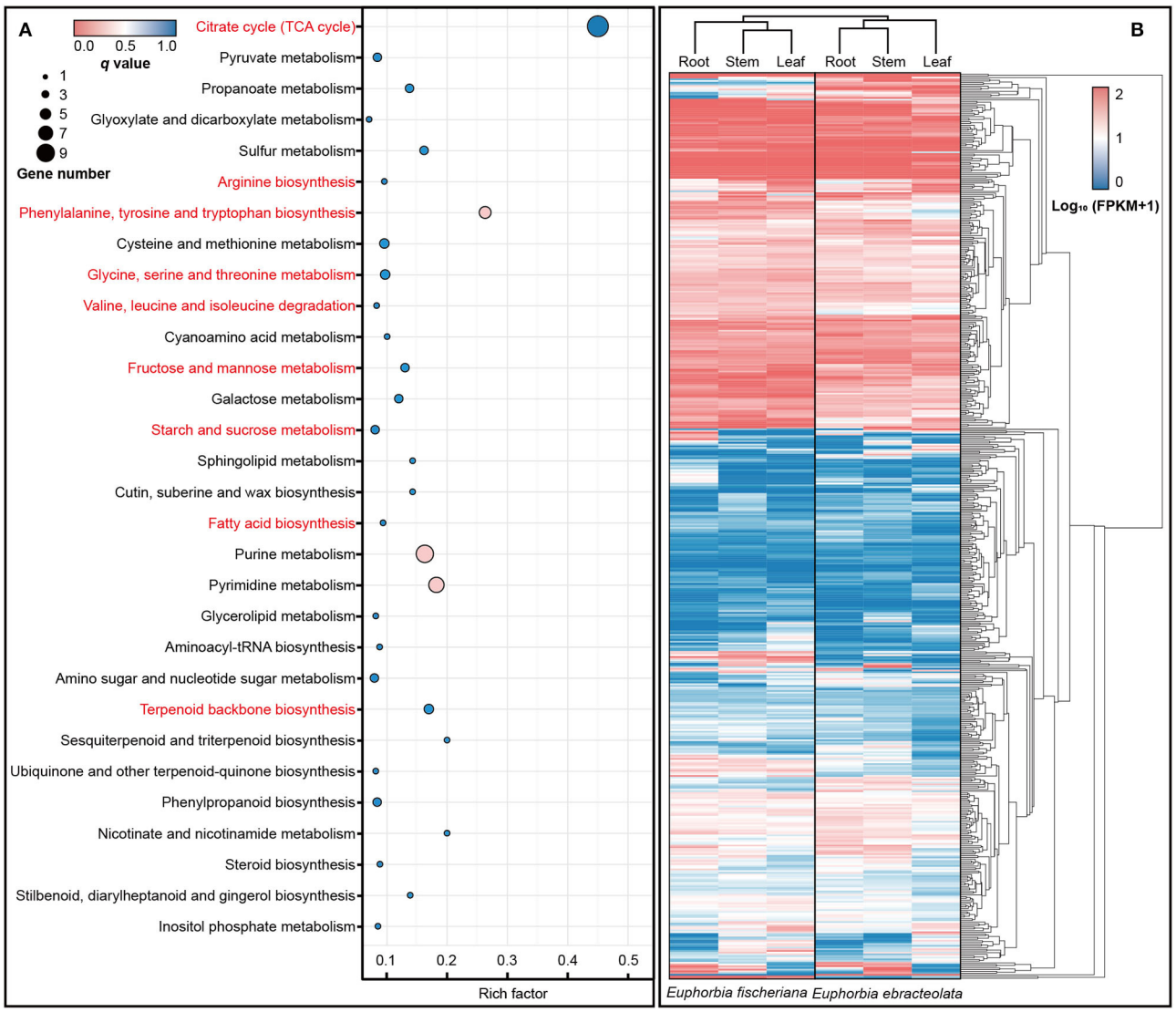


FIGURE 4 | KEGG enrichment and expression differences of orthologous genes. **(A)** KEGG enrichment of orthologous genes (Ka/Ks > 0.5); **(B)** Heat map of orthologous genes (Ka/Ks > 0.5) expression in roots, stems, and leaves.

components in roots and leaves were detected through UPLC–MS analysis, obtaining final data containing 1,220 features in ESI+ mode and 1,506 features in ESI- mode. An unsupervised principal component analysis (PCA) model was used, since it can represent the intrinsic characteristics of the data. The result revealed a clear separation of metabolite samples between *E. fischeriana* and *E. ehracteolata* ($R^2X = 0.724$, $Q^2 = 0.561$ in ESI+ mode; $R^2X = 0.598$, $Q^2 = 0.387$ in ESI– mode) (**Figures 5A,B**). The PCA score plot revealed that the leaves from *E. fischeriana* (blue triangle) and *E. ehracteolata* (red triangle) are not perfectly separated by the 12 samples. Similar to the appearance of the two Langdu leaves, there is less difference in the composition and content of their leaves. But the roots of *E. fischeriana* (blue circle) and *E. ehracteolata* (red circle) were clearly separated. The results indicated that there were significant differences in the compounds of Langdu roots. The PCA model was also applied to obtain a preliminary overview of general similarities and differences between collections. To identify discriminating metabolites and differentiate the four groups, we used the corresponding OPLS-DA analysis (**Figures 5C,D**). The two Langdu species can be separated clearly by only one predictive component deriving from a more sophisticated orthogonal partial least squares discriminate analysis (OPLS–DA) model, since noisy information irrelevant to species was removed prior to model building (Bylesjo et al., 2006). OPLS–DA analysis showed potential markers in different Langdu plants. A total of 98 discriminating metabolites (VIP > 1.0, $p < 0.05$), including 46 in positive mode ($R^2X = 0.758$, $R^2Y = 0.965$, $Q^2 =$

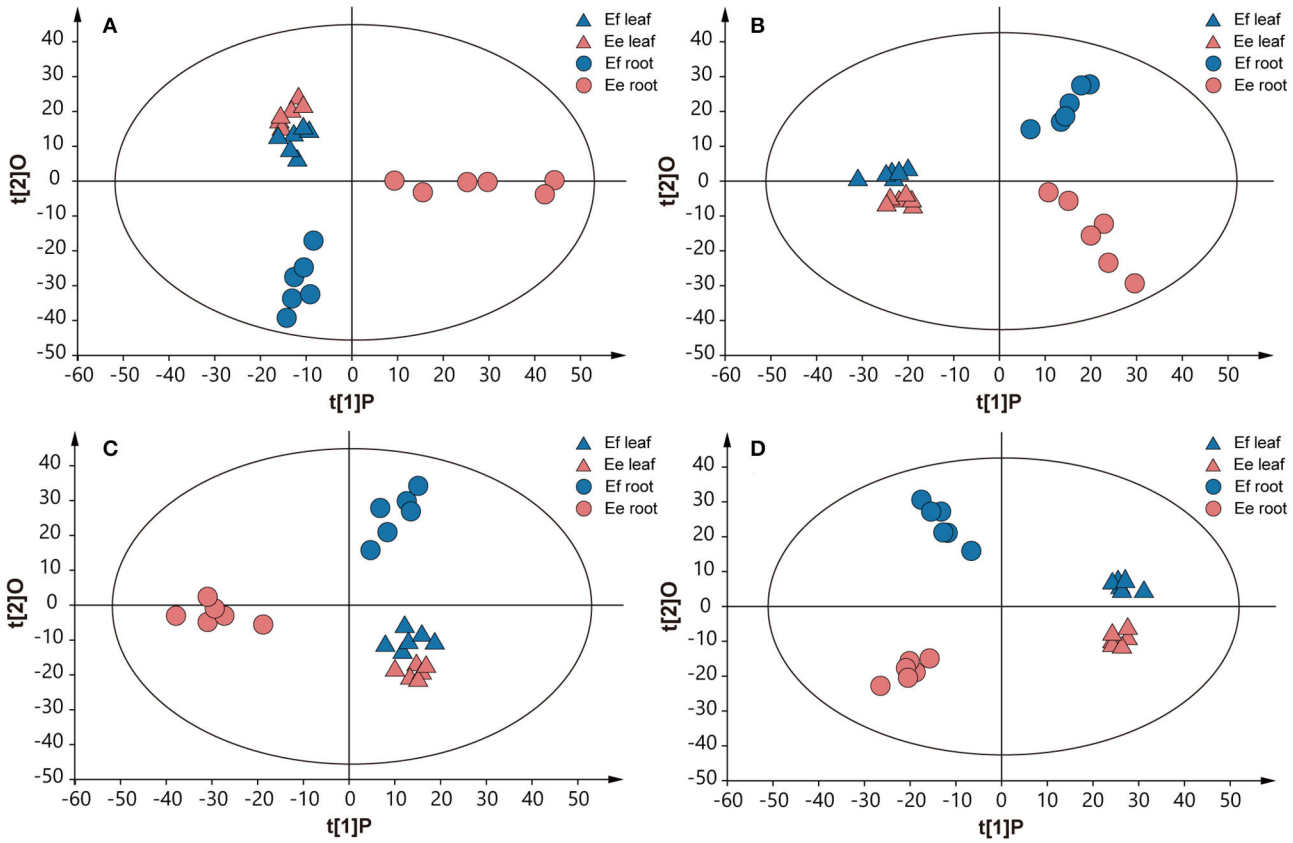


FIGURE 5 | Score plots of PCA and OPLS-DA. **(A)** PCA in positive mode. **(B)** PCA in negative mode. **(C)** OPLS-DA analysis in positive mode. **(D)** OPLS-DA analysis in negative mode. Multivariate analysis of metabolic profiles of root and leaf samples from *E. fischeriana* and *E. ehracteolata* with 6 biological replicates.

0.917) and 52 in negative mode ($R^2X = 0.593$, $R^2Y = 0.981$, $Q^2 = 0.926$), were identified in Langdu roots. The permutation result validated the stability and reliability of this OPLS–DA model. Subsequently, relying on the three criteria—variable importance in the projection (VIP) value of OPLS–DA model ≥ 1.5, *p*-value of *t*-test ≤ 0.01–5 metabolites could be presumably considered as candidate biomarkers. In addition, jolkinolide B and ingenol showed specific accumulation in the two species and could also serve as the biomarkers (**Supplementary Table S4**).

To reveal the effect of environments on the metabolites, the tentatively identified compounds were assigned in the common metabolic pathways according to the literature works (Schauer et al., 2006; Duan et al., 2012). Combined with previous KEGG and GO enrichment results of adaptive evolution genes, we found that the contents of malic acid, fumaric acid, GABA, amino acids, and other compounds in TCA in *E. fischeriana* were significantly lower than in *E. ehracteolata*. But some soluble sugars, such as glucose, sucrose, raffinose, xylulose, and mannitol, were increased in *E. fischeriana* in comparison with the *E. ehracteolata*. Specifically, proline in *E. fischeriana* was significantly higher than that in *E. Ehracteolata*, but unsaturated fatty acids including linoleic acid and α-linolenic acid were lower (**Figure 6**).

Differences in Expression Levels of Orthologous Genes and Active Ingredients

The differences in active ingredients between two Langdu species may result from the differences in genotype and gene expression. *DXR* (OG06739) and *MK* (OG07880) were found in the positive selection library (Ka/Ks > 0.5) mentioned in the previous study, indicating that these two genes had adaptive mutations. We also analyzed the expression levels of MVA and MEP pathways and terpene synthase homologous genes globally and found that the gene expressive abundance of *E. fischeriana* was higher than that of *E. ehracteolata*, especially *DXS*, *DXR*, *CMK*, *HDS*, *HDR*, *HMGS*, *HMGR*, *MK*, and diterpene synthase genes *CPS* and *KSL* in the roots. However, the expression level of *CBS* in *E. ehracteolata* root was significantly higher than that of *E. fischeriana*, and even *CBS* (OG17313) was not expressing in *E. fischeriana* root.

To analyze the differences in active ingredients, a UPLC-MS/MS method simultaneously determining 4 diterpenoids was established, and the root and leaf tissues of the two Langdu species were detected (**Figure 7**). The content of jolkinolide A was about 0.31 mg/g^{-1} in the roots of *E. fischeriana* and 0.10 mg/g^{-1} in the roots of *E. ehracteolata*. Jolkinolide B was significantly accumulated in the roots of *E. fischeriana* (1.13

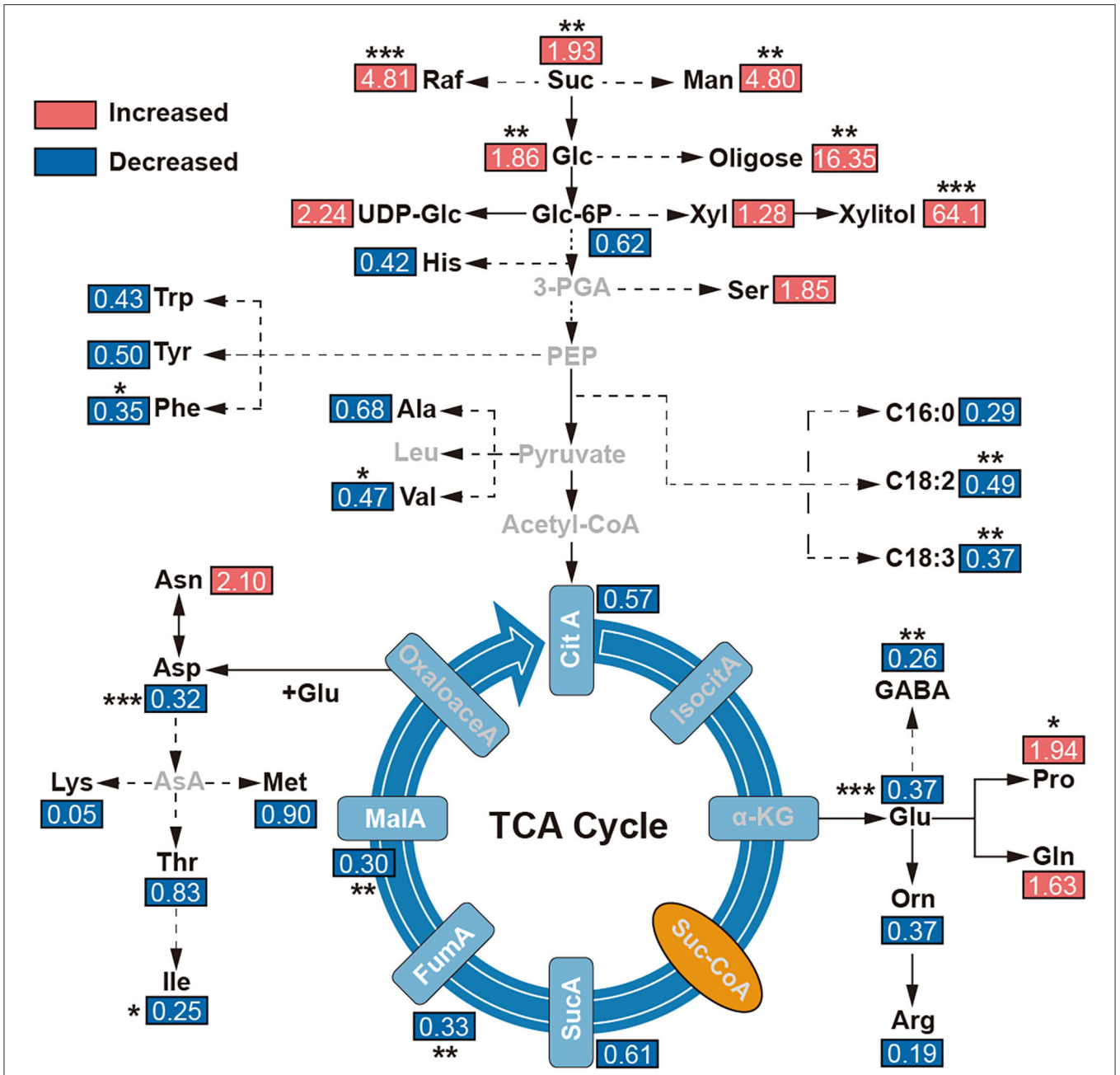


FIGURE 6 | Levels of metabolites in main metabolic maps. The comparisons of metabolite contents were calculated by dividing the metabolite level in *E. fischeriana* with that in *E. ehracteolata*. 3-PGA, glycerate 3-phosphate; α-KG, alpha-ketoglutarate; Arg, arginine; ASA, aspartic acid- β-hemialdehyde, Asn, asparagine; Asp, aspartic acid; Ala, beta-alanine; CitA, Citric acid; PalA, palmitic acid; OctA, octadecanedioic acid; LinA, linoleic acid; α-Lin, α-linolenic acid; FumA, fumarate; GABA, γ-aminobutyric acid; Glc, glucose; Glc3P, glucose-3-phosphate; Gln, glutamine; Glu, glutamate; Gly, glycine; Ile, isoleucine; IsocitA, isocitric acid; Leu, leucine; Lys, lysine; MalA, malate; Man, mannitol; Met, methionine; Orn, ornithine; Pro, proline; Raf, raffinose; Ser, L-serine; Stigm, stigmasterol; Suc, sucrose; SucA, succinic acid; Suc-CoA, succinate coenzyme A; Thr, threonine; Trp, tryptophan; Val, valine; Xly, xylulose. The value represents the ratio of *E. fischeriana* to *E. ehracteolata*. Asterisks denote Student's *t*-test significance: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

mg/g^{-1}) and slightly accumulated in the leaves (0.03 mg/g^{-1}), whereas the content of jolkinolide B in the roots of *E. ehracteolata* was 0.03 mg/g^{-1}. Jolkinolide E reached 0.16 mg/g^{-1} in the root of *E. fischeriana* and 0.06 mg/g^{-1} in *E. ehracteolata*. Conversely, bicyclic diterpenoid ingenol was only detected in the roots of *E. ehracteolata*, about 0.04 mg/g^{-1}. Obviously, there were the differences in active ingredients biosynthesis genes and their expression levels between the two species during the evolution of environmental adaptation, which resulted in the preference of accumulation.

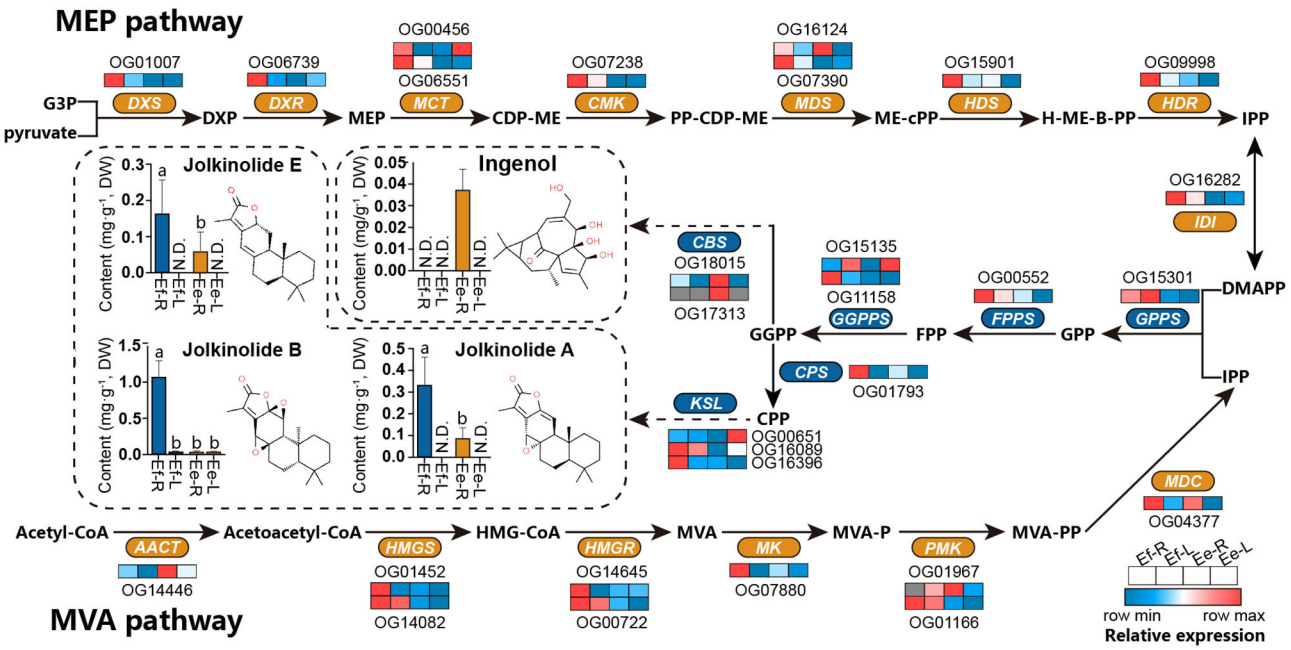


FIGURE 7 | Differences in active ingredients and the expression levels of orthologous genes in *E. fischeriana* and *E. ehracteolata*. MVA pathway, the mevalonate pathway; Acetyl-CoA, acetyl coenzyme A; acetoacetyl-CoA, acetoacetyl coenzyme A; HMG-CoA, 3-hydroxy-3-methylglutaryl CoA; MVA, mevalonate; MVA-P, mevalonate-5-phosphate; MVA-PP, mevalonate-5-diphosphomevalonate; IPP, isopentenyl pyrophosphate; MEP pathway, the 2-C-methyl- d -erythritol 4-phosphate pathway; G3P, glyceraldehyde 3-phosphate; DXP, 1-deoxy-d-xylulose 5-phosphate; MEP, 2-C-methyl-d-erythritol 4-phosphate; CDP-ME, 4-(cytidine 5'-diphospho)-2-C-methyl- d -erythritol; PP-CDP-ME, 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol; ME-cPP, 2-C-methyl-d-erythritol 2,4-cyclodiphosphate; H-ME-B-PP, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; CPP, copalyl diphosphate; AACT, acetoacetyl-CoA transferase; HMGS, hydroxymethyl-glutaryl-CoA synthase; HMGR, 3-hydroxy-3-methylglutaryl CoA reductase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; MDC, mevalonate-5-pyrophosphate decarboxylase; DXS, 1-deoxy-d-xylulose-5-phosphate synthase; DXR, 1-deoxy-d-xylulose-5-phosphate reductoisomerase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase; MDS, 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR: 4-hydroxy-3-methyl-but-2-enyl diphosphate reductase; IDI, Isopentenyl-diphosphate delta-isomerase; GPPS, GPP synthase; FPPS, FPP synthase; GGPS, GGPP synthase; CPS, copalyl diphosphate synthase; KSL, kaurene synthase-like; CBS, casbane synthase. N.D., below detection limit. Different letters denote statistically significant differences (ANOVA/LSD, $P < 0.05$).

Development of EST-SSR Markers in *Euphorbia* Species

Based on the *E. fischeriana* and *E. ehracteolata* transcriptomes, 8,458 and 10,612 EST-SSRs were identified, respectively. The most common repeat types were dinucleotides (44.0 and 44.5%), followed by trinucleotides (42.8 and 43.3%), and the most common repeat motifs were AG/CT, AT/AT, and AC/GT, followed by AAG/CTT and AAT/ATT (**Supplementary Table S5**). To maximize the universal applicability of markers developed in this study and hence reduce their cost, we searched for EST-SSRs in the 6,153 pairs of putative orthologous unigenes and found 967 EST-SSRs distributed among 780 pairs of orthologous unigenes. Taking only those with a repeat unit length of at least 15 bp, 68 pairs of EST-SSRs contained in 532 pairs of orthologs were selected for primer design, and 68 pairs of sequences with conserved, sufficiently long flanking sites were used to design primers successfully. To evaluate the reliability and cross-species transferability of these primers, we tested all the 68 pairs of EST-SSR markers for 9 species of *Euphorbia*. The average commonality ratio of primers in species was 35.69%, 49 pairs produced clear fragments with the expected sizes in both two Langdu and two pairs (OG00193 and OG07454) which produced clear fragments with the expected sizes in all nine species, respectively. A total of nine pairs produced fragments in eight species (**Supplementary Table S6**). Among them, OG07421, OG10880, and OG13687 could be amplified with obvious polymorphism bands in several *Euphorbia* species.

DISCUSSION

The Environmental Adaptation of Two Langdu Is Influenced by Genotype and Metabolites

E. fischeriana mostly grows in northern and northeast China, with high latitude, less rainfall, low annual average temperature. However, *E. ehracteolata* mostly grows in eastern and central China (Li et al., 2008). Temperature, as one of the main environmental variables, makes that the two species exhibit a typical pattern of adaptive evolution and explosive speciation, which may be influenced by genotype and metabolites. The positive selection orthologous genes were enriched to TCA

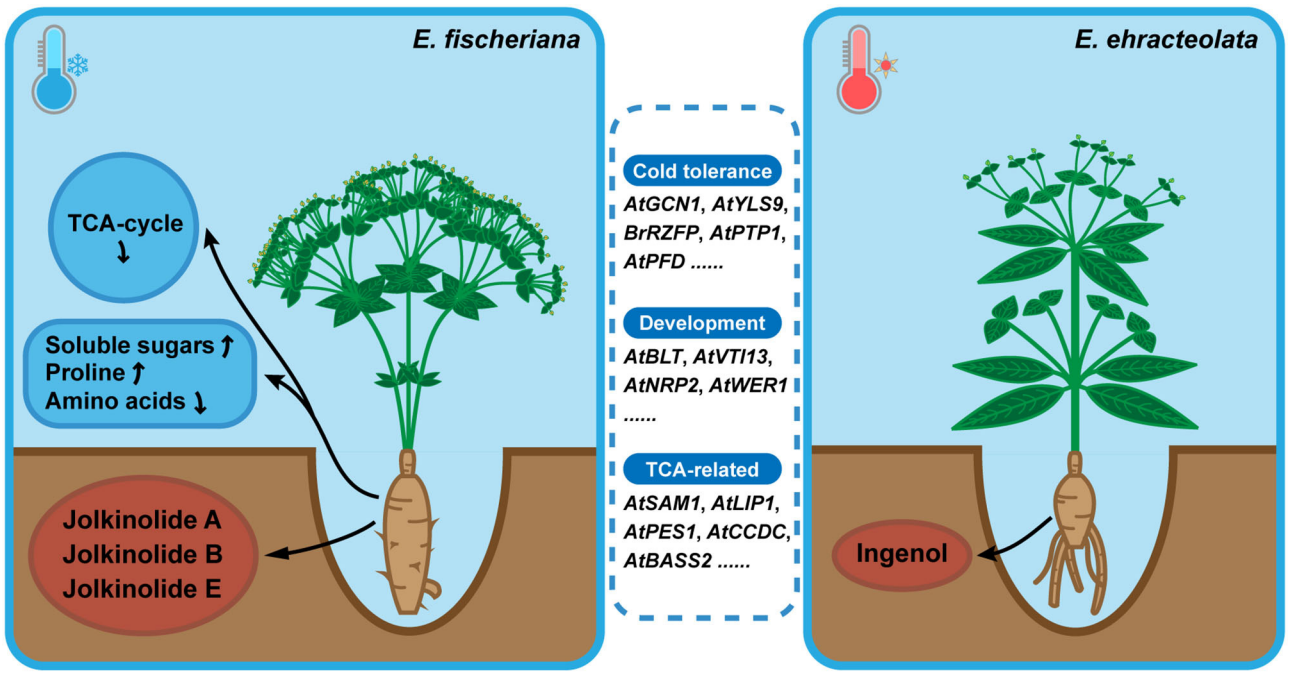


FIGURE 8 | Model for environmental adaptation mechanism of Langdu species.Compared with *E. ehracteolata* growing in eastern and central China, *E. fischeriana* growing in northern and northeast China (higher latitude) has the habitat characteristics of lower annual average temperature, presenting more tolerance of coldness than *E. ehracteolata*. In *E. fischeriana*, *AtGCN1*, *AtYLS9*, *BrRZFP, AtPTP1*, *AtPFD* related to cold stress tolerance, *AtBLT*, *AtVTI13*, *AtNRP2*, *AtWER1* participated in trichome, root hair and epidermal development, *AtSAM1*, *AtLIP1*, *AtPES1*, *AtCCDC*, *AtBASS2* related to TCA-cycle were enriched. These differences in positive selection genes may ultimately affect phenotypes and metabolites accumulation. In addition, content of soluble sugars and proline was significantly higher in *E. fischeriana*, whereas content of amino acids was relatively lower, compared with that in *E. ehracteolata*. Differential accumulation of active ingredients also existed in two Langdu species, during the evolution of environmental adaptation. *E. fischeriana* preferred to accumulate polycyclic diterpenes such as jolkinolide A, jolkinolide B, and jolkinolide E, whereas *E. ebracteolata* preferred to accumulate dicyclic diterpenes such as ingenol, reflecting different accumulation strategies of the same metabolites in related species, which is the result of long-term adaptation to the environment and also the material basis for the formation of pharmacodynamics.

cycle, stress tolerance, development, amino acid biosynthesis and metabolism, fructose and mannose metabolism, starch and sucrose metabolism, and fatty acid biosynthesis pathways (**Figure 8**). Among them, OG06839 is homologous to *AtGCN1*, which mediated phosphorylation of eIF2α and is essential for cold tolerance (Wang et al., 2017b); OG13992 is homologous to *AtYLS9*, which tolerates cold stress by activating an immune response (Griebel et al., 2022); OG06917 is homologous to *BrRZFP*, which regulates germination, fresh weight, and length of shoots and roots to tolerate stress (Jung et al., 2013); OG01773 is homologous to *AtPTP1* and encodes a protein with tyrosine phosphatase activity that is downregulated in response to cold stress (Liu et al., 2016); OG07015 is homologous to *AtPFD* and controls the levels of HY5, a key role in cold acclimation by activating anthocyanin biosynthesis, in response to low temperature (Perea-Resa et al., 2017). The positive selection of the plant physiology-related genes may be one of the reasons why the two species can survive in different cold environments.

Trichomes play important protective roles as against abiotic stressors such as cold, drought, heat, excess of light, and UV radiation (Wagner et al., 2004; Hauser, 2014). For example, OG07214, which is homologous with *AtBLT*, may be a key gene in trichome development (Kasili et al., 2011; Mazie and Baum, 2016). The non-synonymous substitution of OG07214 may cause formation of trichome in *E. fischeriana* fruits (**Figures 1A,F**). In addition, *AtVTI13* (OG07291) is essential for the maintenance of cell wall organization and root hair growth (Larson et al., 2014); *AtNRP2* moderates chromatin structure for proper root hair development (Zhu et al., 2017). *AtWER1* plays an important role in generating the proper balance of downstream transcriptional factors in the gene regulatory network that establishes root epidermal cell fate (Wang et al., 2019). These gene differences may explain the larger root crown and more root epidermis of *E. fischeriana* (**Figures 1D,E**).

The accumulation of pyruvate, polyamines, fatty acid, and starch metabolism helps plants to tolerate cold. *AtSAM1* is responsible for production of S-adenosylmethionine, which promotes the accumulation of polyamines and phytohormone ethylene (Sekula et al., 2020); *AtLIP1* is a triacylglycerol lipases, which negative regulates cold tolerance (Wang et al., 2022); *AtPES1* involves in fatty acid phytyl ester synthesis in chloroplasts (Lippold et al., 2012); *AtCCDC* promotes starch metabolism by interacting with several potential enzyme (Lohmeier-Vogel et al., 2008); *AtBASS2* is a plastidial sodium-dependent pyruvate transporter (Furumoto et al., 2011; Mueller et al., 2014; Lee et al., 2017). Positive selection in above genes leads to non-synonymous

substitutions of key amino acid sites that alter enzyme activity, thereby affecting the accumulation of metabolites and promoting environmental adaptation.

These genetic changes were reflected in the differences in the accumulation of components in the two Langdu species. First, *E. fischeriana* reduces the rate of the TCA cycle and maintains low respiration (Close, 1997). At the same time, most amino acid nutrients are reduced to ensure that carbon sources are mostly used to supply protective substances such as glucose, sucrose, raffinose, and xylulose (**Figure 6**). These soluble sugars are accumulated in large quantities, protecting cell membranes and proteins from freezing and dehydration under cold stress (Klotke et al., 2004). Proline and mannitol also accumulate in *E. fischeriana*, which is due to response to abiotic and biological stresses by affecting osmotic pressure tolerance within plants (Szabados and Savoure, 2010). The accumulation of different metabolites in the roots of two Langdu species may reflect their adaptation to different environments.

Phenotypic formation is the result of both genotype and environmental modification. The phenotype of genuine medicinal materials includes the characteristics, tissue structure, active ingredient composition, and efficacy. The existence of specific genes is the basis for producing specific phenotypes, and the suitable habitat is the driving force for producing specific phenotypes (Yuan and Huang, 2020). Overall, in this study, we detected positive selection genes and metabolites, and these findings will not only shed light on how differentiations between two Langdu species occurs, but also open the door to increased understanding of how plants living in cold environments adapt to different characteristics of high latitude.

Differences in Active Ingredients of Two Langdu Species Might Lead to Transition in Clinical Usage

Active ingredients, as a part of the phenotype, are also influenced by genotype and environmental modifications. Diterpenoids are regarded as the main bioactive constituents of *E. fischeriana* and *E. ebracteolata*, which can be classified into abietane, tigliane, atisane, pimarane, rosane, kaurane, ingenane, and lathyrane types (Li et al., 2021b; Yang et al., 2021). Among them, abietane type is the main diterpenoid type of the two species. A total of 53 compounds have been isolated from *E. fischeriana*, including jolkinolide B, 17-hydroxyjolkinolide B, jolkinolide A, and 26 compounds from *E. ebracteolate*. In this study, it was found that content of jolkinolide A, jolkinolide B, and jolkinolide E in *E. fischeriana* was 3.1, 36.7, and 2.7 times of that in *E. ebracteolate*, respectively. *DXR* (OG06739) and *MK* (OG07880), involved in diterpenoid biosynthesis, were found in the positive selection library (Ka/Ks $>$ 0.5) mentioned in the previous study, indicating that these two genes had adaptive mutations. Such variation may lead to the changes in enzyme activity and thus becomes more efficient in substrate catalysis, which requires further research. We also found that the most gene expression level of *E. fischeriana* was higher than that of *E. ehracteolata*, which may promote the accumulation of Jolkinolide A/B/E (**Figure 7**). Moreover, casbane type is unique to *E. ebracteolate*, including Yuexiandajisu A and Yuexiandajisu B (Xu et al., 1998). Casbene is thought to form bicyclic diterpenoids, which in turn forms ingenanes (Luo et al., 2016). A total of 20 ingenanes (ingenol and Esters) have been isolated from *E. ebracteolate* and only 6 from *E. fischeriana*. However, the expression level of *CBS* in the roots of *E. ebracteolate* was significantly higher than that of *E. fischeriana*, and OG17313 was not detected in the latter. We speculated that this was one of the reasons that the content of ingenol in *E. fischeriana* was higher than that in *E. ehracteolata*. However, more definitive evidence will need to be obtained through genome sequencing and biosynthesis pathway analysis. In conclusion, this reflects the different accumulation strategies of the same metabolites in related species, which is the result of long-term adaptation to the environment and also the material basis for the formation of pharmacodynamics.

Jolkinolide B and derivatives in *E. fischeriana* have been proved to have significant anticancer activity (Wang et al., 2009, 2017a). Pharmacological activity of ingenol in *E. ebracteolata* is not clear (Huang et al., 2019), but various esters of ingenol have shown the remarkable biological properties to mimic diacylglycerol and function as endogenous activators of protein kinase (PKC). Furthermore, they were found to have potential in treatment of pancreatic tumor, actinic keratosis (Siller et al., 2009; Parker et al., 2017), and HIV (Johnson et al., 2008). In conclusion, although *E. fischeriana* and *E. ebracteolata* have been used as the same traditional Chinese medicine for a long time, they should have different focuses on anticancer, anti-HIV, and actinic keratosis treatment in clinical use.

EST-SSRs Are Useful for Population Genetic Analysis Between *Euphorbia*

The development of EST-SSR primers is the further utilization of a large number of EST sequence information and has the characteristics of low cost. *E. fischeriana* and *E. ebracteolata* transcriptome data just provide this data resource. Because EST-SSRs are derived from the relatively conserved transcription part, EST-SSR has a higher translocation between species than genomic SSR, which provides a good tool for studying interspecies population inheritance (Varshney et al., 2005; Kalia et al., 2011). At present, few EST-SSR markers of Langdu have been reported involving *Euphorbia* species (Li et al., 2021a). In this study, 49 of 68 pairs of EST-SSR primers (72.1%) could amplify clear fragments in both Langdu species, 33 pairs (48.5%) were in at least 4 species, 9 were in 8 species, and 2 were in all 9 species (**Supplementary Table S6**). EST-SSR markers developed in *Euphorbia* showed mobility, indicating that these species may have close genetic relationship and that these markers may have new application value. EST-SSR markers developed based on two Langdu could provide genetic information for research on variety identification, genetic diversity analysis, and molecular marker-assisted breeding.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The two RNA-seq library datasets (root, stem,

and leaf) for this study can be found in the National Center for Biotechnology Information (NCBI) BioProject: PRJNA693977, PRJNA693983.

AUTHOR CONTRIBUTIONS

HZ coordinated the project, directed the work, and revised the manuscript. HZ, M-YY, and YH wrote the manuscript. S-FN, F-QL, and HX collected experimental samples. M-YY performed microsection and image processing. M-YY and C-JP performed an in-depth analysis of comparative transcriptome. X-TZ, BT, R-FJ, and KC performed metabolite determination and analysis. L-QH, X-TZ, and YS involved in funding. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by CACMS Innovation Fund (CI2021A041003), Natural Science Foundation of China (82104342), Fundamental Research Funds of CACMS (ZZ15-YQ-060), and Independent Studies Supported by CACMS (ZK2021001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.905275/full#supplementary-material

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Barrero, R. A., Chapman, B., Yang, Y., Moolhuijzen, P., Keeble-Gagnere, G., Zhang, N., et al. (2011). De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genomics* 12, 600. doi: 10.1186/1471-2164-12-600

Bluthgen, N., Brand, K., Cajavec, B., Swat, M., Herzel, H., and Beule, D. (2005). Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform.* 16, 106–115. doi: 10.48550/arXiv.q-bio/0407034

Bylesjo, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., and Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometr.* 20, 341–351. doi: 10.1002/cem.1006

Close, T. J. (1997). Dehydrins: a commonality in the response of plants to dehydration and low temperature. *Physiol. Plant.* 100, 291–296. doi: 10.1111/j.1399-3054.1997.tb04785.x

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, 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

De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778–791. doi: 10.1038/nprot.2007.95

Dieringer, D., and Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* 3, 167–169. doi: 10.1046/j.1471-8286.2003.00351.x

Duan, L.-X., Chen, T.-L., Li, M., Chen, M., Zhou, Y.-Q., Cui, G.-H., et al. (2012). Use of the metabolomics approach to characterize chinese medicinal material Huangqi. *Mol. Plant* 5, 376–386. doi: 10.1093/mp/ssr093

Fiehn, O. (2002). Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.* 48, 155–171. doi: 10.1023/A:1013713905833

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., et al. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211–222. doi: 10.1093/nar/gkp985

Frajman, B., Zaveska, E., Gamisch, A., Moser, T., Schoenswetter, P., Arthofer, W., et al. (2019). Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s.l. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 134, 238–252. doi: 10.1016/j.ympev.2018.10.046

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., et al. (2011). A plastidial sodium-dependent pyruvate transporter. *Nature* 476, 472–U131. doi: 10.1038/nature10250

Graur, D., and Li, W. H. (2000). *Fundamentals of Molecular Evolution. 2nd Edn.* Sunderland, MA: Sinauer Associates.

Griebel, T., Schutte, D., Ebert, A., Nguyen, H. H., and Baier, M. (2022). Cold exposure memory reduces pathogen susceptibility in Arabidopsis based on a functional plastid peroxidase system. *Mol. Plant Microbe Interact.* doi: 10.1094/MPMI-11-21-0283-FI. [Epub ahead of print].

Hauser, M.-T. (2014). Molecular basis of natural variation and environmental control of trichome patterning. *Front. Plant Sci.* 5, 320. doi: 10.3389/fpls.2014.00320

Horn, J. W., van Ee, B. W., Morawetz, J. J., Riina, R., Steinmann, V. W., Berry, P. E., et al. (2012). Phylogenetics and the evolution of major structural characters in the giant genus *Euphorbia* L. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 63, 305–326. doi: 10.1016/j.ympev.2011.12.022

Huang, Y.-S., Lu, Y., Chen, C.-H., Lee, K.-H., and Chen, D.-F. (2019). Potent anti-HIV ingenane diterpenoids from *Euphorbia ebracteolata. J. Nat. Prod.* 82, 1587–1592. doi: 10.1021/acs.jnatprod.9b00088

Johnson, H. E., Banack, S. A., and Cox, P. A. (2008). Variability in content of the anti-AIDS drug candidate prostratin in samoan populations of *Homalanthus nutans. J. Nat. Prod.* 71, 2041–2044. doi: 10.1021/np800295m

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

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

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

Kasili, R., Huang, C.-C., Walker, J. D., Simmons, L. A., Zhou, J., Faulk, C., et al. (2011). *BRANCHLESS TRICHOMES* links cell shape and cell cycle control in *Arabidopsis* trichomes. *Development* 138, 2379–2388. doi: 10.1242/dev.058982

Klotke, J., Kopka, J., Gatzke, N., and Heyer, A. G. (2004). Impact of soluble sugar concentrations on the acquisition of freezing tolerance in accessions of *Arabidopsis thaliana* with contrasting cold adaptation - evidence for a role of raffinose in cold acclimation. *Plant Cell Environ.* 27, 1395–1404. doi: 10.1111/j.1365-3040.2004.01242.x

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248

Koenig, D., Jimenez-Gomez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2655–E2662. doi: 10.1073/pnas.1309606110

Kong, L., Xie, Y., Hu, L., Si, J., and Wang, Z. (2017). Excessive nitrogen application dampens antioxidant capacity and grain filling in wheat as revealed by metabolic and physiological analyses. *Sci. Rep.* 7, 43363. doi: 10.1038/srep43363

Larson, E. R., Domozych, D. S., and Tierney, M. L. (2014). SNARE VTI13 plays a unique role in endosomal trafficking pathways associated with the vacuole and

is essential for cell wall organization and root hair growth in arabidopsis. *Ann. Bot.* 114, 1147–1159. doi: 10.1093/aob/mcu041

Lee, E.-J., Oh, M., Hwang, J.-U., Li-Beisson, Y., Nishida, I., and Lee, Y. (2017). Seed-specific overexpression of the pyruvate transporter BASS2 increases oil content in arabidopsis seeds. *Front. Plant Sci.* 8, 194. doi: 10.3389/fpls.2017.00194

Li, B., Ma, J., Zhu, H., Gilbert, M. G., Esser, H.-J., Dressler, S., et al. (2008). *Flora of China.* Beijing; St. Louis: Science Press and Missouri Botanical Garden.

Li, H., Ma, Y., Pei, F., Zhang, H., Liu, J., and Jiang, M. (2021a). Large-scale advances in SSR markers with high-throughput sequencing in *Euphorbia fischeriana* Steud. *Electron. J. Biotechnol.* 49, 50–55. doi: 10.1016/j.ejbt.2020.11.004

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, Y.-N., He, J., Zhang, J., Shi, Y.-X., Guo, L.-B., Peng, Z.-C., et al. (2021b). Existing knowledge on Euphorbia fischeriana Steud. (Euphorbiaceae): traditional uses, clinical applications, phytochemistry, pharmacology and toxicology. *J. Ethnopharmacol.* 275, 114095. doi: 10.1016/j.jep.2021.114095

Lippold, F., vom Dorp, K., Abraham, M., Holzl, G., Wewer, V., Yilmaz, J. L., et al. (2012). Fatty acid phytyl ester synthesis in chloroplasts of *Arabidopsis*. *Plant Cell* 24, 2001–2014. doi: 10.1105/tpc.112.095588

Liu, S., Chen, H., Li, X., and Zhang, W. (2016). A low-temperature-responsive element involved in the regulation of the *Arabidopsis thaliana At1g71850/At1g71860* divergent gene pair. *Plant Cell Rep.* 35, 1757–1767. doi: 10.1007/s00299-016-1994-y

Lohmeier-Vogel, E. M., Kerk, D., Nimick, M., Wrobel, S., Vickerman, L., Muench, D. G., et al. (2008). Arabidopsis At5g39790 encodes a chloroplast-localized, carbohydrate-binding, coiled-coil domain-containing putative scaffold protein. *BMC Plant Biol.* 8, 120. doi: 10.1186/1471-2229-8-120

Luo, D., Callari, R., Hamberger, B., Wubshet, S. G., Nielsen, M. T., Andersenranberg, J., et al. (2016). Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of *Euphorbia lathyris* L. *Proc. Natl. Acad. Sci. U. S. A.* 16, 34–72. doi: 10.1073/pnas.1607504113

Mao, Y., Zhang, Y., Xu, C., and Qiu, Y. (2016). Comparative transcriptome resources of two *Dysosma* species (Berberidaceae) and molecular evolution of the CYP719A gene in Podophylloideae. *Mol. Ecol. Resour.* 16, 228–241. doi: 10.1111/1755-0998.12415

Mazie, A. R., and Baum, D. A. (2016). Clade-specific positive selection on a developmental gene: *BRANCHLESS TRICHOME* and the evolution of stellate trichomes in *Physaria* (Brassicaceae). *Mol. Phylogenet. Evol.* 100, 31–40. doi: 10.1016/j.ympev.2016.03.027

Miana, G. A., Riaz, M., Shahzad-ul-Hussan, S., Paracha, R. Z., and Paracha, U. Z. (2015). Prostratin: an overview. *Mini Rev. Med. Chem.* 15, 1122–1130. doi: 10.2174/1389557515666150511154108

Mueller, M., Kunz, H.-H., Schroeder, J. I., Kemp, G., Young, H. S., and Neuhaus, H. E. (2014). Decreased capacity for sodium export out of Arabidopsis chloroplasts impairs salt tolerance, photosynthesis and plant performance. *Plant J.* 78, 646–658. doi: 10.1111/tpj.12501

Parker, C. G., Kuttruff, C. A., Galmozzi, A., Jorgensen, L., Yeh, C.-H., Hermanson, D. J., et al. (2017). Chemical proteomics identifies SLC25A20 as a functional target of the ingenol class of actinic keratosis drugs. *ACS Central Sci.* 3, 1276–1285. doi: 10.1021/acscentsci.7b00420

Perea-Resa, C., Rodriguez-Milla, M. A., Iniesto, E., Rubio, V., and Salinas, J. (2017). Prefoldins negatively regulate cold acclimation in *Arabidopsis thaliana* by promoting nuclear proteasome-mediated HY5 degradation. *Mol. Plant* 10, 791–804. doi: 10.1016/j.molp.2017.03.012

Prenner, G., and Rudall, P. J. (2007). Comparative ontogeny of the cyathium in *Euphorbia* (Euphorbiaceae) and its allies: exploring the organ-flower-inflorescence boundary. *Am. J. Bot.* 94, 1612–1629. doi: 10.3732/ajb.94.10.1612

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454. doi: 10.1038/nbt1192

Sekula, B., Ruszkowski, M., and Dauter, Z. (2020). S-adenosylmethionine synthases in plants: structural characterization of type I and II isoenzymes from *Arabidopsis thaliana* and *Medicago truncatula*. *Int. J. Biol. Macromol.* 151, 554–565. doi: 10.1016/j.ijbiomac.2020.02.100

Siller, G., Gebauer, K., Welburn, P., Katsamas, J., and Ogbourne, S. M. (2009). PEP005 (ingenol mebutate) gel, a novel agent for the treatment of actinic keratosis: results of a randomized, double-blind, vehicle-controlled, multicentre, phase IIa study. *Australas. J. Dermatol.* 50, 16–22. doi: 10.1111/j.1440-0960.2008.00497.x

Su, T., Tan, Y., Tsui, M.-S., Yi, H., Fu, X.-Q., Li, T., et al. (2016). Metabolomics reveals the mechanisms for the cardiotoxicity of Pinelliae Rhizoma and the toxicity-reducing effect of processing. *Sci. Rep.* 6, 34692. doi: 10.1038/srep34692

Swanson, W. J., Wong, A., Wolfner, M. F., and Aquadro, C. F. (2004). Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168, 1457–1465. doi: 10.1534/genetics.104.030478

Szabados, L., and Savoure, A. (2010). Proline: a multifunctional amino acid. *Trends Plant Sci.* 15, 89–97. doi: 10.1016/j.tplants.2009.11.009

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

Wagner, G. J., Wang, E., and Shepherd, R. W. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* 93, 3–11. doi: 10.1093/aob/mch011

Wang, C.-J., Yan, Q.-L., Ma, Y.-F., Sun, C.-P., Chen, C.-M., Tian, X.-G., et al. (2017a). ent-Abietane and tigliane diterpenoids from the Roots of *Euphorbia fischeriana* and their inhibitory effects against Mycobacterium smegmatis. *J. Nat. Prod.* 80, 1248–1254. doi: 10.1021/acs.jnatprod.6b00786

Wang, L., Qian, B., Zhao, L., Liang, M.-H., Zhan, X., and Zhu, J. (2022). Two triacylglycerol lipases are negative regulators of chilling stress tolerance in arabidopsis. *Int. J. Mol. Sci.* 23, 3380. doi: 10.3390/ijms23063380

Wang, L. J., Li, H. H., Zhao, C. Z., Li, S. F., Kong, L. Y., Wu, W. W., et al. (2017b). The inhibition of protein translation mediated by AtGCN1 is essential for cold tolerance in *Arabidopsis thaliana*. *Plant Cell Environ.* 40, 56–68. doi: 10.1111/pce.12826

Wang, W., Ryu, K. H., Barron, C., and Schiefelbein, J. (2019). Root epidermal cell patterning is modulated by a critical residue in the WEREWOLF transcription factor. *Plant Physiol.* 181, 1239–1256. doi: 10.1104/pp.19.00458

Wang, Y., Ma, X., Yan, S., Shen, S., Zhu, H., Gu, Y., et al. (2009). 17-Hydroxy-jolkinolide B inhibits signal transducers and activators of transcription 3 signaling by covalently cross-linking janus kinases and induces apoptosis of human cancer cells. *Cancer Res.* 69, 7302–7310. doi: 10.1158/0008-5472.CAN-09-0462

Wanner, L. A., and Junttila, O. (1999). Cold-induced freezing tolerance in Arabidopsis. *Plant Physiol.* 120, 391–399. doi: 10.1104/pp.120.2.391

Wei, G., Dong, L., Yang, J., Zhang, L., Xu, J., Yang, F., et al. (2018). Integrated metabolomic and transcriptomic analyses revealed the distribution of saponins in *Panax notoginseng*. *Acta Pharmaceutica Sinica B* 8, 458–465. doi: 10.1016/j.apsb.2017.12.010

Xu, Z. H., Sun, J., Xu, R. S., and Qin, G. W. (1998). Casbane diterpenoids from *Euphorbia ebracteolata*? *Phytochemistry* 49, 149–151. doi: 10.1016/S0031-9422(97)00900-X

Yang, T., He, J., Yan, Y., Lian, W.-W., Xia, C.-Y., Xu, J.-K., et al. (2021). *Euphorbia ebracteolata* Hayata (Euphorbiaceae): a systematic review of its traditional uses, botany, phytochemistry, pharmacology, toxicology, and quality control. *Phytochemistry* 186, 112736. doi: 10.1016/j.phytochem.2021.112736

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297. doi: 10.1093/nar/gkl031

Yuan, Y., and Huang, L. (2020). Molecular pharmacognosy in Daodi herbs. *Chin. Sci. Bull. Chin.* 65, 1093–1102. doi: 10.1360/TB-2020-0184

Zhang, L., Yan, H.-F., Wu, W., Yu, H., and Ge, X.-J. (2013). Comparative transcriptome analysis and marker development of two closely related Primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics* 14, 329. doi: 10.1186/1471-2164-14-329

Zhao, K. J., X. G., Jin, R. L., and Xu, L. S. (1996). Herbalogieal study on traditional chinese drug langdu and its confused speeies lvru. *Chin. Trad. Herbal Drugs* 27, 554–558. doi: 10.7501/j.issn.0253-2670.1996.9.302

Zhao, S.-Y., Chen, L.-Y., Muchuku, J. K., Hu, G.-W., and Wang, Q.-F. (2016). Genetic adaptation of giant lobelias (*Lobelia aberdarica* and *Lobelia telekil*) to different altitudes in East African mountains. *Front. Plant Sci.* 7, 488. doi: 10.3389/fpls.2016.00488

Zhao, Y., Zhang, J., Zhang, Z., and Xie, W. (2019). Elymus nutans genes for seed shattering and candidate gene-derived EST-SSR markers for germplasm evaluation. *BMC Plant Biol.* 19, 102. doi: 10.1186/s12870-019-1691-4

Zhao, Y.-J., Cao, Y., Wang, J., and Xiong, Z. (2018). Transcriptome sequencing of *Pinus kesiya* var. langbianensis and comparative analysis in the Pinus phylogeny. *BMC Genom.* 19, 725. doi: 10.1186/s12864-018-5127-6

Zheng, H., Jing, L., Jiang, X., Pu, C., Zhao, S., Yang, J., et al. (2021). The ERF-VII transcription factor SmERF73 coordinately regulates tanshinone biosynthesis in response to stress elicitors in Salvia miltiorrhiza. *New Phytol.* 231, 1940–1955. doi: 10.1111/nph.17463

Zhu, Y., Rong, L., Luo, Q., Wang, B., Zhou, N., Yang, Y., et al. (2017). The histone chaperone NRP1 interacts with WEREWOLF to activate *GLABRA2* in arabidopsis root hair development. *Plant Cell* 29, 260–276. doi: 10.1105/tpc.16.00719

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

frontiers | Frontiers in Plant Science

REVIEW
published: 02 June 2022
doi: 10.3389/fpls.2022.924413

Biotechnological Approaches on Engineering Medicinal Tropane Alkaloid Production in Plants

Haiyue Gong[1], Ping He[2], Xiaozhong Lan[3], Lingjiang Zeng[1] and Zhihua Liao[1,2]**

[1]School of Life Sciences, Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City and Southwest University, The Provincial and Ministerial Co-founded Collaborative Innovation Center for R&D in Tibet Characteristic Agricultural and Animal Husbandry Resources, Southwest University, Chongqing, China, [2]Chongqing Academy of Science and Technology, Chongqing, China, [3]Xizang Agricultural and Husbandry College, The Provincial and Ministerial Co-founded Collaborative Innovation Center for R&D in Tibet Characteristic Agricultural and Animal Husbandry Resources, The Center for Xizang Chinese (Tibetan) Medicine Resource, TAAHC-SWU Medicinal Plant Joint R&D Centre, Nyingchi, China

OPEN ACCESS

Edited by:
Yongliang Liu, University of Kentucky, United States

Reviewed by:
Zheyong Xue, Northeast Forestry University, China
Yuan Yuan, China Academy of Chinese Medical Sciences, China

***Correspondence:**
Zhihua Liao
zhliao@swu.edu.cn
Lingjiang Zeng
zengling@swu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *20 April 2022*
Accepted: *09 May 2022*
Published: *02 June 2022*

Citation:
Gong H, He P, Lan X, Zeng L and Liao Z (2022) Biotechnological Approaches on Engineering Medicinal Tropane Alkaloid Production in Plants. Front. Plant Sci. 13:924413. doi: 10.3389/fpls.2022.924413

Hyoscyamine and scopolamine, belonging to medicinal tropane alkaloids (MTAs), are potent anticholinergic drugs. Their industrial production relies on medicinal plants, but the levels of the two alkaloids are very low *in planta*. Engineering the MTA's production is an everlasting hot topic for pharmaceutical industry. With understanding the MTA's biosynthesis, biotechnological approaches are established to produce hyoscyamine and scopolamine in an efficient manner. Great advances have been obtained in engineering MTA's production *in planta*. In this review, we summarize the advances on the biosynthesis of MTAs and engineering the MTA's production in hairy root cultures, as well in plants. The problems and perspectives on engineering the MTA's production are also discussed.

Keywords: biosynthesis, biotechnology, metabolic engineering, synthetic biology, tropane alkaloids

INTRODUCTION

Medicinal plants constitute a treasure for human beings to obtain diverse drugs, flavors and fine chemicals. Of them, medicinal plants of Solanaceae produce a kind of specialized metabolites named tropane alkaloids (TAs), some of which have potent anticholinergic activity. Hyoscyamine and scopolamine are well-known medicinal tropane alkaloids (MTAs) and anticholinergic drugs (Grynkiewicz and Gadzikowska, 2008). They are clinically used to treat asthma, pain, motion sickness, functional gastrointestinal disorders, Parkinson's syndrome and etc. (Huang et al., 2021). Furthermore, hyoscyamine is the key material for industrially producing ipratropium bromide, and scopolamine for tiotropium bromide. Ipratropium bromide, as well as tiotropium bromide, is an essential drug in the treatment of chronic obstructive pulmonary disease (COPD; Griffin et al., 2008). Therefore, the demand for the two alkaloids is huge. However, their production is limited due to low levels in medicinal plants producing TAs. To develop efficient ways to manufacture hyoscyamine and scopolamine, the scientists have been trying their best. Chemists successfully synthesized them, but failed in market because chemical synthesis was not economically availed (Grynkiewicz and Gadzikowska, 2008). Then, the final solution relies on biotechnological approaches, such as metabolic engineering and synthetic biology.

Two amino acids, including ornithine and phenylalanine, are the starting precursors for the two MTAs, hyoscyamine, and scopolamine. The intact biosynthetic pathway of MTAs has been elucidated, which is composed of 13 biosynthesis enzymes (**Figure 1**). Such MTA's biosynthesis enzymes and their corresponding genes provide valuable tools to engineering MTA's production in plants and yeast cells. Although the production of hyoscyamine and scopolamine were obtained in engineered yeast cells, their liter yields are extremely low (Srinivasan and Smolke, 2020). For hyoscyamine and scopolamine, their highest production were, respectively, 80 and 30 $\mu g\,L^{-1}$, very far away from the industry-requiring level, not less than $5\,g\,L^{-1}$ (Srinivasan and Smolke, 2020). Therefore, the industrial production of MTAs is, and will be dependent on plant materials, not only at present but in a long future period. In this review, we summarize engineering MTA's production in planta. The problems and perspectives on engineering MTA's production are discussed, as well.

ENGINEERING MEDICINAL TROPANE ALKALOID PRODUCTION IN HAIRY ROOTS

Hairy root cultures are often used to engineering production of plant natural products. Hyoscyamine, as well as scopolamine, is specifically synthesized in secondary roots of plants (Li et al., 2006; Matthew et al., 2014; Qiu et al., 2021). A lot of publications report engineering the biosynthesis of MTAs in hairy root cultures of various MTA-producing medicinal plants.

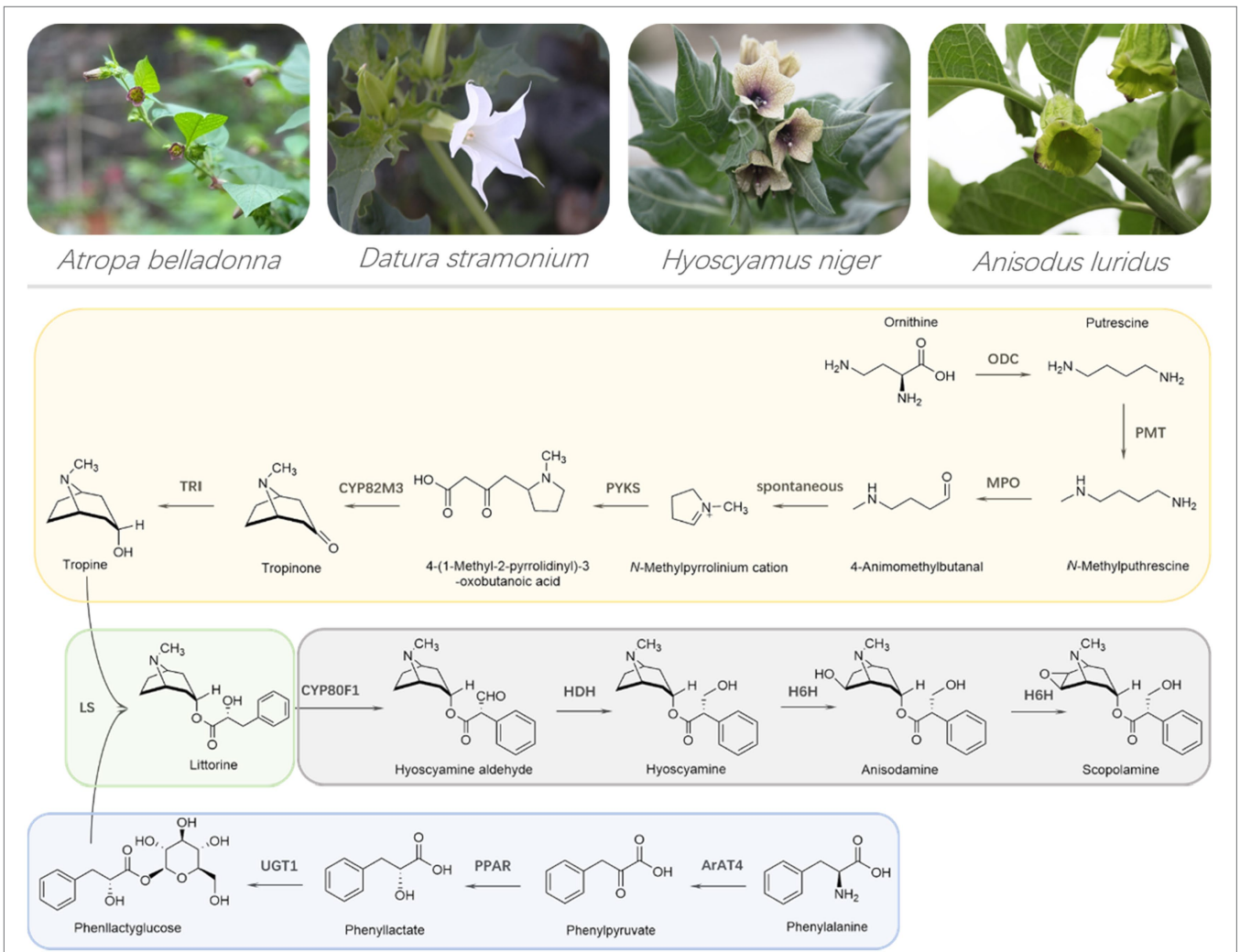


FIGURE 1 | Representative plant species producing medicinal tropane alkaloids and the intact biosynthetic pathway of medicinal tropane alkaloids. ODC, ornithine decarboxylase; PMT, putrescine *N*-methyltransferase; MPO, *N*-methylputrescine oxidase; PYKS, polyketide type III synthase; CYP82M3, tropinone synthase; TRI, tropinone reductase I or tropine-forming reductase; ArAT4, phenylalanine aminotransferase; PPAR, phenylpyruvic acid reductase; UGT1, phenyllactate UDP-glycosyltransferase; LS, littorine synthase; CYP80F1, littorine mutase; HDH, hyoscyamine dehydrogenase or hyoscyamine aldehyde reductase; H6H, hyoscyamine 6β-hydroxylase.

Overexpression of biosynthesis genes is a widely applied method to promote metabolite production. Seven biosynthesis genes required for MTA's biosynthesis are studied in engineering the production of MTAs using hairy root cultures of different plants. Such biosynthesis genes include: *ornithine decarboxylase* (*ODC*), *putrescine N-methyltransferase* (*PMT*), *tropinone reductase I* (*TRI*), *littorine mutase* (*CYP80F1*), *hyoscyamine dehydrogenase* (*HDH*), and *hyoscyamine β-hydroxylase* (*H6H*).

Engineering MTA's Production by Overexpressing a Single MTA Biosynthesis Gene

Ornithine decarboxylase is the rate-limiting enzyme catalyzing the conversion of ornithine to putrescine (Bacchi et al., 1980). In *Atropa belladonna*, the ODC enzyme (AbODC) rather than arginine decarboxylase (AbADC) contributes to the MTA's biosynthesis, and when *A. belladonna* ODC gene was overexpressed in hairy root cultures and plants of *A. belladonna*, the production of hyoscyamine and anisodamine was significantly increased (Zhao et al., 2020). Although ODC is located at the beginning of the MTA's biosynthetic pathway, its role in promoting MTA's production indicates that it is a key regulator directing ornithine into alkaloid metabolism (Zhao et al., 2020). Overexpression of *ODC* increases the putrescine level and provides metabolic flux enough for facilitating the downstream biochemical reactions (Zhao et al., 2020). However, such increased metabolic flux is not enough to support that *ODC* overexpression promotes MTA's production. Another explanation might be polyamine metabolism regulated by *ODC* (Nölke et al., 2008). Polyamines are plant growth regulators involved in diverse cell physiology. Since the *ODC* enzyme is a key primary-metabolism one regulating polyamine biosynthesis, this implies that it may have complicated synergic effects on MTA's biosynthesis through regulating polyamine biosynthesis.

Putrescine *N*-methyltransferase is a committed enzyme in catalyzing putrescine to form *N*-methylputrescine, and this enzyme is a secondary-metabolism enzyme (Hashimoto et al., 1989) Especially, tobacco PMT gene (*NtPMT*) were overexpressed in hairy root cultures of several MTA-producing plants, and such plants include *Duboisia* × hybrid (Moyano et al., 2002), *Hyoscyamus muticus* (Moyano et al., 2003), *Datura metal* (Moyano et al., 2003), *A. belladonna* (Rothe et al., 2003), *Hyoscyamus niger* (Zhang et al., 2004) and *Scopolia parviflora* (Lee et al., 2005). Overexpression of *NtPMT* in root cultures of *D.* hybrid, *A. belladonna*, and *H. niger* did not alter the production of hyoscyamine and scopolamine, while overexpression the same *PMT* gene (*NtPMT*) in hairy root cultures of *S. parviflora*, the levels of both hyoscyamine and scopolamine were significantly elevated. Moyano et al analyzed hyoscyamine and scopolamine content in individual clones of *H. muticus* and *D. metal* hairy root cultures with *NtPMT* overexpression, and they concluded that *NtPMT* overexpression increased hyoscyamine and scopolamine contents in *D. metal*, but only increased hyoscyamine content in *H. muticus*. Subsequently, Rothe et al reported their work on overexpression of *NtPMT* in hairy root cultures of *A. belladonna* and found no significantly increased production of MTAs when *NtPMT* was overexpressed. In the publication of Rothe et al. (2003), they discussed that the individual clones indicated both increased and decreased hyoscyamine and scopolamine content when *NtPMT* was overexpressed in root cultures of *H. muticus* and *Datura metel*. Zhang et al also found that overexpression of *NtPMT* did not promote hyoscyamine and scopolamine content in hairy root cultures of *H. niger* (Zhang et al., 2004). It should be noted that only two *NtPMT*-overexpressing root lines of *A. belladonna* and *S. parviflora* were used for studying MTA's production by Rothe et al. (2003) and Lee et al. (2005). Probably, such a few root lines might not be enough to solidify their conclusion. Such different results above suggest the role of PMT in engineering MTA's production should be carefully studied through intensive experiments.

Tropinone reductase I, also named tropinone-forming reductase, catalyzes tropinone reduction to give tropinone entering formation of littorine, a key intermediate for hyoscyamine biosynthesis (Hashimoto et al., 1992; Dräger, 2006). Richter et al. (2005) reported that overexpression of *Datura stramonium TRI* gene in hairy root cultures of *A. belladonna* markedly increased the TRI activity, tropine production, and contents of hyoscyamine and scopolamine. Kai et al found that overexpression *Anisodus acutangulus TRI* significantly increase the contents of hyoscyamine and scopolamine in hairy root cultures of *Ambrysus acutangulus*. Zhao et al also found that overexpression of *Scopolia lurida* (also named *Anisodus luridus*) *TRI* gene significantly promoted the production of hyoscyamine and scopolamine in hairy root cultures of *S. luridus* (Zhao et al., 2017). These previous reports indicate that TRI is a useful target on engineering MTA's production in hairy root cultures, but there are not reports on *TRI* overexpression in transgenic plants.

Littorine mutase (CYP80F1) catalyzes littorine to form hyoscyamine aldehyde (Li et al., 2006), which is able to be reduced to give hyoscyamine under catalysis by hyoscyamine dehydrogenase (Qiu et al., 2021). The level of hyoscyamine aldehyde is extremely low in plants, suggesting that CYP80F1 might be a rate-limiting enzyme or hyoscyamine aldehyde might be metabolized very quickly to form some other related alkaloids. However, overexpression of *CYP80F1* showed no significant alteration in hyoscyamine and scopolamine production in hairy root cultures of *H. niger* (Li et al., 2006). Our group identified hyoscyamine dehydrogenase (Qiu et al., 2021), also named hyoscyamine aldehyde reductase (Liao et al., 2019), and its role in engineering MTA's production was studied. When *A. belladonna HDH* gene was overexpressed in *A. belladonna* hairy root cultures, the production of hyoscyamine and scopolamine were markedly elevated, indicating that *HDH* is a valuable MTA biosynthesis gene in engineering MTA's production (Qiu et al., 2021).

Hyoscyamine 6*β*-hydroxylase (H6H) catalyzes hyoscyamine to produce anisodamine through 6*β*-hydroxylation, and subsequently oxidizes anisodamine to generate scopolamine (Hashimoto and Yamada, 1986) The *H6H* genes have been identified from *H. niger* (Hashimoto and Yamada, 1986), *A. belladonna* (Li et al., 2012), *A. acutangulus* (Kai et al.,

2011a), and *S. luridus* (Lan et al., 2018). Of these H6H enzymes, the *H. niger* one (HnH6H) showed the highest efficiency in converting hyoscyamine to scopolamine, and HnH6H is the best choice in engineering scopolamine production. Lan et al compared the effects of HnH6H and *S. luridus H6H* gene (*SlH6H*) on engineering scopolamine production, and found that HnH6H was much stronger than SlH6H in promoting scopolamine production (Lan et al., 2018). Without exception, overexpression of *H6H* definitively promoted the scopolamine content in hairy root cultures of different plants producing MTAs. However, the scopolamine production is quite a lot variable in engineered hairy root cultures generated from different plants, even when the same *H6H* gene (*HnH6H*) was overexpressed, indicating that the original capacity on MTA's production depends on plant species themselves. Impressively, When *HnH6H* was overexpressed hairy root cultures of *H. muticus*, the best clone produce scopolamine at a level over 100 times more than the control clones (Jouhikainen et al., 1999). Of these *HnH6H*-overexpressing hairy root cultures, the highest scopolamine production was reported in *H. niger*, reaching 184 mg per liter (Zhang et al., 2004). However, it should be noted that hyoscyamine is not able to be completely converted to scopolamine in root cultures, even under *HnH6H* overexpression. To be concluded, overexpression of *H6H*, especially *HnH6H*, is absolutely necessary for engineering scopolamine production.

Engineering MTA's Production by Overexpressing Combinations of MTA Biosynthesis Genes

From ornithine and phenylalanine to hyoscyamine or scopolamine, more than 10 biosynthesis genes involved. Such a long biosynthetic pathway implies that there are several rate-limiting steps governing by some enzymes. Furthermore, individual MTA's biosynthesis genes may have synergic effects on regulating alkaloid metabolism. Therefore, overexpression of the combinations of MTA's biosynthesis genes is very important in engineering MTA's production. Zhang et al. (2004) reported for the first time that overexpression of both *NtPMT* and *HnH6H* led to much higher levels of scopolamine in hairy root cultures than those overexpressing *NtPMT* or *HnH6H*. Kai et al. (2011b) also reported that the highest production of MTAs was obtained in *A. acutangulus* hairy root cultures overexpressing *AaPMT* and *AaTRI* together, compared with the scopolamine levels in those overexpressing *AaPMT* or *AaTRI*. Overexpression of both *AaTRI* and *AaH6H* also gave higher production of MTAs in hairy root cultures of *A. acutangulus*, than overexpression of *AaTRI* or *AaH6H* (Kai et al., 2012). Such overexpression of two MTA's biosynthesis genes is able to be named a "Pushing-Pulling" strategy of metabolic engineering. When an upstream MTA's biosynthesis gene, such as *PMT* before *TRI* and *H6H*, is overexpressed, its product as precursors for following enzymes is given more, and this pushes more metabolic flux to the downstream reactions. Simultaneously, when a downstream MTA's biosynthesis gene is overexpressed, more precursors are able to be converted, showing a pulling force in biosynthetic pathway. To be concluded, overexpression of combination of MTA's biosynthesis genes usually have better effects on promoting MTA's production than that of a single gene.

Up to date, there are no reports on engineering MTA's production through overexpression of three or more biosynthesis genes. Very recently, our group finished evaluation on *PYKS, CYP82M3, UGT1, LS*, two gene combinations (*PYKS* + *CYP82M3, UGT1* + *LS*), and four gene combination (*PYKS* + *CYP82M3* + *UGT1* + *LS*), in engineering MTA's production, and found interesting and valuable discoveries (submitted). According to the previously reported researches and our submitted ones, it can be draw a conclusion on simultaneous overexpression of several biosynthesis genes is a prior strategy in engineering MTA's production. Although fruitful and marked achievements on engineering MTA's production in hairy root cultures (**Table 1**), industrial production has not been successfully realized using hairy root cultures, because their cultivation in a large or industrial scale is not available technologically at present. So, it is necessary to develop engineered plants with high-yield MTA's production.

ENGINEERING MEDICINAL TROPANE ALKALOID PRODUCTION IN PLANTS

To date, industrial production of medicinal tropane alkaloids is completely dependent on plant cultivation. So, engineered plants with high yield of hyoscyamine and scopolamine are highly desirable and valuable. For development of such engineered plants, genetic transformation with high efficiency is the most important technology. In fact, generated plants are reported in only a very few plant species producing MTAs, including *A. belladonna* and *Datura innoxia*.

Most of engineered plants with high-yield MTAs are reported in *A. belladonna* with overexpressing *HnH6H* for enhancement of scopolamine production. In 1992, transgenic plants of *A. belladonna* with *HnH6H* overexpression were developed by Yun et al. (1992). They established T0 progeny of *HnH6H*-overexpressing plants of *A. belladonna*, and further generated T1 progeny for analyzing the MTA's production in different organs. The *HnH6H*-overexpressing plants of *A. belladonna* were grown under controlled conditions in a growth chamber. Surprisingly, hyoscyamine was almost completely converted to scopolamine in leaves and stems of T1-progeny plants with *HnH6H* overexpression, whereas only part of hyoscyamine was converted to scopolamine in roots. Such difference indicates that higher efficiency converting hyoscyamine to scopolamine in upground organs than that in underground organs. The scopolamine production was greatly elevated in all tested T1-progeny plants, and the highest scopolamine production was over 1.2% dry weight in leaves of the T1-12 line (Yun et al., 1992). Considering that synergic effects between PMT and H6H, our group generated *A. belladonna* plants, in which both *NtPMT* and *HnH6H* were overexpressed, and the scopolamine production was markedly increased as expected (Wang et al., 2011). Then, we further produced T1 progeny of *A. belladonna* plants with overexpression of both *NtPMT*

TABLE 1 | Engineering medicinal tropane alkaloid production in plant hairy root cultures through gene overexpression.

TA biosynthesis genes	Host plants	Metabolites	Content change	References
AbODC	*Atropa belladonna*	Putrescine	↑	Zhao et al., 2020
mouse ODC	*Datura innoxia*	Scopolamine	↑	Singh et al., 2011
NtPMT	*Datura metel*	Hyoscyamine	↑	Moyano et al., 2003
		Scopolamine	↑	
	Hyoscyamus muticus	Hyoscyamine	↑	
		Scopolamine	–	
	Scopolia parviflora	Hyoscyamine	↑	Lee et al., 2005
		Scopolamine	↑	
	A. belladonna	*N*-methylputrescine	↑	Rothe et al., 2003
		Hyoscyamine	–	
		Scopolamine	–	
	Hyoscyamus niger	Hyoscyamine	–	Zhang et al., 2004
		Scopolamine	–	
SlTRI	*Scopolia lurida* or *Anisodus luridus*	Hyoscyamine	↑	Zhao et al., 2017
		Scopolamine	↑	
DsTRI	*A. belladonna*	Hyoscyamine	↑	Richter et al., 2005
		Scopolamine	↑	
HnCYP80F1	*H. niger*	Hyoscyamine	–	Li et al., 2006
		Scopolamine	–	
AbHDH	*A. belladonna*	Hyoscyamine	↑	Qiu et al., 2021
		Anisodamine	↑	
		Scopolamine	↑	
DiH6H	*D. innoxia*	Hyoscyamine	↓	Li et al., 2020
		Scopolamine	↑	
SlH6H	*Scopolia lurida* or *Anisodus luridus*	Anisodamine	↑	Lan et al., 2018
		Scopolamine	↑	
HnH6H		Anisodamine	↑	
		Scopolamine	↑	
HnH6H	*H. muticus*	Scopolamine	↑	Jouhikainen et al., 1999
	Hyoscyamus niger	Scopolamine	↑	Zhang et al., 2004
	Duboisia × *hybrid*	Scopolamine	↑	Li et al., 2020
		Anisodamine	↓	
	Scopolia parviflora	Hyoscyamine	↑	Kang et al., 2005
		Scopolamine	↑	
AaPMT + *AaTRI*	*Anisodus acutangulus*	Hyoscyamine	↑	Kai et al., 2012
		Anisodamine	↑	
		Scopolamine	↑	
AaTRI + *AaH6H*	*Anisodus acutangulus*	Hyoscyamine	↑	Kai et al., 2012
		Anisodamine	↑	
		Scopolamine	↑	
SpPMT + *SpH6H*	*Scopolia parviflora*	Hyoscyamine	↑	Kang et al., 2011
		Scopolamine	↑	
NtPMT + *HnH6H*	*Atropa belladonna*	Hyoscyamine	↑	Yang et al., 2011
		Scopolamine	↑	
	Hyoscyamus niger	Scopolamine	↑	Zhang et al., 2004
	Anisodus acutangulus	Hyoscyamine	↑	Yan et al., 2014
		Scopolamine	↑	

↑, *increased production;* ↓, *decreased production;* –, *no significant changes.*

and *HnH6H*, and investigated MTA's production under the field conditions in the medicinal plant garden of Tibet Agriculture and Animal Husbandry University (Nyingchi, Tibet, China) and Xinpu Town of Guiyang City (Guizhou Province, China). Similar to the results obtained by Yun et al in *HnH6H*-overexpssing plants, hyoscyamine was completely converted to scopolamine in aerial organs and high production of scopolamine was detected. In the plants harvested from the field in Xinpu Town of Guiyang City, the highest production of scopolamine was 0.513% DW in leaves, indicating over 12 times higher than in wild-type levels (Xia et al., 2016). Similar results were obtained when *A. belladonna* plants with overexpressing both *NtPMT* and *HnH6H* (T1 progeny) were grown in the field of Tibet Agriculture and Animal Husbandry University, but the highest scopolamine level in leaves nearly reached 1% DW (Quan et al., 2016).

Compared with *HnH6H*-overexpressing plants generated by Yun et al, the scopolamine production in our transgenic plants was not so high, and possible reasons leading to such difference might have different genotypes of *A. belladonna* used as initial plant materials for transgenic research, different growth conditions, and some others. Notably, an obvious difference between Yun and our studies is the scopolamine production in roots. The scopolamine levels in secondary roots of

HnH6H-overexpressing plants were even much lower than those in wild-type secondary roots, according to Yun et al. (1992). Without exception, overexpression of both *NtPMT* and *HnH6H* in greatly increased the scopolamine content of secondary roots in contrast to wild-type plants (Xia et al., 2016). It was postulated that synergic effects overexpressing both *NtPMT* and *HnH6H* facilitated the conversion from hyoscyamine to scopolamine in the native organ (secondary roots) synthesizing scopolamine (Zhang et al., 2004; Xia et al., 2016).

The biosynthesis genes including *TRI* and *ODC* were also used to engineering MTA's production in transgenic plants. Zhao et al established transgenic plants of *A. belladonna*, in which the *ODC* gene of *A. belladonna* (*AbODC*) was overexpressed, and analyzed putrescine, *N*-methylputrescine, hyoscyamine, anisodamine and scopolamine in roots and leaves (Zhao et al., 2020). The production of putrescine, *N*-methylputrescine, hyoscyamine and anisodamine were significantly increased in roots of *AbODC*-overexpressing plants. The levels of putrescine, hyoscyamine, and anisodamine were also elevated in leaves of *AbODC*-overexpressing plants. The scopolamine production was not significantly altered in roots and leaves, even when *AbODC* was overexpressed. Overexpression of *AbODC* increased the production of hyoscyamine anisodamine, but did not promote scopolamine production, indicating that the epoxidation reaction catalyzed by *A. belladonna* H6H is a key limitation in converting anisodamine to scopolamine (Hashimoto and Yamada, 1986; Kluza et al., 2020). A mouse ODC gene was ectopically expressed in transgenic plants of *D. innoxia*, the polyamine and scopolamine levels were promoted, but hyoscyamine was not analyzed in this research (Singh et al., 2011). Although the scopolamine production was increased in transgenic plants of *D. innoxia*, the highest level for scopolamine was extremely low, 0.25 $\mu g\,g^{-1}$ DW (Singh et al., 2011), suggesting that *D. innoxia* might be not a good candidate for industrially producing scopolamine. Rothe et al developed *NtPMT*-overexpressing plants of *A. belladonna*, and found that *NtPMT* overexpression did not increased the production of scopolamine, hyoscyamine, tropine, pseudotropine, tropinone, and calystegines. It should also be noted that there were only two transgenic plants used for analyzing alkaloid production (Rothe et al., 2003), and this might not be enough to draw a solidified conclusion, because the levels of these metabolites are variable in individual plants.

Notably, all the above transgenic plants with engineering MTA's biosynthetic pathway are generated using kanamycin as selection pressure, given by neomycin phosphotransferase II (NPTII) as marker gene. As it is well known, the *NPTII* gene is useless when transgenic plants are obtained and it brings a biosafety risk for commercial usage of engineering plants. When transgenic plants with high-yield MTAs are developed, it is very important to establish transgenic homozygous lines that are highly valuable for agricultural and industrial application. However, no transgenic homozygous plants producing MTAs were reported before 2021. Therefore, our group took a few years to successfully develop transgenic homozygous plants of *A. belladonna*, with overexpression of both a glyphosate-resistant gene and *HnH6H*, and without antibiotics-resistant gene (Zhang et al., 2021). After obtaining T0 progeny with one-copy insertion, genetic breeding was used to develop homozygous offsprings. The field trials indicated that homozygous transgenic plants of *A. belladonna* grew well, resisted the commercial recommended concentration of glyphosate, produced high-yield scopolamine (7.04 $mg\,g^{-1}$ DW in leaf). These homozygous transgenic plants of *A. belladonna* showed much higher commercial value than wild-type ones, because of their glyphosate resistance and high-yield scopolamine.

The CRISPR/Cas9 system is a powerful gene-editing tool, which is also used to generate *A. belladonna* plants with high-yield hyoscyamine (Zeng et al., 2021). Hyoscyamine is converted to anisodamine, and anisodamine to scopolamine under catalysis of H6H (Hashimoto and Yamada, 1986). Extracts of *A. belladonna* contain the three above alkaloids with similar structures, and it is expensive to separate each from total alkaloid extracts. Our group generated *A. belladonna* plants with disruption of the *H6H* gene through CRISPR/Cas9 system, the *H6H*-disrupting plants produced hyoscyamine at much higher levels than wild-type ones, because hyoscyamine cannot be converted to anisodamine and scopolamine, and they produced neither anisodamine nor scopolamine. The *H6H*-disrupting *A. belladonna* plants are valuable due to high-yield hyoscyamine and convenient separation of hyoscyamine.

Metabolic engineering of MTA's production usually employs their biosynthesis genes. Very recently, our group has a publication on reporting engineering tropane alkaloid production using a novel calmodulin gene in *A. belladonna* plants (Zhang et al., 2022). Increased calcium can promote the TA's production in plants, and accordingly, we postulated that calcium/calmodulin signaling might regulate TA's biosynthesis and then identified a calcium-binding calmodulin gene (*AbCaM1*) highly expressed in secondary roots of *A. belladonna*. Transgenic plants of *A. belladonna*, in which both *AbCaM1* and *G2-EPSPS* were overexpressed, had strong tolerance to glyphosate and high-yield tropane alkaloids, including hyoscyamine, anisodamine and scopolamine. To our knowledge, this publication is the first report on engineering TA's production in plants using a regulator protein. Such transgenic homozygous plants of *A. belladonna* are valuable for industrial production of medicinal TAs, due to their high production of TAs and strong tolerance to herbicide that facilitate manufacturing TAs at low cost.

THE PROBLEMS ON ENGINEERING MTA's PRODUCTION IN PLANTS

Although the MTA's biosynthetic pathway has been completely unveiled, using *A. belladonna* as a model plant producing MTAs. The same enzyme from different MTA-producing plants may have various catalytic efficiency. For example, the H6H enzyme of *H. niger* shows much higher catalytic efficiency in converting hyoscyamine to scopolamine than the reported others. So, it is necessary to screen MTA's biosynthesis enzymes with high catalytic efficiency from different plants producing MTAs. According to previous reports on enzymatic assays, it is suggested

to pay more attention to the MTA's biosynthetic enzymes, such as CYP80F1, phenyllactate UDP-glycosyltransferase and littorine synthase, each of which is a key limitation in the MTA's biosynthesis (Bedewitz et al., 2018; Qiu et al., 2020). Furthermore, over 100 tropane alkaloids are reported in MTA-producing plants, and they compete against the biosynthesis of hyoscyamine and scopolamine. It is largely unknown on these alkaloid biosynthesis, so it is also important to understand their formation, and then genetic operation might be performed in planta to disrupt their biosynthesis and enhance the production of hyoscyamine and scopolamine.

To date, the values of most MTA's biosynthesis genes are unknown in engineering MTA's production in plants. It is of importance to have systematic evaluation on the values of MTA's biosynthesis genes in plant metabolic engineering. And further, their combinations should also be tested in engineering MTA's production. It can be reasonably predicted that combinations of MTA's biosynthesis genes will be the main methods used in engineering MTA's production in the future. Tens of solanaceous plant species are able to produce MTAs, and such abundant resources provide diverse plant chassis for engineering MTA's production through metabolic engineering. However, of these MTA-producing plants, only the plants of *A. belladonna* can be efficiently transformed. Lack of genetic transformation at high efficiency is a key technological limitation for modification of MTA's biosynthetic pathway in plants. So, plant biotechnology, especially transgenic technology, should be strengthened.

THE PROSPECTS IN ENGINEERING MTA's PRODUCTION IN PLANTS

Complete elucidation of MTA's biosynthetic pathway results in synthesis of hyoscyamine and scopolamine in engineered yeast cells using the approaches of synthetic biology. Professor Smolke built the MTA's biosynthetic pathway in yeast cells, and successfully synthesized hyoscyamine and scopolamine in engineered yeast cells (Srinivasan and Smolke, 2020). It is a milestone in manufacturing these valuable alkaloids. However, the production of hyoscyamine and scopolamine is extremely low, of which production levels are very far away from the standards of industrial production. Without doubt, the production of MTAs in engineered yeast cells will be gradually promoted, and probably it might reach or even exceed 5 g per liter, the lowest level for industrially producing TAs based on yeast's fermentation. By the way, synthetic biology's artemisinin fails in market is a typical case, and the most important reason for this failure is plant-base production of artemisinin is much cheaper than synthetic biology's.

Surely, the plant biotechnologists have been trying their best to develop novel plant varieties producing MTAs at higher and higher levels, especially through complicated genetic operations composed of overexpression and suppression/gene edition methods. The next-generation strategy on engineering MTA's production consists of enhancement of MTA's production by overexpressing MTA's biosynthesis genes, and simultaneously disruption of competitive metabolite's production by suppressing or knocking out corresponding biosynthesis genes, in MTA-producing plants. Besides the core target on promoting MTA's production, horticultural traits, such as biomass especially leaf yield, resistance on stress and pathogens and herbicide resistance, are able to be genetically fortified. Engineered plants, with increased MTA's production and strengthened horticultural traits, will be the main source for industrial production of hyoscyamine and scopolamine, and engineered yeast cell factory may be an additional source for MTA's production.

AUTHOR CONTRIBUTIONS

HG, LZ, and ZL conceived the original idea for the review, and edited and reviewed the manuscript. HG and PH drew the figure. XL and LZ checked the chemical structures and took the pictures of plant species. All authors wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the NSFC projects (U1902212 and 31770335), the Fourth National Survey of Traditional Chinese Medicine Resources, Chinese or Tibet Medicinal Resources Investigation in Tibet Autonomous Region (State Administration of Chinese Traditional Medicine 20191217–540124 and 20200501–542329) and the College Students' Innovative Entrepreneurial Training Plan Program (202110635108).

REFERENCES

Bacchi, C. J., Nathan, H. C., Hutner, S. H., McCann, P. P., and Sjoerdsma, A. (1980). Polyamine metabolism: a potential therapeutic target in trypanosomes. *Science* 210, 332–334. doi: 10.1126/science.6775372

Bedewitz, M. A., Jones, A. D., D'Auria, J. C., and Barry, C. S. (2018). Tropinone synthesis via an atypical polyketide synthase and P450-mediated cyclization. *Nat. Commun.* 9, 5281–5213. doi: 10.1038/s41467-018-07671-3

Dräger, B. (2006). Tropinone reductases, enzymes at the branch point of tropane alkaloid metabolism. *Phytochemistry* 67, 327–337. doi: 10.1016/j.phytochem.2005.12.001

Griffin, J., Lee, S., Caiado, M., Kesten, S., and Price, D. (2008). Comparison of tiotropium bromide and combined ipratropium/salbutamol for the treatment of COPD: a UK general practice research database 12-month follow-up study. *Prim. Care Respir. J.* 17, 104–110. doi: 10.3132/pcrj.2008.00024

Grynkiewicz, G., and Gadzikowska, M. (2008). Tropane alkaloids as medicinally useful natural products and their synthetic derivatives as new drugs. *Pharmacol. Reports* 60, 439–463.

Hashimoto, T., Nakajima, K., Ongena, G., and Yamada, Y. (1992). Two tropinone reductases with distinct stereospecificities from cultured roots of *Hyoscyamus niger*. *Plant Physiol.* 100, 836–845. doi: 10.1104/pp.100.2.836

Hashimoto, T., and Yamada, Y. (1986). Hyoscyamine 6β-hydroxylase, a 2-Oxoglutarate-dependent Dioxygenase, Alkaloid-Producing Root Cultures. *Plant Physiol.* 81, 619–625. doi: 10.1104/pp.81.2.619

Hashimoto, T., Yukimune, Y., and Yamada, Y. (1989). Putrescine and putrescine N-methyltransferase in the biosynthesis of tropane alkaloids

in cultured roots of *Hyoscyamus albus*. *Planta* 178, 131–137. doi: 10.1007/BF00392536

Huang, J. P., Wang, Y. J., Tian, T., Wang, L., Yan, Y., and Huang, S. X. (2021). Tropane alkaloid biosynthesis: a centennial review. *Nat. Prod. Rep.* 38, 1634–1658. doi: 10.1039/d0np00076k

Jouhikainen, K., Lindgren, L., Jokelainen, T., Hiltunen, R., Teeri, T. H., and Oksman-Caldentey, K. M. (1999). Enhancement of scopolamine production in *Hyoscyamus muticus* L. hairy root cultures by genetic engineering. *Planta* 208, 545–551. doi: 10.1007/s004250050592

Kai, G., Liu, Y., Wang, X., Yang, S., Fu, X., Luo, X., et al. (2011a). Functional identification of hyoscyamine 6β-hydroxylase from *Anisodus acutangulus* and overproduction of scopolamine in genetically-engineered *Escherichia coli*. *Biotechnol. Lett.* 33, 1361–1365. doi: 10.1007/s10529-011-0575-y

Kai, G., Yang, S., Luo, X., Zhou, W., Fu, X., Zhang, A., et al. (2011b). Co-expression of AaPMT and AaTRI effectively enhances the yields of tropane alkaloids in *Anisodus acutangulus* hairy roots. *BMC Biotechnol.* 11. doi: 10.1186/1472-6750-11-43

Kai, G., Zhang, A., Guo, Y., Li, L., Cui, L., Luo, X., et al. (2012). Enhancing the production of tropane alkaloids in transgenic *Anisodus acutangulus* hairy root cultures by over-expressing tropinone reductase i and hyoscyamine-6β-hydroxylase. *Mol. BioSyst.* 8, 2883–2890. doi: 10.1039/c2mb25208b

Kang, Y. M., Lee, O. S., Jung, H. Y., Kang, S. M., Lee, B. H., Karigar, C., et al. (2005). Overexpression of hyoscyamine 6β-hydroxylase(h6h) gene and enhanced production of tropane alkaloids in *Scopolia parviflora* hairy root lines. *J. Microbiol. Biotechnol.* 15, 91–98.

Kang, Y. M., Park, D. J., Min, J. Y., Song, H. J., Jeong, M. J., Kim, Y. D., et al. (2011). Enhanced production of tropane alkaloids in transgenic *Scopolia parviflora* hairy root cultures over-expressing putrescine N-methyl transferase (PMT) and hyoscyamine-6β-hydroxylase (H6H). *Vitr. Cell. Dev. Biol. Plant* 47, 516–524. doi: 10.1007/s11627-011-9367-2

Kluza, A., Wojdyla, Z., Mrugala, B., Kurpiewska, K., Porebski, P. J., Niedzialkowska, E., et al. (2020). Regioselectivity of hyoscyamine 6β-hydroxylase-catalysed hydroxylation as revealed by high-resolution structural information and QM/MM calculations. *Dalt. Trans.* 49, 4454–4469. doi: 10.1039/d0dt00302f

Lan, X., Zeng, J., Liu, K., Zhang, F., Bai, G., Chen, M., et al. (2018). Comparison of two hyoscyamine 6β-hydroxylases in engineering scopolamine biosynthesis in root cultures of *Scopolia lurida*. *Biochem. Biophys. Res. Commun.* 497, 25–31. doi: 10.1016/j.bbrc.2018.01.173

Lee, O. S., Kang, Y. M., Jung, H. Y., Min, J. Y., Kang, S. M., Karigar, C. S., et al. (2005). Enhanced production of tropane alkaloids in *Scopolia parviflora* by introducing the PMT (putrescine N-methyltransferase) gene. *Vitr. Cell. Dev. Biol. Plant* 41, 167–172. doi: 10.1079/IVP2004621

Li, R., Reed, D. W., Liu, E., Nowak, J., Pelcher, L. E., Page, J. E., et al. (2006). Functional genomic analysis of alkaloid biosynthesis in *Hyoscyamus niger* reveals a cytochrome P450 involved in Littorine rearrangement. *Chem. Biol.* 13, 513–520. doi: 10.1016/j.chembiol.2006.03.005

Li, J., Van Belkum, M. J., and Vederas, J. C. (2012). Functional characterization of recombinant hyoscyamine 6β-hydroxylase from *Atropa belladonna*. *Bioorganic Med. Chem.* 20, 4356–4363. doi: 10.1016/j.bmc.2012.05.042

Li, Q., Zhu, T., Zhang, R., Bu, Q., Yin, J., Zhang, L., et al. (2020). Molecular cloning and functional analysis of hyoscyamine 6β-hydroxylase (H6H) in the poisonous and medicinal plant *Datura innoxia* mill. *Plant Physiol. Biochem.* 153, 11–19. doi: 10.1016/j.plaphy.2020.04.021

Liao, Z., Chen, M., Yang, C., Qiu, F., and Zeng, J. (2019). 莨菪醛还原酶及其应用. The patent number "ZL 2019 1 0826205.8".

Matthew, A. B., Gonzales-Vigil, E., Góngora-Castillo, E., Uebler, J. B., Gonzales-Vigil, E., Wiegert-Rininger, K. E., et al. (2014). A root-expressed l-phenylalanine:4-hydroxyphenylpyruvate aminotransferase is required for tropane alkaloid biosynthesis in *Atropa belladonna*. *Plant Cell* 26, 3745–3762. doi: 10.1105/tpc.114.130534

Moyano, E., Fornalé, S., Palazón, J., Cusidó, R. M., Bagni, N., and Piñol, M. T. (2002). Alkaloid production in Duboisia hybrid hairy root cultures overexpressing the pmt gene. *Phytochemistry* 59, 697–702. doi: 10.1016/S0031-9422(02)00044-4

Moyano, E., Jouhikainen, K., Tammela, P., Palazón, J., Cusidó, R. M., Piñol, M. T., et al. (2003). Effect of pmt gene overexpression on tropane alkaloid production in transformed root cultures of *Datura metel* and *Hyoscyamus muticus*. *J. Exp. Bot.* 54, 203–211. doi: 10.1093/jxb/erg014

Nölke, G., Schneider, B., Agdour, S., Drossard, J., Fischer, R., and Schillberg, S. (2008). Modulation of polyamine biosynthesis in transformed tobacco plants by targeting ornithine decarboxylase to an atypical subcellular compartment. *Open Biotechnol. J.* 2, 183–189. doi: 10.2174/1874070700802010183

Qiu, F., Yan, Y., Zeng, J., Huang, J. P., Zeng, L., Zhong, W., et al. (2021). Biochemical and metabolic insights into Hyoscyamine dehydrogenase. *ACS Catal.* 11, 2912–2924. doi: 10.1021/acscatal.0c04667

Qiu, F., Zeng, J., Wang, J., Huang, J. P., Zhou, W., Yang, C., et al. (2020). Functional genomics analysis reveals two novel genes required for littorine biosynthesis. *New Phytol.* 225, 1906–1914. doi: 10.1111/nph.16317

Quan, H., Xia, K., Zeng, J., Chen, M., Lan, X., and Liao, Z. (2016). Overexpression of NtPMT and HnH6H changed hyoscyamine-rich *Atropa belladonna* to scopolamine-rich varieties. *Acta Pharm. Sin.* 51, 1913–1919. doi: 10.16438/j.0513-4870.2016-0275

Richter, U., Rothe, G., Fabian, A. K., Rahfeld, B., and Dräger, B. (2005). Overexpression of tropinone reductases alters alkaloid composition in *Atropa belladonna* root cultures. *J. Exp. Bot.* 56, 645–652. doi: 10.1093/jxb/eri067

Rothe, G., Hachiya, A., Yamada, Y., Hashimoto, T., and Dräger, B. (2003). Alkaloids in plants and root cultures of *Atropa belladonna* overexpressing putrescine N-methyltransferase. *J. Exp. Bot.* 54, 2065–2070. doi: 10.1093/jxb/erg227

Singh, A., Nirala, N. K., Das, S., Narula, A., Rajam, M. V., and Srivastava, P. S. (2011). Overexpression of odc (ornithine decarboxylase) in *Datura innoxia* enhances the yield of scopolamine. *Acta Physiol. Plant.* 33, 2453–2459. doi: 10.1007/s11738-011-0787-8

Srinivasan, P., and Smolke, C. D. (2020). Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* 585, 614–619. doi: 10.1038/s41586-020-2650-9

Wang, X., Chen, M., Yang, C., Liu, X., Zhang, L., Lan, X., et al. (2011). Enhancing the scopolamine production in transgenic plants of *Atropa belladonna* by overexpressing pmt and h6h genes. *Physiol. Plant.* 143, 309–315. doi: 10.1111/j.1399-3054.2011.01506.x

Xia, K., Liu, X., Zhang, Q., Qiang, W., Guo, J., Lan, X., et al. (2016). Promoting scopolamine biosynthesis in transgenic *Atropa belladonna* plants with pmt and h6h overexpression under field conditions. *Plant Physiol. Biochem.* 106, 46–53. doi: 10.1016/j.plaphy.2016.04.034

Yan, Z., Pan, X., Qiang, W., and Liao, Z. (2014). Improvement of biosynthesis of Tropane alkaloids in *Anisodus acutangulus* by co-transformed PMT and H6H. *J. Southwest Univ.* 36:7

Yang, C., Chen, M., Zeng, L., Zhang, L., Liu, X., Lan, X., et al. (2011). Improvement of tropane alkaloids production in hairy root cultures of *Atropa belladonna* by overexpressing pmt and h6h genes. *Plant Omics* 4, 29–33.

Yun, D. J., Hashimoto, T., and Yamada, Y. (1992). Metabolic engineering of medicinal plants: transgenic *Atropa belladonna* with an improved alkaloid composition. *Proc. Natl. Acad. Sci. U. S. A.* 89, 11799–11803. doi: 10.1073/pnas.89.24.11799

Zeng, L., Zhang, Q., Jiang, C., Zheng, Y., Zuo, Y., Qin, J., et al. (2021). Development of *Atropa belladonna* L. plants with high-yield hyoscyamine and without its derivatives using the CRISPR/Cas9 system. *Int. J. Mol. Sci.* 22, 1–10. doi: 10.3390/ijms22041731

Zhang, L., Ding, R., Chai, Y., Bonfill, M., Moyano, E., Oksman-Caldentey, K. M., et al. (2004). Engineering tropane biosynthetic pathway in *Hyoscyamus niger* hairy root cultures. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6786–6791. doi: 10.1073/pnas.0401391101

Zhang, Q., Liang, M., Liu, Y., Yang, C., Zeng, J., Qin, J., et al. (2021). Development of homozygous transgenic *Atropa belladonna* plants with glyphosate resistance and high-yield scopolamine using metabolic engineering. *Ind. Crop. Prod.* 171:113953. doi: 10.1016/j.indcrop.2021.113953

Zhang, Q., Liang, M., Zeng, J., Yang, C., Qin, J., Qiang, W., et al. (2022). Engineering tropane alkaloid production and glyphosate resistance by overexpressing AbCaM1 and G2-EPSPS in *Atropa belladonna*. *Metab. Eng.* 72, 237–246. doi: 10.1016/j.ymben.2022.03.014

Zhao, T., Li, S., Wang, J., Zhou, Q., Yang, C., Bai, F., et al. (2020). Engineering tropane alkaloid production based on metabolic characterization of ornithine decarboxylase in *Atropa belladonna*. *ACS Synth. Biol.* 9, 437–448. doi: 10.1021/acssynbio.9b00461

Zhao, K., Zeng, J., Zhao, T., Zhang, H., Qiu, F., Yang, C., et al. (2017). Enhancing tropane alkaloid production based on the functional identification of tropine-

forming reductase in *Scopolia lurida*, a Tibetan medicinal plant. *Front. Plant Sci.* 8, 1–11. doi: 10.3389/fpls.2017.01745

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fpls.2022.919025

Identification of Abietane-Type Diterpenoids and Phenolic Acids Biosynthesis Genes in *Salvia apiana* Jepson Through Full-Length Transcriptomic and Metabolomic Profiling

Jiadong Hu[1†], *Feiyan Wang*[1†], *Fengying Liang*[2†], *Ziding Wu*[1], *Rui Jiang*[1], *Jinxing Li*[1], *Junfeng Chen*[1], *Shi Qiu*[1], *Jing Wang*[1], *Yuchen Zhang*[1], *Qing Li*[2*] and *Wansheng Chen*[1,2*]

[1] Center of Chinese Traditional Medicine Resources and Biotechnology, Institute of Chinese Materia Medica, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [2] Department of Pharmacy, Second Affiliated Hospital of Naval Medical University, Shanghai, China

Salvia apiana (*S. apiana*) Jepson is a medicinal plant that is frequently used by the Chumash Indians in southern California as a diaphoretic, calmative, diuretic, or antimicrobial agent. Abietane-type diterpenoids (ATDs) and phenolic acids (PAs) are the main bioactive ingredients in *S. apiana*. However, few studies have looked into the biosynthesis of ATDs and PAs in *S. apiana*. In this study, using metabolic profiling focused on the ATDs and PAs in the roots and leaves of *S. apiana*, we found a distinctive metabolic feature with all-around accumulation of ATDs, but absence of salvianolic acid B. To identify the candidate genes involved in these biosynthesis pathways, full-length transcriptome was performed by PacBio single-molecule real-time (SMRT) sequencing. A total of 50 and 40 unigenes were predicted to be involved in ATDs and PAs biosynthesis, respectively. Further transcriptional profile using Illumina HiSeq sequencing showed that the transcriptional variations of these pathways were consistent with the accumulation patterns of corresponding metabolites. A plant kingdom-wide phylogenetic analysis of cytochromes (CYPs) identified two CYP76AK and two CYP76AH subfamily genes that might contribute for the specific ATDs biosynthesis in *S. apiana*. We also noticed that the clade VII laccase gene family was significantly expanded in *Salvia miltiorrhiza* compared with that of *S. apiana*, indicating their involvements in the formation of salvianolic acid B. In conclusion, our results will enable the further understanding of ATDs and PAs biosynthesis in *S. apiana* and *Salvia* genus.

Keywords: *Salvia apiana* Jepson, biosynthesis, transcriptomics, metabolomics, phenolic acids, abietane-type diterpenoids

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Xiaolong Hao,
Zhejiang Chinese Medical University, China
Yu Fulai,
Chinese Academy of Tropical Agricultural Sciences, China

***Correspondence:**
Wansheng Chen
chenwansheng@shutcm.edu.cn
Qing Li
qli@smmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *13 April 2022*
Accepted: *28 April 2022*
Published: *08 June 2022*

Citation:
Hu J, Wang F, Liang F, Wu Z, Jiang R, Li J, Chen J, Qiu S, Wang J, Zhang Y, Li Q and Chen W (2022) Identification of Abietane-Type Diterpenoids and Phenolic Acids Biosynthesis Genes in Salvia apiana Jepson Through Full-Length Transcriptomic and Metabolomic Profiling. Front. Plant Sci. 13:919025. doi: 10.3389/fpls.2022.919025

INTRODUCTION

Salvia apiana (*S. apiana*) Jepson, commonly known as white sage or bee sage, belongs to the genus *Salvia* L. (Lamiaceae: Nepetoideae: Mentheae: Salviinae). *S. apiana* is specific for California's and Baja California's flora (Will and Claßen-Bockhoff, 2017). It is widely distributed in the California Floristic Province and forms the chaparral and desert sage community together with other eighteen members of Salvia sections such as Echinosphace and Audibertia (Walker et al., 2015). Therefore, its taxonomical position is quite essential for understanding the phylogenetic relationship and origin of genus *Salvia* in the New World. However, the available nuclear genes in *S. apiana* are very limited, which hamper its phylogenetic studies based on molecular data.

The application history of *S. apiana* is unique due to its distribution area. It is widely used by the Chumash Indians as diaphoretic, calmative, diuretic, and antimicrobial agent, as well as a burning sage used in religious practices (called khapshikh or xapcix) (Walker et al., 2015; Krol et al., 2021). In modern times, *S. apiana* is demonstrated to have pharmacological activities, including antimicrobial, anti-inflammatory, gamma-aminobutyric acid (GABA)ergic, analgesic, antioxidant, cytotoxic, and antitumor activities, owing to its special chemical constitution (Khan et al., 2016; Saeed et al., 2016; Srivedavyasasri et al., 2017; Afonso et al., 2019; Krol et al., 2021). Previous phytochemical studies on *S. apiana* have identified substantial amounts of essential oil, accompanied by a variety of triterpenes, C23 terpenoids, diterpenes, flavonoids, and phenolic acids (PAs) (Krol et al., 2021). A distinctive metabolic feature of *S. apiana* is that it can produce abietane-type diterpenoids (ATDs) in the aerial part of the plant, while the other American *Salvia* species can only produce clerodane derivatives (Dentali, 1991; Bisio et al., 2019; Krol et al., 2021). Carnosic acid and its derivatives, such as 16-hydroxycarnosic acid, rosmanol, and 16-hydroxycarnosol, etc., constituted the majority of this chemical group (Luis et al., 1996). Tanshinones, such as tanshinones IIA, tanshinones B, and cryptotanshinone, represent another group of ATDs, are usually regarded as the feature compounds in the East Asia taxa of *Salvia* genus (Guo et al., 2016; Ma et al., 2021). Interestingly, cryptotanshinone was also found in the root of *S. apiana* (González et al., 1992), indicating the presence of other tanshinones chemicals and a much more complex metabolic composition of diterpenes in *S. apiana* as well.

Owing to the restricted distribution range and uncontrolled harvesting, *S. apiana* is now at high risk of rapid decline or extinction (Adlof, 2015). Therefore, studies concerning botany, cultivation, cell culture, and metabolic engineering of *S. apiana* are necessary for efficient propagation of natural resources and production of bioactive metabolites, which largely rely on the understanding of metabolic pathways of these bioactive compounds and the genetic resource at both the transcriptomic and genomic levels (D'Amelia et al., 2017). To date, studies of *S. apiana* have mainly focused on its chemical compositions and biological activities of metabolites. However, little about the genetics of the plant was concerned, which has resulted in the lack of knowledge about the biosynthesis of several pharmacologically active chemical groups in *S. apiana*.

Here, we used single-molecule real-time (SMRT) sequencing on the PacBio Sequel platform to obtain full-length transcriptome information of *S. apiana*. A plant metabolomics and chemometrics analysis based on ultra-high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (UHPLC-Q-TOF-MS/MS) was further applied to characterize the metabolic profiling and screen the chemical markers corresponding to underground and aerial parts of *S. apiana*. Based on the metabolic profiling of ATDs and PAs, the transcriptional patterns of genes in diterpenoids and phenolic acid biosynthesis pathways were elucidated in *S. apiana* using Illumina sequencing.

RESULTS

Metabolic Profiling of *S. apiana*

To have a thoroughly chemical constitutes of ATDs and PAs, ultra-high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (UPLC-Q-TOF-MS) was applied to samples from the aerial (leaves) and underground (roots) parts of *S. apiana* separately (**Figure 1A**). Peak detection, alignment, and normalization were performed by MS-DIAL in both the negative and positive ion modes. Using principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA), we found that the leaf and root groups were significantly separated from each other in both the negative and positive ion modes (**Figures 1B,C**). Among these metabolites, ATDs and PAs were focused for further characterization. Twenty-five metabolites were identified in total based on accurate mass and MS/MS fragment patterns, eight of which were identified using chemical standards and the rest were discovered by analyzing the tR, MS fragment data, and comparing it to literatures (**Figures 1D,E**, **Supplementary Table 1** and **Supplementary Figures 1A,B**). As a result, carnosic acid (19) and its derivatives (6, 7, 8, and 10), along with the common precursors of ATDs such as sugiol (10), 11-hydroxysugiol (11), and 11,20-dihydroxysugiol (13), were discovered in leaves (13), while tanshinones (17, 22, 24, and 25), another group of ATDs, were specially detected in roots. In addition, ATD precursors (10, 11, and 13) were also found in roots. Furthermore, PAs as danshensu (1), caffeic acid (2), and rosmarinic acid (3) were found. The S-plots produced by the OPLS-DA analysis identified the most distinct metabolites [with variable importance in projection (VIP) value >4, **Supplementary Tables 2, 3**] (**Figure 1F**). As expect, ATDs as compound 5, 6, 11, 17, 19, and 21 were the most distinct metabolites between roots and leaves (**Figure 1G**). Notably, *S. apiana* was found to contain both the tanshinones- and carnosic acid-related ATDs, which are quite unique in *Salvias*. Since *Salvia miltiorrhiza* (*S. miltiorrhiza*), for example, just contains tanshinones in roots, while *Salvia officinalis* produces only carnosic acid and related compounds in leaves (Guo et al., 2013; Ignea et al., 2016; Scheler et al., 2016).

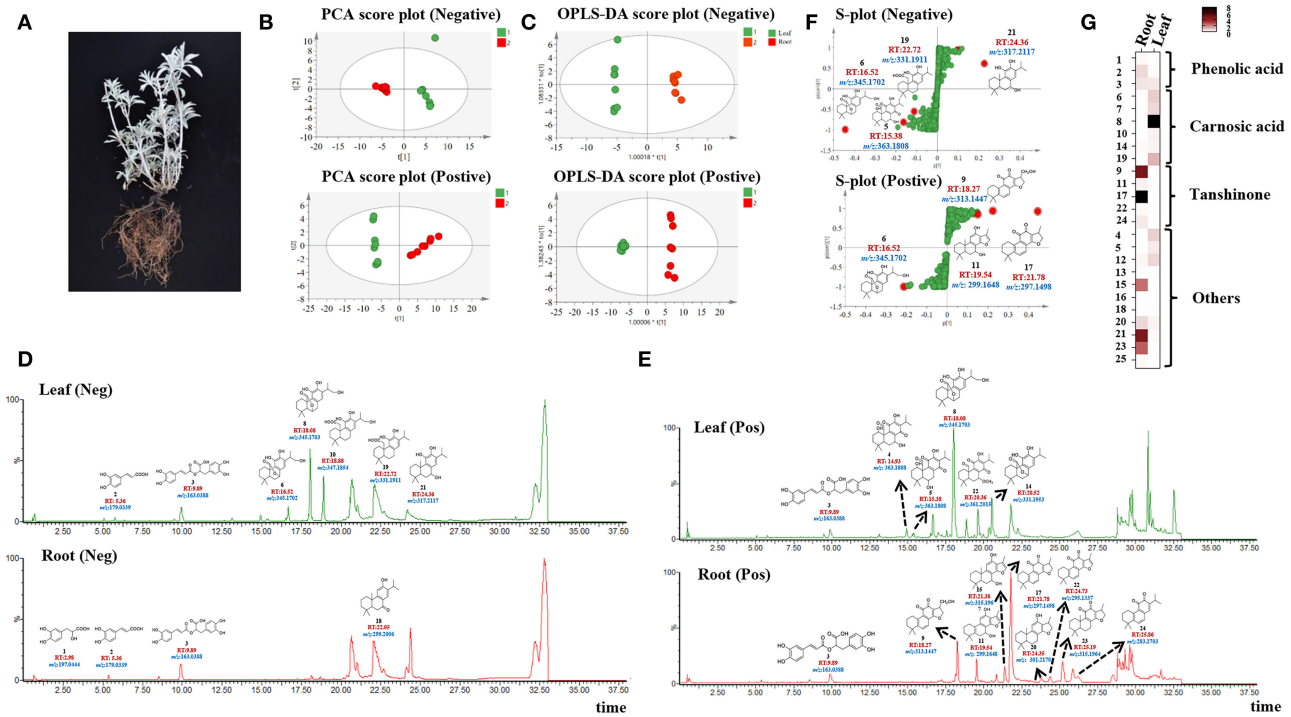


FIGURE 1 | Metabolomics and chemometric analyses of extracts from roots and leaves by QTOF-MS in *Salvia apiana*. **(A)** Image of whole plant of *S. apiana*. **(B)** PCA analysis of roots and leaves in the negative-ion and positive-ion modes. **(C)** OPLS-DA analysis of roots and leaves in the negative-ion and positive-ion modes. **(D)** Total ion current (TIC) chromatograms of leaves and roots in the negative-ion modes annotated with identified metabolites (1, 2, 3, 6, 8, 10, 18, 19 and 21), including corresponding structure, retention time (RT), and mass to charge ratio (m/z). **(E)** Total ion current (TIC) chromatograms of leaves and roots in the positive-ion modes annotated with identified metabolites (3, 4, 5, 8, 9, 11, 12, 14, 15, 17, 20, 23 and 24), including corresponding structure, retention time (RT), and mass to charge ratio (m/z). **(F)** S-plot derived from OPLS-DA analysis in the negative-ion and positive-ion modes annotated with the most distinct metabolites, including corresponding structure, retention time (RT), and mass to charge ratio (m/z). **(G)** Relative content of identified metabolites in roots and leaves. 1, danshensu; 2, caffeic acid; 3, rosmarinic acid; 6, 16-oxhydryl-carnosol; 7, rosmanol; 8, 16-oxhydryl-isocarnosol; 9, 17-oxhydryl-cryptotanshinone; 10, 6-oxhydryl-carnosic acid; 13, 11, 20-dihydroxysugiol; 14, carnosol; 16, 11-hydroxysugiol; 17, cryptotanshinone; 18, sugiol; 19, carnosic acid; 22, tanshinone IIA; 24, miltirone; 25, 11-hydroxyferruginol. Names of metabolite 4, 5, 11, 12, 15, 20, 21 and 23 were listed in **Supplementary Table 1**.

Full-Length Sequencing and Annotation of *S. apiana* Transcriptome

The full-length transcriptome of *S. apiana* was obtained using PacBio SMRT sequencing. A total of 36,953,062 subreads (39.30 Gb) were produced, with the average length of 1,040 bp (**Table 1**). Circular consensus long-read sequencing was further performed to generate highly accurate long high-fidelity (HiFi) reads to improve the quality of the subreads. As a result, a total of 310,713 circular consensus sequences (CCSs) were generated, with an average length of 1,317 bp and the N50 of 1,565 bp (**Table 1**). The full-length non-chimeric (FLNC) reads were identified by screening the coexistence of 5′-primers, 3′-primers, and poly-A tails, which resulted in 189,999 FLNC reads (61.15% of CCS) with an average length of 894 bp and the N50 of 1,255 bp (**Table 1**). After clustering the redundant data, 110,985 final consensus reads were obtained, among which 44,169 (39.80%) consensus reads were over 1,000 bp, 8,522 (19.29%) consensus reads were over 2,000 bp, and 652 (7.65%) consensus reads were over 3,000 bp in length (**Figure 2**).

Gene annotation was carried out by searching against six protein databases [nonredundant (NR), SwissProt, Pfam, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and Clusters of Orthologous Groups of proteins (COG)] based on sequence similarities. A total of 61,659 (79.04% of 78,009) transcripts were mapped to at least one database, with NR having the highest number of hits (61,469 transcripts, 78.80%, **Table 1** and **Figure 2A**). Among these transcripts, 54,710 (70.13%) transcripts were linked to the GO categories, including molecular functions (MFs), cellular components (CCs), and biological processes (BPs) (**Figure 2B**), with CCs as the majority of the GO terms. By the KEGG, 31,950 transcripts were further assigned into 135 biological pathways (**Figure 2C** and **Supplementary Table 4**). Notably, a total of 581 transcripts were sitting in the "metabolism of terpenoids and polyketides" pathway, which would help to reveal the ATDs biosynthesis pathways in *S. apiana* in the future.

Phylogenetic Analysis of *S. apiana*

To determine the phylogenetic position of *S. apiana* within the genus *Salvia*, orthologous genes from other available *Salvia* genomes [*Salvia splendens* (*S. splendens*), *S. miltiorrhiza*, and *Rosmarinus officinalis*], with *Nepeta cataria* (Nepeta, Lamiaceae) as an outgroup (Dong et al., 2018; Bornowski et al., 2020;

Lichman et al., 2020b; Song et al., 2020), were annotated. In total, 130 single-copy genes were selected and used to reconstruct a maximum likelihood phylogenetic tree (**Figure 3**). As expected, *S. apiana* was phylogenetically categorized into the same taxa as *S. splendens* in America, which might diverge from *S. miltiorrhiza*, a species of East Asia taxa, at about 28.24 Mya.

Transcriptome Profiling of *S. apiana* Leaves and Roots

To identify differentially expressed genes (DEGs) in leaves and roots of *S. apiana*, RNA sequencing (RNA-seq) was performed on the Illumina NovaSeq 6000 platform. Three samples of each organ were sequenced, which produced 42.97 Gb clean data in total (**Supplementary Table 5**). After assembly, unigenes derived from Illumina sequencing were mapped to full-length transcriptome datasets with a match percentage ranging from 71.81 to 79.01%. Using DEGs analyses, we showed that the transcriptional levels of 15,152 and 14,954 transcripts were significantly higher in roots and leaves, respectively (**Figure 4A**). To figure out the most significant functional difference between roots and leaves, the GO enrichment of DEGs was further performed (**Figure 4B**). As a result, metabolic process and cellular process were the most enriched terms in the category of molecular function. In the category of cellular components, most DEGs were enriched in membrane part and cell part. In the category of cellular components, catalytic activity and binding were the most different terms. We noticed that in terms of metabolic process and catalytic activity, these were massively enriched, suggesting that the secondary metabolic pathways in roots and leaves were significantly different. Consistent with the physiological function differences between the aerial and underground parts of plants, photosynthesis, primary metabolic, and plant hormones were found as the most different pathways by the KEGG pathway analysis. In the secondary metabolic pathways, phenylpropanoids biosynthesis was the most significant pathway, showing different transcriptional levels of lignins, flavonoids, and phenolic acids. The variation in diterpenoids pathways was not substantial, which could be attributed to the fact that different metabolites share similar upstream pathways (**Figure 4C**).

TABLE 1 | Statistics of *S. apiana* single-molecule real-time (SMRT) sequencing results.

Project	SMRT
Total subreads	36,953,062
Total subread nucleotides (bp)	39,303,737,421
CCS reads	310,713
FLNC reads	189,999
Consensus reads	110,985
Representative transcripts	78,009
Total size (bp) of representative transcripts	81,125,619
GC content	0.44
Transcript length range (bp)	51–6,262
Average transcript size (bp)	1,039.95
Median transcript size (bp)	910
N50 transcript size	1,466
Transcripts with functional annotation	61,659
Nr	61,469
Swiss-Prot	48,135
Pfam	43,373
KEGG (KO)	31,950
COG	55,743
GO	54,710

Transcriptional Variation in Abietane-Type Diterpenoids and Phenolic Acids Biosynthetic Pathways

As the KEGG pathway analysis showed above, metabolism of terpenoids and polyketides (KEGG Ortholog 09109) represented as one of the most enriched pathways with 581 unigenes (**Supplementary Table 6**). According to the elucidated ATDs

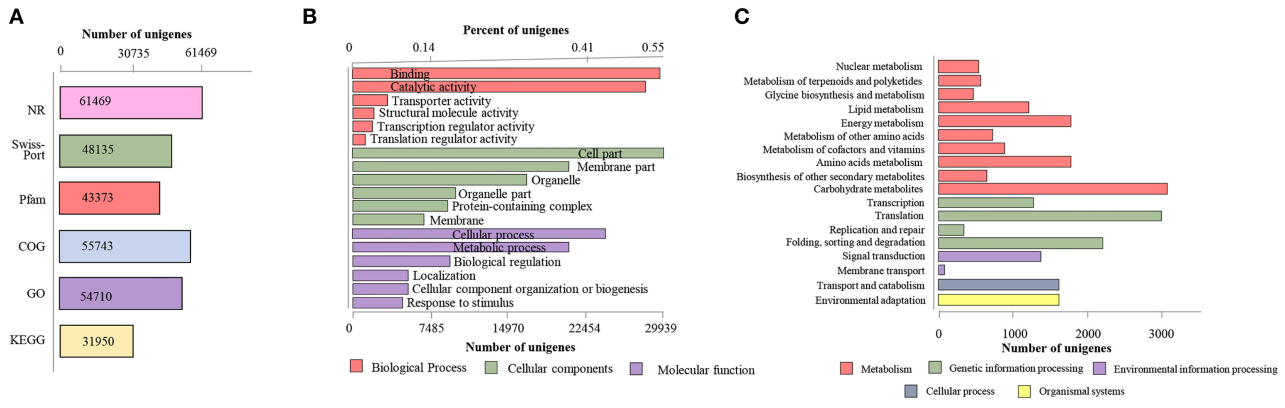


FIGURE 2 | General annotation of *S. apiana* transcriptome. **(A)** Annotation statistics searching against different data banks. **(B)** The most enriched Gene Ontology (GO) terms in three major functional categories. The top six terms of each category were represented. **(C)** The most enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Different databases, the GO categories, and the KEGG pathways are distinguished by colors.

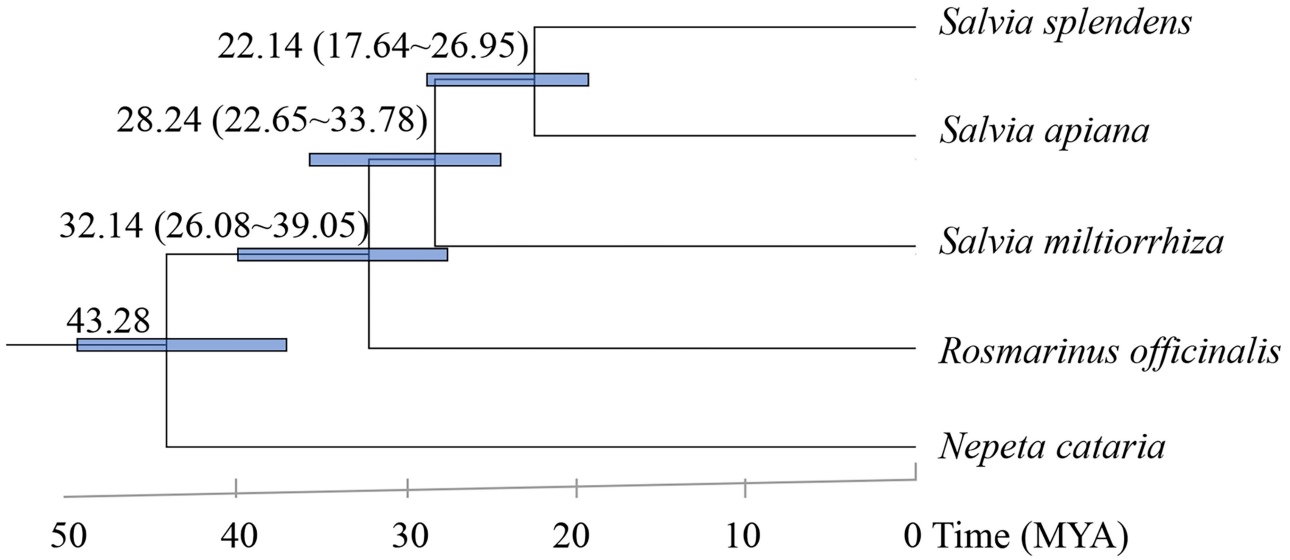


FIGURE 3 | Phylogenetic tree of five plant species. Blue numerical value beside each node shows the estimated divergence time of each node (MYA, million years ago).

FIGURE 4 | Transcription profiles of *S. apiana* roots and leaves. **(A)** Volcano plots represent the differentially expressed genes (DEGs) with higher transcriptional level in roots (blue plots) and leaves (red plots), respectively. The genes involved in terpenoids (green plots) and phenolic acids (yellow plots) biosynthesis are marked. **(B)** Statistical analysis of the most enriched GO terms from three major categories. **(C)** The top twenty KEGG pathways with the most significant variations of DEGs between the roots and leaves. The number of DEGs is represented by the size of circle.

pathway in other *Salvia* plants, including *S. miltiorrhiza*, *S. fruticose*, and *S. officinalis* (Guo et al., 2013; Ignea et al., 2016; Scheler et al., 2016), we predicted the ATD biosynthesis pathway in *S. apiana* (**Figure 5A**). A total of 50 unigenes were supposed to be involved in the ATDs biosynthesis and could be considered as four modules. The geranylgeranyl diphosphate (GGPP) synthesis via the methylerythritol 4-phosphate (MEP) and mevalonate (MVA) pathways, along with the general diterpenoids catalytic steps, provided the common precursor (ferruginol) for ATDs biosynthesis. Then, the pathway was split into the biosynthesis of carnosic acids and tanshinones catalyzed by different CYP76AHs and CYP76AKs, respectively. The transcriptional level of these genes was further profiled by DEG analysis (**Figure 5B**). Most genes in MVA pathway showed high transcriptional levels in root, indicating that they play a prominent role in the production of the terpenoids precursor, isopentenyl diphosphate (IPP). However, the genes in MEP pathway and ferruginol synthesis did not show significant transcription trend. In addition, most *CYP76AH*, *CYP76AK*, and *2-ODD* genes, which were specific for ATDs biosynthesis, showed higher transcriptional level in root, indicating the high accumulation of tanshinones. There was also one *CYP76AK* gene (transcript_30977) expressed highly in leaf, which might contribute to the carnosic acids biosynthesis. Taken together, the transcriptional profile of ATDs pathway was basically consistent with its metabolic properties in roots and leaves of *S. apiana.* Additionally, forty unignes might be involved in the biosynthesis of PAs were annotated (**Figure 6A**), most of which were highly expressed in root, consisting with the high accumulation of phenolic acids in roots (**Figure 6B**). Given that different ATDs and PAs are distributed in roots and leaves, common genes involved in both the metabolic pathways, especially their upstream genes, were found to be highly expressed in both the organs (**Figure 4A**). Genes related to the metabolic branches, on the other hand, were often organ specific. For example, genes in tanshinones biosynthesis (*CPS*, *CYP76AHs*, *CYP76AKs*, *CYP71D375s*, and *2-ODDs*) were mostly highly expressed in roots.

Phylogeny of Cytochrome and Laccase Families

The functional specificity of Lamiaceae cytochromes (CYPs) determines the particular products of ATDs and PAs in *Salvia* plants (Hansen et al., 2021). Therefore, a total of 132 *CYP* genes were discovered in *S. apiana* by Pfam annotation and used for phylogenetic tree reconstruction, together with 578 CYPs from other plants (**Figure 7A**). Among these, CYP98A and CYP76 families were grabbed for detailed analysis. CYP98As are usually related to phenolic acids biosynthesis, while CYP76AKs and CYP76AHs are unique in Lamiaceae, which regulate the specific ATDs biosynthesis. In *S. apiana*, we found six homologous of CYP98A, two homologous each of CYP76AK and CYP76AH, respectively. We further studied the phylogeny of all the CYPs from *S. apiana*, which were classified into 26 families among six clans (**Figure 7B**), with 71 clan representing the richest families and CYP71 as the largest family. Interestingly, although *S. apiana* had more abundant ATD metabolites than that in other *Salvia* plants, CYP76AK and CYP76AH subfamilies did not seem to have experienced gene expansion, suggesting a functional diverse of these CYPs. On the other hand, we found that CYP98A subfamily might expanded, consisting with more complex composition of PAs in *S. apiana*.

Apart from CYP98A-mediated hydroxylation, polymerization of phenolic polymers promotes the metabolic variety of phenolic acids, which were previously supposed to be catalyzed by the laccase family in *Salvia* plants. However, the specific laccase remains to be identified (Di et al., 2013; Li et al., 2019; Zheng et al., 2021). Metabolic profiling revealed the absence of these phenolic polymers, implying the lack or mutation of corresponding laccases in *S. apiana*. Thus, the phylogeny of laccase family was examined in both the *S. apiana* (27 members) and *S. miltiorrhiza* (29 members) (**Figure 7C**). All the laccases were classified into seven clades, but clade VI was absent in *S. apiana*. We also found that the clade VII in *S. miltiorrhiza* showed significant gene expansion. Taken together, these extra laccase genes might contribute to the formation of phenolic polymers such as salvianolic acid B in *S. miltiorrhiza*.

DISCUSSION

S. apiana, known as white sage, is of ritual meaning and great medicinal value in Southern California historically. However, previous studies on *S. apiana* were mainly focused on the pharmaceutical perspective as phytochemistry and biological activities. Here, based on the establishment of full-length transcriptome of *S. apiana*, we were able to investigate its taxonomical position, chemical specificities, as well as genetic characteristics.

The origins of *Salvia* taxa in the New World, as well as their dispersal, are still debated (Will and Claßen-Bockhoff, 2017). According to the phylogenetic relationship of *S. apiana*, *S. splendens*, and *S. miltiorrhiza*, we estimated that the divergence between America *Salvia* taxa and North Asia taxa was at about 28.24 Mya. This was a little earlier than the previous estimation (17.6 Mya) according to the phylogeny between *Salvia bowleyana* (*S. bowleyana*), *S. splendens*, and *S. miltiorrhiza* (Zheng et al., 2021). However, both results suggested that the two lineages were diverged during late Oligocene or early Miocene, which is consisted with previous studies of the *Salvia* lineages (Kriebel et al., 2019; Krol et al., 2021). However, *S. apiana* had not been considered yet, which added nuclear gene information for North America taxa.

Besides phylogenetic position, *S. apiana* appeared to occupy a unique position in the terpenoids biosynthesis metabolic divergence. Based on the metabolic profile analyses, the presence of most of the structures of ATDs seems to be species or lineage specificity, such as tanshinones in the roots of *S. miltiorrhiza* and carnosic acid in *S. officinalis* (Guo et al., 2016; Ignea et al., 2016; Scheler et al., 2016). The production of two classes of ATDs appears to be species or taxonomy specific in such *Salvias*. Since *S. apiana* has all-around catalytic capacity for ATDs, it is supposed to be a suitable model for studying ATD biosynthesis. First, the diverse biosynthesis of different ATDs represented a

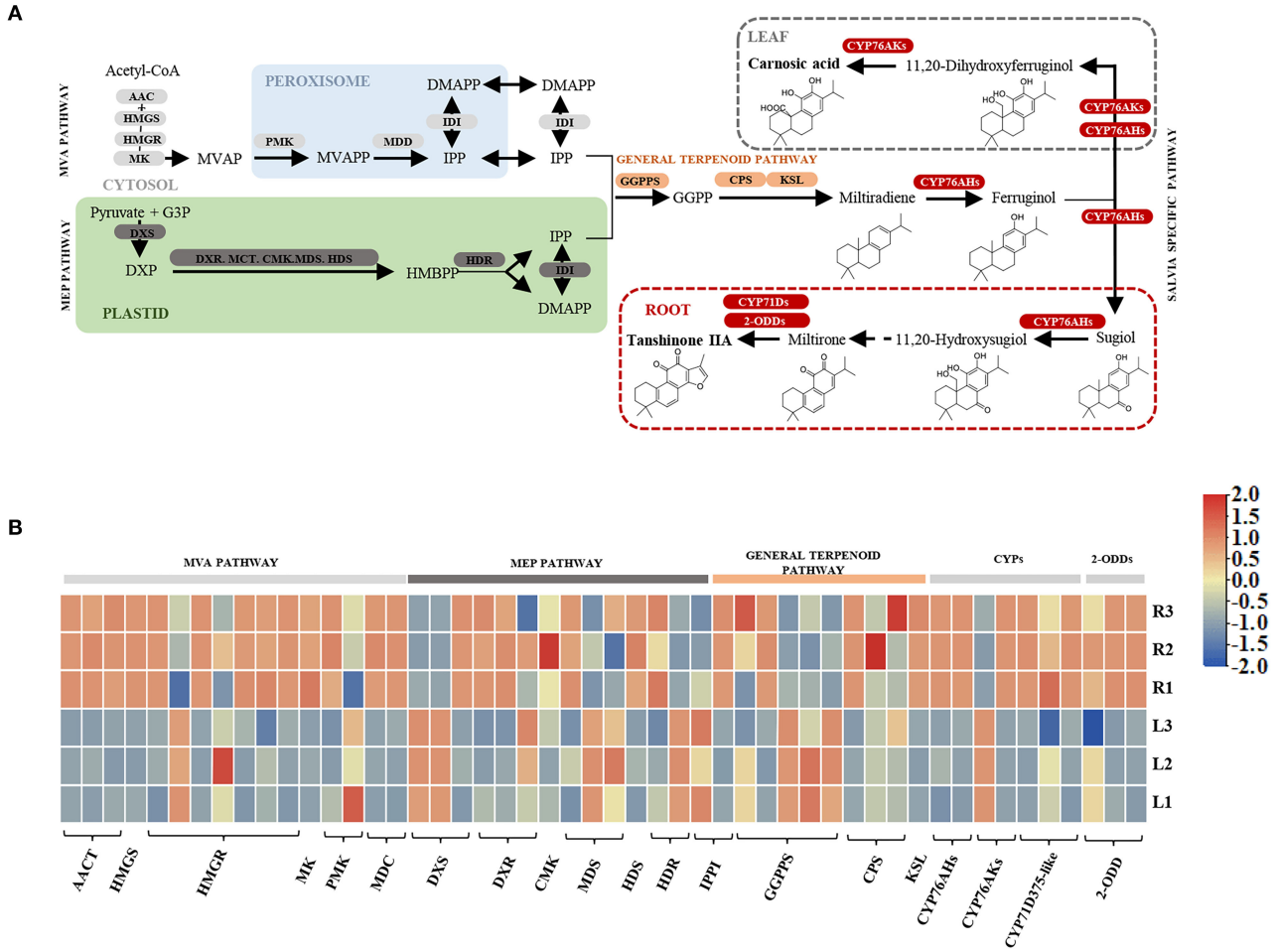


FIGURE 5 | A proposed biosynthetic pathway of abietane-type diterpenoids (ATDs) in *S. apiana*. **(A)** Pathways and genes involved in the ATDs biosynthesis. **(B)** Transcription profile of candidate ATDs biosynthetic genes in roots and leaves of *S. apiana*. Three biological replicates from the leaf (L) and root (R) were plotted individually. AACT, Acetyl-CoA acetyltransferase activity; CYP, cytochrome P450; 2-ODD, 2-oxoglutarate-dependent dioxygenase; HMGS, 3-hydroxy-3-methylglutaryl coenzyme A synthase; HMGR, hydroxymethylglutaryl-CoA reductase; KSL, kaurene synthase-like enzymes; MK, mevalonate phosphotransferase; MVAP, mevalonate 5-phosphate; PMK, phosphomevalonate kinase; MVAPP, mevalonate 5-diphosphate; MDD, mevalonate diphosphate decarboxylase; IPP, isopentenyl diphosphate; IDI, isopentenyl diphosphate delta-isomerase; DMAPP, dimethylallyl diphosphate; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-(cytidine 50-diphospho)-2-C-methyl-D-erythritol kinase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase; HDR, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase; GPPS, geranyl pyrophosphate synthase; GPP, geranyl diphosphate; GGPP, geranylgeranyl diphosphate; MVA, schematic of mevalonate; MEP, 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate.

typical evolutionary event for emergence and loses of metabolic pathway that accompany with speciation within the taxa (Mint Evolutionary Genomics Consortium, 2018; Lichman et al., 2020b). Second, the divergent biosynthesis routes provide direct gene information for elucidating ATD biosynthesis pathways. Based on the current understanding of ATD biosynthesis, two CYP subfamilies, CYP76AKs and CYP76AHs, are critical for the metabolic diversity of ATDs. CYP76AK6, 7, and 8 catalyze successive oxidations at C20 to generate a carboxyl group (Ignea et al., 2016; Scheler et al., 2016), whereas CYP76AK1 only generates an alcohol group at this position (Guo et al., 2016). From our transcriptome of *S. apiana*, two CYP76AKs and two CYP76AHs were discovered. We also found that CYP76AKs and CYP76AHs subfamilies in *S. apiana* were not expanded during evolution, implying that these CYPs have a broad catalytic capability. Thus, further functional investigations and whole-genome duplication (WGD) analyses of *S. apiana* genome might improve the understanding of ATDs biosynthesis in plant and the *S. apiana*'s all-around catalytic capacity. A recently published genome of *Salvia hispanica* (*S. hispanica*), which belongs to America linage as well, revealed the complexities for the genome evolution within the *Saliva* linages, as well as the evolution of ATDs biosynthesis. The size of *S. hispanica* genome is about 347.6 Mb, which is nearly half the size of other known *Salvia* genomes, as *S. miltiorrhiza* (641 Mb), *S. splendens* (711 Mb), and *S. bowleyana* (462 Mb) (Song et al., 2020; Jia et al., 2021;

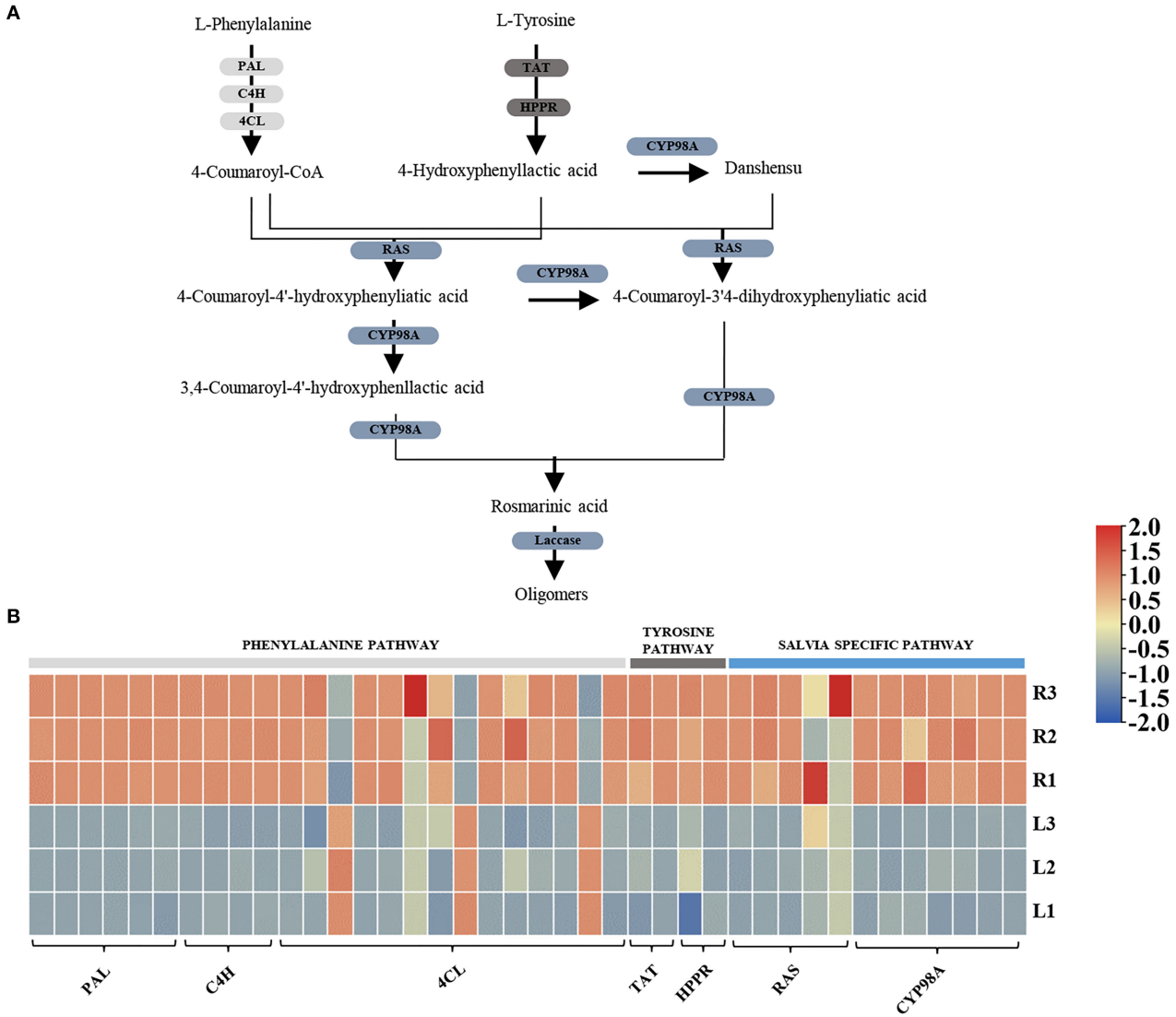


FIGURE 6 | A proposed biosynthetic pathway of phenolic acids (PAs) in *S. apiana*. **(A)** Pathways and genes involved in PAs biosynthesis. **(B)** Expression profile of candidate PAs biosynthetic genes. Three biological replicates from the leaf (L) and root (R) were plotted individually. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:coenzyme A ligase; TAT, tyrosine aminotransferase; HPPR, hydroxyphenylpyruvate reductase; RAS, rosmarinic acid synthase; CYP, cytochrome P450.

Zheng et al., 2021). Compared with *S. splendens*, a recent WGD was absent, thus resulting in the small size of *S. hispanica*. The WGD events during the genome evolution are always been considered as major driven forces for the chemodiversity of plants (Lichman et al., 2020a). Thus, the WGD analyses and the genome of *S. apiana* might help to explain the functional diversities of CYP76AKs and specialized ATDs production in this species. Furthermore, we believe that a lineage- or genus-wide chemical profile of ATDs would be helpful to clarify the metabolic distribution among the *Salvia* species and provide definitive evidences for fully elucidating ATDs biosynthesis.

In contrast to ATDs, phenolic polymers as salvianolic acid B were not produced by *S. apiana*, indicating the lack of corresponding catalytic enzymes (Di et al., 2013; Hou et al., 2013). The biosynthesis of phenolic acids is composed of the general phenylpropane pathway, tyrosine-derived pathway, and additional enzymes (Petersen and Simmonds, 2003). A BAHD acyltransferase member [rosmarinic acid synthase (RAS)] and CYP98A subfamily were reported to contribute to the specific PAs biosynthesis in Lamiaceae by producing rosmarinic acid. Inferring from the indirect evidences, laccase family might catalyze further polymerization (Di et al., 2013; Li et al., 2019; Zheng et al., 2021; Zhou et al., 2021); however, it still needs to be proved. The absence of salvianolic acid B in *S. apiana* implies the lose function of laccase family, which can be considered as a natural mutant. When comparing the laccase family between

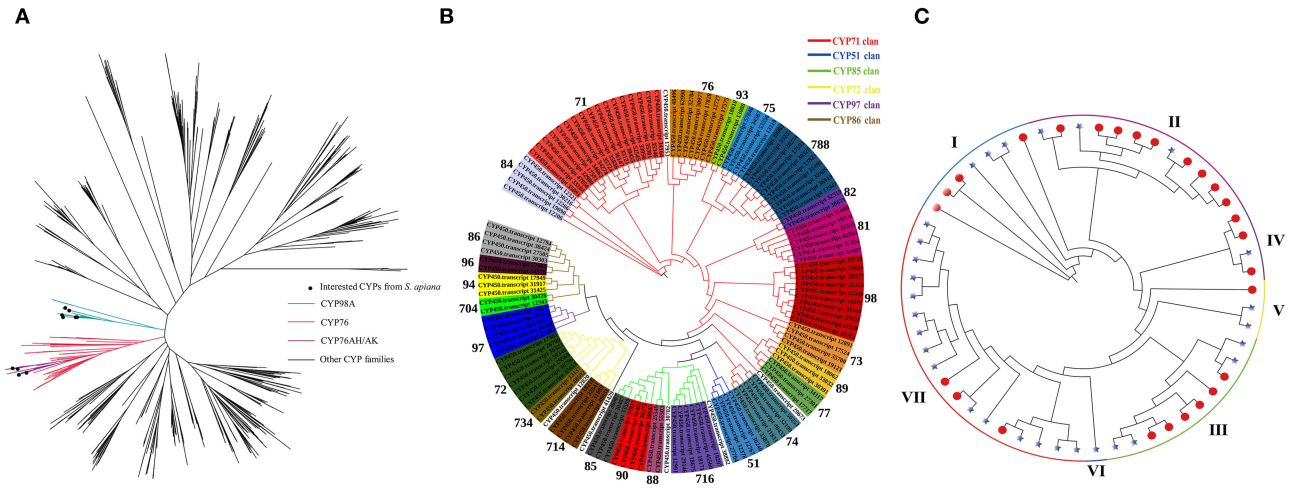


FIGURE 7 | Phylogenetic analysis of cytochrome (CYP) and laccase gene families from *S. apiana*. **(A)** Maximum likelihood (ML) phylogenetic tree of 132 CYPs from *S. apiana* compared to 578 CYPs from other plant species. The blue branch represents the CYP98A subfamily. The red branch represents the CYP76 family, with CYP76AK and CYP76AH subfamilies in purple. **(B)** Maximum likelihood (ML) phylogenetic tree of 132 CYPs from *S. apiana*. Branches in different colors represent different CYP clans: CYP71 (red), CYP51 (blue), CYP85 (green), CYP72 (yellow), CYP97 (purple), and CYP86 (brown). Subfamilies were indicated by background colors. **(C)** ML phylogenetic tree of laccases from *S. apiana* (asterisks) and *Salvia miltiorrhiza* (*S. miltiorrhiza*) (red nodes).

S. apiana and *S. miltiorrhiza*, we discovered that clade VII has expanded in *S. miltiorrhiza*, which included the previously predictions according to mutant experiment of *SmLAC7* and *SmLAC20*. It shows that the range of candidate genes can be effectively narrowed through the comparison of laccase families between *Salvia* species.

CONCLUSION

In summary, we reported the first full-length transcriptome of medicinal plant *S. apiana*. The distribution and contents of ATDs and PAs exhibited tissue-specific patterns (leaf and root) in *S. apiana* by UPLC-MS analysis. Such tissue-specific accumulation of ATDs and PAs was further verified with genes in ATDs and PAs biosynthetic pathways through systematic metabolomic and transcriptomic analyses. Moreover, phylogenetic analysis identified five candidate genes for ATDs synthesis; further functional studies of these genes are needed to verify their roles in ATDs biosynthesis. Together, this study provides novel insights into the molecular basis for the biosynthesis of ATDs and PAs in *S. apiana* and serves as an approach with which to analyze full-length transcriptome data and the biosynthesis process in other *Salvia* plants.

MATERIALS AND METHODS

Plant Materials

Salvia apiana was planted in a greenhouse with 25°C and humidity of 40–70% at Shanghai University of Traditional Chinese Medicine for 6 months (May 2020 to November 2020). The fresh leaves and roots of *S. apiana* were harvested at vegetative stage for both the metabolic and transcriptomic analyses.

Chemical Standards

Chemical standards, including danshensu, caffeic acid, rosmarinic acid, cryptotanshinone, sugiol, carnosic acid, tanshinone IIA, and miltirone, were purchased from Shanghai Yuanye Biotechnology Corporation Ltd. (Shanghai, China). The other chemicals and reagents were purchased as follows: acetonitrile and methanol (HPLC grade), Merck (Darmstadt, Germany); warfarin, Sigma-Aldrich (Madrid, Spain); chloroform, Sinopharm Chemical Reagent (Shanghai, China); leucine encephalin, Waters (Milfor, Massachusetts, USA); pure distilled water, Watsons water (Hong Kong, China); and formic acid (HPLC grade), Fisher Scientific (Fair Lawn, New Jersey, USA).

Ultra-High-Performance Liquid Chromatography Coupled With Quadrupole Time-of-Flight Mass Spectrometry Analysis for the Non-targeted Metabolomics Study

Each of 8 plants was divided into roots and leaves separately, represented as eight biological replicates. All the samples were crushed into homogeneous powder. Ten milligram prepared powder was extracted with 1 ml of 70% methanol (v/v) containing warfarin (5 μg/ml) as the internal reference for 1 h in an ultrasonic bath (53 kHz, 350 W) at 4°C. After centrifuged at 12,000 g under 4°C for 30 min, the supernatant was used for UHPLC-Q-TOF-MS analysis.

Analyses were performed using the ACQUITY UPLC System (Waters) coupled with the Xevo G2-XSQTOF mass spectrometer (Waters). Briefly, samples were first separated using an ACQUITY UPLC T3 column (2.1 × 100 mm, 1.8 μm). The column temperature was constant at 40°C and the flow rate was

0.40 ml/min with an injection volume of 1.0 μl. 0.1% (v/v) formic acid/water (solvent A) and 0.1% (v/v) formic acid/acetonitrile (solvent B) were the mobile phases for gradient elution. The elution gradients were: 98–80% A over 0–7 min; 80–78% A over 7–11 min; 78–40% A over 11–20 min; 40–35% A over 20–25 min; 30–35% A over 25–28 min; 35–5% A over 28–30 min; 5–5% A over 30–33 min; and the final re-equilibration at 2% A for 5 min.

To acquire mass spectrometry data, a Xevo G2–XS with an electrospray ionization source equipped with an electrospray ionization source was used. Mass spectrometry was carried out in both the positive ion and negative ion modes under 30 V cone voltage, with the capillary voltage of 3 kV (positive ion mode) or 2.5 kV (negative ion mode), respectively. The desolvation temperature was set at 450°C with the desolvation gas flow of 600 l/h and the source temperature was set at 150°C with a cone gas flow of 50 l/h. All the data were collected in MS^E mode, with the parameters as follows: MS^E range 50–1,200 m/z, MS^E low energy 6 eV, and MS^E high energy 15–30 eV. To calibrate the instrument, a sodium formate solution (0.5 mM) was used. Leucine enkephalin was continuously acquired and used as an external standard for mass correction. All the data were viewed in MassLynx version 4.2.

Processing of Metabolomics Data

The metabolomics data was first converted to the Analysis Base File (ABF) converter format and imported into the MS-DIAL version 4.60 software (Tsugawa et al., 2015). The parameter settings were as follows: MS^1 and MS^2 tolerance were 0.01 and 0.02 Da, respectively. MS^1 mass range and MS/MS mass range were set between 100 and 1,000 with the MS/MS abundance cutoff at 800 amplitudes. Retention time range was set between 1 and 28 min with the retention time tolerance at 0.15 min. Mass slice width was 0.1 Da. For negative ion mode, adduct types, including $[M-H]^-$, $[M+HCOO]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$, $[2M-H]^-$, and $[2M+FA-H]^-$, were selected. For positive ion mode, $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH4]^+$, and $[2M+H]^+$ were selected. Peak tables generated from MS-DIAL were cleaned and clustered by the MS-CleanR (Fraisier-Vannier et al., 2020) with the following parameters: minimum blank ratio at 0.8; relative mass defect between 50 and 7,000 with the maximum mass difference at 0.005 Da; and maximum relative SD at 30 with the maximum retention time difference at 0.03 min. To generate the final peak table for metabolite identification, the highest intensity in each cluster was retained.

Metabolite Identification and Chemometric Analysis

MS-FINDER was applied to identify the structures of unknown metabolites with reference databases such as UNPD, Northern African Natural Products Database (NANPDB), Human Metabolome Database (HMDB), PlantCyc, KNApSAcK, and LipidMaps (Tsugawa et al., 2016). The structures of compounds were confirmed by matching with theoretical data or public databases.

To obtain normalized metabolomics data, the raw data was processed in MS-DIAL with warfarin (5 μg/ml) as internal standard in each sample. Normalized metabolomics data was directly imported into SIMCA-P 14.1 (Umetrics AB, Umea, Sweden). After Pareto scaling, the orthogonal partial least-squares discriminant analysis (OPLS-DA) was applied to determine the supervised pattern recognition among the defined groups. To determine the unsupervised pattern recognition of leaves and roots, the principal component analysis (PCA) was performed. Only metabolites with variable importance in projection (VIP) value larger than 1.5 obtained from the OPLS-DA model were considered as the potential chemical markers with high contribution.

Ribonucleic Acid Extraction, Quality Assessment, and Quantification

Total RNA was extracted using the TransZol Plus RNA Kit (TransGen, China) according to the manufacturer's protocol. RNA integrity was determined by the 2100 Bioanalyzer (Agilent Technologies Incorporation, Santa Clara, California, USA) and the OD260/280 values were determined with the ND-2000 (NanoDrop Technologies, Wilmington, Delaware, USA). Only high-quality RNA samples were preserved for later Iso-Seq library construction.

PacBio Library Construction and Sequencing

The PacBio libraries were prepared separately using mixed messenger RNA (mRNA) from three leaf or root samples. The Iso-Seq library was prepared using the Clontech SMARTer PCR cDNA Synthesis Kit (Takara Bio, Mountain View, California, USA) and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). Then, sequencing was performed on a PacBio Sequel platform (Pacific Biosciences).

Illumina Complementary DNA Library Construction and Next-Generation Sequencing

The complementary DNA (cDNA) library for next-generation sequencing (NGS) was prepared by using the Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, California, USA) according to the protocol provided by the manufacturer. Sequencing was performed on the Illumina NovaSeq 6000 Sequencer (Illumina, San Diego, California, USA) by Shanghai Majorbio Biopharm Biotechnology Corporation Ltd. (Shanghai, China).

Sequence Assembly and Gene Annotation

The raw paired-end reads were trimmed and quality controlled by fastp (https://github.com/OpenGene/fastp) with default parameters. Then, clean data from the samples of *S. apiana* was *de-novo* assembly into contigs using Trinity assembler (Grabherr et al., 2011).

To annotate the gene function, all the assembled transcripts were searched against the following databases: National Center for Biotechnology Information (NCBI) non-redundant (NR) protein sequences database; protein family (Pfam); Clusters of Orthologous Groups of proteins (KOG/COG);

a manually annotated and reviewed protein sequence database (Swiss-Prot); KEGG Ortholog database (KO); and the Gene Ontology (GO). BLASTX was used for NR, KOG, Swiss-Prot, and the KEGG database analysis with the E-value set at 1e-5 (Camacho et al., 2009). The Pfam search results were generated using HMMER version 3.1b2 (with E < 1e-5).

Phylogenetic Tree Construction and Divergence Time Estimation

A single-copy orthologous nuclear gene set for Lamiaceae was used to construct the phylogenetic tree (Mint Evolutionary Genomics Consortium, 2018). Phylogenetic relationship among these 5 plant species was resolved using the RAxML package (version 8.1.13) (Stamatakis, 2014). Their divergence times were estimated by the program MCMCtree in PAML (version 3.15) (http://abacus.gene.ucl.ac.uk/software/paml.html).

Multiple sequence alignments of CYP targets were performed using the ClustalW program and phylogenetic trees were constructed by MEGA version 7.0. The maximum likelihood statistical method was used to calculate the phylogenetic tree, with 1,000 bootstrap replications.

Differential Expression Analysis and Functional Enrichment

The expression level of each transcript was analyzed by using the transcripts per million reads (TPM) method in RSEM software (Li and Dewey, 2011). The analysis of differential expressed genes (DEGs) was performed by using DESeq2 (Love et al., 2014) with p-adjust <0.05 and $|\log2FC| \geq 1$. In order to identify DEGs, the GO and the KEGG enrichment analyses were performed by Goatools (https://github.com/tanghaibao/Goatools) and KOBAS (Xie et al., 2011), respectively, with significantly enriched in the GO terms and metabolic pathways at Bonferroni-corrected $P \leq 0.05$.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ngdc.cncb.ac.cn/, CRA006765 and CRA006779.

AUTHOR CONTRIBUTIONS

WC and QL were the leading investigators of this research program. QL designed the experiments. JH performed most of experiments and analyzed the data. Other authors assisted in experiments and discussed the results. JH and FL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by the National Key R&D Program of China (2019YFC1711100), the National Natural Science Foundation of China (32070327 and 31770329), the Young Elite Scientists Sponsorship Program by Cast (2021-QNRC1-03), and the Research Project of Science and Technology Commission of Shanghai Municipality (21DZ2202300).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.919025/full#supplementary-material

REFERENCES

Adlof, C. C. (2015). *How Does Harvesting Impact White Sage (Salvia apiana) as a Cultural Resource in Southern California?* [Dissertation]. Northridge: California State University.

Afonso, A. F., Pereira, O. R., Fernandes, N., Calhelha, R. C., Cardoso, S. M., Ferreira, I. C., et al. (2019). The health-benefits and phytochemical profile of *Salvia apiana* and *Salvia farinacea* var. *Victoria Blue* decoctions. *Antioxidants* 8:241. doi: 10.3390/antiox8080241

Bisio, A., Pedrelli, F., D'Ambola, M., Labanca, F., Schito, A. M., Govaerts, R., et al. (2019). Quinone diterpenes from *Salvia* species: chemistry, botany, and biological activity. *Phytochem. Rev.* 18, 665–842. doi: 10.1007/s11101-019-09633-z

Bornowski, N., Hamilton, J. P., Liao, P., Wood, J. C., Dudareva, N., and Buell, C. R. (2020). Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. *DNA Res.* 27:dsaa016. doi: 10.1093/dnares/dsaa016

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

D'Amelia, V., Ruggiero, A., Tranchida-Lombardo, V., Leone, A., Tucci, M., and Docimo, T. (2017). "Biosynthesis of *Salvia* specialized metabolites and biotechnological approaches to increase their production," in *Salvia Biotechnology*, eds V. Georgiev, A Pavlov (Cham: Springer), 241–270. doi: 10.1007/978-3-319-73900-7_7

Dentali, S. J. (1991). *Potential Antiinfective Agents From Eriodictyon Angustifolium Nutt and Salvia apiana Jeps [Dissertation]*. Ann Arbor: University of Arizona. doi: 10.3109/13880209209054003

Di, P., Zhang, L., Chen, J., Tan, H., Xiao, Y., Dong, X., et al. (2013). ^{13}C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. *ACS. Chem. Biol.* 8, 1537–1548. doi: 10.1021/cb3006962

Dong, A. X., Xin, H. B., Li, Z. J., Liu, H., Sun, Y. Q., Nie, S., et al. (2018). High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *Gigascience* 7:giy068. doi: 10.1093/gigascience/giy068

Fraisier-Vannier, O., Chervin, J., Cabanac, G., Puech, V., Fournier, S., Durand, V., et al. (2020). MS-CleanR: a feature-filtering workflow for untargeted LC–MS based metabolomics. *Anal. Chem.* 92, 9971–9981. doi: 10.1021/acs.analchem.0c01594

González, A. G., Aguiar, Z. E., Grillo, T. A., and Luis, J. G. (1992). Diterpenes and diterpene quinones from the roots of *Salvia apiana*. *Phytochemistry* 31, 1691–1695. doi: 10.1016/0031-9422(92)83130-Q

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

Guo, J., Ma, X., Cai, Y., Ma, Y., Zhan, Z., Zhou, Y. J., et al. (2016). Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. *N. Phytol.* 210, 525–534. doi: 10.1111/nph.13790

Guo, J., Zhou, Y. J., Hillwig, M. L., Shen, Y., Yang, L., Wang, Y., et al. (2013). CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12108–12113. doi: 10.1073/pnas.1218061110

Hansen, C. C., Nelson, D. R., Mller, B. L., and Werck-Reichhart, D. (2021). Plant cytochrome P450 plasticity and evolution. *Mol. Plant* 14, 1244–1265. doi: 10.1016/j.molp.2021.06.028

Hou, X., Shao, F., Ma, Y., and Lu, S. (2013). The phenylalanine ammonia-lyase gene family in *Salvia miltiorrhiza*: genome-wide characterization, molecular cloning and expression analysis. *Mol. Biol. Rep*. 40, 4301–4310. doi: 10.1007/s11033-013-2517-3

Ignea, C., Athanasakogloua, A., Ioannoub, E., Georganteab, P., Trikkac, F. A., Loupassak, S., et al. (2016). Carnosic acid biosynthesis elucidated by a synthetic biology platform. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3681–3686. doi: 10.1073/pnas.1523787113

Jia, K. H., Liu, H., Zhang, R. G., Xu, J., Zhou, S. S., Jiao, S. Q., et al. (2021). Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (Lamiaceae) genome. *Hortic. Res.* 8:177. doi: 10.1038/s41438-021-00614-y

Khan, I., Karim, N., Ahmad,W., Abdelhalim, A., and Chebib, M. (2016). GABA-A receptor modulation and anticonvulsant, anxiolytic, and antidepressant activities of constituents from *Artemisia indica* Linn. *Evid. Based Complement. Alternat. Med*. 2016:1215393. doi: 10.1155/2016/1215393

Kriebel, R., Drew, B. T., Drummond, C. P., González-Gallegos, J. G., Celep, F., Mahdjoub, M. M., et al. (2019). Tracking temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data. *Am. J. Bot.* 106, 573–597. doi: 10.1002/ajb2.1268

Krol, A., Kokotkiewicz, A., and Luczkiewicz, M. (2021). White Sage (*Salvia apiana*)–a ritual and medicinal plant of the chaparral: plant characteristics in comparison with other *Salvia* Species. *Planta Med*. doi: 10.1055/a-1453-0964. [Epub ahead of print].

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

Li, Q., Feng, J., Chen, L., Xu, Z., Zhu, Y., Wang, Y., et al. (2019). Genome-wide identification and characterization of *Salvia miltiorrhiza* laccases reveal potential targets for salvianolic acid B biosynthesis. *Front. Plant Sci*. 10:435. doi: 10.3389/fpls.2019.00435

Lichman, B. R., Godden, G. T., and Buell, C. R. (2020a). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of Lamiaceae. *Curr. Opi. Plant Biol.* 55, 74–83. doi: 10.1016/j.pbi.2020.03.005

Lichman, B. R., Godden, G. T., Hamilton, J. P., Palmer, L., Kamileen, M. O., Zhao, D., et al. (2020b). The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv*. 6:eaba0721. doi: 10.1126/sciadv.aba0721

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

Luis, J. G., Lahlou, E. H., Andrés, L. S., Sood, G. H., and Ripoll, M. M. (1996). Apiananes: C23 terpenoids with a new type of skeleton from *Salvia apiana*. *Tetrahedron Lett.* 37, 4213–4216. doi: 10.1016/0040-4039(96)00799-X

Ma, Y., Cui, G., Chen, T., Ma, X., Wang, R., Jin, B., et al. (2021). Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. *Nat. Commun.* 12:685. doi: 10.1038/s41467-021-20959-1

Mint Evolutionary Genomics Consortium (2018). Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* 11, 1084–1096. doi: 10.1016/j.molp.2018.06.002

Petersen, M., and Simmonds, M. S. (2003). Rosmarinic acid. *Phytochemistry* 62, 121–125. doi: 10.1016/S0031-9422(02)00513-7

Saeed, M. E. M., Meyer, M., Hussein, A., and Efferth, T. (2016). Cytotoxicity of South-African medicinal plants towards sensitive and multidrug-resistant cancer cells. *J. Ethnopharmacol*. 186, 209–223. doi: 10.1016/j.jep.2016.04.005

Scheler, U., Brandt, W., Porzel, A., Rothe, K., Manzano, D., Bozic, D., et al. (2016). Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nat. Commun*. 7:12942. doi: 10.1038/ncomms12942

Song, Z., Lin, C., Xing, P., Fen, Y., Jin, H., Zhou, C., et al. (2020). A high-quality reference genome sequence of *Salvia miltiorrhiza* provides insights into tanshinone synthesis in its red rhizomes. *Plant Genome* 13:e20041. doi: 10.1002/tpg2.20041

Srivedavyasasri, R., Hayes, T., and Ross, S. A. (2017). Phytochemical and biological evaluation of *Salvia apiana*. *Nat. Prod. Res.* 31, 2058–2061. doi: 10.1080/14786419.2016.1269096

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

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12, 523–526. doi: 10.1038/nmeth.3393

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Anal. Chem*. 88, 7946–7958. doi: 10.1021/acs.analchem.6b00770

Walker, J. B., Drew, B. T., and Sytsma, K. J. (2015). Unravelling species relationships and diversification within the iconic California Floristic Province sages (*Salvia* Subgenus *audibertia*, Lamiaceae). *Syst. Bot*. 40, 826–844. doi: 10.1600/036364415X689285

Will, M., and Claßen-Bockhoff, R. (2017). Time to split *Salvia* sl (Lamiaceae)–new insights from Old World *Salvia* phylogeny. *Mol. Phylogenet. Evol.* 109, 33–58. doi: 10.1016/j.ympev.2016.12.041

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, 316–322. doi: 10.1093/nar/gkr483

Zheng, X., Chen, D., Chen, B., Liang, L., Huang, Z., Fan, W., et al. (2021). Insights into salvianolic acid B biosynthesis from chromosome-scale assembly of the *Salvia bowleyana* genome. *J. Integr. Plant Biol.* 63, 1309–1323. doi: 10.1111/jipb.13085

Zhou, Z., Li, Q., Xiao, L., Wang, Y., Feng, J., Bu, Q., et al. (2021). Multiplexed CRISPR/Cas9-mediated knockout of laccase genes in *Salvia miltiorrhiza* revealed their roles in growth, development, and metabolism. *Front. Plant Sci.* 12, 647–768. doi: 10.3389/fpls.2021.647768

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 July 2022
doi: 10.3389/fpls.2022.921970

In planta Female Flower Agroinfiltration Alters the Cannabinoid Composition in Industrial Hemp (*Cannabis sativa* L.)

Michihito Deguchi[1], *Seema Dhir*[2], *Shobha Potlakayala*[1], *Sarwan Dhir*[3], *Wayne R. Curtis*[4] *and Sairam Rudrabhatla*[1]*

[1] The Central Pennsylvania Research and Teaching Laboratory for Biofuels, Penn State Harrisburg, Middletown, PA, United States, [2] Biology Department, Fort Valley State University, Fort Valley, GA, United States, [3] Family Sciences and Technology, College of Agriculture, Fort Valley State University, Fort Valley, GA, United States, [4] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, United States

Industrial hemp is a diploid (2n = 20), dioecious plant, and an essential source of various phytochemical productions. More than 540 phytochemicals have been described, some of which proved helpful in the remedial treatment of human diseases. Therefore, further study of hemp phytochemicals in medicine is highly anticipated. Previously, we developed the vacuum agroinfiltration method, which allows the transient gene expression in hemp tissues including female flowers, where cannabinoids are produced and accumulated. In this study, we attempted to alter the composition of total CBD and THC. The RT-PCR and sanger sequence identified eleven copies of the CBDAS gene, two copies of the THCAS gene, and one CBCAS gene. Binary vectors were constructed to overexpress the CBDAS gene and silence the THCAS gene via RNA interference. The Transcript level of the CBDAS gene was increased by more than 10 times than the plants used as a control, which led to a 54% higher total CBD content. The silencing of the THCAS gene led to downregulation of the THCAS gene, with an 80% reduction in transcript levels, and total THC content was reduced to 43% compared with mock plant. These results suggest that hemp vacuum infiltration is highly effective for metabolic engineering of cannabinoids in hemp.

Keywords: agroinfiltration, cannabidiol, *Cannabis sativa* L., metabolic engineering, tetrahydrocannabinol

OPEN ACCESS

Edited by:
Xueqing Fu,
Shanghai Jiao Tong University, China

Reviewed by:
Yi Ma,
University of Connecticut, United States
Jordi Petit Pedró,
Polytechnic University of Valencia, Spain

***Correspondence:**
Sairam Rudrabhatla
svr11@psu.edu

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *17 April 2022*
Accepted: *22 June 2022*
Published: *21 July 2022*

Citation:
Deguchi M, Dhir S, Potlakayala S, Dhir S, Curtis WR and Rudrabhatla S (2022) In planta Female Flower Agroinfiltration Alters the Cannabinoid Composition in Industrial Hemp (Cannabis sativa L.). Front. Plant Sci. 13:921970. doi: 10.3389/fpls.2022.921970

INTRODUCTION

Hemp (*Cannabis sativa* L.) is an important multipurpose crop with a rich source of fiber, essential fatty acids, easily digestible proteins (albumin and edestin), and enhanced levels of the amino acid arginine. This crop's pharmaceutical compounds have been studied more recently since they show potent bioactivities on human diseases (Schluttenhofer and Yuan, 2017).

Cannabinoids are the secondary metabolites in hemp and are produced in capitate stalked glandular trichomes, located predominantly on female flowers (Livingston et al., 2021). So far, 120 cannabinoids have been identified, among which there is mostly a growing interest in two: Cannabidiol (CBD) and tetrahydrocannabinol (THC), due to their high pharmacological properties (Izzo et al., 2009). The cannabinoid pathway is summarized in **Figure 1**. The pathway is initiated by synthesizing olivetolic acid, which originates from the primary metabolite precursor,

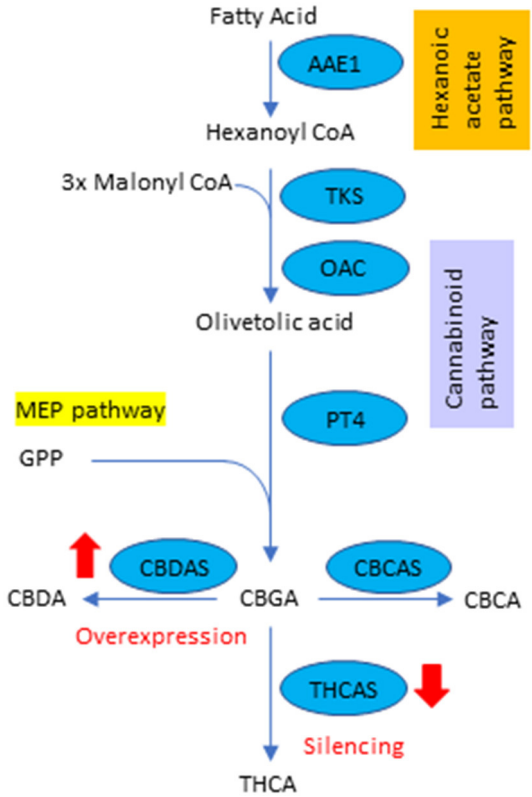


FIGURE 1 | The schematic pathway of phytocannabinoid synthesis and target genes in hemp.

hexanoyl-CoA, by tetraketide synthase (type III polyketide synthase) and olivetolic acid cyclase, in the cytosol (Gülck and Moller, 2020). Olivetolic acid is transported to plastids and converted into cannabigerolic acid (CBGA) by CBGA synthase using geranyl diphosphate (GPP) from the plastidial 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway. Afterward, CBGA is transferred to apoplastic space to synthesize cannabidiolic acid (CBDA), tetrahydrocannabinol acid (THCA), and Cannabichromene acid (CBCA) by CBDA synthase (CBDAS), THCA synthase (THCAS) and CBCA synthase (CBCAS), respectively. Some part of CBDA, THCA and CBCA are decarboxylated to CBD, THC and CBC by non- enzymatic reactions, respectively. Other significant cannabinoids such as cannabinol (CBN) and tetrahydrocannabivarin (THCV) are also present at certain concentrations in female flowers (Tahir et al., 2021).

CBD is the primary cannabinoid in hemp and has shown potential therapeutic agents for several central nervous system diseases such as epilepsy, neurodegenerative diseases, schizophrenia, multiple sclerosis, affective disorders, and the central modulation of feeding behavior (Andre et al., 2016). The first CBD-based product, epidiolex, was recently approved by the U.S. Food and Drug Administration as an anticonvulsant drug (Pauli et al., 2020). On the other hand, hemp also produces low amounts of THC, which is 0.3% on the dry weight basis or less. The THC has therapeutic benefits because it causes relief from nausea and possesses analgesic, neuroprotective, anticancer, anti-inflammatory, and anti-diabetic properties (Hussain et al., 2021). However, THC has been associated with several side effects, such as sensations of euphoria, paranoia and anxiety, cognitive and cholinergic deficits, and immunosuppression (Andre et al., 2016).

Since there is an increased demand for high CBD content and concerns about the presence of a psychoactive agent, we aimed to increase CBD and remove the THC from our hemp cultivar "CRS1" by agroinfiltration-mediated metabolic engineering. Stable transformation is the first choice for plant genetic engineering to add valued traits. However, hemp stable transformation is still a challenge because of its low regeneration efficiency. To combat this, we developed a transient gene expression system in a previous report by vacuum agroinfiltration as an alternative to stable transformation (Deguchi et al., 2020a). Herein, we applied the vacuum agroinfiltration protocol to in planta female flowers of hemp to alter the expression levels of cannabinoid synthase genes and change the cannabinoid contents into desirable profiles.

MATERIALS AND METHODS

Plant Material, Greenhouse Conditions, Generation of Clones, Growth, and Care

The hemp strain, "CRS1," was grown following the approved guidelines for industrial hemp provided by the Pennsylvania Department of Agriculture - Bureau of Plant Industry under the regulated permits IH-16-P-2017 and IH-17-P-2017.

Hemp clones were achieved by collecting a 3-inch segment containing two axillary buds and coating the 45-degree cut with Clonex Rooting gel (Hydrodynamics International, Inc. Lansing, MI). The explant was placed in Root Riot plugs (Hydrodynamics International, Inc. Lansing, MI) and maintained under propagation domes for 2 weeks. Subsequently, they were transferred to four-inch pots containing high porosity soil, HP Mycorrhizae from Pro-Mix (Rivière-du-Loop, Québec, Canada). The clones grew to 15 cm by week four and flowered by week nine. Vegetative cuttings were collected from the same female mother plant to produce genetically identical clones of similar size. The greenhouse conditions were maintained at 25°C with a 14-hour light photoperiod during vegetative growth stage and 12-hours light photoperiod during pre-flowering and flowering stages at 25–40μEm-2s-1.

The humidity for rooting clones was maintained at 65% and decreased gradually to 45% once the clones started to flower. Lost Coast Plant Therapy (Plant Protector, Inc. Loleta, CA) was applied to the clones biweekly at a dilution of 30 mL per 4 liters to control pests.

Cloning, Sequencing, and Phylogenetic Analysis of Cannabinoid Synthase Genes

Total RNA was extracted from 100 mg of each plant sample using the SpectrumTM Plant Total RNA kit (Sigma Aldrich, St. Louis, MO, USA). RNA concentration and absorbance

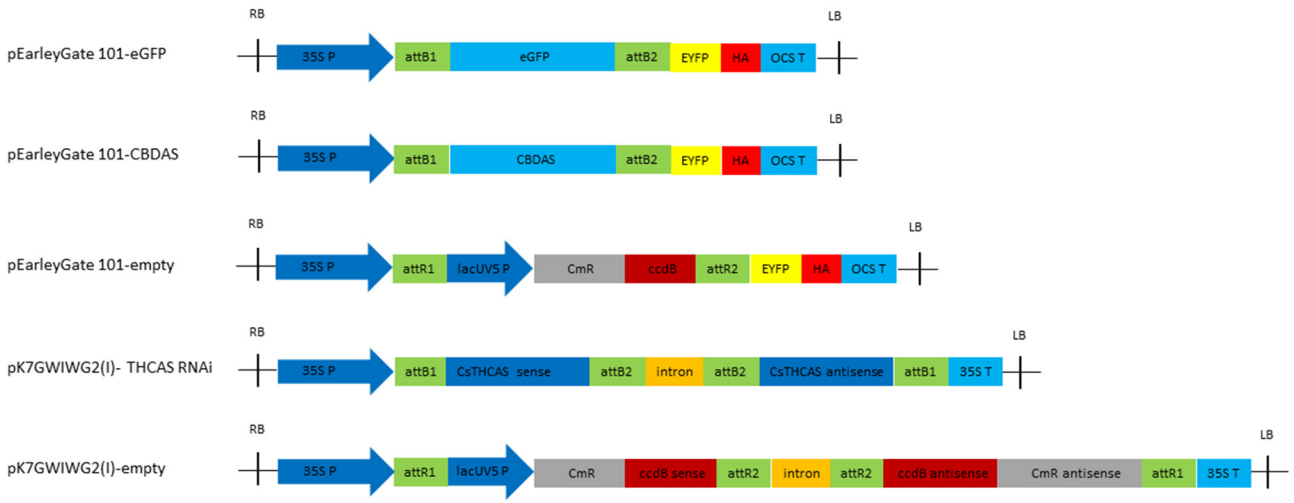


FIGURE 2 | Schematic diagram of the metabolic engineering constructs in this study.

ratios (A260/280 and A260/230) were measured using a NanoVue Plus spectrophotometer (General Electric Healthcare Limited, UK) to measure the quantity and quality of the total RNA. After treatment with DNase I (TaKaRa Bio, Dalian, China) to remove genomic DNA contamination, 2 μg of total RNA was used to synthesize cDNA using the high-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. The cannabinoid synthase genes were amplified with Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA) using mix primers. Forward mix primer sequences were composed of 5'-ATATTTTTCTTTTTCTC-3' and 5'-ATATTTTTCTTTCTCTC-3', and reverse mix primer includes 5'-CTTTGTTCGTTTCTAAA-3' and 5'-CTTTGCTCGTTTCTAAA-3'. The objective size of bands was isolated from agarose gel electrophoresis, purified and cloned into pJET1.2, blunt cloning vector using Clone JET PCR Cloning kit (Fermentas, Burlington, Canada). Those plasmids were sent to Macrogen (Seoul, Korea) for a sanger sequence of insert genes. Phylogenetic analysis was conducted based on the unique CBDAS, THCAS, and CBCAS gene sequences with the Maximum Parsimony method of the MEGA7 software package (Tamura et al., 2011). This software was also used for the alignment of cannabinoid synthase genes.

Construction of Binary Vectors and Agroinfiltration

pEarleyGate 101 vector (Invitrogen, Carlsbad, CA, USA) harboring the eGFP gene and CBDAS gene (AB292682) under the control of a Cauliflower mosaic virus (CaMV) 35S promoter and OCS terminator was used for GFP fluorescence assays and CBDAS overexpression, respectively. An empty pEarleyGate 101 vector was used as a negative control. To silence the THCAS gene, the hairpin RNA-expressing RNAi construct was prepared by inserting sense and antisense partial THCAS fragments into the pK7GWIWG2 gateway vector (Invitrogen). A 209 bp partial sequence (THCAS 992nt−1200nt) was amplified in the sense orientation using cloning primers which include recombination sequences, attB1 and attB2, in forward and reverse primers, respectively, to insert into the entry vector, Gateway pDONR/Zeo/Zeo (Termo Fisher Scientific, Waltham, MA, USA). Targeted gene fragment in the entry clone was then transferred to the binary silencing vector, pK7GWIWG2(I), to form the hairpin RNA, which will be targeted by dicer to synthesize mature small RNA. An empty pK7GWIWG2(I) was used as a negative control. All binary vector cassettes were summarized in **Figure 2**.

The construct was evaluated by sequencing and subsequently transformed into Agrobacterium tumefaciens GV3101 cells. Agrobacterium culture was resuspended in agroinfiltration media containing 10 mM MES, 1x MS, 2% glucose, 0.015% Silwett L-77 (Lehle seeds, Round Rock, USA), 0.05% Pluronic F-68 (Gibco, NY, USA), 5 mM L-Ascorbic acid (Sigma-Aldrich, St. Louis, MO, USA) and 200 μM acetosyringone at a pH 5.6.

Agroinfiltration was performed as previously described (Deguchi et al., 2020a). Hemp female plants produced from the same mother plant were used, and 10–14 days old female flowers in the fourth branch from the top were agroinfiltrated via vacuum infiltration. Three independent female plants were agroinfiltrated per each binary vector. Hemp female plants were placed in a vacuum chamber and female flowers were dipped in *Agrobacterium* solution. The vacuum pump was turned on to decrease the pressure, and the agroinfiltration time was calculated after the vacuum reached 80 mbar. Successful infiltration required bubbles to be flowing up from the hemp female flower. After 10 min at low pressure, the release valve was opened slowly to allow entrance of *Agrobacterium* into the interstitial spaces of the plant tissues. The flower tissues were then washed with distilled water to avoid overgrowth of Agrobacterium, transferred onto a moist filter paper in a petri

dish and placed in a growth room at 21°C. Four days after agroinfiltration, female flowers were harvested and stored in the−80°C freezer until further analysis.

Image Collection, Quantification of Binary Vector, Analysis of Gene Expression and Cannabinoid Content

Image collection and analysis of GFP fluorescence were conducted as previously described (Deguchi et al., 2020a). Four days after vacuum agroinfiltration, the hemp female flowers were subject to GFP fluorescence observation using a Nikon SMZ1500 microscope with Imaging Software NIS Elements F Package (Nikon) equipped with a DS-Ril1 camera (Nikon), an Intensilight C-HGFI Precentered Fiber Illuminator (Nikon) and a GFP2 filter set (Ex. 480 ± 40 nm; Em. 510 nm LP). Agroinfiltration rate was measured by quantifying binary vectors in hemp flowers via qPCR. According to the manufacturer's protocol, DNA was extracted from agroinfiltrated female flowers (100 mg) using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). qPCR was performed with 5 μL of SYBR Select Master Mix (Applied Biosystems, Waltham, MA, USA) in a 10 μL total reaction mixture containing 400 nM of primers and 1 μL of DNA (100 ng/μL). The kanamycin-resistant gene was amplified to quantify the pEarleyGate 101 vector with the following primers (Fw: 5'- ATGGCTAAAATGAGAATATC−3', Rv: 5'- CTAAAACAATTCATCCAGTA−3'), and spectinomycin resistant gene was amplified to quantify pK7GWIWG2(I) vector with following primers (Fw: 5'-ATGGGGGAAGCG GTGATCGC-3', Rv: 5'- TTATTTGCCGACTACCTTGG-3'). To make a calibration curve, Ct values at 5 different concentration series of binary vectors (50 ng/μL, 5 ng/μL, 500 pg/μL, 50 pg/μL, 5 pg/μL, and 500 fg/μL) were measured using Bio-Rad CFX96 system (Bio-Rad, Hercules, CA, USA) under the following reaction conditions: Initial denaturation at 95°C for 10 min, 40 cycles of 95°C for 10 s, and 60°C for 1 min. The copy number of plasmids at each concentration was calculated using the molecular weight and plasmid concentration.

The gene expression of two cannabinoid pathway genes, CBDAS and THCAS was measured using qRT-PCR (Deguchi et al., 2021). Since the CBDAS and THCAS genes have more than 90% homology in their sequences, both sequences were aligned, and a variable region was used to design qRT-PCR primers to avoid off-target effects between them. The primer sequences of CBDAS are Fw: 5'-GATCCGCTGGGCAGAACGGT-3' and Rv: 5'-ATGAGGGAATGGAATTGCTG-3', and the primer sequences of THCAS are Fw: 5'- GATCAGCTGGGAAGAAGACG-3' and Rv: 5'- ATGAGGGAATGGAATTGCTG-3'. The elongation factor 1-α gene was used as a reference gene (Fw: 5'- GCCCTGTCTTTGAGAGCAAC-3', Rv: 5'-CAATCCACTGCTCAATGTGG-3'). Relative gene expression levels of CBDAS and THCAS were calculated using the $2^{-\Delta\Delta Ct}$ method (Deguchi et al., 2021). Statistical analysis was performed using a 1-way ANOVA with Tukey's multiple comparison test ($\alpha = 0.05$). Three independent biological replications were conducted.

Cannabinoid contents were measured by Analytica 360 (Yakima, WA, USA). Collected flower samples were first dried in an convection oven for 6–24 h at 90°C. The 400 mg of dry tissue was extracted with 50% methanol at 60°C for 10 min. The extracts were filtered using a Spin-X column (Corning Inc., Corning, NY, USA). HPLC analysis was performed using a Waters 2695 system coupled to a Waters 3100 single quadrupole mass detector and a 996 photodiode array detector. A reverse-phase Zorbax C-18 column 4.6 mm × 100 mm, 3.5 μ m was eluted by using a mixture (85:15) of methanol and ultrapure water as the mobile phase in an isocratic system with a flow rate of 1.0 ml/min. The standard solutions were prepared to construct calibration curves for the quantification of following cannabinoids: Δ9-THCA, Δ9-THC, Δ8-THC, CBDA, CBD, CBGA, CBG, CBC (Sigma-Aldrich, St. Louis, MO, USA). Peak identification of cannabinoids was performed by comparing the retention times of the samples with those of the standard solutions.

RESULTS

Phylogenetic Analysis and Sequence Alignment of Cannabinoid Synthase Genes

The cannabis genome is highly heterozygous and contains large amounts of repetitive elements. Cannabinoid synthases represented by CBDAS, THCAS, and CBCAS exist in large copy numbers due to gene duplication and divergence (Hurgobin et al., 2021), which requires detailed sequence analysis to select target sequences for the genetic engineering of hemp. We first retrieved functional and non-functional forms of CBDAS, THCAS, and CBCAS from the NCBI database (https://www.ncbi.nlm.nih.gov/) (**Figure 3A**). Next, we designed mix primers at conserved sequence regions adjacent to start and stop codons and carried out RT-PCR to isolate expressed cannabinoid synthase genes from "CRS1." The 1456-nt partial fragments of CBDAS, THCAS and CBCAS were isolated and sequenced. We obtained 11 unique CBDAS copies, two unique THCAS copies, and one unique CBCAS copy (**Figure 3A**). They code 485 amino acids and did not include any stop codon. Notably, non-functional sequences present in the hemp genome were not isolated from any of THCAS, CBDAS and CBCAS, implying that truncated cannabinoid synthase genes are not expressed or expressed at an exceptionally low level. All isolated CBDAS homologs have high homology (>99%) and were identified to contain several single nucleotide polymorphisms (SNPs) leading to different protein translations. CBDAS Kojoma AB292682 which was introduced into yeast for the synthesis of CBDA, was used for the construction of the overexpression vector below (Luo et al., 2019). Two THCAS copies, LC646966 and LC646967, shared high homology (>99%) and possessed five and two SNPs with THCAS Sirikantaramas AB057805, which was used to synthesize THCA in N. tobacco (Sirikantaramas et al., 2005). To find appropriate THCAS fragments for hairpin construction, all isolated CBDAS, THCAS, and CBCAS gene sequences were aligned in MEGA to find variable regions (**Figure 3B**). The 209 nucleotides between 992nd to 1200th from the starting codon

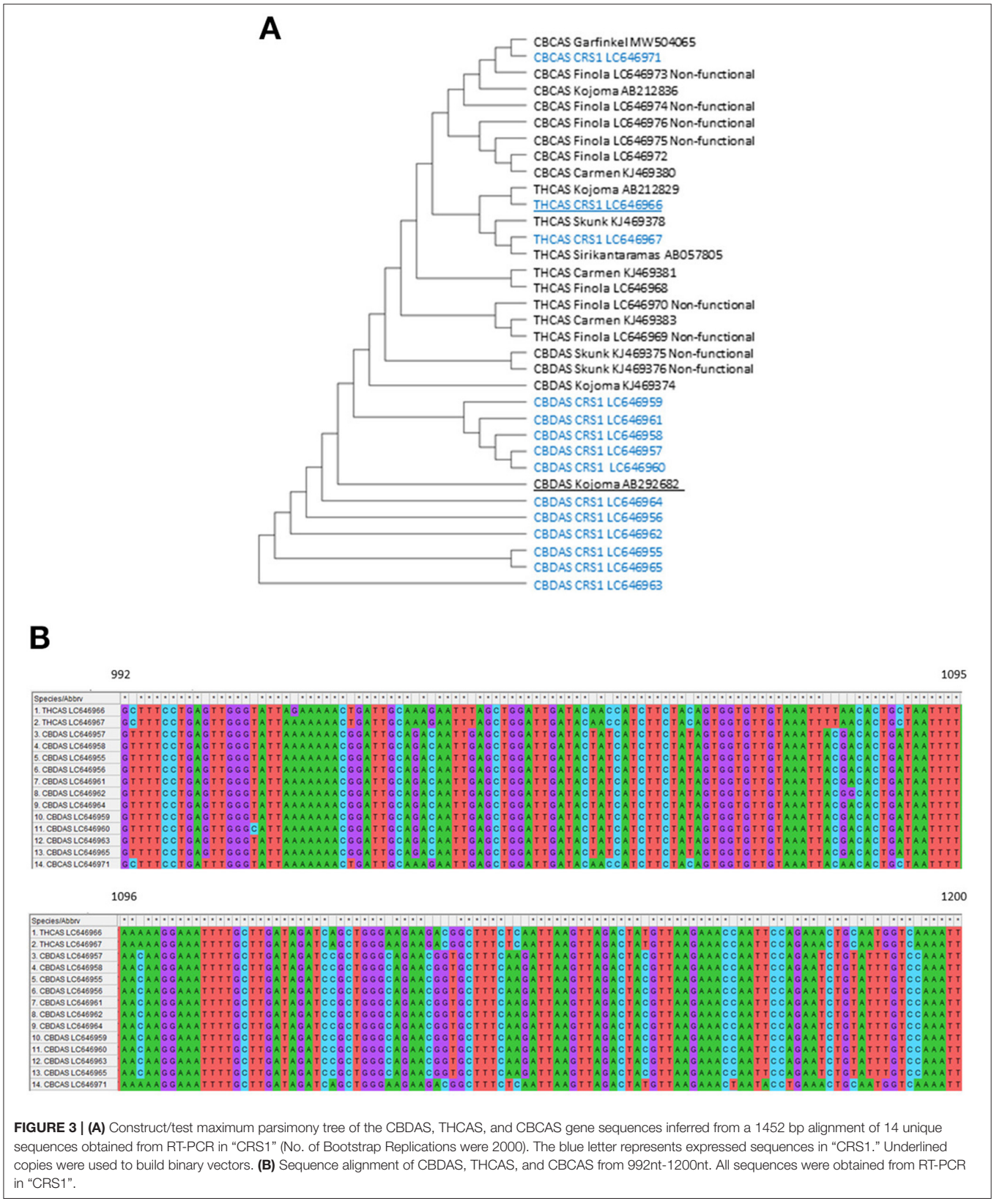


FIGURE 3 | (A) Construct/test maximum parsimony tree of the CBDAS, THCAS, and CBCAS gene sequences inferred from a 1452 bp alignment of 14 unique sequences obtained from RT-PCR in "CRS1" (No. of Bootstrap Replications were 2000). The blue letter represents expressed sequences in "CRS1." Underlined copies were used to build binary vectors. **(B)** Sequence alignment of CBDAS, THCAS, and CBCAS from 992nt-1200nt. All sequences were obtained from RT-PCR in "CRS1".

possess lower homology to CBDAS (83–85%) than other regions. Hence, we used this region to clone a THCAS hairpin cassette.

Female Hemp Flower Vacuum Agroinfiltration

For overexpression of CBDAS, we used full-length cDNA of CBDAS Kojoma AB292682 and cloned into pEarleyGate 101 vector that was driven by CaMV35S promoter and OCS terminator (pEarleyGate 101-CBDAS, **Figure 2**). To silence THCAS, we built pK7GWIWGE(l)-CsTHCAS RNAi, which consists of inverted repeats of THCAS coding sequence separated by an intron. pEarleyGate 101-eGFP was used as a reporter gene (Deguchi et al., 2020a). All binary vectors were transformed into Agrobacterium GV3101 and subjected to vacuum agroinfiltration in mature female flower. Wetting of the flower tissue was observed after agroinfiltration, which indicates successful infiltration (**Figure 4A**). Next, GFP cassette, pEarleyGate 101-eGFP was infiltrated, and GFP fluorescence was observed in whole stalked glandular trichomes including the trichome head (**Figure 4B**).

Absolute Quantification of Binary Vectors in Agroinfiltrated Female Flowers

Afterward, we evaluated the agroinfiltration rate by quantifying the introduced vector units via qPCR (**Figures 5A,B**). In the treatment for CBDAS overexpression, vector unit value (VUV) was high at pEarleyGate 101-CBDAS line 1 (VUV = 1.84), 2 (VUV = 1.75), and 5 (VUV = 1.33), whereas line 3 (VUV = 0.19), 4 (VUV = 0.21) and 6 (VUV = 0.36) showed lower VUV (**Figure 5A**). In mock plants for CBDAS overexpression, VUV was higher at pEarleyGate-empty line 1 (VUV = 1.1), 3 (VUV = 1.4), 5 (VUV = 1.6), and lower at line 2 (VUV = 0.26), 4 (VUV = 0.50) and 6 (VUV = 0.22). For the silencing of THCAS, VUV was higher at pK7GWIWG2(l)-THCAS RNAi line 3 (VUV = 4.58), 4 (VUV = 3.89) and 5 (VUV = 4.88), although line 1 (VUV = 2.19), 2 (VUV = 1.94) and 6 (VUV = 0.93) showed lower VUV (**Figure 5B**). In mock plants for the silencing of THCAS, pK7GWIWG2(l)-empty lines 2 (4.17), 4 (VUV = 3.99), and 6 (VUV = 3.46) showed higher VUV, and line 1 (VUV = 0.55), 3 (VUV = 1.73) and 5 (VUV = 1.66) showed lower VUV. As a result, those 12 lines showing higher agroinfiltration rates were subjected to the following gene expression and cannabinoid content analysis.

Gene Expression Analysis of *CBDAS* and *THCAS*

The overexpression of CBDAS resulted in CBDAS expression being more than ten times higher than that of mock plants (**Figure 6A**). There was no significant difference in THCAS expression between pEarleyGate 101-CBDAS infiltrated plants, and pEarleyGate 101-empty infiltrated plants. On the other hand, pK7GWIWG2(l)-THCAS RNAi caused downregulation of THCAS expression with an 80% reduction (**Figure 6B**). There was no significant difference in CBDAS expression between pK7GWIWG2(l)-THCAS RNAi infiltrated plants and pK7GWIWG2(l)-empty infiltrated plants, demonstrating that there was no off-target effect of hairpin THCAS RNA on the CBDAS expression.

Cannabinoid Contents in Agroinfiltrated Female Flowers

In hemp female flowers, Δ9-THCA, CBDA, CBD, CBGA were detected, and Δ9-THC, Δ8-THC, CBG and CBC were below the limit of quantification.

Overexpression of CBDAS gene resulted in 54% higher total cannabinoid and total CBD (**Figure 7A**). There was no effect of CBDAS overexpression on total THC and CBG. Likewise, silencing of the THCAS gene brought a significant reduction of total THC, which is equivalent to 57% of total THC in mock plants. In contrast, it did not affect total cannabinoid or total CBD (**Figure 7B**). Notably, CBG was also decreased with a 71% reduction compared with the mock plants.

DISCUSSION

There are several successful reports for Cannabis stable transformation (Galán-Ávila et al., 2021; Hesami et al., 2021; Zhang et al., 2021), however, they still show very low efficiency and genotype dependence. More experimental research is required, including optimizing various parameters such as age and type of explant, type and concentration of plant growth regulators, macro and micronutrients, and vitamins (Hesami et al., 2021). As an alternative method to stable transformation, we previously reported that vacuum agroinfiltration is helpful for both overexpression and silencing in excised hemp tissues/organs. Here, applying this agroinfiltration system to the in planta female flowers, the expression of cannabinoid synthase genes was altered, which led to the modification of cannabinoid composition. This work is the first successful metabolic engineering of cannabinoids in Cannabis sativa, demonstrating that our agroinfiltration system enabled gene expression changes at high levels to alter the composition of phytochemicals.

The copy number variations of cannabinoid synthases have been reported in the Cannabis genome (Laverty et al., 2019; McKernan et al., 2020) and are likely to have resulted from genome duplication and tandem rearrangement of long terminal repeat retrotransposons (Grassa et al., 2021; Hurgobin et al., 2021). CBDAS exists in significantly larger copy numbers than THCAS and CBCAS (Weiblen et al., 2015; Vergara et al., 2017). This is consistent with our results that eleven copies of CBDAS were isolated from cDNA libraries synthesized from female flowers whereas only two copies of THCAS and one copy of CBCAS were isolated (**Figure 3A**). Additionally, we found the functional copies of CBDAS to be highly similar (>99% nucleotide identity) among accessions. This might be because intensive breeding practices have been performed to select these desirable cannabinoid synthase loci, which led to them being less polymorphic. Notably, there are many non-functional copies of cannabinoid synthase genes in the Cannabis genome. Nevertheless, no non-functional copy was identified in RT-PCR and sequencing. These results suggest that one or a few copies are predominantly responsible for synthesizing CBDA,

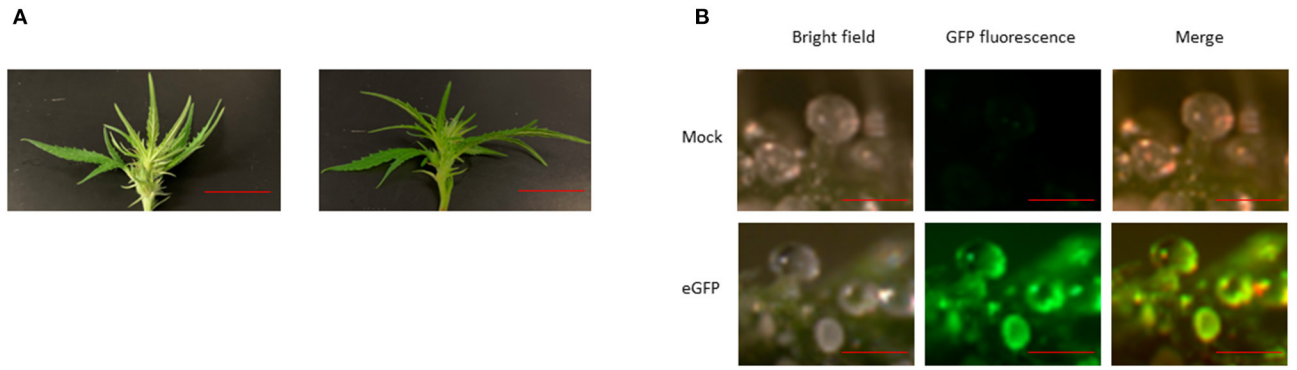


FIGURE 4 | (A) Female hemp flower before (left) and after (right) agroinfiltration. Bar 1.5 cm. **(B)** Transient GFP expression in the female flower. Bar 1.2 mm.

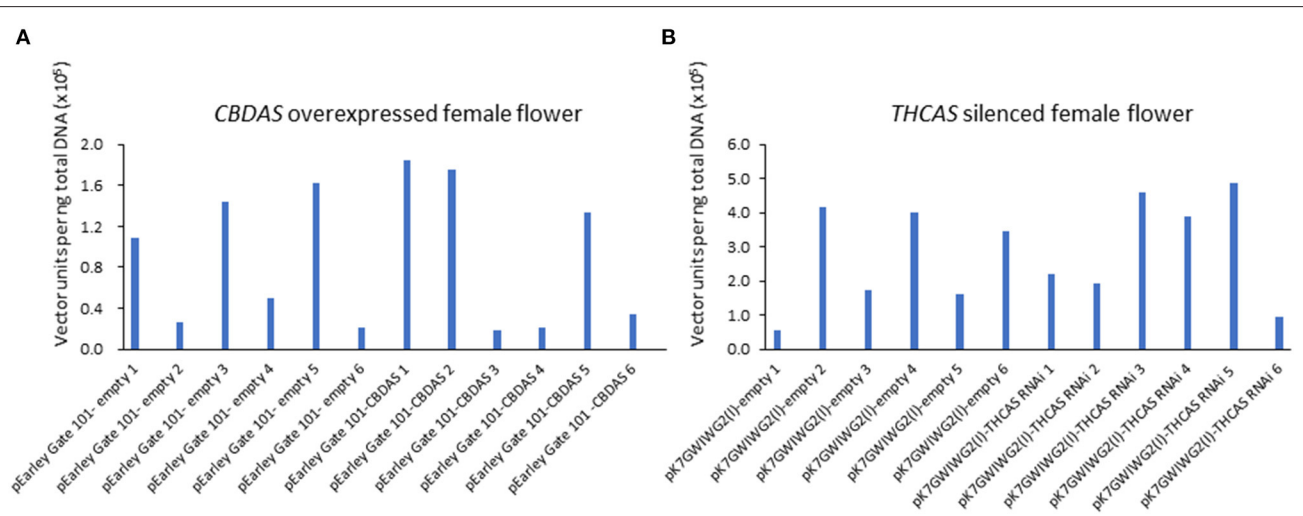


FIGURE 5 | Absolute quantitation of vector units in agroinfiltrated hemp flowers in *CBDAS* overexpressed female flowers **(A)** and THCAS silenced female flowers **(B)**. The vector accumulation was determined by qPCR using a calibration curve.

THCA, and CBCA, as suggested previously (Hurgobin et al., 2021).

In planta flower agroinfiltration has been performed for the study of the flower petal color in ornamental plants (Shang et al., 2007; Hussein et al., 2013; Ratanasut et al., 2014; Fresquet-Corrales et al., 2017; Zeinipour et al., 2018). Unlike previous studies on in planta flower agroinfiltration, we performed agroinfiltration on female hemp flowers because it produces a high amount of cannabinoids in stalked glandular trichomes: a cellular metabolite factory (Tanney et al., 2021), which is abundant on the calyces and bracts in the female flower. The cannabinoids, such as CBDA and THCA, have high toxicity to the plant cell and thus induce death via apoptosis (Andre et al., 2016). To avoid the cannabinoid's toxicity, they are synthesized within secretory disk cells that line the base of the glandular trichome head. Afterward, they are transferred and stored in the secretory cavity, which was developed between secretory disk cells and the cuticle. When the pEarleyGate 101-eGFP vector was expressed via agroinfiltration, GFP fluorescence was detected in both the secretory disk cell and the secretory cavity, suggesting that our agroinfiltration system will help alter the expression of genes responsible for cannabinoid synthesis, transport and storage in stalked glandular trichome (**Figure 4B**).

To evaluate the efficiency of agroinfiltration in each plant, we measured the vector concentration in female flowers using the absolute quantification method. One indicator of successful agroinfiltration is wetting in infiltrated tissues (King et al., 2015). We confirmed that entire female flowers became wet because of the penetration of Agrobacterium solution into the tissues (**Figure 4A**). Nevertheless, infiltration efficiency varied on the individual plant, and the difference between the highest and lowest efficiency of infiltration was as much as 10 times in the same treatment. This clarified that more efforts are necessary to improve the *in planta* female flower agroinfiltration in hemp; If this method would be applied to the high-throughput study of gene function like commonly used model plants for transient expression such as Arabidopsis, Nicotiana benthamiana and Nicotiana tobacco. In previous report, we vigorously studied the effects of surfactant, antioxidant, vacuum time, Agrobacterium, and hemp strain on the efficiency of agroinfiltration (Deguchi

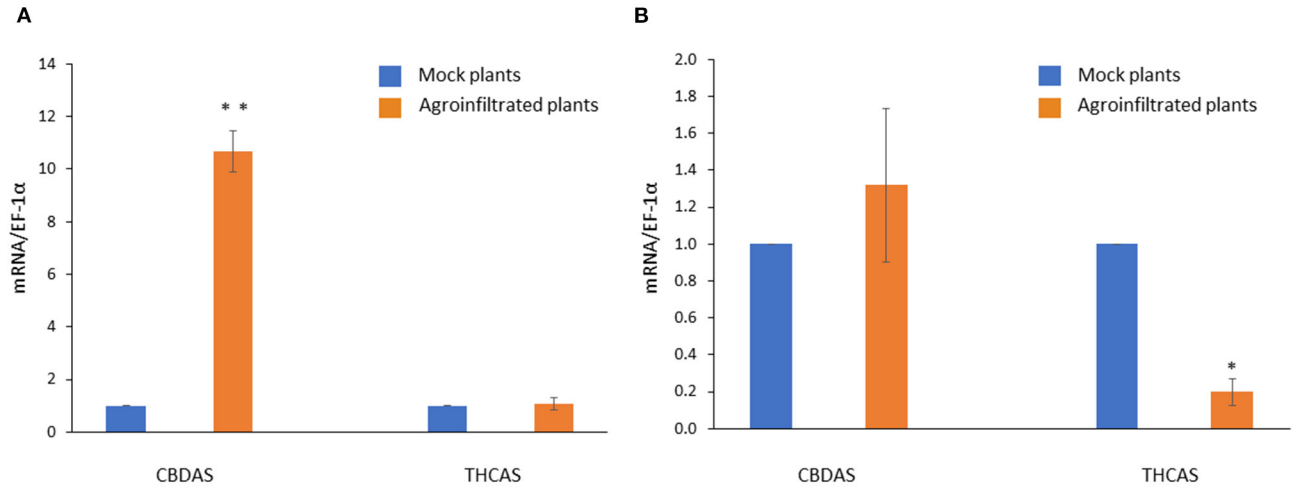


FIGURE 6 | Gene expression profiles of *CBDAS* and *THCAS* in *CBDAS* overexpressed female flowers **(A)** and *THCAS* silenced female flowers **(B)**. Data is presented as mean ± S.E. * and ** indicates statistically significant differences at $P < 0.05$ and $P < 0.01$, respectively, compared with mock plants, according to the statistical analysis in paired *t*-tests.

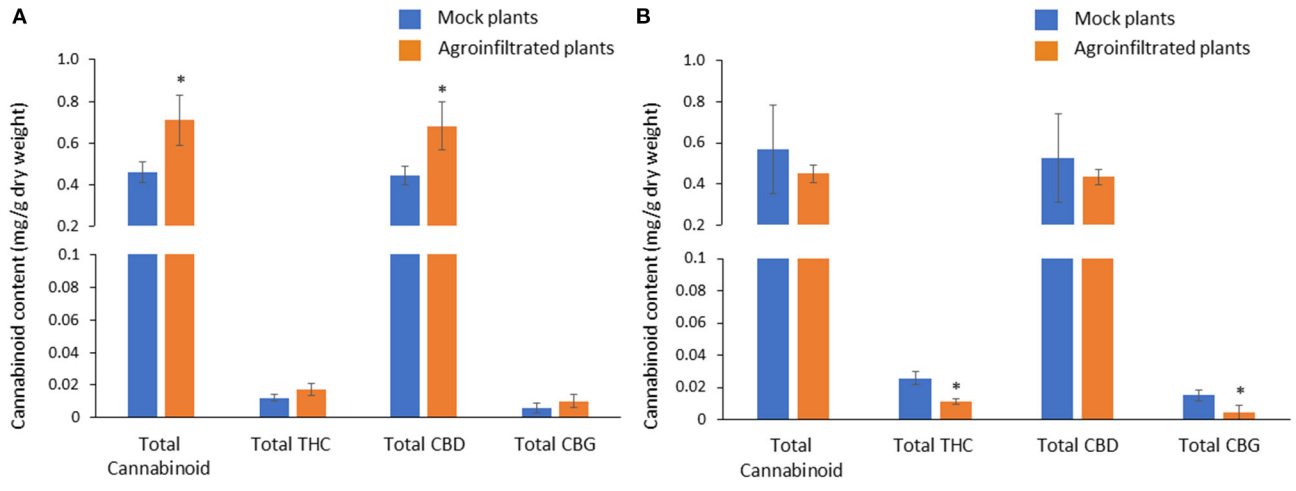


FIGURE 7 | Comparison of cannabinoid contents between mock and agroinfiltrated plants in *CBDAS* overexpressed female flowers **(A)** and *THCAS* silenced female flowers **(B)**. The cannabinoid contents were quantified with HPLC. Data is presented as mean ± S.E. *Indicates statistically significant differences at $P < 0.05$ compared with mock plants, according to the statistical analysis in paired *t*-tests.

et al., 2020a). To further improve this transient gene expression protocol, keeping plants in the dark after infiltration, a suitable developmental stage and temperature for plant growth would need to be optimized (Burman et al., 2020; Zhang et al., 2020).

To date, agroinfiltration has been proven to be effective for metabolic engineering of representative secondary metabolites such as terpenoid (Reed and Osbourn, 2018; Mani et al., 2021), alkaloid (Li et al., 2019; Mora-Vasquez et al., 2022) carotenoid (Rodriguez-Concepcion and Daros, 2022), anthocyanin (Fresquet-Corrales et al., 2017) as well as minor secondary metabolites (van Herpen et al., 2010; Pan et al., 2021), most of which were achieved in N. benthamiana. This plant species shares typical cellular compartmentalization, cofactor, and coenzymes with other plants, enabling it to introduce new synthetic pathways without extensively optimizing the system. Moreover, N. benthamiana produces recombinant proteins at high levels and is amenable to Agrobacterium; thus, it does not cause necrosis. Because of these reasons, N. benthamiana usually becomes the first choice of a platform for heterologous phytochemical production. Gülck et al. (2020) attempted heterologous production of CBDA by overexpression of five cannabinoid pathway genes– AAE1, TKS, OAC, PT, CBDAS– via agroinfiltration in N. benthamiana. Although high expression of these exogenous genes was detected, none of the cannabinoids were synthesized. This is likely due to the different localization of these enzymes; more specifically, olivetolic acid is synthesized by AAE, TKS and OAC at cytosol. Whereas, CBGA synthesis from olivetolic acid and GPP occurred by PT at plastid, and CBDA is

synthesized from CBGA and geranyl diphosphate by CBDAS at apoplast in Cannabis sativa (Hurgobin et al., 2021).

On the other hand, we previously optimized and established hemp agroinfiltration methods (Deguchi et al., 2020a). In this work, this protocol was applied to in planta agroinfiltration on female flowers, leading to a drastic increase in CBDAS expression (**Figure 6A**) and 54% higher CBDA content than control plants (**Figure 7A**). The effect of enhanced CBDAS expression on total CBD content was not as high as expected despite CBDAS expression being highly increased (**Figures 5A**, **6A**). These results suggest that a couple of agroinfiltrations from the earlier flowering stage might be effective for the continuous high CBDAS expression to achieve a higher accumulation of total CBD. Based on metabolite analysis, the precursors for cannabinoid synthesis are not likely to be abundant (**Figure 7A**; Gagne et al., 2012), thus, further supply of precursors from primary metabolite will also be essential. AAE1 is a first step enzyme in the cannabinoid pathway and is responsible for the conversion of fatty acids to hexanoyl-CoA. Stout et al. (2012) revealed that the concentration of hexanoyl-CoA paralleled the accumulation of the CBDA, indicating the AAE1 gene will be a rate-limiting step in CBDA synthesis. Hence, simultaneous overexpression of both CBDAS and AAE1 will lead to a higher accumulation of CBDA than CBDAS overexpression alone.

Moreover, in planta agroinfiltration to female flowers was highly effective to silence the THCAS gene via RNA interference (**Figure 7B**). The copy number of the THCAS gene is likely to be one or two, based on genome sequences which were consistent with the fact that we obtained only two different copies of THCAS (Matchett-Oates et al., 2021a). We used a 209 bp fragment (THCAS 992nt-1200nt) to make a hairpin construct which seemed to result in a non-off target effect on the CBDAS (**Figure 5B**). There is approximately only 85% homology within this region between THCAS copies and CBDAS copies. Interestingly, Matchett-Oates et al. (2021b) reported successful downregulation of CBDAS, THCAS, and CBCAS via RNAi using agroinfiltration. They used a 603 bp fragment of the THCAS gene for siRNA generation. Their RNAi construct did not only show a significant reduction of THCAS with 60% downregulation but also 70% and 40% downregulation of CBDAS and CBCAS, respectively, because of off-target effects. These results imply that smaller RNAi constructed using variable sequence regions might be more effective to avoid or reduce off-target effects on the silencing of cannabinoid synthase genes. Additionally, total CBG content was also decreased in *THCAS* silenced female flower, however, there was no increase in total CBD or total CBC. These results implies that the synthesis of other cannabinoids, for example, CBN, THCV and CBDVA might have been activated by silencing of *THCAS*.

In this work, we successfully increased total CBD content and decreased THC content by the vacuum agroinfiltration method in female hemp flowers. Previous reports proposed that CBDAS is a rate-determining step for CBDA synthesis (Taura et al., 2007; Husain et al., 2019), and our result was consistent with this hypothesis. Although there is high copy number variation on CBDAS gene in the hemp genome sequence, 11 CBDAS copies isolated from female flowers had more than 99% homology, and truncated copies were not detected. Likewise, only two individual THCAS copies (>99% homology) were isolated from female flowers. These results highly encourage us to challenge knocking out the CBDAS gene and the THCAS gene to boost the production of CBGA or eliminate THCA from female flowers, which is desired by the hemp industry. To perform complete and specific downregulation of cannabinoid synthase genes, genome editing technology, particularly CRISPR/Cas9, would be the best choice. This technology enables the knockout of several homologous genes via a single editing step (Jacobs et al., 2017). Recently, several successful cannabis transformations have been reported (Galán-Ávila et al., 2021; Zhang et al., 2021). A combination of CRISPR/Cas technology with hemp regeneration and stable transformation protocols with high efficiency will generate a lot of new strains with added values (Shiels et al., 2022). These hemp plants will be classified into cisgenic plants, which are easier to obtain public acceptance for commercialization (Deguchi et al., 2020b).

CONCLUSION

Here, our vacuum agroinfiltration protocol proved to be very useful for gene overexpression/silencing studies to alter the composition of cannabinoids in female hemp flowers. The protocol will also expedite the function studies on genes responsible for not only secondary metabolite synthesis but also flowering time and sex determination, which are associated with important traits for hemp growth but remain to be unraveled (Petit et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MD contributed to the experimental design, prepared plant materials, analyzed the data, prepared the figures, and wrote the original manuscript of the paper. SR, SP, SeDhir, and SaDhir revised manuscript. SR was responsible for project administration, funding acquisition, resources, and supervision of this study. All authors read and approved the final version of the paper.

FUNDING

We want to acknowledge support from the P.A. Options for Wellness and Penn State Harrisburg School of Science, Engineering, and Technology.

ACKNOWLEDGMENTS

We thank Dr. Taura (Faculty of pharmaceutical sciences in Kyushu university) for providing full length cDNA of the *CBDAS* gene.

REFERENCES

Andre, C. M., Hausman, J. F., and Guerriero, G. (2016). Cannabis sativa: The plant of the thousand and one molecules. *Front. Plant Sci.* 7:19. doi: 10.3389/fpls.2016.00019

Burman, N., Chandran, D., and Khurana, J. P. (2020). A rapid and highly efficient method for transient gene expression in rice plants. *Front. Plant Sci.* 11:584011. doi: 10.3389/fpls.2020.584011

Deguchi, M., Bogush, D., Weeden, H., Spuhler, Z., Potlakayala, S., Kondo, T., et al. (2020a). Establishment and optimization of a hemp (*Cannabis sativa L.*) agroinfiltration system for gene expression and silencing studies. *Sci. Rep.* 10:3504. doi: 10.1038/s41598-020-60323-9

Deguchi, M., Kane, S., Potlakayala, S., George, H., Proano, R., Sheri, V., et al. (2020b). Metabolic engineering strategies of industrial hemp (*Cannabis sativa* L.): A Brief Review of the Advances and Challenges. *Front. Plant Sci.* 11:580621. doi: 10.3389/fpls.2020.580621

Deguchi, M., Potlakayala, S., Spuhler, Z., George, H., Sheri, V., Agili, R., et al. (2021). Selection and validation of reference genes for normalization of qRT-PCR data to study the cannabinoid pathway genes in industrial hemp. *PLoS ONE* 16:e0260660. doi: 10.1371/journal.pone.0260660

Fresquet-Corrales, S., Roque, E., Sarrion-Perdigones, A., Rochina, M., Lopez-Gresa, M. P., Diaz-Mula, H. M., et al. (2017). Metabolic engineering to simultaneously activate anthocyanin and proanthocyanidin biosynthetic pathways in Nicotiana spp. *PLoS ONE* 12:e0184839. doi: 10.1371/journal.pone.0184839

Gagne, S. J., Stout, J. M., Liu, E., Boubakir, Z., Clark, S. M., and Page, J. E. (2012). Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proc. Natl. Acad. Sci. USA.* 109, 12811–12816. doi: 10.1073/pnas.1200330109

Galán-Ávila, A., Gramazio, P., Ron, M., Prohens, J., and Herraiz, F. J. (2021). A novel and rapid method for *Agrobacterium*-mediated production of stably transformed *Cannabis sativa* L. plants. *Indus. Crops Products* 170:e113691. doi: 10.1016/j.indcrop.2021.113691

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., et al. (2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* 230:1665–1679. doi: 10.1111/nph.17243

Gülck, T., Booth, J. K., Carvalho, Â., Khakimov, B., Crocoll, C., Motawia, M. S., et al. (2020). Synthetic biology of cannabinoids and cannabinoid glucosides in nicotiana benthamiana and saccharomyces cerevisiae. *J. Nat. Prod.* 83, 2877–2893. doi: 10.1021/acs.jnatprod.0c00241

Gülck, T., and Moller, B. L. (2020). Phytocannabinoids: Origins and biosynthesis. *Trends Plant Sci.* 25, 985–1004. doi: 10.1016/j.tplants.2020.05.005

Hesami, M., Baiton, A., Alizadeh, M., Pepe, M., Torkamaneh, D., and Jones, A. M. P. (2021). Advances and perspectives in tissue culture and genetic engineering of cannabis. *Int. J. Mol. Sci.* 22:e115671. doi: 10.3390/ijms22115671

Hurgobin, B., Tamiru-Oli, M., Welling, M. T., Doblin, M. S., Bacic, A., Whelan, J., et al. (2021). Recent advances in Cannabis sativa genomics research. *New Phytol.* 230, 73–89. doi: 10.1111/nph.17140

Husain, R., Weeden, H., Bogush, D., Deguchi, M., Soliman, M., Potlakayala, S., et al. (2019). Enhanced tolerance of industrial hemp (*Cannabis sativa* L.) plants on abandoned mine land soil leads to overexpression of cannabinoids. *PLoS ONE* 14, e0221570. doi: 10.1371/journal.pone.0221570

Hussain, T., Jeena, G., Pitakbut, T., Vasilev, N., and Kayser, O. (2021). Cannabis sativa research trends, challenges, and new-age perspectives. *iScience* 24:103391. doi: 10.1016/j.isci.2021.103391

Hussein, G. M., Abu El-Heba, G. A., Abdou, S. M., and Abdallah, N. A. (2013). Optimization of transient gene expression system in Gerbera jemosonii petals. *GM Crops Food* 4, 50–57. doi: 10.4161/gmcr.23925

Izzo, A. A., Borrelli, F., Capasso, R., Di Marzo, V., and Mechoulam, R. (2009). Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb. *Trends Pharmacol. Sci.* 30, 515–527. doi: 10.1016/j.tips.2009.07.006

Jacobs, T. B., Zhang, N., Patel, D., and Martin, G. B. (2017). Generation of a collection of mutant tomato lines using pooled CRISPR Libraries. *Plant Physiol.* 174, 2023–2037. doi: 10.1104/pp.17.00489

King, J. L., Finer, J. J., and McHale, L. K. (2015). Development and optimization of agroinfiltration for soybean. *Plant Cell Rep.* 34, 133–140. doi: 10.1007/s00299-014-1694-4

Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., et al. (2019). A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res.* 29, 146–156. doi: 10.1101/gr.242594.118

Li, J., Mutanda, I., Wang, K., Yang, L., Wang, J., and Wang, Y. (2019). Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in Nicotiana benthamiana. *Nat. Commun.* 10:4850. doi: 10.1038/s41467-019-12879-y

Livingston, S. J., Bae, E. J., Unda, F., Hahn, M. G., Mansfield, S. D., Page, J. E., et al. (2021). Cannabis Glandular Trichome Cell Walls Undergo Remodeling to Store Specialized Metabolites. *Plant Cell Physiol.* 62, 1944–1962. doi: 10.1093/pcp/pcab127

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567, 123–126. doi: 10.1038/s41586-019-0978-9

Mani, V., Park, S., Kim, J. A., Lee, S. I., and Lee, K. (2021). Metabolic perturbation and synthetic biology strategies for plant terpenoid production-an updated overview. *Plants* 10:2179. doi: 10.3390/plants10102179

Matchett-Oates, L., Braich, S., Spangenberg, G. C., Rochfort, S., and Cogan, N. O. I. (2021a). *In silico* analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation. *PLoS ONE* 16:e0257413. doi: 10.1371/journal.pone.0257413

Matchett-Oates, L., Spangenberg, G. C., and Cogan, N. O. I. (2021b). Manipulation of Cannabinoid Biosynthesis via Transient RNAi Expression. *Front. Plant Sci.* 12:773474. doi: 10.3389/fpls.2021.773474

McKernan, K. J., Helbert, Y., Kane, L. T., Ebling, H., Zhang, L., Liu, B., et al. (2020). Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes.

Mora-Vasquez, S., Wells-Abascal, G. G., Espinosa-Leal, C., Cardineau, G. A., and Garcia-Lara, S. (2022). Application of metabolic engineering to enhance the content of alkaloids in medicinal plants. *Metab. Eng. Commun.* 14:e00194. doi: 10.1016/j.mec.2022.e00194

Pan, Z., Bajsa-Hirschel, J., Vaughn, J. N., Rimando, A. M., Baerson, S. R., and Duke, S. O. (2021). In vivo assembly of the sorgoleone biosynthetic pathway and its impact on agroinfiltrated leaves of Nicotiana benthamiana. *New Phytol.* 230, 683–697. doi: 10.1111/nph.17213

Pauli, C. S., Conroy, M., Vanden Heuvel, B. D., and Park, S. H. (2020). Cannabidiol drugs clinical trial outcomes and adverse effects. *Front. Pharmacol.* 11:63. doi: 10.3389/fphar.2020.00063

Petit, J., Salentijn, E. M. J., Paulo, M. J., Denneboom, C., and Trindade, L. M. (2020). Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa L.*): A genome-wide association study. *Front. Plant Sci.* 11:569958. doi: 10.3389/fpls.2020.569958

Ratanasut, K., Monmai, C., and Piluk, P. (2014). Transient hairpin RNAi-induced silencing in floral tissues of Dendrobium Sonia 'Earsakul' by agroinfiltration for rapid assay of flower colour modification. *Plant Cell Tissue Organ Culture* 120, 643–654. doi: 10.1007/s11240-014-0632-z

Reed, J., and Osbourn, A. (2018). Engineering terpenoid production through transient expression in Nicotiana benthamiana. *Plant Cell Rep.* 37, 1431–1441. doi: 10.1007/s00299-018-2296-3

Rodriguez-Concepcion, M., and Daros, J. A. (2022). Transient expression systems to rewire plant carotenoid metabolism. *Curr. Opin. Plant Biol.* 66:102190. doi: 10.1016/j.pbi.2022.102190

Schluttenhofer, C., and Yuan, L. (2017). Challenges towards Revitalizing Hemp: A Multifaceted Crop. *Trends Plant Sci.* 22, 917–929. doi: 10.1016/j.tplants.2017.08.004

Shang, Y., Schwinn, K. E., Bennett, M. J., Hunter, D. A., Waugh, T. L., Pathirana, N. N., et al. (2007). Methods for transient assay of gene function in floral tissues. *Plant Methods* 3:1. doi: 10.1186/1746-4811-3-1

Shiels, D., Prestwich, B. D., Koo, O., Kanchiswamy, C. N., O'Halloran, R., and Badmi, R. (2022). Hemp genome editing-challenges and opportunities. *Front Genome Ed* 4:823486. doi: 10.3389/fgeed.2022.823486

Sirikantaramas, S., Taura, F., Tanaka, Y., Ishikawa, Y., Morimoto, S., and Shoyama, Y. (2005). Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of

the glandular trichomes. *Plant Cell Physiol.* 46, 1578–1582. doi: 10.1093/pcp/pci166

Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W., and Page, J. E. (2012). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. *Plant J.* 71, 353–365. doi: 10.1111/j.1365-313X.2012.04949.x

Tahir, M. N., Shahbazi, F., Rondeau-Gagne, S., and Trant, J. F. (2021). The biosynthesis of the cannabinoids. *J. Cannabis Res.* 3:7. doi: 10.1186/s42238-021-00062-4

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

Tanney, C. A. S., Backer, R., Geitmann, A., and Smith, D. L. (2021). Cannabis glandular trichomes: a cellular metabolite factory. *Front. Plant Sci.* 12:721986. doi: 10.3389/fpls.2021.721986

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., and Morimoto, S. (2007). Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. *FEBS Lett.* 581, 2929–2934. doi: 10.1016/j.febslet.2007.05.043

van Herpen, T. W., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H. J., and Beekwilder, J. (2010). Nicotiana benthamiana as a production platform for artemisinin precursors. *PLoS ONE* 5, e14222. doi: 10.1371/journal.pone.0014222

Vergara, D., Baker, H., Clancy, K., Keepers, K. G., Mendieta, J. P., Pauli, C. S., et al. (2017). Genetic and Genomic Tools for Cannabis sativa. *CRC. Crit. Rev. Plant Sci.* 35, 364–377. doi: 10.1080/07352689.2016.1267496

Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., et al. (2015). Gene duplication and divergence affecting drug content in Cannabis sativa. *New Phytol.* 208, 1241–1250. doi: 10.1111/nph.13562

Zeinipour, M., Azadi, P., Majd, A., Kermani, M. J., Irian, S., Hosseini, S. M., et al. (2018). Agroinfiltration: a rapid and reliable method to select suitable rose cultivars for blue flower production. *Physiol. Mol. Biol. Plants* 24, 503–511. doi: 10.1007/s12298-018-0516-5

Zhang, X., Xu, G., Cheng, C., Lei, L., Sun, J., Xu, Y., et al. (2021). Establishment of an *Agrobacterium*-mediated genetic transformation and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (*Cannabis Sativa* L.). *Plant Biotechnol. J.* 19, 1979–1987. doi: 10.1111/pbi.13611

Zhang, Y., Chen, M., Siemiatkowska, B., Toleco, M. R., Jing, Y., Strotmann, V., et al. (2020). A highly efficient *agrobacterium*-mediated method for transient gene expression and functional studies in multiple plant species. *Plant Commun.* 1:100028. doi: 10.1016/j.xplc.2020.100028

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

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 July 2022
DOI 10.3389/fpls.2022.936244

OPEN ACCESS

EDITED BY
Fangyuan Zhang,
Southwest University, China

REVIEWED BY
Xin Fang,
Kunming Institute of Botany (CAS), China
Lei Zhang,
Second Military Medical University, China

*CORRESPONDENCE
Chanchan Liu
liuchanchan@njucm.edu.cn
Qinan Wu
wuqn@njucm.edu.cn

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

RECEIVED 05 May 2022
ACCEPTED 08 July 2022
PUBLISHED 29 July 2022

CITATION
Zhou P, Dang J, Shi Z, Shao Y, Sang M, Dai S, Yue W, Liu C and Wu Q (2022) Identification and characterization of a novel gene involved in glandular trichome development in *Nepeta tenuifolia*. *Front. Plant Sci.* 13:936244. doi: 10.3389/fpls.2022.936244

Identification and characterization of a novel gene involved in glandular trichome development in *Nepeta tenuifolia*

Peina Zhou[1,2], Jingjie Dang[1,2], Zunrui Shi[1,2], Yongfang Shao[1,2], Mengru Sang[1,2], Shilin Dai[1,2], Wei Yue[1,2], Chanchan Liu[1,2,3]* and Qinan Wu[1,2,3]*

[1]College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, China, [2]Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing, China, [3]National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine, Nanjing, China

Nepeta tenuifolia is a medicinal plant rich in terpenoids and flavonoids with antiviral, immunoregulatory, and anti-inflammatory activities. The peltate glandular trichome (PGT) is a multicellular structure considered to be the primary storage organ for monoterpenes; it may serve as an ideal model for studying cell differentiation and the development of glandular trichomes (GTs). The genes that regulate the development of GTs have not yet been well studied. In this study, we identified NtMIXTA1, a GT development-associated gene from the R2R3 MYB SBG9 family. *NtMIXTA1* overexpression in tobacco resulted in the production of longer and denser GTs. Virus-induced gene silencing of *NtMIXTA1* resulted in lower PGT density, a significant reduction in monoterpene concentration, and the decreased expression of genes related to monoterpene biosynthesis. Comparative transcriptome and widely targeted metabolic analyses revealed that silencing *NtMIXTA1* significantly influenced the expression of genes, and the production of metabolites involved in the biosynthesis of terpenoids, flavonoids, and lipids. This study provides a solid foundation describing a mechanism underlying the regulation of GT development. In addition, this study further deepens our understanding of the regulatory networks involved in GT development and GT development-associated metabolite flux, as well as provides valuable reference data for studying plants with a high medicinal value without genetic transformation.

KEYWORDS

glandular trichome development, *Nepeta tenuifolia*, VIGS, transcriptomic (RNA-seq), metabolome

Introduction

Trichomes are specialized organs that originate from the epidermis and are made up of single or multiple cells. Trichomes are classified into two types based on their potential for secretion: glandular trichomes (GTs) and non-glandular trichomes (NGTs) (Werker, 2000). GTs are regarded as cell biofactories because they can synthesize and accumulate a variety of valuable metabolites, including polysaccharides, terpenoids (such as artemisinin and sclareol), methyl ketones, and acyl sugars (Tissier, 2012; Wang, 2015). These metabolites serve as chemical barriers that protect plants from herbivorous insects and pathogens. They are also widely used as medicines, spices, food additives, and pesticides (Duke et al., 2000). Recent research has focused on the molecular mechanisms underlying trichome formation and development. The specific regulators of these processes have been characterized in *Artemisia annua*, tomato, and cucumber (Chalvin et al., 2020). The members of the R2R3 MYB family, which is the largest MYB subfamily, have diverse functions in regulating secondary metabolism, cell development, responses to multiple stressors, and hormonal signal transport (Stracke et al., 2001; Feller et al., 2011). The R2R3 MYB family is divided into 22 subgroups based on sequence and functional similarities (Dubos et al., 2010). Several studies have implied that members from subgroup 9 of the R2R3 MYB family (R2R3 SBG9) are important for epidermal differentiation, particularly that associated with trichomes (Brockington et al., 2013). They are represented by *AtMYB16*, *AtMYB106*, and *AtMYB17* in *Arabidopsis thaliana, AmMIXTA-LIKE 1*, *AmMIXTA-LIKE 2* in *Antirrhinum majus*, *GhMYB25* in *Gossypium hirsuta*, *AaMIXTA1* in *A. annua,* and *SlMX1* in tomato, which are involved in trichome formation and development (Perez-Rodriguez et al., 2005; Walford et al., 2011; Ewas et al., 2016; Shi et al., 2018).

Nepeta tenuifolia Briq. is a member of the Lamiaceae family. It has an intense aroma and has been widely used as a traditional medicine in Asia (Liu et al., 2018). Due to its excellent antiviral, immunoregulatory, and anti-inflammatory properties, *N. tenuifolia* has been included as a component of the Chinese medicine Jingfang Baidu powder for the treatment of severe acute respiratory syndrome (SARS) and coronavirus disease 2019 (COVID-19) (Zhang et al., 2003; He et al., 2013; Feng Q. et al., 2020). Previous phytochemical studies have reported that, among the numerous compounds found in *N. tenuifolia*, volatile oils, primarily monoterpenes, exhibit the most significant biological activity. *N. tenuifolia* leaves, stems, and spikes contain three forms of GTs: peltate GTs (PGTs), where monoterpenes are specifically synthesized, capitate GTs, and digitiform GTs (Liu et al., 2018). There is a strong correlation between the number of PGTs and the volatile oil content. In addition, gene regulation of monoterpene production in PGTs has been elucidated (Liu et al., 2018, 2021). Nonetheless, the molecular mechanisms underlying the development of PGTs remain unknown.

Previous research has demonstrated that GT initiation may be related to the production of its specialized metabolites. For example, *SlMYC1* in tomatoes regulates the initiation of the type VI secretion system, which also modulates terpene biosynthesis (Xu et al., 2019). In *A. annua*, artemisinin content is proportional to GT density. Repressing the expression of *AaMIXTA1*, a positive regulator of GTs, decreases the number of GSTs, artemisinin content, and cuticle deposition (Shi et al., 2018). GT initiation is also implicated in cuticle biosynthesis; however, the relationship between trichome formation and cutin and wax biosynthesis remains unknown. Intriguingly, in addition to terpenoids, GT initiation also affects flavonoid biosynthesis, even though these two pathways operate independently (Sugimoto et al., 2021). The tomato mutant *odorless-2 (od-2)* exhibits defects in the development and density of GTs, as well as a disruption in the production of terpenes and flavonoids (Kang et al., 2010). *AaTAR2* of *A. annua* positively regulates trichome development and biosynthesis of artemisinin and flavonoids (Zhou et al., 2020). Researches have demonstrated that members of the MYB family play a coordinated metabolic role in flavonoid and terpenoid biosynthesis pathways (Bedon et al., 2010; Zvi et al., 2012). However, the gene regulatory networks involved in GT formation and metabolite variations remain unclear.

Gene characterization and genetic research in plants are typically inseparable from the development of transgenic systems. However, in the field of medicinal plants, few genetic transformation techniques are available, and their establishment requires lengthy timelines, posing challenges to the molecular property research. Recently, virus-induced gene silencing (VIGS) has been widely used for gene functional studies. The method is available and effective for no stable system of genetic transformation (Unver and Budak, 2009; Courdavault et al., 2020; Zang et al., 2021).

In this study, we identified and characterized the gene *NtMIXTA1*, which is a member of the R2R3 MYB SBG9 transcription factor family. This *NtMIXTA1*-overexpressed tobacco showed an increase in GT length and density. Silencing *NtMIXTA1* expression resulted in the decrease in PGT density and menthane monoterpenoids. We also investigated the expression of related genes and variation in the levels of metabolites using the transcriptomic and widely targeted metabolic analysis of *NtMIXTA1*-silenced plants. The findings demonstrated that *NtMIXTA1* may be a positive regulator of PGT initiation and that the biosynthesis of terpenoids, flavonoids, and lipids was significantly affected. Our findings provide valuable information for future research on the initiation of trichome formation and the associated metabolic pathway.

Results

Isolation and characterization of *NtMIXTA1*

To identify the genes belonging to the R2R3 SBG9, 32 sequences were screened on the basis of the family domains and characterized protein sequences of members from the R2R3 SBG9 (from the NCBI) (Brockington et al., 2013). Phylogenetic analysis of the 32 candidate genes revealed that only *Sch000019825* clustered with sequences of the SBG9 R2R3 MYB members. The function of SBG9 R2R3 MYB members in regulating epidermal cell differentiation has been elucidated, especially in trichomes (Baumann et al., 2007) (Figure 1A). The multiple sequence alignment of these sequences revealed that Sch000019825 shared a common region "HxAQWESARLxAExRLxRxS" near the amino (N) terminal and a highly conserved R2R3-MYB domain close to the carboxy (C) terminal, demonstrating that it can be assigned to the R2R3 SBG9 (Zhang et al., 2016; Shi et al., 2018; Qin et al., 2021) (Figure 1B). The functions in trichomes and epidermal development of AmMYBML2, AmMYBML3, PhMYB1, and AaMIXTA1 have been characterized (Baumann et al., 2007; Jaffe et al., 2007; Shi et al., 2018), and Sch000019825 is similar to above proteins. Based on the aforementioned information, *Sch000019825* was named *NtMIXTA1* and was analyzed further. The open reading frame (ORF) of *NtMIXTA1* is 1,134 bp in size, encoding 377 amino acids. Based on ExPASy (https://www.expasy.org/) analysis, the formula of the encoded protein is $C_{1767}H_{2752}N_{518}O_{575}S_{10}$; the theoretical isoelectric point and molecular weight of the protein are 6.17 and 44.773 kDa, respectively. Furthermore, the protein may be unstable, with an instability index of 51.19.

Expression profiles and subcellular localization of *NtMIXTA1*

The transcript levels of *NtMIXTA1* were investigated by RT-qPCR using cDNA from different tissues of root, stem, young leaf, and spike, and different levels of leaf 0–9. The results showed that *NtMIXTA1* exhibited higher expression levels in the young leaves and spikes, where PGTs are abundant. Among the different levels of leaves, *NtMIXTA1* was highly expressed in the youngest leaves (containing shoot leaf 0), and its expression decreased with leaf aging (Figures 2A,B); this was consistent with the PGT density variation (Jiang et al., 2016; Liu et al., 2021). The nuclear localization signal (NLS) of NtMIXTA1 was determined by sequence analysis. The subcellular localization of *NtMIXTA1* was investigated in *N. benthamiana* leaf cells. The GFP was fused to the N-terminus of NtMIXTA1, driven by the *35S* promoter. The GFP fluorescence of NtMIXTA1 was observed in the nucleus of *N. benthamiana* epidermal cells, while the fluorescence of the control was observed in whole cell (Figure 2C), suggesting that the NtMIXTA1 protein was located in the nucleus.

NtMIXTA1 overexpression in tobacco influences GT development

To explore the role of NtMIXTA1 in GT development, *NtMIXTA1* was overexpressed in a model plant, tobacco (*Nicotiana tabacum*). After genetic transformation, we identified 10 transgenic lines, named *NtMIXTA1*-OE1∼10. The expression level of *NtMIXTA1* in these transgenic plants was higher than that in the wild-type (WT) tobacco, with the highest expression being ∼40-fold higher than that in WT tobacco (Supplementary Figure S1A). We chose the lines OE8, OE9, and OE1 and collected their seeds (T1). Phenotypic analysis of leaves from T1 revealed that the *NtMIXTA1*-overexpressed tobacco plants showed a higher GT density and GT length than the WT plants which had a significant change between WT and OE lines with a p-value < 0.05 (Supplementary Figures S1B–E). In the *NtMIXTA1*-overexpressed lines, the density and length of GTs were increased around 1.59- and 1.65-fold compared with the control. These results also showed that NtMIXTA1 affected GT development and regulated GT length.

NtMIXTA1 silencing validates its involvement in PGT formation

The *NtMIXTA1* expression was silenced in *N. tenuifolia* seedlings by VIGS to determine its role in PGT initiation and development. Here, a marker gene, phytoene desaturase (*PDS*), was used to evaluate the gene-silencing efficacy by observing leaf photobleaching (Supplementary Figure S2). Based on the high *PDS*-silencing efficacy and albino leaf phenotype, the leaves of the fourth and fifth nodes were used as the experimental materials for VIGS. A fragment ∼399 bp in size was PCR-amplified using primers specific for *NtMIXTA1* (Supplementary Table S1) as the target sequence, which was inserted into the VIGS vector pTRV2 to generate a new recombinant plasmid for gene silencing. The RT-qPCR analysis of *NtMIXTA1* showed that the expression of *NtMIXTA1* in pTRV*2-NtMIXTA1*-infected leaves was reduced by 64%, compared with that in pTRV2-infected leaves (Figure 3A) at 21 days post-infiltration, indicating that *NtMIXTA1* was effectively silenced in VIGS plants. *NtMIXTA1*-silenced *N. tenuifolia* exhibited a significant reduction in PGT density, whereas the morphology of the PGTs remained unchanged (Figures 3B–E). In detail, *NtMIXTA1*-silenced plants exhibited 49.94% reduction in leaf 4 and 39.00% reduction in leaf 5 in PGT density. These

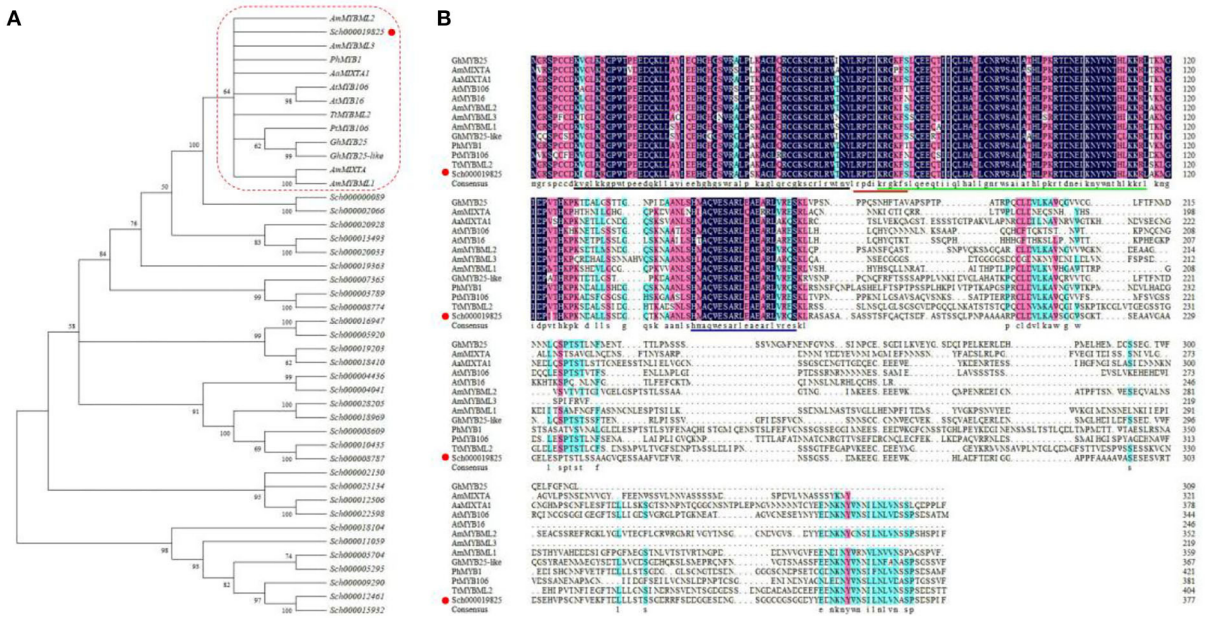


FIGURE 1
Sequence analysis of NtMIXTA1. **(A)** ML (Maximum-likelihood) tree of R2R3 SBG9 proteins. *Arabidopsis thaliana* (At), *Petunia hybrida* (Ph), *Artemisia annua* (Aa), *Gossypium hirsutum* (Gh), *Thalictrum thalictroides* (Tt), *Antirrhinum majus* (Am), and *Populus trichopoda* (Pt). **(B)** Alignment of the protein sequences of NtMIXTA1 and reported R2R3 SBG9 proteins. The R2 and R3 domain characteristics are shown with back and green lines; the conserved domain of subgroup 9 is blue; the red line indicates the putative nuclear localization signal (NLS).

results suggested that the gene may play crucial role in regulating PGT formation.

Monoterpene biosynthesis is linked to PGT formation

PGTs are the predominant source of essential oils, especially menthane monoterpenoids, in *N. tenuifolia* (Liu et al., 2018). Monoterpene biosynthesis begins with the breakdown of geranyl diphosphate (GPP), which is catalyzed by limonene synthase (LS); this yields limonene, which then forms trans-isopiperitenol *via* a reaction catalyzed by L3OH (limonene-3-hydroxylase). Through the catalysis of isopiperitenol dehydrogenase (IPD) and isopiperitenone reductase (IPR), trans-isopiperitenol is converted into pulegone and then transformed into isomenthone or menthone by pulegone reductase (PR) (Figure 4A). Pulegone and limonene are the main ingredients in essential oils; thus, their contents are considered markers of essential oil content (Liu et al., 2018). Using GC analysis, we calculated the relative contents of these two compounds based on the peak areas with correction by referring to an internal standard. The peaks of pulegone and limonene were identified by using standard substances (Supplementary Figure S3). Compared with pTRV2-infected leaves, there was a significant reduction (>50%) in the relative levels of pulegone and limonene in the leaves of plants subjected to VIGS-mediated *NtMIXTA1* silencing (Figure 4B; Supplementary Figure S3; Supplementary Table S2). We inferred that fewer PGTs in *NtMIXTA1*-silenced plants reduced the levels of essential oils. Based on our previous study, the expression of key genes involved in menthane monoterpenoid biosynthesis may also be affected (Liu et al., 2018, 2021). Thus, we measured the expression levels of *LS*, *L3OH*, *IPR*, *IPD*, and *PR* by RT-qPCR. The expression levels of these genes in *NtMIXTA1*-silenced *N. tenuifolia* were substantially lower than those of pTRV2-treated plants (Figure 4C), which was consistent with our inference.

Differential gene expression and differential metabolites in *NtMIXTA1* -silenced plants

To explore the regulatory networks that may underlie PGT formation and the related metabolic variation influenced by *NtMIXTA1*, transcriptomic and widely targeted metabolic analyses (LC-MS and GC-MS) were performed using leaves from TRV2 and VIGS-silenced TRV2-*NtMIXTA1* plants. The PCA analysis of RNA-seq results and metabolites showed good reproducibility for the samples whose biological replicates were closely grouped (Supplementary Figure S4).

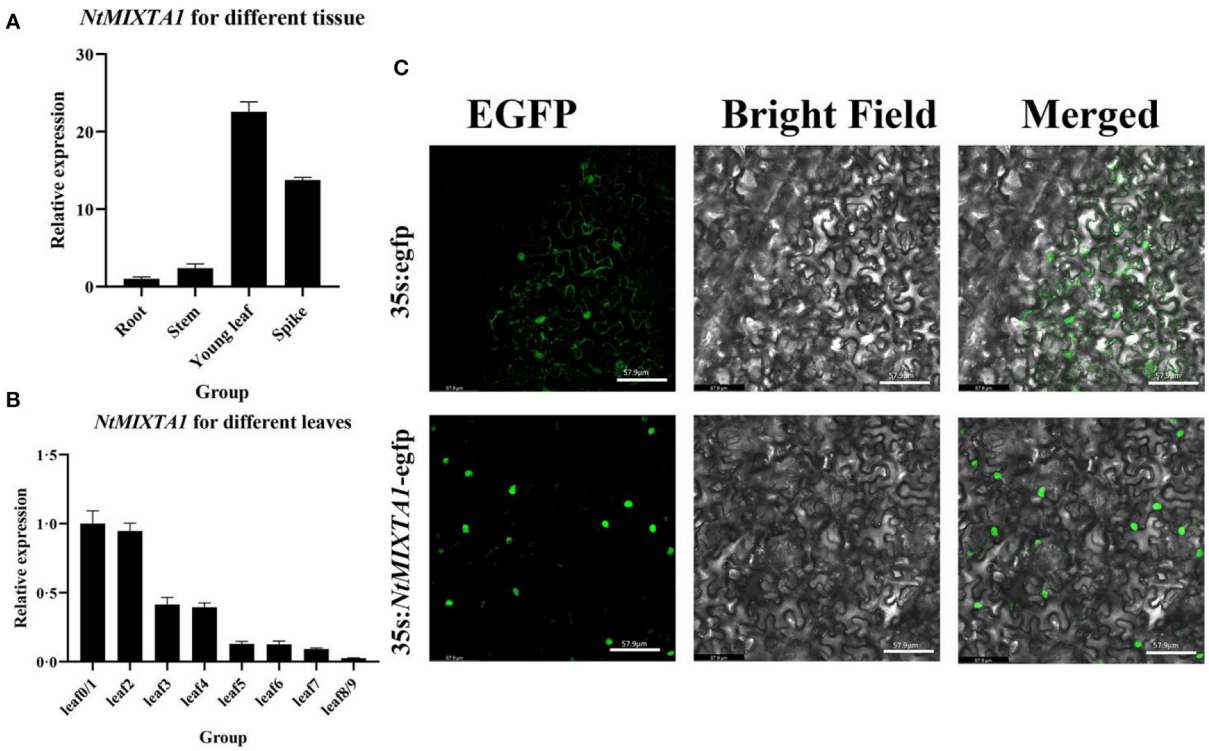


FIGURE 2
Relative expression of *NtMIXTA1* in different tissues **(A)** and leaves at different developmental stages **(B)** with *β-actin* as a reference gene. The relative expression levels were represented as mean ± SD ($n = 3$). **(C)** Subcellular localization of the StMIXTA protein in *Nicotiana benthamiana* epidermal cells. Bars = 57.9 μm.

For metabolic analysis, the VIGS samples were analyzed using widely targeted LC-MS (in both the negative and positive ion modes) and by GC-MS metabolic profiling. All the samples were mixed for quality control (QC) to confirm the repeatability and reliability of the method (Xiao et al., 2021). The total ion chromatograms (TICs) obtained after GC-MS and LC-MS showed that our data were reproducible and reliable (Supplementary Figures S5, S6). The relative contents of the corresponding compounds were calculated from each peak area. There were 1,705 identified metabolites, including 301 terpenoids, 201 flavonoids, 154 phenolic acids, 135 esters, 96 lipids, and others (Supplementary Table S2; Supplementary Figure S7). To identify the differentially expressed metabolites (DEMs) in the TRV2- and *NtMIXTA1*-silenced plants, metabolites were screened using the following criteria: variable importance in projection (VIP) value > 1 and p-value < 0.05 from the OPLS-DA model. Among these metabolites, 406 were downregulated and 183 were upregulated in *NtMIXTA1*-silenced plant, compared with the case in TRV2-silenced plant (Supplementary Table S2; Supplementary Figure S8). Terpenoids, flavonoids, and lipids accounted for 25%, 9%, and 8% of the total metabolites, respectively (Supplementary Figure S8B). These DEMs were mapped to 82 KEGG pathways (Supplementary Table S3).

For the construction of the transcriptome atlas, TRV2 and *NtMIXTA1*-silenced plants (three biological replicates) were analyzed; this yielded 40.98 Gb of clean data with high-quality reads (Q30 > 91%) without an adaptor and an average GC content > 48% for all libraries. After the genome alignment and *de novo* assembly of novel genes, we detected 20,930 genes, including 23,168 transcripts and 3,313 novel genes, with a fragments per kilobase of transcript per million fragments mapped (FPKM) value > 0 (Supplementary Table S4). In the TRV2 and *NtMIXTA1*-silenced plants, 1,523 differentially expressed genes (DEGs) were generated after screening, with 761 upregulated and 762 downregulated genes in *NtMIXTA1*-silenced leaves (Supplementary Table S4; Supplementary Figure S9). The DEGs were assigned to 122 KEGG pathways (Supplementary Table S3). We speculated that these DEGs and DEMs may be influenced by alterations in *NtMIXTA1* expression.

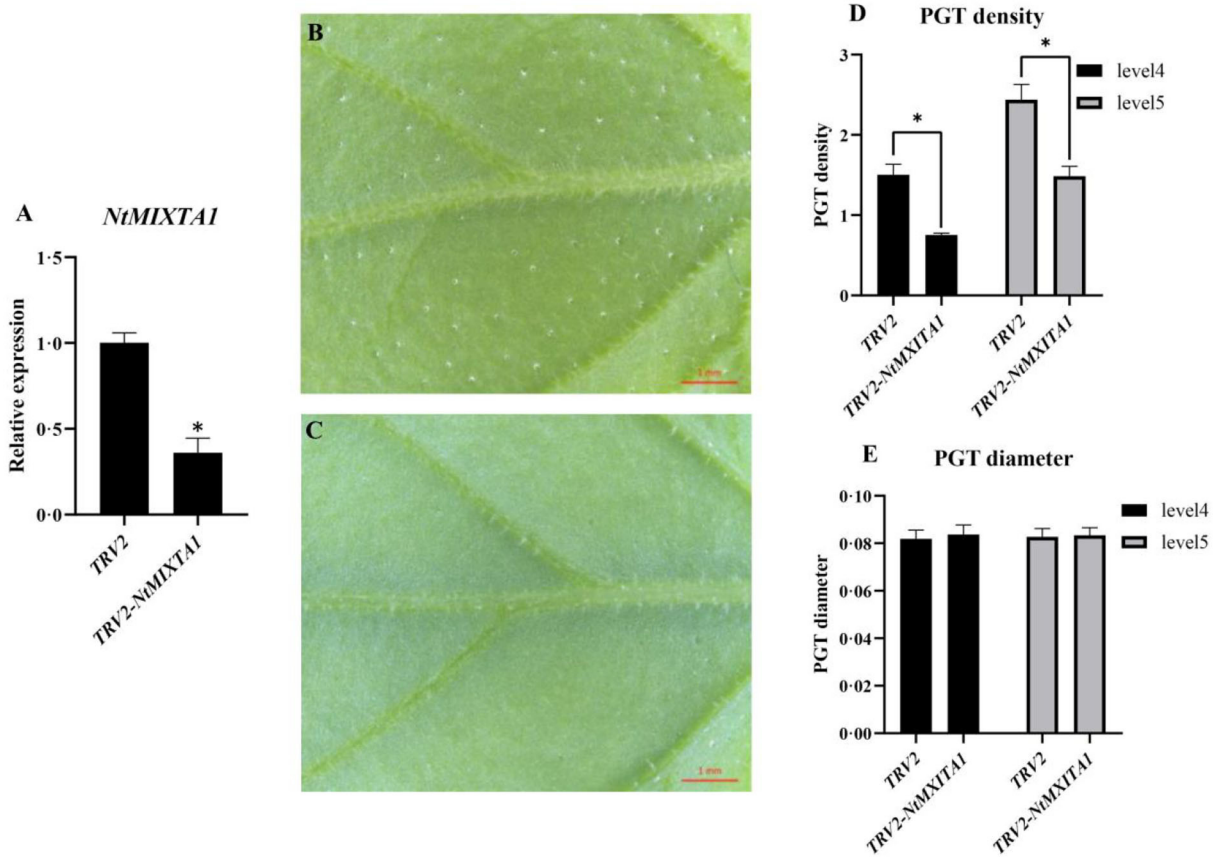


FIGURE 3
Gene expression and phenotype analysis of VIGS plants. **(A)** RT-qPCR analysis of *NtMIXTA1* expression with *β-actin* as a reference gene; the relative expression levels were represented as mean ± SD ($n = 3$); phenotypes of pTRV2 **(B)** and pTRV2-*NtMXITA1* **(C)** showed the presence and decline of PGTs; statistics of PGT density of pTRV2 and pTRV2-*NtMXITA1* **(D)** with p-value < 0.05; statistics of PGT diameter of pTRV2 and pTRV2-*NtMXITA1* **(E)**; phenotypes, the PGT diameter, and density were represented as mean ± SD ($n = 5$). *was represented as p-value < 0.05.

NtMIXTA1 silencing influences the biosynthesis of flavonoids, terpenoids, and lipids

Transcriptome and metabolic profiling data were integrated to explore the influence of gene expression and metabolite variation on the occurrence of lower PGT density phenotypes resulting from *NtMIXTA1* silencing. Among the DEMs, 69% were downregulated in *NtMIXTA1*-silenced plants (Supplementary Figure S8A). The top three DEMs were terpenoids, flavonoids, and lipids (Supplementary Figure S8B), suggesting a close relationship between PGT formation and terpenoid, flavonoid, and lipid biosynthesis. In detail among the DEMs, 146 terpenoids, 55 flavonoids, and 49 lipids were identified (Figures 5A,B) (Supplementary Table S5). Cluster analysis showed that most of the terpenoids (118 downregulated) and lipids (32 downregulated) were detected at low levels in the *NtMIXTA1*-silenced plants (Figure 5C). Flavonoids comprised 26 downregulated and 29 upregulated metabolites (Supplementary Figure S10). The biosynthesis of flavonoids is known to include the following enriched KEGG pathways: biosynthesis of flavonoids (ko00941), flavones and flavonols (ko00944), anthocyanins (ko00942), and isoflavonoids (ko00943). Terpenoid-related biosynthesis includes the following enriched KEGG pathways: biosynthesis of terpenoid backbone (ko00900), monoterpenoids (ko00902), diterpenoids (ko00904), sesquiterpenoids, and triterpenoids (ko00909). Lipid-related biosynthesis includes the following enriched KEGG pathways: fatty acid biosynthesis (ko00061), fatty acid elongation (ko00062), fatty acid degradation (ko00071), cutin, suberin, and wax biosynthesis (ko00073), biosynthesis of unsaturated fatty acids (ko01040), fatty acid metabolism (ko01212), α-linolenic acid metabolism (ko00592), and linoleic acid metabolism (ko00591). Most of the biosynthesis pathways mentioned above were enriched in KEGG analysis of the transcriptome and metabolome data (Figure 5A). Therefore, we investigated the DEGs and DEMs involved in these biosynthetic pathways. In total, 54 DEGs and 30 DEMs were related to

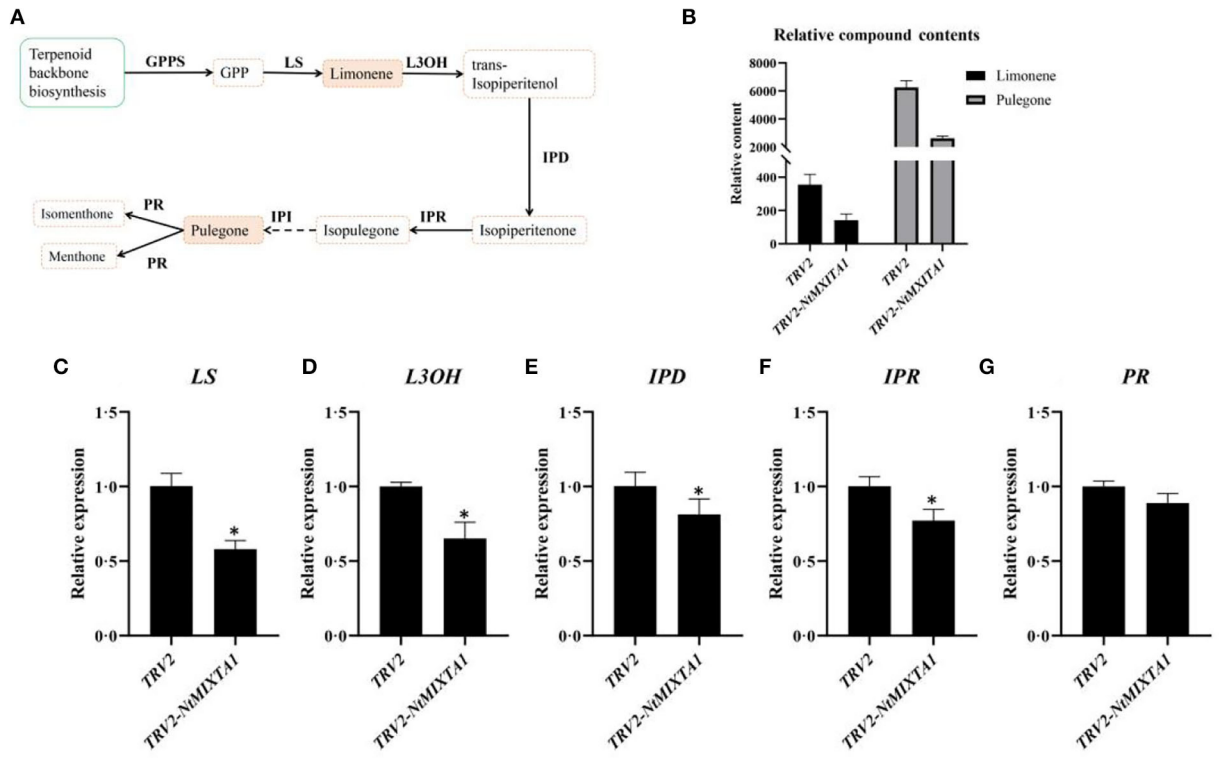


FIGURE 4
Variation of gene expression levels and compound contents of menthane monoterpenes biosynthesis. **(A)** The biosynthesis of menthane monoterpenes; **(B)** the relative contents of pulegone and limonene in TRV2 and in TRV2-*NtMIXTA1*; **(C–G)** the expression levels of *LS, L3OH, IPR, IPD,* and *PR* in TRV2 and in TRV2-*NtMIXTA1*, p-value < 0.05 with β-*actin* as a reference gene ($n = 3$); enzyme abbreviations: GPP, geranyl diphosphate; GPPS, geranyl diphosphate synthase; LS, limonene synthase; L3OH, limonene-3-hydroxylase; IPD, isopiperitenol dehydrogenase; IPR, isopiperitenone reductase; PR, pulegone reductase. *was represented as p-value < 0.05.

terpenoid biosynthesis, 30 genes and 12 metabolites were related to flavonoid biosynthesis, and 27 genes and 19 metabolites were related to lipid biosynthesis (Supplementary Table S5). Combined with the results of the KEGG-mapped DEGs and DEMs, we found that most terpenoid biosynthesis- and lipid biosynthesis-associated genes were notably downregulated in *NtMIXTA1*-silenced leaves; this finding was in accordance with the observed reduction in the contents of related metabolites (Figures 5B,C; Supplementary file 2).

First, we focused on the biosynthesis of menthane monoterpenoids, which produced the main constituents of volatile oils in *N. tenuifolia* (Liu et al., 2018; Bai et al., 2021). Metabolites such as pulegone, limonene, and isopulegone were significantly downregulated in TRV2-*NtMIXTA1*, with similar trends in the expression of genes involved in the menthane monoterpenoid biosynthesis pathways (Figure 6). This finding is in accordance with the results of the GC analysis of the pulegone and limonene contents (Figure 4B) and RT-qPCR analysis of *LS*, *L3OH*, *IPR*, *IPD*, and *PR* expression levels (Figure 4C). The mevalonate (MVA) and 2-methyl-D-erythritol-4-phosphate (MEP) pathways generate the key precursors required for synthesizing several terpenoids: isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Lange and Ahkami, 2013). As expected, the genes associated with these pathways, such as *DXS* and *DXR* (the core genes of MEP), and *HMGS* and *HMGR* (the core genes of MVA), were downregulated in the *NtMIXTA1*-silenced plants (Figure 6A). This may result in the reduced biosynthesis of IPP and DMAPP, leading to lower levels of terpenoids in *NtMIXTA1*-silenced plants (Figure 5B).

Second, trichome development may influence cuticle and wax biosynthesis (Panikashvili et al., 2007; Shi et al., 2018). Genes related to the above biosynthesis such as *CYP77A* (cytochrome P450 family 77 subfamily A) and *HHT1* (omega-hydroxypalmitate O-feruloyl transferase) were downregulated; the levels of cutin-related metabolites, such as hexadecanoic acid and octadecanoic acid, were reduced in *NtMIXTA1*-silenced plants (Supplementary file 2). Fatty acids are precursors of cuticle and wax (Panikashvili et al., 2007), and as a result, the reduction of the fatty acid contents in *NtMIXTA1*-silenced plants may influence cuticle and wax biosynthesis. In addition to the above genes, we also identified genes homologous to

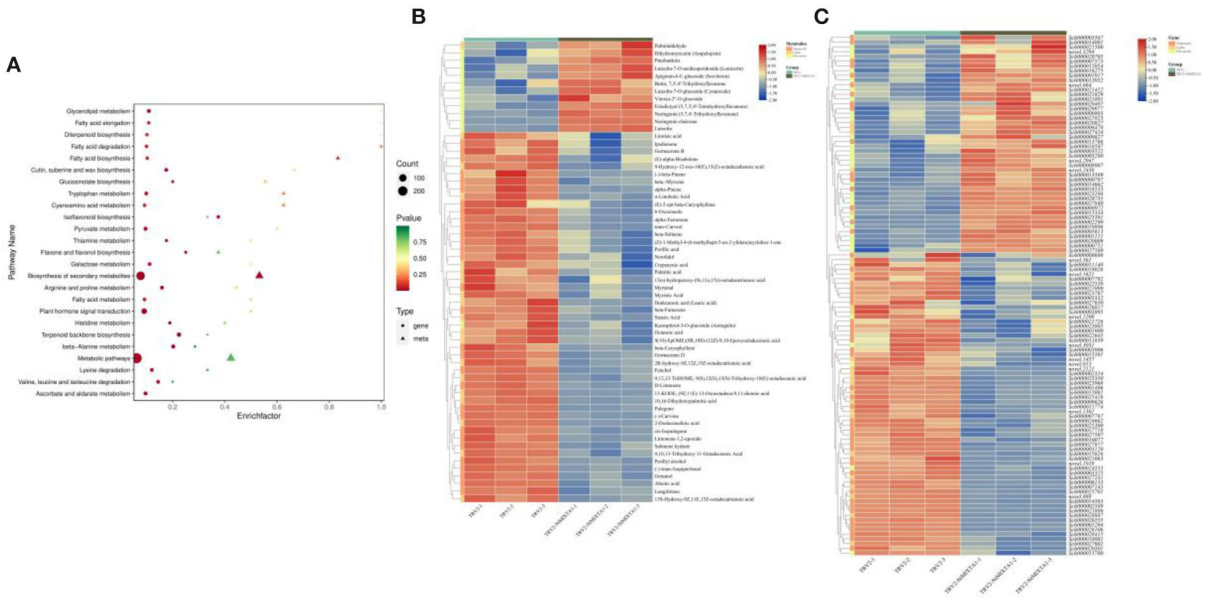


FIGURE 5
KEGG enrichment of top 50 pathways of combined analysis **(A)**. Heatmap of DEMs involving terpenoids, flavonoids, and lipids **(B)** and related gene expression **(C)**.

cutin and wax biosynthesis-related genes, such as *LAS2* (long-chain acyl-CoA synthetase 2), *KCS12* (3-ketoacyl-CoA synthase 12), *CYP77A1* (cytochrome P450 77A1), *CER10* (3-oxo-5-alpha-steroid 4-dehydrogenase family protein), and *KCS10* (3-ketoacyl-CoA synthase 10) (Supplementary Figure S11B); these were mostly downregulated in *NtMIXTA1*-silenced plants. In conclusion, our results demonstrated that the repression of *NtMIXTA1* transcript levels suppressed cuticle biosynthesis in *N. tenuifolia*.

Next, trichome formation is also associated with flavonoid biosynthesis (Kang et al., 2014; Li et al., 2016). According to our metabolic data, flavonoids can be classified into flavones, flavonols, flavanones, flavanonols, flavonoid carbonosides, isoflavones, and chalcones. There was no specific conversion of the different flavonoid structures (Supplementary Figure S12). Interestingly, the metabolites detected in flavonoid biosynthesis (ko00941), including naringenin chalcone, naringenin, eriodictyol, luteolin, and dihydromyricetin (Supplementary Figures S12A,B), were upregulated in *NtMIXTA1*-silenced plants. The genes encoding the enzymes that catalyze the formation of the compounds mentioned above, including *C4H* (cinnamic acid 4-hydroxylase), *F3H* (flavonoid 3-hydroxylase), and *F3'H* (flavonoid 3'-hydroxylase) (Supplementary Figure S12C), were also upregulated in *NtMIXTA1*-silenced plants. Understanding the relationship between flavonoid flux and PGT formation may require further research on the regulatory networks associated with these processes.

Discussion

Analysis of gene function using VIGS

VIGS is a useful tool for investigating gene functions in species where stable genetic transformation is difficult or impossible to achieve (Baulcombe, 1999; Lange et al., 2013; Fei et al., 2021). This method requires less energy and time to execute because it achieves transient transformation (Lu et al., 2003). In addition, VIGS can efficiently identify genes that cannot be examined using other techniques. The method has been widely used in diverse areas of plant research, including evolutionary developmental biology, analyses of secondary metabolism, symbiotic interactions, and plant–pathogen interactions (Dommes et al., 2019). Recently, TRV has been used to determine the functions of genes associated with GT formation and development (Janga et al., 2019; Zang et al., 2021). TRV1 and TRV2 are one set of the most widely used vectors (Ratcliff et al., 2001) that have been used to silence genes in *Nicotiana benthamiana,* tomato, petunia, *Arabidopsis*, cotton, and cucumber (Liu et al., 2004; Fu et al., 2005; Fang et al., 2021). The *PDS* gene is a well-known reporter used in the VIGS system, and *PDS*-silenced plants exhibit the photobleaching phenotypes (Liu and Page, 2008). Both the *GoSPGF* in cotton, and the *CsHOX3* and *CsbHLH1* in *Cucumis sativus* L have been shown to play a role in trichome formation (Janga et al., 2019; Dong et al., 2022).

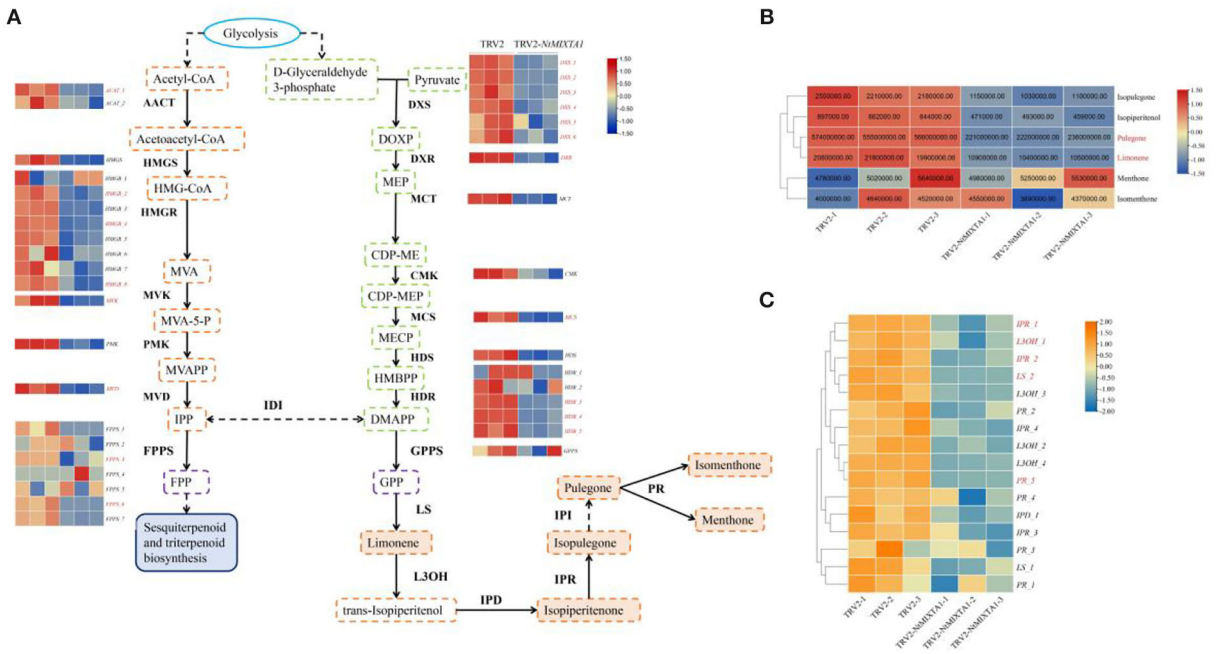


FIGURE 6
Biosynthesis of menthane monoterpene and variations involving related genes and metabolites; **(A)** biosynthetic pathway of menthane monoterpene biosynthesis; **(B)** heatmap of menthane monoterpenes content; **(C)** heatmap of gene expression in menthane monoterpene biosynthesis. Enzyme abbreviations: AACT, acetyl-CoA C-acetyltransferase; HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MVD, diphosphomevalonate decarboxylase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT, 2-C-methyl-D-eryth-ritol 4-phosphate cytidylyltransferase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; IDI, isopentenyl-diphosphate Delta-isomerase; FPPS, farnesyl diphosphate synthase. GPPS, geranyl diphosphate synthase; LS, Limonene synthase; L3OH, limonene-3-hydroxylase; IPD, isopiperitenol dehydrogenase; IPR, isopiperitenone reductase; PR, pulegone reductase. The DEGs and DEMs are shown in red in the heatmap.

In this study, we established a VIGS system in *N. tenuifolia* using *PDS* as a control for VIGS efficiency. *PDS* silencing persisted for ∼40 days and its silencing efficiency was the highest in leaves of the fourth and fifth nodes (from root to shoot), with an obvious albino leaf phenotype (Supplementary Figure S2). These leaves were therefore collected as the plant material for analysis. The silencing effect of VIGS decreased with plant growth. RT-qPCR results demonstrated that *PDS* expression was silenced until the 9th leaf pair (Supplementary Figure S2). There was no difference in leaf phenotypes between *PDS*-silenced and TRV2-infected plants. In addition, for the specificity of VIGS silencing *NtMXITA1*, the similarity of the nucleic acid sequence (399 bp) between VIGS fragments and other genes was examined. The VIGS fragments exhibited a similarity of 69, 40, and 8% to Sch000016802, Sch000000071, and Sch000018969, respectively. Gene count and RT-qPCR revealed that the transcript levels of these three additional genes did not differ between TRV2 and TRV2-NtMXITA1 lines (Supplementary Figure S13). These results indicated that *NtMIXTA1* VIGS had no off-target effects. As a result, based on the VIGS system, *NtMIXTA1* was silenced, leading to a decrease in PGT density and terpenoids levels. Then, RNA-seq and metabolic analysis of *NtMIXTA1*-silenced plants were performed to investigate regulatory networks involved in GT development and GT development-associated genes and metabolites.

NtMIXTA1 positively influences PGT formation and monoterpene biosynthesis

R2R3 MYB SBG9 genes regulate the differentiation of epidermal projections into trichomes and conical cells, including petal epidermis regulation of AmMIXTA, trichome development regulation of *AmMYBML1*, *AmMYBML2*, *AtMYB106*, *AtMYB16*, and Aa*MIXTA1* (Perez-Rodriguez et al., 2005; Baumann et al., 2007; Jakoby et al., 2008; Shi et al., 2018). In this study, 32 candidate genes for the R2R3 MYB SBG9 were identified. According to the ML phylogenetic tree and protein alignment, the gene *NtMIXTA1* was identified and further research was conducted (Figure 1A). We attempted to silence *NtMIXTA1* in *N. tenuifolia* seedlings using VIGS, which resulted

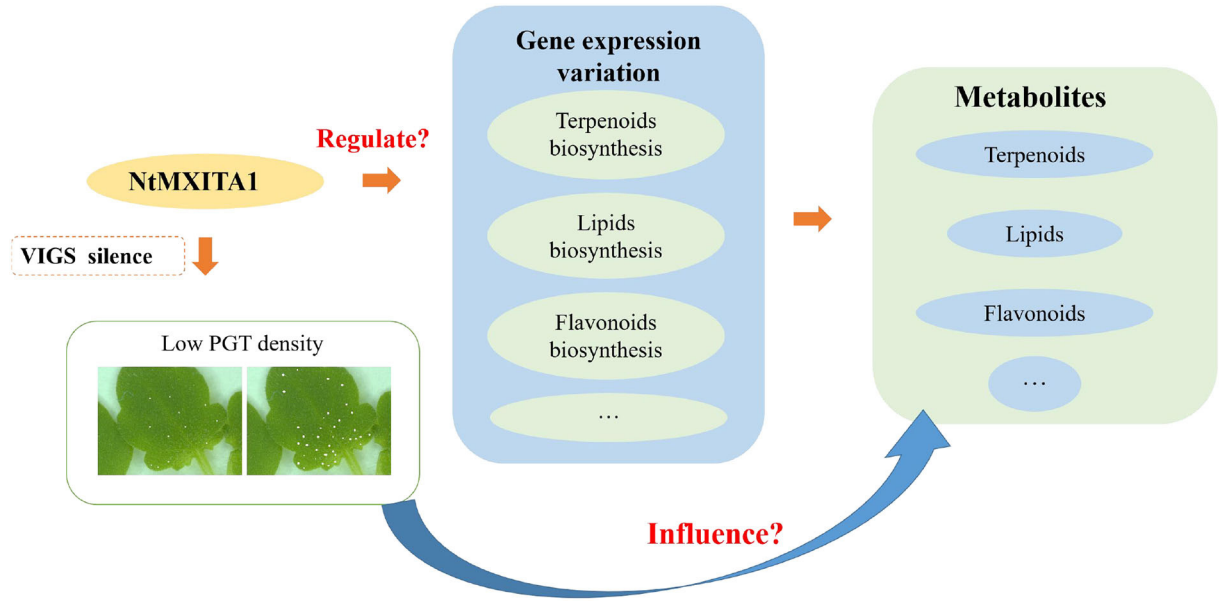


FIGURE 7
Hypothesis model of *N. tenuifolia* PGT initiation and metabolite variation.

in a decreased density of PGTs compared to the controls but did not affect PGTs morphology (Figures 3B–E).

AaHD1 and AaHD8 also promote GT initiation in *A. annua* (Yan et al., 2017; Xie et al., 2021). *HD1* and *HD8* genes in *N. tenuifolia* are homologous to *AaHD1* and *AaHD8*, and their expression was also reduced in *NtMIXTA*-silenced plants (Supplementary Figure S11A). AaHD8 interacts with AaMIXTA1 and promotes the expression of *AaHD1*, which regulates GT initiation and cuticle development, *via* the HD ZIP-MYB complex in *A. annua* (Yan et al., 2018). In tomatoes, a *SlMIXTA*-like gene also cooperates with HD-ZIP IV TFs to regulate cuticle biosynthesis and epidermal cell formation (Lashbrooke et al., 2015). The relationship between NtMIXTA1, HD1, and HD8 may also influence GT development, requiring additional research.

Previous research has demonstrated that menthane monoterpenoid biosynthesis in *N. tenuifolia* occurs mainly in the PGTs (Liu et al., 2021). Metabolic studies of *NtMIXTA1*-silenced plants demonstrated that the production of monoterpene biosynthesis was downregulated in *N. tenuifolia* (Figures 4, 6B,C). RNA-seq analysis was also used to evaluate the expression levels of genes involved in menthane monoterpenoid biosynthesis, including *LS*, *L3OH*, *IPD*, *IPR*, and *PR*; these findings were consistent with the RT-qPCR results showing downregulated (Figures 4C, 6C). In addition, the gene transcription of MVA and MEP, which act upstream of terpenoid biosynthesis, was repressed in *NtMIXTA1*-silenced plants resulting in low terpenoid levels (Figure 6A). Based on the above variation, we inferred that *NtMIXTA1* may affect PGT initiation and the expression of metabolic genes.

A complex network of PGT formation, terpenoids, flavonoids, and lipids

Fatty acids are the primary components of lipids such as cutin and wax. Previously research has shown that the morphology of the cuticle and trichome (both GTs and NGTs) is altered in both trichome and cuticle mutants, implying that the trichome and cuticle are involved in such an interaction (Berhin et al., 2021). For example, the function loss of AtGL1 and Atmyb106 showed a reduction in cutin load and abnormal trichomes (Xia et al., 2010; Oshima et al., 2013); SlMX1, SlWOOLLY, and AaMIXTA all regulate wax biosynthesis and GT initiation (Mohamed et al., 2016; Chang et al., 2018; Shi et al., 2018; Xiong et al., 2020). These findings imply an interaction between trichome and cuticle development. In this study, a similar trend was observed with PGT development and cutin deposition. The expression of genes related to cutin and wax biosynthesis, including *LAS2*, *KCS12*, *CYP77A1*, *CER10*, and *KCS10*, as detected by RNA-seq analysis, and cutin-related metabolites, including fatty acids, was reduced in *NtMIXTA1*-silenced plants (Figure 5C; Supplementary Figure S11B; Supplementary file 2).

In addition to the cuticle and trichome, the impact of flavonoid and terpenoid biosynthetic pathways, and trichome development was recently demonstrated in tomato and *A. annua*

(Kang et al., 2014; Shi et al., 2018; Sugimoto et al., 2021). The *af* mutated tomato, *AaYABBY5*-overexpressed, and *AaTAR2*-overexpressed *A. annua* all affect trichome density, flavonoid, and terpenoid biosynthesis (Kayani et al., 2019, 2021; Zhou et al., 2020; Sugimoto et al., 2021). Our findings demonstrated that terpenoid-related genes and terpenoid metabolites were downregulated in plants with fewer PGTs, compared to controls; however, the variation in flavonoid contents in control plants and plants with fewer PGTs was not significant. The reason may be that the VIGS method involved transient transformation, which resulted in temporary gene silencing. Although the expression of genes involved in flavonoid biosynthesis was upregulated, protein translation may not have been affected at the sampling time. Flavonoids and terpenoids are specifically synthesized and accumulated in the tVI-GTs (type VI GTs) of tomatoes; however, flavonoids have not been detected in the PGTs of *N. tenuifolia*. The aforementioned reasons may lead to differentiation in metabolites levels between our study and previous reports. Our metabolic analysis detected 201 flavonoids in total, 79 of which were flavonoid glycosides generated by glycosyltransferases (UFGTs); therefore, the levels of UFGTs would also be influenced. Any correlation between the metabolic pathways of flavonoids and terpenoids in future research would be valuable.

There were 146 DEGs were annotated as TFs (77 downregulated and 69 upregulated). The top four downregulated TF families were AUX/IAA, MYB, WRKY, and AP2/ERF-ERF. The MYB and WRKY were essential in regulating plant development and secondary metabolism (Ma and Constabel, 2019; Chalvin et al., 2020; Xie et al., 2020; Fu et al., 2021). The AP2/ERF was an important regulator in plant morphogenesis, stress response mechanisms, hormone signal transduction, and metabolite regulation (Feng K. et al., 2020). The AUX/IAA typically affected plant growth and secondary metabolism *via* endogenous hormones (Salehin et al., 2019; Wei et al., 2021). Therefore, the other possibility is that NtMIXTA1 can combine with other TFs to regulate the above biosynthesis and PGT initiation. The following hypothesis model is based on the above results (Figure 7). Silencing of the *NtMIXTA1* resulted in lower PGT density, gene expression variation of terpenoids, lipids, flavonoids, and other biosynthesis, which finally resulted in the changes in the levels of terpenoids, lipids, and flavonoids. The regulated network will be investigated in subsequent studies.

Conclusion

In summary, in this study, the R2R3 MYB SBG9 transcription factor *NtMIXTA1* was identified and characterized. The function of *NtMIXTA1* in GT formation and development was investigated by VIGS and transgenic tobacco. The results suggested NtMIXTA1 positively influenced PGT formation, and changes in the levels of terpenoids, lipids, and flavonoids. The VIGS system was established for *N. tenuifolia* to explore gene function rapidly. This study provides valuable information regarding the relationship between GT initiation and metabolic flux.

Materials and methods

Plant materials

The *N. tenuifolia* plants were grown at the Nanjing University of Chinese Medicine; the seeds were obtained from Hebei Province, China (Liu et al., 2018). The plants were grown in a greenhouse under 10,000 lux intensity and 50% humidity with a light:dark photoperiod of 16:8 h at 25°C. Plant tissues were carefully removed, immediately snap-frozen in liquid nitrogen, and stored at −80°C for RNA extraction. *Nicotiana benthamiana* (K326) seeds were grown in MS (Murashige and Skoog) medium and used in transgene experiments.

Gene expression analysis, phylogenetic tree construction, and amino acid sequence alignment

Genes from the R2R3 MYB SBG9 are known to be the main regulators of epidermal cell development in various tissues, such as in the trichomes (Brockington et al., 2013). Genes belonging to the SBG9 R2R3 MYB were identified from transcriptome data. The gene expression levels were represented by count in RNA-Seq. A phylogenetic tree was constructed using the maximum-likelihood (ML) method using MEGA X software with 1,000 bootstrap-based JTT + G + I amino acid substitution models (Nei and Kumar, 2000; Kumar et al., 2018). Protein sequences of SBG9 R2R3 MYB genes from other species whose functions in epidermal differentiation are clear were downloaded from the National Center for Biotechnology Information (NCBI). Multiple amino acid sequence alignments of *NtMIXTA1* and the other proteins were performed using DNAMAN. The PCR primers used to amplify the *NtMIXTA1* ORF are listed in Supplementary Table S1.

Subcellular localization of *NtMIXTA1*

The ORFs of *NtMIXTA1* were amplified by PCR, and this ORF was cloned into the pNC-cam1304-subN vector with the GFP protein-coding sequence (Yan et al., 2019) using the ClonExpress recombination reaction (Vazyme Biotech, Nanjing, China Nanjing, China Nanjing, China). The recombinant plasmid was then transformed into A*grobacterium tumefaciens* strain GV3101 and injected into *N. benthamiana* leaves

(Sparkes et al., 2006). After administration, GFP signals were observed using a Leica TCS SP8 microscope under excitation and emission wavelengths of 488 and 507 nm, respectively (Leica Microsystems, Wetzlar, Germany); three biological replicates were analyzed for each sample.

Over-expression vector construction and the transformation of WT and transgenic tobacco plants

The *NtMIXTA1* ORF was inserted into the pNC-Cam1304-35S to generate a recombinant overexpression plasmid. The correctness of the recombinant constructs was verified using colony PCR and DNA sequencing. A positive colony was used to extract the correct recombinant plasmid, which was then introduced into EHA105 cells. The overexpression vector transformed into tobacco was used, as described in a previous study (Ma et al., 2019).

RNA was extracted from the transgenic plants, and the universal primer pNC-Cam1304-35S (F: agcggataacaatttcacacagga; R: cgccagggttttcccagtcacgac) was used to select the positive transgenic plants. The expression of *NtMIXTA1* in the transgenic plants was quantified using RT-qPCR; the top three plants were grown in a greenhouse, and transgenic seeds (T1) from these plants were collected. Then, the density and length of the GTs from the transgenic plants T1 were observed and compared with those of the GTs from the WT plants.

VIGS assay

A 399-bp fragment of *NtMIXTA1* was amplified by PCR from cDNA using a 2× Phanta® Max Master Mix (Vazyme Biotech, Nanjing, China Nanjing, China Nanjing, China). The primers used are listed in Supplementary Table S1. The target genes were inserted into pTRV2, which is referred to as pTRV2-*NtMIXTA1*. VIGS was performed as described in a previous study (Zang et al., 2021). In our experiment, cotyledons of 10-day-old seedlings of *N. tenuifolia* were injected. Empty pTRV1 and pTRV2 vectors were injected to represent the controls. The phytoene desaturase *(PDS)* gene was used as a marker for the silencing effect (Yamamoto et al., 2021). The leaves were collected for analysis approximately 21 days after injection.

Microscopy and RT-qPCR analysis

After VIGS administration, leaves of the fourth and fifth nodes (counted from root to shoot) were used to observe PGT density, number, and morphology (Liu et al., 2018). The diameters of the PGTs were also recorded. These experiments were performed using five sample replicates.

The biological replicates were three times selected randomly from five VIGS samples. RNA isolation and cDNA generation for RT-qPCR were performed as described previously (Zhou et al., 2021). RT-qPCR was performed on a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific Inc., Waltham, MA) using the ChamQ Universal SYBR qPCR kit (Vazyme Biotech). The RT-qPCR analysis and calculation of the relative gene expression levels were performed as described in a previous study (Zhou et al., 2021). All the primers used for the RT-qPCR analyses are listed in Supplementary Table S1.

Measurement of the levels of essential oils

Leaves from plants subjected to VIGS were harvested for microscopic analysis. Each sample (0.1 g) was accurately weighed and extracted using n-hexane and camphor as an internal standard (final concentration 30 ng/μL). The mixture was ground and treated with ultrasound three times (60 Hz, 30 s). The extracts were collected in bottles after dehydration. The supernatant was collected after centrifuging the mixture and stored at −80°C. The contents of the volatile components were measured using a GC-FID instrument (7890A, Agilent Technologies, California) using an HP-5 column (30 m × 320 μm × 0.25 μm; Agilent 19091 J-413) with helium as a carrier gas. The detailed conditions for GC-MS and metabolite identification have been described in a previous study (Liu et al., 2018). The peak areas of the compounds were corrected using an internal standard, and the relative contents of each substance were calculated. All the experiments were performed in triplicate.

RNA-seq of the VIGS plants

Total RNA was isolated from the leaves of the plants subjected to VIGS, with three biological replicates sampled for each tissue. Quality control for RNA and RNA-seq library construction was performed as described in a previous study (Liu et al., 2021). After the quality control of the raw data, a total of 40.98 Gb of clean data were generated. The FPKM values were noted for ascertaining gene expression, and DEGs were analyzed using featureCounts v1.6.2 and DESeq2 v1.22.1. A p-value < 0.05 and $|\log2\text{foldchange}| \geq 0.5$ (pTRV2-NtMIXTA1 vs. pTRV2) were used as thresholds for significant difference in expression. Gene functions were annotated using the GO, KEGG, Swiss-Prot, and NCBI-NR (non-redundant) databases.

Metabolite extraction and analysis

Sample extraction of the widely targeted metabolome for LC-MS and GC-MS analyses has been described previously (Chen et al., 2013; Yuan et al., 2022). The detailed methods are listed in Supplementary file 1.

The volatile compounds were identified based on their spectra and retention times (RTs) using commercially available standards or characteristic fragment ions in an independent database constructed based on standard compounds. Specific ion-detection modes were used for accurate scanning, including the scanning typical RTs and qualitative and quantitative ions. As previously described, one quantitative ion and two–three qualitative ions were selected for each compound (Yuan et al., 2022). Quantitative ions were selected to integrate and correct the peak areas for further data processing. The metabolites detected by LC-MS were identified based on a standard compounds database, per standard metabolic operating procedures. Multiple-reaction monitoring (MRM) was used to measure the levels of the metabolites. Correction and calculation of the peak areas of the metabolites were performed as described in a previous study (Fraga et al., 2010). All the samples were analyzed in triplicate. Unsupervised principal component analysis (PCA), hierarchical cluster analysis (HCA), and orthogonal partial least squares discriminant analysis (OPLS-DA) of the identified metabolites were performed using R software. The metabolites were identified as differentially expressed metabolites when the following criteria were fulfilled: variable importance in projection (VIP) value ≥ 1 and p-value < 0.5 (Zeng et al., 2020).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA813002.

Author contributions

PZ, CL, and QW designed the experiments and wrote the manuscript. PZ performed experiments with the help from JD, SD, WY, and ZS. PZ, JD, and ZS analyzed the data. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of China (81903756 for CL, 81973435 and 81473313 for QW), 2017 Chinese Medicine Public Health Service Subsidy Project National General Survey of Traditional Chinese Medicine Resources [Finance Society (2017) No. 66], and 2018 Study on Variety research and quality characteristics of Dao-di herbs produced in Jiangsu Province for QW and Ecological Planting and Quality Assurance Project of Authentic Medicinal Materials (2021) for QW. The Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_1759 and KYCX22_2031) for PZ.

Acknowledgments

We thank Chenghao Fei for the help in identifying compounds and Congling Jia, Juan Shu, Qiujuan Li, Dishuai Li, and Xuemei Cheng for their assistance in the VIGS experiment.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.936244/full#supplementary-material

References

Bai, X., Liu, L., Zhang, J., Chen, L., Wu, T., Aisa, H. A., et al. (2021). Spectrum-effect relationship between GC-QTOF-MS fingerprint and antioxidant, anti-inflammatory activities of *Schizonepeta tenuifolia* essential oil. *Biomed. Chromatogr.* 35, e5106. doi: 10.1002/bmc.5106

Baulcombe, D. C. (1999). Fast forward genetics based on virus-induced gene silencing. *Curr. Opin. Plant Biol.* 2, 109–113. doi: 10.1016/S1369-5266(99)80022-3

Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., et al. (2007). Control of cell and petal morphogenesis by R2R3

MYB transcription factors. *Development* 134, 1691–1701. doi: 10.1242/dev.02836

Bedon, F., Bomal, C., Caron, S., Levasseur, C., Boyle, B., Mansfield, S. D., et al. (2010). Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *J. Exp. Bot.* 61, 3847–3864. doi: 10.1093/jxb/erq196

Berhin, A., Nawrath, C., and Hachez, C. (2021). Subtle interplay between trichome development and cuticle formation in plants. *New Phytol.* 5, 2036–2046. doi: 10.1111/nph.17827

Brockington, S. F., Alvarez-Fernandez, R., Landis, J. B., Alcorn, K., Walker, R. H., Thomas, M. M., et al. (2013). Evolutionary analysis of the MIXTA gene family highlights potential targets for the study of cellular differentiation. *Mol. Biol. Evol.* 30, 526–540. doi: 10.1093/molbev/mss260

Chalvin, C., Drevensek, S., Dron, M., Bendahmane, A., and Boualem, A. (2020). Genetic control of glandular trichome development. *Trends Plant Sci.* 5, 477–487. doi: 10.1016/j.tplants.2019.12.025

Chang, J., Yu, T., Yang, Q., Li, C., Xiong, C., Gao, S., et al. (2018). Hair, encoding a single C2H2 zinc-finger protein, regulates multicellular trichome formation in tomato. *Plant J.* 96, 90–102. doi: 10.1111/tpj.14018

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., and Liu, X. (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

Courdavault, V., Besseau, S., Oudin, A., Papon, N., and O'Connor, S. E. (2020). Virus-induced gene silencing: hush genes to make them talk. *Trends Plant Sci.* 25, 714–715. doi: 10.1007/978-1-0716-0751-0

Dommes, A. B., Gross, T., Herbert, D. B., Kivivirta, K. I., and Becker, A. (2019). Virus-induced gene silencing: empowering genetics in non-model organisms. *J. Exp. Bot.* 70, 757–770. doi: 10.1093/jxb/ery411

Dong, M., Xue, S., Bartholomew, E. S., Zhai, X., Sun, L., Xu, S., et al. (2022). Transcriptomic and functional analysis provides molecular insights into multicellular trichome development. *Plant Physiol.* 1, 301–314. doi: 10.1093/plphys/kiac050

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

Duke, S. O., Canel, C., Rimando, A. M., Telle, M. R., Duke, M. V., and Paul, R. N. (2000). Current and potential exploitation of plant glandular trichome productivity. *Adv. Botan. Res.* 31, 121–151. doi: 10.1016/S0065-2296(00)31008-4

Ewas, M., Gao, Y., Wang, S., Liu, X., Zhang, H., Nishawy, E. M. E., et al. (2016). Manipulation of SlMXl for enhanced carotenoids accumulation and drought resistance in tomato. *Sci. Bull.* 61, 1413–1418. doi: 10.1007/s11434-016-1108-9

Fang, L., Wei, X. Y., Liu, L. Z., Zhou, L. X., Tian, Y. P., Geng, C. (2021). A tobacco ringspot virus-based vector system for gene and microRNA function studies in cucurbits. *Plant Physiol.* 186, 853–864. doi: 10.1093/plphys/kiab146

Fei, Y., Pyott, D. E., and Molnar, A. (2021). Temperature modulates virus-induced transcriptional gene silencing via secondary small RNAs. *New Phytol.* 232, 356–371. doi: 10.1111/nph.17586

Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66, 94–116. doi: 10.1111/j.1365-313X.2010.04459.x

Feng, K., Hou, X. L., Xing, G. M., Liu, J. X., and Duan, A. Q., et al. (2020). Advances in AP2/ERF super-family transcription factors in plant. *Crit. Rev. Biotechnol.* 40, 750–776. doi: 10.1080/07388551.2020.1768509

Feng, Q., Guan, Y., Huang, Z., Ye, S., Cheng, G., Yao, J., et al. (2020). Study on active ingredients of Jingfang Baidu San for preventing COVID-19 based on network pharmacology and molecular docking. *J. Pharm. Pract.* 38, 485–491. doi: 10.12206/j.issn.1006-0111.202005078

Fraga, C. G., Clowers, B. H., Moore, R. J., and Zink, E. M. (2010). Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-mass spectrometry, XCMS, and chemometrics. *Anal. Chem.* 82, 4165–4173. doi: 10.1021/ac1003568

Fu, D. Q., Zhu, B. Z., Zhu, H. L., Jiang, W. B., and Luo, Y. B. (2005). Virus-induced gene silencing in tomato fruit. *Plant J.* 43, 299–308. doi: 10.1111/j.1365-313X.2005.02441.x

Fu, X., Peng, B., Hassani, D., Xie, L., and Liu, H., et al.. (2021). AaWRKY9 contributes to light- and jasmonate-mediated to regulate the biosynthesis of artemisinin in *Artemisia annua*. *New Phytol.* 231, 1858–1874. doi: 10.1111/nph.17453

He, T., Tang, Q., Zeng, N., Gou, L., Liu, J., Yang, J., et al. (2013). Study on effect and mechanism of volatile oil of *Schizonepetae* Herba and its essential components against influenza virus. *China J. Chin. Mater. Med.* 38, 1772–1777. doi: 10.4268/cjcmm20131125

Jaffe, F. W., Tattersall, A., and Glover, B. J. (2007). A truncated MYB transcription factor from *Antirrhinum majus* regulates epidermal cell outgrowth. *J. Exp. Bot.* 58, 1515–1524. doi: 10.1093/jxb/erm020

Jakoby, M. J., Falkenhan, D., Mader, M. T., Brininstool, G., Wischnitzki, E., Platz, N., et al. (2008). Transcriptional profiling of mature *Arabidopsis* trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiol.* 148, 1583–1602. doi: 10.1104/pp.108.126979

Janga, M. R., Pandeya, D., Campbell, L. M., Konganti, K., Villafuerte, S. T., Puckhaber, L., et al. (2019). Genes regulating gland development in the cotton plant. *Plant Biotechnol. J.* 17, 1142–1153. doi: 10.1111/pbi.13044

Jiang, Z., Wang, H., Zhang, X. L., Yue, W., Dai, S. L., and Qi-Nan, W. U. (2016). Research on the Metabolism of *Schizonepetae Spica* Glandular Scale Inclusion. *Plant Sci. J.*

Kang, J. H., Liu, G., Shi, F., Jones, A. D., Beaudry, R. M., and Howe, G. A. (2010). The tomato odorless-2 mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.* 154, 262–272. doi: 10.1104/pp.110.160192

Kang, J. H., McRoberts, J., Shi, F., Moreno, J. E., Jones, A. D., and Howe, G. A. (2014). The flavonoid biosynthetic enzyme chalcone isomerase modulates terpenoid production in glandular trichomes of tomato. *Plant Physiol.* 164, 1161–1174. doi: 10.1104/pp.113.233395

Kayani, S., Shen, Q., Ma, Y., Fu, X., Xie, L., Zhong, Y., et al. (2019). The YABBY family transcription factor AaYABBY5 directly targets cytochrome P450 monooxygenase (CYP71AV1) and double-bond reductase 2 (DBR2) involved in artemisinin biosynthesis in *Artemisia Annua*. *Front. Plant Sci.* 10, 1084. doi: 10.3389/fpls.2019.01084

Kayani, S., Shen, Q., Rahman, S., Fu, X., Li, Y., Wang, C., et al. (2021). Transcriptional regulation of flavonoid biosynthesis in *Artemisia annua* by AaYABBY5. *Hortic. Res.* 8, 257–272. doi: 10.1038/s41438-021-00693-x

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/msy,096

Lange, B. M., and Ahkami, A. (2013). Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes-current status and future opportunities. *Plant Biotechnol. J.* 11, 169–196. doi: 10.1111/pbi.12022

Lange, M., Yellina, A. L., Orashakova, S., and Becker, A. (2013). Virus-induced gene silencing (VIGS) in plants: an overview of target species and the virus-derived vector systems. *Methods Mol. Biol.* 975, 1–14. doi: 10.1007/978-1-62703-278-0_1

Lashbrooke, J., Adato, A., Lotan, O., Alkan, N., Tsimbalist, T., Rechav, K., et al. (2015). The tomato MIXTA-like transcription factor coordinates fruit epidermis conical cell development and cuticular lipid biosynthesis and assembly. *Plant Physiol.* 169, 2553–2571. doi: 10.1104/pp.15.01145

Li, Y., Shan, X., Gao, R., Yang, S., Wang, S., Gao, X., et al. (2016). Two IIIf Clade-bHLHs from *Freesia* hybrida play divergent roles in flavonoid biosynthesis and trichome formation when ectopically expressed in *Arabidopsis*. *Sci. Rep.* 6, 30514. doi: 10.1038/srep30514

Liu, C., Srividya, N., Parrish, A. N., Yue, W., Shan, M., Wu, Q., et al. (2018). Morphology of glandular trichomes of Japanese catnip (Schizonepeta tenuifolia Briquet) and developmental dynamics of their secretory activity. *Phytochemistry* 150, 23–30. doi: 10.1016/j.phytochem.2018.02.018

Liu, E., and Page, J. E. (2008). Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. *Plant Methods* 4, 5. doi: 10.1186/1746-4811-4-5

Liu, L., Yin, M., Lin, G., Wang, Q., Zhou, P., Dai, S., et al. (2021). Integrating RNA-seq with functional expression to analyze the regulation and characterization of genes involved in monoterpenoid biosynthesis in *Nepeta tenuifolia* Briq. *Plant Physiol. Biochem.* 167, 31–41. doi: 10.1016/j.plaphy.2021.07.026

Liu, Y., Nakayama, N., Schiff, M., Litt, A., Irish, V. F., and Dinesh-Kumar, S. P. (2004). Virus induced gene silencing of a DEFICIENS ortholog in *Nicotiana benthamiana*. *Plant Mol. Biol.* 54, 701–711. doi: 10.1023/B:PLAN.0000040899.53378.83

Lu, R., Martin-Hernandez, A. M., Peart, J. R., Malcuit, I., and Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods* 30, 296–303. doi: 10.1016/S1046-2023(03)00037-9

Ma, D., and Constabel, C. P. (2019). MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003

Ma, Y. Q., Li, Q., Pu, Z. Q., Lu, M. X., Yao, J. W., Feng, J. C., et al. (2019). Constitutive expression of NtabSPL6-1 in tobacco and *Arabidopsis* could change the structure of leaves and promote the development of trichomes. *J. Plant Physiol.* 240, 152991. doi: 10.1016/j.jplph.2019.152991

Mohamed, E., Yanqiang, G., Shouchuang, W., Xianqing, L., Hongyan, Z., Elsayed, M. E. N., et al. (2016). Manipulation of SlMXl for enhanced carotenoids accumulation and drought resistance in tomato. *Sci. Bull.* 61, 1413–1418.

Nei, M., and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York, NY: Oxford University Press.

Oshima, Y., Shikata, M., Koyama, T., Ohtsubo, N., Mitsuda, N., and Ohme-Takagi, M. (2013). MIXTA-like transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle development in *Arabidopsis* and *Torenia fournieri*. *Plant Cell* 25, 1609–1624. doi: 10.1105/tpc.113.110783

Panikashvili, D., Savaldi-Goldstein, S., Mandel, T., Yifhar, T., Franke, R. B., Hofer, R., et al. (2007). The *Arabidopsis* DESPERADO/AtWBC11 transporter is required for cutin and wax secretion. *Plant Physiol.* 145, 1345–1360. doi: 10.1104/pp.107.105676

Perez-Rodriguez, M., Jaffe, F. W., Butelli, E., Glover, B. J., and Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132, 359–370. doi: 10.1242/dev.01584

Qin, W., Xie, L., Li, Y., Liu, H., Fu, X., Chen, T., et al. (2021). An R2R3-MYB transcription factor positively regulates the glandular secretory trichome initiation in *Artemisia annua* L. *Front. Plant Sci.* 12, 657156. doi: 10.3389/fpls.2021.657156

Ratcliff, F., Martin-Hernandez, A. M., and Baulcombe, D. C. (2001). Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J.* 25, 237–245. doi: 10.1046/j.0960-7412.2000.00942.x

Salehin, M., Li, B., Tang, M., Katz, E., and Song, L., et al.. (2019). Auxin-sensitive Aux/IAA proteins mediate drought tolerance in *Arabidopsis* by regulating glucosinolate levels. *Nat. Commun.* 10, 4021. doi: 10.1038/s41467-019-12002-1

Shi, P., Fu, X., Shen, Q., Liu, M., Pan, Q., Tang, Y., et al. (2018). The roles of AaMIXTA1 in regulating the initiation of glandular trichomes and cuticle biosynthesis in *Artemisia annua*. *New Phytol.* 217, 261–276. doi: 10.1111/nph.14789

Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* 1, 2019–2025. doi: 10.1038/nprot.2006.286

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi: 10.1016/S1369-5266(00)00199-0

Sugimoto, K., Zager, J. J., Aubin, B. S., Lange, B. M., and Howe, G. A. (2021). Flavonoid deficiency disrupts redox homeostasis and terpenoid biosynthesis in glandular trichomes of tomato. *Plant Physiol.* 3, 1450–1468. doi: 10.1093/plphys/kiab488

Tissier, A. (2012). Glandular trichomes: what comes after expressed sequence tags? *Plant J.* 70, 51–68. doi: 10.1111/j.1365-313X.2012.04913.x

Unver, T., and Budak, H. (2009). Virus-induced gene silencing, a post transcriptional gene silencing method. *Int. J. Plant Genomics* 2009, 198680. doi: 10.1155/2009/198680

Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. (2011). GhMYB25-like: a key regulator of early cotton fiber development. *Plant J.* 65, 789–797. doi: 10.1111/j.1365-313X.2010.04464.x

Wang, G. (2015). Recent progress in secondary metabolism of plant glandular trichomes. *Plant Biotechnol. Nar.* 31, 353–361. doi: 10.5511/plantbiotechnology.14.0701a

Wei, S., Chen, Y., Hou, J., Yang, Y., and Yin, T. (2021). Aux/IAA and ARF gene families in *salix suchowensis*: Identification, evolution, and dynamic transcriptome profiling during the plant growth process. *Front. Plant Sci.* 12, 666310. doi: 10.3389/fpls.2021.666310

Werker, E. (2000). Trichome diversity and development. *Adv. Bot. Res.* 31, 1–35. doi: 10.1016/S0065-2296(00)31005-9

Xia, Y., Yu, K., Navarre, D., Seebold, K., Kachroo, A., and Kachroo, P. (2010). The glabra1 mutation affects cuticle formation and plant responses to microbes. *Plant Physiol.* 154, 833–846. doi: 10.1104/pp.110.161646

Xiao, J., Gu, C., He, S., Zhu, D., Huang, Y., and Zhou, Q. (2021). Widely targeted metabolomics analysis reveals new biomarkers and mechanistic insights on chestnut *(Castanea mollissima Bl.)* calcification process. *Food Res. Int.* 141, 110128. doi: 10.1016/j.foodres.2021.110128

Xie, L., Yan, T., Li, L., Chen, M., Hassani, D., Li, Y., et al. (2021). An HD-ZIP-MYB complex regulates glandular secretory trichome initiation in *Artemisia annua*. *New Phytol.* 5, 2050–2064. doi: 10.1111/nph.17514

Xie, L., Yan, T., Li, L., Chen, M., and Ma, Y., et al.. (2020). The WRKY transcription factor AaGSW2 promotes glandular trichome initiation in *Artemisia annua*. *J. Exp. Bot.* doi: 10.1093/jxb/eraa523

Xiong, C., Xie, Q., Yang, Q., Sun, P., Gao, S., Li, H., et al. (2020). WOOLLY, interacting with MYB transcription factor MYB31, regulates cuticular wax biosynthesis by modulating CER6 expression in tomato. *Plant J.* 103, 323–337. doi: 10.1111/tpj.14733

Xu, J., van Herwijnen, Z. O., Dräger, D. B., Sui, C., Haring, M. A., and Schuurink, R. C. (2019). SlMYC1 regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. *Plant Cell* 30, 2988–3005. doi: 10.1105/tpc.18.00571

Yamamoto, K., Grzech, D., Koudounas, K., Stander, E. A., Caputi, L., Mimura, T., et al. (2021). Improved virus-induced gene silencing allows discovery of a serpentine synthase gene in *Catharanthus roseus*. *Plant Physiol.* 187, 846–857. doi: 10.1093/plphys/kiab285

Yan, P., Zeng, Y., Shen, W., Tuo, D., Li, X., and Zhou, P. (2019). Nimble cloning: a simple, versatile, and efficient system for standardized molecular cloning. *Front. Bioeng. Biotechnol.* 7, 460. doi: 10.3389/fbioe.2019.00460

Yan, T., Chen, M., Shen, Q., Li, L., Fu, X., Pan, Q., et al. (2017). Homeodomain protein 1 is required for jasmonate-mediated glandular trichome initiation in *Artemisia annua*. *New Phytol.* 213, 1145–1155. doi: 10.1111/nph.14205

Yan, T., Li, L., Xie, L., Chen, M., Shen, Q., Pan, Q., et al. (2018). A novel HD-ZIP IV/MIXTA complex promotes glandular trichome initiation and cuticle development in *Artemisia annua*. *New Phytol.* 218, 567–578. doi: 10.1111/nph.15005

Yuan, H., Cao, G., Hou, X., Huang, M., Du, P., Tan, T., et al. (2022). Development of a widely targeted volatilomics method for profiling volatilomes in plants. *Mol. Plant* 15, 189–202. doi: 10.1016/j.molp.2021.09.003

Zang, Y., Xu, C., Xuan, L., Ding, L., Zhu, J., Si, Z., et al. (2021). Identification and characteristics of a novel gland-forming gene in cotton. *Plant J.* 108, 781–792. doi: 10.1111/tpj.15477

Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., et al. (2020). Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. *Mol. Plant* 13, 112–127. doi: 10.1016/j.molp.2019.10.009

Zhang, C., Fan, Y., Lian, J., Zhu, S., Cheng, K., Chai, K., et al. (2003). Exert the advantages of TCM to prevent SARS. *J. Zhejiang Coll. Trad. Chin. Med.* 3, 1–5. doi: 10.3969/j.issn.1005-5509.2003.03.001

Zhang, Y., Wu, J., Wu, Y., Tang, M., and Xie, Y. (2016). Progresses and perspective of the function of MYB transcription factor MIXTA and its orthologous gene. *Sci. Agric. Sin.* 49, 1230–1241. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=ZNYK201607002&DbName=CJFQ2016

Zhou, P., Yin, M., Dai, S., Bao, K., Song, C., Liu, C., et al. (2021). Multi-omics analysis of the bioactive constituents biosynthesis of glandular trichome in *Perilla frutescens*. *BMC Plant Biol.* 21. doi: 10.1186/s12870-021-03069-4

Zhou, Z., Tan, H., Li, Q., Li, Q., Wang, Y., Bu, Q., et al. (2020). Trichome and artemisinin regulator 2 positively regulates trichome development and artemisinin biosynthesis in *Artemisia annua*. *New Phytol.* 228, 932–945. doi: 10.1111/nph.16777

Zvi, M., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S., et al. (2012). PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytol.* 195, 335–345. doi: 10.1111/j.1469-8137.2012.04161.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 August 2022
DOI 10.3389/fpls.2022.967819

OPEN ACCESS

EDITED BY
Fangyuan Zhang,
Southwest University, China

REVIEWED BY
Zhichao Xu,
Northeast Forestry University, China
Jinping Si,
Zhejiang Agriculture and Forestry University, China
Weimin Jiang,
Hengyang Normal University, China

*CORRESPONDENCE
Aizhong Liu
liuaizhong@mail.kib.ac.cn

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

RECEIVED 13 June 2022
ACCEPTED 11 July 2022
PUBLISHED 04 August 2022

CITATION
Wang Y, Xu J and Liu A (2022) Identification of the carotenoid cleavage dioxygenase genes and functional analysis reveal *DoCCD1* is potentially involved in beta-ionone formation in *Dendrobium officinale*. *Front. Plant Sci.* 13:967819. doi: 10.3389/fpls.2022.967819

Identification of the carotenoid cleavage dioxygenase genes and functional analysis reveal *DoCCD1* is potentially involved in beta-ionone formation in *Dendrobium officinale*

Yue Wang[1,2], Jianchu Xu[1] and Aizhong Liu[3]*

[1]Key Laboratory of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [2]Bio-Innovation Center of DR PLANT, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [3]Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming, China

The carotenoids are the most widely distributed secondary metabolites in plants and can be degraded by carotenoid cleavage dioxygenase (CCD) to form apocarotenoids including an important C13 compound beta-ionone. Volatile beta-ionone can confer the violet and woody fragrance to plant essential oils, flowers, fruits, and vegetables, which therefore has been used in various industries. *Dendrobium officinale* is a traditional medicinal plant. However, there was limited information on the key enzymes involved in the biosynthesis of beta-ionone in *D. officinale*. In the present study, beta-ionone was detected in stems and leaves of *D. officinale* and genome-wide identification and expression profiles of *CCD* genes were subsequently carried out. There were nine DoCCD members in *D. officinale*. According to the phylogenetic relationship, DoCCD proteins were classified into six subfamilies including CCD1, CCD4, CCD7, CCD8, nine-cis-epoxycarotenoid dioxygenase (NCED) and zaxinone synthase (ZAS). *DoCCD* genes showed distinctive expression profiles and *DoCCD1* gene was abundantly expressed in eight tissues. Induced expression of *DoCCD1* gene resulted in discoloration of *Escerichia coli* strains that can accumulate carotenoids. Analysis of Gas Chromatography/Mass Spectrometer showed that DoCCD1 enzyme can cleave lycopene to produce 6-methyl-5-hepten-2-one and pseudoionone and also catalyze beta-carotene to form beta-ionone. Expression of *DoCCD1* gene in *Nicotiana benthamiana* leaf resulted in production of

abundant beta-ionone. Overall, the present study first provides valuable information on the CCD gene family in *D. officinale*, function of *DoCCD1* gene as well as production of beta-ionone through genetic modification.

KEYWORDS

***DoCCD1*, beta-ionone, carotenoid cleavage dioxygenase, *Dendrobium officinale*, secondary metabolism, apocarotenoids**

Introduction

Carotenoids are C40 isoprenoids that are the most widely distributed secondary metabolites in plants. Due to abundant double bonds, carotenoids tend to be oxidized through catalytic action by enzymes in plants or induction by ROS, which results in formation of various products that are named as apocarotenoids (Auldridge et al., 2006a; Moreno et al., 2021). Apocarotenoids include important hormones abscisic acid (ABA), strigolactones as well as other small molecules such as ionone, zaxinone, anchorene (Felemban et al., 2019). These apocarotenoids have been shown to play important roles in plant growth, development, response to various stresses and signaling communication (Felemban et al., 2019; Moreno et al., 2021). Among them, volatile compound beta-ionone (C13) has violet and woody notes, which can confer aromas to flowers, fruits, and vegetables (Paparella et al., 2021). Moreover, pharmacological activities of beta-ionone have been confirmed (Aloum et al., 2020; Paparella et al., 2021). With approved security, beta-ionone has been used in the food and fragrance industries (Aloum et al., 2020). Due to its importance and wide application, the biosynthesis of beta-ionone has attracted great attentions.

In plants, carotenoids can be catalyzed by carotenoid cleavage dioxygenases (CCDs) at specific bond sites to produce various apocarotenoids, including aromatic volatile beta-ionone (Simkin et al., 2004b). CCD protein as non-heme iron enzymes is usually characteristics of four conserved histidines and a conserved peptide sequence at the C-terminus (Auldridge et al., 2006a). Based on the substrate difference, the CCD family consisting of multiple members in plants can be classified into CCD and 9-cis-Epoxycarotenoid Dioxygenase (NCED) types (Tan et al., 2003). According to different functions, CCD can be further divided into different subfamilies, such as CCD1, CCD4, CCD2, CCD7, CCD8, and ZAS. In Arabidopsis, there are nine CCD members, including five NCEDs (NCED2, NCED3, NCED5, NCED6, and NCED9) and four CCDs (CCD1, CCD4, CCD7, and CCD8) (Tan et al., 2003). Classification of CCDs in other plants is mainly based on the study in Arabidopsis and some CCD members with new functions were found in other plants such as CCD2 in crocus and *Freesia hybrida* (Ahrazem et al., 2016a; Fang et al., 2020), ZAS in rice (Wang et al., 2019).

Although CCD enzymes can cleave carotenoids, the CCD enzymes from different subfamilies show great differences in substrates and cleavage sites, and therefore result in function diversification (Auldridge et al., 2006a). The first identified CCD was Viviparous 14 in maize (Schwartz et al., 1997). Viviparous 14 and its homologs in Arabidopsis NCEDs can cleave 9-cis-violaxanthin or 9-cis-neoxanthin at 11, 12 (11′, 12′) sites and directly participate in the biosynthesis of ABA (Schwartz et al., 1997; Tan et al., 2003). CCD2 was found in limited plants such as crocus and *Freesia hybrida* and can degrade zeaxanthin at 7, 8, or 7′, 8′ sites for the biosynthesis of safranal and crocin (Frusciante et al., 2014; Ahrazem et al., 2016a; Fang et al., 2020). CCD7 and CCD8 are mainly related to the formation of strigolactones (Alder et al., 2012). ZAS can produce zaxinone with 3-OH-β-apo-10′-carotenal as the substrate, and control the contents of strigolactones in Arabidopsis and rice (Wang et al., 2019; Ablazov et al., 2020).

The functions of CCD1 and CCD4 have been deeply investigated and their substrates are mainly carotenoids. However, the cytoplasmic CCD1 has more abundant substrates, including zeaxanthin, lutein, etc. (Auldridge et al., 2006a; Vogel et al., 2008; Ahrazem et al., 2016b; Zheng et al., 2021; **Figure 1**). The plastid-located CCD4 has higher substrate specificity (Us-Camas et al., 2022). CCD1 and CCD4 cleave carotenoids to increase the content of apocarotenoids and therefore have important influences on plant quality, such as the carotenoid contents in crops (Ko et al., 2018; Thakur et al., 2021), the color and flavors in vegetables and fruits (Campbell et al., 2010; Fukamatsu et al., 2013; Ilg et al., 2014), the flower color and fragrance (Ohmiya et al., 2006; Huang et al., 2009; Chiou et al., 2010; Zhang et al., 2015; Gao et al., 2021) and tea quality (Wang et al., 2020a,b). CCD1 in tomato was responsible for the formation of volatile isoprenes including beta-ionone, pseudoionone, and geranylacetone (Simkin et al., 2004a). CCD1 and CCD4 in *Bixa orellana* were recently found to be involved in the bixin biosynthesis (Us-Camas et al., 2022). Moreover, CCD1 have great effects on the emission of beta-ionone. PhCCD1 can control the emission of beta-ionone in petunia flower (Simkin et al., 2004b) and the similar phenomena were also found in other plants such as *Osmanthus fragrans* (Baldermann et al., 2010), tomato (Simkin et al., 2004a), tea (Wang et al., 2020a), Brassica species (Zhang et al., 2015).

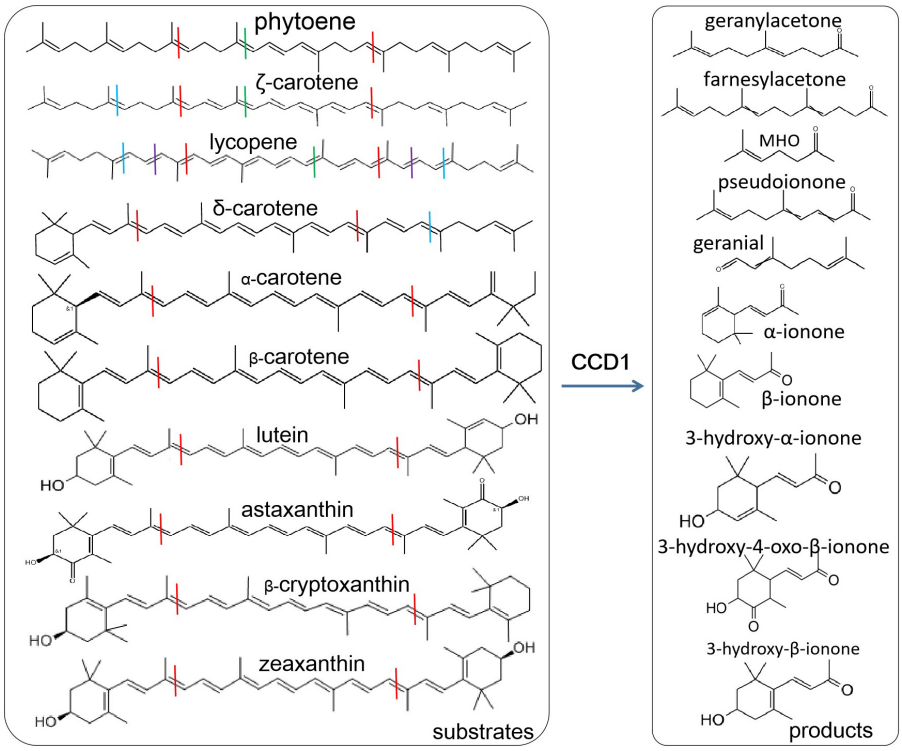


FIGURE 1
Various carotenoid substrates and products by CCD1 cleavage. Lines with different colors indicate the cleavage sites, blue at 5, 6 and 5′, 6′; red at 9, 10 and 9′, 10′; dark red at 13, 14 and 13′, 14′; purple at 7, 8, and 7′, 8′.

Dendrobium officinale is a traditional medicinal plant (Ng et al., 2012). Modern studies have shown that there are abundant bioactive compounds, such as alkaloids, bibenzyl, polysaccharides, and phenolic compounds (Tang et al., 2017; Wang Y. et al., 2022). However, there is limited knowledge on the biosynthesis of beta-ionone. In the present study, beta-ionone was detected in stems and leaves through headspace GC/MS and then genome-wide identification and characterization of the CCD gene family were performed in *D. officinale*. The expression of *DoCCD* genes was analyzed in various tissues and the function of DoCCD1 was investigated in *E. coli* as well as in plants. These results provide the valuable information on the CCD gene family as well as the function of DoCCD1 in the biosynthesis of beta-ionone.

Materials and methods

Plant materials

12-month-old plants *in vitro* of *Dendrobium officinale* Kimura et Migo were grown in the tissue culture room of the Kunming Institute of Botany, Chinese Academy of Sciences (Kunming, Yunnan, China). The growth condition was at 25 ± 2°C with a photoperiod of 12 h light/12 h dark. Healthy leaves and stems were harvested, quickly frozen in liquid nitrogen and kept at −80°C for future RNA extraction. After 50 days grown in soils, *Nicotiana benthamiana* plants were used to *Agrobacteria*-mediated injection.

Identification and phylogenetic analysis of carotenoid cleavage dioxygenase genes in *Dendrobium officinale*

The RPE65 (retinal pigment epithelial membrane protein) domain was used to search the putative CCD proteins in *D. officinale* genome (Zhang et al., 2016) with HMMER (Finn et al., 2015) and E value < $1e^{-100}$. After removing the repeated sequences, the sequences were remained with the conserved domain confirmed by SMART and NCBI. The isoelectric point and molecular weight of DoCDD deduced proteins were calculated with the ProtParam tool (Wilkins et al., 1999). Potential localization of DoCDD proteins was predicted by online iPSORT Prediction[1] (Bannai et al., 2002).

1 https://ipsort.hgc.jp/index.html

Moreover, the sequences of nine *Arabidopsis thaliana AtCCD* genes were obtained from TAIR and the sequences of 13 rice (*Oryza sativa*) *OsCCD* genes were obtained from phytozome.[2] The sequences of CCD proteins from three species were aligned through ClustalW (McWilliam et al., 2013) and a phylogenetic tree was constructed using MEGA 7.0 (Kumar et al., 2016) with the neighbor-joining method and a bootstrap of 1,000 replicates.

Conserved motifs and gene structures of *DoCCD genes*

Ten conserved motifs embedded in the CDD proteins were searched with MEME[3] (Bailey et al., 2009) and potential functions of these motifs were further analyzed through the InterPro database[4] (Apweiler et al., 2001). The gene structure of *CCD* genes was obtained by comparing the genomic sequences with the coding sequences with online GSDS[5] (Hu et al., 2015).

Expression analysis of *DoCCD* genes in various tissues

The expression patterns of nine *DoCCD* genes were analyzed based on the previously published transcriptomic data (PRJNA348403, Zhang et al., 2017). Eight tissues included column, flower buds, green root tip, leaf, lip, sepal, stem, and white root. The FPKM method was used to estimate the transcript abundance (Wang and Liu, 2020) and a heatmap was obtained through R package pheatmap (v1.0.10).[6]

Amplification of *DoCCD1* gene

The total RNA was extracted from leaf sample using the RNAprep Pure Plant Plus Kit (Cat DP441, Tiangen, China). The cDNA was obtained with the one-step cDNA Synthesis Kit (Cat AT311, Transgen, China). The coding fragments were amplified using high fidelity Q5 polymerase (NEB) with the specific primers of *DoCCD1* gene. The sequences of forward and reverse primers were 5′-ATGGAGAAGGAAAATGGAA-3′ and 5′-CTTTCCCTGTTGTTGAAGTTGTTCC-3′, respectively.

2 https://phytozome.jgi.doe.gov/pz/portal.html
3 http://meme.nbcr.net/meme3/mme.html
4 http://www.ebi.ac.uk/interpro
5 http://gsds.cbi.pku.edu.cn/
6 https://rdrr.io/cran/pheatmap/

Expression of *DoCCD1* gene in *Escerichia coli* strains

The *DoCCD1* fragment was cloned into the pThio-Dan1 vector through the ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China), which produced the vector pThio-Dan1-DoCCD1. *DoCCD1* gene was driven under the arabinose-induced promoter. The expression vector pThio-Dan1-DoCCD1 and the empty vector pThio-Dan1 were transformed into different *E. coli* strains BL21-AI, respectively. The BL21-AI strains carry different carotenoid-producing genes and therefore accumulate different types of carotenoids, including lycopene (pACCRT-EIB), beta-carotene (pACCAR16△crt) and zeaxanthin (pACCAR25△crtX) (Misawa et al., 1995). When the OD_{600} value of cultures was up to 0.6, 0.1 mM arabinose was added. The strains were cultured at 16°C for 20 h to induce protein expression and strain cultures were harvested for further analysis.

Transient expression of *DoCCD1* gene in *Nicotiana benthamiana*

The *DoCCD1* fragment was cloned into plant expression vector pHREAC through the ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China), therefore producing pHREAC-DoCCD1. *DoCCD1* gene was driven under 35S promoter. The empty vector pHREAC and the expression vector pHREAC-DoCCD1 were transformed into *Agrobacteria* strain GV3101, respectively. *N. benthamiana* leaves were injected, as described by Bruückner and Tissier (2013). Briefly, the *Agrobacteria* were cultured and then resuspended with infiltration buffer with OD_{600} of 0.5. The plants were infiltrated with 1 mL syringe on the abaxial side of leaves and maintained at the growth chamber. After 3 days leaves were harvested, immediately frozen and maintained at −80°C for further analysis.

Headspace solid-phase microextraction-gas chromatograph/mass spectrometer for volatile analysis

Fresh plant materials after ground in liquid nitrogen or strain cultures were used for headspace volatile analysis. The solid-phase microextraction (SPME) was connected to the headspace bottle and the samples were treated at 40°C for 30 min for volatile collection. The analysis of volatiles was carried out by Agilent7890-5975 Gas Chromatograph/Mass Spectrometer (GC/MS). The separation column was DB-5MS

(30 m × 0.25 mm × 0.25 μm) and helium was used as the carrier gas with a flow velocity of 1.1 mL/min. The temperature at the sample inlet was 250°C. The GC oven was first maintained at 40°C for 1 min, and then followed by 2°C/min up to 60°C and finally 10°C/min up to 325°C. The temperature of ion source was kept at 250°C. The MS data was acquired through scanning the range of m/z 33–500. The products were qualitatively analyzed based on the retention index (RI) and mass spectra of standards.

Analysis of carotenoids by high performance liquid chromatography

Carotenoids produced in 50 mL *E. coli* strain cultures were obtained through extraction with sonication-mediated acetone. The analysis of 5 μL sample was carried out on the Waters e2695 Alliance High Performance Liquid Chromatography (HPLC). The column was WaterNova-PAK-C18 (3.9 m × 150 mm × 4μm) at 30°C. The absorbance was detected at 450 nm. The separation was performed a flow rate of 0.5 mL/min by isopropanol as solvent A and acetonitrile:water (8:2, v/v) as solvent B. The gradient elution was used as following 100% B (0 min); 100% A (40 min). The equilibrium time was 6 min. Peaks were qualitatively identified according to the retention index of standards in the database.

Results

Abundant beta-ionone in *Dendrobium officinale*

The stems and leaves of *D. officinale* are main resource materials for medicines and foods. Therefore, their volatile compounds were detected through headspace SPME-GC/MS. As shown in **Figure 2**, it was found that there were two and three most significant peaks in leaf and stem samples, respectively. According to the RI and mass spectra, four peaks (1–4) were qualitatively identified (**Supplementary Figure 1**). The most abundant compound in leaf was 2-hexenal (peak 1) that has fresh green leaf notes while the most abundant compound in stem was 2-octenal (peak 3). Importantly and interestingly, beta-ionone was found to be the commonly dominant compound of volatiles in both the leaf and stem (peak 2 and peak 4) and the levels in two tissues were similar. The beta-ionone with violet and woody notes has been widely used in various industries, which is therefore worth further investigating on its biosynthesis in *D. officinale*.

Identification and phylogenetic analysis of *DoCCD* members in *Dendrobium officinale*

CCD proteins have a typical domain, namely, RPE65 (retinal pigment epithelial membrane protein) domain. To obtain the potential genes are potentially involved in the production of beta-ionone in *D. officinale*, identification of *CCD* genes was first carried out based on its genome using RPE65 domain as query with *E*-value < $1e^{-100}$ through HMM3.0. After removal of repeated sequences and confirmation of present RPE65 domain, a total of nine CCD candidates were obtained. According to the homologs in Arabidopsis and rice, these CCD members were named as DoCCD1, DoCCD4, DoCCD7a, DoCCD7b, DoCCD7c, DoCCD8, DoNCED2, DoNCED3, and DoZAS. The detailed information was listed in **Supplementary Table 1**. The lengths of coding sequences ranged from 750 to 1,902 bp, while the lengths of genomic DNA sequences varied from 750 to 4,1354 bp. The lengths of putative proteins were 249–633 AA, with molecular weights of 27.16–71.81 kD and pI of 5.97–9.39. CCD proteins usually include a chloroplast transit peptide (cTP) domain at the N-terminal, which can facilitate proteins to the plastid (Tan et al., 2003). The prediction of the cTP domain showed that seven DoCCD proteins contain the cTP domain while DoCCD1 and DoCCD4 lack the cTP domain (**Supplementary Table 1**).

To reveal the relationship between CCD proteins, a phylogenetic tree was constructed including nine DoCCDs, nine AtCCDs, and 13 OsCCDs (**Figure 3A**). These CCD proteins were clearly classified into six subfamilies, including CCD1, CCD4, CCD7, CCD8, ZAS, and NCED, which was the same as in rice (Wang et al., 2019) and *Gossypium* (Zhang et al., 2021). In *D. officinale*, there were three members in the CCD7 subfamily while there was only one member in Arabidopsis and rice, respectively. DoCCD7a was close to AtCCD7 and OsCCD7 proteins whereas DoCCD7b and DoCCD7c formed a new branch. There were great differences in the number of NCED members among species, such as five in Arabidopsis, three in rice, and two in *D. officinale*. For other four subfamilies, there was only one member in *D. officinale*. The ZAS subfamily was newly identified (Wang et al., 2019). There were one and four members in *D. officinale* and rice, respectively, while no member was found in Arabidopsis. These results showed CCD proteins are conservatively present in plants and may maintain their conserved functions.

Conserved motifs and gene structure of *DoCCD* genes

To further characterize *DoCCD* genes, the gene structure and conserved motifs were analyzed. It was found that there were 10 conserved motifs, representing part of the RPE65

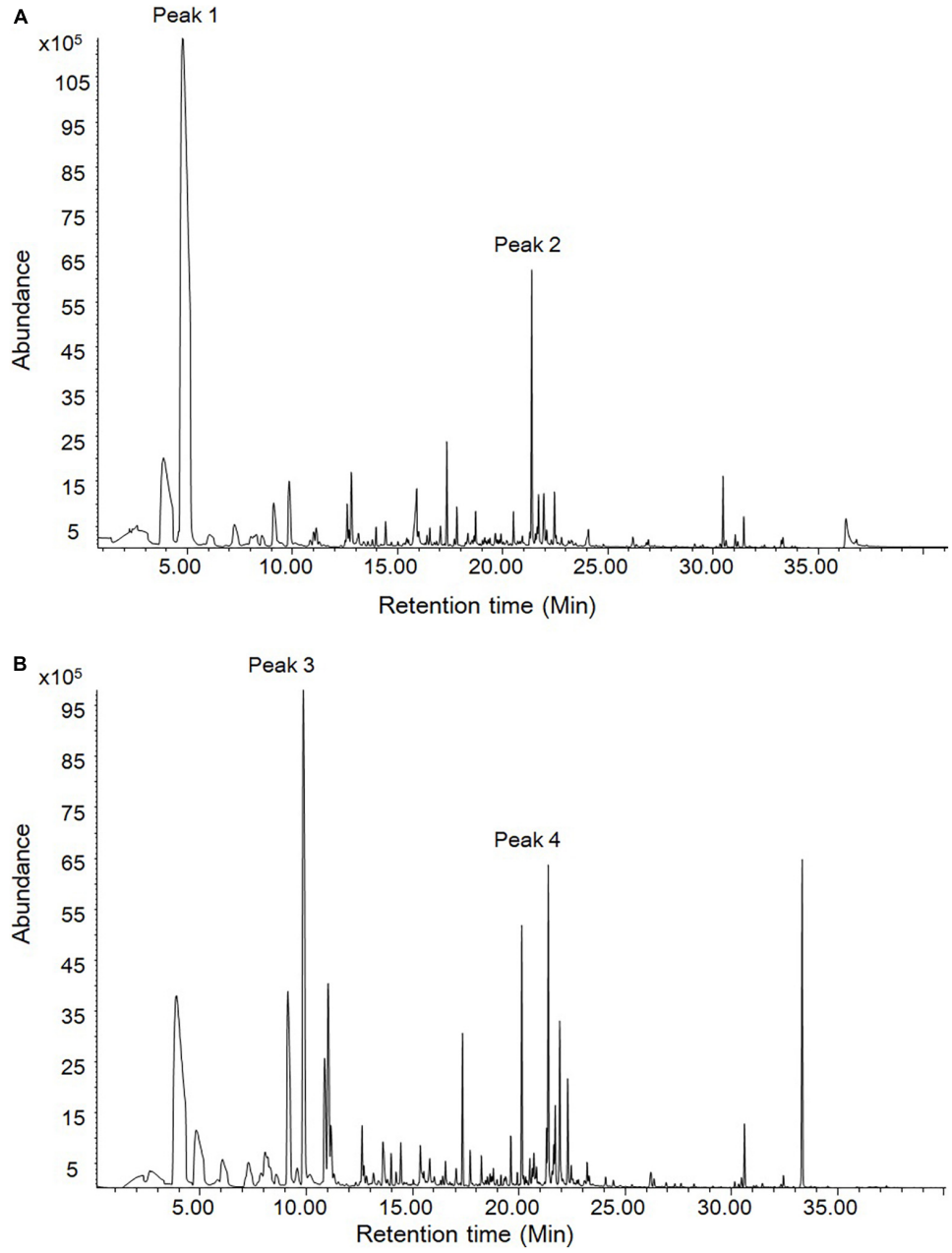


FIGURE 2
The volatile compounds in leaves and stems of *D. officinale* through headspace SPME-GC/MS analysis. **(A)** The leave volatiles were detected. **(B)** The stem volatiles were detected. According to RI and mass spectra, peak 2 and peak 4 were qualitatively identified as beta-ionone.

domain (Figure 3B). Most of the members (25) contained all the 10 motifs, while other members lacked some motifs, such as DoCCD7b and OsCCD7. Members from the CCD8 subfamily contained only five or six motifs. The analysis of gene structures showed that there were notable differences between subfamilies while the members in the same subfamily showed the same or similar gene structures (Figure 3C). The members from the NCED and CCD4 subfamilies were intronless, while other members contained varying introns. All the members from the CCD1 and ZAS subfamilies contained 12 introns, six introns were embedded in CCD8 members while most members of the CCD7 subfamily included six introns, expect *DoCCD7b* having no intron. Notably, although *CCD* genes in the same subfamily had the same number of introns, the introns in *DoCCD* genes were generally longer than those in Arabidopsis and rice. These results suggest that *CCD* genes maintain highly conserved gene structures during evolution.

FIGURE 3
Phylogenetic analysis and characterization of *CCD* genes from three species. **(A)** The phylogenetic tree of 31 CCD proteins. CCD proteins were aligned with ClustalW and an unrooted tree was constructed using MEGA 7.0 with the neighbor-joining method and bootstraps of 1,000 replicates. **(B)** Ten conserved motifs were analyzed in CCD proteins. **(C)** The exon-intron structures of *CCD* genes were analyzed. *At*, *Do* and *Os* stand for *Arabidopsis thaliana*, *Oryza sativa* and *Dendrobium officinale*, respectively.

Expression patterns of *DoCCD* genes in various tissues

The patterns of gene expression can provide important information for their functions. According to the previously transcriptomic data (Zhang et al., 2017), nine *DoCCD* genes were analyzed in eight tissues, including column, flower buds, green root tip, leaf, lip, sepal, stem, and white root. As shown in **Figure 4**, these genes exhibited distinct expression profiles. Among them, *DoCCD7a*, *DoCCD8,* and *DoZAS* genes were hardly detected in these tissues. *DoCCD7b* was expressed in column, flower buds and lip with varying levels and the transcripts in lip were the most abundant. *DoCCD7c* gene was almost not expressed in six tissues, but specifically and highly expressed in green root tip and white root. Two *DoNCED* genes showed differential expression patterns. *DoNCED3* gene was expressed with very low or no detectable levels while *DoNCED2* gene was expressed in various tissues with low levels. The highest transcript of *DoNCED2* gene was in flower buds and the second was in root, suggesting their divergent functions. *DoCCD4* gene was extensively expressed in eight tissues, and the transcripts in leaf were the most abundant with an FPKM of 137.10 and followed by sepal and column. *DoCCD1* gene was highly and extensively detected in eight tissues and the highest transcript was found in leaf with an FPKM of 338.06 and the lowest transcript in the white root with an FPKM of 68.37.

Function analysis of *DoCCD1* enzyme

Due to low solubility of carotenoids in water, it is usually difficult to carry out experiments *in vitro*. An alternative is to detect the CCD activity in *E. coli* strains that can accumulate the specific type of carotenoids. To detect potential function of the enzyme encoded by *DoCCD1* gene, the CDS sequences were cloned to produce an inducible pThio-Dan1-DoCCD1 vector. The expression of *DoCCD1* gene embedded in this vector can be induced once arabinose is added into the culture solution. The plasmid pThio-Dan1-DoCCD1 and empty vector pThio-Dan1 were transformed to three BL21-AI strains that can accumulate lycopene (pACCRT-EIB), beta-carotene (pACCAR16△crt), and zeaxanthin (pACCAR25△crtX) (Misawa et al., 1995), respectively. After induced expression of *DoCCD1* gene by arabinose, it was found that the strain cultures exhibited remarkable changes in colors, varying from lighter color to almost white (**Figure 5**), while the controls (empty vector) kept pigmentation. These results suggest that *DoCCD1* gene was functional in *E. coli* and DoCCD1 enzyme may use multiple substrates such as lycopene, beta-carotene, and zeaxanthin, which then can result in lack of carotenoid pigmentation. The carotenoids extracted from these strain cultures were further detected by HPLC. As shown in **Figure 5**, abundant carotenoids (lycopene, beta-carotene, and zeaxanthin) were found in the control cultures while almost no abundance was detected in the strain cultures with expression of *DoCCD1* gene. These results demonstrate that DoCCD1 can catalyze these three carotenoids to result in color changes.

To determine the possible products in carotenoid accumulation strains with expression of *DoCCD1* gene, the volatiles were detected by headspace SPME-GC/MS (**Figure 6**). After *DoCCD1* gene was expressed in the strains that can yield lycopene, two new peaks were remarkably observed, but their abundances were low in the control (**Figure 6A**). According to the RI and mass spectra in the database, these

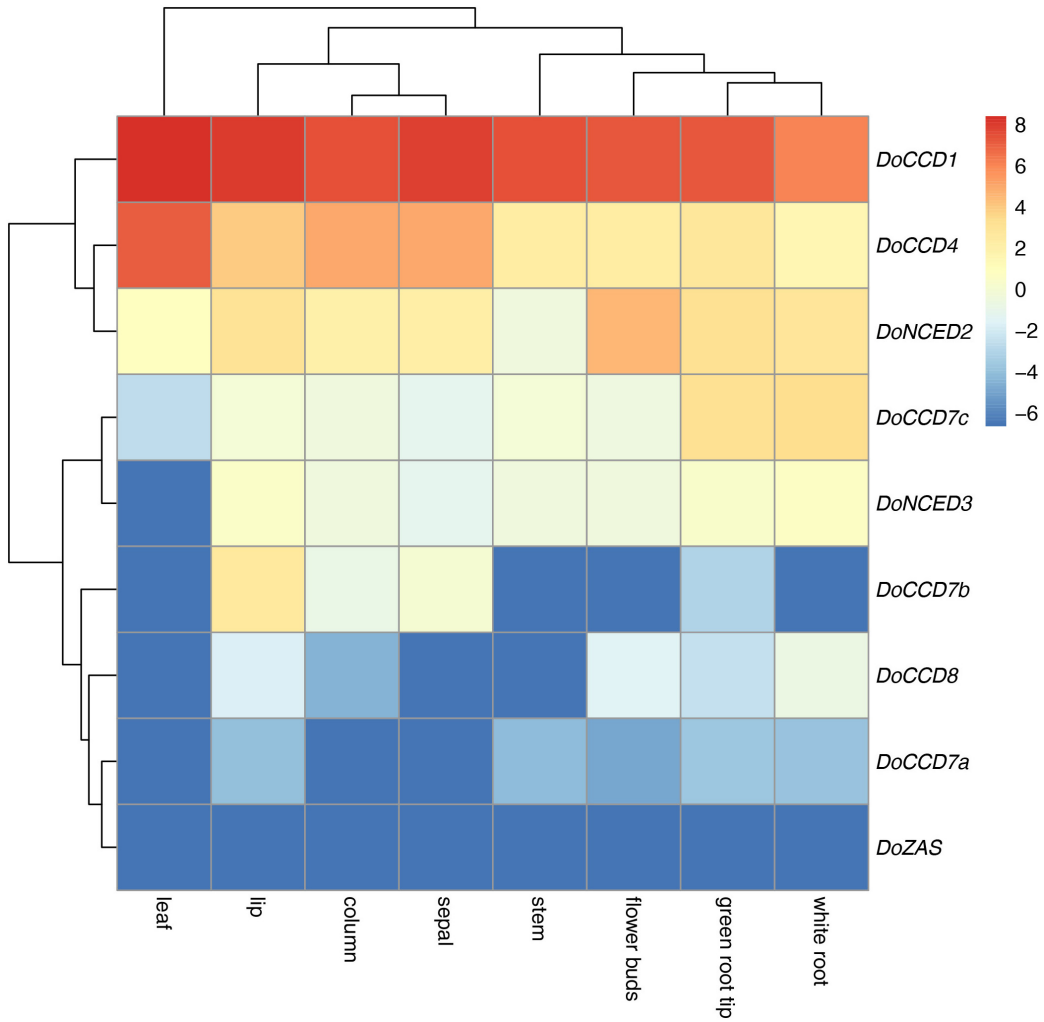


FIGURE 4
Expression heatmap of nine *DoCCD* genes in eight tissues. The data was from the RNA-Seq (Zhang et al., 2017) including tissues such as column, flower buds, lip, sepal, leaf, stem, white root, and green root tip. R package was used to construct the heatmap and the color scale stands for the normalized values.

two products were qualitatively identified as pseudoionone and 6-methyl-5-hepten-2-one (MHO), respectively. These results demonstrate that DoCCD1 can cleave the lycopene at different bond sites. Pseudoionone was formed through the cleavage at 9, 10 and 9′, 10′ sites while MHO was formed through the cleavage at 5, 6 and/or 5′, 6′ sites. Notably, the abundance of MHO was much higher than that of pseudoionone, which implies that more lycopene was cleaved at 5, 6 and/or 5′, 6′ sites by DoCCD1. After *DoCCD1* gene was expressed in the beta-carotene accumulation strains, a new peak was obviously observed, but the peak was much low in the control (**Figure 6B**). The product was qualitatively identified as beta-ionone based on RI and mass spectra. These results demonstrate that DoCCD1 can catalyze beta-carotene at the 9, 10 and 9′, 10′ bond sites to produce beta-ionone. After *DoCCD1* gene was expressed in zeaxanthin accumulation strains, no significant peak was detected compared with the control, although color change was found. These results suggest that DoCCD1 may cleave zeaxanthin but produce undetectable products by GC/MS.

To determine the function of *DoCCD1* in *planta*, the fragment of *DoCCD1* gene was cloned into plant expression vector pHREAC and under 35S promoter, producing engineered GV3101 strain. The volatiles in *N. benthamiana* leaves were analyzed by headspace SPME-GC/MS (**Figure 7**). The results showed that a remarkable peak appeared in the *N. benthamiana* leaves with *DoCCD1* expression, while the peak in the control was very low. This peak was qualitatively to be beta-ionone. These results suggest that *DoCCD1* gene was functional in *N. benthamiana* leaf and DoCCD1 enzyme can catalyze the *N. benthamiana* carotenoids to produce beta-ionone.

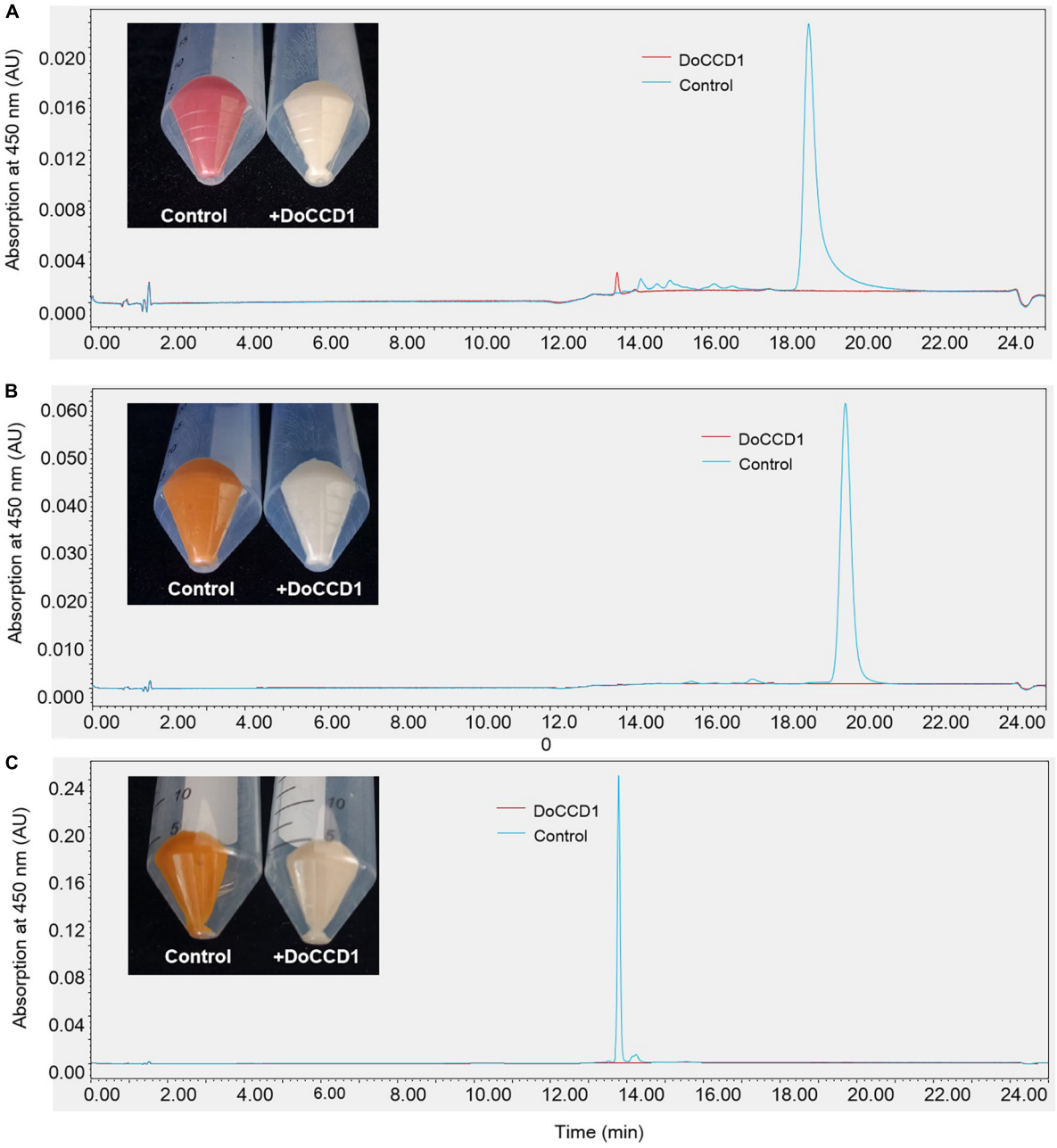


FIGURE 5
The color changes and HPLC analysis of *E. coli* strains. **(A)** *E. coli* strain can accumulate lycopene (pACCRT-EIB). **(B)** *E. coli* strain can accumulate beta-carotene (pACCAR16△crt). **(C)** *E. coli* strain can accumulate zeaxanthin (pACCAR25△crtX). Control stands for the strain containing empty vector pThio-Dan1 while + DoCCD1 stands for the strain containing plasmid pThio-Dan1-DoCCD1.

Discussion

Carotenoids and their derivatives apocarotemoids play important roles in plant life including growth, development and responses to environmental changes. Among them, beta-ionone, a C13 volatile compound with a relatively low odor threshold is one of the key components for plant essential oils, flower and fruit flavors (Huang et al., 2009; Baldermann et al., 2010; Paparella et al., 2021). The bioactive effects of beta-ionone such as anti-cancer have been demonstrated (Paparella et al., 2021). *D. officinale* has been used as a medicinal, edible, and ornamental plant, and is rich in polysaccharides, phenols, and terpene compounds (Tang et al., 2017; Wang Y. et al., 2022). In our present study, among the volatile components in the

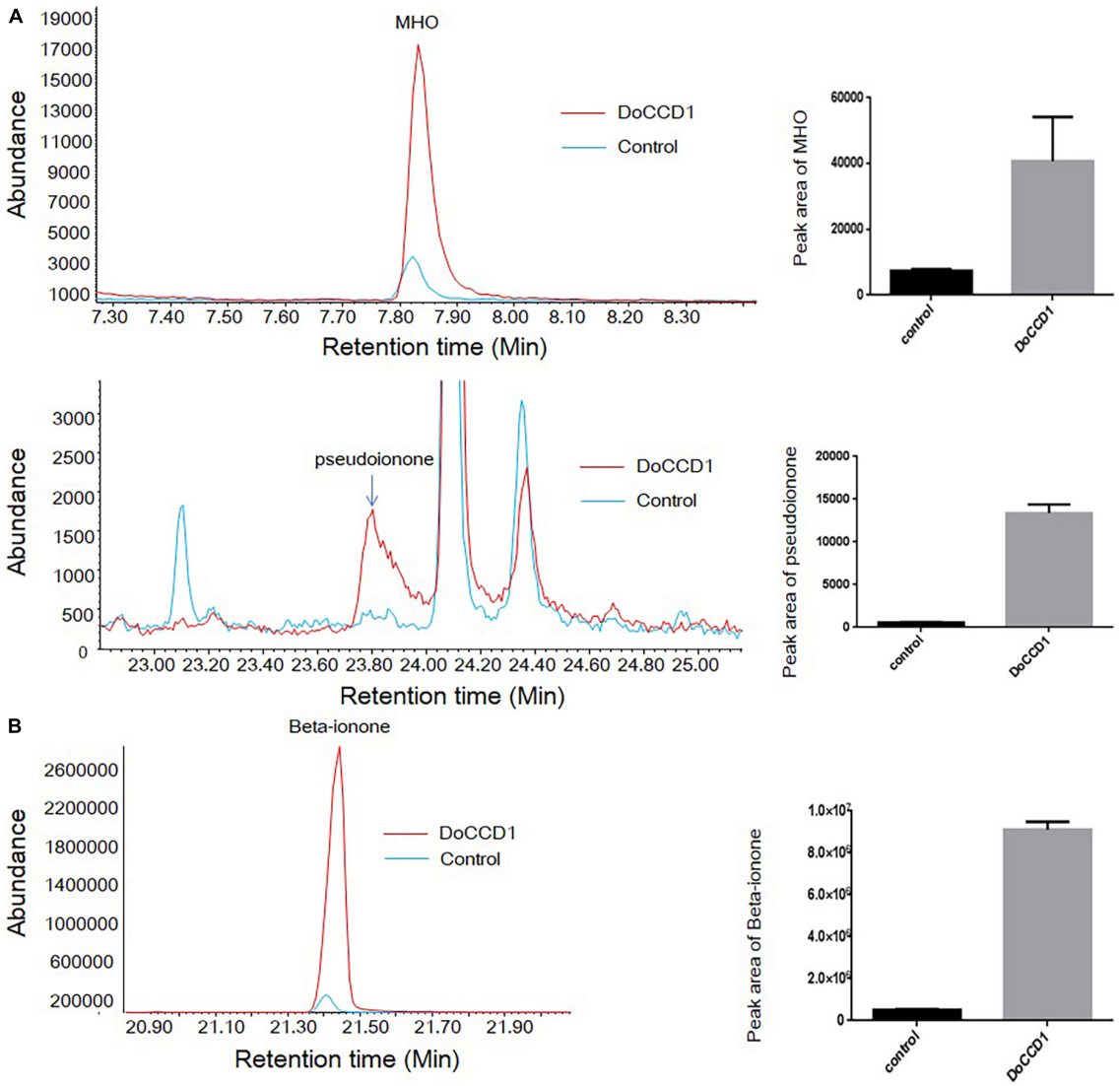


FIGURE 6
Products in carotenoid-producing *E. coli* strains by headspace SPME-GC/MS. **(A)** *E. coli* strains can accumulate lycopene. Two peaks were identified as MHO (up) and pseudoionone (down), respectively. **(B)** *E. coli* strains can accumulate beta-carotene. The peak was identified as beta-ionone. Red lines stand for the strain containing *DoCCD1* gene. Blue lines stand for the control strain without *DoCCD1* gene.

stem and leaf, the relatively abundant beta-ionone was found (**Figure 2**), which was similar to the previous report (Ma et al., 2018). Beta-ionone was also detected in the fresh flowers of *D. loddigesii* (Wang S. et al., 2022). No beta-ionone was detected in some reports (Chen et al., 2016; Hu et al., 2020), which might be due to the different methods applied. The headspace SPME-GC/MS method used in the present work is effective to detect beta-ionone. The presence of beta-ionone in *D. officinale* can provide important basis for use of leaves and stems and their effective values.

The *CCD* genes encode the CCDs that can catalyze carotenoids into formation of apocarotenoids, such as beta-ionone. In our study, a total of nine *CCD* gene members were first identified in *D. officinale*, which can be classified into six subfamilies including CCD1, CCD4, CCD7, CCD8, NCED, and ZAS as in rice (Wang et al., 2019). However, Arabidopsis lacks the ZAS subfamily (Tan et al., 2003). The numbers of *CCD* genes in different species vary greatly, which could be related to the genome ploidy, gene replication and species evolution. There were nine *CCD* gene members in *D. officinale* and Arabidopsis (Tan et al., 2003), respectively, which were less than the CCD members in other species, such as rice (Tan et al., 2003), tobacco (Zhou et al., 2019), cotton (Zhang et al., 2021), Saccharum (Su et al., 2021). In dicotyledonous plants Arabidopsis and cotton, there were more NCED members than those of the CCD type, suggesting the expansion of NCED and

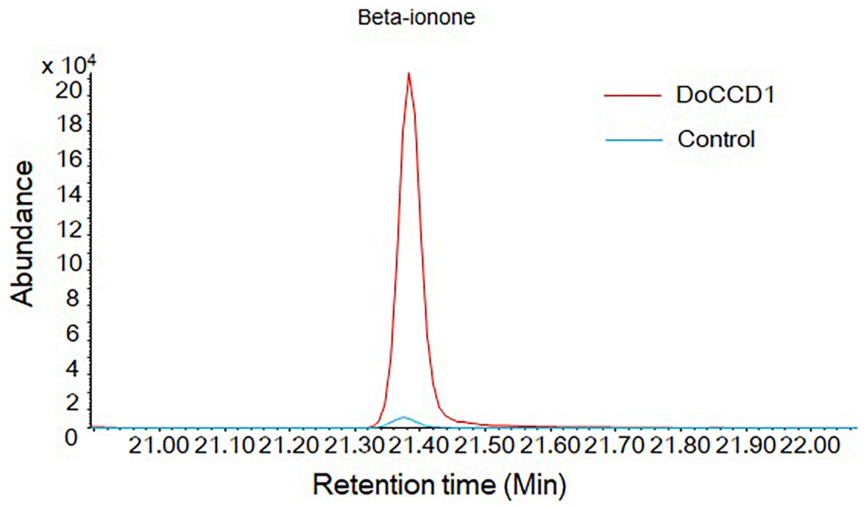


FIGURE 7
Products in *N. benthamiana* leaves with expression of *DoCCD1* by headspace SPME-GC/MS. The peak was identified as beta-ionone. The red line stands for *N. benthamiana* leaves with expression of *DoCCD1* gene. The blue line stands for control leaves without *DoCCD1* gene.

their important roles. However, in monocotyledonous plants *D. officinale* and rice, there were only two or three NCED members. These results imply that duplication of *NCED* genes may happen after differentiation of the monocot and eudicot (Vallabhaneni et al., 2010). Notably, there were three members in the CCD7 subfamily in *D. officinale*. Among them, DoCCD7a was closely clustered with AtCCD7 and OsCCD7, suggesting their conserved function. However, DoCCD7b and DoCCD7c formed a new branch, which may imply their new unknown functions. The analysis of gene structures revealed that the *DoCCD* genes from the same subfamily shared the same or similar patterns while the *DoCCD* genes from the different subfamilies exhibited different patterns (**Figure 3**). The genes from the NCED and CCD4 subfamilies had no intron while other *DoCCD* genes included varying numbers of introns. The similar profiles have been observed in Arabidopsis and rice (**Figure 3**). The intronless genes usually showed a faster evolution rate, which can facilitate their quick response to the environmental changes (Tan et al., 2003; Zhang et al., 2021).

The patterns of gene expression can provide important information for their roles in plants. *DoCCD* genes showed distinct expression profiles (**Figure 4**). *DoCCD1* gene was expressed in various tissues, as other *CCD1* genes from other plants, including Arabidopsis (Auldridge et al., 2006b), tomato (Simkin et al., 2004a), petunia (Simkin et al., 2004b), strawberry (García-Limones et al., 2008), cotton (Zhang et al., 2021), which were expressed in leaf, flower, stem, fruit, and root. The extensive expression of *CCD1* genes in plant tissues shows their conserved function. The high expression of *CCD1* gene in flower and fruit can enhance emission of volatile apocarotenoids, which therefore was helpful to attract pollinators and propagators (García-Limones et al., 2008). The expression of *CCD1* genes in vegetative tissues can produce some apocarotenoids with antimicrobial activities, which can prevent plants from infection of fungal pathogens (Simkin et al., 2004b; Auldridge et al., 2006b). *CCD7* and *CCD8* genes were mainly expressed in the roots of rice, Arabidopsis, and petunia (Vallabhaneni et al., 2010) and involved in the biosynthesis of strigolactone (Alder et al., 2012). In *D. officinale*, three *DoCCD7* genes showed differential expression patterns. Among them, *DoCCD7c* gene was specifically expressed in root tissues, indicating its possible roles in the biosynthesis of strigolactone. However, *DoCCD7b* gene was expressed in some flower tissues, implying its putative functions in flower development. The *NCED* genes are mainly related to the biosynthesis of ABA (Schwartz et al., 1997). *DoNCED3* was not expressed while *DoNCED2* gene was expressed with low levels in various tissues, which suggests that the function of DoNCED3 may be lost while DoNCED2 may mainly play roles in the formation of ABA in root and unknown roles in flowers.

Previous studies have shown that CCD1 enzyme can utilize a series of compounds as substrates, including acyclic and cyclic carotenoids (Vogel et al., 2008). Due to low solubility of carotenoids in water, the activity of CCD1 enzymes was mainly performed on *E. coli* strains that can accumulate different carotenoids (Misawa et al., 1995). In our study, DoCCD1 can cleave three substrates (lycopene, beta-carotene, and zeaxanthin) to result in discoloration (**Figure 5**). When zeaxanthin was used as the substrate, no product was detected by GC/MS or HPLC, which has been observed before (Auldridge et al., 2006b). The possible explanation was the volatile ability of products was too low or the products were catabolized by the strains (Auldridge et al., 2006b). Previous studies have shown that CCD1 can cleave the substrates at the 9, 10 (9′, 10′) bond

sites, such as CCD1 from *A. thaliana* (Schwartz et al., 2001), tea plant (Wang et al., 2020a), *Morus notabilis* (Qi et al., 2021), *Petunia hybrida* (Simkin et al., 2004b), *Lycopersicon esculentum* (Simkin et al., 2004a), *Rosa damascena* (Huang et al., 2009), *Osmanthus fragrans* (Baldermann et al., 2010). When beta-carotene was the substrate for CCD1, the cleavage at the 9, 10 (9′, 10′) bond sites can produce beta-ionone (Schwartz et al., 2001), which can influence the fragrance and flavors of flowers, fruits, and essential oils (Schwartz et al., 2001; Simkin et al., 2004a,b; Brandi et al., 2011). When lycopene was the substrate, CCD1 enzymes from different species showed various cleavage sites. DoCCD1 can cleave lycopene at 5, 6 and/or 5′, 6′ bond sites to produce 6-methyl-5-hepten-2-one (MHO), and also can produce pseudoionone due to the cleavage at 9, 10 (9′, 10′) bond sites. However, the cleavage at 5, 6 and/or 5′, 6′ bond sites may be more preferential than at 9, 10 (9′, 10′) bond sites because more abundant MHO was detected (**Figure 6A**). Other CCD1 enzymes from Arabidopsis, *Bixa Orellana*, *Zea mays,* and *Lycopersicon esculentum* can also cleave lycopene at 5, 6 and/or 5′, 6′ bond sites to form MHO (Vogel et al., 2008). However, MnCCD1 failed to produce MHO but produced pseudoionone due to the cleavage at 9, 10 (9′, 10′) sites (Qi et al., 2021). CCD1 from carrot didn't use lycopene as a substrate (Yahyaa et al., 2013). OsCCD1 can cleave lycopene to produce pseudoionone and MHO, and geranial was also detected due to its cleavage at 7, 8 (7′, 8′) sites (Ilg et al., 2009). In some plants, there were two *CCD1* genes and they also showed different substrates and cleavage sites (Simkin et al., 2004a). These studies demonstrate the conserved cleavage activity of CCD1 enzymes as well as specificity in different plants.

In plants CCD1 was localized in the cytoplasm and can use various carotenoids and apocarotenoids as substrates (Auldridge et al., 2006b; Ilg et al., 2010), which can have great influences on the flavor, fragrance and quality. Loss of AtCCD1 results in increases in seed carotenoid contents but plants had no other changed phenotype (Auldridge et al., 2006b). In the maize endosperm, high levels of *CCD1* gene lead to dominant *White cap* mutant and deficient carotenoids in kernels (Timmermans et al., 2004). LeCCD1A and LeCCD1B in tomato can cleave lycopene that is responsible for the red fruit and produce C14 dialdehydes and pseudoionone and suppression of these two genes resulted in the decreased levels of beta-ionone and geranylacetone in fruits (Simkin et al., 2004a; Ilg et al., 2014). In our study, expression of *DoCCD1* gene in *N. benthamiana* leaves enhanced the level of beta-ionone (**Figure 7**), which can provide an alternative method for production of beta-ionone in plants.

To summarize, the present study confirmed the presence of beta-ionone in *D. officinale*, identified the CCD gene family members and analyzed their expression patterns. Further functional analyses demonstrate that DoCCD1 can catalyze different substrates to produce apocarotenoids. These data elucidate the potential role of DoCCD1 in the formation of beta-ionone in *D. officinale*.

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

AL: conceptualization. YW and JX: experiments and data analysis. YW: writing—original draft. AL: writing—review and revising. All authors have read and agreed to the published version of the manuscript.

Funding

We are grateful for funds obtained from the Yunnan Provincial Science and Technology Department (grant no. 202101AT070189), the National Natural Science Foundation of China (grant no. 31701465), and the Beijing DR PLANT Biotechnology Co., Ltd. (grant no. E0514832C1). This study received funding from Beijing DR PLANT Biotechnology Co. Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.967819/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The mass spectra of detected compounds.

References

Ablazov, A., Mi, J., Jamil, M., Jia, K. P., Wang, J. Y., Feng, Q., et al. (2020). The apocarotenoid zaxinone is a positive regulator of strigolactone and abscisic acid biosynthesis in Arabidopsis roots. *Front. Plant Sci.* 11:578. doi: 10.3389/fpls.2020.00578

Ahrazem, O., Gómez-Gómez, L., Rodrigo, M. J., Avalos, J., and Limón, M. C. (2016b). Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: features and functions. *Int. J. Mol. Sci.* 17:1781. doi: 10.3390/ijms17111781

Ahrazem, O., Rubio-Moraga, A., Berman, J., Capell, T., Christou, P., Zhu, C., et al. (2016a). The carotenoid cleavage dioxygenase CCD2 catalysing the synthesis of crocetin in spring crocuses and saffron is a plastidial enzyme. *N. Phytol.* 209, 650–663. doi: 10.1111/nph.13609

Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., et al. (2012). The path from β-carotene to carlactone, a strigolactone-like plant hormone. *Science* 335, 1348–1351. doi: 10.1126/science.1218094

Aloum, L., Alefishat, E., Adem, A., and Petroianu, G. (2020). Ionone is more than a violet's fragrance: a review. *Molecules* 25:5822. doi: 10.3390/molecules25245822

Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., et al. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29, 37–40.

Auldridge, M. E., McCarty, D. R., Klee, H. J.(2006a). Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr. Opin. Plant Biol.* 9, 315–321. doi: 10.1016/j.pbi.2006.03.005

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., et al. (2006b). Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant J.* 45, 982–993. doi: 10.1111/j.1365-313X.2006.02666.x

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

Baldermann, S., Kato, M., Kurosawa, M., Kurobayashi, Y., Fujita, A., Fleischmann, P., et al. (2010). Functional characterization of a carotenoid cleavage dioxygenase 1 and its relation to the carotenoid accumulation and volatile emission during the floral development of *Osmanthus fragrans* Lour. *J. Exp. Bot.* 61, 2967–2977. doi: 10.1093/jxb/erq123

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signal. *Bioinformatics* 18, 298–305.

Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., et al. (2011). Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* 11:24. doi: 10.1186/1471-2229-11-24

Bruückner, K., and Tissier, A. (2013). High-level diterpene production by transient expression in *Nicotiana benthamiana*. *Plant Methods* 9:46. doi: 10.1186/1746-4811-9-46

Campbell, R., Ducreux, L. J., Morris, W. L., Morris, J. A., Suttle, J. C., Ramsay, G., et al. (2010). The metabolic and developmental roles of carotenoid cleavage dioxygenase4 from potato. *Plant Physiol.* 154, 656–664. doi: 10.1104/pp.110.158733

Chen, N. D., You, T., Li, J., Bai, L. T., Hao, J. W., and Xu, X. Y. (2016). A comparative study of three tissue-cultured Dendrobium species and their wild correspondences by headspace gas chromatography-mass spectrometry combined with chemometric methods. *J. Food Drug Anal.* 24, 839–847. doi: 10.1016/j.jfda.2016.05.006

Chiou, C. Y., Pan, H. A., Chuang, Y. N., and Yeh, K. W. (2010). Differential expression of carotenoid-related genes determines diversified carotenoid coloration in floral tissues of Oncidium cultivars. *Planta* 232, 937–948. doi: 10.1007/s00425-010-1222-x

Fang, Q., Li, Y., Liu, B., Meng, X., Yang, Z., Yang, S., et al. (2020). Cloning and functional characterization of a carotenoid cleavage dioxygenase 2 gene in safranal and crocin biosynthesis from *Freesia hybrida*. *Plant Physiol. Biochem.* 154, 439–450. doi: 10.1016/j.plaphy.2020.06.035

Felemban, A., Braguy, J., Zurbriggen, M. D., and Al-Babili, S. (2019). Apocarotenoids involved in plant development and stress response. *Front. Plant Sci.* 10:1168. doi: 10.3389/fpls.2019.01168

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–W38. doi: 10.1093/nar/gkv397

Frusciante, S., Diretto, G., Bruno, M., Ferrante, P., Pietrella, M., Prado-Cabrero, A., et al. (2014). Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12246–12251. doi: 10.1073/pnas.1404629111

Fukamatsu, Y., Tamura, T., Hihara, S., and Oda, K. (2013). Mutations in the CCD4 carotenoid cleavage dioxygenase gene of yellow-flesh peaches. *Biosci. Biotechnol. Biochem.* 77, 2514–2516. doi: 10.1271/bbb.130626

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

García-Limones, C., Schnäbele, K., Blanco-Portales, R., Luz Bellido, M., Caballero, J. L., Schwab, W., et al. (2008). Functional characterization of FaCCD1: a carotenoid cleavage dioxygenase from strawberry involved in lutein degradation during fruit ripening. *J. Agric. Food Chem.* 56, 9277–9285. doi: 10.1021/jf801096t

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hu, J., Huang, W., Zhang, F., Luo, X., Chen, Y., and Xie, J. (2020). Variability of volatile compounds in the medicinal plant *Dendrobium officinale* from different regions. *Molecules* 25:5046. doi: 10.3390/molecules25215046

Huang, F. C., Horváth, G., Molnár, P., Turcsi, E., Deli, J., Schrader, J., et al. (2009). Substrate promiscuity of RdCCD1, a carotenoid cleavage oxygenase from *Rosa damascena*. *Phytochemistry* 70, 457–464. doi: 10.1016/j.phytochem.2009.01.020

Ilg, A., Beyer, P., and Al-Babili, S. (2009). Characterization of the rice carotenoid cleavage dioxygenase 1 reveals a novel route for geranial biosynthesis. *FEBS J.* 276, 736–747. doi: 10.1111/j.1742-4658.2008.06820.x

Ilg, A., Bruno, M., Beyer, P., and Al-Babili, S. (2014). Tomato carotenoid cleavage dioxygenases 1A and 1B: relaxed double bond specificity leads to a plenitude of dialdehydes, mono-apocarotenoids and isoprenoid volatiles. *FEBS Open Bio* 4, 584–593. doi: 10.1016/j.fob.2014.06.005

Ilg, A., Yu, Q., Schaub, P., Beyer, P., and Al-Babili, S. (2010). Overexpression of the rice carotenoid cleavage dioxygenase 1 gene in Golden Rice endosperm suggests apocarotenoids as substrates in planta. *Planta* 232, 691–699. doi: 10.1007/s00425-010-1205-y

Ko, M. R., Song, M. H., Kim, J. K., Baek, S. A., You, M. K., Lim, S. H., et al. (2018). RNAi-mediated suppression of three carotenoid-cleavage dioxygenase genes, OsCCD1, 4a, and 4b, increases carotenoid content in rice. *J. Exp. Bot.* 69, 5105–5116. doi: 10.1093/jxb/ery300

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

Ma, Y. N., Hou, Z. N., Liang, Z. S., and Liu, L. (2018). Quality evaluation of the rare medicinal plant *Dendrobium officinale* based on volatile constituents, methanol extracts and polysaccharides. *Curr. Phaemaceutical Anal.* 14, 121–132. doi: 10.2174/1573412913666161201122530

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 41, W597–W600.

Misawa, N., Satomi, Y., Kondo, K., Yokoyama, A., Kajiwara, S., Saito, T., et al. (1995). Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol.* 177, 6575–6584. doi: 10.1128/jb.177.22.6575-6584.1995

Moreno, J. C., Mi, J., Alagoz, Y., and Al-Babili, S. (2021). Plant apocarotenoids: from retrograde signaling to interspecific communication. *Plant J.* 105, 351–375. doi: 10.1111/tpj.15102

Ng, T. B., Liu, J., Wong, J. H., Ye, X., Sze, S. C. W., Tong, Y., et al. (2012). Review of research on *Dendrobium*, a prized folk medicine. *Appl. Microbiol. Biotechnol.* 93, 1795–1803.

Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., and Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. *Plant Physiol.* 142, 1193–1201. doi: 10.1104/pp.106.087130

Paparella, A., Shaltiel-Harpaza, L., and Ibdah, M. (2021). β-Ionone: its occurrence and biological function and metabolic engineering. *Plants* 10:754. doi: 10.3390/plants10040754

Qi, Z., Fan, X., Zhu, C., Chang, D., Pei, J., and Zhao, L. (2021). Overexpression and characterization of a novel plant carotenoid cleavage dioxygenase 1 from *Morus notabilis*. *Chem. Biodivers.* 19: e202100735. doi: 10.1002/cbdv.202100735

Schwartz, S. H., Qin, X., and Zeevaart, J. A. (2001). Characterization of a novel carotenoid cleavage dioxygenase from plants. *J. Biol. Chem.* 276, 25208–25211. doi: 10.1074/jbc.M102146200

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

Simkin, A. J., Schwartz, S. H., Auldridge, M., Taylor, M. G., and Klee, H. J. (2004a). The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *Plant J.* 40, 882–892. doi: 10.1111/j.1365-313X.2004.02263.x

Simkin, A. J., Underwood, B. A., Auldridge, M., Loucas, H. M., Shibuya, K., Schmelz, E., et al. (2004b). Circadian regulation of the PhCCD1 carotenoid cleavage dioxygenase controls emission of beta-ionone, a fragrance volatile of petunia flowers. *Plant Physiol.* 136, 3504–3514. doi: 10.1104/pp.104.049718

Su, W., Zhang, C., Feng, J., Feng, A., You, C., Ren, Y., et al. (2021). Genome-wide identification, characterization and expression analysis of the carotenoid cleavage oxygenase (CCO) gene family in Saccharum. *Plant Physiol. Biochem.* 162, 196–210. doi: 10.1016/j.plaphy.2021.02.041

Tan, B. C., Joseph, L. M., Deng, W. T., Liu, L., Li, Q. B., Cline, K., et al. (2003). Molecularcharacterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. *Plant J.* 35, 44–56. doi: 10.1046/j.1365-313x.2003.01786.x

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

Thakur, N., Flowerika, Thakur, N., Khan, S., Pandey, A. K., and Tiwari, S. (2021). Carotenoid cleavage dioxygenases (HD-CCD1A and B) contribute as strong negative regulators of beta-carotene in Indian bread wheat (cv. HD2967). *3 Biotech* 11:221. doi: 10.1007/s13205-021-02775-y

Timmermans, M. C., Brutnell, T. P., and Becraft, P. W. (2004). The 46th annual maize genetics conference. unlocking the secrets of the maize genome. *Plant Physiol.* 136, 2633–2640. doi: 10.1104/pp.104.048702

Us-Camas, R., Aguilar-Espinosa, M., Rodríguez-Campos, J., Vallejo-Cardona, A. A., Carballo-Uicab, V. M., Serrano-Posada, H., et al. (2022). Identifying *Bixa orellana* L. new carotenoid cleavage dioxygenases 1 and 4 potentially involved in bixin biosynthesis. *Front. Plant Sci.* 13:829089. doi: 10.3389/fpls.2022.829089

Vallabhaneni, R., Bradbury, L. M., and Wurtzel, E. T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. *Arch. Biochem. Biophys.* 504, 104–111. doi: 10.1016/j.abb.2010.07.019

Vogel, J. T., Tan, B. C., McCarty, D. R., and Klee, H. J. (2008). The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. *J. Biol. Chem.* 283, 11364–11373. doi: 10.1074/jbc.M710106200

Wang, J. Y., Haider, I., Jamil, M., Fiorilli, V., Saito, Y., Mi, J., et al. (2019). The apocarotenoid metabolite zaxinone regulates growth and strigolactone biosynthesis in rice. *Nat. Commun.* 10:810. doi: 10.1038/s41467-019-08461-1

Wang, J., Wu, B., Zhang, N., Zhao, M., Jing, T., Wu, Y., et al. (2020a). Dehydration-induced carotenoid cleavage dioxygenase 1 reveals a novel route for beta-ionone formation during tea (*Camellia sinensis*) withering. *J. Agric. Food Chem.* 68, 10815–10821. doi: 10.1021/acs.jafc.0c04208

Wang, J., Zhang, N., Zhao, M., Jing, T., Jin, J., Wu, B., et al. (2020b). Carotenoid cleavage dioxygenase 4 catalyzes the formation of carotenoid-derived volatile beta-ionone during tea (*Camellia sinensis*) withering. *J. Agric. Food Chem.* 68, 1684–1690. doi: 10.1021/acs.jafc.9b07578

Wang, S., Du, Z. H., Yang, X. Y., Wang, L. L., Xia, K. F., and Chen, Z. L. (2022). An integrated analysis of metabolomics and transcriptomics reveals significant differences in floral scents and related gene expression between two varieties of *Dendrobium loddigesii*. *Appl. Science Basel.* 12:1262. doi: 10.3390/app12031262

Wang, Y., and Liu, A. (2020). Genomic characterization and expression analysis of basic helix-loop-helix (bHLH) family genes in traditional Chinese herb *Dendrobium officinale*. *Plants* 9:1044. doi: 10.3390/plants9081044

Wang, Y., Tong, Y., Adejobi, O. I., Wang, Y., and Liu, A. (2022). Research advances in multi-omics on the traditional Chinese herb *Dendrobium officinale*. *Front. Plant Sci.* 12:808228. doi: 10.3389/fpls.2021.808228

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., et al. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552.

Yahyaa, M., Bar, E., Dubey, N. K., Meir, A., Davidovich-Rikanati, R., Hirschberg, J., et al. (2013). Formation of norisoprenoid flavor compounds in carrot (*Daucus carota* L.) roots: characterization of a cyclic-specific carotenoid cleavage dioxygenase 1 gene. *J. Agric. Food Chem.* 61, 12244–12252. doi: 10.1021/jf404085k

Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., et al. (2015). Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in Brassica species. *N. Phytol.* 206, 1513–1526. doi: 10.1111/nph.13335

Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R., Hsiao, Y. Y., Niu, S. C., et al. (2017). The *Apostasia* genome and the evolution of orchids. *Nature* 549, 379–383.

Zhang, G. Q., Xu, Q., Bian, C., Tsai, W. C., Yeh, C. M., Liu, K. W., et al. (2016). The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* 6:19029. doi: 10.1038/srep19029

Zhang, S., Guo, Y., Zhang, Y., Guo, J., Li, K., Fu, W., et al. (2021). Genome-wide identification, characterization and expression profiles of the CCD gene family in *Gossypium* species. *3 Biotech* 11:249. doi: 10.1007/s13205-021-02805-9

Zheng, X., Yang, Y., and Al-Babili, S. (2021). Exploring the diversity and regulation of apocarotenoid metabolic pathways in plants. *Front. Plant Sci.* 12:787049. doi: 10.3389/fpls.2021.787049

Zhou, Q., Li, Q., Li, P., Zhang, S., Liu, C., Jin, J., et al. (2019). Carotenoid cleavage dioxygenases: identification, expression, and evolutionary analysis of this gene family in tobacco. *Int. J. Mol. Sci.* 20:5796. doi: 10.3390/ijms20225796

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 August 2022
doi: 10.3389/fpls.2022.932966

Protein Engineering of a Germacrene A Synthase From *Lactuca sativa* and Its Application in High Productivity of Germacrene A in *Escherichia coli*

Rong Chen[1,2], Yuheng Liu[1], Shu Chen[1], Ming Wang[1], Yao Zhu[1], Tianyuan Hu[1], Qiuhui Wei[1], Xiaopu Yin[1] and Tian Xie[1]**

[1] Key Laboratory of Elemene Class Anti-cancer Chinese Medicine of Zhejiang Province, Engineering Laboratory of Development and Application of Traditional Chinese Medicine from Zhejiang Province, School of Pharmacy, Hangzhou Normal University, Hangzhou, China, [2] School of Public Health, Hangzhou Normal University, Hangzhou, China

OPEN ACCESS

Edited by:
Fangyuan Zhang,
Southwest University, China

Reviewed by:
Tianwen Wang,
Xinyang Normal University, China
Bei Bei Ge,
Institute of Plant Protection (CAAS), China

***Correspondence:**
Xiaopu Yin
yinxp@hznu.edu.cn
Tian Xie
tianxie@hznu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 30 April 2022
Accepted: 24 June 2022
Published: 11 August 2022

Citation:
Chen R, Liu Y, Chen S, Wang M, Zhu Y, Hu T, Wei Q, Yin X and Xie T (2022) Protein Engineering of a Germacrene A Synthase From Lactuca sativa and Its Application in High Productivity of Germacrene A in Escherichia coli. Front. Plant Sci. 13:932966. doi: 10.3389/fpls.2022.932966

Germacrene A (GA) is a key intermediate for the synthesis of medicinal active compounds, especially for β-elemene, which is a broad-spectrum anticancer drug. The production of sufficient GA in the microbial platform is vital for the precursors supply of active compounds. In this study, *Escherichia coli* BL21 Star (DE3) was used as the host and cultivated in SBMSN medium, obtaining a highest yield of FPP. The GA synthase from *Lactuca sativa* (LTC2) exhibited the highest level of GA production. Secondly, two residues involved in product release (T410 and T392) were substituted with Ser and Ala, respectively, responsible for relatively higher activities. Next, substitution of selected residues S243 with Asn caused an increase in activity. Furthermore, I364K-T410S and T392A-T410S were created by combination with the beneficial mutation, and they demonstrated dramatically enhanced titers with 1.90-fold and per-cell productivity with 5.44-fold, respectively. Finally, the production titer of GA reached 126.4 mg/L, and the highest productivity was 7.02 mg/L.h by the I364K-T410S mutant in a shake-flask batch culture after fermentation for 18 h. To our knowledge, the productivity of the I364K-T410S mutant is the highest level ever reported. These results highlight a promising method for the industrial production of GA in *E. coli*, and lay a foundation for pathway reconstruction and the production of valuable natural sesquiterpenes.

Keywords: germacrene A, β-elemene, germacrene A synthase, host optimization, site-directed mutagenesis

INTRODUCTION

In nature, sesquiterpenes represent the diverse C15 terpene classes of plant natural products. They have many important pharmacological, physiological, and ecological effects, and are widely used in many fields such as medicine, food, and cosmetics. Germacrene A (GA), which is easily bound to sesquiterpene cyclase, acts as an intermediate for the biosynthesis of various compounds, such as patchoulol and phytoalexins (de Kraker et al., 1998). GA can be converted to germacranes, further producing elemanes, which are an important group of sesquiterpenes widely occurring in nature. Meanwhile, GA can be oxidized into germacrene A carboxylic acid, which is further oxidized to produce the lactone ring, and then functionalized and/or cyclized to the respective guaianolide,

eudesmanolide, and germacranolide sesquiterpene lactones (Bouwmeester et al., 2002). GA itself is unstable *in vitro*. GA is particularly susceptible to perform Cope rearrangement toward β-elemene at high temperatures or during freezer storage (de Kraker et al., 1998). Moreover, the step of thermal conversion of GA to ß-elemene was successfully developed by Zhang et al. (2018) which makes the cost of β-elemene only 0.15% of that from plant extraction. β-elemene has been shown to exert activity against a wide range of cancers, including brain, breast, liver, lung, and prostate cancers, as well as drug-resistant tumors. Moreover, Chemo-radiotherapy can be greatly improved by the use of β-elemene, which has few adverse effects on normal tissue cells in clinical trials (Adio, 2009a,b; Wang et al., 2012; Chang et al., 2017). In addition, multidrug-resistant tumor-repopulating cells could be reversed with injections of β-elemene (Zhai et al., 2018, 2019). Because of the important roles of GA both as an intermediate or as an end product, its efficient production in platform microorganisms should be explored using the methods of metabolic engineering and synthetic biology.

Microorganisms have been used for the production of various commercial compounds for a long time. The most common microbial cell factories are the yeast *Saccharomyces cerevisiae,* the bacteria *Escherichia coli,* etc. Since the endogenous mevalonate (MVA) pathway and FPP synthase ERG20 in *S. cerevisiae* could produce farnesyl pyrophosphate (FPP) as the direct precursor for sesquiterpene synthesis (Borodina and Nielsen, 2014), the yeast has been used as a chassis to be engineered for industrial GA production. A yeast platform was established to produce GA with a time-space yield of 2.65 mg/L h in shake flasks (Hu et al., 2017). Subsequently, a time-space yield of 3.44 mg/L h GA was obtained by the best engineered yeast in shake flasks (Zhang et al., 2018). The engineered *S. cerevisiae* expressing a cyanobacterial germacrene a synthase mutant (AvGAS F23W) further improved the production of GA, reaching 4.30 mg/L h in shake-flask batch culture (Zhang et al., 2021). However, no attempts have been made to explore the production of GA in *E. coli*. As the most commonly used host strain in the application of bulk chemical production, *E. coli* has huge advantages such as clear physiological and genetic characteristics, fast growth in minimal salt medium, utilization of various substrates, and easy genetic modification (Liu et al., 2016). Therefore, in this study, we aimed to construct and engineer *E. coli* as cell factories for the highly efficient production of GA.

Escherichia coli generally produces C5 precursors through the endogenous DXP pathway, and then forms FPP, which is used for quinone and cell wall biosynthesis. However, this approach to produce isoprenoid precursors remains ineffective due to regulation mechanisms present in the native host (Yang et al., 2012). Instead, engineering of the heterologous MVA pathway into *E. coli* has been reported to supply sufficient FPP and further improve the productivity of sesquiterpenes (Martin et al., 2003). The evolution of germacrene A synthase (GAS) provides a feasible way to evaluate the catalytic efficiency of the conversion from FPP to GA. In this study, we first introduced the heterologous MVA pathway operons (pMM) in five host *E. coli* hosts, and then screened the optimal culture conditions, to obtain the highest FPP precursor yield. Secondly, screening of GAS from a variety of sources with a high GA yield was performed. Finally, site-directed mutagenesis of the key and selected residues was performed to generate beneficial mutants, and double mutations were combined to further improve the enzymatic activity (**Scheme 1**). The I364K-T410S was created and demonstrated a dramatically enhanced time-space-yield at 7.02 mg/L.h, which is 1.63∼2.65 fold higher than those previously reported in *S. cerevisiae* on a shake flask fermentation level. Moreover, this study is the first report of GA-production in *E. coli*.

MATERIALS AND METHODS

Strains and Medium

Five *E. coli* strains, including BL21 (DE3), BL21 Star (DE3), BW25113, JM109 (DE3), and BL21 trxB (DE3) were used for gene expression or GA production. Strains were cultivated in SBMSN, LB, and YM9 medium, respectively. The composition of YM9 medium was as follows (L^{-1}): yeast extract 2.0 g, Na_2HPO_4 6.0 g, KH_2PO_4 3.0 g, NaCl 0.5 g, NH_4Cl 1.0 g, $MgSO_4$ 1.0 mM, and $CaCl_2$ 0.1 mM (Lei et al., 2021). The composition of SBMSN medium was (L^{-1}) tryptone 12 g, yeast extract 24.0 g, KH_2PO_4 1.7 g, K_2HPO_4 11.42 g, $MgCl_2{\cdot}6H_2O$ 1.0 g, ammonium oxalate 1.42 g and Tween-80 2.0 g (Lei et al., 2021). Antibiotics, 25 mg/L chloramphenicol and 10 mg/L tetracycline were added when necessary.

Plasmids and Strains Construction

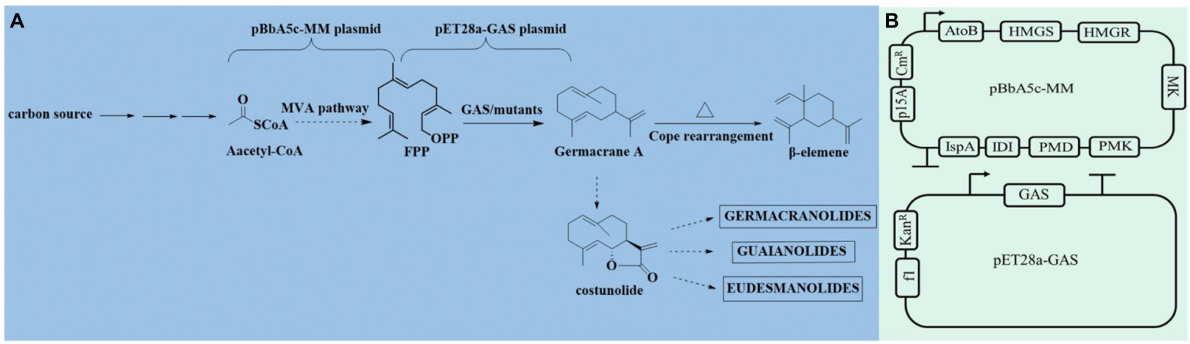
The pBbA5c-MevT–MBIS (abbreviated as pMM) was constructed by the Jay Keasling group and shared from Addgene[1] (Peralta-Yahya et al., 2011). The pMM, harboring the mevalonate pathway operons and ispA gene, was transformed into *E. coli* BL21 Star (DE3) to generate the engineering strain BL21 Star-MM. The resulting strain could convert acetyl-CoA into FPP. The pMM was respectively transformed into *E. coli* competent cells, including BL21(DE3), BL21 Star (DE3), BW25113, JM109 (DE3), and BL21 trxB (DE3).

To screen the candidate GAS, two literature reported GAS (CbGAS and LTC2) and seven putative GAS genes by BLASTp analysis in the NCBI database were selected, codon-optimized for *E. coli* expression and synthesized by Qinke Biotech Corporation (Nanjing, China). pET28a was used as the expression vector by ligation with these GASs, and the resulting plasmids (pET28a-GAS) were individually transformed into *E. coli* BL21 Star (DE3), generating the expression strains. For *in vivo* yield analysis, each GAS expression vector and pMM were transformed into *E. coli* BL21 Star-MM, generating the co-expressed strains.

GC Analysis of Farnesol Yield

For *in vivo* farnesol yield analysis, the strains carrying pMM were cultivated in the five different type of medium as previously prepared with 25 μg/mL of chloramphenicol at 37°C. When the cell density at 600 nm reached 0.6–0.8, 0.4 mM IPTG was

[1]www.addgene.org



SCHEME 1 | Biosynthesis of germacrene A and its precursors. FPP is produced from acetyl-CoA through the MVA pathway and then form β-elemene or other sesquiterpene lactones **(A)**. The plasmids pBbA5c-MM and pET28a-GASs **(B)**.

added to induce protein expression. All flasks were immediately added to 10% (vol/vol) n-dodecane. After 18 h of cultivation, the upper n-dodecane layer was collected, filtered by a 0.22 μm Millipore filter membrane, and diluted with n-hexane at 1: 20 for GC analysis.

GC analysis was performed using an Agilent 2010 series GC system with a flame ionization detector (FID). A GC column HP-5MS (30 m × 250 μm i.d. × 0.25 μm film thickness, Agilent) was employed for separating samples. The system was equilibrated for 2 min at 80°C prior to the subsequent analysis. The temperature program was set as follows: initiation at 80°C, ramp to 140°C at a rate of 40°C per min followed by a 2 min hold, ramp to 260°C at a rate of 40°C per min followed by a 1 min hold. The injector temperatures was set at 260°C while detector temperature was 305°C. Commercial farnesol at the highest purity available (> 99%) was used as control sample (Titan, Shanghai, China). Three biological replicates were performed for collection and GC analysis.

GC–MS Analysis of Germacrene A Synthase-Catalytic Products

For *in vivo* products analysis, the co-expressing strains, carried pMM and pET28a-GAS, were cultivated in SBMSN medium with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol at 37°C. IPTG induction and product enrichment were performed as described in the farnesol-producing section. Identification of the product was performed by GC–MS as previously described with a slight alternation (Chen et al., 2021). The temperature program was set the same as described above. Commercial β-elemene at the highest purity available (> 99%) was used as control sample (Titan, Shanghai, China).

Site-Directed Mutagenesis of LTC2

Mutagenesis was generated by PCR using the quick-change site-directed mutagenesis method (Costa et al., 1996). The plasmid pET28a-LTC2 was used as the template and the primers were listed in **Table 1**. The amplified products were digested with DpnI to remove the original template and then transformed into *E. coli* DH5α competent cells. The positive clones were screened by PCR amplification and sequenced for further verification. The resulting plasmids with mutation were transformed into *E. coli* BL21(DE3) or *E. coli* BL21 Star (DE3)-MM competent cells for further assay.

Protein Preparation and *in vitro* Enzyme Assay

The *E. coli* BL21(DE3) strains carrying pET28a-LTC2 or mutant sites were grown in LB medium with 50 μg/mL kanamycin at 37°C until the cell density at 600 nm reached 0.6, followed by induction with 0.1 mM IPTG overnight to produce N-terminal His_6-tagged recombinant proteins. The expression level of His-fusion proteins was detected by SDS–PAGE. The protein concentration was quantified using the Bradford method. Activity assays of LTC2 and mutants were performed in a volume of 1 mL containing 20 μg of protein, 0.1 M Tris-HCl (7.0), 3 mM $MgCl_2$, 0.1 M DTT, 0.15 M 50% glycerin, and 1 μL *E.E*-FPP (Sigma–Aldrich) (Bennett et al., 2002). The reaction mixture was conducted at 37°C for 1 h, and the products were collected by headspace solid phase micro extraction (SPME) fiber. After 60 min of sampling, the fiber was removed and immediately transferred to the injection port of the GC. GC analysis was performed as described in the farnesol-producing section with β-elemene as the standard.

Protein Modeling and Docking Analysis

The three-dimensional homology model of LTC2 was generated on the Swiss-Model server.[2] The X-ray structure of tobacco 5-epi-aristolochene synthase (PDB ID: 5EAU) was chosen as the best template (Starks et al., 1997). Docking of the FPP ligand into the model of the LTC2 active site cavity was performed *via* setting the parameters of "copy ligand from templates" by Discovery Studio 2020 software (Accelrys, San Diego, CA, United States). The GA was docked into the LTC2 structure by AutoDock Vina program (Trott and Olson, 2010). The generated model was geometrically refined by the Princeton TIGRESS 2.0 (Khoury et al., 2017). The final model was evaluated using Procheck program, and the lowest energy conformational model was chosen for docking studies (Chen et al., 2016).

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

TABLE 1 | Plasmids, strains, and primers used in this study.

Plasmids, strains and primers	Description	Source or reference
Plasmids		
pET-28a	Integrative plasmid,kan^r	Library stock
pET-28a-LTC2	kan^r,pET-28a-P_{T7}-LTC2-T_{T7}	This study
pET-28a-CbGAS	kan^r,pET-28a-P_{T7}-CbGAS-T_{T7}	This study
pET-28a-HbSAS	kan^r,pET-28a-P_{T7}-HbSAS-T_{T7}	This study
pET-28a-FtSAS	kan^r,pET-28a-P_{T7}-FtSAS-T_{T7}	This study
pET-28a-MmGDS	kan^r,pET-28a-P_{T7}-MmGDS-T_{T7}	This study
pET-28a-NpGAS	kan^r,pET-28a-P_{T7}-NpGAS-T_{T7}	This study
pET-28a-CmGAS	kan^r,pET-28a-P_{T7}-CmGAS-T_{T7}	This study
pET-28a-CfGAS	kan^r,pET-28a-P_{T7}-CfGAS-T_{T7}	This study
T410S	kan^r,pET-28a-P_{T7}-LTC2-T410S-T_{T7}	This study
T410V	kan^r,pET-28a-P_{T7}-LTC2-T410V-T_{T7}	This study
T410A	kan^r,pET-28a-P_{T7}-LTC2-T410A-T_{T7}	This study
T392A	kan^r,pET-28a-P_{T7}-LTC2-T392A-T_{T7}	This study
T392V	kan^r,pET-28a-P_{T7}-LTC2-T392V-T_{T7}	This study
T38S	kan^r,pET-28a-P_{T7}-LTC2-T38S-T_{T7}	This study
L58V	kan^r,pET-28a-P_{T7}-LTC2-L58V-T_{T7}	This study
A229S	kan^r,pET-28a-P_{T7}-LTC2-A229S-T_{T7}	This study
S243N	kan^r,pET-28a-P_{T7}-LTC2-S243N-T_{T7}	This study
I364K	kan^r,pET-28a-P_{T7}-LTC2-I364K-T_{T7}	This study
I492K	kan^r,pET-28a-P_{T7}-LTC2-I492K-T_{T7}	This study
S243N-T410S	kan^r,pET-28a-P_{T7}-LTC2-S243N-T410S-T_{T7}	This study
I364K-T410S	kan^r,pET-28a-P_{T7}-LTC2-I364K-T410S-T_{T7}	This study
T392A-T410S	kan^r,pET-28a-P_{T7}-LTC2-T392A-T410S-T_{T7}	This study
T392V-T410S	kan^r,pET-28a-P_{T7}-LTC2-T392V-T410S-T_{T7}	This study
NKS	kan^r,pET-28a-P_{T7}-LTC2-S243N-I364K-T410S-T_{T7}	This study
KAS	kan^r,pET-28a-P_{T7}-LTC2-I364K-T392A-T410S-T_{T7}	This study
pBbA5c-MM	Cm^r,pBbA5c-P_{LacUV5}-atoB-HMGS(CO)-HMGR(CO)-MK(CO)-PMK(CO)-PMD-idi-ispA-T_{rrnbT1}	Peralta-Yahya et al., 2011
Strains		
E. coli BL21(DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3)	Tsingke Co.
E. coli BL21 Star(DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm rne 131*(DE3)	Tsingke Co.
E. coli BW25113	F$^-$, DE(*ara*D-*ara*B)567, *lac*Z4787(*del*)::*rrn*B-3, LAM-, *rph*-1, DE(*rha*D-*rha*B)568, *hsd*R514	Huayueyang Co.
E. coli JM109(DE3)	*end*A1 *rec*A1 *gyr*A96 thi-1 *hsd*R17 (r_k^-,m_k^+) *rel*A1 *sup*E44 D (*lac-pro*AB) [F' *tra*D36 *pro*AB *laqI*^{q}*Z*ΔM15](DE3)	Weidi Co.
E. coli BL21 trxB(DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm trxB*15::kan(DE3)	Huayueyang Co.
Primers		
T410-F	aaaatggtctgattnbwgcgcatataatg	This study
T410-R	wvnaatcagaccatttttttcatattccgg	This study
T392A/V-F	agaagcagaatgggynaatagcgttatg	This study
T392A/V-R	nrcccattctgcttcttccagataaccacg	This study
T38S-F	gatcgttttctgagctttagcctggataat	This study
T38S-R	ctaaagctcagaaaacgatcaccccaaa	This study
L58V-F	aagcaccgaaagaagaagtgcgtcgtct	This study
L58V-R	cttcttctttcggtgcttccattgcttttgcat	This study
A229S-F	ttatagcgaagaatgtagcacccatgaat	This study
A229S-R	gctacattcttcgctataattgctaaaatac	This study
S243N-F	ctggcaaaactgcattttaactatctggaact	This study
S243N-R	ttaaaatgcagttttgccagtttcagcagac	This study
I364K-F	aaaaacagctggcaaaagaaggtcgtg	This study
I364K-R	ttttgccagctgtttttccagttctgcat	This study
I492K-F	gcaattgatgaactgaaaaaaatgattgaaa	This study
I492K-R	ttttttcagttcatcaattgcttcttttcgc	This study

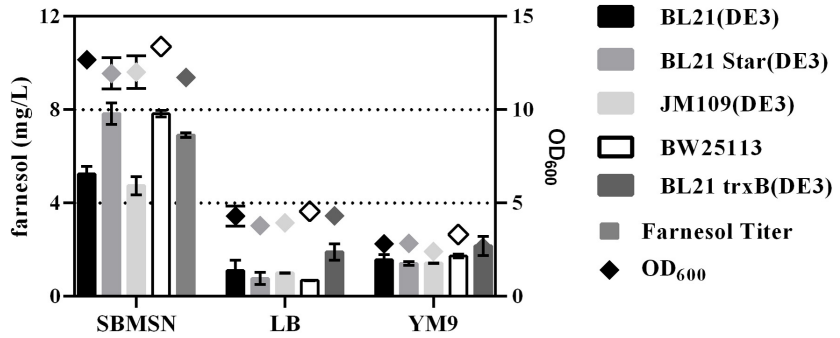


FIGURE 1 | Effect of medium and strains on farnesol yield and OD_{600} in *E. coli* hosts carrying pMM.

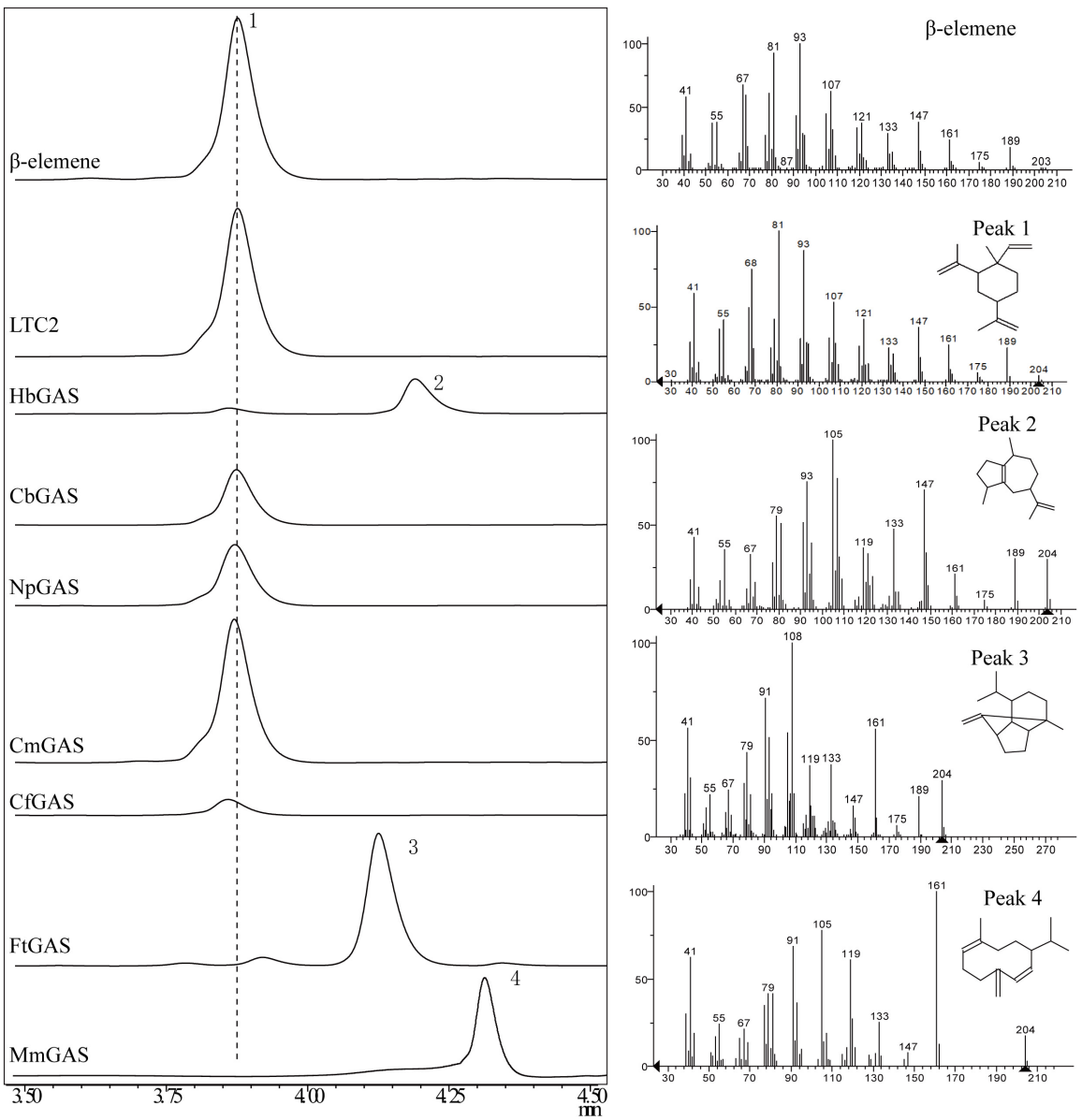


FIGURE 2 | GC-MS analysis of products in *E. coli* BL21Star (DE3) pMM- GASs. Peaks 1, β-elemene; 2, guaiene; 3, sativene; 4, germacrene D.

RESULTS

Effect of Strains and Medium on Farnesol Yield

The expression level of heterologous gene was affected by many factors such as host strains, cell growth, inducer concentration, etc. (Yin et al., 2015). It is an effective way to improve exogenous gene expression by optimizing the host and culture medium. Because FPP can be spontaneously hydrolyzed into farnesol by endogenous phosphatase or pyrophosphatase (Wang et al., 2016), the level of farnesol was used to evaluate the ability of FPP production in *E. coli*. As shown in **Figure 1**, five different host strains carrying pMM could produce more farnesol in SBMSN medium than those in LB and YM9 medium. For example, the engineered strain BL21(DE3)-MM in SBMSN medium produced 4.88-fold and 3.34-fold farnesol than those in LB and YM9 medium, respectively. Farnesol exhibited higher levels in BL21(DE3) Star, BW25113, and BL21 trxB (DE3) strains grown in SBMSN medium. Overall, The BL21Star (DE3) strain showed the highest per-cell productivity of farnesol yield (1.55 mg/g. DCW). *E. coli* BL21 Star (DE3) is characterized by containing a mutated rne131 gene encoding ribonuclease E (RNaseE), which can enhance the stability of messenger RNA, which in turn elevates the expression level of target proteins (Lee et al., 2020). Hence, the *E. coli* BL21 Star (DE3) carrying the plasmid pMM was used as a chassis and cultivated in SBMSN medium to evaluate the yield of GA.

Screening of Germacrene A Synthase

For selecting an efficient GAS, two literature reported GASs (LTC2 and CbGAS) were overexpressed in *E. coli* BL21 Star (DE3)-MM engineered strain. LTC2 was identified from *Lactuca sativa* while CbGAS was identified from the bacterium *Nostoc* sp. PCC 7120 (Bennett et al., 2002; Agger et al., 2008). To mine candidate GAS, BLSATP was performed with CbGAS in the Genbank database. Finally, six genes with 37–92% amino-acid sequence similarities were selected for yield analysis, including *HbGAS* (*Hassallia byssoidea*), *MmGAS* (*Methyloglobulus morosus*), *FtGAS* (*Fischerella thermalis*), *NpGAS* (*Nostoc parmelioides*), *CmGAS* (*Calothrix membranacea*), and *CfGAS* (*Calothrix* sp. FI2-JRJ7). GC–MS confirmed that LTC2, CbGAS, NpGAS, CmGAS, and CfGAS could produce β-elemene as the sole product in the engineered *E. coli*. The main product of HbGAS was guaiene, concomitant with a small amount of β-elemene. FtGAS and MmGAS generated other sesquiterpenes, sativene and germacrene D, respectively (**Figure 2**). Among them, LTC2 obtained the highest yield of β-elemene with 75.45 mg/L after 18 h of fermentation and growth (OD_{600} = 12.1) (**Figure 3**). However, LTC2 showed no β-elemene-producing capability in yeast cells (Zhang et al., 2021). We inferred that the mildly acidic intracellular environment of yeast cells led to low activity for LTC2, consequently, the effect of pH on LTC2 was performed. At the pH conditions (about 6.8) in yeast cytosol, LTC2 maintained only 40% relative activity. The optimal pH of LTC2 was 7.5 (**Supplementary Figure 1**). SDS–PAGE confirmed that these bacterial origin GASs were largely expressed as inclusion bodies in *E. coli*, resulting in a low content of soluble recombinant enzymes and eventually a low β-elemene yield. LTC2 could be expressed as a soluble protein with a few amount of inclusion in the whole cell lysates (**Supplementary Figure 2**). Hence, LTC2 was selected for further research.

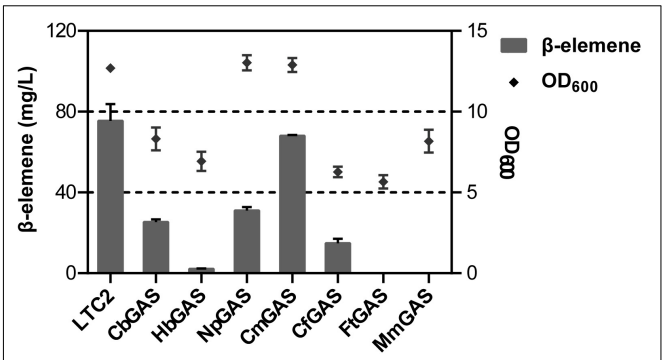


FIGURE 3 | β-elemene titer and OD_{600} analysis of *E. coli* BL21Star (DE3) pMM- GASs.

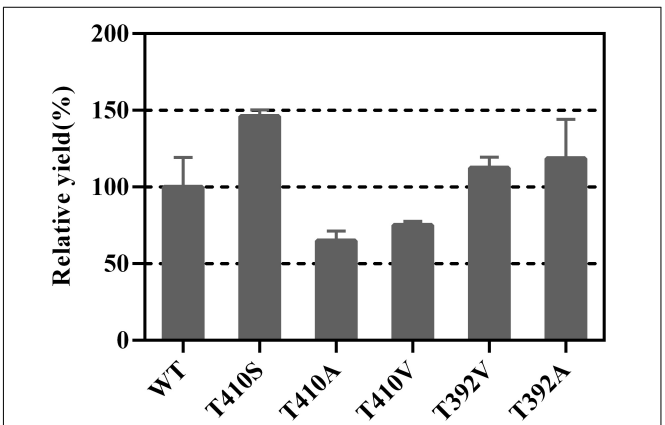


FIGURE 4 | The relative yields of T410 and T392 mutants compared with WT.

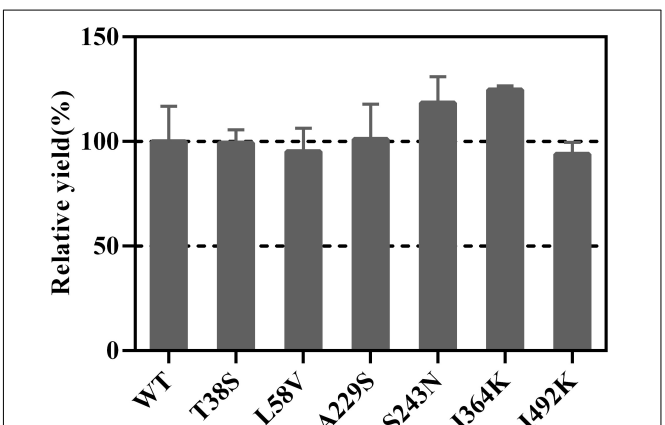


FIGURE 5 | The relative yields of selected mutants compared with WT.

TABLE 2 | The selected residue positions in LTC2 and plant-derived GASs.

Species	Name	Amino acid position					
		38	58	229	243	364	492
Lactuca sativa	LTC2	T	L	A	S	I	I
Barnadesia spinosa	BsGAS1	S	V	S	N	K	K
Helianthus annuus L.	HaGAS1	S	V	S	N	K	K
	HaGAS2	S	V	S	N	K	K
Cichorium intybus	CiGASsh	S	V	S	N	K	K
	CiGASlo	S	V	S	N	K	K
Taraxacum officinale	ToGAS2	S	V	S	N	K	K

T410S and T392A Mutants Improve the Germacrene A Yield

First, we sorted some mutation sites with enhancing activity in sesquiterpene synthase family. For example, the threonine residue at the 399 position of Artemisia annua amorpha-4,11-diene synthase (AaADS) was replaced with serine (T399S), resulting in a catalytic efficiency improvement (Li et al., 2013). The leucine residue at the 381 position of *A. annua* α-bisabolol synthase (AaBOS) was substituted with alanine (L381A), exhibiting an approximate twofold increase in the production of γ-humulene compared with the wide type (WT) (Li et al., 2013). Corresponding to T399 of AaADS and L381 of AaBOS, T410 and T392 of LTC2 were chosen to be replaced with Ser and Ala, respectively. Ala and Val with a small side chain may lead to the easy release of product. Hence, T410A, T410V, T392A, and T392V were constructed. All mutants of T410 and T392 showed primarily soluble expression in *E. coli* BL21 Star (DE3) (**Supplementary Figure 3**). *In vitro* enzymatic analysis verified that the relative yield of T410S raised to 146% compared with the WT, while the T410A and T410V mutants partially lost their activities. The T392A and T392V had slightly increasing activities toward FPP (**Figure 4**). Overall, the T410S and T392A exhibited relatively increasing activities compared with the WT.

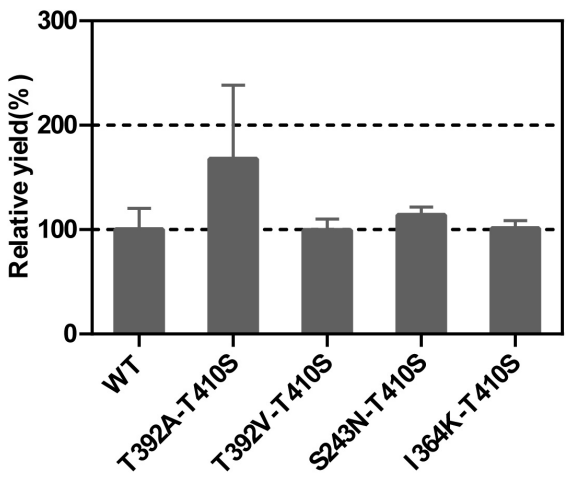


FIGURE 6 | The relative yields of LTC2 double mutants compared with WT.

Effects of Selected Mutants on Germacrene A Yield

Mutation of residues involved in the active center or conserved sites was disregarded, which possibly led to enzyme inactivation, product profile alteration and catalytic efficiency lowering (Bibi et al., 2020). Plasticity residues in or around the active center may increase the product diversity or improve the catalytic efficiency (Yoshikuni et al., 2006; O'Maille et al., 2008). However, the mutation of GAS exists some difficulties, such as the complex catalytic mechanisms, the lack of appropriate crystal structure, and an effective high-throughput screening approach.

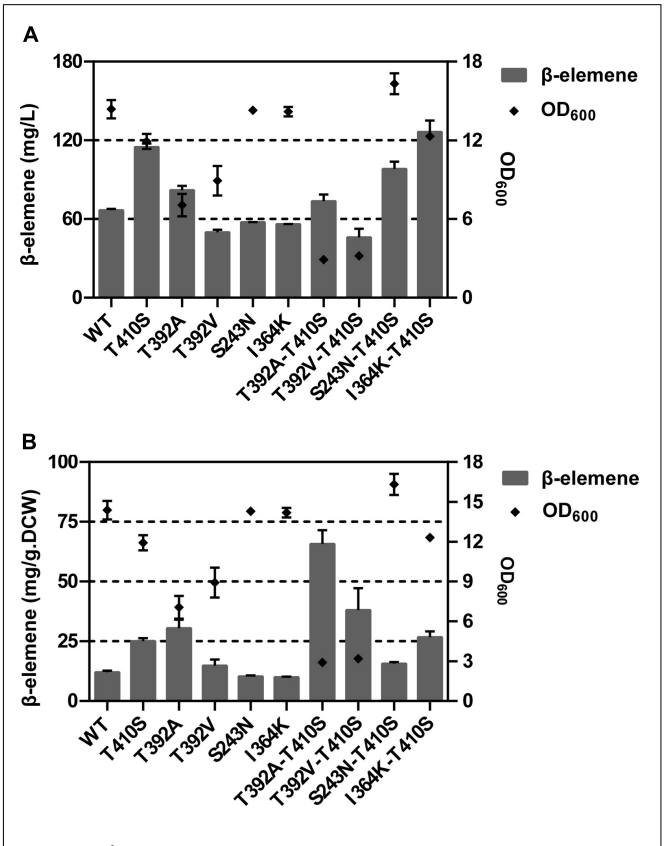


FIGURE 7 | The titer **(A)** and per-cell productivity **(B)** of β-elemene in *E. coli* BL21 Star (DE3) pMM – LTC2 mutants.

Hence, those residues, which are conserved but non-critical for catalysis in GAS, were chosen for mutation. Sequence alignment was performed using LTC2 and literature-reported plant-derived GASs, including BsGAS1, HaGAS1, HaGAS2, CiGASsh, CiGASlo, and ToGAS2 (Bouwmeester et al., 2002; Göpfert et al., 2009; Huber et al., 2016; Nguyen et al., 2016).

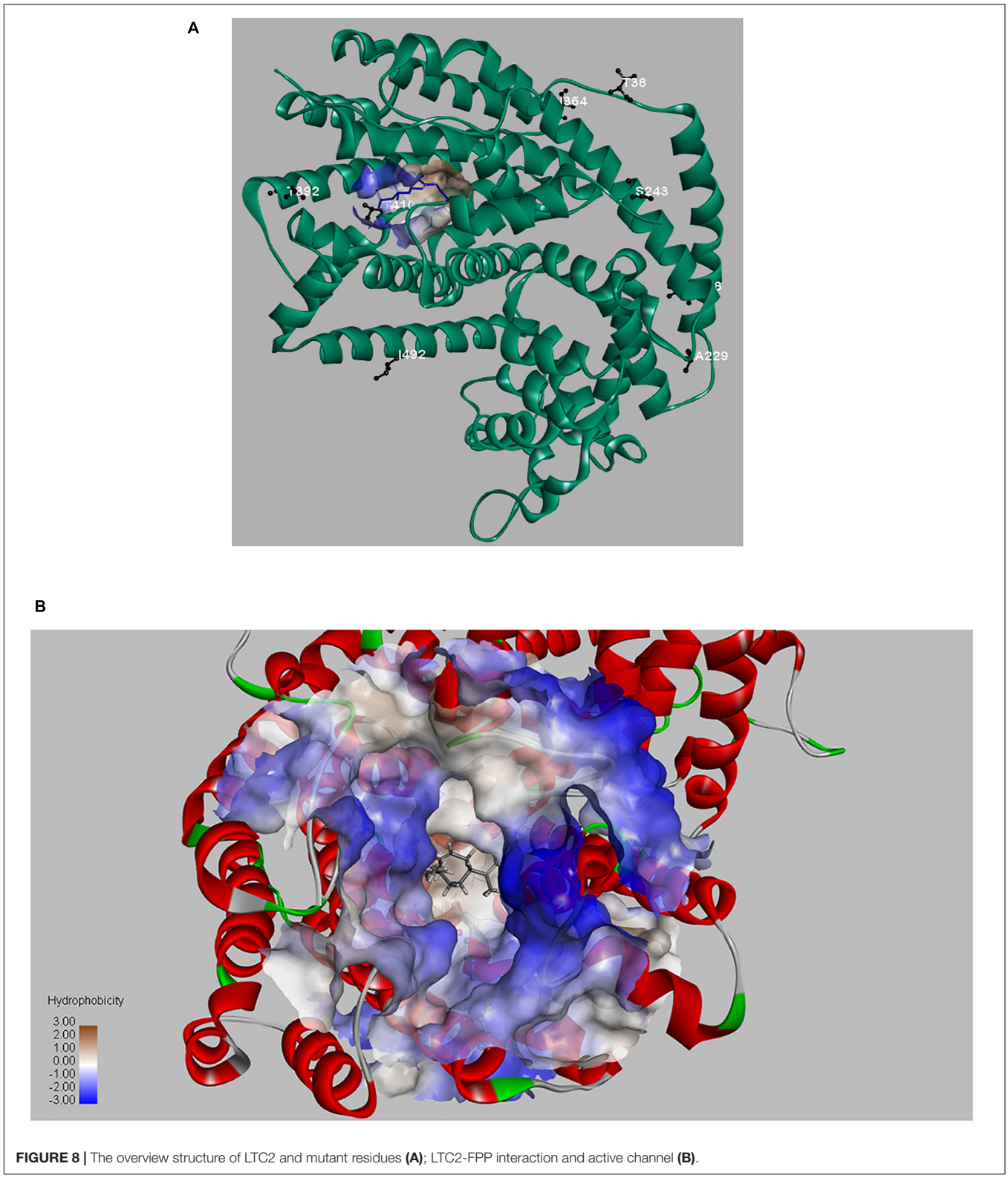


FIGURE 8 | The overview structure of LTC2 and mutant residues **(A)**; LTC2-FPP interaction and active channel **(B)**.

TABLE 3 | Comparison of data for the production of β-elemene in engineered strains.

Chassis	GAS	Titer (mg/L)	Time (h)	Productivity (mg/L h)	References
Saccharomyces cerevisiae	AvGAS-F23V	309.8	72	4.30	Zhang et al., 2021
Saccharomyces cerevisiae	LTC2	190.7	72	2.65	Hu et al., 2017
Saccharomyces cerevisiae	LTC2	469	144	3.26	Zhang et al., 2018
E. coli	LTC2-I364K-T410S	126.4	18	7.02	This research

The sequence alignment revealed that six residues were identical at the corresponding positions in these GASs but distinct in LTC2 (**Table 2**). Therefore, the T38S, L58V, A229S, S243N, I364K, and I492K mutants were taken into consideration for mutation. All mutants were expressed solubly in *E. coli* (**Supplementary Figure 4**) and retained activities toward FPP *in vitro*. Among them, I364K and S223N were slightly increased in relative yield compared with WT which were chosen for further mutation (**Figure 5**).

Combination Analysis of Mutation Sites

It's an effective way to improve enzyme performance by combinant mutations, which often produce an additive effect (Wang et al., 2015). Therefore, we used the T410S mutant as the template and conducted site-directed mutagenesis to generate double mutants that further improved the activity of LTC2. The *E. coli* BL21 Star (DE3) carrying mutant plasmids were expressed solubly in *E. coli* (**Supplementary Figure 5**). *In vitro* enzymatic analysis showed that the T392A-T410S mutant demonstrated a significantly increased relative yield while others did not compared with the WT (**Figure 6**).

In vivo Evaluation of Mutants for Germacrene A Production

Next, the enrichment in n-dodecane and quantitative analysis of volatile GA by GC after 18 h cultivation was performed for all mutants. The results of the combinatorial mutagenesis showed that the GA yield of I364K-T410S was 126.4 mg/L, while that of WT was 66.7 mg/L. The I364K-T410S was 1.90 time in the amount of β-elemene production, compared with WT. We found that these engineered strains after 18 h of induced culture exhibited a large growth difference, so the value of β-elemene yield was also reported as micrograms per gram of dried weight (mg/g DCW). The T392A-T410S mutant exhibited the highest yield of β-elemene with 65.7 mg/g DCW, a 5.44 fold of the WT (**Figure 7**). These results demonstrated the synergistic effects of beneficial mutations.

DISCUSSION

In recent years, many protein engineering approaches have been developed, such as rational design, directed evolution, *de novo* design, computational-aided approaches etc., which have enabled novel proteins with improved properties, including better enzymatic properties, better stability, improved catalytic activity or advanced applications (Sinha and Shukla, 2019). Because of the complexity of enzymatic catalysis, especially the terpene synthase (TPS), a combination of protein engineering approaches may be advantageous (Yoshikuni et al., 2006; Fang et al., 2017). In our research, improved LTC2 mutants were obtained by protein engineering approach. To better understand the positive effects of these selected sites on LTC2 function, we performed structural simulation analysis (**Figure 8A**). The three-dimensional homology model of LTC2 was obtained by comparative modeling using the tobacco 5-epi-aristolochene synthase (TEAS, PDB ID: 5EAU). LTC2 shared 40.6% sequence-similarity with TESA. The obtained models of LTC2 was structurally similar to TEAS, comprising two structural domains: the C-terminal catalytic domain and the N-terminal domain with unknown function (Starks et al., 1997; Rudolf and Chang, 2020). These highly conserved motifs of DDXXD, RXR, and NSE/DTE of terpene synthase family were found in its primary sequence. As reported in TPS, the DDXXD motif binds with metal Mg^{2+}, playing a role in promoting multiple orientations of the substrate alkyl chain. The RXR motif, located in the entrance of the enzymatic pocket, acts in protecting the hydrophobic active center from water molecules invading (Liu et al., 2018). The catalytic mechanism of terpene synthase is complicated and may include the following steps: enzyme-substrate binding and folding, the generation and stabilization of high-energy carbocations, the acidic/basic catalysis to form specific skeletons or metabolites, as well as product release (Li et al., 2013). Moreover, the last step of product release is often a rate-limiting step (Rising et al., 2000). Deprotonation is a general way to speed up the formation of olefinic products for most terpene synthases (TPS) (Fang et al., 2017). In this study, the substitution T of 410 with S in LTC2 led to a 46% increase in catalytic activity and a 72% increase of GA yield. The serine residue is structurally similar to the threonine residue. Both have a polar side chain and a hydroxy group, but the serine is smaller than the threonine in structure. Hence, the change from threonine to serine in LTC2 may act a favored effect in the last reaction step-deprotonation process. Moreover, the serine has a more polar side chain ($–CH_2OH$) than the threonine ($–CH(OH)CH_3$), resulting in a more hydrophilic pocket environment to accelerate dissociation rate of hydrophobic product (Li et al., 2013). As we anticipated, the threonine at 392 of LTC2 was located on the rim of active pocket and may affect the conformation of the FPP entry or GA exit channel. The replacement of T392 with alanine may decrease the steric hindrance on the rim of FPP entrance or GA exit of the pocket (**Figure 8B**). Thus, T410S and T392A of LTC2 exerted favorable effects (Li et al., 2013).

Usually, functionally important amino acids are evolutionarily conserved in the primary structure. However, mutation of these

key residues in or around the active center would result in inactivation of enzyme (Zabala et al., 2012; Bibi et al., 2020). Therefore, substitution of residues that are conserved but non-critical to catalysis may exert a beneficial effect. Based on multiple sequence alignment of LTC2 and other GA-produced GASs, six residues were found to be identical at the corresponding positions in literature reported GASs but distinct in LTC2. Ultimately, only purified S243N and I364K mutants exerted slightly favorable effects *in vitro*. As shown in **Figure 8A**, T38, L58, A229, and I492 were located in the N-terminal α-barrel domain and far from the active center, leading to no effect on the enzymatic activity. S243N and I364K were predicted to be located in the C-terminal α-helix domain but still far from the enzyme active center (**Figure 8A**). However, their elevated levels were limited *in vitro*, even with a minor decrease in *E. coli* BL21 Star (DE3)-MM strains. Furthermore, a combination analysis of beneficial mutants was performed. As a result, the T392A-T410S and I364K-T410S presented increased productivities and titers of GA, respectively. Actually, the triple mutants of KAS(I364K-T293A-T410S) and NKS(S264N-I364K-T410S) were also constructed but showed unfavorable titers of GA, compared with their corresponding double mutants (**Supplementary Figure 6**).

The combined analysis of GA titer and OD_{600} revealed that these strains expressed T392A alone or together with T410S showed lower biomass. Therefore, the T392A-T410S exhibited the highest per-cell productivity of 65.8 mg/g DCW. Whereas, I364K-T410 showed the highest product yield of 126.0 mg/L after 18 h of fermentation and a similar biomass compared with WT. Overall, the per-cell productivity of the strain carrying T392A-T410S exhibited the highest level of GA, probably results of improvement in thermo-stability or catalytic efficiency that need to be further verified by experiments. Furthermore, the crystal structures of WT and mutant enzymes alone or with FPP substrate/GA product should be performed to explain the mechanism.

Germacrene A is a key intermediate for the synthesis of various active compounds, especially for β-elemene, a broad-spectrum anticancer drug. The thermal conversion of GA to β-elemene undergoes very easily. At present, the only way to obtain elemene is extraction from the ginger plant *Curcuma wenyujin*. However, these factors, such as low content, complex extraction procedure and high extraction cost, have largely restricted its practical applications. Consequently, we used host/medium optimization and protein engineering approaches to generate an engineered *E. coli* BL21 Star (DE3)-MM-$LTC2_{mutant}$ strain to yield GA.

First, to achieve a higher FPP production, we developed the extraction fermentation with n-dodecane as overlay, and optimized cultivation for 18 h and IPTG concentration at 0.4 mM. Under this condition, the strain BL21(DE3) Star carrying pMM produced the highest titer of about 7.83 mg/L FPP in SBMSN medium. On the contrary, the engineering strain BL21(DE3) Star-pMM had a lower biomass and produced a low titer of FPP in LB or YM9 medium. An integrated consideration of these factors, the strain BL21(DE3) Star-pMM was used as the best host and cultivated in SBMSN medium, aiming to supply sufficient precursors. Finally, co-expressed with GAS mutant, the strain BL21(DE3) Star-pMM-$GAS_{I364K-T410S}$ mutant exhibited a time-space-yield of 7.02 mg/L.h, which is 1.63∼2.65 fold of those previously reported in *S. cerevisiae* on a shake flask fermentation level (**Table 3**). Although the final titer of the engineered strain was lower than that of *S. cerevisiae*, more biotechnologies and engineering strategies should be developed for increasing GA production. Usually, the successful implementation of new bioprocesses requires the maximization of product yield, titer, and productivity. However, these performance indicators cannot be maximized simultaneously due to the inherent trade-off between the biomass and the target product (Klamt et al., 2018). Some optimization strategies have been successfully established. For example, high cell density fermentation could increase biomass reaching an OD_{600} of 160–180, using a chemically defined fermentation medium for *E. coli* growth (Yang et al., 2020; Perl et al., 2022). Alternatively, two-stage fermentation, composed of cell growth and production, makes it easy to achieve the improved productivity of biotechnological production processes (Klamt et al., 2018). It can be expected that the yield of GA could be further improved to a very high level by these optimizing strategies.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

RC: conceptualization, methodology, writing – original draft, and writing – review and editing. YL: methodology, formal analysis, and data curation. SC, MW, and YZ: methodology. TH: writing – original draft. QW: formal analysis. XY: conceptualization, supervision, and funding acquisition. TX: conceptualization, supervision, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (82173919) and the Natural Science Foundation of Zhejiang Province (LY21C050004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.932966/full#supplementary-material

REFERENCES

Adio, A. M. (2009a). (–)-trans-β-Elemene and related compounds: occurrence, synthesis, and anticancer activity. *Tetrahedron* 65, 5145–5159. doi: 10.1016/j.tet.2009.04.062

Adio, A. M. (2009b). Germacrenes A-E and related compounds: thermal, photochemical and acid induced transannular cyclizations. *Tetrahedron* 65, 1533–1552. doi: 10.1016/j.tet.2008.11.050

Agger, S. A., Lopez-Gallego, F., Hoye, T. R., and Schmidt-Dannert, C. (2008). Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *J. Bacteriol.* 190, 6084–6096. doi: 10.1128/JB.00759-08

Bennett, M. H., Mansfield, J. W., Lewis, M. J., and Beale, M. H. (2002). Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). *Phytochemistry* 60, 255–261. doi: 10.1016/s0031-9422(02)00103-6

Bibi, N., Ullah, A., Darwesh, L., Khan, W., Khan, T., Ullah, K., et al. (2020). Identification and computational analysis of novel TYR and SLC45A2 gene mutations in pakistani families with identical non-syndromic oculocutaneous albinism. *Front. Genet.* 11:749. doi: 10.3389/fgene.2020.00749

Borodina, I., and Nielsen, J. (2014). Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* 9, 609–620. doi: 10.1002/biot.201300445

Bouwmeester, H. J., Kodde, J., Verstappen, F. W. A., Altug, I. G., de Kraker, J. W., and Wallaart, T. E. (2002). Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiol.* 129, 134–144. doi: 10.1104/pp.001024

Chang, Z. W., Gao, M., Zhang, W. J., Song, L. J., Jia, Y. X., and Qin, Y. R. (2017). Beta-elemene treatment is associated with improved outcomes of patients with esophageal squamous cell carcinoma. *Surg. Oncol.* 26, 333–337. doi: 10.1016/j.suronc.2017.07.002

Chen, R., Deng, J., Lin, J. P., Yin, X. P., Xie, T., Yang, S. L., et al. (2016). Assessing the stereoselectivity of carbonyl reductases toward the reduction of OPBE and docking analysis. *Biotechnol. Appl. Biochem.* 63, 465–470. doi: 10.1002/bab.1397

Chen, R., Wei, Q. H., Liu, Y. H., Wei, X., Chen, X. B., Yin, X. P., et al. (2021). Transcriptome sequencing and functional characterization of new sesquiterpene synthases from *Curcuma wenyujin*. *Arch. Biochem. Biophys.* 709:108986. doi: 10.1016/j.abb.2021.108986

Costa, G. L., Bauer, J. C., McGowan, B., Angert, M., and Weiner, M. P. (1996). Site-directed mutagenesis using a rapid PCR-based method. *Methods Mol. Biol.* 57, 239–248. doi: 10.1385/0-89603-332-5:239

de Kraker, J. W., Franssen, M. C., de Groot, A., Konig, W. A., and Bouwmeester, H. J. (1998). (+)-Germacrene A biosynthesis. The committed step in the biosynthesis of bitter sesquiterpene lactones in chicory. *Plant Physiol.* 117, 1381–1392. doi: 10.1104/pp.117.4.1381

Fang, X., Li, J. X., Huang, J. Q., Xiao, Y. L., Zhang, P., and Chen, X. Y. (2017). Systematic identification of functional residues of *Artemisia annua* amorpha-4,11-diene synthase. *Biochem. J.* 474, 2191–2202. doi: 10.1042/BCJ20170060

Göpfert, J., Bülow, A., and Spring, O. (2009). Identification and functional characterization of a new sunflower germacrene A synthase (HaGAS3). *Nat. Prod. Commun.* 5, 705–715. doi: 10.1186/1745-6150-7-18

Hu, Y. T., Zhou, Y. J., Bao, J. C., Huang, L. Q., Nielsen, J., and Krivoruchko, A. (2017). Metabolic engineering of Saccharomyces cerevisiae for production of germacrene A, a precursor of beta-elemene. *J. Ind. Microbiol. Biotechnol.* 44, 1065–1072. doi: 10.1007/s10295-017-1934-z

Huber, M., Epping, J., Gronover, C. S., Fricke, J., Aziz, Z., Brillatz, T., et al. (2016). A latex metabolite benefits plant fitness under root herbivore attack. *PLoS Biol.* 14:e1002332. doi: 10.1371/journal.pbio.1002332

Khoury, G. A., Smadbeck, J., Kieslich, C. A., Koskosidis, A. J., Guzman, Y. A., Tamamis, P., et al. (2017). Princeton_TIGRESS 2.0: high refinement consistency and net gains through support vector machines and molecular dynamics in double-blind predictions during the CASP11 experiment. *Proteins* 85, 1078–1098. doi: 10.1002/prot.25274

Klamt, S., Mahadevan, R., and Hädicke, O. (2018). When do two-stage processes outperform one-stage processes? *Biotechnol. J.* 13:1700539. doi: 10.1002/biot.201700539

Lee, H. J., Cho, A., Hwang, Y. J., Park, J. B., and Kim, S. K. (2020). Engineering of a microbial cell factory for the extracellular production of catalytically active phospholipase A2 of *Streptomyces violaceoruber*. *J. Microbiol. Biotechnol.* 30, 1244–1251. doi: 10.4014/jmb.2001.01052

Lei, D. W., Qiu, Z. T., Wu, J. H., Qiao, B., Qiao, J. J., and Zhao, J. R. (2021). Combining metabolic and monoterpene synthase engineering for de novo production of monoterpene alcohols in *Escherichia coli*. *ACS Synth. Biol.* 10, 1531–1544. doi: 10.1021/acssynbio.1c00081

Li, J. X., Fang, X., Zhao, Q., Ruan, J. X., Yang, C. Q., Wang, L. J., et al. (2013). Rational engineering of plasticity residues of sesquiterpene synthases from Artemisia annua: product specificity and catalytic efficiency. *Biochem. J.* 451, 417–426. doi: 10.1042/BJ20130041

Liu, P. P., Zhu, X. N., Tan, Z. G., Zhang, X. L., and Ma, Y. H. (2016). Construction of *Escherichia Coli* cell factories for production of organic acids and alcohols. *Adv. Biochem. Eng. Biotechnol.* 155, 107–140. doi: 10.1007/10_2014_294

Liu, Y., Chen, J. G., Qian, J. Y., Lin, H., Sun, N., and Huang, Z. N. (2018). Evolutionary analysis and structural characterization of *Aquilaria sinensis* sesquiterpene synthase in agarwood formation: a computational study. *J. Theor. Biol.* 456, 249–260. doi: 10.1016/j.jtbi.2018.08.006

Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802. doi: 10.1038/nbt833

Nguyen, T. D., Faraldos, J. A., Vardakou, M., Salmon, M., O'Maille, P. E., and Ro, D. K. (2016). Discovery of germacrene A synthases in *Barnadesia spinosa*: the first committed step in sesquiterpene lactone biosynthesis in the basal member of the Asteraceae. *Biochem. Biophys. Res. Commun.* 479, 622–627. doi: 10.1016/j.bbrc.2016.09.165

O'Maille, P. E., Malone, A., Dellas, N., Hess, B. A. Jr., Smentek, L., Sheehan, I., et al. (2008). Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* 4, 617–623. doi: 10.1038/nchembio.113

Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* 2:483. doi: 10.1038/ncomms1494

Perl, A., Dalton, H., Yoo, Y. J., and Koffas, M. A. G. (2022). Methods for the development of recombinant microorganisms for the production of natural products. *Methods Mol. Biol.* 2396, 1–17. doi: 10.1007/978-1-0716-1822-6_1

Rising, K. A., Starks, C. M., Noel, J. P., and Chappell, J. (2000). Demonstration of germacrene A as an intermediate in 5-epi-aristolochene synthase catalysis. *J. Am. Chem. Soc.* 122, 1861–1866. doi: 10.1021/ja993584h

Rudolf, J. D., and Chang, C. Y. (2020). Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. *Nat. Prod. Rep.* 37, 425–463. doi: 10.1039/c9np00051h

Sinha, R., and Shukla, P. (2019). Current trends in protein engineering: updates and progress. *Curr. Protein. Pept. Sci.* 20, 398–407. doi: 10.2174/1389203720666181119120120

Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277, 1815–1820. doi: 10.1126/science.277.5333.1815

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Wang, B., Peng, X. X., Sun, R., Li, J., Zhan, X. R., Wu, L. J., et al. (2012). Systematic review of β-elemene injection as adjunctive treatment for lung cancer. *Chin. J. Integr. Med.* 18, 813–823. doi: 10.1007/s11655-012-1271-9

Wang, C. L., Park, J. E., Choi, E. S., and Kim, S. W. (2016). Farnesol production in *Escherichia coli* through the construction of a farnesol biosynthesis pathway-application of PgpB and YbjG phosphatases. *Biotechnology. J.* 11, 1291–1297. doi: 10.1002/biot.201600250

Wang, H. L., Gao, W. Y., Sun, H. H., Chen, L. F., Zhang, L. J., Wang, X. D., et al. (2015). Protein engineering of a nitrilase from *Burkholderia cenocepacia* J2315 for efficient and enantioselective production of (R)-o-chloromandelic acid. *Appl. Environ. Microbiol.* 81, 8469–8477. doi: 10.1128/AEM.02688-15

Yang, D., Park, S. Y., Park, Y. S., Eun, H., and Lee, S. Y. (2020). Metabolic engineering of *Escherichia coli* for natural product biosynthesis. *Trends Biotechnol.* 38, 745–765. doi: 10.1016/j.tibtech.2019.11.007

Yang, J. M., Xian, M., Su, S. Z., Zhao, G., Nie, Q. J., Jiang, X. L., et al. (2012). Enhancing production of bio-isoprene using hybrid MVA pathway and

isoprene synthase in *E. coli*. *PLoS One* 7:e33509. doi: 10.1371/journal.pone.0033509

Yin, H., Zhuang, Y. B., Li, E. E., Bi, H. P., Zhou, W., and Liu, T. (2015). Heterologous biosynthesis of costunolide in *Escherichia coli* and yield improvement. *Biotechnol. Lett.* 37, 1249–1255. doi: 10.1007/s10529-015-1784-6

Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006). Designed divergent evolution of enzyme function. *Nature* 440, 1078–1082. doi: 10.1038/nature04607

Zabala, A. O., Cacho, R. A., and Tang, Y. (2012). Protein engineering towards natural product synthesis and diversification. *J. Ind. Microbiol. Biotechnol.* 39, 227–241. doi: 10.1007/s10295-011-1044-2

Zhai, B. T., Zeng, Y. Y., Zeng, Z. W., Zhang, N. N., Li, C. X., Zeng, Y. J., et al. (2018). Drug delivery systems for elemene, its main active ingredient β-elemene, and its derivatives in cancer therapy. *Int. J. Nanomedicine.* 13, 6279–6296. doi: 10.2147/IJN.S174527

Zhai, B. T., Zhang, N. N., Han, X. M., Li, Q. J., Zhang, M. M., Chen, X. Y., et al. (2019). Molecular targets of β-elemene, a herbal extract used in traditional Chinese medicine, and its potential role in cancer therapy: a review. *Biomed. Pharmacother.* 114:108812. doi: 10.1016/j.biopha.2019.108812

Zhang, W. X., Guo, J. Q., Wang, Z., Li, Y. W., Meng, X. F., Shen, Y., et al. (2021). Improved production of germacrene A, a direct precursor of ß-elemene, in engineered Saccharomyces cerevisiae by expressing a cyanobacterial germacrene A synthase. *Microb. Cell. Fact.* 20:7. doi: 10.1186/s12934-020-01500-3

Zhang, X. L., Huang, L. Q., Dai, Z. B., Wang, D., Zhang, L. L., Guo, J., et al. (2018). *Recombinant Yeast and Thereof*. Patent China CN108060092A. Available online at: https://www.patent9.com/patent/201711064197.5.html (accessed May 22, 2018).

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

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fpls.2022.932594

OPEN ACCESS

EDITED BY
Supaart Sirikantaramas,
Chulalongkorn University, Thailand

REVIEWED BY
Wei Zhou,
Zhejiang Chinese Medical University, China
Mihir Kumar Mandal,
University of California, Davis, United States

*CORRESPONDENCE
Ling Yuan
lyuan3@uky.edu
Yongliang Liu
yongliangliu@uky.edu

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

RECEIVED 30 April 2022
ACCEPTED 18 July 2022
PUBLISHED 18 August 2022

CITATION
Wu Z, Singh SK, Lyu R, Pattanaik S, Wang Y, Li Y, Yuan L and Liu Y (2022) Metabolic engineering to enhance the accumulation of bioactive flavonoids licochalcone A and echinatin in *Glycyrrhiza inflata* (Licorice) hairy roots.
Front. Plant Sci. 13:932594.
doi: 10.3389/fpls.2022.932594

Metabolic engineering to enhance the accumulation of bioactive flavonoids licochalcone A and echinatin in *Glycyrrhiza inflata* (Licorice) hairy roots

Zhigeng Wu[1,2], Sanjay Kumar Singh[3], Ruiqing Lyu[3], Sitakanta Pattanaik[3], Ying Wang[1,2], Yongqing Li[1,2], Ling Yuan[1,3]* and Yongliang Liu[1,3]*

[1]Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement and Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Department of Plant and Soil Sciences and Kentucky Tobacco Research and Development Center, University of Kentucky, Lexington, KY, United States

Echinatin and licochalcone A (LCA) are valuable chalcones preferentially accumulated in roots and rhizomes of licorice (*Glycyrrhiza inflata*). The licorice chalcones (licochalcones) are valued for their anti-inflammatory, antimicrobial, and antioxidant properties and have been widely used in cosmetic, pharmaceutical, and food industries. However, echinatin and LCA are accumulated in low quantities, and the biosynthesis and regulation of licochalcones have not been fully elucidated. In this study, we explored the potential of a R2R3-MYB transcription factor (TF) *AtMYB12*, a known regulator of flavonoid biosynthesis in *Arabidopsis*, for metabolic engineering of the bioactive flavonoids in *G. inflata* hairy roots. Overexpression of *AtMYB12* in the hairy roots greatly enhanced the production of total flavonoids (threefold), echinatin (twofold), and LCA (fivefold). RNA-seq analysis of *AtMYB12*-overexpressing hairy roots revealed that expression of phenylpropanoid/flavonoid pathway genes, such as *phenylalanine ammonia-lyase* (*PAL*), *chalcone synthase* (*CHS*), and *flavanone 3′-hydroxylase* (*F3′H*), is significantly induced compared to the control. Transient promoter activity assay indicated that AtMYB12 activates the *GiCHS1* promoter in plant cells, and mutation to the MYB-binding motif in the *GiCHS1* promoter abolished activation. In addition, transcriptomic analysis revealed that *AtMYB12* overexpression reprograms carbohydrate metabolism likely to increase carbon flux into flavonoid biosynthesis. Further, AtMYB12 activated the biotic defense pathways possibly by activating the salicylic acid and jasmonic acid signaling, as well as by upregulating WRKY TFs. The transcriptome of *AtMYB12*-overexpressing hairy roots serves as a valuable

source in the identification of potential candidate genes involved in LCA biosynthesis. Taken together, our findings suggest that *AtMYB12* is an effective gene for metabolic engineering of valuable bioactive flavonoids in plants.

KEYWORDS

***Glycyrrhiza inflata*, flavonoids, echinatin, licochalcone A, AtMYB12 gene, metabolic engineering**

Introduction

Glycyrrhiza species of the family Fabaceae, including *Glycyrrhiza glabra* L., *Glycyrrhiza uralensis* Fisch., and *Glycyrrhiza inflata* Bat., are valued greatly for their roots and rhizomes (licorice), which are widely used in cosmetics and herbal medicines (Zhang and Ye, 2009; Jiang et al., 2020). The bioactivity of licorice is mainly attributed to two groups of specialized metabolites:, namely, triterpene saponins and flavonoids (Wang D. et al., 2020; Wang Z.-F. et al., 2020). Glycyrrhizin is the most abundant saponin in licorice and has long been recognized as a potent sweetening agent (Pandey and Ayangla, 2018). Glycyrrhizin also has been explored for anti-coronavirus properties in the current COVID-19 pandemic (Luo et al., 2020; Chrzanowski et al., 2021). The other major group of bioactive components present in licorice are flavonoids (Zhu et al., 2016; Cheng et al., 2021). The licorice flavonoids are known to possess anti-inflammatory, antioxidant, and antimicrobial properties (Wang Z.-F. et al., 2020; Husain et al., 2021). Among the different flavonoids, echinatin and licochalcone A (LCA) are predominantly present in *G. inflata* (Lin et al., 2017; Rizzato et al., 2017; Song et al., 2017). A high LCA cosmetic formulation reduces UV-induced erythema formation in human healthy volunteers possibly by modulation of dendritic cell activity (Kolbe et al., 2006). Because of its anti-inflammatory and antimicrobial properties, LCA has been used for the treatment of facial skin diseases such as acne and rosacea (Schoelermann et al., 2016; Yang et al., 2018). Therefore, there is a great demand of LCA in cosmetic industries (Nguyen et al., 2020; Cerulli et al., 2022). However, LCA is naturally accumulated at low levels in wild *G. inflata*, even less in cultivated *G. inflata* plants. Metabolic engineering is thus viewed as a rational alternative to increase LCA production.

Chalcones are a subgroup of polyphenol compounds that are synthesized through the phenylpropanoid pathway (**Figure 1A**). The precursor phenylalanine (Phe) is derived from the primary metabolic pathways, including glycolysis, the shikimate pathway, and Phe biosynthetic pathway (Tzin and Galili, 2010). Chalcone synthase (CHS) is the first rate-limiting enzyme specific for flavonoid pathway (Saito et al., 2013). Regulation of the flavonoid biosynthesis has been extensively studied in numerous plant species including *Arabidopsis thaliana* (Saito et al., 2013). In *Arabidopsis*, three closely related MYB transcription factors (TFs), MYB11, MYB12, and MYB111, from subgroup 7 of the R2R3-MYB family, redundantly regulate the biosynthesis of flavonoids, especially flavonols (Mehrtens et al., 2005; Stracke et al., 2007). These MYBs bind to the promoters of key flavonoid biosynthetic pathway genes, such as *CHS*, to activate expression (Mehrtens et al., 2005; Stracke et al., 2007). The three MYBs exhibit distinct expression patterns, and AtMYB12 mainly controls flavonoid biosynthesis in *Arabidopsis* roots (Stracke et al., 2007). The regulatory role of AtMYB12 on flavonoid pathway has been further investigated through heterologous expression in tobacco leaves and tomato fruits (Luo et al., 2008; Pandey et al., 2015; Zhang et al., 2015). AtMYB12 induces the accumulation of flavonoids in tomato fruits by reprogramming the primary metabolism and directing the carbon flux toward flavonoid pathway (Zhang et al., 2015). Chlorogenic acid (CGA) is a subclass of polyphenols present in *Solanaceous* species (tomato and tobacco) and coffee, but not in *Arabidopsis* (Luo et al., 2008; Naveed et al., 2018). In addition to flavonoids, ectopic expression of *AtMYB12* in tobacco significantly increases CGA biosynthesis (Luo et al., 2008; Zhang et al., 2015). AtMYB12 also activates CGA biosynthetic genes in tomato fruits (Luo et al., 2008; Zhang et al., 2015). AtMYB12 overexpression in kale increases total flavonoid and phenolics in leaves (Lännenpää, 2014). These findings suggest that AtMYB12 is a potential candidate for metabolic engineering to induce flavonoids and flavonoid-derived metabolites in heterologous plant species.

As the biosynthetic pathway and gene regulation of licorice chalcones are not well elucidated, we aimed to explore the potential of AtMYB12 for metabolic engineering of licorice chalcones in *G. inflata*. We hypothesized that ectopic expression of *AtMYB12* in *G. inflata* will lead to higher accumulation of licorice chalcones and identification of potential chalcone pathway genes. As protocol for the generation of stable transgenic lines is not established in *G. inflata*, we thus generated hairy roots overexpressing *AtMYB12*. Molecular and biochemical analyses of *AtMYB12*-overexpressing hairy roots showed higher expression of phenylpropanoid/flavonoid pathway genes, including *GiCHS*, and increased accumulation

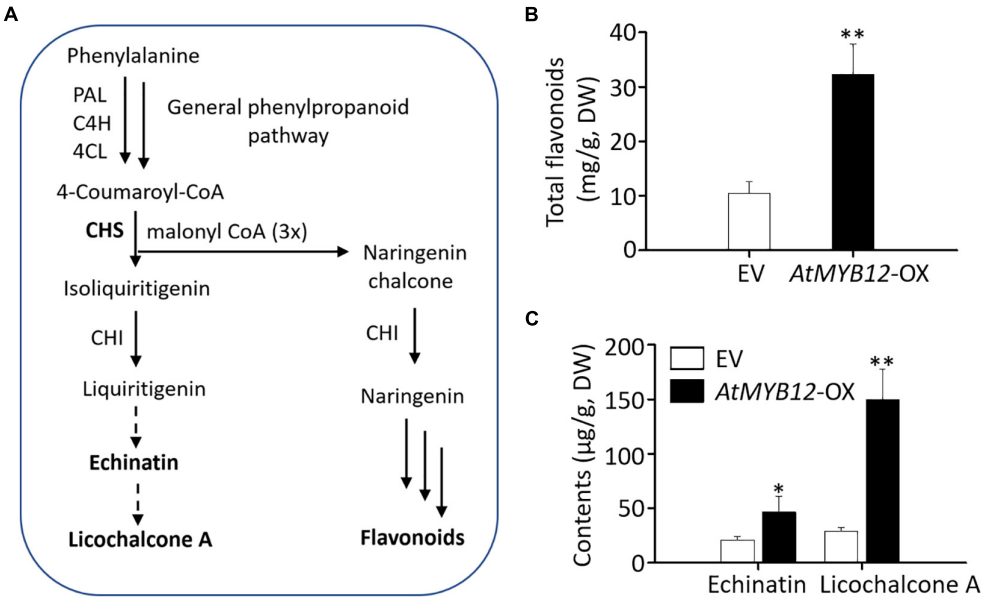


FIGURE 1
Proposed biosynthetic pathway of licorice chalcones and measurement of metabolites in empty vector (EV) control and *AtMYB12*-OX hairy roots of *G. inflata*. **(A)** A simplified, schematic diagram of the proposed licorice chalcone and general flavonoid biosynthetic pathway. PAL, phe ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate; CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase. Solid arrows indicate known enzymes; dotted arrows indicate pathway enzymes are not known. **(B,C)** Echinatin and licochalcone A contents on dry weight basis. Data are presented as the mean ± SD (n = 3). Asterisks indicate statistically significant differences compared with EV lines (*$p < 0.05$, **$p < 0.01$, Student's t-test). Total flavonoids based on dry weight (DW).

of total flavonoids and licorice-specific flavonoids, such as echinatin and LCA, confirming the regulatory roles of AtMYB12 on early flavonoid pathway genes in a heterogeneous plant species. In addition, RNA-seq data showed that the carbon flux was reprogrammed toward the flavonoid pathway. Our findings suggest that AtMYB12 is an effective regulator for engineering the production of licorice chalcones in *G. inflata*.

Materials and methods

Plant materials

Arabidopsis thaliana Col-0 accession was used for RNA isolation and *AtMYB12* cloning. *G. inflata* seeds were provided by Gansu Jin You Kang Pharmaceutical Technology Co., Ltd., Lanzhou, China. *G. inflata* seeds were soaked in H_2SO_4 for 30 min, washed with water five times, then treated with 1% NaClO for 10 min, and washed with sterilized distilled water five times. Surface-sterilized seeds were germinated on Murashige and Skoog (MS) medium and kept in dark for 2 days before being transferred to light condition. Then, 8-day-old *G. inflata* seedlings were used for DNA, RNA isolation, transient gene expression, and generation of transgenic hairy roots.

Generation of transgenic *AtMYB12* hairy roots

AtMYB12 was amplified from *Arabidopsis* cDNA and cloned into the pCAMBIA2301 vector containing *CaMV*35S promoter and *rbcS* terminator to generate pCAMBIA2301-AtMYB12. Primers used for cloning of *AtMYB12* and other genes in this study are all listed in **Supplementary Table 1**. The empty vector (EV; pCAMBIA2301) and pCAMBIA2301-AtMYB12 plasmids were separately transformed into *Agrobacterium rhizogenes* R1000 by freeze–thaw method. The hypocotyl segments from 8-day-old *G. inflata* seedlings were submerged in the *A. rhizogenes* R1000 suspension for 30 min, blot-dried on sterile filter paper, and then placed on MS medium at 22°C in darkness. After co-cultivation for 2 days, the hypocotyl segments were transferred to MS medium supplemented with 400 mg/l cefotaxime. After 2–3 weeks of culture, hairy roots developed from hypocotyls and the rapidly growing hairy roots were excised and cultivated individually on solid MS medium supplemented with 400 mg/l cefotaxime and 100 mg/l kanamycin for 2 weeks at 25°C in dark. Rapidly growing root lines that showed kanamycin resistance were selected for further analysis. These hairy root lines were cultured in 125 ml flasks each containing 10 ml MS liquid medium on an orbital shaker at 100 rpm at 25°C. The hairy roots clones were routinely subcultured every 2 weeks and harvested after 2 months for RNA isolation and metabolite extraction.

cDNA synthesis and determination of transgenic status of the hairy roots

Total RNA was isolated from EV control and *AtMYB12*-overexpressing seedlings and hairy roots using the RNeasy Plant Mini Kit following the instructions of the manufacturer (QIAGEN, United States). Approximately 2 μg of total RNA was used for DNase I digestion. Synthesis of first-strand cDNA was performed using Superscript III reverse transcriptase (Invitrogen) in a total volume of 20 μl. To verify the transgenic status of *AtMYB12*-OX and EV control hairy root lines, gene-specific primers were used to PCR-amplify the *rol B*, *rol C*, *vir C*, and kanamycin-resistant (*nptII*) genes. PCR products were analyzed on a 1% ethidium bromide-stained agarose gel.

Determination of the contents of flavonoids in hairy roots

Total flavonoid contents were determined by sodium nitrite–aluminum nitrate colorimetric method using rutin as standard (Hao et al., 2018). The standards rutin, echinatin, and LCA were purchased from Biosynth Carbosynth, United States. The contents of echinatin and LCA were determined by LC-MS/MS.

Library construction and RNA sequencing

Three independent lines of both EV and *AtMYB12*-OX hairy roots were used for RNA-seq. Total RNA was isolated from hairy roots using the RNeasy Plant Mini Kit (QIAGEN, United States) following the instructions of the manufacturer. The RNA samples with RNA integrity number (RIN) 8 or above were used for library preparation and sequencing. The TruSeq RNA Sample Prep Kit (Illumina, United States) was used for making libraries according to the protocol of the manufacturer. Individually indexed libraries were combined at equal proportions and loaded onto a single lane of a flow cell. A 50-cycle single-end sequencing run was performed on the Illumina HiSeq2500 at the Duke Center for Genomic and Computational Biology.

Data processing, identification of differentially expressed genes, and gene ontology enrichment analysis

Raw Illumina sequence reads were processed as described previously (Singh et al., 2015). Read mapping was performed by Bowtie2 (Langmead and Salzberg, 2012) using an in-house-generated *G. inflata* transcriptome (unpublished data). Differential gene expression analysis was carried out using the DESeq2 Bioconductor package in R (Love et al., 2014). The differentially expressed genes (DEGs) were identified following two criteria: (i) fold change ≥ 2 and (ii) false discovery rate *p*-value correction of ≤ 0.05. Heatmaps were constructed using the Complex Heatmap (Gu et al., 2016) function in R through the Bioconductor package (R Core Team, 2022). Functional annotation of DEGs was performed with eggNOG 4.5 (Huerta-Cepas et al., 2016) database. Gene Ontology (GO) analysis of the enriched functional categories was performed using BiNGO (version 2.44) (Maere et al., 2005).

Reverse transcription quantitative PCR

Reverse transcription quantitative PCR (RT-qPCR) was used to measure transcripts levels of *GiCHS* genes. The *GiActin* gene was used as an internal control. Relative gene expression was measured as previously described (Liu et al., 2019). All qRT-PCRs were performed in triplicate and repeated twice. Primers used in qRT-PCR are listed in **Supplementary Table 1**.

Transient overexpression of *AtMYB12* in *Glycyrrhiza inflata* seedlings

The EV and pCAMBIA2301-AtMYB12 were transformed into *Agrobacterium tumefaciens* GV3101 by freeze–thaw method and was plated on Luria–Bertani (LB) medium containing 100 μg ml^{-1} kanamycin, 50 μg ml^{-1} gentamicin, and 30 μg ml^{-1} rifampicin. A single colony was transferred to 1 ml liquid LB medium containing the same antibiotics and incubated at 250 rpm and 28°C overnight. The overnight culture was diluted in 25 ml liquid LB medium and grown for 16 h at 250 rpm and 28°C. The cells were then centrifuged, and the pellet was resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, 100 μM acetosyringone) to an OD_{600} of 1.0, and incubated at 28°C for at least 3 h. Then, 8-day-old *G. inflata* seedlings were immersed in the infiltration solution under vacuum pressure for 1 h. After vacuum infiltration, seedlings were washed five times with sterile distilled water and laid on sterile wet filter papers in Petri dishes. After 5 days of incubation at room temperature, the transfected seedlings were collected for RNA isolation.

Cloning of the *GiCHS1* promoter

Genomic DNA was extracted from *G. inflata* seedlings for promoter cloning. A forward primer (CHS1-pro-F) was designed based on the genomic sequence upstream of the coding region of *G. uralensis* homolog of *GiCHS1*. A reverse primer (GiCHS1-cds-R) was designed within the coding sequence of

GiCHS1. PCR product of GuCHS1-pro-F and GiCHS1-cds-R was sequenced. *CHS1* promoter sequences of *G. inflata* and *G. uralensis* were aligned using ClustalW software (Thompson et al., 2003). Based on the *GiCHS1* promoter sequence, another pair of primers (GiCHS1-pro-F2 and GiCHS1-pro-R2) was designed for vector construction.

Promoter activity assay in tobacco protoplasts

Tobacco cell line described before (Pattanaik et al., 2010) was used for protoplast isolation and promoter activity assay. The effector plasmid was constructed by cloning *AtMYB12* into a modified pBS vector under the control of the *CaMV*35S promoter and *rbcS* terminator. The reporter plasmid was generated by cloning *GiCHS1* promoter in a modified pUC vector containing the firefly *luciferase* (*LUC*) reporter and *rbcS* terminator. The MYB binding motif in *GiCHS1* promoter was mutated by site-directed mutagenesis to generate mutant promoter. The *GUS* reporter driven by *CaMV*35S promoter and *rbcS* terminator was used as an internal control in the protoplast assay. The reporter, effector, and internal control plasmids were electroporated into tobacco protoplasts in different combinations; luciferase and GUS activities in transfected protoplasts were measured as described previously (Pattanaik et al., 2010).

Results

Ectopic expression of *AtMYB12*-induced total flavonoids and licorice-specific chalcones in *Glycyrrhiza inflata* hairy roots

We generated transgenic hairy roots overexpressing *AtMYB12* (*AtMYB12*-OX) aiming to increase flavonoid production. EV hairy root lines (EV-1, EV-2, and EV-3) served as control. Three independent *AtMYB12*-OX hairy root lines (*AtMYB12*-OX-1, *AtMYB12*-OX-2, and *AtMYB12*-OX-3) were selected for further analysis. The transgenic status of the independent EV and *AtMYB12*-OX hairy root lines was verified by PCR (**Supplementary Figure 1**). Total flavonoid contents of the three *AtMYB12*-OX lines were significantly higher (threefold) than those of the EV lines (**Figures 1A,B**). While echinatin and LCA in EV hairy root lines were approximately 18–24 and 25–31 ng mg^{-1}, respectively (**Figures 1A,C**); *AtMYB12*-OX lines showed a significant increase in the accumulation of echinatin (30–59 ng mg^{-1}; ∼2.2-fold increase) and LCA (119–174 ng mg^{-1}; ∼5.2-fold increase) (**Figure 1C**). The metabolic outcomes of *AtMYB12* overexpression suggest that *AtMYB12* is an effective gene for metabolic engineering of the licorice flavonoid pathway.

AtMYB12-induced expression of phenylpropanoid/flavonoid pathway genes in *Glycyrrhiza inflata* hairy roots

The metabolic outcomes of *AtMYB12*-OX hairy roots prompted us to generate and analyze the transcriptome data of EV and *AtMYB12*-OX lines. Sequencing of RNA libraries of EV and *AtMYB12*-OX lines generated a total of 1,110 million (M) clean reads. Each biological replicate was represented by an average of more than 170 M reads. On average, more than 70% of the total reads from EV and overexpression line libraries were successfully mapped to the *G. inflata* transcriptome (**Supplementary Figure 2**). Compared to the EV lines, 3,236 genes were differentially expressed in *AtMYB12*-OX lines, in which 1,722 genes were upregulated and 1,514 genes were downregulated (**Supplementary Table 2**). We particularly examined genes in the phenylpropanoid/flavonoid pathway. CHS is a key rate-limiting enzyme in flavonoid biosynthetic pathway (Zhang et al., 2017). Noticeably, 13 *GiCHSs* were identified among the DEGs, and 12 of them were upregulated in *AtMYB12*-OX hairy roots (**Figure 2A** and **Table 1**). In addition, we identified three *G. inflata phenylalanine ammonia-lyase* (*PAL*) and 9 *flavanone 3'-hydroxylase* (*F3'H*) genes among the DEGs, and all of them were induced in *AtMYB12*-OX hairy roots (**Table 1**). To verify the expression of selected DEGs in RNA-seq, we conducted RT-qPCR to measure the expression of two *CHS* genes, *Gin33862* (hereafter designed as *GiCHS1*) and *Gin35437*, using independently isolated RNAs from the *AtMYB12*-OX hairy roots. The results confirmed the induction of both *CHS* genes in *AtMYB12*-OX hairy root lines (**Figure 2B**). These results suggest that the upregulation of *GiPALs*, *GiCHSs,* and *GiF3'Hs* likely leads to the enrichment of flavonoids in *G. inflata* hairy roots.

GiCHS1 is highly expressed in roots

LCA and echinatin are preferentially accumulated in *G. inflata* roots and rhizomes. We therefore analyzed the transcriptomes (SRA accession: PRJNA574093) of *G. inflata* leaves and roots, collected from two geographical locations in China (Guangzhou and Ningxia), to determine the tissue-specific expression of *CHS*. Among the 12 *GiCHSs* upregulated in *AtMYB12-OX* roots, expression of two *GiCHSs* was not detected in leaf and root transcriptomes. Among the other 10 *GiCHSs*, *Gin33862* (*GiCHS1*) is preferentially expressed in *G. inflata* roots from both locations (**Figure 2C**). Two other *GiCHSs*, *Gin35437* and *Gin16453*, showed increased expression only in the roots collected from Guangzhou (**Figure 2C**).

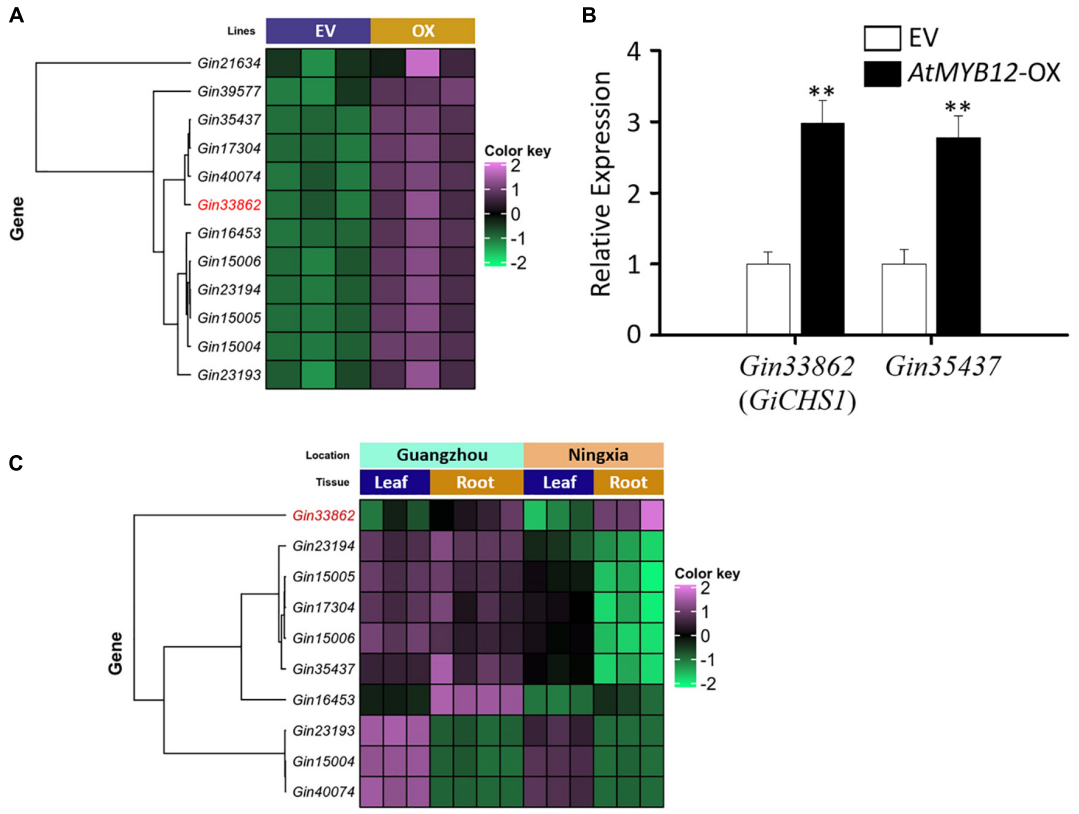


FIGURE 2
Expression profiles of the *G. inflata CHS* genes induced by AtMYB12. **(A)** Heatmap shows the expression patterns of 12 *GiCHS* genes in EV and *AtMYB12*-OX (OX) hairy root lines. *Gin33862* (*GiCHS1*) is highlighted in red. **(B)** Relative expression of two selected *GiCHS* genes in EV and *AtMYB12*-OX hairy root lines measured by qRT-PCR. Values were normalized to the expression level of an internal control, the *actin* gene (*GiActin*). Data presented are the means of three biological replicates ± SD ($n = 3$). Asterisks indicate statistically significant differences compared with EV lines (**$p < 0.01$, Student's *t*-test). **(C)** Heatmap shows the expression patterns of 10 *GiCHS* genes in two tissues (root and leaf) collected from two locations (Guangzhou and Ningxia) in China.

Transient overexpression of *AtMYB12* in *Glycyrrhiza inflata* seedlings induced *GiCHSs* expression

To further verify the effect of AtMYB12 on LCA biosynthesis in *G. inflata*, we developed an *Agrobacteria*-mediated transient gene expression assay in *G. inflata* seedlings. Similar to that in the hairy roots, expression of *GiCHS1* and *Gin35437* was significantly induced by ectopic expression of *AtMYB12* in *G. inflata* seedlings (**Figures 3A,B**). These results indicate that the heterologous AtMYB12 positively regulates flavonoid biosynthesis in *G. inflata* plants.

AtMYB12 directly activates the *GiCHS1* promoter activity

We next asked whether AtMYB12 directly activates the flavonoid pathway gene promoters in *G. inflata*. As *GiCHS1* expression is upregulated by AtMYB12 and highly expressed in roots, we cloned the *GiCHS1* promoter for activity assay. Due to the lack of genomic sequences for *G. inflata*, the *GiCHS1* promoter was cloned based on the *G. uralensis* genome sequences (Mochida et al., 2017) as *G. inflata* and *G. uralensis* are two closely related species. The amino acid sequence identity between *GiCHS1* and its *G. uralensis* homolog is 99%. The promoter of *GiCHS1* also shares high sequence identity (96%) with that of *G. uralensis CHS1* (**Supplementary Figure 3**). As shown in **Figure 3D**, transcriptional activity of the *GiCHS1* promoter (*GiCHS1*-pro) was significantly induced by AtMYB12, suggesting that AtMYB12 directly activates the *GiCHS1* promoter in plant cells. To further confirm the activation of the *GiCHS1* promoter by AtMYB12, we surveyed the promoter sequence for MYB recognition element (MRE) (A[A/C]CTACC) and identified a putative MRE (AACTACC) at -204 to -198 relative to ATG. This MRE is conserved among the *Arabidopsis CHS*, *CHI*, and *FLS* promoters and also present in the *CHS* promoters from other plants (**Figure 3C**). It is predicted to be targeted by R2R3 MYBs, including MYB11, MYB111, and

TABLE 1 Phenylpropanoid/flavonoid pathway genes identified in DEGs.

Glycyrrhiza inflata gene ID	Log2 fold change	Arabidopsis homolog	Description
Gin06540	1.618	AT2G37040	Phenylalanine
Gin12084	2.728	AT2G37040	ammonia-lyase (PAL)
Gin12083	2.426	AT3G10340	
Gin31596	-2.659	AT5G13930	
Gin15004	1.081	AT5G13930	
Gin15005	1.223	AT5G13930	
Gin15006	1.121	AT5G13930	
Gin16453	2.211	AT5G13930	
Gin17304	1.242	AT5G13930	
Gin21634	1.167	AT5G13930	Chalcone synthase
Gin23193	1.166	AT5G13930	(CHS)
Gin23194	1.258	AT5G13930	
Gin33862	1.156	AT5G13930	
Gin35437	1.139	AT5G13930	
Gin39577	1.862	AT5G13930	
Gin40074	1.627	AT5G13930	
Gin02931	1.589	AT5G07990	
Gin09638	3.125	AT5G07990	
Gin09639	4.233	AT5G07990	
Gin09641	3.840	AT5G07990	Flavanone
Gin13176	4.344	AT5G07990	3′-hydroxylase (F3′H)
Gin13177	4.586	AT5G07990	
Gin13178	4.038	AT5G07990	
Gin13179	4.270	AT5G07990	
Gin13180	3.766	AT5G07990	

MYB12 (Stracke et al., 2007). We speculated that this MRE (AACTACC) in the *GiCHS1* promoter is targeted by AtMYB12. We mutated this motif (to AAGGGGG) to generate the mutant *GiCHS1* promoter (*GiCHS1m*-pro) (**Figure 3D**). Results of promoter activity assay showed that AtMYB12 is unable to activate *GiCHS1m*-pro (**Figure 3D**), suggesting that AtMYB12 directly binds to the MYB binding site in the *GiCHS1* promoter.

RNA-seq revealed reprogramming of carbohydrate metabolism in *AtMYB12*-OX lines

Carbon resources of phenylpropanoids are derived from monosaccharides, such as glucose. The monosaccharides are directed to phenylpropanoid pathway through several primary pathways, including pentose phosphate pathway, glycolysis, and the shikimate pathway (Zhang et al., 2015). We observed that three genes related to glycolysis and shikimate pathways were induced by AtMYB12 in *G. inflata* hairy roots (**Supplementary Table 3**). Glucose also serves as a precursor of the polysaccharide cellulose, the major component of plant cell wall (Taylor, 2008; Yang et al., 2018). Further, GO enrichment analysis (**Figure 4** and **Supplementary Table 4**) showed that several pathway genes related to cellulose synthesis and cell wall synthesis, including "cell wall biogenesis" and "cellulose metabolic process," are downregulated in *AtMYB12*-OX lines. Cellulose production during cell wall biosynthesis has been shown to be dependent on cellulose synthase A (CESA). We identified seven *G. inflata CESA* genes in the DEGs that are downregulated in *AtMYB12*-OX lines (**Table 2**). These results indicate an increased carbon flux toward the phenylpropanoid pathway at the cost of cellulose synthesis in *AtMYB12*-OX hairy roots.

Pathogen defense response genes were activated in *AtMYB12*-OX lines

GO enrichment analysis showed that, in *AtMYB12*-OX lines, most of the upregulated pathways are related to pathogen defense responses, including "defense response to bacterium," "defense response to oomycetes," "response to fungus," and "innate immune response" (**Figure 4**). Salicylic acid (SA) and jasmonic acid (JA) are two important phytohormones that are particularly involved in pathogen defense (Yang et al., 2019; Chen et al., 2020). We observed that the SA and JA signaling pathways, including "response to salicylic acid," "cellular response to salicylic acid stimulus," and "response to jasmonic acid," were activated (**Figure 4** and **Supplementary Table 4**). In addition, a number of TF families were identified among the DEGs. In particular, members of the *WRKY* TFs are enriched in the DEGs (**Figure 5**). A growing body of research suggests that WRKY TFs are involved in pathogen resistance (Wani et al., 2021). Therefore, it is possible that AtMYB12-mediated defense responses are activated through SA and JA signaling, as well as the activation of WRKY TFs.

Discussion

TFs are ideal candidates for metabolic engineering because of their broad regulatory roles in metabolic pathways (Broun, 2004; Lu et al., 2016). Increasing evidence suggests that many TF functions and regulatory mechanisms are conserved across the species (Feller et al., 2011; Schluttenhofer and Yuan, 2015). For instance, ectopic expression of the maize bHLH TF *Lc* induces anthocyanin accumulation in tobacco and *Arabidopsis* (Lloyd et al., 1992). Similarly, expression of snapdragon R2R3 MYB *Rosea1* and bHLH TF *Delila,* both driven by a fruit-specific promoter, induces anthocyanin accumulation in tomato fruit (Butelli et al., 2008). Ectopic expression of the *Arabidopsis* R2R3 MYB *PAP1* induces anthocyanin accumulation in tobacco

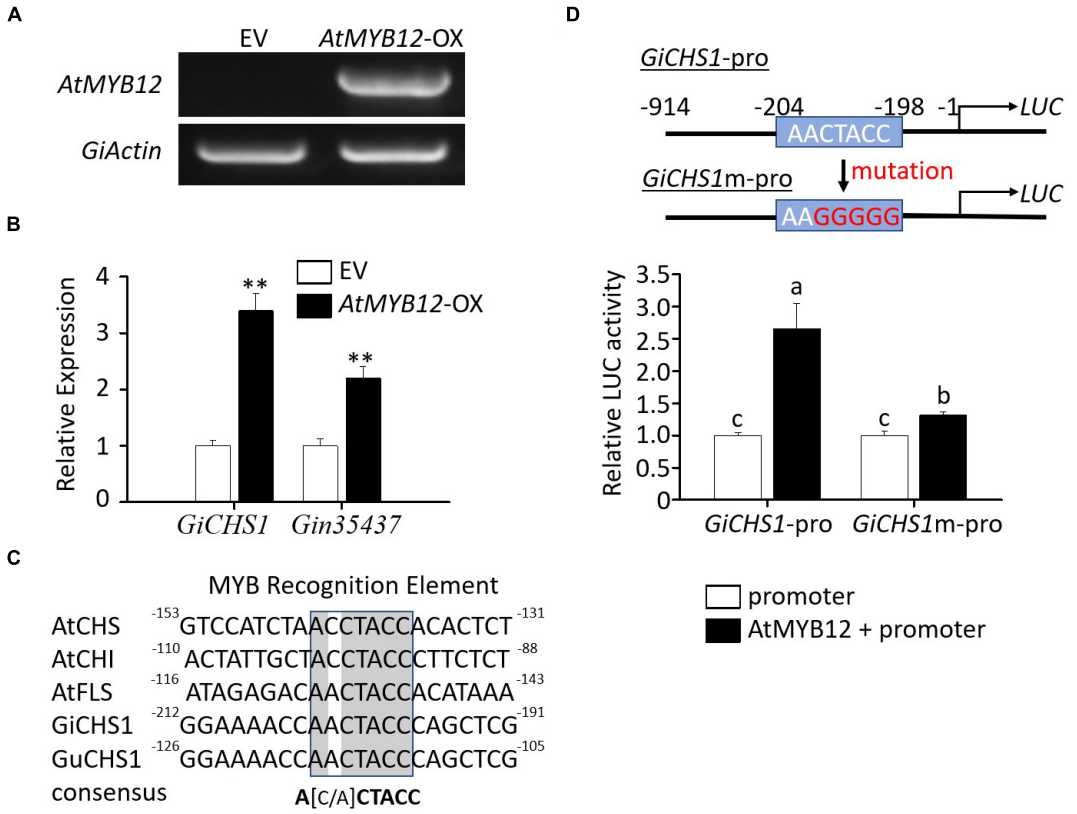


FIGURE 3
Molecular analysis of *G. inflata* seedlings transiently overexpressing *AtMYB12* and transactivation assay of the *GiCHS1* promoter. **(A)** RT-PCR analysis showed *AtMYB12* expression in *G. inflata AtMYB12*-OX seedlings but not in EV seedlings. The *G. inflata* actin gene (*GiActin*) served as internal control. **(B)** Relative expression of the two *GiCHS* genes in *AtMYB12*-OX seedlings was measured using qRT-PCR. *GiActin* was used as an internal control for normalization. Data are presented as the mean of three biological replicates ± SD. Asterisks indicate statistically significant differences compared with EV lines (**$p < 0.01$, Student's *t*-test). **(C)** Similar to the promoters of *Arabidopsis CHS*, *CHI*, and *FLS*, *GiCHS1* and *GuCHS1* promoters also contain the MYB recognition elements (MRE). Numbers by the ends of the DNA sequences, such as -130 at the right end of *AtCHS* promoter sequence, represent positions relative to translation start site. **(D)** The diagram in the right shows the MRE in the *GiCHS1* promoter (*GiCHS1*-pro) and the mutated MRE sequence in the mutant promoter (*GiCHS1m*-pro). *LUC*, the reporter *luciferase* gene. Left panel shows transactivation of *GiCHS1*-pro and *GiCHS1m*-pro after infiltration of the promoter vector alone or in combination with the AtMYB12-expression vector into tobacco cells. Data presented as the mean of biological replicates ± SD ($n = 3$). Asterisks indicate statistically significant differences compared with EV lines (**$p < 0.01$, Student's *t*-test).

(Borevitz et al., 2000). These findings support the conserved regulatory roles of TFs in metabolic pathways in plants. Licorice chalcones, including echinatin and LCA, are a characteristic group of flavonoids that are exclusively produced in *G. inflata* roots and rhizomes (Lin et al., 2017; Rizzato et al., 2017; Song et al., 2017). Although echinatin and LCA possess several important bioactive properties (Wang D. et al., 2020; Wang Z.-F. et al., 2020), key enzymes involved in the biosynthesis and molecular mechanism of regulation have not been fully elucidated. AtMYB12, along with AtMYB11 and AtMYB111, regulates flavonoid biosynthesis in *Arabidopsis* (Stracke et al., 2007). Heterologous expression of *AtMYB12* in tomato not only activates flavonoid pathway genes but also induces genes in the upstream primary metabolic pathways (Luo et al., 2008; Zhang et al., 2015). Here, we demonstrated that ectopic expression of *AtMYB12* upregulates the expression of *GiCHSs* and enhances echinatin and LCA accumulation in *G. inflata* roots (**Figures 1**, **3**).

A transformation protocol to generate stable transgenic plants is not available for *G. inflata*. As licorice-specific flavonoids are preferentially accumulated in rhizomes and roots of *G. inflata* (Lin et al., 2017; Rizzato et al., 2017; Song et al., 2017), we used transgenic hairy roots to explore the potential of AtMYB12 to enhance the accumulation of echinatin and LCA. RNA-seq analysis of three independent hairy root lines showed upregulation of flavonoid pathway genes, including *PAL*, *CHS*, and *F3'H* (**Table 1**). CHS belongs to the plant polyketide synthase superfamily and is a key enzyme in the flavonoid pathway (Dao et al., 2011). CHS catalyzes the condensation of one molecule cinnamic acid or its derivatives and three molecules of malonyl Co-A to produce the narigenin chalcone, which

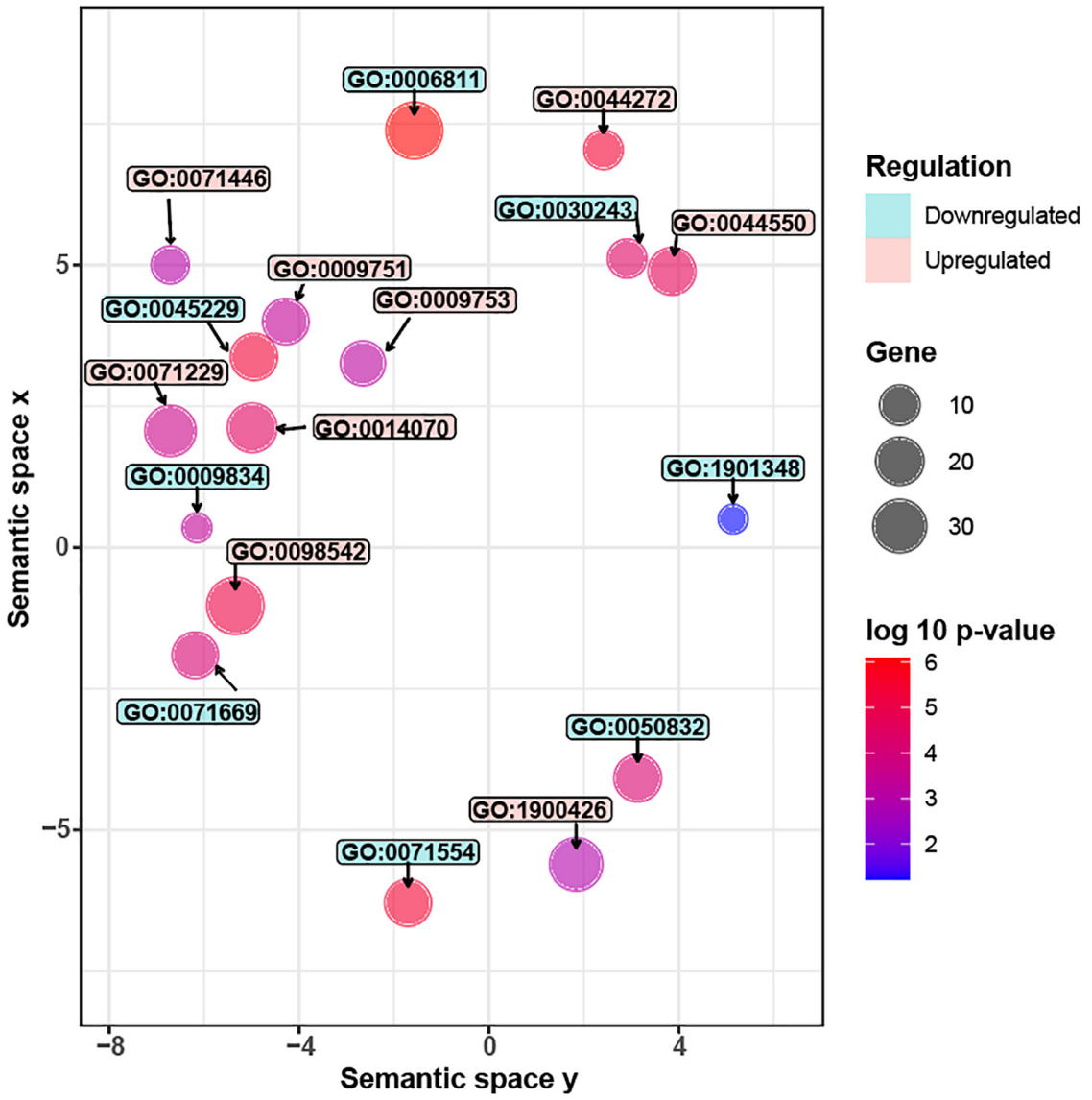


FIGURE 4
Significantly enriched GO terms in *G. inflata AtMYB12*-OX hairy roots. Gene Ontology (GO) analyses of differentially expressed genes (DEGs). Upregulated GO terms are colored in *red* while downregulated terms are in *blue*. Each circle represents one GO term. The circle size represents the number of genes in each GO category while the color represents the significance level. Description of the upregulated GO terms (from top to bottom): GO:0042742, defense response to bacterium; GO:0071446, cellular response to salicylic acid stimulus; GO:0044550, secondary metabolite biosynthetic process; GO:0009751, response to salicylic acid; GO:0009753, response to jasmonic acid; GO:0071229, cellular response to acid chemical; GO:0014070, response to organic cyclic compound; GO:0098542, defense response to other organism; GO:1900426, positive regulation of defense response to bacterium. Description of the downregulated GO terms (from top to bottom): GO:0006811, ion transport; GO:0030243, cellulose metabolic process; GO:0045229, external encapsulating structure organization; GO:0009834, plant-type secondary cell wall biogenesis; GO:1901348, positive regulation of secondary cell wall biogenesis; GO:0071669, plant-type cell wall organization or biogenesis; GO:0050832, defense response to fungus; GO:0071554, cell wall organization or biogenesis.

serves as a precursor for diverse sets of flavonoids (Dao et al., 2011). Our RNA-seq data showed that expression of 12 *G. inflata CHSs* is upregulated by AtMYB12 (**Figure 2A** and **Table 1**).

Agrobacterium-mediated transient transformation of whole seedlings has been used to study the regulation of metabolic pathways in different plant species, such as *Catharanthus roseus* (Liu et al., 2019; Mortensen et al., 2019). We transiently transformed *G. inflata* seedlings with *Agrobacterium* harboring *AtMYB12* and measured the expression of selected *CHSs*. Similar to the hairy roots, expression of two selected Gi*CHSs* was significantly induced in seedlings transformed with *AtMYB12* (**Figure 3B**). We demonstrated that AtMYB12 activates the *GiCHS1* promoter in plant cells by binding

TABLE 2 *Cellulose synthase A* (*CESA*) genes identified in DEGs.

Glycyrrhiza inflata gene ID	Log2 fold change	Arabidopsis homolog	Description
Gin23691	−1.032	AT4G39350	Cellulose synthase A2 (CESA2)
Gin14785	−1.563	AT5G44030	Cellulose synthase A4 (CESA4)
Gin34862	−1.408	AT5G44030	Cellulose synthase A4 (CESA4)
Gin05128	−1.259	AT5G17420	Cellulose synthase A7 (CESA7)
Gin32817	−1.666	AT5G17420	cellulose synthase A7 (CESA7)
Gin06050	−3.691	AT4G18780	Cellulose synthase A8 (CESA8)
Gin19934	−1.230	AT4G18780	Cellulose synthase A8 (CESA8)

to a MYB-recognition element (MRE), and mutation of the MRE abolished the activation (**Figure 3D**). In *Arabidopsis* (Mehrtens et al., 2005) and tomato (Zhang et al., 2015), AtMYB12 is shown to bind and activate the *CHS* promoter, further reinforcing the functional conservation of AtMYB12.

In tomato fruits, in addition to flavonoids, ectopic expression of *AtMYB12* upregulates CGA biosynthetic pathway genes and increases CGA accumulation (Luo et al., 2008; Zhang et al., 2015). Given the fact that echinatin and LCA contents were higher in *AtMYB12*-OX hairy roots (**Figure 1**), we speculated that licorice chalcone biosynthetic genes are regulated by AtMYB12. Although genes encoding enzymes in licorice chalcone pathway are not fully characterized, structural differences between echinatin and LCA suggest two modification steps: O-methylation and prenylation likely occur in the biosynthetic steps from echinatin to LCA. From our RNA-seq data set, we identified four genes that are annotated as O-methyltransferases and two as prenyltransferases, all differentially expressed in *AtMYB12*-OX hairy roots (**Supplementary Table 5**). These genes are thus putative candidates of future investigation of LCA biosynthesis.

In tomato fruits, other than increased accumulation of flavonoids, ectopic expression of *AtMYB12* decreases the contents of carbon resources, including glucose and fructose (Zhang et al., 2015). Genes in the primary metabolic pathways were mostly induced by AtMYB12 (Zhang et al., 2015). Our results confirmed the induction of the genes in the primary pathways upstream of flavonoids (**Supplementary Table 3**). We also found that the genes related to cellulose synthesis were downregulated (**Figure 4** and **Table 2**), suggesting that, in both tomato fruits (Zhang et al., 2015) and *G. inflata* hairy roots, AtMYB12 redirects carbon flux from other carbohydrate resources toward the flavonoid pathway. Therefore, we suggest that the functionally conserved nature of AtMYB12 makes it a promising candidate for metabolic engineering in other plant species.

AtMYB12 overexpression improves pathogen resistance in transgenic tobacco plants (Ding et al., 2021). In transgenic tobacco, *AtMYB12* induces the production of flavonoid compounds, such as rutin, as well as reactive oxygen species, H_2O_2, and NO (Pandey et al., 2015; Ding et al., 2021). Similar to tobacco, our results revealed that the genes involved in pathogen resistance pathways (**Figure 4**), defense-associated plant hormone signaling (**Figure 4**), and WRKY TFs (**Figure 5**) were significantly enriched in *AtMYB12*-OX hairy roots. SA and JA play essential roles in plant defense against different pathogens (Yang et al., 2019; Chen et al., 2020). WRKY TFs are among the largest families of transcriptional regulators and contribute to various plant processes, including disease defense (Wani et al., 2021). Some WRKYs like AtWRKY33 could regulate SA/JA biosynthesis while others are regulated by SA/JA signaling (Birkenbihl et al., 2012; Wani et al., 2021). We speculate that AtMYB12 activated the SA/JA-WRKY network that contributes to the pathogen defense responses in *G. inflata* hairy roots.

Metabolic engineering offers an excellent approach for producing various bioactive, health-promoting phytochemicals

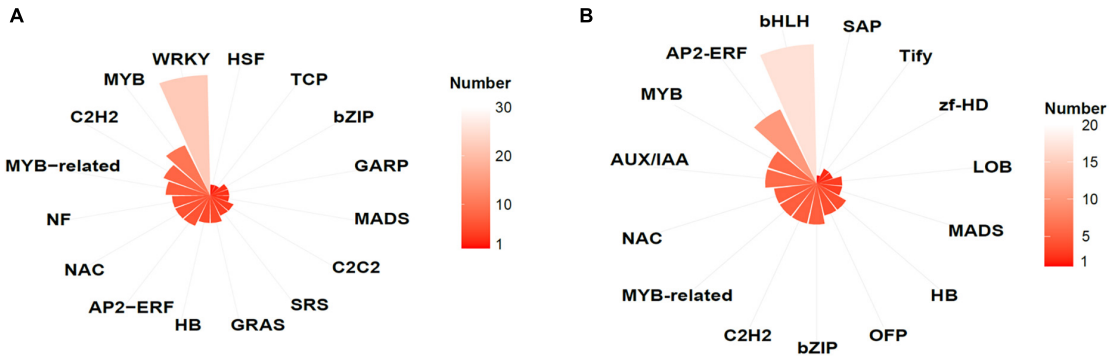


FIGURE 5
Transcription factor (TF) families identified in DEGs in EV and AtMYB12 transcriptomes. **(A)** Upregulated TFs represented by gene numbers (the lighter the color, the higher the number); e.g., the most upregulated TFs belong to WRKY family. **(B)** Downregulated TFs represented by gene numbers (the lighter the color, the higher the number); e.g., the bHLH family is most affected.

in plants. This study underscores the importance of metabolic engineering for enhancing accumulation of valuable metabolites, such as licorice chalcones in *G. inflata*, through heterologous expression of a known flavonoid regulator. Metabolic pathways of non-model plants are relatively less studied. Transcriptomic and genomic resources will help unravel the biosynthetic pathways in non-model plants, such as *G. inflata*, and aid the bioengineering of bioactive compounds.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number: PRJNA842240.

Author contributions

ZW and YLL performed the experiments, analyzed the data, and wrote the article. SS analyzed the RNA-seq data. RL, SP, YW, YQL, and LY evaluated the experiments and revised the article. YW, YQL, and LY initiated and supervised the project. All authors contributed to the article and approved the submitted version.

Funding

This study was partially supported by the National Key R&D Program of China (2019YFC1711100), by the grant nos. 2018530000241001 and 2019530000241005 from the Yunnan Tobacco Company, Harold R. Burton Endowed Professorship to LY and the Kentucky Tobacco Research and Development Center.

Acknowledgments

We thank Megan Combs (Department of Civil Engineering and Environmental Research Training Laboratories, University of Kentucky) for assistance on LC-MS/MS analyses.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.932594/full#supplementary-material

References

Birkenbihl, R. P., Diezel, C., and Somssich, I. E. (2012). *Arabidopsis* WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection. *Plant Physiol.* 159, 266–285. doi: 10.1104/pp.111.192641

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 12, 2383–2393. doi: 10.1105/tpc.12.12.2383

Broun, P. (2004). Transcription factors as tools for metabolic engineering in plants. *Curr. Opin. Plant Biol.* 7, 202–209. doi: 10.1016/j.pbi.2004.01.013

Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26, 1301–1308.

Cerulli, A., Masullo, M., Montoro, P., and Piacente, S. (2022). Licorice (*Glycyrrhiza glabra*, *G. uralensis*, and *G. inflata*) and their constituents as active cosmeceutical ingredients. *Cosmetics* 9:7.

Chen, J., Clinton, M., Qi, G., Wang, D., Liu, F., and Fu, Z. Q. (2020). Reprogramming and remodeling: Transcriptional and epigenetic regulation of salicylic acid-mediated plant defense. *J. Exp. Bot.* 71, 5256–5268. doi: 10.1093/jxb/eraa072

Cheng, M., Zhang, J., Yang, L., Shen, S., Li, P., Yao, S., et al. (2021). Recent advances in chemical analysis of licorice (Gan-Cao). *Fitoterapia* 149:104803. doi: 10.1016/j.fitote.2020.104803

Chrzanowski, J., Chrzanowska, A., and Graboń, W. (2021). Glycyrrhizin: An old weapon against a novel coronavirus. *Phytother. Res.* 35, 629–636. doi: 10.1002/ptr.6852

Dao, T., Linthorst, H., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* 10, 397–412.

Ding, X., Zhang, H., Li, M., Yin, Z., Chu, Z., Zhao, X., et al. (2021). AtMYB12-expressing transgenic tobacco increases resistance to several phytopathogens and aphids. *Front Agron.* 3:694333. doi: 10.3389/fagro.2021.694333

Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66, 94–116. doi: 10.1111/j.1365-313X.2010.04459.x

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313

Hao, P.-Y., Feng, Y.-L., Zhou, Y.-S., Song, X.-M., Li, H.-L., Ma, Y., et al. (2018). Schaftoside interacts with NlCDK1 protein: A mechanism of rice resistance to brown planthopper, Nilaparvata lugens. *Front. Plant Sci.* 9:710. doi: 10.3389/fpls.2018.00710

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Husain, I., Bala, K., Khan, I. A., and Khan, S. I. (2021). A review on phytochemicals, pharmacological activities, drug interactions, and associated toxicities of licorice (Glycyrrhiza sp.). *Food Front.* 2, 449–485. doi: 10.1002/fft2.110

Jiang, M., Zhao, S., Yang, S., Lin, X., He, X., Wei, X., et al. (2020). An "essential herbal medicine"—licorice: A review of phytochemicals and its effects in combination preparations. *J. Ethnopharmacol.* 249:112439. doi: 10.1016/j.jep.2019.112439

Kolbe, L., Immeyer, J., Batzer, J., Wensorra, U., Mundt, C., Wolber, R., et al. (2006). Anti-inflammatory efficacy of Licochalcone A: Correlation of clinical potency and in vitro effects. *Arch. Dermatol. Res.* 298, 23–30. doi: 10.1007/s00403-006-0654-4

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lännenpää, M. (2014). Heterologous expression of AtMYB12 in kale (Brassica oleracea var. acephala) leads to high flavonol accumulation. *Plant Cell Rep.* 33, 1377–1388. doi: 10.1007/s00299-014-1623-6

Lin, Y., Kuang, Y., Li, K., Wang, S., Song, W., Qiao, X., et al. (2017). Screening for bioactive natural products from a 67-compound library of *Glycyrrhiza inflata*. *Bioorg. Med. Chem.* 25, 3706–3713. doi: 10.1016/j.bmc.2017.05.009

Liu, Y., Patra, B., Pattanaik, S., Wang, Y., and Yuan, L. (2019). GATA and phytochrome interacting factor transcription factors regulate light-induced vindoline biosynthesis in Catharanthus roseus. *Plant Physiol.* 180, 1336–1350. doi: 10.1104/pp.19.00489

Lloyd, A. M., Walbot, V., and Davis, R. W. (1992). *Arabidopsis* and Nicotiana anthocyanin production activated by maize regulators R and C1. *Science* 258, 1773–1775. doi: 10.1126/science.1465611

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, X., Tang, K., and Li, P. (2016). Plant metabolic engineering strategies for the production of pharmaceutical terpenoids. *Front. Plant Sci.* 7:1647. doi: 10.3389/fpls.2016.01647

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., et al. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: Expression in fruit results in very high levels of both types of polyphenol. *Plant J.* 56, 316–326. doi: 10.1111/j.1365-313X.2008.03597.x

Luo, P., Liu, D., and Li, J. (2020). Pharmacological perspective: Glycyrrhizin may be an efficacious therapeutic agent for COVID-19. *Int. J. Antimicrob. Agents* 55:105995. doi: 10.1016/j.ijantimicag.2020.105995

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096. doi: 10.1104/pp.104.058032

Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., et al. (2017). Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *Plant J.* 89, 181–194. doi: 10.1111/tpj.13385

Mortensen, S., Bernal-Franco, D., Cole, L. F., Sathitloetsakun, S., Cram, E. J., and Lee-Parsons, C. (2019). EASI transformation: An efficient transient expression method for analyzing gene function in Catharanthus roseus seedlings. *Front. Plant Sci.* 10:755. doi: 10.3389/fpls.2019.00755

Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., et al. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomed. Pharmacother.* 97, 67–74. doi: 10.1016/j.biopha.2017.10.064

Nguyen, J. K., Masub, N., and Jagdeo, J. (2020). Bioactive ingredients in Korean cosmeceuticals: Trends and research evidence. *J. Cosmet. Dermatol.* 19, 1555–1569. doi: 10.1111/jocd.13344

Pandey, A., Misra, P., Choudhary, D., Yadav, R., Goel, R., Bhambhani, S., et al. (2015). AtMYB12 expression in tomato leads to large scale differential modulation in transcriptome and flavonoid content in leaf and fruit tissues. *Sci. Rep.* 5:12412. doi: 10.1038/srep12412

Pandey, D. K., and Ayangla, N. J. P. R. (2018). Biotechnological aspects of the production of natural sweetener glycyrrhizin from Glycyrrhiza sp. *Phytochem. Rev.* 17, 397–430.

Pattanaik, S., Werkman, J. R., Kong, Q., and Yuan, L. (2010). Site-directed mutagenesis and saturation mutagenesis for the functional study of transcription factors involved in plant secondary metabolite biosynthesis. *Methods Mol. Biol.* 643, 47–57. doi: 10.1007/978-1-60761-723-5_4

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Rizzato, G., Scalabrin, E., Radaelli, M., Capodaglio, G., and Piccolo, O. (2017). A new exploration of licorice metabolome. *Food Chem.* 221, 959–968.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: Structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001

Schluttenhofer, C., and Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiol.* 167, 295–306. doi: 10.1104/pp.114.251769

Schoelermann, A., Weber, T., Arrowitz, C., Rizer, R., Qian, K., and Babcock, M. (2016). Skin compatibility and efficacy of a cosmetic skin care regimen with licochalcone A and 4-t-butylcyclohexanol in patients with rosacea subtype I. *J. Eur. Acad. Dermatol. Venereol.* 30, 21–27. doi: 10.1111/jdv.13531

Singh, S. K., Wu, Y., Ghosh, J. S., Pattanaik, S., Fisher, C., Wang, Y., et al. (2015). RNA-sequencing reveals global transcriptomic changes in Nicotiana tabacum responding to topping and treatment of axillary-shoot control chemicals. *Sci. Rep.* 5:18148. doi: 10.1038/srep18148

Song, W., Qiao, X., Chen, K., Wang, Y., Ji, S., Feng, J., et al. (2017). Biosynthesis-based quantitative analysis of 151 secondary metabolites of licorice to differentiate medicinal Glycyrrhiza species and their hybrids. *Anal. Chem.* 89, 3146–3153. doi: 10.1021/acs.analchem.6b04919

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. *New Phytol.* 178, 239–252.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* 2:2.3.

Tzin, V., and Galili, G. (2010). The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book* 8:e0132. doi: 10.1199/tab.0132

Wang, D., Liang, J., Zhang, J., Wang, Y., and Chai, X. (2020). Natural chalcones in Chinese materia medica: Licorice. *Evid. Based Complementary Altern. Med.* 2020, 1–14.

Wang, Z.-F., Liu, J., Yang, Y.-A., and Zhu, H.-L. (2020). A review: The anti-inflammatory, anticancer and antibacterial properties of four kinds of licorice flavonoids isolated from licorice. *Curr. Med. Chem.* 27, 1997–2011. doi: 10.2174/0929867325666181001104550

Wani, S. H., Anand, S., Singh, B., Bohra, A., and Joshi, R. (2021). WRKY transcription factors and plant defense responses: Latest discoveries and future prospects. *Plant Cell Rep.* 40, 1071–1085. doi: 10.1007/s00299-021-02691-8

Yang, G., Lee, H. E., Yeon, S. H., Kang, H. C., Cho, Y. Y., Lee, H. S., et al. (2018). Licochalcone A attenuates acne symptoms mediated by suppression of NLRP3 inflammasome. *Phytother. Res.* 32, 2551–2559. doi: 10.1002/ptr.6195

Yang, J., Duan, G., Li, C., Liu, L., Han, G., Zhang, Y., et al. (2019). The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. *Front. Plant Sci.* 10:1349. doi: 10.3389/fpls.2019.01349

Zhang, Q., and Ye, M. (2009). Chemical analysis of the Chinese herbal medicine Gan-Cao (licorice). *J. Chromatogr A* 1216, 1954–1969.

Zhang, X., Abrahan, C., Colquhoun, T. A., and Liu, C.-J. (2017). A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in *Arabidopsis*. *Plant Cell* 29, 1157–1174. doi: 10.1105/tpc.16.00855

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., et al. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* 6:8635. doi: 10.1038/ncomms9635

Zhu, Z., Tao, W., Li, J., Guo, S., Qian, D., Shang, E., et al. (2016). Rapid determination of flavonoids in licorice and comparison of three licorice species. *J. Sep. Sci.* 39, 473–482. doi: 10.1002/jssc.201500685

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fpls.2022.973591

OPEN ACCESS

EDITED BY
Yongliang Liu,
University of Kentucky, United States

REVIEWED BY
Weimin Jiang,
Hengyang Normal University, China
Dongfeng Yang,
Zhejiang Sci-Tech University, China

*CORRESPONDENCE
Zhihua Liao
zhliao@swu.edu.cn
Yueli Tang
tangyueli_999@163.com

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

RECEIVED 20 June 2022
ACCEPTED 05 August 2022
PUBLISHED 02 September 2022

CITATION
Xiang L, He P, Shu G, Yuan M, Wen M, Lan X, Liao Z and Tang Y (2022) AabHLH112, a bHLH transcription factor, positively regulates sesquiterpenes biosynthesis in *Artemisia annua*. *Front. Plant Sci.* 13:973591. doi: 10.3389/fpls.2022.973591

AabHLH112, a bHLH transcription factor, positively regulates sesquiterpenes biosynthesis in *Artemisia annua*

Lien Xiang[1,2], Ping He[3], Guoping Shu[2], Mingyuan Yuan[2], Mengling Wen[2], Xiaozhong Lan[4], Zhihua Liao[2,3]* and Yueli Tang[2]*

[1]College of Environmental Science and Engineering, China West Normal University, Nanchong, China, [2]Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City and Southwest University, Tibet Agriculture and Animal Husbandry College and Southwest University (TAAHC-SWU) Medicinal Plant Joint R&D Centre, School of Life Sciences, Southwest University, Chongqing, China, [3]Chongqing Academy of Science and Technology, Chongqing, China, [4]The Provincial and Ministerial Co-founded Collaborative Innovation Center for R&D in Tibet Characteristic Agricultural and Animal Husbandry Resources, Food Science College, Tibet Agriculture and Animal Husbandry University, Nyingchi, China

The bHLH transcription factors play important roles in the regulation of plant growth, development, and secondary metabolism. β-Caryophyllene, *epi*-cedrol, and β-farnesene, three kinds of sesquiterpenes mainly found in plants, are widely used as spice in the food industry and biological pesticides in agricultural production. Furthermore, they also have a significant value in the pharmaceutical industry. However, there is currently a lack of knowledge on the function of bHLH family TFs in β-caryophyllene, *epi*-cedrol, and β-farnesene biosynthesis. Here, we found that AabHLH112 transcription factor had a novel function to positively regulate β-carophyllene, *epi*-cedrol, and β-farnesene biosynthesis in *Artemisia annua*. Exogenous MeJA enhanced the expression of *AabHLH112* and genes of β-caryophyllene synthase (*CPS*), *epi*-cedrol synthase (*ECS*), and β-farnesene synthase (*BFS*), as well as sesquiterpenes content. Dual-LUC assay showed the activation of *AaCPS, AaECS*, and *AaBFS* promoters were enhanced by AabHLH112. Yeast one-hybrid assay showed AabHLH112 could bind to the G-box (CANNTG) *cis*-element in promoters of both *AaCPS* and *AaECS*. In addition, overexpression of *AabHLH112* in *A. annua* significantly elevated the expression levels of *AaCPS*, *AaECS*, and *AaBFS* as well as the contents of β-caryophyllene, *epi*-cedrol, and β-farnesene, while suppressing *AabHLH112* expression by RNAi reduced the expression of the three genes and the contents of the three sesquiterpenes. These results suggested that *AabHLH112* is a positive regulator of β-caryophyllene, *epi*-cedrol, and β-farnesene biosynthesis in *A. annua*.

KEYWORDS

***Artemisia annua*, AabHLH112, β-caryophyllene, *epi*-cedrol, β-farnesene, metabolic regulation**

Introduction

Artemisia annua, a traditional Chinese medicinal plant, not only produces the well-known artemisinin, but also produces a number of sesquiterpenes that play an important role in plants' stress resistance, communication, and growth regulation (Chadwick et al., 2013). β-caryophyllene, *epi*-cedrol, and β-farnesene, three kinds of sesquiterpenes, are used as spice in the food industry and biological pesticides in agricultural production (Yu X. et al., 2012; Cheng et al., 2022). Besides, they also have a significant value in the pharmaceutical industry (Cheng et al., 2022). β-Caryophyllene was found to has good curative effect on colitis (Bento et al., 2011), cerebral ischemia (Chang et al., 2013), diabetes (Basha and Sankaranarayanan, 2014), anxiety and depression (Bahi et al., 2014), liver fibrosis (Mahmoud et al., 2014), and osteoarthritis (Rufino et al., 2015). In cancer studies, β-caryophyllene demonstrated synergy with the chemotherapy drug paclitaxel on human tumor cell lines, and alone it stimulates apoptosis and suppresses tumor growth (Legault and Pichette, 2007). *epi*-Cedrol is widely used in pharmaceutical industry for its sedative, anti-inflammatory, and cytotoxic activities (Zhang et al., 2016; Luo et al., 2019). β-farnesene, an important acyclic and volatile sesquiterpenes acting as the main component of the aphid alarm pheromones for many pest aphids, can be used to create new biological pesticides (Wang et al., 2014). Because of their pharmacological activity and commercial value, numerous studies have focused on improving the production of sesquiterpenes in *A. annua*.

β-Caryophyllene, *epi*-cedrol, and β-farnesene are biosynthesized *via* both the cytoplasmic mevalonate (MVA) pathway and plastidial methylerythritol phosphate (MEP) pathway (Navale et al., 2019). Plant cells use one molecule of dimethylallyl diphosphate (DMAPP) and two molecules of isopentenyl diphosphate (IPP) to produce farnesyl diphosphate (FPP) in a reaction catalyzed by FPP synthase (FPS) (Navale et al., 2019). Then the sesquiterpene synthase takes FPP as the substrate to synthesize a variety of sesquiterpenes (**Figure 1**).

Transcription factors (TFs) are sequence-specific DNA-binding proteins which interact with the promoter regions of target genes and regulate their transcription (Yang et al., 2012). It is a useful tool to improve the production of secondary metabolites in pharmaceutically important plants, such as flavonoids or alkaloids (Gantet and Memelink, 2002). Recently, a number of AP2/ERF, WRKY, and bHLH transcription factors have been found to have global regulatory functions in pharmaceutical terpenoids biosynthesis (Lu et al., 2016; Li et al., 2017). In *A. annua*, overexpression of *AaERF1* and *AaERF2* significantly promoted the transcription of *ADS* and *CYP71AV1*, resulting in an increase of artemisinin content (Yu Z. et al., 2012). Overexpression of *SmERF1L1* significantly increased tanshinones production in transgenic *Salvia miltiorrhiza* hairy roots (Huang et al., 2019). In addition, overexpression of *SmWRKY1* significantly elevated the transcription of *SmDXS* and *SmDXR*, increasing tanshinone production in *S. miltiorrhiza* (Cao et al., 2018).

The bHLH family is the second largest transcription factor family in plants, which plays an important role in the regulation of pharmaceutical terpenoids biosynthesis (Ji et al., 2014). It was reported that *SmbHLH3* acts as a transcription repressor for both phenolic acids and tanshinones biosynthesis in *S. miltiorrhiza* hairy roots (Zhang et al., 2020). In *Arabidopsis*, AtMYC2, a JA-responsive TF, can enhance sesquiterpenes biosynthesis by binding to the promoters of *TPS11* and *TPS21* (Hong et al., 2012). Chuang et al. (2018) reported that overexpression of *PbbHLH4* markedly increased terpenes content in *Phalaenopsis*. JA-responsive *SlMYC1* differentially regulates monoterpene and sesquiterpenes biosynthesis in tomato, which positively regulates monoterpene biosynthesis in leaf and stem but negatively regulates sesquiterpenes biosynthesis in stem (Xu et al., 2018). *CrMYC2* directly activates the expression of *ORCA* genes through binding to the G-box in its promoter, and promotes vinblastine and vincristine biosynthesis in *Catharanthus roseus* (Zhang et al., 2011). *BpMYC4* and *BpbHLH9* from *Betula platyphylla Suk* are involved in regulating the biosynthesis of triterpenes (Yin et al., 2017). Furthermore, studies also showed that bHLH TFs could act as negative regulators in terpenes biosynthesis, including the biosynthesis of the diterpene paclitaxel and monoterpenoid indole alkaloid in *Taxus cuspidata* (Lenka et al., 2015) and *Catharanthus roseus* (Patra et al., 2018), respectively. One hundred and twenty two putative bHLH TFs were found in *A. annua* by genome-wide identification (Xiang et al., 2019) and only three bHLH TFs-*AabHLH1*, *AabHLH112*, and *AaMYC2*, respectively, have been reported to play regulatory roles in artemisinin biosynthesis (Ji et al., 2014; Shen et al., 2016; Xiang et al., 2019). Now, although our knowledge of the regulation of artemisinin biosynthesis is increasing, little is known about the regulation of non-artemisinin sesquiterpenes biosynthesis in *A. annua*. In this study, we found that the *AabHLH112* has a novel function in improving the accumulation of β-caryophyllene, *epi*-cedrol, and β-farnesene in *A. annua*, and revealed the mechanism of AabHLH112-mediated upregulation of sesquiterpenes biosynthesis. Our results suggested that this transcription factor could be a useful tool to improve sesquiterpenes production in *A. annua*.

Materials and methods

Plant material and growth conditions

Seeds were harvested from wild-type *A. annua* grown in the experimental field of Southwest University (Chongqing, China) for this study. These seeds were surface-sterilized with 20% sodium hypochlorite solution for 20 min, and then washed three times with sterile water. Subsequently, seeds were germinated on

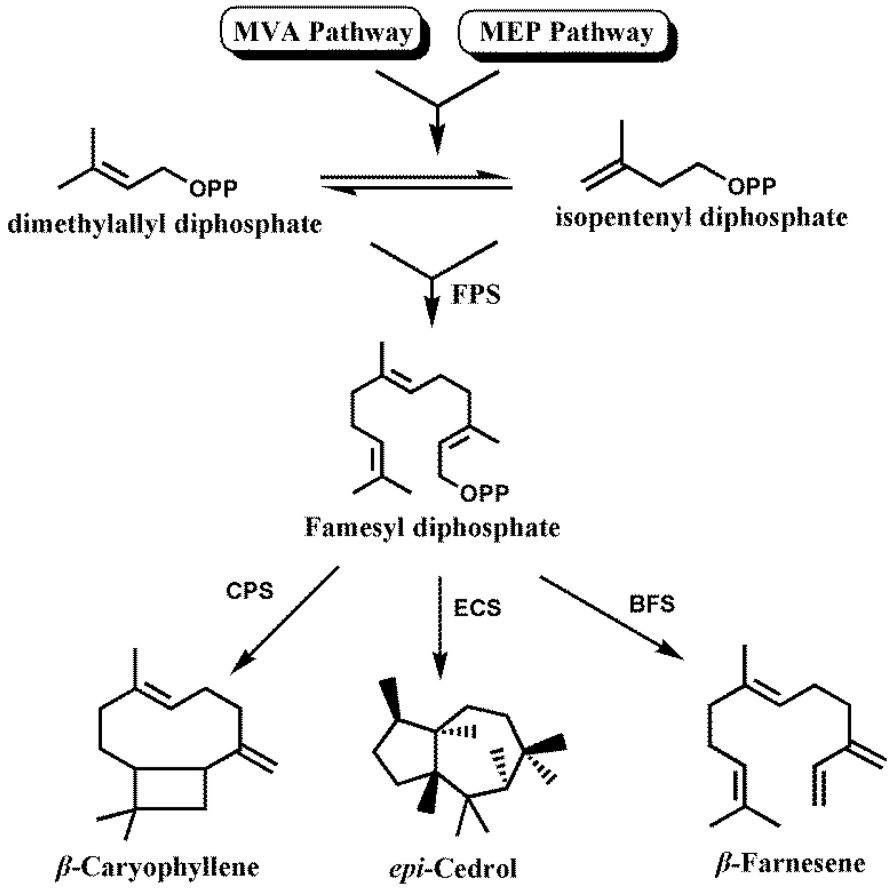


FIGURE 1
Biosynthetic pathway for sesquiterpenes in *Artemisia annua*. FPS, farnesyl diphosphate synthase; CPS, β-caryophyllene synthase; ECS, *epi*-cedrol synthase; BFS, β-farnesene synthase.

1/2 MS solid medium at 23 ± 2°C under a light period of 16-h-light/8-h-dark. All seedlings were grown in pots with organic substrates in an artificial climate room at 23 ± 2°C under a light period of 16-h-light/8-h-dark. *Nicotiana benthamiana* seeds were sown directly on solid and their growth conditions are consistent with *A. annua*. *Nicotiana benthamiana* plants grown for 30 days were used for dual-luciferase assays.

Exogenous methyl jasmonate treatment and sample collection

One-month old wild-type *A. annua* plants were treated according to the report (Xiang et al., 2015) with modifications. 300 μM exogenous MeJA solutions containing 0.4% ethanol was sprayed on the plant surfaces. Then the leaves of *A. annua* plants at 0, 3, 6, 12, and 24 h after exogenous MeJA treatment were collected respectively and then put in liquid nitrogen immediately for RNA extraction and determination of sesquiterpenes contents.

Artemisia annua transformation

The coding sequence of *AabHLH112* was amplified and inserted into the pHB binary plasmid vector through *Bam*HI and *Pst*I sites to construct the plant overexpression vector, pHB-AabHLH112. Besides, the 352bp fragment from the *AabHLH112* was amplified and inserted into the intermediate pHANNIBAL vector, and then the expression cassette was recombined into the pBin19 plasmid for constructing the *AabHLH112*-RNAi vector. pHB-AabHLH112 and pBin19-RNAi-AabHLH112 vectors were transferred into *A. tumefaciens* strain EHA105 to form engineering strains. These strains were grown on YEP solid medium containing related antibiotics for 48 h. Subsequently, positive monoclonal strain was inoculated into YEP liquid medium containing related antibiotics to culture OD_{600} = 0.5∼0.6. The supernatant was discarded after centrifugation, and then resuspended in the 1/2 MS liquid to OD_{600} = 0.3∼0.5. Then the culture was shaken at 200 rpm/28°C for 30 min. Cultured engineered strains were used to transform *A. annua via Agrobacterium*-mediated

transformation as described previously (Xiang et al., 2019). After that, the transformed seedlings were transplanted into pots with organic substrates and cultured in an artificial climate room at 23 ± 2°C under a light period of 16 h-light/8 h-dark.

Quantitative real-time polymerase chain reaction

Quantitative real-time PCR was performed to analyze genes' expression in this study. Total RNA of root, stem, leaf and flower was extracted using the total plant RNA Extract Kit (Tiangen, China), and then reversely transcribed into cDNA using Fast King RT Kit (Tiangen, China). Subsequently, the cDNA was used as the template for detecting the expression levels of *AabHLH112* and sesquiterpene synthase genes *AaCPS*, *AaECS*, and *AaBFS* by qPCR experiment. The qPCR amplification conditions were 95°C for 3 min, followed by 40 cycles of 95°C for 15 s, 55°C for 20 s, and 72°C for 20 s. β-*actin* of *A. annua* was used as the reference gene in this study (Ma et al., 2018), and the relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). In addition, the expression levels of sesquiterpenes-related genes were detected in *AabHLH112*-overexpressing and RNAi-*AabHLH112* transgenic *A. annua* using the above method. The primer sequences in qPCR experiment were listed in **Supplementary Table 1**.

Plant sesquiterpenes extraction and gas chromatography-mass spectrometry analysis

The sesquiterpenes of *A. annua* leaves were analyzed by gas chromatography-mass spectrometry (GC-MS) as described before with some modifications (Fu et al., 2017). The fresh leaves 1–8 from the main stem of 3 months old *AabHLH112*-overexpressing transgenic and wild-type *A. annua* plants were harvested, frozen immediately in liquid nitrogen, and then freeze-dried for 72 h at -80°C. 100 mg leaf powder was soaked in 4 mL *n*-hexane in 15 mL centrifuge tube with 35 μL trans-farnesol (80 μg/mL) as the internal standard, and ultrasonically extracted for 45 min at 28°C in an ultrasonic processor (KQ-500DE; Kunshan Ultrasonic Instrument Co. Ltd., Kunshan, China). Plant extract was centrifuged at 1,000 rpm for 10 min. The supernatant was filtered through 0.22 μm pore size filters and then analyzed by gas chromatography-mass spectrometry (GCMS-QP2010 Ultra; Shimadzu) with the temperature program: initial temperature of 70°C (1 min hold), increase to 160°C at 10°C/min, ramp to 240°C at 5°C/min, and finally increase to 280°C at 20°C/min (17 min hold). Helium was used as a carrier gas and 1 μL sample was injected in split mode; split rate, 2:1; ion source temperature, 230°C; ionization voltage, 70 eV with scanning from *m/z* 33 to 500. Qualitative analysis of all compounds was fulfilled by comparing with NIST (National Institute of Standards and Technology) database and Wiley libraries. The relative contents of sesquiterpenes were calculated by comparing peak areas with that of the internal standards. Trans-farnesol was purchased from Sigma-Aldrich in this study.

Cloning and analysis of the promoter regions of sesquiterpene synthase genes

To isolate the promoter fragments of sesquiterpene synthase genes (*AaCPS*, *AaECS*, and *AaBFS*), the total DNA of *A. annua* leaves was extracted as the template using cetyltrimethylammonium bromide (CTAB) method. All primer sequences were designed according to the genome of *A. annua*. To avoid non-specific amplification, forward and reverse primer sequences are located in their promoter and ORF regions, respectively. These promoter fragments were obtained by nested PCR using the Pro Taq DNA polymerase (Aikerui, Changsha, China) according to the manufacturer's instructions. Agarose gel electrophoresis was used to detect PCR products. Subsequently, these fragments were inserted into the pJET2.1 vector (Thermo Fisher Scientific, Waltham, MA, United States) and sequenced respectively. The *cis*-elements of promoter region in *AaCPS*, *AaECS*, and *AaBFS* were predicted using plantCARE online website.[1] The promoter sequences of sesquiterpene synthase genes were shown in **Supplementary Data Sheet 1**. Primers were listed in **Supplementary Table 1**.

Dual-luciferase assay

Dual-LUC assays were performed using the methods reported previously (Xiang et al., 2019). The promoter sequences of *AaCPS* (OP056317), *AaECS* (OP056318), and *AaBFS* (OP056319) genes were inserted into pGreenII 0800-LUC plasmid to generate pAaCPS:LUC, pAaECS:LUC, and pAaBFS:LUC constructs as the reporter vector, respectively. Subsequently, these reporter vectors were transferred into *A. tumefaciens* strain GV3101 together with the pSoup plasmid. The *AabHLH112* (MG872820) was inserted into the pHB plasmid driven by CaMV 35S promoter as the effector vector and transferred into *A. tumefaciens* strain GV3101. Meanwhile, the pHB-YFP plasmid was transferred into GV3101 as a negative control. All engineering and control strains were inoculated into YEP liquid selective medium and cultured overnight at 28°C. The agrobacterium cells were

1 http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

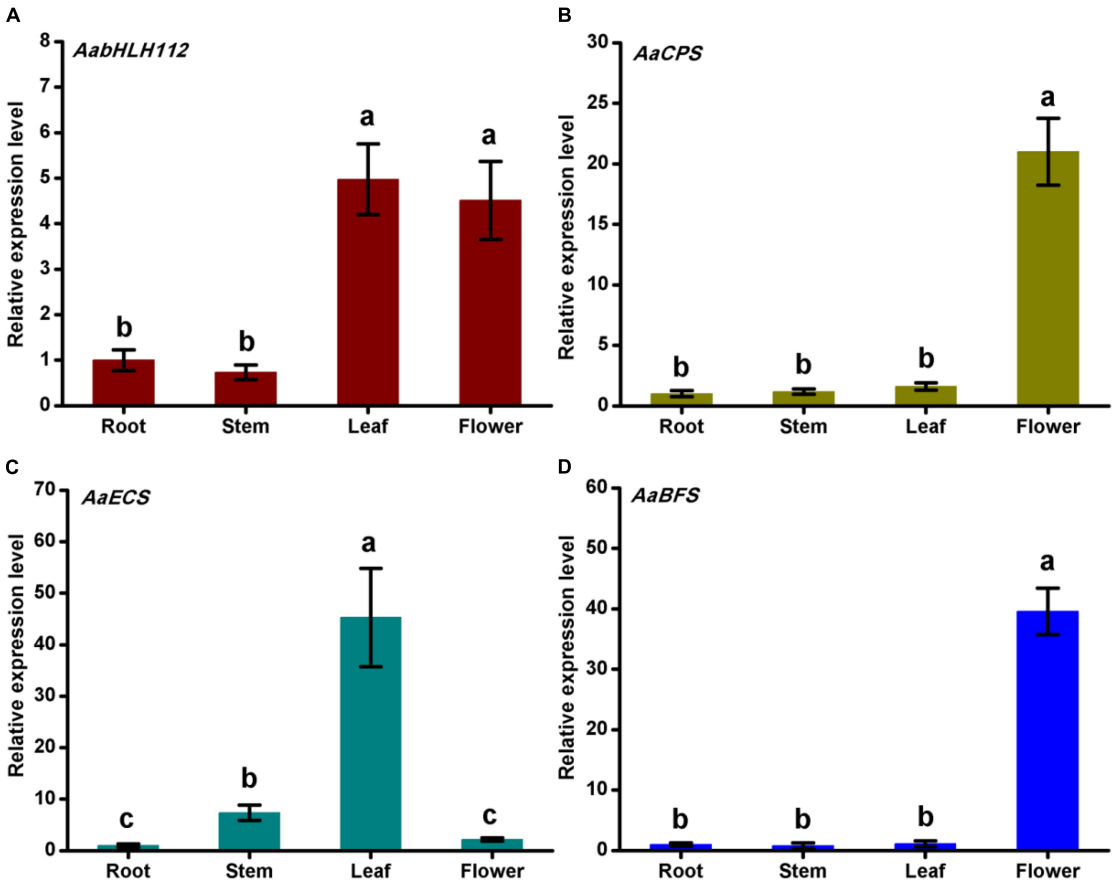


FIGURE 2
Genes expression levels of *AabHLH112* **(A)**, *AaCPS* **(B)**, AaECS **(C)**, and AaBFS **(D)** in root, stem, leaf, and flower of *Artemisia annua*. Bars are means ± SD from three independent biological replicates. One-way ANOVA was tested for significant differences among the means (indicated by different letters at $p < 0.01$).

collected by centrifuge at 4,500 rpm for 10 min and resuspended in the MS liquid to OD_{600} = 0.6 ± 0.05. The acetosyringone (As, 100 mM, 1:500, v:v) and 2-(*N*-morpholino) ethanesulfonic acid [MES, 0.5 M (pH = 5.7), 1:50, v:v] were added to the resuspension and then were injected into tobacco leaves after being placed for 4 h at room temperature. Tobacco plants injected with agrobacterium cells were exposed to weak light for 48–72 h. Infiltration and detection were performed as described previously (Ma et al., 2018). All experiments were repeated five times for each combination. Primers are listed in **Supplementary Table 1**.

Yeast one-hybrid assay

To investigate how AabHLH112 regulates the expression of *AacCPS*, *AaECS*, and *AaBFS*, yeast one-hybrid assays were fulfilled as described previously (Xiang et al., 2019). The coding sequence of *AabHLH112* was amplified and inserted into pB42AD plasmid with the GAL4 activation domain (AD) through *Eco*RI and *Xho*I sites to generate pB42AD-*AabHLH112* constructs as the prey. The 45∼55 bp fragments containing one G-box motifs from *AaCPS*, *AaECS*, and *AaBFS* promoters, named pAaCPS-G1 (−884∼−840), pAaCPS-G2 (−838∼−794), pAaCPS-G3 (−806∼−762), pAaCPS-G4 (−430∼−386), pAaECS-G1 (−1228∼−1178), pAaECS-G2 (−915∼−866), pAaECS-G3 (−385∼−336), pAaBFS-G1 (−1045∼−996), pAaBFS-G2 (−985∼−935), pAaBFS-G3 (−66∼−15) and the mutant of G-box were inserted into pLacZ plasmid through *Kpn*I and *Xho*I sites as the bait, respectively. The pB42AD-*AabHLH112* plasmid was co-transformed into yeast strain EGY48 with each individually bait constructs, respectively. The yeast cells were grown on SD-Ura-Trp selective medium for 48 h at 30°C. All independent yeast cells were shifted into SD-Ura-Trp liquid medium and cultured overnight at 30°C, and then these cells were collected by microcentrifugation and resuspended in 100 μL sterile water. Resuspended cells were grown on SD/-Ura-Trp medium

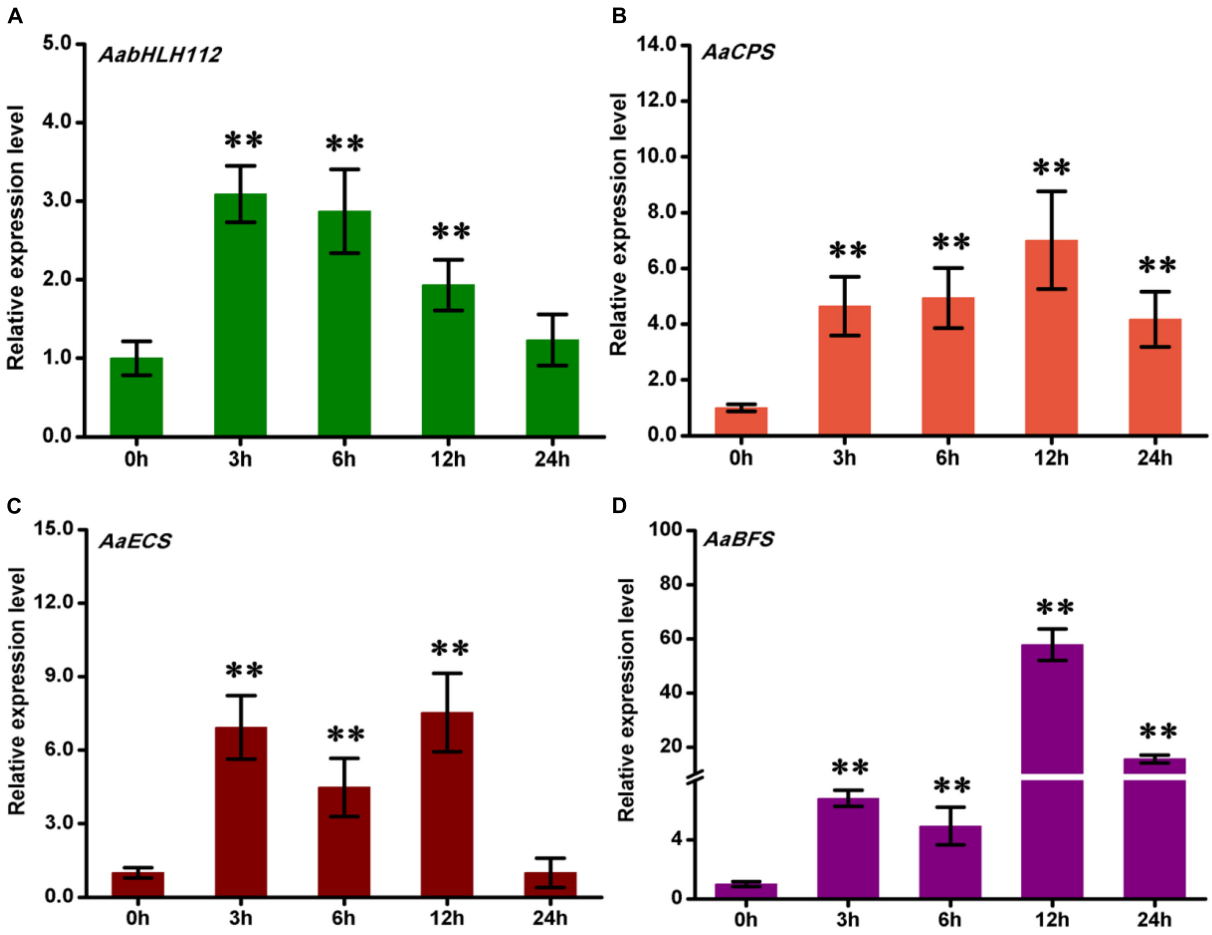


FIGURE 3
The expression of *AabHLH112* **(A)**, *AaCPS* **(B)**, *AaECS* **(C)**, and *AaBFS* **(D)** in response to MeJA treatment. Error bars represent the standard deviations of three technical replicates. Statistical significance was assessed with Student's *t*-test (**$p < 0.01$).

with 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) for 24 48 h at 30°C. The empty pB42AD and pLacZ plasmids were used as the negative control. Five independent biological replicates were implemented for each experiment in this study. The primers are listed in **Supplementary Table 1**.

Results

Tissues expression patterns of AabHLH112 and three sesquiterpene synthase genes

To study the tissue expression patterns of *AabHLH112* and sesquiterpenes biosynthesis genes, we measured their expression in various *A. annua* tissues including roots, stems, leaves, and flowers. *AabHLH112* was highly expressed in flowers and leaves, and lowly expressed in roots and stems (**Figure 2A**). *AaCPS* and *AaBFS* genes' expression has a tissue-specific mode, with a high expression in flowers and low expression in leaves, stems, and roots (**Figures 2B,D**). By comparison, *AaECS* was highly expressed in leaves and lowly expressed in roots and flowers (**Figure 2C**). Together, *AabHLH112*, *AaCPS*, *AaECS*, and *AaBFS* were expressed in all of the detected tissues.

Methyl jasmonate treatment induced the expression of *AabHLH112* and three sesquiterpene synthase genes

The phytohormone JA plays a critical role in plant stress resistance and secondary metabolism (Wasternack and Song, 2017; Kianersi et al., 2021). To investigate the response of *AabHLH112*, *AaCPS*, *AaECS*, and *AaBFS* to MeJA treatment, wild type *A. annua* plants were treated with exogenous MeJA, and all the genes' expression was

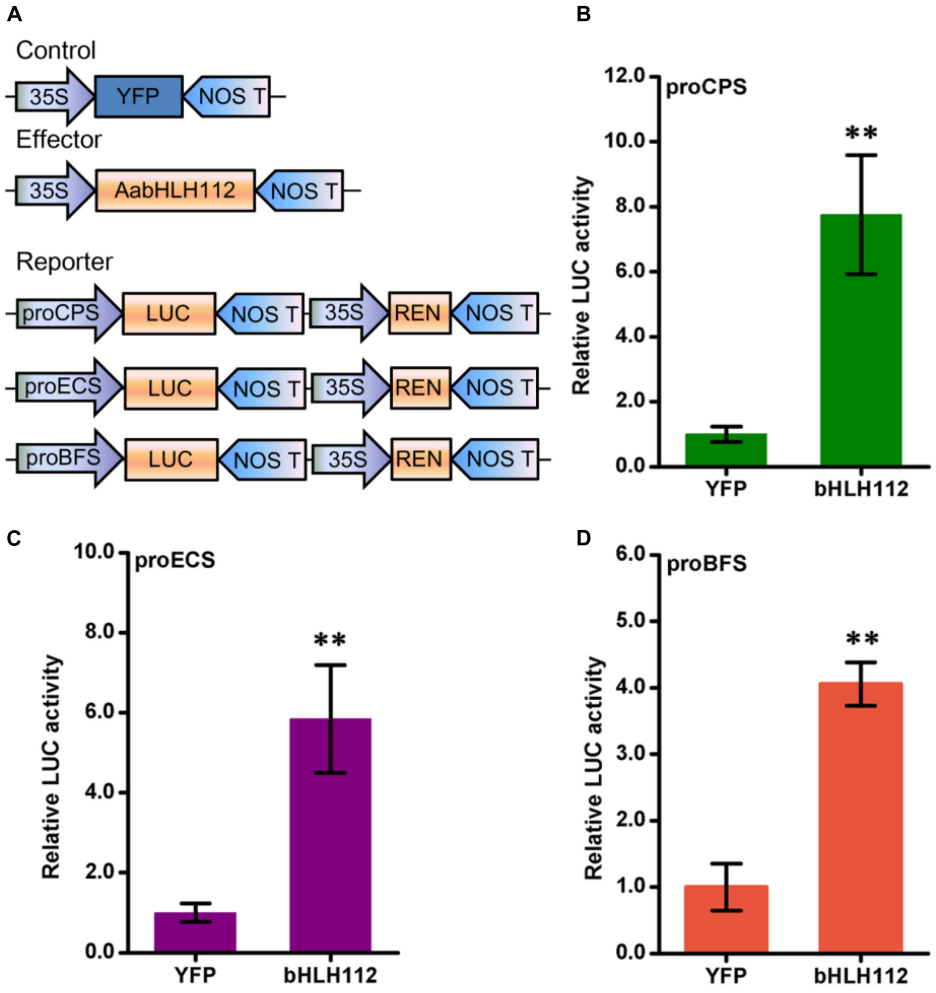


FIGURE 4
Biochemical assays between AabHLH112 and sesquiterpenes synthase genes. **(A)** Diagrams of reporter and effector constructs in transient dual-luciferase assays (REN, renilla luciferase; LUC, firefly luciferase). **(B–D)** Effects of AabHLH112 on activities of the *AaCPS*, *AaECS*, and *AaBFS* promoters in *Nicotiana benthamiana* leaves. The YFP effector was used as a negative control. The data represents the means ± SD (n = 3), $^{**}p < 0.01$ in student's t-test.

quantified by qPCR. The expression of *AabHLH112* was induced by MeJA, rapidly increased and peaked after 3 h of treatment, followed by a gradual decrease (Figure 3A). This result indicated that *AabHLH112* is involved in MeJA response in *A. annua*, as in the case with the *Arabidopsis* homologous *AtbHLH33* (Li et al., 2020). Several studies have shown that the expression of many genes involved in terpenes biosynthesis is significantly increased under JA treatment. Here, we found the expression levels of *AaECS*, *AaCPS*, and *AaBFS* were significantly upregulated under exogenous MeJA treatment (Figures 3B–D); besides, the accumulation of β-caryophyllene, *epi*-cedrol, and β-farnesene, direct products of the three enzymes, also showed an upward trend after MeJA treatment (Supplementary Figure 1). These results suggest that MeJA can induce the expression of these sesquiterpenes synthase genes, thus enhancing the biosynthesis of β-caryophyllene, *epi*-cedrol and β-farnesene in *A. annua*.

AabHLH112 activates the expression of the *AaECS*, *AaCPS*, and *AaBFS*

Dual-LUC assays can be used to study the transcriptional regulation of transcription factors on target genes at the transcriptional level (Ma et al., 2018). The coding sequence of *AabHLH112* was inserted into the plant overexpression vector pHB as an effector construct, and the promoter sequences of *AaCPS* (Supplementary Figure 2), *AaECS* (Supplementary Figure 3), and *AaBFS* (Supplementary Figure 4) genes were inserted into pGreenII 0800-LUC plasmid to generate pAaCPS:LUC, pAaECS:LUC, and pAaBFS:LUC constructs as

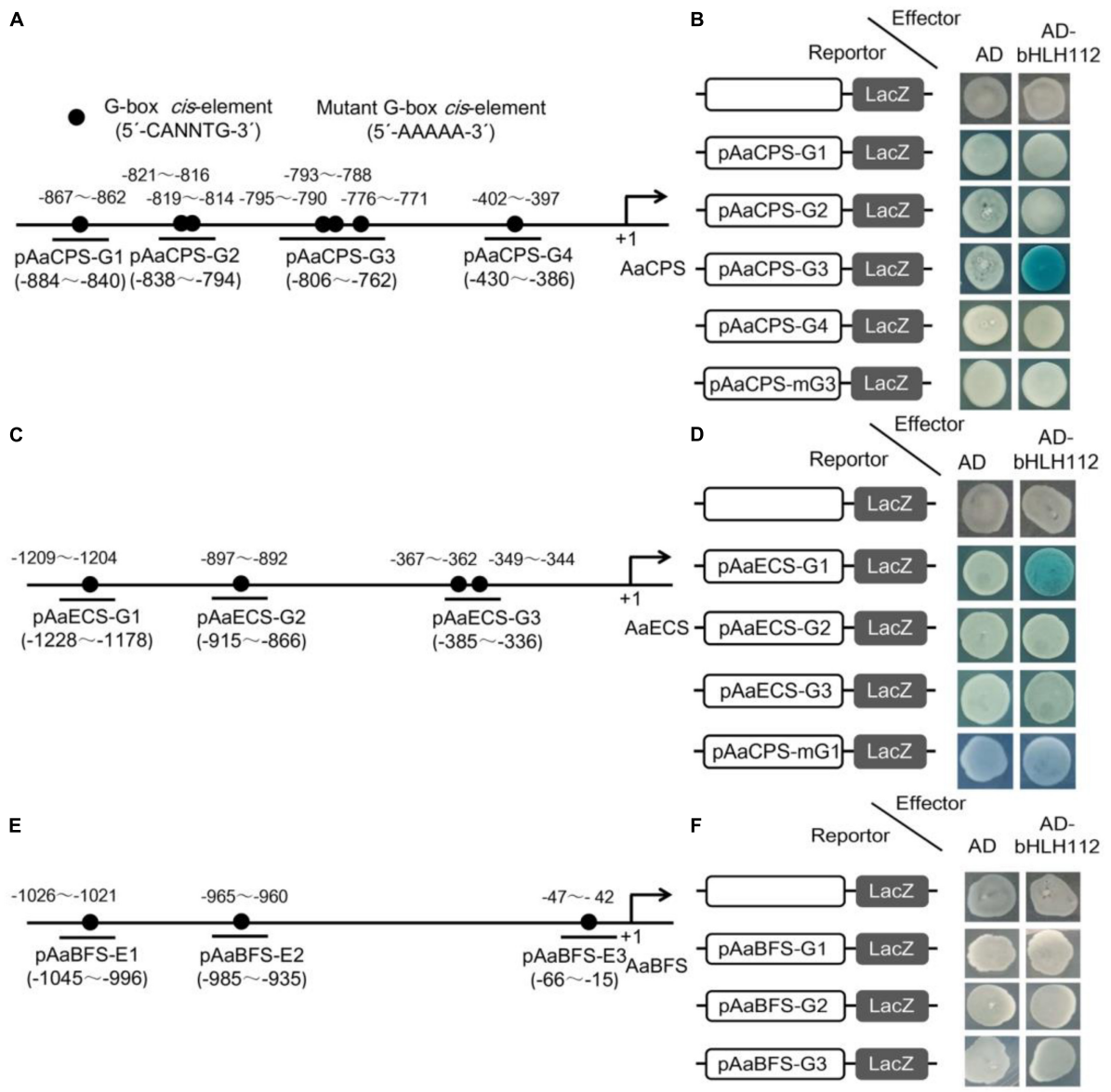


FIGURE 5
Yeast one-hybrid assays between AabHLH112 and G-box *cis*-elements of sesquiterpene synthase gene promoters, blue plaques indicate protein-DNA interactions. **(A)** Schematic diagram of the *AaCPS* promoter; **(B)** Y1H assay between AabHLH112 and *AaCPS* promoter; **(C)** Schematic diagram of the *AaECS* promoter; **(D)** Y1H assay between AabHLH112 and *AaECS* promoter; **(E)** Schematic diagram of the *AaBFS* promoter; **(F)** Y1H assay between AabHLH112 and *AaBFS* promoter; the black circles were the potential G-box elements in three promoters.

the reporter vector, respectively. Meanwhile, the pHB-YFP (yellow fluorescent protein) plasmid was used as a negative control (Figure 4A). The effector and reporter constructs were transiently expressed in *N. benthamiana* leaves using *Agrobacterium tumefaciens*-mediated co-infiltration. The dual-LUC assay results showed that the promoter activity of the above sesquiterpene synthase genes *AaCPS*, *AaECS*, and *AaBFS* was significantly increased, with the LUC/REN value increased by 7.75, 5.58, and 4.05-folds than the YFP control (Figures 4B–D), respectively.

AabHLH112 bind to G-box *cis*-elements in the promoters of *AaCPS* and *AaECS*

The core DNA sequence motif recognized by bHLH proteins is a consensus hexanucleotide sequence known as the G-box (5-CANNTG-3). It was previously reported that AaMYC2 can specifically bind to the G-box *cis*-elements in the promoters of key enzyme genes of artemisinin biosynthesis, including *AaCYP71AV1* and *AaDBR2* (Shen et al., 2016).

Recently, Xiang et al. (2019) reported that AabHLH112 can directly bind to G-box in the *AaERF1* promoter and promote artemisinin biosynthesis. The above dual-LUC results showed that AabHLH112 could activate the transcription of *AaCPS*, *AaECS*, and *AaBFS*. In order to further investigate how AabHLH112 regulates the transcription of these sesquiterpene synthase genes, the promoter sequences of *AaCPS*, *AaECS*, and *AaBFS* were analysed by PlantCARE online website, and multiple G-box motifs were found in each promoter (**Figures 5A,C,E**). Following that, yeast one-hybrid assay was used to evaluate whether AabHLH112 can directly bind to the G-box motifs in these promoters. 45 50 bp fragments containing G-box motifs from each promoter were inserted into pLacZ plasmid as the bait, and *AabHLH112* was inserted into pB42AD plasmid with the GAL4 activation domain to generate pB42AD-*AabHLH112* constructs as the prey. The results indicated that AabHLH112 can directly bind to pAaCPS-G3 and pAaECS-G1 fragments containing the G-box in *AaCPS* and *AaECS* promoters (**Figures 5B,D**), but not the pAaCPS1-G1, pAaCPS-G2, pAaCPS-G4, pAaECS-G2, pAaECS-G3, and pAaBFS-G1, pAaBFS-G2, pAaBFS-G3 fragments. When the G-box *cis*-elements in the pAaCPS-G3 and pAaECS-G1 fragments were mutated to the 5'-AAAAA-3' sequences resulting in pAaCPS-mG3 and pAaECS-mG1, respectively, the binding signals of AabHLH112 and pAaCPS-mG3 and pAaECS-mG1 disappeared (**Figures 5B,D**). These results demonstrate that AabHLH112 enhances *AaCPS* and *AaECS* transcription by directly binding to the specific G-box *cis*-elements in their promoters.

Generation of *AabHLH112*-overexpression and RNAi trangenic *Artemisia annua*

To better explore the function of *AabHLH112* in sesquiterpenes biosynthesis in *A. annua*, *AabHLH112*-overexpressing and *AabHLH112*-RNAi transgenic *A. annua* plants were generated (**Supplementary Figure 5**). Primers designed according to the sequences of rubisco gene (*rbc*), *AabHLH112* and marker gene (hygromycin-resistant) were used to determine the regeneration plants by genomic PCR. The 877bp fragment of the hygromycin-resistant gene and coding sequence of AabHLH112 with partial *rbc* terminator (1383bp) were specifically amplified from the transformed plants and the positive control (pHB-*AabHLH112*), while these specific fragments could not be amplified from wild-type plants (**Supplementary Figures 5B,C**). Besides, *AabHLH112*-RNAi transgenic plants were also generated and detected by genomic PCR. Specific detection primers were designed according to the sequences of *AabHLH112*, *OCS* terminator and *NptII* (kanamycin) genes. As shown in **Supplementary Figures 5D,E**, the partial *AabHLH112* sequence and *NptII* gene were detected in the *AabHLH112*-RNAi transgenic plants.

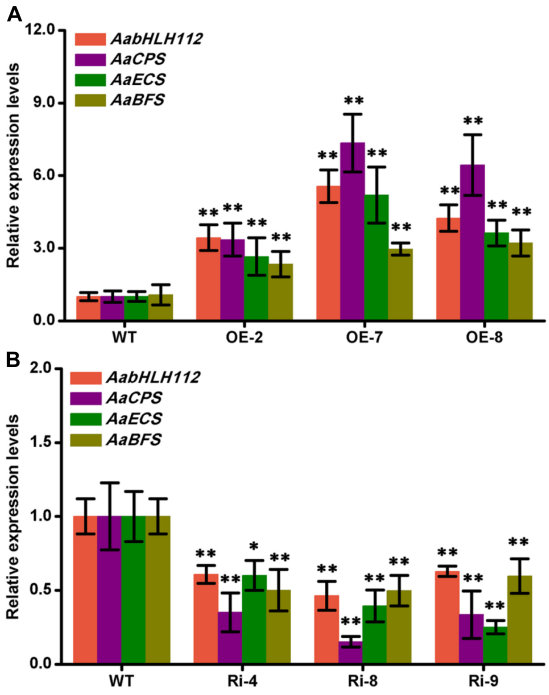


FIGURE 6
Genes relative expression levels in wild-type, *AabHLH112*-overexpression and *AabHLH112*-RNAi *Artemisia annua* plants. **(A)**: Relative expression levels of *AabHLH112*, *AaCPS*, *AaECS*, and *AaBFS* in wild type and *AabHLH112*-overexpression transgenic plants; **(B)**: Relative expression levels of *AabHLH112*, *AaCPS*, *AaECS*, and *AaBFS* in wild type and *AabHLH112*-RNAi transgenic plants. Error bars represent the standard deviations of three technical replicates. Statistical significance was assessed with Student's *t*-test (*$p < 0.05$ and **$p < 0.01$).

AabHLH112 positively regulates the expression of *AaCPS*, *AaECS*, and *AaBFS* in *Artemisia annua*

The expression level of *AabHLH112* in transgenic *A. annua* lines was determined by qPCR. Compared with the wild type (WT) *A. annua*, *AabHLH112* presented higher expression level in *AabHLH112*-overexpression lines of OE-2, OE-7, and OE-8 with 3.44-fold, 5.56-fold, and 4.24-fold higher, respectively (**Figure 6A**). Meanwhile, the expression levels of *AaCPS*, *AaECS*, and *AaBFS* in the three lines (OE-2, OE-7, and OE-8) were approximately 3.35-7.34 fold, 2.65-9.15 fold, and 2.24-2.55 fold higher compared to the WT respectively (**Figure 6A**). On the contrary, The expression level of *AabHLH112* were markedly decreased in *AabHLH112*-RNAi lines of Ri-4, Ri-8, and Ri-9 (**Figure 6B**), and the expression of *AaCPS*, *AaECS*, and *AaBFS* was also significantly reduced (**Figure 6B**) in the three *AabHLH112*-RNAi lines compared to WT plants. These results suggest that AabHLH112 positively regulates the expression of *AaCPS*, *AaECS* and *AaBFS* in *A. annua*.

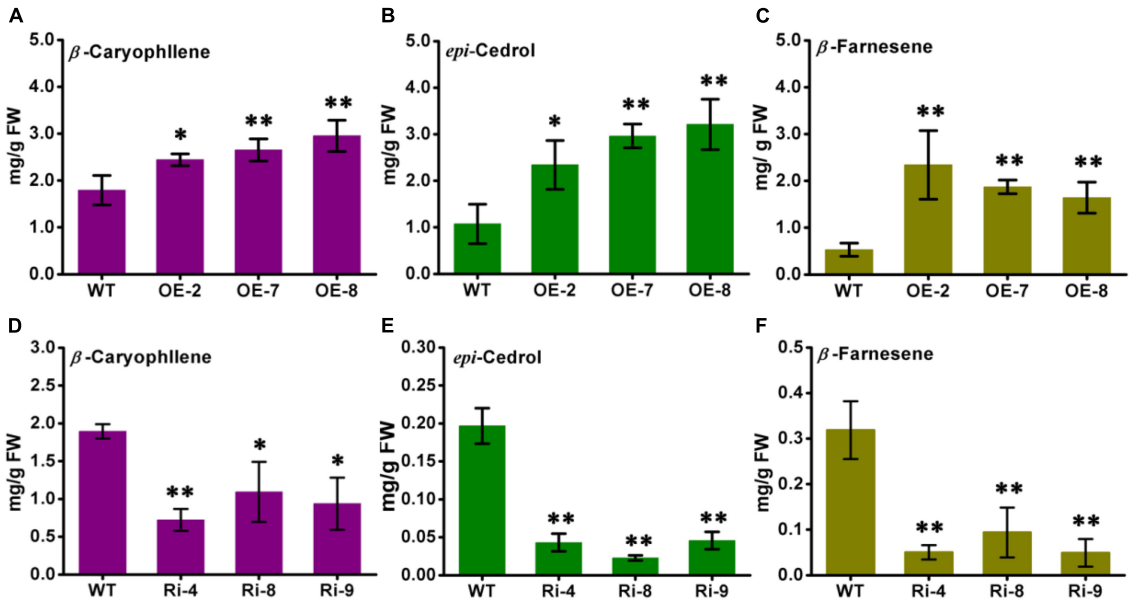


FIGURE 7
The contents of β-caryophyllene, *epi*-cedrol, and β-farnesene in wild-type, *AabHLH112*-overexpression, and *AabHLH112*-RNAi Artemisia annua plants. **(A)** β-caryophllene, **(B)** epi-cedrol, and **(C)** β-farnesene in *AabHLH112*-overexpressing lines, WT, wild type; OE-2,OE-7 and OE-8 are independent lines of *AabHLH112*-overexpressing lines; **(D)** β-caryophllene, **(E)** epi-cedrol and **(F)** β-farnesene in *AabHLH112*-RNAi lines, WT, wild type; Ri-4, Ri-8, and Ri-9 are independent lines of *AabHLH112*-RNAi lines, the data represents the means ± SD (n = 3), *p < 0.05, **p < 0.01 in student's t-test.

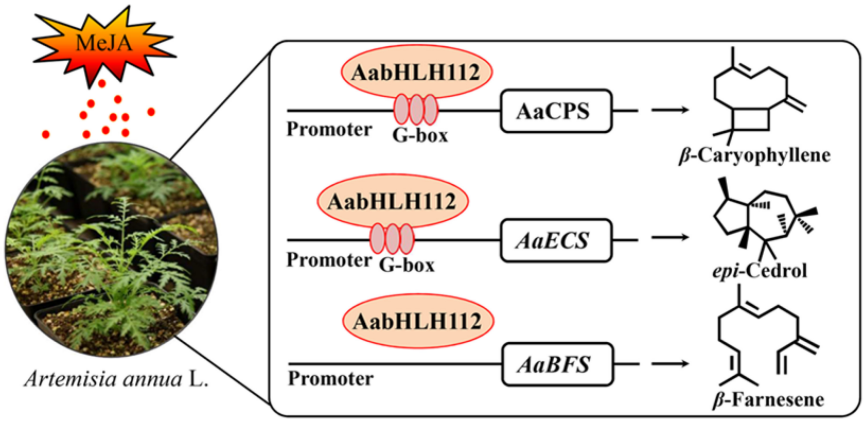


FIGURE 8
A simplified model for the regulation of sesquiterpenes biosynthesis by AabHLH112 in *Artemisia annua*.

AabHLH112 positively regulates β-caryophyllene, *epi*-cedrol, and β-farnesene biosynthesis in *Artemisia annua*

The contents of β-caryophyllene, *epi*-cedrol, and β-farnesene in the *AabHLH112*-overexpression and RNAi transgenic lines were detected by the GC-MS method. The wild type *A. annua* produced 1.87 mg/g FW (fresh weight) of β-caryophyllene, 1.07 mg/g FW of *epi*-cedrol and 0.53 mg/g FW of β-farnesene, while the three independent *AabHLH112*-overexpression lines (OE-2, OE-7, and OE-8) exhibited significant increase in sesquiterpenes contents. OE-2, OE-7, and OE-8 contained β-caryophyllene at levels of 2.34 mg/g FW, 2.45 mg/g FW, and 2.95 mg/g FW, respectively, representing 25.13, 31.01, and 57.75% increases in content compared with WT plants (Figure 7A). Compared to WT plants, the content of *epi*-cedrol was increased by 118, 176, and 200% in the three

AabHLH112-overexpression lines respectively (**Figure 7B**); and the content of β-farnesene was increased by 341, 253, and 201% in the three *AabHLH112*-overexpression lines (**Figure 7C**). In the *AabHLH112*-RNAi lines of Ri-4, Ri-8 and Ri-9, the contents of β-caryophyllene, *epi*-cedrol and β-farnesene were significantly reduced, compared to that in WT plants. Ri-4, Ri-8, and Ri-9 lines contained β-caryophyllene at levels of 0.73 mg/g FW, 1.09 mg/g FW, and 0.94 mg/g FW respectively, representing 61.91, 42.28, and 50.54% reduction in content compared with WT plants (**Figure 7D**). *Epi*-cedrol content was reduced by 78.12, 88.45, and 76.82% in Ri-4, Ri-8 and Ri-9 lines respectively, relative to that in WT plants (**Figure 7E**). β-farnesene content in Ri-4, Ri-8, and Ri-9 lines was reduced by 84.27, 70.55, and 84.58% respectively, compared with that in WT plants (**Figure 7F**). These results showed that AabHLH112 is a positive regulator of β-caryophyllene, *epi*-cedrol, and β-farnesene biosynthesis in *A. annua*.

Discussion

Artemisia annua is a traditional Chinese medicinal plant, not only producing the well-known drug artemisinin, but also producing other important sesquiterpenes of medicinal value, such as β-caryophyllene and *epi*-cedrol. The bHLH TFs are the second largest transcription factor family in the plants, which play an important role in plant growth, development, and secondary metabolism. Previous research revealed that bHLH TFs play an important part in the regulation of terpenoids biosynthesis. In tomato, overexpression of CubHLH1 resulted in a significant increase in carotenoid content (Endo et al., 2016). MYC, one subgroup of bHLH family, plays an important role in JA, ABA, and GA signaling and is involved in the regulation of terpenoids biosynthesis in plants (Hong et al., 2012; Du et al., 2017). In *A. thaliana*, *AtMYC2* could directly bind to promoters of the sesquiterpene synthase genes *TPS21* and *TPS11* and activate their expression, thus increasing sesquiterpenes biosynthesis (Hong et al., 2012). *AaMYC2* overexpression could activate the transcription of *CYP71AV1* and *DBR2*, resulting in an increase of artemisinin content in *A. annua* (Shen et al., 2016). Besides, AabHLH1 positively regulated biosynthesis of artemisinin *via* activating the promoters of *ADS* and *CYP71AV1* (Ji et al., 2014). In this study, we found that AbHLH112 positively regulates the biosynthesis of β-caryophyllene, *epi*-cedrol, and β-farnesene in *A. annua* besides upregulating artemisinin biosynthesis.

The gene expression was synchronous with its product. In this study, we measured the expression levels of *AabHLHH112*, *AaCPS*, *AaECS*, and *AaBFS* in roots, stems, leaves and flowers. And we found that *AabHLH112* has the highest expression in leaves and flowers and the lowest expression in stems and roots, which has a similar expression pattern with *AaADS* involved in artemisinin biosynthesis (Pu et al., 2013). *AaCPS* and *AaBFS*, involved in β-caryophyllene and β-farnesene biosynthesis respectively, have a tissue-specific expression pattern, with the highest expression in flowers and the lowest expression in roots, stems and leaves. By contrast, *AaECS* is highly expressed in leaves and lowly expressed in roots and flowers, showing a tissue difference in the expression of these genes.

Methyl jasmonate (MeJA) is one of the most potent elicitors that can induce overaccumulation of many natural products in plants, such as artemisinin, tanshinones, vinblastine, and so on (Xiang et al., 2015; Shi et al., 2016; Zhang et al., 2018). Later studies showed that exogenous MeJA activated the transcription of some genes leading to the products' overaccumulation (Wang et al., 2020). In this study, the 1-month-old *A. annua* seedlings were treated with MeJA for up to 24 h and the four genes expression were quantified by qPCR. The results showed that the genes detected had a higher expression after the MeJA treatment, indicating that exogenous MeJA treatment can induce the expression of sesquiterpene biosynthesis genes in *A. annua*. In addition, we found that the expression level of *AabHLH112* was induced as early as 3 h after the treatment, and then began to decline, while the highest expression levels of *AaECS* and *AaCPS* were detected at 12 h after MeJA treatment. This result suggested that these sesquiterpenes synthase genes might be regulated by *AabHLH112*. To further explore the function of AabHLH112 in the regulation of sesquiterpene biosynthesis genes expression, dual-LUC assay was taken to determine the activation of *AaCPS*, *AaECS*, and *AaBFS* promoters by *AabHLH112*. The results indicated AabHLH112 can significantly enhance the transcriptional activity of *AaCPS*, *AaECS*, and *AaBFS* promoters in *N. benthamiana* leaves (**Figures 4B–D**).

It is known that G-box is a putative recognition site for bHLH TFs. We have reported in a previous study that AabHLH112 can bind to the G-box in the *AaERF1* promoter and enhanced its transcription to upregulate artemisinin biosynthesis (Xiang et al., 2019). In this study, yeast one-hybrid assay (Y1H) showed that AabHLH112 can directly bind to pAaCPS-G3 and pAaECS-G1 fragments containing the G-box in the promoters of *AaCPS* and *AaECS*, but not bind to pAaCPS1-G1, pAaCPS-G2, pAaCPS-G4, pAaECS-G2, pAaECS-G3, and pAaBFS-G1, pAaBFS-G2, pAaBFS-G3 fragments (**Figures 5B,D,F**). Based on these results, we estimated that the flanking sequence of G-box *cis*-elements may affect the recognition of AabHLHL112 protein. In addition, there are three G-box *cis*-elements in the pAaCPS-G3 fragment and they are very close to each other. So, a profound study should be carried out to analyze which G-box is the real recognition site by AabHLH112. Interestingly, AabHLH112 positively regulated *AaBFS* expression to increase β-farnesene accumulation, but Y1H result showed there was no direct interaction between AabHLH112 and *AaBFS* promoter, suggesting that AabHLH112 upregulated β-farnesene biosynthesis through an indirect pathway, which deserves further exploration.

Taken together, our study indicated that AabHLH112, which is induced by MeJA, positively regulates the expression of *AaCPS*, *AaECS*, and *AaBFS* through directly binding to their promoters or an indirect pathway, thereby enhancing the biosynthesis of β-caryophyllene, *epi*-cedrol, and β-farnesene in *Artemisia annua* (**Figure 8**). Our study revealed the mechanism of AabHLH112-mediated upregulation of sesquiterpenes biosynthesis, and provided a potentially useful transcription factor that could be used to improve sesquiterpenes production in *A. annua*.

Data availability statement

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

Author contributions

LX, ZL, and YT conceived and coordinated the study and wrote the manuscript. LX performed the gene cloning, vector construction, and plant transformation work. GS, MY, and MW did the gene expression, luciferase analysis, and Y1H experiments. GS, XL, and PH helped with the metabolites analysis by GC-MS. All authors reviewed the results and approved the final version of the manuscript.

Funding

This research was financially supported by the Science Funding of Sichuan Province (2020YJ0171, China), the NSFC projects (81973420 and 81803660), the National Key Research and Development Project (2019YFE0108700, China), and the Forth National Survey of Traditional Chinese Medicine Resources, Chinese or Tibet Medicinal Resources Investigation in Tibet Autonomous Region (20191217-540124, 20191223-540126, and 20200501-542329).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.973591/full#supplementary-material

References

Bahi, A., Mansouri, S., Memari, E., Ameri, M., Nurulain, S., and Ojha, S. (2014). β-Caryophyllene, a CB2 receptor agonist produces multiple behavioral changes relevant to anxiety and depression in mice. *Physiol. Behav.* 135, 119–124. doi: 10.1016/j.physbeh.2014.06.003

Basha, R., and Sankaranarayanan, C. (2014). β-Caryophyllene, a natural sesquiterpene, modulates carbohydrate metabolism in streptozotocin-induced diabetic rats. *Acta Histochem.* 116, 1469–1479. doi: 10.1016/j.acthis.2014.10.001

Bento, A., Marcon, R., Dutra, R., Claudino, R., Cola, M., Leite, D., et al. (2011). β-Caryophyllene inhibits dextran sulfate sodium-induced colitis in mice through CB2 receptor activation and PPARγ pathway. *Am. J. Pathol.* 178, 1153–1166. doi: 10.1016/j.ajpath.2010.11.052

Cao, W., Wang, Y., Shi, M., Hao, X., Zhao, W., Wang, Y., et al. (2018). Transcription factor SmWRKY1 positively promotes the piosynthesis of tanshinones in *Salvia miltiorrhiza*. *Front. Plant Sci.* 9:554. doi: 10.3389/fpls.2018.00554

Chadwick, M., Trewin, H., Gawthrop, F., and Wagstaff, G. (2013). Sesquiterpenoids lactones: benefits to plants and people. *Int. J. Mol. Sci.* 14, 12780–12805. doi: 10.3390/ijms140612780

Chang, H., Kim, J., Lee, J., Kim, W., and Chun, H. (2013). Protective effect of β-caryophyllene, a natural bicyclic sesquiterpene, against cerebral ischemic injury. *J. Med. Food* 16, 471–480. doi: 10.1089/jmf.2012.2283

Cheng, T., Zhang, K., Guo, J., Yang, Q., Li, T., Xiao, M., et al. (2022). Highly efficient biosynthesis of β-caryophyllene with a new sesquiterpene synthase from tobacco. *Biotechnol. Biofuels Bioproducts* 15:39. doi: 10.1186/s13068-022-02136-8

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

Du, M., Zhao, J., Tzeng, D., Liu, Y., Deng, W., Ynag, T., et al. (2017). MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. *Plant Cell* 29, 1883–1906. doi: 10.1105/tpc.16.00953

Endo, T., Fujii, H., Sugiyama, A., Nakano, M., Nakajima, N., and Ikoma, Y. (2016). Overexpression of a citrus basic helix-loop-helix transcription factor (CubHLH1), which is homologous to Arabidopsis activation-tagged bri1 suppressor 1 interacting factor genes, modulates carotenoid metabolism in transgenic tomato. *Plant Sci.* 243, 35–48. doi: 10.1016/j.plantsci.2015.11.005

Fu, X., Shi, P., He, Q., Shen, Q., Tang, Y., Pan, Q., et al. (2017). AaPDR3, a PDR transporter 3, is involved in sesquiterpene β-caryophyllene transport in Artemisia annua. *Front. Plant Sci.* 8:723. doi: 10.3389/fpls.2017.00723

Gantet, P., and Memelink, J. (2002). Transcription factors: tools to engineer the production of pharmacologically active plant metabolite. *Trends Pharmacol. Sci.* 23, 563–569. doi: 10.1016/s0165-6147(02)02098-9

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

Huang, Q., Sun, M., Yuan, T., Wang, Y., Shi, M., and Lu, S. (2019). The AP2/ERF transcription factor SmERF1L1 regulates the biosynthesis of tanshinones and phenolic acids in Salvia miltiorrhiza. *Food Chem.* 274, 368–375. doi: 10.1016/j.foodchem.2018.08.119

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

Kianersi, F., Pour-Aboughadareh, A., Majdi, M., and Poczai, P. (2021). Effect of methyl jasmonate on thymol, carvacrol, phytochemical accumulation, and expression of key genes involved in thymol/carvacrol biosynthetic pathway in some iranian thyme species. *Int. J. Mol. Sci.* 22:11124. doi: 10.3390/ijms222011124

Legault, J., and Pichette, A. (2007). Potentiating effect of β-caryophyllene on anticancer activity of α-humulene, isocaryophyllene and paclitaxel. *J. Pharmacy Pharmacol.* 59, 1643–1647. doi: 10.1211/jpp.59.12.0005

Lenka, S., Nims, N., Vongpaseuth, K., Boshar, R., Roberts, S., and Walker, E. (2015). Jasmonate-responsive expression of paclitaxel biosynthesis genes in Taxus cuspidata cultured cells is negatively regulated by the bHLH transcription factors TcJAMYC1. TcJAMYC2, and TcJAMYC4. *Front. Plant Sci.* 6:115. doi: 10.3389/fpls.2015.00115

Li, X., Xu, Y., Shen, S., Yin, X., Klee, H., Zhang, B., et al. (2017). Transcription factor CitERF71 activates the terpene synthase gene CitTPS16 involved in the synthesis of E-geraniol in sweet orange fruit. *J. Exp. Bot.* 68, 4929–4938. doi: 10.1093/jxb/erx316

Li, Y., Li, L., Ding, W., Li, H., Shi, T., Yang, X., et al. (2020). Genome-wide identification of osmanthus fragrans bHLH transcription factors and their expression analysis in response to abiotic stress. *Environ. Exp. Bot.* 172:103990.

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

Lu, X., Tang, K., and Li, P. (2016). Plant metabolic engineering strategies for the production of pharmaceutical terpenoids. *Front. Plant Sci.* 7:1647. doi: 10.3389/fpls.2016.01647

Luo, F., Ling, Y., Li, Den, Tang, T., Liu, Y. C., Liu, Y., et al. (2019). Characterization of a sesquiterpene cyclase from the glandular trichomes of Leucosceptrum canum for sole production of cedrol in *Escherichia coli* and Nicotiana benthamiana. *Phytochemistry* 162, 121–128. doi: 10.1016/j.phytochem.2019.03.009

Ma, Y., Xu, D., Li, L., Zhang, F., Fu, X., Shen, Q., et al. (2018). Jasmonate promotes artemisinin biosynthesis by activating the TCP14-ORA complex in Artemisia annua. *Sci. Adv.* 4:eaas9357. doi: 10.1126/sciadv.aas9357

Mahmoud, M., Swefy, S., Hasan, R., and Ibrahim, A. (2014). Role of cannabinoid receptors in hepatic fibrosis and apoptosis associated with bile duct ligation in rats. *Eur. J. Pharmacol.* 742, 118–124. doi: 10.1016/j.ejphar.2014.08.021

Navale, G., Sharma, P., Said, M., Ramkumar, S., Dharne, M., Thulasiram, H., et al. (2019). Enhancing epi-cedrol production in *Escherichia coli* by fusion expression of farnesyl pyrophosphate synthase and epi-cedrol synthase. *Eng. Life Sci.* 19, 606–616. doi: 10.1002/elsc.201900103

Patra, B., Pattanaik, S., Schluttenhofer, C., and Yuan, L. (2018). A network of jasmonate-responsive bHLH factors modulate monoterpenoid indole alkaloid biosynthesis in Catharanthus roseus. *New Phytol.* 4, 1566–1581. doi: 10.1111/nph.14910

Pu, G., Ma, D., Wang, H., Ye, H., and Liu, B. (2013). Expression and localization of amorpha-4,11-diene synthase in Artemisia annua L. *Plant Mol. Biol. Rep.* 31, 32–37. doi: 10.1007/s11105-012-0472-0

Rufino, A., Ribeiro, M., Sousa, C., Judas, F., Salgueiro, L., Carlos Cavaleiro, C., et al. (2015). Evaluation of the anti-inflammatory, anti-catabolic and pro-anabolic effects of E-caryophyllene, myrcene and limonene in a cell model of osteoarthritis. *Eur. J. Pharmacol.* 750, 141–150. doi: 10.1016/j.ejphar.2015.01.018

Shen, Q., Lu, X., Yan, T., Fu, X., Lv, Z., Zhang, F., et al. (2016). The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in Artemisia annua. *New Phytol.* 210, 1269–1281. doi: 10.1111/nph.13874

Shi, M., Zhou, W., Zhang, J., Huang, S., Wang, H., and Kai, G. (2016). Methyl jasmonate induction of tanshinone biosynthesis in Salvia miltiorrhiza hairy roots is mediated by JASMONATE ZIM-DOMAIN repressor proteins. *Sci. Rep.* 15:20919. doi: 10.1038/srep20919

Wang, G., Yu, D., Fan, J., Wang, C., and Xia, L. (2014). Expressing an (E)-β-farnesene synthase in the chloroplast of tobacco affects the preference of green peach aphid and its parasitoid. *Plant Biol. J. Integr.* 57, 770–782. doi: 10.1111/jipb.12319

Wang, J., Song, L., Gong, X., Xu, J., and Li, M. (2020). Functions of jasmonic acid in plant regulation and response to abiotic stress. *Int. J. Mol. Sci.* 21:1446. doi: 10.3390/ijms21041446

Wasternack, C., and Song, S. (2017). Jasmonates: biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. *J. Exp. Bot.* 68, 1303–1321. doi: 10.1093/jxb/erw443

Xiang, L., Jian, D., Zhang, F., Yang, C., Bai, G., Lan, X., et al. (2019). The cold-induced transcription factor bHLH112 promotes artemisinin biosynthesis indirectly via ERF1 in Artemisia annua. *J. Exp. Bot.* 70, 4835–4848. doi: 10.1093/jxb/erz220

Xiang, L., Zhu, S., Zhao, T., Zhang, M., Liu, W., Chen, M., et al. (2015). Enhancement of artemisinin content and relative expression of genes of artemisinin biosynthesis in Artemisia annua by exogenous MeJA treatment. *Plant Growth Regul.* 75, 435–441. doi: 10.1007/s10725-014-0004-z

Xu, J., Herwijnen, Z. O., Dräger, D., Sui, C., Haring, M., and Schuurink, R. (2018). SlMYC1 regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. *Plant Cell* 30, 2988–3005. doi: 10.1105/tpc.18.00571

Yang, C., Fang, X., Wu, X., Mao, Y., Wang, L., and Chen, X. Y. (2012). Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* 54, 703–712. doi: 10.1111/j.1744-7909.2012.01161.x

Yin, J., Li, X., Zhan, Y., Li, Y., Qu, Z., Sun, L., et al. (2017). Cloning and expression of BpMYC4 and BpbHLH9 genes and the role of BpbHLH9 in triterpenoid synthesis in birch. *BMC Plant Biol.* 17:214. doi: 10.1186/s12870-017-1150-z

Yu, X., Pickett, J., Ma, Y., Bruce, T., Napier, J., Huw, D., et al. (2012). Metabolic engineering of plant-derived (E)-β-farnesene synthase genes for a novel type of aphid-resistant genetically modified crop plants. *J. Integr. Plant Biol.* 54, 282–299. doi: 10.1111/j.1744-7909.2012.01107.x

Yu, Z., Li, J., Yang, C., Hu, W., Wang, L., and Chen, X. (2012). The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in Artemisia annua L. *Mol. Plant* 5, 353–365. doi: 10.1093/mp/ssr087

Zhang, C., Xing, B., Yang, D., Ren, M., Guo, H., Yang, S., et al. (2020). SmbHLH3 acts as a transcription repressor for both phenolic acids and tanshinone biosynthesis in Salvia miltiorrhiza hairy roots. *Phytochemistry.* 169:112183. doi: 10.1016/j.phytochem.2019.112183

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus. *Plant J.* 67, 61–71. doi: 10.1111/j.1365-313X.2011.04575.x

Zhang, X., Liu, J., Liu, Y., Wang, Y., Abozeid, A., Yu, Z., et al. (2018). Metabolomics Analysis reveals that ethylene and methyl jasmonate regulate different branch pathways to promote the accumulation of terpenoid indole alkaloids in Catharanthus roseus. *J. Nat. Products* 81, 335–342. doi: 10.1021/acs.jnatprod.7b00782

Zhang, Y., Han, L., Chen, S., Guan, J., Qu, F. Z., Zhao, Y. Q., et al. (2016). Hair growth promoting activity of cedrol isolated from the leaves of Platycladus orientalis. *Biomed. Pharmacother.* 83, 641–647. doi: 10.1016/j.biopha.2016.07.022

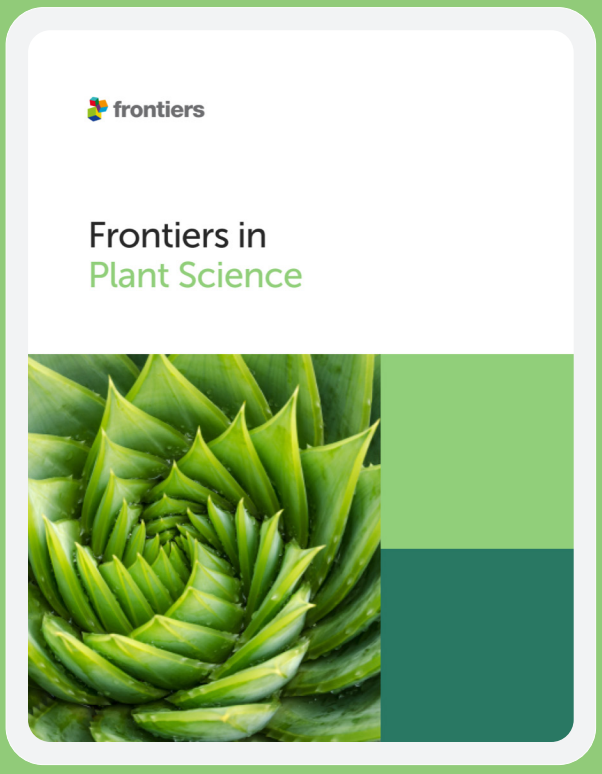
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact