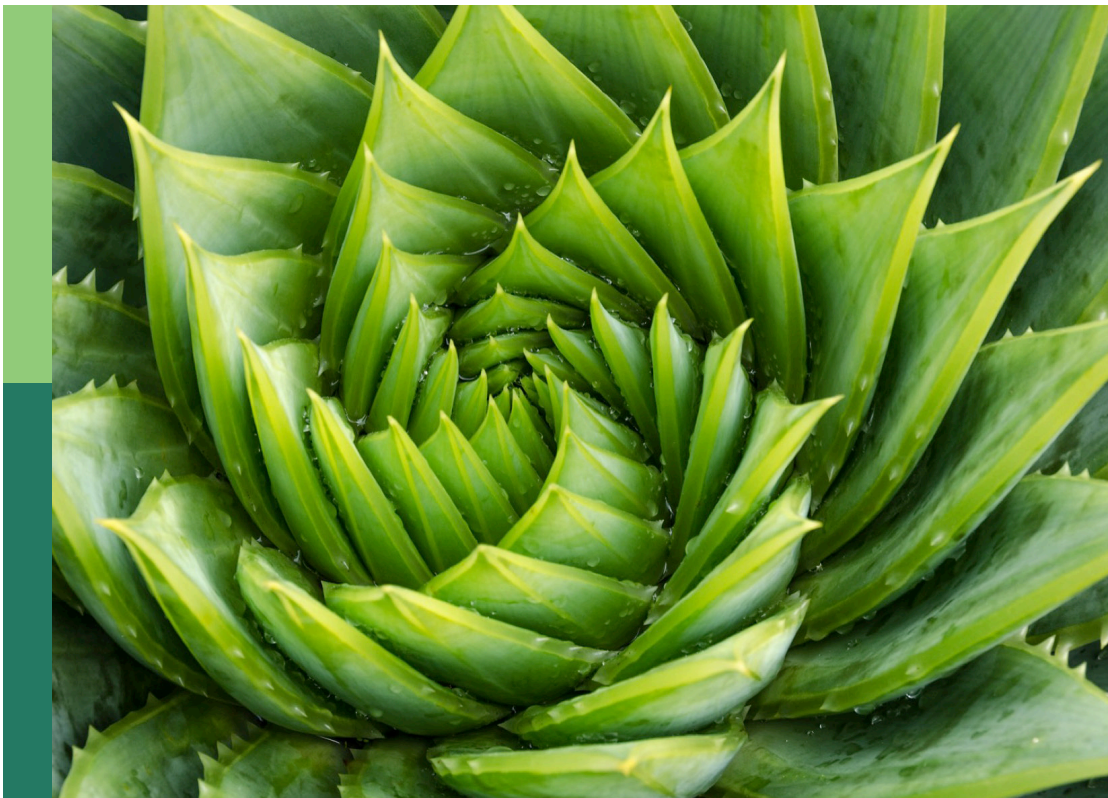
frontiers | Research Topics

Plant secondary metabolite biosynthesis

Edited by
Xiang Pu, Naoki Kitaoka, Shilin Chen and Carlos E. Rodríguez-López

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-5390-9
DOI 10.3389/978-2-8325-5390-9

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 metabolite biosynthesis

Topic editors

Xiang Pu — Sichuan Agricultural University, China

Naoki Kitaoka — Hokkaido University, Japan

Shilin Chen — Institute of Chinese Materia Medica - China Academy of Chinese Medical Sciences, China

Carlos E. Rodríguez-López — Monterrey Institute of Technology and Higher Education (ITESM), Mexico

Citation

Pu, X., Kitaoka, N., Chen, S., Rodríguez-López, C. E., eds. (2024). *Plant secondary metabolite biosynthesis*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-5390-9

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 21 August 2024
DOI 10.3389/fpls.2024.1477551

OPEN ACCESS

EDITED AND REVIEWED BY
Laigeng Li,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Xiang Pu
puxiang@sicau.edu.cn
Shilin Chen
slchen@cdutcm.edu.cn

RECEIVED 08 August 2024
ACCEPTED 12 August 2024
PUBLISHED 21 August 2024

CITATION
Pu X, Kitaoka N, Rodríguez-López CE and Chen S (2024) Editorial: Plant secondary metabolite biosynthesis.
Front. Plant Sci. 15:1477551.
doi: 10.3389/fpls.2024.1477551

Editorial: Plant secondary metabolite biosynthesis

Xiang Pu[1]*, Naoki Kitaoka[2], Carlos E. Rodríguez-López[3,4] and Shilin Chen[5]*

[1]College of Science, Sichuan Agricultural University, Ya'an, China, [2]Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan, [3]Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Monterrey, Mexico, [4]Tecnologico de Monterrey, The Institute for Obesity Research, Monterrey, Mexico, [5]Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu, China

KEYWORDS

plant, secondary metabolite, biosynthetic pathway, gene characterization, synthetic biotechnology

Editorial on the Research Topic
Plant secondary metabolite biosynthesis

Introduction

Plant secondary metabolites (PSM) are a diverse group of compounds that contribute to many important biological and ecological functions. They are synthesized by plants to interact with the biotic and abiotic environments, playing roles in plant defense, growth, and development (Erb and Kliebenstein, 2020). Additionally, PSM have widespread applications in human industries, including food additives, cosmetics, dyes, insecticides, and drugs. The biosynthesis of PSM is complex and dynamic, with more than one million PSM estimated to be present in terrestrial and aquatic plants (Afendi et al., 2012). Despite their diversity, plants produce limited quantities of PSM in a metabolic cost-saving way. This has limited their commercial production, and the overexploitation of source plants has raised concerns about their sustainability and highlighted the need for advanced research.

Recent advances in genomics (Siadjeu and Pucker, 2023), transcriptomics (Voelckel et al., 2017), metabolomics (Li et al., 2024), and other omics technologies have revolutionized our understanding of plant biology, enabling the discovery of new PSM and their biosynthetic pathways. Functional genomics approaches, such as genome-wide association studies (Luo et al., 2020), transcriptome analysis (Liu et al., 2021), and gene editing (Nasti and Voytas, 2021), have facilitated the identification and characterization of genes involved in the biosynthesis of PSM. Metabolic engineering and synthetic biology approaches have enabled the manipulation of plant secondary metabolism to improve the yield and quality of specific metabolites of interest or to produce them in heterologous cultures (Zhang et al., 2022). These advances have created new opportunities for the sustainable production and utilization of PSM.

This Research Topic on Plant Secondary Metabolite Biosynthesis illustrates a comprehensive and up-to-date view of the biosynthesis, regulation, and biotechnological production of PSM, and to promote interdisciplinary and cross-disciplinary research collaborations in this field for sustainable and efficient utilization of these valuable compounds.

Identifying and characterizing novel genes involved in the biosynthesis of PSM

Combining bioinformatics method, gene quantitative analysis, and evolutionary analysis, Wang et al. retrieve 103 *BAHD* genes from the Ginseng Genome Data resource. Phylogenetic analysis indicates that these *BAHD* genes are clustered in three clades. Most *PgBAHD*s contain cis-acting elements associated with abiotic stress response and plant hormone response. Among them, 34 *PgBAHD*s are clustered with genes that display malonyl transferase activity. Seven *PgBAHD*s (*PgBAHD4*, *PgBAHD45*, *PgBAHD65*, *PgBAHD74*, *PgBAHD90*, *PgBAHD97*, and *PgBAHD99*) are designated as the potential candidates involved in the biosynthesis of malonyl ginsenosides based on co-expression analysis. To fully elucidate the biosynthetic pathways of orobanchol derivatives in Fabaceae plants, Homma et al. probe the metabolomic pathways downstream of orobanchol in cowpea, barrel medic, and pea via substrate feeding experiment after different enzyme inhibitor (fluridone, uniconazole-P, and prohexadione) treatment. Subsequently, *DOX* and *BAHD* acyltransferases responsible for converting orobanchol to their derivatives are mined from the public dataset and screened using co-expression analysis. Enzymatic assays of heterologously expressed proteins reveal that the DOX in barrel medic converts orobanchol to medicaol, the DOX and BAHD acyltransferase in pea convert orobanchol to fabacol and acetylated fabacol, the favacol acetyltransferase and its homolog in cowpea acetylate orobanchol. These findings shed light on the molecular mechanisms underlying the structural diversity of strigolactones.

Hendrickson et al. characterize series of terpene synthases (TPS) from *Medicago truncatula*, a model legume. Thirty-nine *MtTPS* candidates are obtained from the Mt4.0v1 genome in Phytozome (tps://phytozome-next.jgi.doe.gov). They assess the *MtTPS* activity using a modular metabolic engineering system in *E. coli* and characterize the resultant metabolite using GC-MS. Analyses of the resultant metabolite reveal the production of an assortment of sesquiterpenes. This work also establishes the gene-to-metabolite relationships for sesquiterpene synthase in *M. truncatula*. To explore the glycosylation step for rutin biosynthesis in *Solanum melongena*, Gan et al. identify 195 UDP-glycosyltransferase (UGT) candidate genes from the *S. melongena* genome V4.0 (ttps://solgenomics.net). These *UGT* genes are classified into 17 subgroups and the members of Groups A, B, D, E, and L are associated with flavonol biosynthesis. A hierarchical clustering analysis of gene expression profiles reveals that the expression patterns of *SmUGT* genes in Clusters 7-10 are closely related to those of rutin biosynthetic pathway genes. Among them, *SmUGT89B2* is verified to encode the final enzyme in rutin biosynthesis via virus-induced gene silencing and transient overexpression assay in eggplant. Rosmarinic acid (RA) is one of the major bioactive components of *Prunella vulgaris*. Yan et al. identified 51 RA biosynthesis-related genes from the chromosome-level genome of *P. vulgaris*. Bioinformatic analyses and gene expression pattern indicate that 17 of them could be involved in RA biosynthesis. *In vitro* enzymatic assay reveals that PvRAS3 catalyzes the condensation of *p*-coumaroyl-CoA and caffeoyl-CoA with pHPL and DHPL, and PvRAS4 only catalyzes the condensation of *p*-coumaroyl-CoA with pHPL and DHPL. Generation of *pvras3* homozygous mutants through CRISPR/Cas9 technology and subsequent chemical compound analysis confirm that PvRAS3 is the main enzyme catalyzing the condensation of acyl donors and acyl acceptors during RA biosynthesis in *P. vulgaris*. This work supports the existence of at least four RA biosynthetic pathways, with the role of *PvRAS4* appears minor in this plant.

Developing and employing innovative biotechnological techniques to improve the yield of PSM

Zheng et al. assemble a high-quality genome of *Fagopyrum dibotrys*. Based on evolutionary relationship analysis, the authors speculate that *F. dibotrys* has originated in the high-altitude Tibetan Plateau region. The homologues of genes involved in the biosynthesis of flavonoids are annotated. This study could reveal the genetic background and facilitate the cultivation of high-yielding *F. dibotrys* varieties. Li et al. employ single-cell RNA sequencing to profile the transcriptomes of protoplasts derived from *Gynostemma pentaphyllum* shoot apexes and leaves. Examining gene expression patterns across various cell types reveal that genes related to gypenoside biosynthesis are predominantly expressed in mesophyll cells. They also explore the impact of transposable elements (TE) on *G. pentaphyllum* transcriptomic landscapes. This study not only provides new insights into the spatiotemporal organization of gypenoside biosynthesis and TE activity in shoot apexes and leaves, but also offers valuable cellular and genetic resources for improving the yield of gypenoside in *G. pentaphyllum*.

Mentha spicata (spearmint) possesses peltate glandular trichomes (PGT) where valuable essential oils are produced. Reddy et al. identify a novel non-canonical Aux/IAA gene, *MsIAA32*, from spearmint, which lacks the TIR1-binding domain and regulates the development of PGT. Using yeast two-hybrid studies, two canonical Aux/IAAs (*MsIAA3*, *MsIAA4*) and an ARF (*MsARF3*) are identified as the preferred binding partners of *MsIAA32*. The possible role of *MsIAA32* in non-glandular trichome formation is confirmed by ectopic expression in Arabidopsis. Undoubtedly, identifying new gene targets controlling PGT numbers in spearmint will provide new ways to increase volatile/scented PSM production. Dendrobine, a noteworthy alkaloid found in *Dendrobium nobile*, possesses valuable pharmaceutical potential. Sarsaiya et al. develop innovative approaches to enhance dendrobine production by utilizing endophytic fungi. Using test bottles (EGTB), the experimental group (12-month-old *D. nobile* seedling), co-cultured with *Trichoderma longibrachiatum* (MD33), displays a 2.6-fold denrobine increase (1804.23 ng/ml) compared to the control group (685.95 ng/ml). Co-culturing *D. nobile* seedlings with *T. longibrachiatum* MD33 in the temporary immersion bioreactor systems (EGTIBS) leads to a substantial 9.7-fold

dendrobine increase (4415.77 ng/ml) compared to the control (454.01 ng/ml) after 7 days. Scaling up the TIBS approach for commercial dendrobine production could provide an accessible platform for dendrobine production.

Phoebe zhennan is widely cultivated in China and the price of nanmu wood is expensive. Yang et al. explore the composition and content of essential oil and fragrance compounds in *P. zhennan* wood at different tree ages. The yield of essential oil from 30a wood is significantly greater than that from 10a and 80a wood. In total, 596 (LC) and 76 (GC) features are annotated using chromatography-coupled mass spectrometry in the essential oil and wood. This research determines that the main components of the wood fragrance are sesquiterpenoids. The types and relative contents of sesquiterpenoids from wood increase with tree age. These results suggest that choosing wood from trees of a suitable age will significantly improve the yield of essential oil. Qin et al. successfully improve the seed tocopherol concentration by altering chlorophyll metabolism. They find that RNAi suppression of *CHLSYN* coupled with homogentisate phytyltransferase (HPT) overexpression increase tocopherol concentration by more than two-fold in Arabidopsis seeds. Additional three-fold increase in seed tocopherol are observed by engineering homogentisate production via overexpression of bacterial *TyrA*, which encodes chorismate mutase/prephenate dehydrogenase. They further increase total tocochromanol concentration by overexpression of the gene for hydroxyphenylpyruvate dioxygenase. These biofortification approaches shift metabolism towards increased amounts of tocotrienols relative to tocopherols. This study provides a theoretical basis for genetic improvement of total tocopherol concentrations in green oilseeds.

Plant secondary metabolite biosynthesis, the road ahead

The number of PSM is still expanding due to the diversity of plant species and the rapid advances in analytical technology. Undoubtedly, PSM has served as a natural compound resource for human industries. But there is a contradiction between the insatiable human needs for PSM and the existing chemical-based production methods, such as plant source-dependent extraction and sophisticated chemical synthesis. Metabolic engineering and synthetic biology approaches have been verified as one of the best ways for PSM manufacture. These innovative biotechnological techniques depend on fully elucidating the biosynthetic pathway and biochemically charactering the gene elements responsible for PSM biosynthesis one by one. In recent years, characterization and heterologous reconstitution of the biosynthetic pathway for several valuable PSM including artemisinic acid, thebaine, colchicine, and baccatin III have been reported. These findings are intriguing and encouraging for researchers focusing on PSM biosynthesis. However, there is still a long way to go. A huge number of biosynthetic steps for PSM remain unknown. Thousands of gene elements responsible for PSM biosynthesis remain uncharacterized. Modern omics technologies, novel gene manipulation approaches, and artificial intelligence technology has expanded our knowledge on PSM biosynthesis and will facilitate the biomanufacture of valuable PSM.

Author contributions

XP: Writing – original draft. NK: Writing – review & editing. CR-L: Writing – review & editing. SC: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. XP received funding from the Department of Science and Technology of Sichuan Province, PR China (2021ZYD0059; 2022NSFC1551) and the National Natural Science Foundation of China (21708028). CR-L received funding from Tecnológico de Monterrey Challenge-Based Research Funding Program.

Acknowledgments

The authors wish to thank to all the contributors and the FPS editorial team for their efficient work. We would like to express our deep gratitude to all reviewers for their valuable participation in assuring the scientific quality of this 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

References

Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53, e1. doi: 10.1093/pcp/pcr165

Erb, M., and Kliebenstein, D. J. (2020). Plant secondary metabolites as defenses, regulators, and primary metabolites: the blurred functional trichotomy. *Plant Physiol.* 184, 39–52. doi: 10.1104/pp.20.00433

Li, Y. H., Hirai, M. Y., and Nakamura, Y. (2024). Plant metabolomics. *J. Exp. Bot.* 75, 1651–1653. doi: 10.1093/jxb/erae047

Liu, M. Y., Yang, L., Cai, M. M., Feng, C., Zhao, Z. M., Yang, D. P., et al. (2021). Transcriptome analysis reveals important candidate gene families related to oligosaccharides biosynthesis in *Morinda officinalis*. *Plant Physiol. Biochem.* 167, 1061–1071. doi: 10.1016/j.plaphy.2021.09.028

Luo, T. T., Xia, W., Gong, S. F., Mason, A. S., Li, Z. Y., Liu, R., et al. (2020). Identifying vitamin E biosynthesis genes in *Elaeis guineensis* by genome-wide association study. *J. Agric. Food Chem.* 68, 678–685. doi: 10.1021/acs.jafc.9b03832

Nasti, R. A., and Voytas, D. F. (2021). Attaining the promise of plant gene editing at scale. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2004846117. doi: 10.1073/pnas.2004846117

Siadjeu, C., and Pucker, B. (2023). Medicinal plant genomics. *BMC Genomics* 24, 429. doi: 10.1186/s12864-023-09542-8

Voelckel, C., Gruenheit, N., and Lockhart, P. (2017). Evolutionary transcriptomics and proteomics: Insight into plant adaptation. *Trends Plant Sci.* 22, 462–471. doi: 10.1016/j.tplants.2017.03.001

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrübbers, L., et al. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 609, 341–347. doi: 10.1038/s41586-022-05157-3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 December 2023
DOI 10.3389/fpls.2023.1301084

OPEN ACCESS

EDITED BY
Zhichao Wu,
National Institutes of Health (NIH),
United States

REVIEWED BY
Yang Chu,
China Academy of Chinese Medical Sciences, China
Qinggang Yin,
China Academy of Chinese Medical Sciences, China
Yuhan Fang,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Peng Di
di@jlau.edu.cn
Limin Yang
yanglimin@jlau.edu.cn

†These authors have contributed equally to this work

RECEIVED 24 September 2023
ACCEPTED 30 November 2023
PUBLISHED 14 December 2023

CITATION
Wang P, Yan Y, Yan M, Piao X, Wang Y, Lei X, Yang H, Zhang N, Li W, Di P and Yang L (2023) Identification and analysis of BAHD superfamily related to malonyl ginsenoside biosynthesis in *Panax ginseng*. *Front. Plant Sci.* 14:1301084. doi: 10.3389/fpls.2023.1301084

Identification and analysis of BAHD superfamily related to malonyl ginsenoside biosynthesis in *Panax ginseng*

Ping Wang†, Yan Yan†, Min Yan, Xiangmin Piao, Yingping Wang, Xiujuan Lei, He Yang, Nanqi Zhang, Wanying Li, Peng Di* and Limin Yang*

State Local Joint Engineering Research Center of Ginseng Breeding and Application, College of Chinese Medicinal Materials, Jilin Agricultural University, Changchun, China

Introduction: The BAHD (benzylalcohol O-acetyl transferase, anthocyanin O-hydroxycinnamoyl transferase, N-hydroxycinnamoyl anthranilate benzoyl transferase and deacetylvindoline 4-O-acetyltransferase), has various biological functions in plants, including catalyzing the biosynthesis of terpenes, phenolics and esters, participating in plant stress response, affecting cell stability, and regulating fruit quality.

Methods: Bioinformatics methods, real-time fluorescence quantitative PCR technology, and ultra-high-performance liquid chromatography combined with an Orbitrap mass spectrometer were used to explore the relationship between the BAHD gene family and malonyl ginsenosides in *Panax ginseng*.

Results: In this study, 103 BAHD genes were identified in *P. ginseng*, mainly distributed in three major clades. Most PgBAHDs contain cis-acting elements associated with abiotic stress response and plant hormone response. Among the 103 genes, 68 PgBAHDs are WGD (whole-genome duplication) genes. The significance of malonylation in biosynthesis has garnered considerable attention in the study of malonyltransferases. The phylogenetic tree results showed 34 PgBAHDs were clustered with genes that have malonyl characterization. Among them, seven PgBAHDs (*PgBAHD4*, *45*, *65*, *74*, *90*, *97*, and *99*) showed correlations > 0.9 with crucial enzyme genes involved in ginsenoside biosynthesis and > 0.8 with malonyl ginsenosides. These seven genes were considered potential candidates involved in the biosynthesis of malonyl ginsenosides.

Discussion: These results help elucidate the structure, evolution, and functions of the *P. ginseng* BAHD gene family, and establish the foundation for further research on the mechanism of BAHD genes in ginsenoside biosynthesis.

KEYWORDS

***Panax ginseng*, BAHD gene family, malonyltransferase, malonyl ginsenoside, biosynthesis**

1 Introduction

The BAHD acyltransferase family mainly uses coenzyme A thioesters as acyl donors and alcohols or amines as acceptors to catalyze acylation reactions to form various acylation products. (St-Pierre and Luca, 2000). A comprehensive comparative analysis of the amino acid sequences of the currently identified BAHD acyltransferase family members reveals that the family proteins' amino acid sequences all contain two conserved regions, HXXXD and DFGWG (Molina and Kosma, 2015). The HXXXD conserved region located in the active center of the enzyme is also distributed in other acyltransferase families, such as chloramphenicol acetyltransferase type I, II, III (chloramphenicol acetyltransferase, CAT) and choline/carnitine O-acyltransferase. The DFGWG conserved region is located at the C-terminal. In addition, acyltransferases related to anthocyanin/flavonoid biosynthesis also contain YFGNC conserved sequences (Suzuki et al., 2004; Unno et al., 2007; Yu et al., 2009).

Acylation modifications mediated by acyltransferases are prevalent in the structural modification of natural products. They are essential for enriching the structural diversity of plant secondary metabolites, enhancing the stability, lipid solubility, and improving the bioavailability of compounds (Matern et al., 1986; Ardhaoui et al., 2004; Mellou et al., 2006; Taguchi et al., 2010). The BAHD acyltransferase family is a class of proteins unique to plants for acylation modification of secondary metabolites. It is vital in the biosynthesis of a wide range of active acylated natural products (Yu et al., 2009). Researchers have explored the role of BAHD acyltransferase in the biosynthesis of saponin adjuvants from the soapbark tree. Identified one enzyme (encoded by Qs0206480) that generated a product consistent with the addition of an acetyl group, and two enzymes (encoded by Qs0023500 and Qs0213660) that likely corresponded to the addition of L-rhamnose and D-glucose, respectively. The study provides insights into the biosynthetic pathway of saponin adjuvants and highlights the role of BAHD acyltransferases in the modification of the heptasaccharide scaffold (Reed et al., 2023). There is research explored the role of BAHD acyltransferase in the biosynthesis of aescin and aesculin in horse chestnut. AcBAHD3 and AcBAHD6 were able to acetylate the hydroxyl group of aescin precursor, yielding a product called 22-O-acetylprotoaescigenin, and also found that these enzymes could use acetyl-CoA as a donor to catalyze the desacetylation of aescin, resulting in the formation of aesculin. This study provides evidence for the involvement of BAHD acyltransferase in the acylation and deacetylation of triterpenoid compounds in horse chestnut. Highlighted the role of BAHD enzymes in the diversification of triterpenoid compounds in plants (Sun et al., 2023). In the study conducted by GAME36, a BAHD-type acyltransferase, it was found that SGA-acetylation occurs in both cultivated and wild tomatoes. This process involves the conversion of α-tomatine to Esculeoside A, which is non-bitter and less toxic. The researchers successfully elucidated the biosynthesis pathway of core Esculeoside A in ripe tomatoes (Sonawane et al., 2023).

Panax ginseng is a perennial herbaceous plant with a long growth period. Modern pharmacological studies have shown that ginsenosides are the main medicinal components of ginseng. Malonyl ginsenosides are natural ginsenosides that contain a malonyl group attached to a glucose unit of the corresponding neutral ginsenosides (Wang et al., 2016). In ginseng, the proportion of malonyl ginsenosides m-Rb1, m-Rb2, m-Rc, and m-Rd in the total ginsenosides ranged from 35% to 60% (Liu et al., 2012). More than 20 malonyl ginsenosides have been identified by high-resolution mass spectrometry (Sun et al., 2012; Wan et al., 2015). The role of BAHD acyltransferase, a key enzyme in the acylation of secondary metabolites, in the biosynthesis of malonyl ginsenosides has not yet been reported, so the identification and analysis of the ginseng BAHD genes is of great significance.

In this study, a total of 103 BAHD family genes were identified, and the analysis of evolutionary relationships indicated that PgBAHDs could be divided into five evolutionary clades. The gene structure, chromosomal localization, inter-gene, and inter-species collinearity of PgBAHDs were further investigated. In addition, the expression profiles of PgBAHD genes in different tissues and the trend of response to various abiotic treatments were investigated. Co-expression analysis of ginsenoside biosynthesis pathway genes with PgBAHD genes and secondary metabolite malonyl ginsenoside with PgBAHD genes was performed. Eventually eight PgBAHDs were identified as genes that may be involved in malonyl ginsenoside biosynthesis. This study provides a reliable basis for further metabolic regulation of the ginsenoside biosynthetic pathway and for conducting corresponding synthetic biology studies and molecular breeding.

2 Materials and methods

2.1 BAHD sequence retrieval and identification

The candidate BAHD genes were initially acquired from the Ginseng Genome Data resource (Wang et al., 2022). The Hidden Markov Models (HMMs) for the conserved BAHD domain (Pfam: PF02458) were extracted from the Pfam database (http://pfam.xfam.org). The HMMER 3.2.1 software was employed to detect the BAHD genes obtained from the ginseng genome, with an E-value threshold set at 10^{-2}. To ensure the presence of the BAHD PF02458 domain, all candidate PgBAHDs were further validated using the SMART data resource (http://smart.embl.de/), NCBI-Conserved Domain Database (CDD), and PlantTFDB (Plant

Abbreviations: HMMs, Hidden Markov Models; CDD, Conserved Domain Database; MW, Molecular weight; pI, Isoelectric point; aa, amino acid; kDa, Kilodaltons; CDS, Coding sequences; MEME, Multiple expectation maximization for motif elicitation; TPM, Transcripts Per Million; HMM, Hidden Markov model; MEGA, Molecular evolutionary genetics analysis; MYA, Million years ago; AR, adventitiou roots; CT, callus; RG, fibrous roots; SA, Salicylic acid; IAA, Indole-3-acetic acid; GA3, Gibberellin; MeJA, Methyl Jasmonate; ABA, Abscisic acid; chr, Chromosome; Ka, Non-synonymous; K_S, Synonymous; WGD, Whole-genome duplication; TD, tandem duplication; PD, proximal duplication; SD, singleton duplication; DSD, dispersal duplication; RNA-Seq, RNA sequencing; qRT-PCR, Quantitative real-time PCR; H-ESI, High Energy Spark-Induced Breakdown Ionization.

Transcription Factor Database) (http://planttfdb.cbi.pku.edu.cn). Subsequently, their HXXXD and DFGWG motifs were examined.

Furthermore, the online Sequence Manipulation Suite was utilized to predict the theoretical pI and MW of PgBAHD proteins (http://www.detaibio.com/sms2/reference.html) (Stothard, 2000).

2.2 Phylogenetic analysis

Mafft (https://mafft.cbrc.jp/alignment/software/) was used with default parameters to perform multiple alignments of ginseng BAHD sequences and multiple alignments of BAHD, among other species. The ginseng BAHD phylogenetic tree was constructed using the maximum likelihood method IQ-TREE based on the JTTDCMut+F+R4 model (Nguyen et al., 2015), and the nodes were tested 1000 times using bootstrap analysis. Further annotation of the phylogenetic tree results was conducted using Evolview (https://evolgenius.info/).

2.3 Gene structure and cis-acting elements analysis

TBtools 1.053 was employed to demonstrate the gene structure (Chen et al., 2020). Conserved motifs of PgBAHDs were identified using MEME (Multiple expectation maximization for motif elicitation) native software (version 4.12.0) in Linux with a maximum of 10 mismatches and an optimal motif width of 6-100 amino acid residues.

The sequence of 2000 bp upstream of the start codon of PgBAHDs was obtained for promoter analysis. PlantCARE (http://bionformatics.psb.ugent.be/webtools/plantcare/html) was used to predict cis-acting elements in the promoter region, and PlantTFDB software (http://planttfdb.cbi.pku.edu.cn/) was used online to predict the distribution of promoter transcription factor binding sites (p-value ≤ $1e^{-6}$).

2.4 Chromosomal location, duplication, synteny and evolution analyses

The MCScanX program was utilized to conduct inter- and intra-species collinearity analysis of proteins, using an E value threshold of $1e^{-5}$. Furthermore, the Duplicate Gene Classifier script within the MCScanX program was employed to quantify different types of duplication. The results were visualized using Circos (Krzywinski et al., 2009; Wang et al., 2012).

Using KaK_S-Calculator-2.0, this study calculates replicated gene pairs' Ka and K_S. The analysis aims to assess the environmental selection pressure by examining the Ka/K_S ratio (Wang et al., 2010).

2.5 Meta-expression analysis

To analyze gene expression among different tissues and responses to various abiotic treatments, RNA-Seq datasets from 14 different tissues were obtained from NCBI (accession number PRJNA302556) (Wang et al., 2015). Additionally, 15 RNA-Seq datasets for abiotic treatment were retrieved (No.24-38 in the ginseng transcriptome data resource, http://ginsengdb.snu.ac.kr/transcriptome.php) from the Ginseng Genome Data Resource (http://ginsengdb.snu.ac.kr/). The clean reads were aligned to the ginseng genome using Hisat2 software. Assembly and calculation of expression values for each transcript were performed using Hisat2, StringTie, and Ballgown.

In a previous study, the RG, AR, and CT cDNA libraries were established. These nine cDNA libraries were subsequently sequenced on HiSeq 2500 (Illumina) with the PE125 strategy. The TPM was calculated using the same protocol as the other 16 RNA-Seq datasets. Finally, the heatmap was generated using the R package 'Heatmap'.

2.6 Quantitative real-time PCR analysis

Total RNA was extracted from the samples using an EasyPure Plant RNA Kit (TransGen Biotech), with the inclusion of RNase-free DNase I (TransGen Biotech) to eliminate DNA contamination. The concentration and quality of the RNA samples were assessed using a NanoPhotometer N50 (Implen, GER). Subsequently, the Perfect-Start Uni RT-qPCR Kit (TransGen Biotech) was employed to reverse transcribe RNA into cDNA, followed by two-step quantitative real-time PCR using a Roche Light Cycler 96 (SYBR-GREEN I; No Passive Reference Dye). The β-Actin gene was utilized as the internal control (Hou et al., 2014). Data analyses involved the use of the $2^{-\Delta\Delta CT}$ method for determining the relative expression of PgBAHD genes (Livak and Schmittgen, 2023). Primers for qRT-PCR were synthesized by Sangon Biotech (ShangHai, China), and their sequences are listed in Supplementary Table S7.

2.7 Metabolome samples and chemicals preparation

Each dried powdered sample, weighing 0.5 g and with a mesh size of < 40, was accurately weighed. The samples were then sonicated with 10 ml of 80% ethanol for 40 min (100 W, 40 KHz) and centrifugation at 10,000 rpm for 10 min. This process was repeated three times, and the resulting supernatants were combined and transferred into a 10 ml volumetric flask. The volumetric flask was adjusted to a final volume of 10 ml using 80% ethanol and thoroughly mixed. Before injection, the solution was filtered twice through a 0.45 μm organic membrane.

Reference standards of ginsenosides, purchased from Shanghai Yuanye Biotechnology Co., Ltd (Shanghai, China) with purities not less than 98.0%, were used. Approximately 5 mg of each standard was taken into a 5 ml volumetric flask, dissolved with methanol, and diluted to the scale to obtain a reserve solution of each standard with a 1 mg/ml concentration. An appropriate amount of each stock solution was dispensed, diluted with methanol, and adjusted to a final concentration of about 50 μg/ml to create the mixed standard solution.

2.8 UHPLC-orbitrap MS conditions

The sample components were separated using a Thermo Fisher Vanquish liquid chromatography system (Thermo Fisher Scientific, San Jose, CA, USA). The chromatographic column used was a Hypersil Gold Vanquish UHPLC column (100×2.1 mm, 1.9 μm; Thermo). In this experiment, the mobile phase consisted of acetonitrile (A) and 0.1% formic acid in water. The gradient elution program followed the following steps: 0-34.0 min, 15%-55% A; 34-35 min, 55%-98% A; 35-36 min, 98% A; 36-37 min, 98%-15% A; 37-40 min, 15% A. The column temperature was maintained at 35°C, the flow rate was set at 0.30 ml/min, and the injection volume for each sample was 1 μl. During the separation process, the column temperature remained at 35°C, and the flow rate used was 0.30 ml/min, with an injection volume of 1 μl for each sample.

MS spectrometric detection was performed on an Orbitrap Fusion mass spectrometer (FSN10450, Thermo Fisher, USA) equipped with an H-ESI source operating in negative ion modes (Neg Ion Spray Voltage) at 2700 V. Each sample was analyzed separately using an orbitrap full scan in the first stage (full scan, with a resolution of 60,000, RF Lens 50%). For quantitative characterization, MS/MS data were acquired using data-dependent ms2 scans (DDA, resolution 15000, HCD Collision Energy 40%). Ion source conditions were as follows: sheath gas flow, 40 Arb (Arbitrary units); auxiliary gas flow, 5 Arb; ion transfer tube temperature, 320°C; vaporizer temperature, 320°C. Full-scan MS data were collected from 150 to 1500 m/z. Mass data were recorded with Xcalibur 4.0 software.

2.9 Correlation coefficient analysis

The TPM (Transcripts Per Million) values of PgBAHDs and ginsenoside biosynthesis pathway genes from AR (adventitiou roots), CT (callus), and RG (fibrous roots) were used to calculate the Pearson's correlation between these types of genes using the R package 'Hmisc'. Similarly, the TPM values of PgBAHD genes and malonyl ginsenoside from AR, CT, and RG were used to calculate the Pearson's correlation between PgBAHD genes and metabolites using the R package 'Hmisc'.

3 Results

3.1 Identification and phylogenetic analysis of the PgBAHD gene families

To identify ginseng BAHD family genes, a search of the ginseng genome using a Hidden Markov Model (HMM) of the transferase family identified 103 genes. All 103 BAHD proteins contained the conserved structural domains HXXXD and DFGWG that characterize the BAHD family. The complete data of these genes, including gene ID, protein length, gene length, molecular weight (MW), and isoelectric point (pI), are shown in Supplementary Table 1. The lengths of these BAHD genes ranged from 741 bp (*PgBAHD3* and *80*) to 2682 bp (*PgBAHD84*), while protein lengths ranged from 246 to 893 amino acids (aa). The molecular weight (MW) varied from 27.27 kDa (*PgBAHD3*) to 98.77 kDa (*PgBAHD84*). The isoelectric point (pI) is an indicator used to determine the pH, and it varied among PgBAHD proteins from 5.22 (*PgBAHD25*) to 9.30 (*PgBAHD65*). Overall, 75 proteins were predicted to have low isoelectric points (pI < 7).

A phylogenetic tree of 103 ginseng BAHDs was constructed to categorize and explore the evolutionary relationships of PgBAHD genes (Figure 1). The analysis included PgBAHD genes and genetically and biochemically characterized BAHD acyltransferases (the information on foreign genes in the phylogenetic tree can be found in Supplementary Table 2) (D'Auria, 2006). The phylogenetic tree results showed that all genes were divided into five clades, of which 103 PgBAHD genes were clustered in three. Clade I contain functionally characterized members that are almost all related to the structural modification of phenolic glycosides, and most of these genes are involved in the acylation of anthocyanins. For example, Dv3MaT is an anthocyanin-like malonyltransferase, and in addition, NtMAT1 catalyzes the malonylation of phenolic glycosides and flavonoid glycosides in *Nicotiana tabacum* (Suzuki et al., 2002; Taguchi et al., 2005). Twenty-six of the 103 PgBAHD genes were in this clade. Clade III, focuses on acyltransferases related to the biosynthesis of volatile esters in mature fruits and tissues such as flowers and leaves, and 35 of the 103 PgBAHD genes are in this clade (Shalit et al., 2003). In this clade, there are 8 PgBAHD genes closely related to Ss5MaT2 (*PgBAHD14*, *30*, *46*, *61*, *62*, *73*, *74* and *88*). In clade V, benzyl alcohol/phenylethanol benzoyltransferase (BPBT), methanol acyltransferase (AMAT), and tigloyl transferase (HMT/HLT) were mainly related to the biosynthesis of volatile ester compounds (Wang and Luca, 2005; Okada et al., 2005). There were also genes associated with the biosynthesis of paclitaxel and the biosynthesis of ρ-coumaryl shikimate/quinate esters, and 42 PgBAHD genes were in this clade. In addition, clade II contains only Glossy2 from *Zea mays* and CER2 from *Arabidopsis thaliana* (Xia et al., 1997; Costaglioli et al., 2005). Clade IV has only one gene derived from Hordeum vulgare, agmatine coumaroyltransferase (ACT) (Burhenne et al., 2003). This study focuses on 35 PgBAHD genes, including 26 PgBAHDs in cladeI and eight PgBAHDs closely related to Ss5MaT2.

3.2 Gene structure and cis-regulatory element analysis of PgBAHDs

All PgBAHDs contained motif 2and motif 4, annotated as the classic BAHD DNA-binding domain HXXXD and DFGWG, respectively. The 103 PgBAHDs contain three to 10 conserved motifs (Figures 2A; Supplementary Figure 1). Of these, 98 PgBAHDs incorporated six to eight motifs. In addition, the *PgBAHD80* has three motifs, the *PgBAHD3*, *32*, and *55* have five motifs, and the *PgBAHD93* contains 15 motifs. The PgBAHDs belonging to the same clade have a similar motif composition. For example, in clade I, among the 16 PgBAHD genes, all possess motif 7, and none contain motif 6. Additionally, within this clade, only

10.3389/fpls.2023.1301084

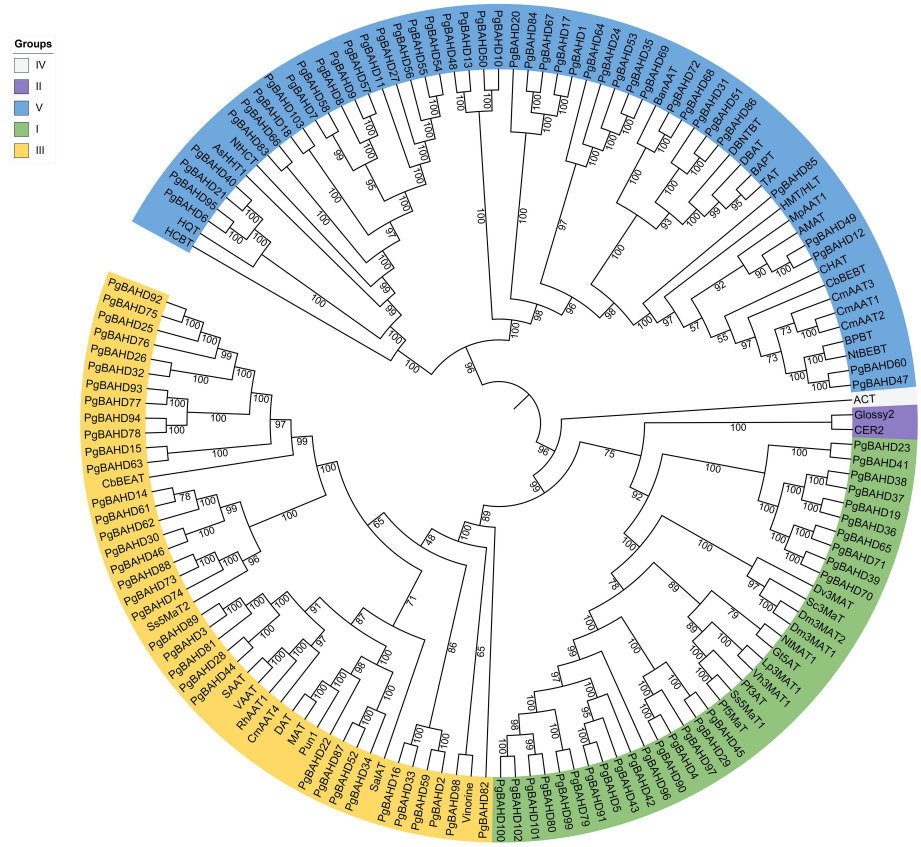


FIGURE 1
Phylogenetic analysis of PgBAHD genes with genetically or biochemically characterized BAHD acyltransferases.

seven genes have motif 6. Motif 8 is unique to clade V. Motif 9 exclusively occurs in clade III, but *PgBAHD44*, *77*, *82* and *98* in this clade do not contain motif 9.

To further investigate the gene structure of PgBAHDs, an exon/intron structure diagram was constructed. Of the 103 PgBAHDs, the majority had one to two exons, and 80 PgBAHDs (77.7%) had this structure. A total of 19 PgBAHDs (18.5%) had three to four exons. In addition, *PgBAHD15* contained eight exons, *PgBAHD32* and *102* contained nine exons, and *PgBAHD84* contained 15 exons (Figure 2B). The PgBAHD genes had relatively similar exon numbers.

Several key response elements were identified in the promoter region of PgBAHDs (Supplementary Table 3). The most salient response factors included those related to stress (37%), hormones (27%), light (29%), and growth (7%) (Figure 3A), but no corresponding promoters were identified in *PgBAHD37*. Several hormone regulatory sites were identified in the study, such as SA (Salicylic acid), IAA (Indole-3-acetic acid), GA3 (Gibberellin), MeJA (Methyl Jasmonate), and ABA (Abscisic acid) (Figure 3B). It was found that the PgBAHD genes might be more guided by ABA and MeJA, depending on the distribution of cis-acting elements in its promoter region. In addition, regulatory elements were identified for various conditions such as dehydration, low temperature, salt stress-responsive, anoxic specific inducibility, wounding responsiveness, and anaerobic induction (Figure 3C).

3.3 Duplication, synteny and evolution analyses of PgBAHD gene members

In Figure 4A, PgBAHD genes are irregularly distributed on all 24 ginseng chromosomes. Subgenome A includes chr2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 18, 23, and subgenome B includes chr1, 10, 12, 14, 15, 16, 17, 19, 20, 21, 22, 24. Among them, chr 5 had the highest density of PgBAHD genes with 10 PgBAHD genes, followed by chr 12 with nine PgBAHD genes. In addition, the vast majority of chromosomes possessed genes ranging from two to six in number. Most genes are distributed primarily at the ends of chromosomes. In addition, to investigate the collinearity of ginseng genes with members of the same family, the genomic collinearity of BAHD was analyzed in *P. ginseng* and *Panax quinquefolium*, and *P. ginseng* and *Panax notoginseng* (Figure 4B). Seventy-six PgBAHD genes showed collinearity with *P. quinquefolium* BAHD genes, and 63 PgBAHD genes showed collinearity with *P. notoginseng* BAHD genes. The results indicated that the PgBAHD genes was more closely related to the *P. quinquefolium* BAHD genes.

The 103 PgBAHD in ginseng showed 29 pairs of genes with collinearity. By calculating the non-synonymous substitution rate (Ka) and synonymous substitution rate (K_S) of the two protein-coding genes (Supplementary Table 4), the K_S value reflects the rate of substitution of background nucleobase during the evolutionary process, and the Ka/K_S value determines the selective pressure of

10.3389/fpls.2023.1301084

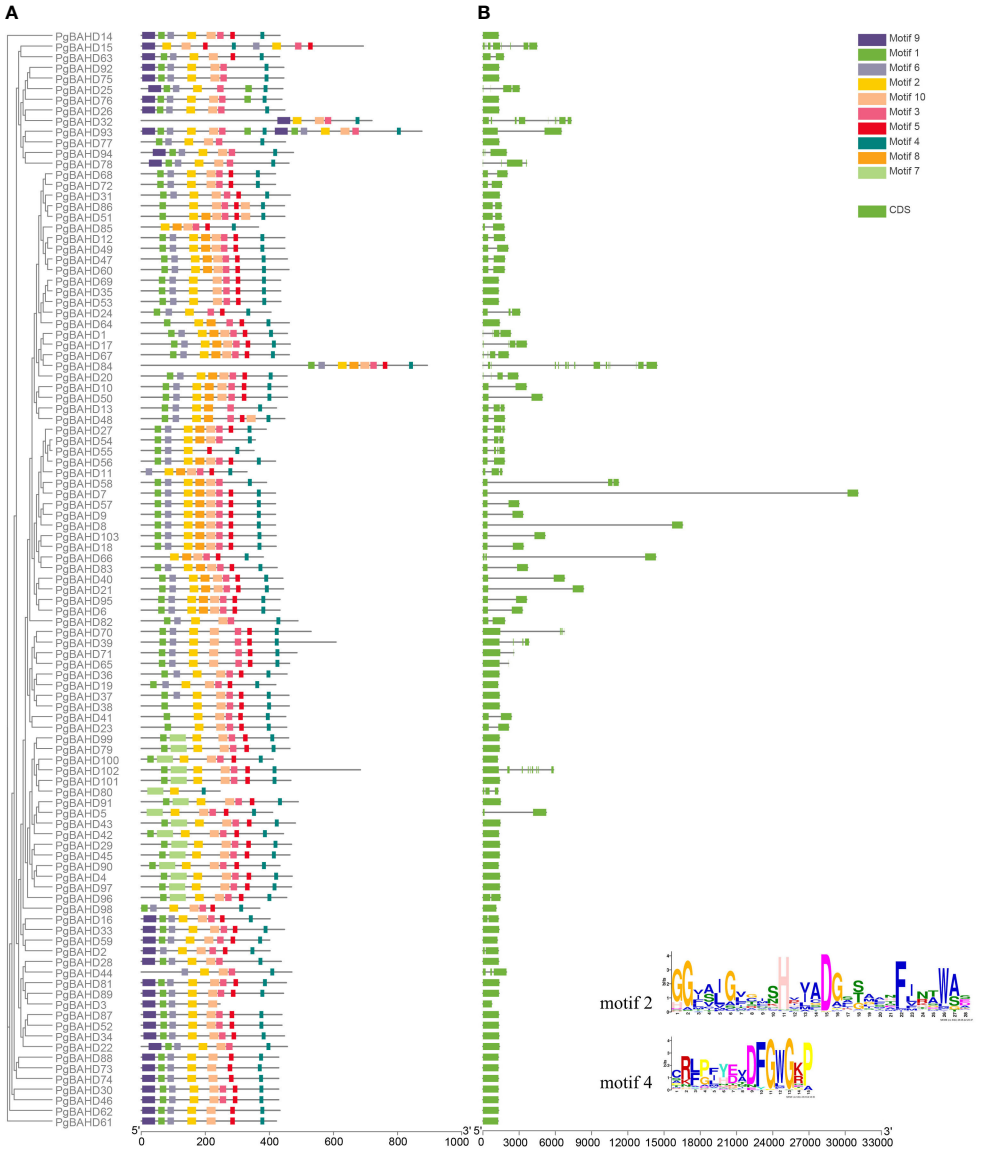


FIGURE 2
Conserved motif and gene structure analysis of PgBAHDs. **(A)** Motif distribution of PgBAHDs. **(B)** Gene structure of 103 PgBAHDs. The Motif composition of PgBAHD was analyzed by the MEME tool. The detailed information of the ten motifs is in Supplementary Figure 1.

these genes during the genetic evolutionary process. The K_S values of the PgBAHD gene families ranged from 0.0273 to 0.8226. The Ka/K_S values ranged from 0.0692 to 0.8231, indicating that the PgBAHD gene families evolved under purify selection.

Different patterns of gene duplication collectively contribute to the evolution of gene families and are responsible for their functional expansion and diversification. These include whole-genome duplication (WGD) or segmental duplication, tandem duplication (TD), proximal duplication (PD), singleton duplication (SD), and dispersal duplication (DSD). Duplicated BAHD family gene pairs in the ginseng genome were analyzed, and all BAHD gene family members were assigned to WGD, PD, TD, or DSD. Sixty-eight (66.0%) of the 103 ginseng BAHD genes were WGD, 21 (20.4%) were TD, eight were PD, and the remaining six PgBAHDs were DSD. WGD and TD are the duplication patterns that have mainly influenced the evolution of the BAHD superfamily in ginseng.

3.4 Expression profiles of PgBAHD genes in different tissues

TPM data from existing transcriptome data (Wang et al., 2015) were used to calculate the expression profiling of the PgBAHD genes in the main root cortex, main root epiderm, leaf blade, leaflet pedicel, fruit peduncle, stem, leaf peduncle, rhizome, leg root, fiber root, arm root, fruit pedicel, fruit flesh, and seed in different tissues.

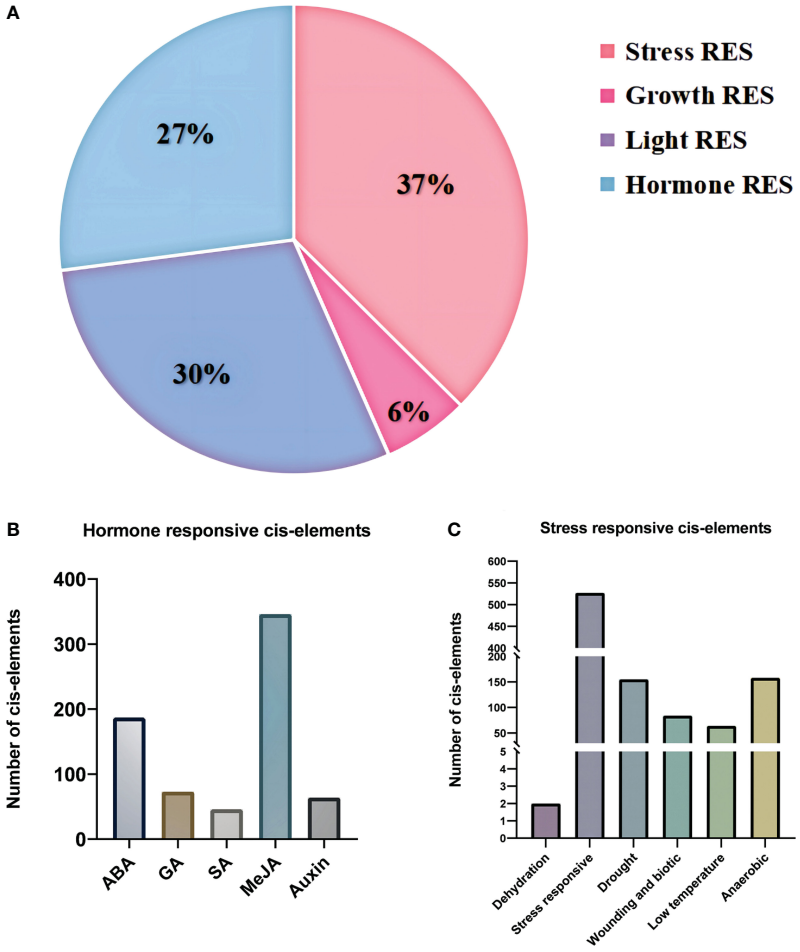


FIGURE 3
Distribution of cis-regulatory elements into promoter regions of PgBAHD genes. **(A)** Classification of identified regulatory elements based on function and their response to hormone, light, stress, and growth. **(B)** Distribution of different types of hormone-related cis-regulatory elements. **(C)** Cis-regulatory elements are related to various types of stresses.

A heat map of BAHD gene expression was produced (Figure 5A; Supplementary Table 5). The results showed that 24 of the 103 PgBAHD genes were not expressed in any tissue (≈ 23.3%; TPM < 1). A total of 56 PgBAHD genes were expressed in at least one tissue (≈ 54.4%; TPM ≥ 1). A total of 23 PgBAHD genes were expressed in all tissues (≈ 22.3%; TPM ≥ 1). The expression patterns of PgBAHDs are low-level, tissue-distinct, and constitutive (Cheng et al., 2019; Liu et al., 2020a). Twelve PgBAHD genes were highly expressed in tissues of the above-ground portion (average TPM > 25), *PgBAHD4*, *9*, *39*, *57*, *79*, *95* and *99* were highly expressed in fruit flesh (TPM > 50). *PgBAHD60* was highly expressed in the leaf peduncle (TPM = 91.5). Fruit pedicel had *PgBAHD4*, *79* and *95* expressions with TPM values greater than 110. In addition, the genes that were highly expressed in the stem were *PgBAHD65* (TPM = 72.2), and the genes that were highly expressed in the rhizome were *PgBAHD39*, *65* and *103* (TPM > 65). While thirteen PgBAHD genes had high expression (TPM > 20) in five tissues in the underground, among them *PgBAHD103* was expressed in the leg root, fiber root, and arm root with TPM > 98. Similarly, *PgBAHD6* was highly expressed in these three tissues (TPM > 50). *PgBAHD18* in the arm root also had high expression (TPM = 91.5). *PgBAHD57* was expressed in the main root cortex with TPM = 59.3. *PgBAHD6* was low expressed in the main root cortex. PgBAHD family genes expression was mainly in fruit pedicel, fruit flesh, arm root, and fiber root.

Among the three different morphology samples (fibrous roots, RG; adventitiou roots, AR; and callus, CT) of ginseng samples (Figure 5B; Supplementary Table 6), A total of 27 of 103 PgBAHD genes were expressed in at least one sample (TPM ≥ 1). Forty-four PgBAHD genes were expressed in all samples (≈ 33.9%; TPM ≥ 1). There were 65 PgBAHD genes expressed in AR, 61 in RG, and 50 in CT (TPM ≥ 1). In addition, only 32 genes were not expressed (TPM < 1) in all three samples. The highest expression of *PgBAHD4* was found in AR (TPM = 150.9), and this gene also had an enriched expression in RG (TPM = 131.7). The gene with the highest expression in CT was *PgBAHD65*, with an TPM value of 324.1. However, the expression level of this gene is relatively low in the other two tissues. Overall, most genes showed high expression levels in AR and CT samples. Among the 35 key PgBAHD genes mentioned in result 3.1, a total of 13 genes (*PgBAHD4*, *19*, *36*, *39*, *45*, *65*, *71*, *74*, *79*, *90*, *97*, *99* and *102*) in this section had a TPM > 20 in at least one tissue, and a TPM > 5 in at least one type of ginseng

10.3389/fpls.2023.1301084

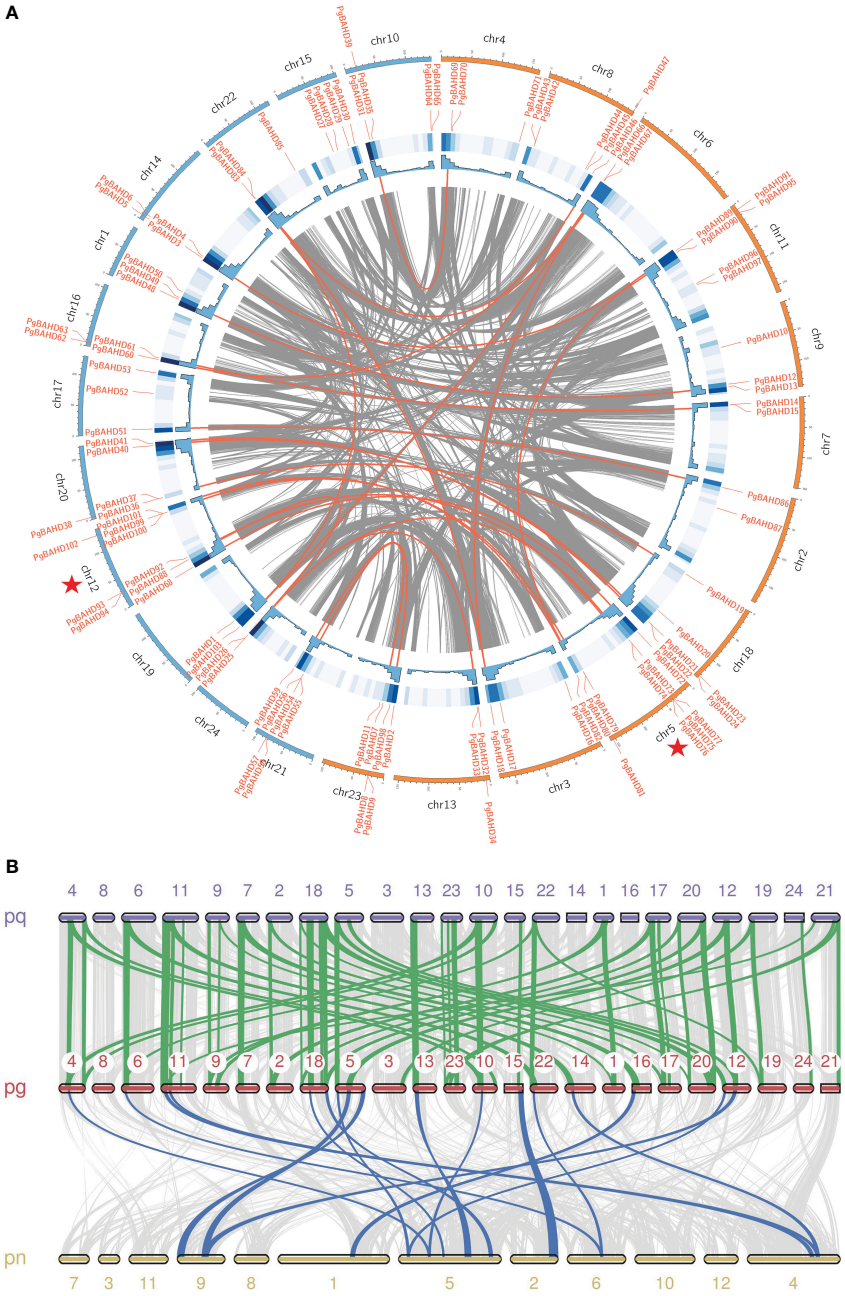


FIGURE 4
(A) Chromosomal locations and collinearity analysis of the PgBAHD gene family. The red lines indicate probably duplicated PgBAHD gene pairs.
(B) Collinearity relationship of BAHD genes among *P. ginseng*, *P. quinquefolium* and *P. notoginseng*.

sample. Further qRT PCR analysis was performed on the candidate PgBAHD genes. The results showed that the expression patterns of seven candidate genes (*PgBAHD4*, *45*, *65*, *74*, *90*, *97*, and *99*) were basically consistent with RNA seq data (Supplementary Figure 2).

3.5 Expression analysis of PgBAHD genes under different abiotic stresses

Different abiotic treatments of ginseng, such as cold, heat, drought, and salt, were applied in a previous study to investigate the response of PgBAHD genes to abiotic stresses. These treatments offer valuable insights for further studies in this area (Supplementary Figure 33; Supplementary Table 8). The results showed that the expression of seven PgBAHDs (TPM = 0; Fold change = 0) remained unchanged under four different stress conditions, and four of these genes (*PgBAHD13*, *22*, *48*, and *77*) did not have any expression in the 14 ginseng tissues and in the different morphology samples described in the previous two section as well. In the previous section, we focused on 13 PgBAHD genes that were significantly upregulated under three weeks of high temperature and drought stress. Among them, *PgBAHD39*, *45*, *65*,

10.3389/fpls.2023.1301084

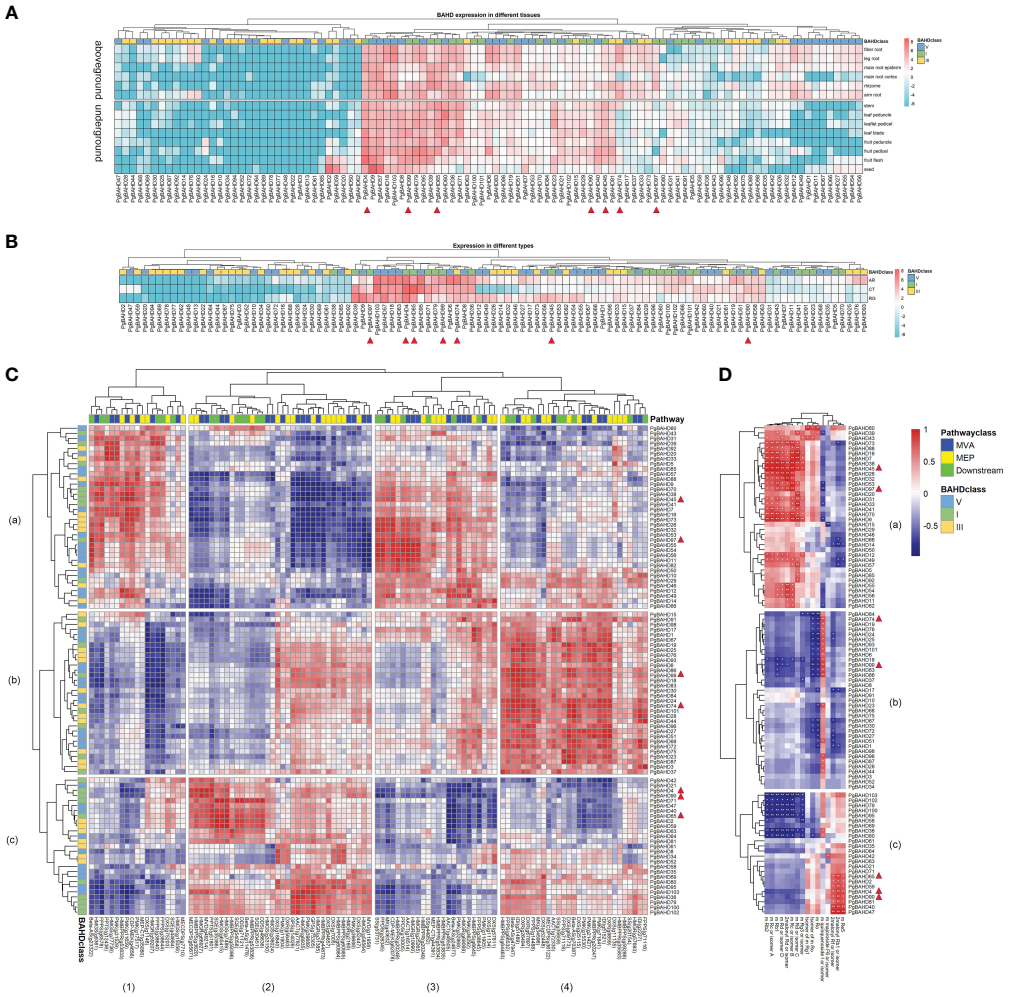


FIGURE 5
(A) Expression heat map of PgBAHD genes in different ginseng tissues. 14 samples were used for expression analysis, including main root cortex, main root epiderm, leaf blade, leaflet pedicel, fruit peduncle, stem, leaf peduncle, rhizome, leg root, fiber root, arm root, fruit pedicel, fruit flesh, and seed. Expression values from RNA-seq data were log2-transformed and are displayed as filled blocks in blue to red. **(B)** Heatmap of PgBAHD genes in different morphology samples of AR, CT and RG based on transcriptome data. Three biological replicates were set up in this samples. Note: Blue: Low expression level; Red: High expression level; The gene expression values present as log2-transformed normalized TPM values. **(C)** Coexpression between PgBAHD genes and ginsenoside biosynthesis pathway genes. Red indicates high correlation, and blue indicates low correlation. **(D)** Correlation between PgBAHD genes and malonyl ginsenosides. Red indicates high correlation, and blue indicates low correlation (*$p < 0.05$, **$p < 0.01$).

and *71* were also relatively upregulated under low temperature and salt stress. However, one week of high temperature stress seemed to have no effect on these genes.

About 43 (≈ 41.8%) genes were up-regulated in expression under cold treatment compared to the control, of which 14 genes showed more significant changes (fold change > 2), and six genes showed more than five-fold changes. In addition, an 11.4-fold change was observed in the *PgBAHD43* gene, while *PgBAHD35* and *69* only showed relatively high expression (fold change > 6) under cold treatment. About 36 (≈ 35.0%) PgBAHDs had up-regulated gene changes under drought treatment compared to control, of which about 19 genes had more than two-fold changes and ten genes showed about five-fold changes after drought treatment. There were also five genes with more than ten-fold change in expression, including *PgBAHD20* with 85.4-fold change. Salt treatment resulted in the up-regulation of the expression of 33 (≈ 32.0%) PgBAHD genes compared to the control, of which eight PgBAHDs were up-regulated (fold change >2), four PgBAHDs were significantly up-regulated (fold change >5), and in addition, the expression of the gene *PgBAHD39* was very significantly up-regulated (fold change >15). While the other PgBAHDs were slightly changed. The findings indicate that the expression patterns of PgBAHDs significantly changed under drought treatments, while they only showed slight changes in response to cold and salt treatments.

PgBAHDs exhibited varying response patterns to different treatments. Among the heat treatments, eight PgBAHDs (fold change >2) displayed significant changes between the control and one-week heat treatments. Notably, *PgBAHD43* exhibited a remarkable change of more than 23-fold. In the three-week heat-treated group, about 27 genes showed significant up-regulation of expression (fold change >2) compared to the control group, of

which 19 PgBAHDs had more than four-fold changes, with *PgBAHD39*, *43*, *65*, and *79* changes most especially above 20-fold. A total of 20 PgBAHDs were found to be changed at both one and three weeks of treatment, and 32 genes with a significantly higher amount of gene change (fold change > 2) were found in the three-week heat treatment compared with the one-week heat treatment. The expression trend analysis indicates that several PgBAHDs might play a role in the response to heat treatment after three weeks. Expression profiles also confirmed that PgBAHDs have a variety of functional and physicochemical properties.

3.6 Co-expression analysis of PgBAHD genes with ginsenoside pathway genes and malonyl ginsenosides

To investigate the relationship between PgBAHD genes and ginsenoside biosynthesis, co-expression analysis of PgBAHDs and ginsenoside biosynthesis pathway genes was performed. This pathway includes the upstream MEP and MVA pathways and the downstream pathway (Figure 5C; Supplementary Table 9). This section focuses on the 13 key PgBAHD genes mentioned earlier. Among them, a total of 12 genes have a correlation greater than 0.9 with at least one pathway gene. Among these, *PgBAHD39*, *45*, and *97* were predominantly localized within modules a1 and a3, *PgBAHD19*, *74*, and *99* exhibit primary concentration within module b4, whereas *PgBAHD4*, *36*, *65*, *79*, and *90* were primarily situated in module c2. *PgBAHD45* was mainly highly correlated with CDP-MEK (*g4656*) and PMK (*g11659*), while *PgBAHD97* was highly correlated with Beta-AS (*g23322*) and CDP-MEK (*g62078*). In c2 module, *PgBAHD4* and *90* were highly correlated with MVD (*g49214*) and HMGR(*g66927*), and *PgBAHD65* with HMGS (*g65415*), HMGS(*g7489*), SQE(*g35386*) and PPTS(*g51801*) were highly correlation. *PgBAHD74* and *99* were highly correlated with Beta-AS (*g47947*), DXR (*g26494*), and IDI (*g51489*). In addition, *PgBAHD74* is also highly correlated with Beta-AS (*g70387*), DXS (*g69173*), FPPS (*g14488*), and DXR (*g9859*).

To research the metabolites associated with malonyl ginsenoside biosynthesis, this study used ultra-high-performance liquid chromatography combined with an Orbitrap mass spectrometer (UHPLC-Orbitrap MS) for non-targeted detection of ginsenosides in ginseng tissues of different types (fibrous roots, RG; adventitious root, AR; and callus, CT). A total of 16 metabolites (malonyl ginsenosides) were identified in nine samples (Supplementary Table 10). To further validate the relationship between PgBAHD genes and malonyl ginsenoside biosynthesis, coexpression analysis was performed (Figure 5D; Supplementary Table 11). Selected *PgBAHD4*, *45*, *65*, *74*, *90*, *97*, and *99* as candidate genes, which have a correlation greater than 0.8 with at least one malonyl ginsenoside. The two genes *PgBAHD45* and *97* screened in the a1 and a3 modules mentioned above were mainly distributed in the a module in the analysis of their correlation with saponins (Figure 5D). They were highly correlated with m-Rc or isomer-B, m-Rb, and 2malonyl-Rd or isomer. Among them, 45 was also highly correlated with m-Rd or isomer-C, m-Rc or isomer-A, m-Rb2, and m-Rd or isomer-D. The *PgBAHD74* and *99* screened by the b4 module are highly correlated only with m-quinquenode I or isolmer, and they are distributed in module b in Figure 5D. The genes in module c in Figure 5D correspond to module c2 in 5C, where *PgBAHD4*, *65*, and *90* are highly correlated with 2malonyl-Rb1 or isomer, and *PgBAHD4* is also highly correlated with m-Re5. The results suggest that PgBAHDs are related to the metabolism of malonyl ginsenosides, and these PgBAHDs may promote the biosynthesis of malonyl ginsenosides in ginseng.

4 Discussion

BAHD acyltransferase plays a widespread role in acylation modification (Kusano et al., 2019; Luo et al., 2007; Grienenberger et al., 2009; Li et al., 2018). It is important for the biosynthesis of various active acylated natural products (Liu et al., 2017). Acylation products include lignin monomers, anthocyanins, terpenes, and esters. It also involves plant growth and development, environmental stress response, and fruit ripening (D'Auria, 2006; Zheng et al., 2009). Therefore, there is a need to analyze the possible roles of BAHD gene families in ginseng systematically. In this study, 103 BAHD genes were identified from ginseng. The number of genes in gene families also varied by species (Abdullah et al., 2021). In previous studies, 52 BAHD genes were identified in *A. thaliana* with a genome size of 0.125 Gb, *Prunus avium* (125; 0.125 Gb), *Rubus mesogaeus* (69, 0.24 Gb), *Brachypodium distachyon* (15; 0.26 Gb), *Musa acuminata* (46; 0.43 Gb), *P. ginseng* (103; 2.98 Gb), Hordeum vulgare (116; 5.1 Gb). Therefore, The number of BAHD genes did not seem to be correlated with genome size, which is in line with the results of the previous studies (Xu et al., 2021).

The BAHD family genes of ginseng were clustered into five evolutionary clades (I, II, III, IV, and V), and the results were consistent with the previously reported evolutionary clades of the plant BAHD acyltransferase family (D'Auria, 2006). Based on the results of sequence alignment of the encoded proteins, *D'Auria J C* performed a phylogenetic analysis of the relationships of 46 BAHD acyltransferase genes that have been functionally characterized (D'Auria, 2006). The clades were distinguished from each other based on differences in substrate and enzyme activities. This distinction revealed the evolutionary progression of BAHD family members' functions and aided in predicting the activities of enzymes with unknown functions. NtMAT1 catalyzes the malonylation of phenolic and flavonoid glycosides in *N. tabacum* and belongs to the clade I. In addition, based on a conserved sequence, YFGNC (motif 3), is shared by all the clade members. The motif is an acyltransferase associated with the biosynthesis of anthocyanins/flavonoid compounds (Unno et al., 2007; Yu et al., 2008). Successfully cloned a new enzyme Dv3MaT from *Ahlia variabilis* flowers, which is a cDNA coding 3-glucoside-specific malonyltransferase for anthocyanins (Suzuki et al., 2002). The hydroxycinnamoylation at positions three and five of the glycan chain of anthocyanins deepens the color, while malonylation increases stability (D'Auria et al., 2007; Luo et al., 2007). *Nicotiana benthamiana* BAHD family malonyltransferase NbMaT1 displays significant substrate tolerance to a wide range of natural products with different glycosyl substitutions at different

positions in the flavonoid, coumarin, and phenylethylchromone skeletons (Liu et al., 2017). There are 26 ginseng BAHD genes in clade I, so it is assumed that these PgBAHD genes are closely related to the malonylation of ginsenosides. Malonyltransferase Ss5MaT2, although also associated with acyl modification, is classified in clade III because it does not contain conserved sequences common to clade I. Therefore, attention will also be paid to *PgBAHD14*, *30*, *46*, *61*, *62*, *73*, *74* and *88*. Based on sequence structure, physicochemical properties, function, and distribution on chromosomes, the identified PgBAHDs showed a high degree of diversity, which is consistent with previous reports on BAHD gene families (Moglia et al., 2016; Zhang et al., 2019; Ahmad et al., 2020; Liu et al., 2020b).

Gene sequences and molecular weights differed significantly, but the characteristic structural domains and constituent motifs were relatively conserved. The exon/intron composition analysis showed that most 103 PgBAHDs had one to four exons. The exon numbers of PgBAHD genes within the same group were relatively similar, which is a phenomenon that most plants have(D'Auria, 2006; Tuominen et al., 2011; Liu et al., 2020b). Conservative motif and gene structure analyses showed that the identification and grouping of PgBAHDs was reliable. The presence of cis-acting elements related to stress responses in the promoter region of the PgBAHD gene indicates that this gene family may be activated by transcription factors associated with stress stimuli, as observed in previous studies (Ahmadizadeh and Heidari, 2014; Heidari et al., 2019). At the promoter site, a high frequency of the cis-acting elements related to responsive hormones such as ABA and MeJA were detected, indicating that stress-related hormonal signals primarily induce the PgBAHD gene.

Gene duplication patterns can be classified into five distinct types (Maher et al., 2006; Qiao et al., 2015). Each gene duplication pattern contributes differently to the expansion of gene families (Freeling, 2009). WGD, TD, and DSD are considered significant features of eukaryotic genome evolution, primarily propelling the emergence of novel functionalities within genomes and genetically evolved systems (Friedman and Hughes, 2001; Moore and Purugganan, 2003). Studies have estimated that WGD account for approximately 90% of the genetic expansion observed in the Arabidopsis lineage (Maere et al., 2005). The Hydroxycinnamoyltransferase and SWEET gene families have primarily expanded through WGD and DSD (Li et al., 2017; Ma et al., 2017).TD is the primary driving force behind the expansion of the AP2/ERF gene family (Du et al., 2013; Guo et al., 2014). Research results indicate that WGD and TD are the primary expansion modes for the ginseng BAHD gene family. Ginseng possibly experienced two WGD events between 2.2 million and 28 million years ago (Kim et al., 2018; Liu et al., 2022). TD and SD contribute to the domestication, survival, and resistance to both biotic and abiotic stresses in plants. These duplications lead to the creation of structural and functional diversity within genes (Schilling et al., 2020; Zan et al., 2020; Liu et al., 2021). The Ka, K_S results show that the K_S values of the PgBAHD gene pairs are all less than one. Among them, 17 gene pairs have relatively low K_S values (< 0.1), suggesting that these genes have undergone fewer mutations in a short period and may possess stable functions during the evolutionary process. On the other hand, 12 gene pairs have relatively high K_S values (> 0.1), indicating that these genes have undergone significant evolution over a more extended period, potentially facilitating functional evolution and family expansion. The Ka/K_S ratio is less than one, suggesting that purifying selection pressure has been acting on the ginseng BAHD gene family, leading to relatively stable expression. This implies that the gene family plays an important role in the growth and development of ginseng.

The molecular weight of BAHD proteins ranges from 48-55 kDa. They utilize CoA thioesters as acyl donors to transfer acetyl, malonyl, tigloyl, benzoyl, and hydroxycinnamoyl moieties, thereby regulating the structure and content of compounds in secondary metabolic pathways, ultimately influencing and modifying their properties (Kuate et al., 2008). Malonyl transferase is a class of acyltransferases that transfer malonyl molecules to sugar moieties, forming a wide range of biologically active natural acylated glycosidic constituents. Some malonyl ginsenosides have been isolated in ginseng in studies. For example, reports indicate the presence of malonyl ginsenosides Rb1, Rb2, Rc, and Rd in both *P. ginseng* and *P. quinquefolius* (Nakai and Yamamoto, 1984). Malonyl notoginsenoside R4 and malonyl-ginsenoside Ra3 have been isolated from the fresh roots of *P. ginseng*, respectively (Ruan et al., 2010). The malonyl ginsenosides isolated above are all derived from protopanaxadiol (PPT) type ginsenosides (Wan et al., 2015). Wang et al. isolated the PPT-type M-Re (malonyl-ginsenoside Re) from the flower buds of *P. ginseng* for the first time (Wang et al., 2016). The malonylation of plants primarily serves several purposes: stabilizing unstable structures, increasing the solubility of target compounds in water, and facilitating the transfer of target compounds into the vacuoles (Suzuki et al., 2002; Luo et al., 2007; Zhao et al., 2011; Matern et al., 1983). During ginsenoside biosynthesis, malonylation may also be associated with the transfer of ginsenosides into the vacuoles, thus affecting the accumulation of total ginsenosides in ginseng. Therefore, the specific functions and modes of action of different PgBAHDs in ginseng need to be further investigated.

5 Conclusion

This study subjected the BAHD gene family in ginseng to identification, phylogenetic construction, gene structure analysis, chromosomal localization, expression pattern analysis, and co-expression analysis. The results showed a high correlation between PgBAHDs and the critical enzyme genes of the ginsenoside biosynthesis pathway and with malonyl ginsenosides. This study provides a reliable basis for further metabolic regulation of the ginsenoside biosynthesis pathway, synthetic biology research, and molecular breeding.

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

PW: Writing – original draft, Conceptualization, Data curation. YY: Conceptualization, Writing – original draft. MY: Data curation, Writing – review & editing. XP: Methodology, Writing – review & editing. YW: Data curation, Writing – review & editing. XL: Data curation, Writing – review & editing. HY: Conceptualization, Writing – review & editing. NZ: Methodology, Writing – review & editing. WL: Data curation, Writing – review & editing. PD: Conceptualization, Writing – review & editing. LY: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Jilin Province Science and Technology Development Project (20210101190JC), Key Research and Development Project of Jilin Province Science and Technology Development Program (20230204001YY), National Key Research and Development Program of China (2021YFD1600900) and Natural Science Foundation of China (U21A20405).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1301084/full#supplementary-material

References

Abdullah, F. S., Heidari, P., and Poczai, P. (2021). The BAHD gene family in cacao (Theobroma cacao, malvaceae): genome-wide identification and expression analysis. *Front. Ecol. Evol.* 9, 707708. doi: 10.3389/fevo.2021.707708

Ahmad, M. Z., Zeng, X., Dong, Q., Manan, S., Jin, H., Li, P., et al. (2020). Global dissection of the BAHD acyltransferase gene family in soybean: Expression profiling, metabolic functions, and evolution. *Res. Square* doi: 10.21203/rs.2.21482/v2

Ahmadizadeh, M., and Heidari, P. (2014). Bioinformatics study of transcription factors involved in cold stress. *Biharean Biol.* 8, 83–86.

Ardhaoui, M., Falcimaigne, A., Ognier, S., Engasser, J. M., Moussou, P., Pauly, G., et al. (2004). Effect of acyl donor chain length and substitutions pattern on the enzymatic acylation of flavonoids. *J. Biotechnol.* 110, 265–272. doi: 10.1016/j.jbiotec.2004.03.003

Burhenne, K., Kristensen, B. K., and Rasmussen, S. K. (2003). A new class of N-hydroxycinnamoyltransferases: Purification, cloning, and expression of a barley agmatine coumaroyltransferase (EC 2.3.1.64). *J. Biol. Chem.* 278, 13919–13927. doi: 10.1074/jbc.M213041200

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

Cheng, Y., Ahammed, G. J., Yao, Z., Ye, Q., Ruan, M., Wang, R., et al. (2019). Comparative genomic analysis reveals extensive genetic variations of WRKYs in Solanaceae and functional variations of CaWRKys in pepper. *Front. Genet.* 10, 492. doi: 10.3389/fgene.2019.00492

Costaglioli, P., Joubès, J., Garcia, C., Stef, M., Arveiler, B., Lessire, R., et al. (2005). Profiling candidate genes involved in wax biosynthesis in Arabidopsis thaliana by microarray analysis. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1734, 247–258. doi: 10.1016/j.bbalip.2005.04.002

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

D'Auria, J. C., Reichelt, M., Luck, K., Svatoš, A., and Gershenzon, J. (2007). Identification and characterization of the BAHD acyltransferase malonyl CoA: Anthocyanidin 5-O-glucoside-6″-O-malonyltransferase (At5MAT) in Arabidopsis thaliana. *FEBS Lett.* 581, 872–878. doi: 10.1016/j.febslet.2007.01.060

Du, D., Hao, R., Cheng, T., Pan, H., Yang, W., Wang, J., et al. (2013). Genome-wide analysis of the AP2/ERF gene family in prunus mume. *Plant Mol. Biol. Rep.* 31, 741–750. doi: 10.1007/s11105-012-0531-6

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

Friedman, R., and Hughes, A. L. (2001). Pattern and timing of gene duplication in animal genomes. *Genome Res.* 11, 1842–1847. doi: 10.1101/gr.200601

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., et al. (2009). A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* 58, 246–259. doi: 10.1111/j.1365-313X.2008.03773.x

Guo, C., Guo, R., Xu, X., Gao, M., Li, X., Song, J., et al. (2014). Evolution and expression analysis of the grape (Vitis vinifera L.) WRKY gene family. *J. Exp. Bot.* 65, 1513–1528. doi: 10.1093/jxb/eru007

Heidari, P., Ahmadizadeh, M., Izanlo, F., and Nussbaumer, T. (2019). In silico study of the CESA and CSL gene family in Arabidopsis thaliana and Oryza sativa: Focus on post-translation modifications. *Plant Gene* 19, 100189. doi: 10.1016/j.plgene.2019.100189

Hou, S. L., Han, M., Liu, C. J., and Yang, L. M. (2014). Development of real-time fluorescence quantitative RT-PCR assay for β-actin gene of Panax ginseng. *Chin. Tradit. Herb. Drugs* 45, 2530–2533. doi: 10.7501/j.issn.0253-2670.2014.17.020

Kim, N. H., Jayakodi, M., Lee, S. C., Choi, B. S., Jang, W., Lee, J., et al. (2018). Genome and evolution of the shade-requiring medicinal herb Panax ginseng. *Plant Biotechnol. J.* 16, 1904–1917. doi: 10.1111/pbi.12926

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kuate, S. P., Pádua, R. M., Eisenbeiss, W. F., and Kreis, W. (2008). Purification and characterization of malonyl-coenzyme A: 21-hydroxypregnane 21-O-malonyltransferase (Dp21MaT) from leaves of Digitalis purpurea L. *Phytochemistry* 69, 619–626. doi: 10.1016/j.phytochem.2007.08.025

Kusano, H., Li, H., Minami, H., Kato, Y., Tabata, H., and Yazaki, K. (2019). Evolutionary developments in plant specialized metabolism, exemplified by two transferase families. *Front. Plant Sci.* 10, 794. doi: 10.3389/fpls.2019.00794

Li, G., Jones, K. C., Eudes, A., Pidatala, V. R., Sun, J., Xu, F., et al. (2018). Overexpression of a rice BAHD acyltransferase gene in switchgrass (Panicum virgatum L.) enhances saccharification. *BMC Biotechnol.* 18, 1–10. doi: 10.1186/s12896-018-0464-8

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

Liu, A., Liu, C., Lei, H., Wang, Z., Zhang, M., Yan, X., et al. (2020a). Phylogenetic analysis and transcriptional profiling of WRKY genes in sunflower (Helianthus annuus L.): Genetic diversity and their responses to different biotic and abiotic stresses. *Ind. Crops Prod.* 148, 112268. doi: 10.1016/j.indcrop.2020.112268

Liu, C., Qiao, X., Li, Q., Zeng, W., Wei, S., Wang, X., et al. (2020b). 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, C., Wu, Y., Liu, Y., Yang, L., Dong, R., Jiang, L., et al. (2021). Genome-wide analysis of tandem duplicated genes and their contribution to stress resistance in pigeonpea (Cajanus cajan). *Genomics* 113, 728–735. doi: 10.1016/j.ygeno.2020.10.003

Liu, L., Li, X., Li, B., Sun, M.y., and Li, S.x. (2022). Genome-wide analysis of the GRF gene family their expression profiling in peach (Prunus persica). *J. Plant Interact.* 17, 437–449. doi: 10.1080/17429145.2022.2045370

Liu, Y., Wang, X., Mo, T., Yan, Y., Song, Y., Zhao, Y., et al. (2017). Identification and functional application of a new malonyltransferase NbMaT1 towards diverse aromatic glycosides from Nicotiana benthamiana. *RSC Adv.* 7, 21028–21035. doi: 10.1039/C7RA01940H

Liu, Z., Li, Y., Li, X., Ruan, C. C., Wang, L. J., and Sun, G. Z. (2012). The effects of dynamic changes of malonyl ginsenosides on evaluation and quality control of Panax ginseng C.A. Meyer. *J. Pharm. Biomed. Anal.* 64–65, 56–63. doi: 10.1016/j.jpba.2012.02.005

Livak, K. J., and Schmittgen, T. D. (2023). Analysis of relative gene expression data using real- time quantitative PCR and the 2 ⁻ᐤᐤ C T method. *Methods* 408, 402–408. doi: 10.1006/meth.2001.1262

Luo, J., Nishiyama, Y., Fuell, C., Taguchi, G., Elliott, K., Hill, L., et al. (2007). Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana. *Plant J.* 50, 678–695. doi: 10.1111/j.1365-313X.2007.03079.x

Ma, C., Zhang, H., Li, J., Tao, S., Qiao, X., Korban, S. S., et al. (2017). Genome-wide analysis and characterization of molecular evolution of the HCT gene family in pear (Pyrus bretschneideri). *Plant Syst. Evol.* 303, 71–90. doi: 10.1007/s00606-016-1353-z

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., et al. (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5454–5459. doi: 10.1073/pnas.0501102102

Maher, C., Stein, L., and Ware, D. (2006). Evolution of Arabidopsis microRNA families through duplication events. *Genome Res.* 16, 510–519. doi: 10.1101/gr.4680506

Matern, U., Heller, W., and Himmelspach, K. (1983). Conformational changes of apigenin 7-O-(6-O-malonylglucoside), a vacuolar pigment from parsley, with solvent composition and proton concentration. *Eur. J. Biochem.* 133, 439–448. doi: 10.1111/j.1432-1033.1983.tb07483.x

Matern, U., Reichenbach, C., and Heller, W. (1986). Efficient uptake of flavonoids into parsley (Petroselinum hortense) vacuoles requires acylated glycosides. *Planta* 167, 183–189. doi: 10.1007/BF00391413

Mellou, F., Loutrari, H., Stamatis, H., Roussos, C., and Kolisis, F. N. (2006). Enzymatic esterification of flavonoids with unsaturated fatty acids: Effect of the novel esters on vascular endothelial growth factor release from K562 cells. *Process Biochem.* 41, 2029–2034. doi: 10.1016/j.procbio.2006.05.002

Moglia, A., Acquadro, A., Eljounaidi, K., Milani, A. M., Cagliero, C., Rubiolo, P., et al. (2016). Genome-wide identification of bahd acyltransferases and in *vivo* characterization of HQT-like enzymes involved in caffeoylquinic acid synthesis in globe artichoke. *Front. Plant Sci.* 7, 1424. doi: 10.3389/fpls.2016.01424

Molina, I., and Kosma, D. (2015). Role of HXXXD-motif/BAHD acyltransferases in the biosynthesis of extracellular lipids. *Plant Cell Rep.* 34, 587–601. doi: 10.1007/s00299-014-1721-5

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

Nakai, Y., and Yamamoto, (1984). NII-electronic library service. *Chem. Pharm. Bull.* 32, 685–691. doi: 10.1248/cpb.32.685

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

Okada, T., Hirai, M. Y., Suzuki, H., Yamazaki, M., and Saito, K. (2005). Molecular characterization of a novel quinolizidine alkaloid O-tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in Lupinus plants. *Plant Cell Physiol.* 46, 233–244. doi: 10.1093/pcp/pci021

Qiao, X., Li, M., Li, L., Yin, H., Wu, J., and Zhang, S. (2015). Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (Pyrus bretschneideri) and five other Rosaceae species. *BMC Plant Biol.* 15, 1–16. doi: 10.1186/s12870-014-0401-5

Reed, J., Orme, A., El-Demerdash, A., Owen, C., Martin, L. B. B., Misra, R. C., et al. (2023). Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Sci.* 379, 1252-1264. doi: 10.1126/science.adf3727

Ruan, C. C., Liu, Z., Li, X., Liu, X., Wang, L. J., Pan, H. Y., et al. (2010). Isolation and characterization of a new ginsenoside from the fresh root of panax ginseng. *Molecules* 15, 2319–2325. doi: 10.3390/molecules15042319

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122

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

Sonawane, P. D., Gharat, S. A., Jozwiak, A., Barbole, R., Heinicke, S., Almekias-Siegl, E., et al. (2023). A BAHD-type acyltransferase concludes the biosynthetic pathway of non-bitter glycoalkaloids in ripe tomato fruit. *Nat. Commun.* 14, 4540. doi: 10.1038/s41467-023-40092-5

Stothard, P. (2000). Internet on-ramp internet on-ramp. *Biotechniques* 28, 1102–1104. doi: 10.2144/00286ir01

St-Pierre, B., and Luca, V. (2000). Chapter Nine Evolution of acyltransferase genes: Origin and diversification fo the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Adv. Phytochem.* 34, 285–315. doi: 10.1016/S0079-9920(00)80010-6

Sun, B. S., Xu, M. Y., Li, Z., Wang, Y. B., and Sung, C. K. (2012). UPLC-Q-TOF-MS/MS analysis for steaming times-dependent profiling of steamed Panax quinquefolius and its ginsenosides transformations induced by repetitious steaming. *J. Ginseng Res.* 36, 277–290. doi: 10.5142/jgr.2012.36.3.277

Sun, W., Yin, Q., Wan, H., Gao, R., Xiong, C., Xie, C., et al. (2023). Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. *Nat. Commun.* 14, 1–15. doi: 10.1038/s41467-023-42253-y

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M. A., et al. (2002). cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A:anthocyanidin 3-O-glucoside-6′-O-malonyltransferase from dahlia flowers. *Plant Physiol.* 130, 2142–2151. doi: 10.1104/pp.010447

Suzuki, H., Sawada, S., Watanabe, K., Nagae, S., Yamaguchi, M. A., Nakayama, T., et al. (2004). Identification and characterization of a novel anthocyanin malonyltransferase from scarlet sage (Salvia splendens) flowers: An enzyme that is phylogenetically separated from other anthocyanin acyltransferases. *Plant J.* 38, 994–1003. doi: 10.1111/j.1365-313X.2004.02101.x

Taguchi, G., Shitchi, Y., Shirasawa, S., Yamamoto, H., and Hayashida, N. (2005). Molecular cloning, characterization, and downregulation of an acyltransferase that catalyzes the malonylation of flavonoid and naphthol glucosides in tobacco cells. *Plant J.* 42, 481–491. doi: 10.1111/j.1365-313X.2005.02387.x

Taguchi, G., Ubukata, T., Nozue, H., Kobayashi, Y., Takahi, M., Yamamoto, H., et al. (2010). Malonylation is a key reaction in the metabolism of xenobiotic phenolic glucosides in Arabidopsis and tobacco. *Plant J.* 63, 1031–1041. doi: 10.1111/j.1365-313X.2010.04298.x

Tuominen, L. K., Johnson, V. E., and Tsai, C. J. (2011). Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. *BMC Genomics* 12, 236. doi: 10.1186/1471-2164-12-236

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., et al. (2007). Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *J. Biol. Chem.* 282, 15812–15822. doi: 10.1074/jbc.M700638200

Wan, J. Y., Fan, Y., Yu, Q. T., Ge, Y. Z., Yan, C. P., Alolga, R. N., et al. (2015). Integrated evaluation of malonyl ginsenosides, amino acids and polysaccharides in fresh and processed ginseng. *J. Pharm. Biomed. Anal.* 107, 89–97. doi: 10.1016/j.jpba.2014.11.014

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinforma.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Wang, J., and Luca, V. (2005). The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including "foxy" methylanthranilate. *Plant J.* 44, 606–619. doi: 10.1111/j.1365-313X.2005.02552.x

Wang, K., Jiang, S., Sun, C., Lin, Y., Yin, R., Wang, Y., et al. (2015). The spatial and temporal transcriptomic landscapes of ginseng, Panax ginseng C. A. Meyer. *Sci. Rep.* 5, 1–12. doi: 10.1038/srep18283

Wang, Y. S., Jin, Y. P., Gao, W., Xiao, S. Y., Zhang, Y. W., Zheng, P. H., et al. (2016). Complete1H-NMR and13C-NMR spectral assignment of five malonyl ginsenosides from the fresh flower buds of Panax ginseng. *J. Ginseng Res.* 40, 245–250. doi: 10.1016/j.jgr.2015.08.003

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, 1–14. doi: 10.1093/nar/gkr1293

Wang, Z.-H., Wang, X.-F., Lu, T., Li, M.-R., Jiang, P., Zhao, J., et al. (2022). Reshuffling of the ancestral core-eudicot genome shaped chromatin topology and epigenetic modification in Panax. *Nat. Commun.* 13, 1–12. doi: 10.1038/s41467-022-29561-5

Xia, Y., Nikolau, B. J., and Schnable, P. S. (1997). Developmental and hormonal regulation of the Arabidopsis CER2 gene that codes for a nuclear-localized protein required for the normal accumulation of cuticular waxes. *Plant Physiol.* 115, 925–937. doi: 10.1104/pp.115.3.925

Xu, Y., Tie, W., Yan, Y., Xu, B., Liu, J., Li, M., et al. (2021). Identification and expression of the BAHD family during development, ripening, and stress response in banana. *Mol. Biol. Rep.* 48, 1127–1138. doi: 10.1007/s11033-020-06132-9

Yu, X. H., Chen, M. H., and Liu, C. J. (2008). Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in Medicago truncatula. *Plant J.* 55, 382–396. doi: 10.1111/j.1365-313X.2008.03509.x

Yu, X.-H., Gou, J.-Y., and Liu, C.-J. (2009). BAHD superfamily of acyl-CoA dependent acyltransferases in Populus and Arabidopsis: bioinformatics and gene expression. *Plant Mol. Biol.* 70, 421–442. doi: 10.1007/s11103-009-9482-1

Zan, T., Li, L., Xie, T., Zhang, L., and Li, X. (2020). Genome-wide identification and abiotic stress response patterns of abscisic acid stress ripening protein family members in Triticum aestivum L. *Genomics* 112, 3794–3802. doi: 10.1016/j.ygeno.2020.04.007

Zhang, T., Huo, T., Ding, A., Hao, R., Wang, J., Cheng, T., et al. (2019). Genome-wide identification, characterization, expression and enzyme activity analysis of coniferyl alcohol acetyltransferase genes involved in eugenol biosynthesis in Prunus mume. *PloS One* 14, 1–18. doi: 10.1371/journal.pone.0223974

Zhao, J., Huhman, D., Shadle, G., He, X. Z., Sumner, L. W., Tang, Y., et al. (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. *Plant Cell* 23, 1536–1555. doi: 10.1105/tpc.110.080804

Zheng, Z., Qualley, A., Fan, B., Dudareva, N., and Chen, Z. (2009). An important role of a BAHD acyl transferase-like protein in plant innate immunity. *Plant J.* 57, 1040–1053. doi: 10.1111/j.1365-313X.2008.03747.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 December 2023
DOI 10.3389/fpls.2023.1310080

OPEN ACCESS

EDITED BY
Xiang Pu,
Sichuan Agricultural University, China

REVIEWED BY
Tian Xie,
Chinese Academy of Sciences (CAS), China
Qinggang Yin,
China Academy of Chinese Medical Sciences, China

*CORRESPONDENCE
Shuangshuang Yan
ssyan@scau.edu.cn
Bihao Cao
caobh01@scau.edu.cn

†These authors have contributed equally to this work

RECEIVED 09 October 2023
ACCEPTED 21 November 2023
PUBLISHED 22 December 2023

CITATION
Gan Y, Yu B, Liu R, Shu B, Liang Y, Zhao Y, Qiu Z, Yan S and Cao B (2023) Systematic analysis of the UDP-glucosyltransferase family: discovery of a member involved in rutin biosynthesis in *Solanum melongena*. *Front. Plant Sci.* 14:1310080. doi: 10.3389/fpls.2023.1310080

Systematic analysis of the UDP-glucosyltransferase family: discovery of a member involved in rutin biosynthesis in *Solanum melongena*

Yuwei Gan†, Bingwei Yu†, Renjian Liu, Bingbing Shu, Yonggui Liang, Yafei Zhao, Zhengkun Qiu, Shuangshuang Yan* and Bihao Cao*

Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (South China), Ministry of Agriculture and Rural Affairs/Guangdong Vegetable Engineering and Technology Research Center/ College of Horticulture, South China Agricultural University, Guangzhou, China

Eggplant (*Solanum melongena*) is an economically important crop and rich in various nutrients, among which rutin that has positive effects on human health is found in eggplant. Glycosylation mediated by UDP-glycosyltransferases (UGTs) is a key step in rutin biosynthesis. However, the *UGT* gene has not been reported in eggplant to date. Herein, 195 putative *UGT* genes were identified in eggplant by genome-wide analysis, and they were divided into 17 subgroups (Group A-P and Group R) according to the phylogenetic evolutionary tree. The members of Groups A, B, D, E and L were related to flavonol biosynthesis, and rutin was the typical flavonol. The expression profile showed that the transcriptional levels of *SmUGT* genes in Clusters 7-10 were closely related to those of rutin biosynthetic pathway genes. Notably, *SmUGT89B2* was classified into Cluster 7 and Group B; its expression was consistent with rutin accumulation in different tissues and different leaf stages of eggplant. SmUGT89B2 was located in the nucleus and cell membrane. Virus-induced gene silencing (VIGS) and transient overexpression assays showed that *SmUGT89B2* can promote rutin accumulation in eggplant. These findings provide new insights into the UGT genes in eggplant, indicating that SmUGT89B2 is likely to encode the final enzyme in rutin biosynthesis.

KEYWORDS

eggplant, UGT superfamily, rutin, biosynthesis, functional identification

Introduction

As an important horticultural crop, eggplant (*Solanum melongena* L.) is the third most cultivated *Solanaceae* species in the world, next to potato (*Solanum tuberosum* L.) and tomato (*Solanum lycopersicum* L.) (Barchi et al., 2021). Eggplant is planted on 1.85 million hectares with a worldwide production of 56.30 million tons, which is mainly cultivated in

India and China, and consumed worldwide (FAO, see www.fao.org) (Martínez-Ispizua et al., 2021; Wang et al., 2022). Flavonoids are one of the most valued compounds in eggplant fruit: anthocyanins are the highest, followed by flavonol (Martínez-Ispizua et al., 2021). Currently, all kinds of flavonols have been identified in eggplant fruit, including quercetin, rutin, myricetin, quercetin-3-diglucoside, kaempferol-3,7-diglucoside, kaempferol-diglucoside and quercetin-3-rhamnoside (Singh et al., 2017; Sulli et al., 2021). Flavonols are plant-specialized metabolites with important functions in regulate plant development, signaling, and stress response (Daryanavard et al., 2023).

Rutin is a typical flavonol and plays a key role in defense responses against abiotic and biotic stresses, such as resistance to low temperatures, adaptation to environmental change, and tolerance to fungal and viral pathogens (Suzuki et al., 2005; Yang et al., 2022). Rutin has wide range of pharmacological activities and high antioxidant capacity. It is widely used in clinical medicine, e.g., for anticarcinogenic, antiviral and anti-inflammation treatment and for the protection of nerves and the cardiovascular system (Glasser et al., 2002; Jimenez-Aliaga et al., 2011; Luca et al., 2020). Rutin is the diglycoside form of quercetin, which is the main flavonol in vegetables and fruits. It is the most popular dietary flavonoid compound in the human diet (Hosseinzadeh and Nassiri-Asl, 2014; Tobar-Delgado et al., 2023). Although the biosynthetic pathway of flavonol is clear, the enzyme in the final two glycosylation steps of quercetin conversion into rutin remains elusive (Figure S1A) (Vogt, 2010; Zhang et al., 2017; Shen et al., 2022). *UGT* encodes the final enzyme in the rutin biosynthetic pathway and determines the pharmaceutical functions of rutin (Jing et al., 2016).Glycosylation is one of the most important modifications. Glycosylation catalyzed by glycosyltransferase is an important modification process in plants and is one of the final steps involved in secondary metabolite synthesis (Caputi et al., 2012). Glycosyltransferase catalyzes the transfer of glycosyl from activated donor molecules onto various small molecule receptors, leading to the production of di-saccharides, poly-glycosides, and diversified glycosides of noncarbohydrate moieties (Caputi et al., 2012; Wilson and Tian, 2019). The plant secondary product glycosyltransferase (PSPG) box is the binding site of the sugar donor. It consists of a conserved sequence of 44 amino acid residues in the UGT of plants near the C-terminus (Gachon et al., 2005; Offen et al., 2006). The UDP-glycosyltransferase superfamily is a large family that possesses redundant functions and complex structures and presents distinct but overlapping substrate specificities (Mackenzie et al., 1997). Members of the UGT superfamily are named using the acknowledged convention (Figure S1B). Numbers from 1 to 50 are utilized for animals, 51 to 70 for yeasts and amoebae, 71 to 100 for plants and 101 to 200 for bacteria (Figure S1B) (Ross et al., 2001). With the development of molecular biology, an increasing number of *UGT* genes have been cloned and characterized, which are enrich of rutin biosynthesis. Rutin is formed by the two-step glycosylation of quercetin. For example, FaGT6 was proven to be responsible for the first step, which catalyzed quercetin to 3-*O*-glucosides (Griesser et al., 2008). *UGT* genes with similar functions in the first step of glycosylation have been reported in both soybean (Kovinich et al., 2010) and grape (Mizohata et al., 2013). GmUGT79A6 was identified as rhamnose-glycosyltransferase involved in the second step of rutin glycosylation (Rojas et al., 2014). In addition, FeUGT79A8 (Koja et al., 2018) and FtUGT73BE5 (Yin et al., 2020) were reported to promote rutin accumulation in buckwheat. Many *UGT* genes involved in key secondary metabolite biosynthesis have been successfully identified and investigated in different plant species, including *Arabidopsis* (Paquette et al., 2003), pomegranate (Wilson et al., 2019), buckwheat (Yin et al., 2020), soybean (Yin et al., 2017), and tea (Cui et al., 2016; Dai et al., 2022). However, the genes of rutin biosynthesis are still unknown in eggplant.

In this study, the pangenome of eggplant was used to identify the UGT family and obtain potential *UGT* genes that might be related to rutin biosynthesis. A total of 195 *UGT* candidate genes were identified in eggplant, and their gene structure and evolutionary relationship were analyzed. Five *UGT* candidate genes were identified with rutin biosynthesis *via* expression pattern and correlation analysis. Furthermore, *SmUGT89B2* was proven to participate in rutin biosynthesis in eggplant by using VIGS and transient overexpression assays.

Material and methods

Identification of UGTs in the eggplant genome and their chromosomal locations

UGT families were retrieved from the eggplant genome V4.0 downloaded from the Solanaceae genomics database (https://solgenomics.net/) by using the following two methods: 44 amino acid sequences of PSPG-box were used as query sequences to blast the eggplant, and the Hidden Markov Model (HMM) profile of PF00201. The members of UGT families needed to meet the requirement of the two methods. The NCBI Conserved Domain Search Service (Batch CD-Search) and SMART were used to filter redundant sequences. The full-length amino acid sequences, molecular weight (MW), theoretical isoelectric point (PI) and instability index of the UGT proteins were predicted by the ExPASy server. Based on the gff3 file of eggplant, the distribution of *UGT* genes on chromosomes was retrieved and visualized by using TBtools (Chen et al., 2020).

Sequence alignment and phylogenetic analysis

The full-length amino acid sequences of all SmUGTs were aligned by Clustal X2 software with default values used for all parameters. The phylogenetic evolutionary tree was constructed using the maximum likelihood method through MEGA X software with 2000 bootstrap replications (Kumar et al., 2018). The 122 UGT protein sequences of Arabidopsis used in the tree were downloaded from the site (http://www.p450.kvl.dk/index).

Gene structure and conserved motif analysis

Based on the relationship of the coding sequence and its corresponding genomic DNA sequence from the gff3 file, the gene structure of *SmUGTs* was visualized by using TBtools. Conserved motifs or domains were searched by MEME (https://meme-suite.org/meme/tools/meme)(Bailey et al., 2015), with parameters as follows: number of motif predictions, 20; minimum motif width, 6; maximum motif width, 50 (Bailey et al., 2009), and then visualized by using TBtools.

RNA-seq data analysis

The transcriptome data at different developmental stages of eggplant fruit were obtained from previous eggplant genome research (Barchi et al., 2019). A heatmap reflecting the expression patterns of *SmUGTs* at different developmental stages of eggplant fruit was constructed by TBtools.

Plant materials

The E801 inbred line used in the study was provided by our lab. Eggplant seeds were sown in hole trays when the seedlings grew to the three-leaf stage and transplanted to the experimental farm (South China Agricultural University, Qilin). The different tissues of eggplant, including the roots, stems, leaves, flowers and commercial fruits, were collected, immediately frozen in liquid nitrogen and then stored at -80°C for RNA extraction and determination of rutin content. The different developmental stages of eggplant fruits at 5 DPA, 12 DPA, 15 DPA, 22 DPA and 30 DPA were collected, and immediately frozen in liquid nitrogen and then stored at -80°C for RNA extraction and determination of rutin content (Figure S2).

Total RNA extraction and quantitative real-time PCR

Total RNA was extracted from eggplant tissues by an Eastep Super Total RNA Extraction Kit (Promega, WI, USA) according to Liu et al. (Liu et al., 2023). First-strand cDNA was synthesized by using a cDNA Kit PLUS (B0003) (EZBioscience, Roseville, MN, USA) according to the manufacturer's instructions. All primer sequences are listed in Table S2, including reference genes. The relative expression of UGT genes was calculated as described by Song et al. (Song et al., 2020). The qPCRs using different tissues were performed with biological triplicates.

Determination of rutin content

The rutin content was detected according to Dong et al. (Dong et al., 2020). Fresh tissues (0.1 g) were ground into a fine powder with liquid nitrogen, extracted with 2 mL ethanol (60%) and centrifuged for 10 min at 12000 rpm. The supernatant was transferred to a new 2 mL centrifuge tube and placed in a metal bath at 80 °C for 2 h. After heating, 0.4 mL of supernatant was mixed with 0.4 mL of $AlCl_3$ (0.1 mol/L) and 0.6 mL of CH_3COOK (1 mol/L) solution and then added to 0.6 mL of ethanol (60%). The mixture was shaken and allowed to stand for 30 minutes. The absorbance of the solution was measured using a microplate reader at 420 nm. Determination of rutin content was performed in triplicate.

Subcellular localization analysis

To confirm the Subcellular localization of *SmUGT89B2*, the coding DNA sequence (CDS) of SmUGT89B2 was cloned from eggplant. Its coding sequence without a terminator codon was cloned and inserted into a vector fused with EGFP carrying the CaMV 35S promoter to generate the 35S: SmUGT89B2-EGFP construct by using one-step cloning (C112, Vazyme, China) according to the manufacturer's instructions. The recombinant plasmid was transformed into competent cells of *Agrobacterium* strain GV3101. *Agrobacterium tumefaciens* GV3101 containing the corresponding constructs was incubated with YEP medium at 28°C. The bacterial cells were collected and modulated 0.4-0.6 at 600 nm using MS medium (10 mM 2-(4-morpholino)-ethane sulfonic acid, 10 mM $MgCl_2$, and 200 μM acetosyringone, pH = 5.6) and then placed in darkness at 28°C for 2 h. *Agrobacterium tumefaciens* GV3101 containing the 35S:SmUGT89B2-EGFP construct and empty vector were mixed with the NLS-DsRed construct in equal proportions (1:1). The mixed solution was infiltrated into the abaxial side of an *Nicotiana benthamiana* leaf. Three days later, the GFP signal of leaves that were inoculated with *Agrobacterium tumefaciens* was observed by using fluorescence microscopy (Leica, Germany) (Song et al., 2023). The primers used are listed in Table S2.

Virus-induced gene silencing assays

The full-length CDS of SmUGT89B2 was predicted to be a hypothetically optimal fragment (300 bp) by using the Solanaceae database (http://vigs.solgenomics.net/). The 300 bp fragment was amplified and ligated to the pTRV2 vector. pTRV2-*SmUGT89B2*, pTRV1 and pTRV2 were transformed into competent cells of Agrobacterium strain GV3101. *Agrobacterium tumefaciens* GV3101 containing pTRV2-*SmUGT89B2* and pTRV1 were mixed in equal proportions, and the bacterial solution was infiltrated into the abaxial side of 30-day-old seedlings of eggplant with a 1 mL syringe. The eggplant plants were grown in a growth chamber under growth conditions as follows: 60–70% humidity, 22-25°C, 16 h/8 h (light/dark cycle). The primers used are listed in Table S2.

Transient overexpression analysis

Transient overexpression analysis was performed according to a previous method (Wang et al., 2022). The full-length coding sequence

of *SmUGT89B2* was cloned and inserted into the PC1300 vector. The recombinant vector was transformed into competent cells of Agrobacterium strain GV3101. The bacterial solution containing the corresponding constructs was infiltrated into the abaxial side of eggplant (E801) leaves at the five-leaf stage, and the plants were grown in a growth chamber. After 5 days, the leaves of eggplant were collected, frozen in liquid nitrogen and stored at −80°C for RNA extraction and determination of rutin content. The primers used are listed in Table S2.

Results

Identification and chromosomal localization of eggplant UGT genes

Based on the V4 version of the eggplant genome, the HMM model was used for searching *UGT* genes. After redundancy was removed by SMART and CDD-Search, a total of 195 *UGT* genes were identified. All candidate *UGT* genes contained a conserved UDPGT domain, which encoded amino acid sequences ranging from 154 aa to 1197 aa in length. The molecular weight (MW) of these proteins ranged from 17.02 kDa to 134.07 kDa, and the scope of the theoretical pI was 4.81 to 9.23. Most of the proteins were stable, and only fifty-seven (29% of the total) of SmUGT proteins were unstable (Table S1). Analysis of their chromosome location distribution showed that all candidate *UGT* genes were mapped on chromosomes 1-12, except for SMEL4_00g012200, which was not assembled to any chromosomes due to sequencing and assembly technology limitations. Among these chromosomes, chromosome 11 contained the largest number of SmUGT members (36). Chromosome 1, chromosome 5, and chromosome 10 contained 17, 23, and 26 members, respectively. Chromosome 12 (7) included the fewest number of SmUGT members (7) (Figure 1).

Phylogenetic analysis of eggplant UGT family proteins

To clarify the phylogenetic relationship of the UGT family in eggplant, a phylogenetic tree was constructed by the maximum likelihood method with all of the identified SmUGT protein sequences and those in other plants, including 122 *Arabidopsis UGTs*, two *UGT* genes of Group P in rice (Li et al., 2014) and four *UGT* genes of Groups Q and O in maize (Li et al., 2014; Huang et al., 2015). These UGT proteins from rice and maize were conducive to the classification of the SmUGT phylogenetic group. According to the phylogenetic tree, SmUGTS were clustered into 17 phylogenetic groups (A-P, R) and an outgroup (Figure 2A). As reported in *Arabidopsis*, *AtUGT84B1* from Group L was involved in glycosylation of IAA (Mateo-Bonmati et al., 2021). Both *AtUGT79B1* and *AtUGT91A1* from Group A were proven to mediate anthocyanin modification (Ono et al., 2010). *AtUGT89C1*, *AtUGT73C6* and *AtUGT78D1*, which belonged to Groups B, D and E, respectively, were related to glycosylation of flavonol (Jones et al., 2003; Yonekura-Sakakibara et al., 2007). It was likely that members of Groups A, B, D, E and L were related to phytohormone regulation and secondary metabolite biosynthesis.

To further understand the distribution of the UGTs across the phylogenetic range covered in different plant species, the number of UGT proteins distributed in phylogenetic groups is summarized in Table 1. Notably, some significant differences were observed in these phylogenetic groups. Compared with tomato, the members of eggplant UGTs were absent in Group Q, whereas the members of Solanaceae UGTs were present in Groups O, P and R compared with Arabidopsis, especially the members of Group R only found in Solanaceae. Group A contained the largest numbers of SmUGT members (35), followed by Groups O (30) and D (29). Groups F, I, J and N included a small number of UGT members in these plants. Compared with that of other species, the members of Group O in Solanaceae species had significant expansion, as well as tomato. The results indicated that Group O might play a vital role in Solanaceae plants.

The PSPG-box was located in the UDPGT conserved domain. The PSPG box was distinct in different phylogenetic groups, in which highly conserved residues were observed at positions 1 (W), 4 (Q), 8 (L), 10 (H), 19-24 (HC/SGWNS), 27 (E), and 44 (Q) (Figure 2B). Some amino acids could interact directly with UDP-sugar donors to form hydrogen bonds, which were related to the evolution and function of enzymes. The amino acid residue Q (44) has been reported to determine the sugar donor specificity of UGTs, which affects the catalytic efficiency of glucosyl transfer activity (Ono et al., 2010; Caputi et al., 2012).

Gene structure and conserved motif analysis of *SmUGTs*

To investigate the structural features of SmUGT proteins, conserved amino acid sequences and gene structure were analyzed. Except for the outgroup, there were 89 members without introns (46%). The number of intron regions ranged from 1 to 7 in the rest of the members (Figure 3B). SMEL4_05g000470 (from Group O) was one of the most complex genes and had 7 introns. All members of Groups G, H, I, J, K, N, and P had introns. All SmUGT amino acid sequences were submitted to MEME, and a total of 20 conserved motifs were obtained (Table S3). These motifs possessed 11-50 residues and were named motif 1-motif 20. Motif 1, located in the PSPG box, existed in all SmUGT genes and served as a specific motif of the UGT family. In general, the same group has a similar motif distribution, whereas different groups possess a specific distribution of motifs. As shown in Figure 3A, motif 14 was mainly distributed in Groups O and A, while motif 16 and motif 18 were primarily present in Groups D and M. Motif 13 and motif 19 were found only in Group O.

Expression analysis of *SmUGTs* in different developmental stages of eggplant fruit

Expression patterns can predict the biological functions of genes to a certain extent. To investigate the underlying functions of

10.3389/fpls.2023.1310080

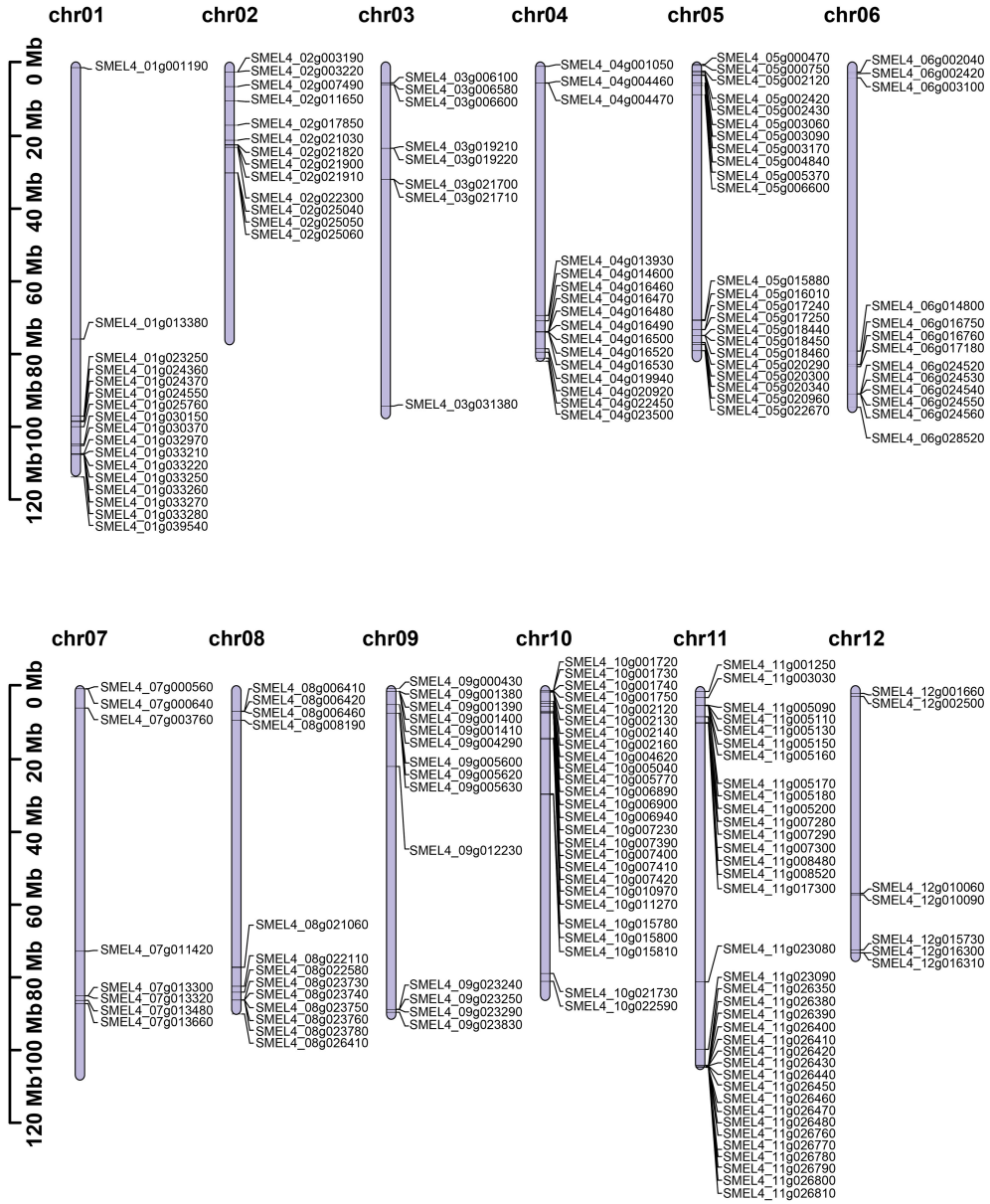


FIGURE 1
The distribution of chromosome locations. The cylinder represents the chromosomes, and the numbers above the column represent the number of chromosomes.

candidate *UGT* genes during rutin biosynthesis processes, transcriptome data of eggplant at different developmental stages from a public database were downloaded (Barchi et al., 2019). The expression patterns of candidate *SmUGTs* and rutin biosynthesis genes in different developmental stages of eggplant fruit were analyzed. The rutin biosynthesis genes were highly expressed in stage 3 (Figure 4A). A total of 23 *SmUGTs* (12%) were expressed at a level that could not be detected. Based on the similar expression trend, all *SmUGTs* in different developmental fruit stages were hierarchically clustered and divided into 10 clusters (Figure 4B). The expression patterns of Clusters 7, 8, 9 and 10 were similar to the expression patterns of rutin biosynthesis genes, and their expression levels gradually increased with fruit development. Therefore, the members of Clusters 7–10 served as novel candidates for the regulation of rutin biosynthesis.

Expression patterns of *SmUGTs* in different tissues and different developmental stages of fruits and leaves in eggplant

To further identify *SmUGT* genes involved in rutin biosynthesis in eggplant, five *SmUGT* genes that belonged to Groups A-F and L with a higher expression level were selected from candidate Clusters 7 to 10 to perform qRT–PCR assays. These five genes were renamed by the UGT Nomenclature Committee, called *SmUGT71BA3*,

10.3389/fpls.2023.1310080

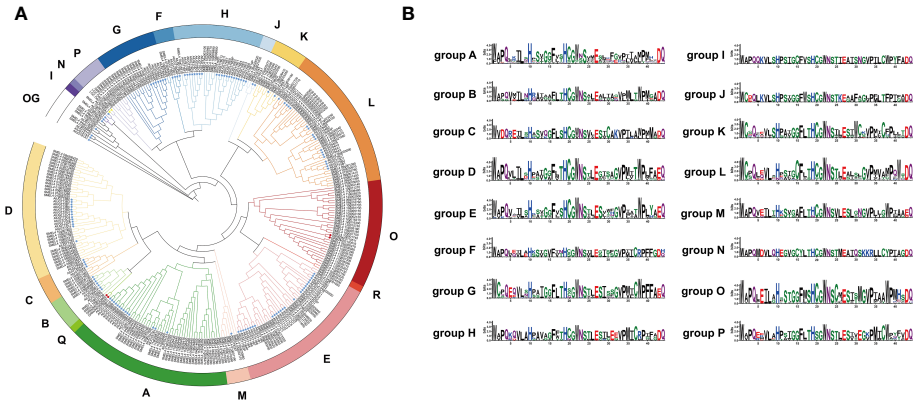


FIGURE 2
(A) Phylogenetic tree of the UGT family members. The tree consisted of all the identified *UGT* genes from eggplant, two *UGT* genes from rice (Li et al., 2014), four *UGT* genes from maize (Li et al., 2014; Huang et al., 2015) and all *Arabidopsis UGT* genes (Wilson and Tian, 2019). The tree was constructed using the maximum likelihood method. Different colored branches represent different groups. The blue star represents *Arabidopsis*. The yellow triangle represents rice, and the red circle represents maize. **(B)** Logos of the PSPG motif of phylogenetic Groups A - P generated by WebLogo (Crooks et al., 2004). The height of the amino acid indicates the frequency observed in multiple alignment.

SmUGT72B88, *SmUGT73A69*, *SmUGT75L62* and *SmUGT89B2*, respectively (Table S2). Rutin accumulated in all tissues of eggplant, and the highest content was found in leaves, followed by flowers and pericarp. Meanwhile, the highest content was found at 5 DPA, and with the development of fruit, the rutin content gradually decreased (Figure 5A). The expression of *SmUGT72B88*, *SmUGT73A69* and *SmSmUGT89B2* was the highest in leaves, followed by flowers and fruits, which was consistent with the accumulation of rutin in different tissues of eggplant (Figure 5B). However, *SmUGT71BA3* was highly expressed in roots and fruits

TABLE 1 Summary of each group of UGT families in plant species.

Group	Eggplant	Tomato	*Arabidopsis*	Rice	Maize
A	35	24	14	14	8
B	5	2	4	9	3
C	4	2	4	8	5
D	29	17	13	26	18
E	17	16	25	38	34
F	2	2	4	/	2
G	11	13	7	20	12
H	6	6	21	7	9
I	1	2	1	9	9
J	2	1	2	3	3
K	8	5	2	1	1
L	25	20	18	23	23
M	5	3	1	5	3
N	1	1	1	2	4
O	30	25	/	6	5
P	7	6	/	9	1
Q	/	1	/	/	7
R	2	1	/	/	/
OG	5	15	5	/	/
Total	195	162	122	180	147

10.3389/fpls.2023.1310080

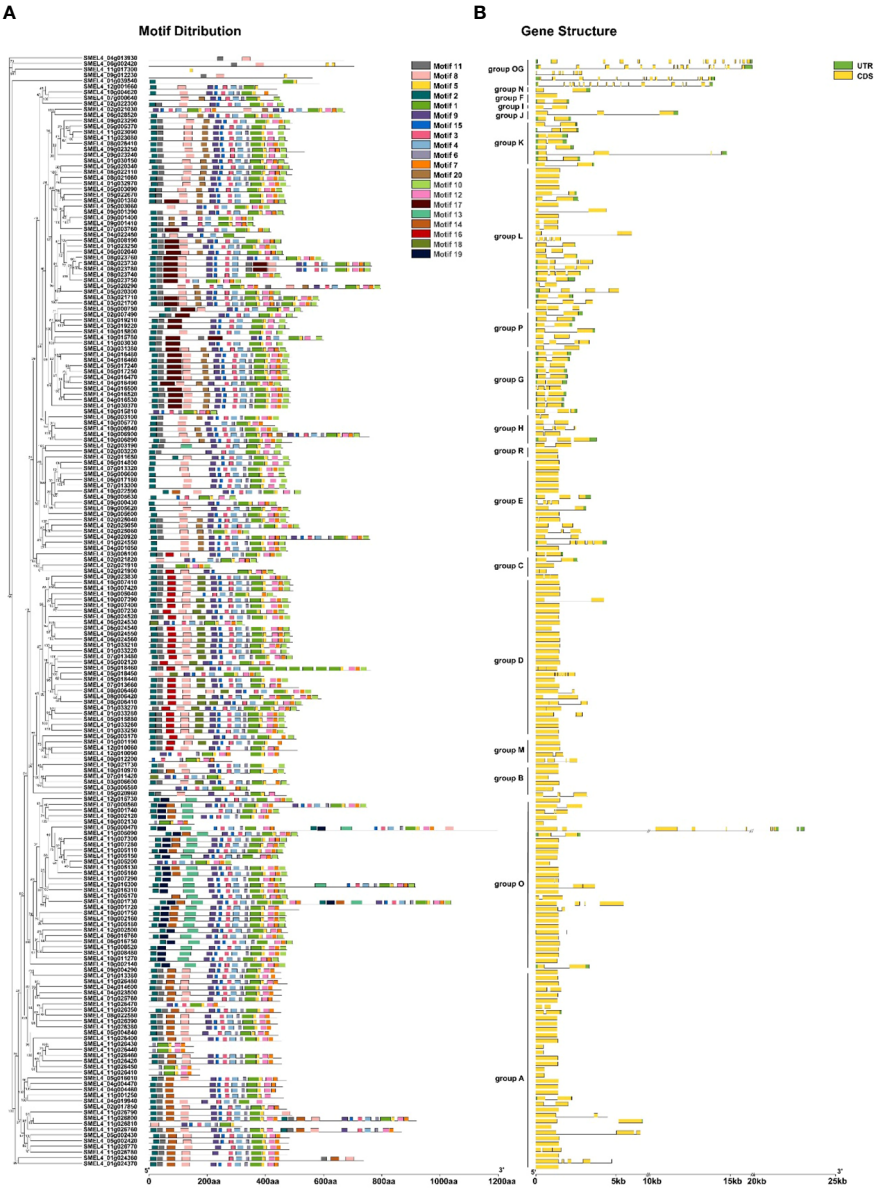


FIGURE 3
Conserved motifs and gene structure of *SmUGTs*. The phylogenetic tree was constructed by all *UGT* gene sequences using the maximum likelihood method, and the conserved motif **(A)** and gene structure **(B)** were placed orderly on the right side of the phylogenetic tree. Different colored boxes represent different motifs, and the box length represents the motif length in the motif distribution. The yellow box indicates exons; the green box indicates UTRs; and the black line indicates introns.

compared with that in other tissues. The expression of *SmUGT75L62* was highest in pericarp, whereas it was transcribed at a low level in stem that was not detectable. The expression pattern of *SmUGT71BA3* was inversely represented in pericarp and pulp, showing a significant decrease in pericarp and a marked increase in pulp from 5 DPA to 30 DPA. In the other four genes, including *SmUGT72B88*, *SmUGT73A69*, *SmUGT75L62* and *SmUGT89B2*, their own expression levels were similar in the pericarp and pulp. The expression levels of *SmUGT73A69* and *SmUGT75L62* exhibited a growing tendency with the process of fruit development, while *SmUGT72B88* and *SmUGT89B2* were observed to have higher expression at 5 DPA and then showed a decrease at 12 DPA. Their expression levels were not significantly different from 12 DPA to 30 DPA, except for that of *SmUGT89B2*, which increased at 30 DPA (Figure 5B).

To further explore the expression patterns of these five genes, leaves at different developmental stages were used for verification. During the course of leaf senescence, the rutin content gradually increased from young leaves to mature leaves and then decreased, with a peak at mature leaves (Figure 5C). As shown in Figure 5D, the transcription levels of *SmUGT72B88*, *SmUGT73A69* and *SmUGT75L62* increased with the process of leaf senescence. The

10.3389/fpls.2023.1310080

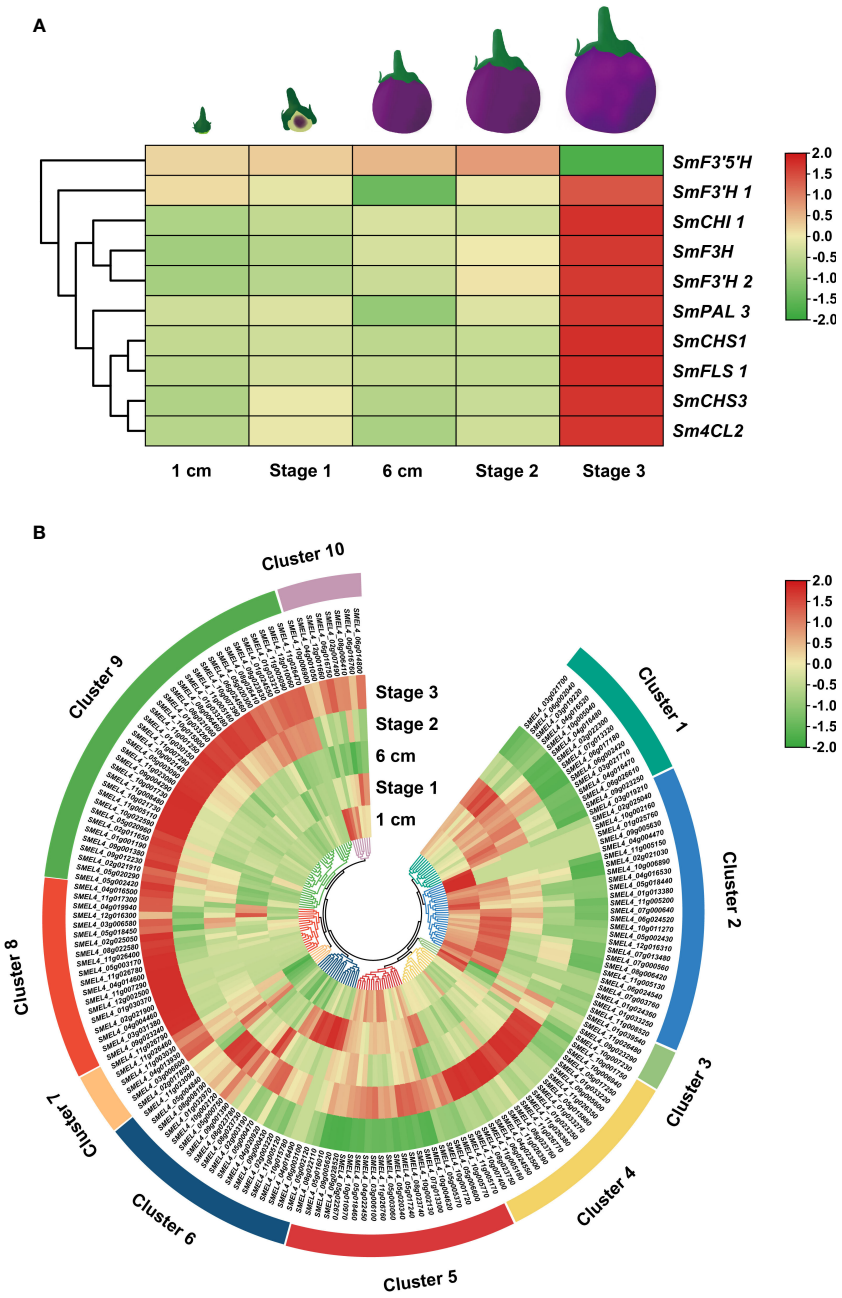


FIGURE 4
Expression patterns of rutin biosynthesis **(A)** and *UGT* **(B)** genes. Fruit 1 cm represents 3-4 days after flowering (DAF); fruit stage 1 represents 8-14 DAF; fruit 6 cm represents 18-21 DAF; fruit stage 2 represents approximately 38 DAF; fruit stage 3 represents 55-60 DAF.

expression patterns of *SmUGT89B2* maintained good agreement with the accumulation of rutin in different developing leaves.

Sequence analysis and subcellular location of *SmUGT89B2*

To determine whether *SmUGT89B2* is involved in rutin biosynthesis in eggplant, the amino acid sequence of SmUGT89B2 was analyzed. The full-length CDS of *SmUGT89B2* was cloned from the E801 inbred line. Its protein encoded 483 amino acids. The molecular weight (MW) of SmUGT89B2 was 53.83 kDa, and theoretical PI was 6.44. It is an unstable protein, and the instability index was 42.04. The aliphatic index (95.51) was predicted that SmUGT89B2 was a hydrophilic protein (Table S1). Several UGT proteins that have been reported to be involved in flavonol glycosylation were obtained (Carbonell-Bejerano et al., 2014; Yin et al., 2017; Koja et al., 2018; Yin et al., 2020), and some UGT proteins from other Solanaceae species that shared the highest sequence identity with SmUGT89B2 were retrieved from

10.3389/fpls.2023.1310080

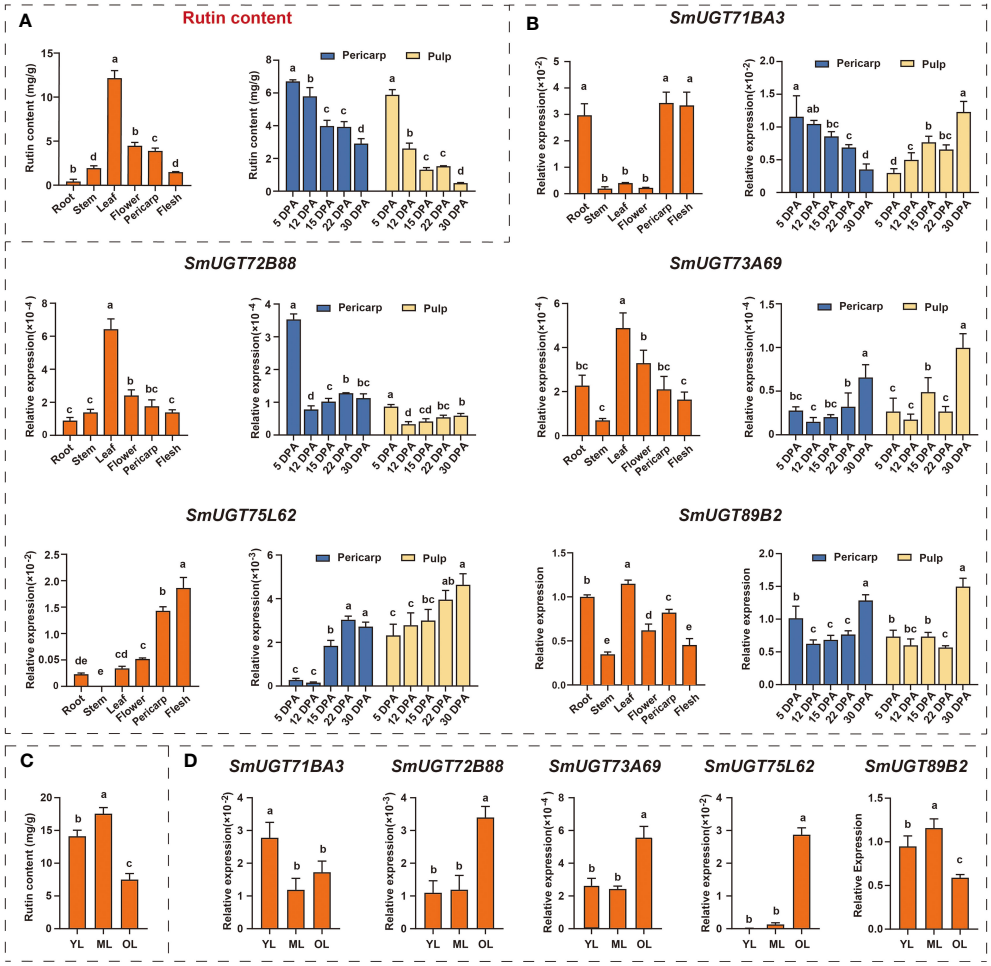


FIGURE 5
The expression patterns of five *SmUGT* genes in different tissues and different developmental stages of leaves and fruit in eggplant (E801). **(A)** Rutin content in different eggplant tissues (roots, stems, leaves, flowers and pericarp and pulp). **(B)** Expression pattern of five *SmUGT* genes in different eggplant tissues (roots, stems, leaves, flowers, pericarp and flesh). Data are the mean ± SE (n = 3). Different lowercase letters in the figures represent significantly different values ($P < 0.05$, Tukey's test). **(C)** The rutin content at different developmental stages of leaves in E801. **(D)** Expression of five *SmUGT* genes at different developmental stages of leaves. Different lowercase letters in the figures represent significantly different values ($P < 0.05$, Tukey's test).

public databases (NCBI, https://www.ncbi.nlm.nih.gov/). All of them were used to construct an unrooted phylogenetic tree based on their multiple sequence alignment. As shown in the phylogenetic tree, the UGT genes involved in flavonol glycosylation in the different species had a distant evolutionary relationship with SmUGT89B2. The conserved PSPG-box motif existed in all UGT amino acid sequences, in which some amino acids were highly conserved (Figure 6A) (Caputi et al., 2012). To clarify the subcellular localization of SmUGT89B2, its sequence was fused in frame with the EGFP gene to generate the 35S:SmUGT89B2-EGFP construct (Figure 6B). The DsRed gene served as the nuclear marker by fusion with a nuclear localization signal (NLS) (Sun et al., 2020). The coexpression of either 35S: EGFP or 35S:SmUGT89B2-EGFP with DsRed in epidermal cells of *Nicotiana benthamiana* leaves indicated that SmUGT89B2 was localized in the nucleus and cell membrane, as well as in the cytoplasm (Figure 6B). It was likely that SmUGT89B2 functioned in the whole cell.

SmUGT89B can positively promote rutin accumulation in eggplant

To determine whether *SmUGT89B2* was associated with rutin accumulation in eggplant, the transcriptional level of *SmUGT89B2* was silenced in the E801 inbred line through virus-induced gene silencing (VIGS). Three silenced eggplant lines that possessed the highest silencing efficiency were selected to determine *SmUGT89B2* expression and the rutin content. As shown in Figure 7B, the expression level of *SmUGT89B2* was significantly decreased in silenced lines, reducing by approximately 60% of that in the control plants. Simultaneously, the rutin content in the silenced leaves decreased by approximately 40% of that in the control leaves. To further confirm whether *SmUGT89B2* plays a role in rutin biosynthesis in eggplant, a transiently overexpressed assay was performed in eggplant seedlings at the five-leaf stage. The *SmUGT89B2*

10.3389/fpls.2023.1310080

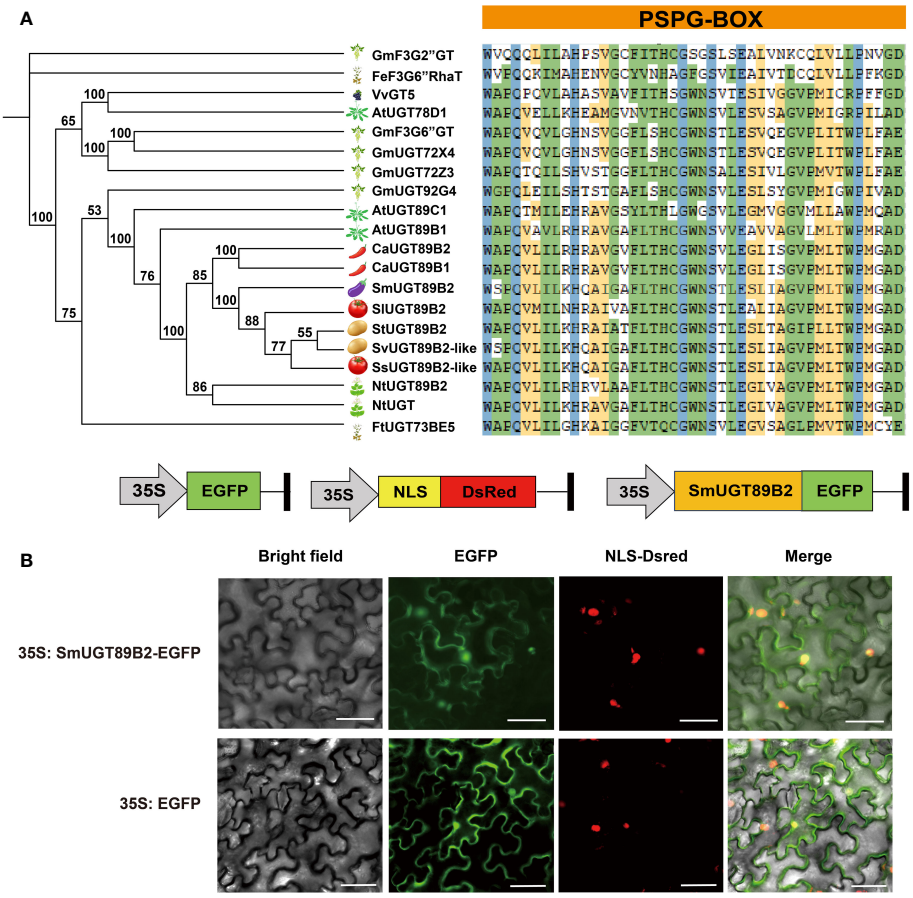


FIGURE 6
Sequence analysis and subcellular localization of SmUGT89B2. **(A)** Multiple alignment of PSPG-box motif (right) and a phylogenetic tree (left) of *SmUGT89B2* with homologous genes from other plant species. The sequences of other plant species were obtained from NCBI and soybean database. A phylogenetic tree was built *via* MEGA X using the maximum likelihood method with 2000 bootstrap replicates. **(B)** Subcellular localization of SmUGT89B2. EGFP represents enhanced GFP. NLS-DsRed served as a nuclear marker, and fluorescence was visualized using an inverted fluorescence microscope. Scale bar = 50 μm.

transcription levels increased 8.8-28-fold in the overexpressed leaf compared with those in the control, consistent with the 1.2-1.7-fold increase in rutin content in overexpressed leaf (Figure 7D). These results suggested that *SmUGT89B2* played an important role in rutin biosynthesis in eggplant.

Discussion

Glycosyltransferase is a supergene family, and plant *UGTs* are classified into the first family (GT1) of the glycosyltransferase (GT) family. Glycosylated metabolites generated by UDP-dependent glycosyltransferases play important roles in plant interactions with the environment as well as human and animal nutrition (Wilson and Tian, 2019). The UGT family has been identified in various plants, including *Arabidopsis* (123) (Li et al., 2001), cotton (274) (Xiao et al., 2019), maize (147) (Li et al., 2014), peach (168) (Wu et al., 2017), buckwheat (106) (Yin et al., 2020), soybean (212) (Yin et al., 2017), tea (178) (Cui et al., 2016), pomelo (145) (Wu et al., 2020), and *Brassica rapa* (147) (Yu et al., 2017). The numbers of *UGT* genes present significant differences between various plants. In this study, a total of 195 *UGT* genes identified from the eggplant genome were divided into 17 groups by constructing a phylogenetic tree with other plant *UGTs*, containing 14 highly conserved phylogenetic groups (A-N) and three new phylogenetic groups (O, P, and R) (Figure 2A). Group O contained 30 *UGT* members in eggplant, while it contained approximately the same amount of UGT members in tomato (25) (Wilson and Tian, 2019) and tabacoo (40) (Yang et al., 2023). In contrast, the number of *UGT* members was dramatically reduced in rice (6), maize (5) and other dicotyledon plants (Wilson and Tian, 2019). Thus, the members of Group O were significantly expanded in Solanaceae plants and seem to be unique to Solanaceae species. *UGT709G2* from Group P shows a substrate preference for UDP-Glc and catalyzes myricetin 3-O-rhamnoside into myricetin 3-O-glucosyl rhamnoside in montbretia (Irmisch et al., 2018). The members of Group R were recently discovered and first identified in buckwheat. *FeUGT708C1* from Group R could recognize and catalyze the C-glycosylation of trihydroxy acetophenone-like compounds (Nagatomo et al., 2014).

10.3389/fpls.2023.1310080

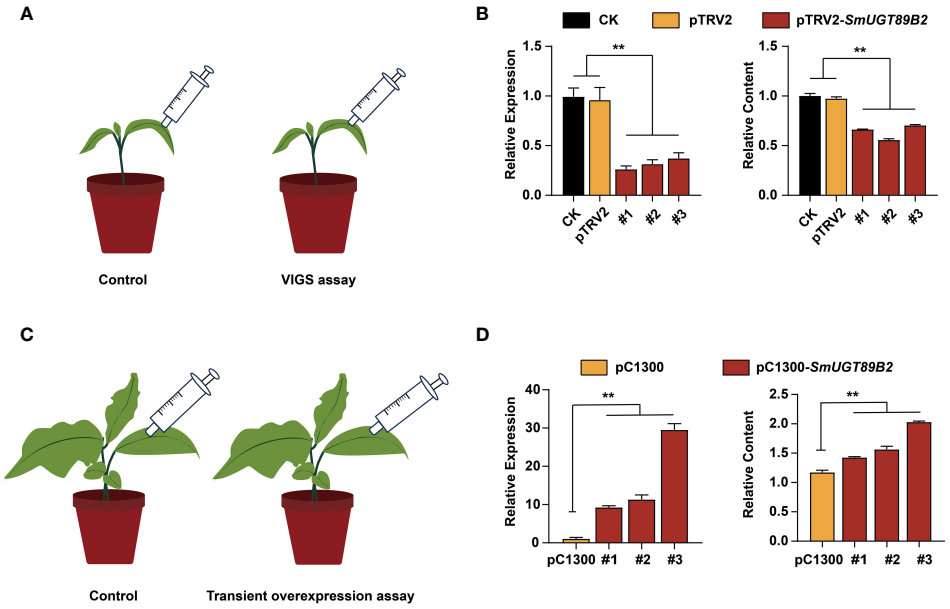


FIGURE 7
Silencing and transient overexpression of *SmUGT89B2* in eggplant. **(A)** Schematic diagram of the SmUGT89B VIGS assay. **(B)** The expression of *SmUGT89B* and relative rutin content in leaves of *SmUGT89B*-silenced eggplants. CK were injected with water. pTRV2 was injected with *Agrobacterium* harboring an empty vector. Data are the mean ± SE (n = 3). **(C)** Schematic diagram of the transient overexpression assay for *SmUGT89B2*. **(D)** The expression of *SmUGT89B* and relative rutin content in leaves of *SmUGT89B*-overexpressed eggplants. pC1300 were the plants whose leaves were injected with *Agrobacterium* harboring an empty vector. mean ± SE (n = 3). Student's t test was used to indicate significant differences compared to those in the empty vector control (**P < 0.01).

It appears that the members of new phylogenetic groups are related to the glycosylation of secondary compounds in different plants. Group R was absent in all Brassicales but present in Solanaceae species. which was probably responsible for the distant relationship between them. According to the phylogenetic evolutionary tree, five *SmUGT* genes with two *Arabidopsis* genes (*AtUGT80A2* and *AtUGT80B1*) were classified into a separate out group (OG) (Figure 2A), in which the PSPG-box was less conserved. Members of the OG group have been proposed to modify sterols and lipids in other plants (Paquette et al., 2003). It is likely that the five genes may be involved in the formation of sterols and lipids in eggplant.

The members defined in the same phylogenetic subgroup generally possess similar functions (Soler et al., 2015; Song et al., 2020). Some *UGT* genes belonging to Groups B, D, E and L have been identified. *AtUGT73C6* (Group D) can catalyze the transfer of glucose from UDP-glucose to kaempferol-3-O-rhamnoside and quercetin-3-O-rhamnoside (Jones et al., 2003). *FaGT6*, which belongs to the UGT71 family, catalyzes quercetin into 3-O-glucoside (Griesser et al., 2008). UGT75 family members are mostly flavonoid-5-O-glycosyltransferases, which are mainly involved in the synthesis of plant anthocyanins (Noguchi et al., 2009; He et al., 2015). Therefore, the members from Groups B, D, E and L might possess the underlying function of participating in flavonol biosynthesis.

Based on the sequence analysis of a multigene family, its gene expression compared with that in the expression profiles of known genes and complementation of the metabolite pattern in combination with reverse genetics are conducive to understanding the underlying function of a single member from the superfamily (Offen et al., 2006). The expression patterns of SmUGT members from Clusters 7 to 10 exhibited good agreement with the transcription levels of rutin biosynthetic genes in eggplant (Figure 4B). Thus, the members of Clusters 7 to 10 acted as novel candidates for rutin biosynthesis. *SmUGT89B2* found in Cluster 7 was classified into Group B, as well as *AtUGT89C1*. *AtUGT89C1* has been proposed to be involved in flavonol biosynthesis in *Arabidopsis* (Yonekura-Sakakibara et al., 2007). It seems that SmUGT89B2 may be involved in flavonol biosynthesis in eggplant. The other four UGT genes observed in RNA-Seq data with high RPKM values, including *SmUGT71BA3* and *SmUGT72B88* (Group E), *SmUGT73A69* from Group D and *SmUGT75L62* from Group L, which were found in Cluster 9, were selected for qRT–PCR assays, as well as *SmUGT89B2* (Figure 4B).

The expression pattern of these five genes was unlike the rutin accumulation pattern at different developmental stages of fruit in eggplant (Figure 5A). The rutin content was the highest at the young fruit stage. It seems that plants can activate self-protection processes via the production of secondary metabolites against undesirable circumstances, such as UV irradiation, cold stress, and desiccation stress (Suzuki et al., 2005; Yang et al., 2012). This defense reaction was previously reported in anthocyanin accumulation (Naing and Kim, 2021). A large number of secondary metabolites are needed to protect plants, especially in tender tissues. The expression of *SmUGT89B2* was similar to the

accumulation pattern of rutin in eggplant (Figure 5). Therefore, *SmUGT89B2* served as a novel candidate gene for rutin biosynthesis in this study. The subcellular localization results showed that SmUGT89B2 was localized in the nucleus and cell membrane, as well as in the cytoplasm (Figure 6B). Most *UGT* genes have multiple subcellular localizations to control where specific pathway products accumulate (Yi et al., 2020; Daryanavard et al., 2023). Transient overexpression and silencing expression of *SmUGT89B2* in eggplant indicated that *SmUGT89B2* could affect rutin accumulation (Figure 7). It was likely that SmUGT89B2 encoded the final enzyme in the rutin biosynthetic pathway. Although the functions of UGT89B2 involved in rutin biosynthesis in eggplants *in vivo* have been identified in this study, the enzyme activity and biochemical characteristics of UGT89B2 should be further identified. Which step of SmUGT89B2 involved in two glycosylation steps of quercetin transform to rutin also need to be verified by catalytic experiments. But the exploration of this catalytic experiment is certainly the most difficult because UGT glycosyltransferase possesses substrate diversity. In the future, enzymatic tests of the SmUGT89B2 will be the key works for our studies.

Conclusion

Overall, all *UGT* genes from the eggplant genome through the HMM model and PSPG-box motif were searched in the study. A total of 195 *SmUGT* genes were identified, which were divided into 17 phylogenetic subgroups according to the classification of *Arabidopsis*, maize and rice. The conserved motifs and gene structures of candidate *UGT* genes were analyzed. According to the expression profile analysis, Clusters 7-10 served as candidate clusters, in which the five genes *SmUGT71BA3*, *SmUGT72B88*, *SmUGT73A69*, *SmUGT75L62* and *SmUGT89B2* were selected as candidate genes. However, among these selected genes, *SmUGT89B* had a higher transcription level in eggplant, and its expression pattern was similar to the accumulation trend in eggplant by qRT–PCR assay. VIGS and transient overexpression demonstrated that *SmUGT89B2* could positively promote rutin biosynthesis in eggplant.

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

YG: Data curation, Formal analysis, Writing – original draft, Visualization, Writing – review & editing. BY: Writing – original draft. RL: Formal analysis, Visualization, Writing – original draft. BS: Data curation, Validation, Writing – original draft. YL: Data curation, Validation, Writing – original draft, Formal analysis. YZ: Writing – original draft. ZQ: Data curation, Project administration, Writing – original draft. SY: Project administration, Writing – original draft, Writing – review & editing. BC: Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Key R&D Projects in Guangdong Province (2022B0202080003), Seed Industry Revitalisation Project of Guangdong (2022-NPY-00-026), and Vegetable Industry System Innovation Team Project of Guangdong (2022KJ110).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1310080/full#supplementary-material

References

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). Meme suite: tools for motif discovery and searching. *Nucleic. Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The meme suite. *Nucleic. Acids Res.* 43 (W1), W39–W49. doi: 10.1093/nar/gkv416

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* 9 (1), 11769. doi: 10.1038/s41598-019-47985-w

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., et al. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant. J.* 107 (2), 579–596. doi: 10.1111/tpj.15313

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases

revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* 69 (6), 1030–1042. doi: 10.1111/j.1365-313X.2011.04853.x

Carbonell-Bejerano, P., Diago, M. P., Martinez-Abaigar, J., Martinez-Zapater, J. M., Tardaguila, J., and Nunez-Olivera, E. (2014). Solar ultraviolet radiation is necessary to enhance grapevine fruit ripening transcriptional and phenolic responses. *BMC Plant Biol.* 14, 183. doi: 10.1186/1471-2229-14-183

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 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). Weblogo: a sequence logo generator. *Genome Res.* 14 (6), 1188–1190. doi: 10.1101/gr.849004

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 sinensi)*. *J. Exp. Bot.* 67 (8), 2285–2297. doi: 10.1093/jxb/erw053

Dai, X., Shi, X., Yang, C., Zhao, X., Zhuang, J., Liu, Y., et al. (2022). Two UDP-glycosyltransferases catalyze the biosynthesis of bitter flavonoid 7-O-neohesperidoside through sequential glycosylation in tea plants. *J. Agric. Food. Chem.* 70 (7), 2354–2365. doi: 10.1021/acs.jafc.1c07342

Daryanavard, H., Postiglione, A. E., Muhlemann, J. K., and Muday, G. K. (2023). Flavonols modulate plant development, signaling, and stress responses. *Curr. Opin. Plant Biol.* 72, 2350. doi: 10.1016/j.pbi.2023.102350

Dong, R., Yu, B., Yan, S., Qiu, Z., Lei, J., Chen, C., et al. (2020). Analysis of vitamin P content and inheritance models in eggplant. *Hortic. Plant J.* 6 (4), 240–246. doi: 10.1016/j.hpj.2020.05.005

Gachon, C. M. M., Langlois-Meurinne, M., and Saindrenan, P. (2005). Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* 10 (11), 542–549. doi: 10.1016/j.tplants.2005.09.007

Glasser, G., Graefe, E. U., Struck, F., Veit, M., and Gebhardt, R. (2002). Comparison of antioxidative capacities and inhibitory effects on cholesterol biosynthesis of quercetin and potential metabolites. *Phytomedicine* 9 (1), 33–40. doi: 10.1078/0944-7113-00080

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×ananassa*) achene and receptacle. *J. Exp. Bot.* 59 (10), 2611–2625. doi: 10.1093/jxb/ern117

He, F., Chen, W., Yu, K., Ji, X., Duan, C., Reeves, M. J., et al. (2015). Molecular and biochemical characterization of the UDP-glucose: anthocyanin 5-O-glucosyltransferase from *Vitis amurensis*. *Phytochemistry* 117, 363–372. doi: 10.1016/j.phytochem.2015.06.023

Hosseinzadeh, H., and Nassiri-Asl, M. (2014). Review of the protective effects of rutin on the metabolic function as an important dietary flavonoid. *J. Endocrinol. Invest.* 37 (9), 783–788. doi: 10.1007/s40618-014-0096-3

Huang, J., Pang, C., Fan, S., Song, M., Yu, J., Wei, H., et al. (2015). Genome-wide analysis of the family 1 glycosyltransferases in cotton. *Mol. Genet. Genomics* 290 (5), 1805–1818. doi: 10.1007/s00438-015-1040-8

Irmisch, S., Jo, S., Roach, C. R., Jancsik, S., Yuen, M. M. S., Madilao, L. L., et al. (2018). Discovery of UDP-glycosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite montbretin a. *Plant Cell.* 30 (8), 1864–1886. doi: 10.1105/tpc.18.00406

Jimenez-Aliaga, K., Bermejo-Bescos, P., Benedi, J., and Martin-Aragon, S. (2011). Quercetin and rutin exhibit antiamyloidogenic and fibril-disaggregating effects in *vitro* and potent antioxidant activity in APPswe cells. *Life Sci.* 89 (25-26), 939–945. doi: 10.1016/j.lfs.2011.09.023

Jing, R., Li, H. Q., Hu, C. L., Jiang, Y. P., Qin, L. P., and Zheng, C. J. (2016). Phytochemical and pharmacological profiles of three *Fagopyrum* buckwheats. *Int. J. Mol. Sci.* 17 (4), 589. doi: 10.3390/ijms17040589

Jones, P., Messner, B., Nakajima, J. I., Schaffner, A. R., and Saito, K. (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in arabidopsis thaliana. *J. Biol. Chem.* 278 (45), 43910–43918. doi: 10.1074/jbc.M303523200

Koja, E., Ohata, S., Maruyama, Y., Suzuki, H., Shimosaka, M., and Taguchi, G. (2018). Identification and characterization of a rhamnosyltransferase involved in rutin biosynthesis in *Fagopyrum esculentum* (common buckwheat). *Biosci. Biotechnol. Biochem.* 82 (10), 1790–1802. doi: 10.1080/09168451.2018.1491286

Kovinich, N., Saleem, A., Arnason, J. T., and Miki, B. (2010). Functional characterization of a UDP-glucose: flavonoid 3-O-glucosyltransferase from the seed coat of black soybean *(Glycine max* (L.) Merr.). *Phytochemistry* 71 (11-12), 1253–1263. doi: 10.1016/j.phytochem.2010.05.009

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 (6), 1547–1549. doi: 10.1093/molbev/msy096

Li, Y., Baldauf, S., Lim, E., and Bowles, D. J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.* 276 (6), 4338–4343. doi: 10.1074/jbc.M007447200

Li, Y., Li, P., Wang, Y., Dong, R., Yu, H., and Hou, B. (2014). Genome-wide identification and phylogenetic analysis of family-1 UDP glycosyltransferases in maize (*Zea mays*). *Planta* 239 (6), 1265–1279. doi: 10.1007/s00425-014-2050-1

Liu, R., Shu, B., Wang, Y., Yu, B., Wang, Y., Gan, Y., et al. (2023). Transcriptome analysis reveals key genes involved in the eggplant response to high-temperature stress. *Environ. Exp. Bot.* 211, 105369. doi: 10.1016/j.envexpbot.2023.105369

Luca, S. V., Macovei, I., Bujor, A., Miron, A., Skalicka-Woźniak, K., Aprotosoaie, A. C., et al. (2020). Bioactivity of dietary polyphenols: the role of metabolites. *Crit. Rev. Food. Sci. Nutr.* 60 (4), 626–659. doi: 10.1080/10408398.2018.1546669

Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., et al. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7 (4), 255–269. doi: 10.1097/00008571-199708000-00001

Martínez-Ispizua, E., Calatayud, Á., Marsal, J. I., Mateos-Fernández, R., Díez, M. J., Soler, S., et al. (2021). Phenotyping local eggplant varieties: commitment to biodiversity and nutritional quality preservation. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.696272

Mateo-Bonmati, E., Casanova-Saez, R., Simura, J., and Ljung, K. (2021). Broadening the roles of UDP-glycosyltransferases in auxin homeostasis and plant development. *New Phytol.* 232 (2), 642–654. doi: 10.1111/nph.17633

Mizohata, E., Okuda, T., Hatanaka, S., Nakayama, T., Horikawa, M., Nakayama, T., et al. (2013). Crystallization and preliminary X-ray crystallographic analysis of UDP-glucuronic acid: flavonol-3-O-glucuronosyltransferase (Vvgt5) from the grapevine *Vitis vinifera*. *Acta Crystallogr. F.* 69 (Pt 1), 65–68. doi: 10.1107/S1744309112045095

Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M., and Taguchi, G. (2014). Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (Buckwheat) cotyledon. *Plant J.* 80 (3), 437–448. doi: 10.1111/tpj.12645

Naing, A. H., and Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: their role in tolerance to abiotic stresses. *Physiol. Plant* 172 (3), 1711–1723. doi: 10.1111/ppl.13373

Noguchi, A., Horikawa, M., Fukui, Y., Fukuchi-Mizutani, M., Iuchi-Okada, A., Ishiguro, M., et al. (2009). Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in lamiales. *Plant Cell.* 21 (5), 1556–1572. doi: 10.1105/tpc.108.063826

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., et al. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO. J.* 25 (6), 1396–1405. doi: 10.1038/sj.emboj.7600970

Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., Imai, H., Takahashi, S., et al. (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). *Plant Cell.* 22 (8), 2856–2871. doi: 10.1105/tpc.110.074625

Paquette, S., Moller, B. L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62 (3), 399–413. doi: 10.1016/s0031-9422(02)00558-7

Rojas, R. F., Rodriguez, T. O., Murai, Y., Iwashina, T., Sugawara, S., Suzuki, M., et al. (2014). Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 → 6) rhamnosyltransferase. *Plant Mol. Biol.* 84 (3), 287–300. doi: 10.1007/s11103-013-0133-1

Ross, J., Li, Y., Lim, E., and Bowles, D. J. (2001). Higher plant glycosyltransferases. *Genome Biol.* 2 (2), REVIEWS3004. doi: 10.1186/gb-2001-2-2-reviews3004

Shen, N., Wang, T., Gan, Q., Liu, S., Wang, L., and Jin, B. (2022). Plant flavonoids: classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* 383, 132531. doi: 10.1016/j.foodchem.2022.132531

Singh, A. P., Wang, Y., Olson, R. M., Luthria, D., Banuelos, G. S., Pasakdee, S., et al. (2017). LC-MS-MS analysis and the antioxidant activity of flavonoids from eggplant skins grown in organic and conventional environments. *Food Nutr. Sci.* 08 (09), 873–888. doi: 10.4236/fns.2017.89063

Soler, M., Camargo, E. L., Carocha, V., Cassan-Wang, H., San, C. H., Savelli, B., et al. (2015). The eucalyptus grandis R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytol.* 206 (4), 1364–1377. doi: 10.1111/nph.13039

Song, J., Chen, C., Zhang, S., Wang, J., Huang, Z., Chen, M., et al. (2020). Systematic analysis of the capsicum ERF transcription factor family: identification of regulatory factors involved in the regulation of species-specific metabolites. *BMC Genomics* 21 (1), 573. doi: 10.1186/s12864-020-06983-3

Song, J., Sun, B., Chen, C., Ning, Z., Zhang, S., Cai, Y., et al. (2023). An R-R-type MYB transcription factor promotes non-climacteric pepper fruit carotenoid pigment biosynthesis. *Plant J.* 115 (3), 724–741. doi: 10.1111/tpj.16257

Sulli, M., Barchi, L., Toppino, L., Diretto, G., Sala, T., Lanteri, S., et al. (2021). An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.638195

Sun, B., Zhou, X., Chen, C., Chen, C., Chen, K., Chen, M., et al. (2020). Coexpression network analysis reveals an MYB transcriptional activator involved in capsaicinoid biosynthesis in hot peppers. *Hortic. Res.* 7 (1), 162. doi: 10.1038/s41438-020-00381-2

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 (5), 1303–1307. doi: 10.1016/j.plantsci.2005.01.007

Tobar-Delgado, E., Mejía-España, D., Osorio-Mora, O., and Serna-Cock, L. (2023). Rutin: family farming products' extraction sources, industrial applications and current

trends in biological activity protection. *Molecules* 28 (15), 5864. doi: 10.3390/molecules28155864

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3 (1), 2–20. doi: 10.1093/mp/ssp106

Wang, Y., Li, C., Yan, S., Yu, B., Gan, Y., Liu, R., et al. (2022). Genome-wide analysis and characterization of eggplant F-box gene superfamily: gene evolution and expression analysis under stress. *Int. J. Mol. Sci.* 23 (24), 16049. doi: 10.3390/ijms232416049

Wilson, A. E., and Tian, L. (2019). Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. *Plant J.* 100 (6), 1273–1288. doi: 10.1111/tpj.14514

Wilson, A. E., Wu, S., and Tian, L. (2019). PgUGT95B2 preferentially metabolizes flavones / flavonols and has evolved independently from flavone / flavonol UGTs identified in *Arabidopsis thaliana*. *Phytochemistry* 157, 184–193. doi: 10.1016/j.phytochem.2018.10.025

Wu, B., Gao, L., Gao, J., Xu, Y., Liu, H., Cao, X., et al. (2017). Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00389

Wu, B., Liu, X., Xu, K., and Zhang, B. (2020). Genome-wide characterization, evolution and expression profiling of UDP-glycosyltransferase family in pomelo (*Citrus grandis*) fruit. *BMC Plant Biol.* 20 (1), 459. doi: 10.1186/s12870-020-02655-2

Xiao, X., Lu, Q., Liu, R., Gong, J., Gong, W., Liu, A., et al. (2019). Genome-wide characterization of the UDP-glycosyltransferase gene family in upland cotton. *3 Biotech.* 9 (12), 453. doi: 10.1007/s13205-019-1984-1

Yang, C., Fang, X., Wu, X., Mao, Y., Wang, L., and Chen, X. (2012). Transcriptional regulation of plant secondary metabolism f. *J. Integr. Plant Biol.* 54 (10), 703–712. doi: 10.1111/j.1744-7909.2012.01161.x

Yang, L., Kang, Y., Liu, J., Li, N., Sun, H., Ao, T., et al. (2022). Foliar spray with rutin improves cadmium remediation efficiency excellently by enhancing antioxidation and phytochelatin detoxification of *Amaranthus hypochondriacus*. *Int. J. Phytoremediat.* 24 (10), 1060–1070. doi: 10.1080/15226514.2021.1999902

Yang, Q., Zhang, Y., Qu, X., Wu, F., Li, X., Ren, M., et al. (2023). Genome-wide analysis of UDP-glycosyltransferases family and identification of UDP genes involved in abiotic stress and flavonol biosynthesis in *Nicotiana tabacum*. *BMC Plant Biol.* 23 (1), 204. doi: 10.1186/s12870-023-04208-9

Yi, S., Kuang, T., Miao, Y., Xu, Y., Wang, Z., Dong, L., et al. (2020). Discovery and characterization of four glycosyltransferases involved in anthraquinone glycoside biosynthesis in *Rubia yunnanensis*. *Org. Chem. Front.* 7 (17), 2442–2448. doi: 10.1039/D0QO00579G

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., Di, S., Fan, C., Chang, Z., and Pang, Y. (2017). Genome-wide identification and functional characterization of UDP-glucosyltransferase genes involved in flavonoid biosynthesis in glycine max. *Plant Cell Physiol.* 58 (9), 1558–1572. doi: 10.1093/pcp/pcx081

Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in arabidopsis by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* 282 (20), 14932–14941. doi: 10.1074/jbc.M611498200

Yu, J., Hu, F., Dossa, K., Wang, Z., and Ke, T. (2017). Genome-wide analysis of UDP-glycosyltransferase super family in *Brassica rapa* and *Brassica oleracea* reveals its evolutionary history and functional characterization. *BMC Genomics* 18 (1), 474. doi: 10.1186/s12864-017-3844-x

Zhang, L., Li, X., Ma, B., Gao, Q., Du, H., Han, Y., et al. (2017). The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol. Plant* 10 (9), 1224–1237. doi: 10.1016/j.molp.2017.08.013

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 January 2024
DOI 10.3389/fpls.2024.1302817

OPEN ACCESS

EDITED BY
Xiang Pu,
Sichuan Agricultural University, China

REVIEWED BY
Jing Zhou,
Hainan University, China
Xiao Ding,
Chinese Academy of Sciences (CAS), China
Puja Ray,
Presidency University, India

*CORRESPONDENCE
Jishuang Chen
bihu_zmu@zmu.edu.cn
Jingshan Shi
shijs@zmu.edu.cn

†These authors share first authorship

RECEIVED 27 September 2023
ACCEPTED 12 January 2024
PUBLISHED 29 January 2024

CITATION
Sarsaiya S, Jain A, Shu F, Yang M, Pu M, Jia Q, Gong Q, Wu Q, Qian X, Shi J and Chen J (2024) Enhancing dendrobine production in *Dendrobium nobile* through mono-culturing of endophytic fungi, *Trichoderma longibrachiatum* (MD33) in a temporary immersion bioreactor system. *Front. Plant Sci.* 15:1302817. doi: 10.3389/fpls.2024.1302817

Enhancing dendrobine production in *Dendrobium nobile* through mono-culturing of endophytic fungi, *Trichoderma longibrachiatum* (MD33) in a temporary immersion bioreactor system

Surendra Sarsaiya[1,2†], Archana Jain[1†], Fuxing Shu[2,3], Mingfa Yang[2], Mengxuan Pu[2], Qi Jia[2], Qihai Gong[1], Qin Wu[1], Xu Qian[3], Jingshan Shi[1*] and Jishuang Chen[1,2,3*]

[1]Key Laboratory of Basic Pharmacology and Joint International Research Laboratory of Ethnomedicine of Ministry of Education, Zunyi Medical University, Zunyi, China, [2]Bioresource Institute for Healthy Utilization (BIHU), Zunyi Medical University, Zunyi, China, [3]College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing, China

Introduction: Dendrobine, a valuable alkaloid found in *Dendrobium nobile*, possesses significant pharmaceutical potential.

Methods: In this study, we explored innovative approaches to enhance dendrobine production by utilizing endophytic fungi in a Temporary Immersion Bioreactor System (TIBS, Nanjing BioFunction Co. Ltd., China) and traditional test bottles. Dendrobine was unequivocally identified and characterised in *D. nobile* co-culture seedlings through UHPLC analysis and LC-MS qTOF analysis, supported by reference standards.

Results: The CGTB (control group) and EGTB (experimental group) 12-month-old *D. nobile* seedlings exhibited similar peak retention times at 7.6±0.1 minutes, with dendrobine identified as $C_{16}H_{25}NO_2$ (molecular weight 264.195). The EGTB, co-cultured with *Trichoderma longibrachiatum* (MD33), displayed a 2.6-fold dendrobine increase (1804.23 ng/ml) compared to the CGTB (685.95 ng/ml). Furthermore, a bioanalytical approach was applied to investigate the mono-culture of *T. longibrachiatum* MD33 with or without *D. nobile* seedlings in test bottles. The newly developed UHPLC-MS method allowed for dendrobine identification at a retention time of 7.6±0.1 minutes for control and 7.6±0.1 minutes for co-culture. Additionally, we explored TIBS to enhance dendrobine production. Co-culturing *D. nobile* seedlings with *Trichoderma longibrachiatum* (MD33) in the TIBS system led to a substantial 9.7-fold dendrobine increase (4415.77 ng/ml) compared to the control (454.01 ng/ml) after just 7 days. The comparative analysis of dendrobine concentration between EGTB and EGTIBS highlighted the remarkable potential of TIBS for optimizing dendrobine production. Future research may focus on scaling up the TIBS approach for commercial dendrobine production and investigating the underlying

mechanisms for enhanced dendrobine biosynthesis in *D. nobile*. The structural elucidation of dendrobine was achieved through ^{1}H and ^{13}C NMR spectroscopy, revealing a complex array of proton environments and distinct carbon environments, providing essential insights for the comprehensive characterization of the compound.

Discussion: These findings hold promise for pharmaceutical and industrial applications of dendrobine and underline the role of endophytic fungi in enhancing secondary metabolite production in medicinal plants.

KEYWORDS

dendrobine, *Dendrobium nobile*, endophytic fungi, temporary immersion bioreactor system (TIBS), alkaloid production

Introduction

Dendrobium nobile, an enduring epiphytic herb belonging to the Orchidaceae family, holds a significant place in traditional Chinese medicine. This remarkable plant is primarily found in regions south of the Yangtze River, encompassing Guizhou, Yunnan, Guangxi, and other sub-tropical areas. With a global diversity of over 25,000 plant species within the *Dendrobium* genus (Nie et al., 2020; Sarsaiya et al., 2020a; Sarsaiya et al., 2020b), *Dendrobium nobile* stands out for its invaluable contributions to the history of drug development. Among its many compounds, dendrobine is a noteworthy alkaloid found in *Dendrobium nobile*, typically comprising about 0.5% of its total weight. This colorless solid, reminiscent of the picrotoxin family of natural products, has thus far been exclusively sourced from *D. nobile* and through chemical synthesis (Sarsaiya et al., 2020b). Over the decades, dendrobine has been harnessed for its therapeutic potential in addressing a range of health concerns. It has been explored in the treatment of conditions such as tumors, hyperglycemia, and hyperlipidemia and has displayed promising effects in mitigating metastatic cancer. Additionally, it has shown potential for addressing ailments associated with ageing, including Alzheimer's disease and nervous system disorders (Nie et al., 2020; Sarsaiya et al., 2020b). However, the production of dendrobine by the *D. nobile* plant species is limited due to its slow growth rate. Consequently, the quantity of dendrobine derived from *D. nobile* remains insufficient to meet the current demands of both the industrial and research sectors. Robust and thriving plants host intricate and diverse microbial ecosystems that wield significant influence over the physiology and growth of these plants. In parallel, the variety within these plant-associated microbial populations is intricately intertwined with external environmental conditions. Beyond their primary role in nourishing plants, plant metabolites serve multifaceted functions, encompassing defence mechanisms against pathogens, pests, and herbivores (Zhao et al., 2023).

Most of the previous research pertaining to orchid-fungus associations has primarily focused on the fungal dynamics within terrestrial and temperate orchids. However, the intricate interplay between fungi and tropical orchids, exemplified by *D. nobile*, extends beyond the norm. Diverse fungal taxa inhabit various parts of *D. nobile*, encompassing roots, stems, buds, and leaves (Chen et al., 2023a). This rich fungal community includes mutualistic mycorrhizal partners as well as endophytic fungi, alongside a notably varied assembly of nonmycorrhizal fungal associates. Despite this breadth of association, the precise role these plant-associated fungi play in stimulating phytochemical production remains enigmatic (Sarsaiya et al., 2020a). In addition to their intriguing ecological roles, endophytic fungi confer substantial benefits upon their host, *D. nobile*. These advantages range from assisting the plant in coping with both biotic and abiotic stresses to strengthening its resistance against pathogens. This was accomplished through the secretion of secondary metabolites boasting antimicrobial properties and various phytohormones, as elucidated by Hajong and Kapoor (2020). Existing exploration of Dendrobium myco-endophytes (DMEs) has predominantly centered on *in vitro* symbiotic practices, utilizing fungal endophyte strains isolated from fully matured Dendrobium roots, buds, stems, and leaves. Unravelling the nature of these endophytic fungi—whether they assume pathogenic, conditionally pathogenic, or non-pathogenic roles for the host plant—carries significant importance (Sarsaiya et al., 2019). Nonetheless, the scope of orchid seed germination in its natural habitat likely entails a more intricate tapestry of interactions, potentially involving a broader array of non-mycorrhizal fungi and even other organism types, as posited by Salazar et al. (2020) and Hernández-Ramírez et al. (2023).

In the realm of plant-fungal co-culturing, a well-recognized phenomenon is the dormancy of numerous gene clusters. These silent gene clusters encode secondary metabolites that, under conventional laboratory culture conditions, remain unexpressed

and, consequently, unproduced. To unlock the potential of these hidden metabolites, researchers have devised strategies to induce their expression. One such promising method involves co-culturing endophytic fungi with plants in the same growth medium. This intriguing approach capitalizes on the synergy between plant and fungal species, potentially leading to the activation of gene clusters and the subsequent production of metabolites (Chen et al., 2023b). In the natural world, microorganisms coexist within intricate communities, engaging in a myriad of interactions. Central to these interactions are the secondary metabolites that serve as chemical signals for communication or as tools in the competition for limited resources. These metabolites also play a pivotal role in fortifying defence mechanisms. Thus, emulating the complexity of the natural environment through the mixed fermentation of different microorganisms, often referred to as co-cultivation or co-culture, holds the promise of amplifying compound production. Moreover, this co-cultivation strategy may catalyze the awakening of dormant biosynthetic pathways, ultimately resulting in the accumulation of novel natural products (Akone et al., 2016; Gegenbauer et al., 2023).

To overcome these limitations, researchers have explored the possibility of cultivating microorganisms under conditions that mimic the complex communities found in nature, commonly referred to as the "microbiome." Within these microbial communities, bioactive secondary metabolites assume a pivotal role in mediating interactions among various microorganisms and may have a direct influence on the production of phytochemicals in plants (Vinale et al., 2017; ZhiPing et al., 2023). Although previous studies have explored how *D. nobile* and endophytic fungi interact in terms of acquiring nutrients, promoting growth, and providing protection against pathogens, the specific contribution of these fungi in co-cultivation to boost the production of natural compounds has not been extensively investigated. In this study, we undertook the collection and analysis of *D. nobile* seedlings to quantify dendrobine concentrations. We conducted this analysis using LC-MS chromatography for two groups: the control group, consisting of *D. nobile* seedlings grown under standard conditions, and the experimental group, involving the co-culturing of *D. nobile* with *Trichoderma longibrachiatum* MD33. Importantly, this co-culturing was conducted within a temporary immersion bioreactor system. Our study specifically investigated the impact of *T. longibrachiatum* MD33 on the accumulation of dendrobine in *D. nobile* under controlled experimental conditions.

Materials and methods

Test fungal strain and culture conditions

Trichoderma longibrachiatum MD33 (hereafter referred to as MD33) was the fungus strain used in this study. It was found on healthy wild *D. nobile* plants collected in Jinshishi, Chishui, Guizhou, China (Sarsaiya et al., 2020a; Sarsaiya et al., 2020b). The strain MD33 has been duly documented and is registered in the NCBI Library under the accession number MN826683. For cultivation purposes, MD33 was nurtured in PD (potato dextrose) broth, maintaining a pH level of 6.5. The cultivation process occurred at a temperature of 28°C over a span of 120 hours, with constant agitation set at 120 revolutions per minute (rpm) (MQD-B2NRG, Shanghai Yuquan Instrument Co. Ltd., Shaghai, China).

Sample collection and its processing

Sample collection and processing involved the acquisition of *D. nobile* seedlings (8 months old) from the Bioresource Institute for Healthy Utilization (BIHU) at Zunyi Medical University (ZMU) in Zunyi, Guizhou, China. Prior to experimentation, these samples were authenticated to ensure their identity (Zhang et al., 2015; Hu et al., 2016; Yu et al., 2017; Cao et al., 2018). To serve as a growth medium for co-culturing experiments, water agar (0.65% w/v) was used and dispensed into glass bottles. Notably, no supplementary nutrients were introduced into the medium; it consisted solely of double-distilled water and agar. After thorough sterilization of the glass bottles containing the growth medium, the *D. nobile* seedlings were introduced into individual bottles, all under meticulously aseptic conditions.

Co-culturing endophytic fungi on *D. nobile* seedlings in test bottles

To conduct experiments quantifying dendrobine, we utilized the previously identified endophytic fungal mycelium, *T. longibrachiatum* (MD33) (Sarsaiya et al., 2020a and Sarsaiya et al., 2020b). We processed the contents of endophytic fungal flask cultures to isolate the mycelium suspension. Subsequently, this mycelium suspension was concentrated and introduced into incisions made at the terminal leaves, where they attach to the stem. In parallel, for seedling co-culturing experiments, a sterilized cotton plug was used as a control to establish a baseline reference. To serve as a control, we established triplicate samples under identical conditions but without the addition of endophytic fungal culture. Following the inoculation process, the test co-culturing bottles were relocated to a plant growth light incubator instrument (MGC-300B, Shanghai Yiheng Scientific Instruments Co. Ltd., Shanghai, China), where they were maintained at a temperature of 25°C under a light/dark cycle of 14 hours of light and 10 hours of darkness, with a light intensity of 2000 lux. The evaluation of all test bottles took place after 21 days.

Co-culturing endophytic fungi with *D. nobile* seedlings in a temporary immersion bioreactor system

To facilitate the co-culturing of endophytic fungi *T. longibrachiatum* (MD33) with *D. nobile* seedlings, approximately 300 seedlings were evenly distributed into each temporary immersion bioreactor. These seedlings were placed in a modified ½ MS culture medium supplemented with ½ MS (2.47 g/L), sugar (30.0 g/L), NAA (0.5 mg/L), and potato extract (30 g/L).

The pH of the medium was maintained between 5.8 to 6.0. These bioreactors were then positioned in a well-light growth chamber rack at a constant temperature of 25 ± 2°C with an 11-hour light cycle and a light intensity of 2000 lux. To regulate the immersion of the medium with the seedlings, all bioreactors were connected to the BioF II machine (TIBS, Nanjing BioFunction Co. Ltd., China). The machine controlled the air supply in cycles of 300 seconds, repeated four times with intervals of six hours throughout the day. The experiment was divided into two distinct sets: Set A, where *T. longibrachiatum* (MD33) was introduced, and Set B, which served as the control group without any addition of endophytic fungi. Each set consisted of triplicate bioreactors containing *D. nobile* seedlings. In Set A, approximately 0.25 ± 0.1 grams of wet mycelial growth from pure cultures were introduced into the bioreactors under aseptic conditions. This facilitated the interaction between *T. longibrachiatum* (MD33) mycelium and the *D. nobile* seedlings. In Set B, no endophytic fungal cultures were added to the bioreactors. The interactions between *D. nobile* seedlings and endophytic fungi were observed at seven-day intervals. This included monitoring for mortality rates and changes in the coloration of the *D. nobile* seedlings. After a 21-day cultivation period, the *D. nobile* seedlings were carefully extracted, washed with distilled water, and dissected to separate the stems and leaves for subsequent histopathological analysis. The remaining *D. nobile* seedlings were prepared for dendrobine concentration quantification using the LC-MS technique.

Extraction of the reference plant dendrobine

To extract dendrobine alkaloids, oven-dried *D. nobile* stem powder was utilized. The dried powder was combined with absolute ethanol and subjected to boiling at 90°C for a duration of 2 hours. Following the extraction process, the mixture was carefully filtered through Whatman filter paper. The filtered sample, obtained through a 0.22 µm filter, was transferred to sample tubes for LC-MS analysis. This analysis involved comparing the peak retention time of the extracted material with a reference standard of dendrobine. The standard dendrobine used was sourced from Chengdu DeSiTe Biological Technology Co., Ltd., located in Chengdu, China, and possessed a purity level exceeding 99%.

Reference standard dendrobine solution preparation

To create the reference standard dendrobine solution, an initial stock solution was formulated with a concentration of 20.0 µg/ml in chloroform. Dendrobine working standard solutions were prepared at concentrations of 20, 40, 60, 80, and 100 ng/ml from the stock solution. This was achieved through precise dilution of the dendrobine stock solution, using chloroform as the diluent. The dendrobine standard utilized in this procedure was procured from Chengdu DeSiTe Biological Technology Co., Ltd., located in Chengdu, China, and boasted a purity level exceeding 99%.

Quantification of dendrobine concentration through LC-MS

For the LC-MS analysis, the *D. nobile* seedlings from each bioreactor were subjected to a meticulous sample preparation process. The seedlings were first freeze-dried at a temperature of -40°C for a duration of 10 hours. This freeze-drying process was carried out using a ModulyoD-230 Freeze Dryer from Thermo Electron Corporation, based in Milford, MA. Subsequently, 100 mg of the freeze-dried *D. nobile* seedling parts were finely powdered in a mortar and pestle. The powdered material was then soaked in 50 ml of chloroform for a period of 12 hours while being agitated at 120 rpm. Following this soaking process, 40 ml of the liquid portion was carefully separated from the mixture. The chloroform phase was isolated from the aqueous phase and subjected to evaporation at a controlled temperature of 35°C using a rotary evaporator. This process resulted in the formation of a residual substance. To prepare the sample for analysis, the remaining residue was re-dissolved in 5 ml of chloroform. Prior to the analysis, this solution was meticulously filtered through a 0.22 µ m filter to ensure clarity and purity.

The detection of dendrobine was carried out using an advanced UHPLC system, specifically the Thermo Scientific Dionex UltiMate 3000 from the United States. The system was equipped with a column measuring 150 x 2.1 mm and featuring particles with a size of 1.9 µm. For the mobile phase, a composition of 0.1% formic acid and acetonitrile was employed in a ratio of 95:5 (v/v). The flow rate was set at 0.3 ml/min. Additionally, sheath gas flowed at a rate of 35, auxiliary gas at a rate of 15, and the spray voltage was set at 3.5 kV. The capillary temperature was maintained at 350°C, while the auxiliary gas heater temperature was set at 400°C. During the analysis, a 10 µL volume of the prepared sample was injected. The identification of dendrobine was achieved by comparing retention times, molecular weight (264.195), and LC-MS fragmentation patterns with an authentic chemical reference standard. The dendrobine standard used in this analysis was purchased from Chengdu DeSiTe Biological Technology Co. Ltd. in China, and it possessed a purity level exceeding 99%. To quantify the concentration of dendrobine, a range of different dendrobine concentrations were meticulously prepared. These served as the basis for constructing a standard graph, enabling the accurate measurement of dendrobine concentration in the samples.

Structural analysis using nuclear magnetic resonance spectroscopy

The TIBS co-cultured *D. nobile* seedling samples were used to study the structure of dendrobine using NMR spectroscopy. The compound under investigation was dissolved in deuterated solvents, including CD_3OD (δ_H 3.31, δ_C 49.00), to obtain clear and homogeneous solutions for NMR spectral tests. NMR spectra were acquired on an Agilent DD2400-MR spectrometer, operating at a proton frequency of 400 MHz and a carbon frequency of 101 MHz. The instrument, sourced from Agilent Technologies (Santa Clara, CA), was equipped with state-of-the-

art technology to ensure high-resolution spectra. Proton NMR spectra were recorded using a 7 mm NMR tube with the samples dissolved in deuterated solvents. Carbon NMR spectra were acquired using the same samples in 7 mm NMR tubes. The acquired NMR data were processed using Agilent's proprietary software, and spectra were analyzed using MestReNova (version 14.2.1-27684) to extract relevant information about the chemical structure of dendrobine.

Statistical analysis

Every experiment conducted in this study was carried out in triplicate, ensuring the reliability and robustness of the results. The data obtained from these experiments were presented as the mean value accompanied by the standard deviation (SD), calculated from the triplicate values. For statistical analysis, Duncan's multiple range test (DMRT) was employed. This statistical test helps assess significant differences among multiple groups. The entire statistical analysis was performed using the SPSS V16.0 statistical package, a well-established software program developed by SPSS Inc. in Chicago, USA.

Results

Quantification of dendrobine in *D. nobile* seedlings using test bottles

Dendrobine quantification in *D. nobile* seedlings using test bottles revealed the presence of picrotoxan-type sesquiterpenoids. Dendrobine alkaloids were identified and characterized through unambiguous comparisons with reference standards, incorporating retention time, accurate mass, and MS/MS fragment ions (Figures 1, 2). The chemical composition and details, including retention time and precursor ions, are summarized in Table 1, which presents the induced phytodendrobine yield under different conditions in the control (CGTB) and experimental (EGTB) groups. Both groups, utilizing 8-month-old *D. nobile* seedlings, showed similar peak retention times (7.6 ± 0.1 minutes), with dendrobine identified as $C_{16}H_{25}NO_2$ and a molecular weight of 264.195. A bioanalytical approach was used to investigate the effect of cultivating mono-culture of MD33 with and without *D. nobile* seedlings on dendrobine production in solid media test bottles. The UHPLC-MS method, applied to chloroform extracts from *D. nobile* seedlings, illustrated a peak intensity (7.6 ± 0.1 minutes) between the control and co-culture sets (Figure 3). Confirmation through LC-MS analysis in positive ion mode, compared with known chemical reference dendrobine, revealed a substantial increase in dendrobine accumulation when MD33 was co-cultured with *D. nobile* seedlings. Co-culture test bottles demonstrated a remarkable 2.6-fold increase in average dendrobine production (1804.23 ng/ml) compared to the control group (685.95 ng/ml) (Table 1). The mass of dendrobine in both groups was 264.195, corresponding to the molecular formula $C_{16}H_{25}NO_2$. *D. nobile* seedlings with dendrobine were identified by comparing the peak retention time with that of the chemical reference dendrobine.

Quantification of dendrobine in *D. nobile* seedlings using a temporary immersion bioreactor system

We investigated utilizing a temporary immersion bioreactor system (TIBS) to increase dendrobine accumulation in *D. nobile* by co-culturing with MD33. MD33 generally aided *D. nobile* seedling growth. In addition, the experimental and control groups exhibited significantly different dendrobine levels. After 7 incubation days, the experimental group had significantly more dendrobine (Table 2; Figures 4, 5) than the control group. In particular, *D. nobile* phytodendrobine accumulation increased 9.7-fold in the experimental group to 4415.77 ng/ml, compared to 454.01 ng/ml

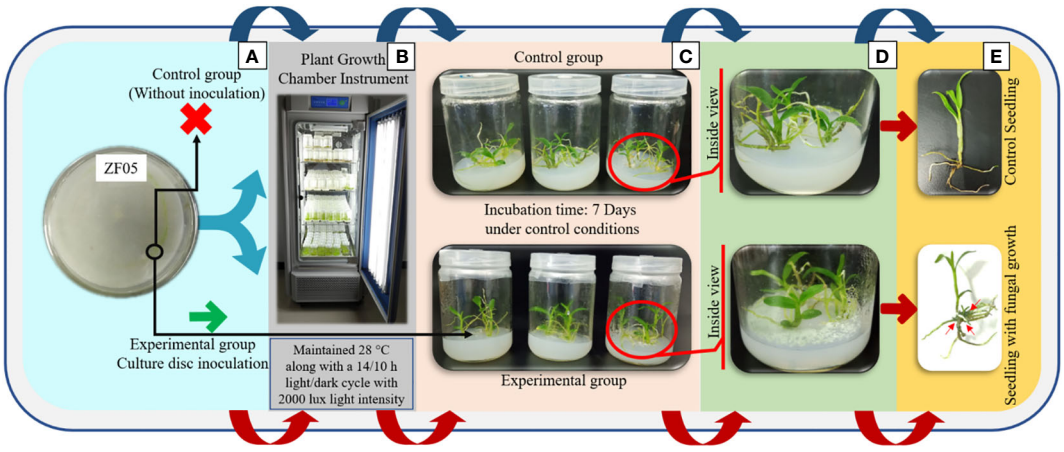


FIGURE 1
Co-culturing of *Trichoderma longibrachiatum* MD33 with *D nobile* seedlings for high yield of dendrobine concentration. **(A)** Pure culture of *Trichoderma longibrachiatum* MD33; **(B)** Plant growth chamber for maintain the growth conditions; **(C)** Test bottles after 7 days incubation; **(D)** Inside view of test bottles; **(E)** Observation of single seedling.

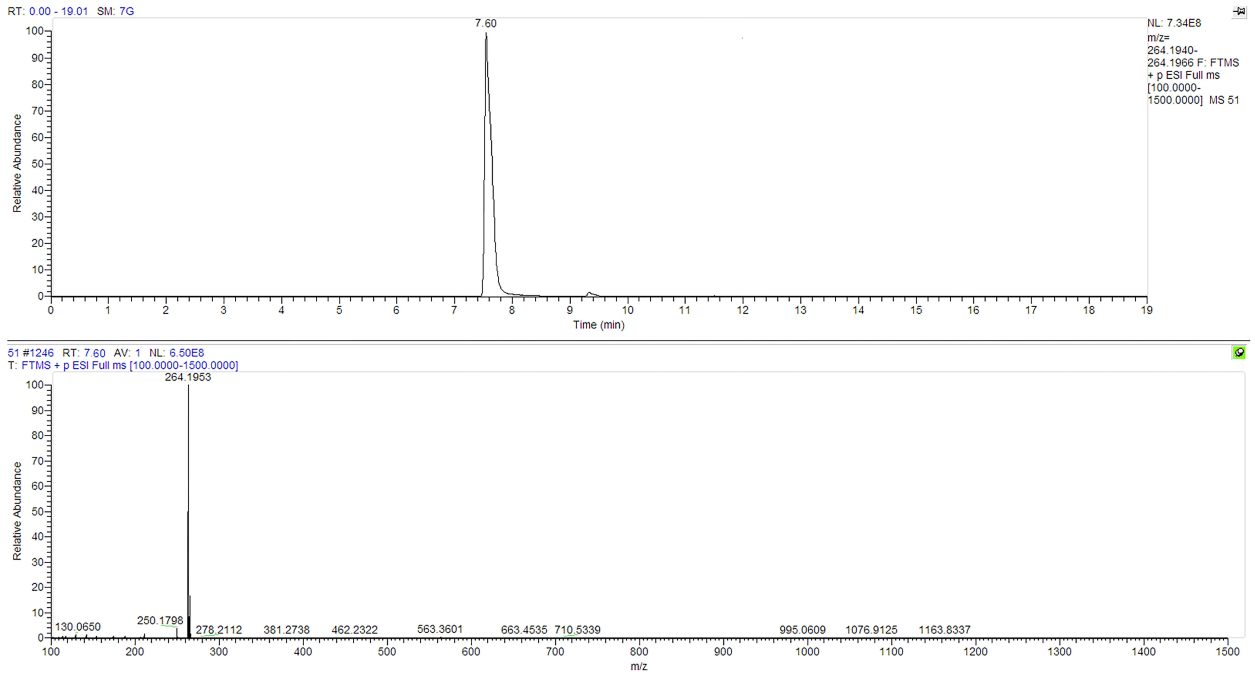


FIGURE 2
Chemical reference dendrobine (CRD) LC-MS Chromatogram for authentic reference dendrobine peak and molecular weight.

in the control group. Chromatography revealed a notable difference in the peak intensity between the MD33 fungal experimental group (13,085,216,134) and the control group (1,359,280,548). The retention times did not differ between the experimental and control groups. Mass spectrometry confirmed the presence of dendrobine in both groups, matching the molecular formula $C_{16}H_{25}NO_2$, with a precise molecular weight of 264.195 in both experimental and control groups. Additionally, the chemical reference dendrobine produced a mass of 264.195 Da with a distinctive peak, and a retention time of 7.6.

Comparative analysis of co-cultured experiments

Valuable insights into strategies for enhancing dendrobine production were revealed through a comparative analysis of dendrobine concentration in co-cultured *D. nobile* seedlings using an Experiment with Test Bottles (EGTB) and an Experiment with Temporary Immersion Bioreactor System (EGTIBS) (Table 3). Moreover, utilizing the prospective endophytic fungus MD33 to co-cultivate *D. nobile* seedlings in the EGTB experiment resulted in a noteworthy 2.6-fold elevation in dendrobine concentration, as confirmed by LC-MS analysis. The effectiveness of MD33 as a facilitator of dendrobine synthesis in *D. nobile* seedlings was demonstrated using this method. However, the outcomes of EGTIBS were even more significant. Dendrobine concentration in seedlings of *D. nobile* cultured in a Temporary Immersion Bioreactor System (TIBS) increased by 9.7-fold compared to that in the control group. The experimental group that utilized TIBS achieved a noteworthy dendrobine concentration of 4415.77 ng/ml, demonstrating the exceptional capacity of this cultivation technique to enhance dendrobine levels. The significant disparity in dendrobine concentrations between the two experimental methodologies

TABLE 1 Yield of induced *D. nobile* dendrobine in the test bottles.

Test Group	Plant material	SAM	Detection Technique	Peak retention time (min)	Chemical composition	Structural elucidation	Molecular weight (MW)	Dendrobine concentration (ng/ml)	Fold increased
CGTB	*D. nobile* seedlings	12	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	685.95	–
EGTB	*D. nobile* seedlings	12	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	1804.23	2.6*
CRD	–	–	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	–	–

CGTB, Control Group Test Bottles; EGTB, Experimental Group Test Bottles; CRD, Chemical reference dendrobine; SAM, Seedlings age in months; ng/ml, nanogram per millilitre; *as compared to CGTB.

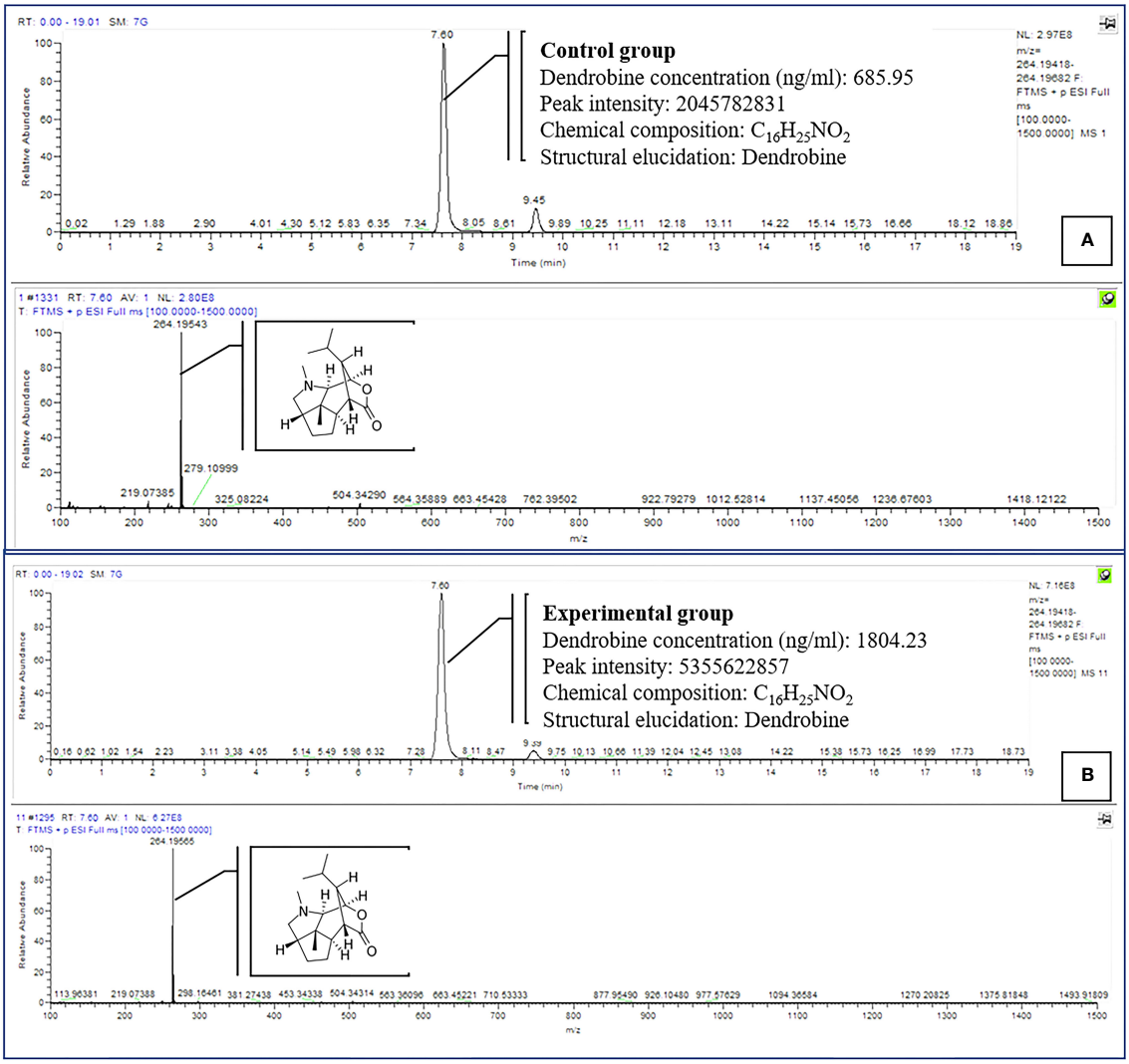


FIGURE 3
Comparison of dendrobine profiles produced by CGTB **(A)** and EGTB **(B)** *D. nobile* seedlings dendrobine. Tentative LCMS identification of the dendrobine is highlighted in the chromatogram (Molecular weight 264.195).

highlights the efficacy of the TIBS system for maximizing dendrobine production. Given its exceptional yield potential, future research avenues should consider expanding the TIBS method to commercial dendrobine production because it is particularly advantageous for pharmaceutical and industrial applications. Further investigations into the mechanisms governing the increased production of dendrobine in the TIBS system may yield significant knowledge for optimizing the biosynthesis of dendrobine in *D. nobile*.

TABLE 2 Yield of induced *D. nobile* dendrobine in the temporary immersion bioreactor system (TIBS).

Test Group	Plant material	Liquid medium	SAM	Detection Technique	Peak retention time (min)	Chemical composition	Structural elucidation	Molecular weight (MW)	Dendrobine concentration (ng/ml)	Fold increased
CGTIBS	*D. nobile* seedlings	Modified ½ MS	10	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	454.01	–
EGTIBS	*D. nobile* seedlings	medium	10	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	4415.77	9.7
CRD	–	–	–	UHPLC	7.6 ± 0.1	$C_{16}H_{25}NO_2$	Dendrobine	264.195	–	–

CGTIBS, Control Group; EGTIBS, Experimental Group; CRD, Chemical reference dendrobine; SAM, Seedlings age in months; ng/ml, nanogram per millilitre.

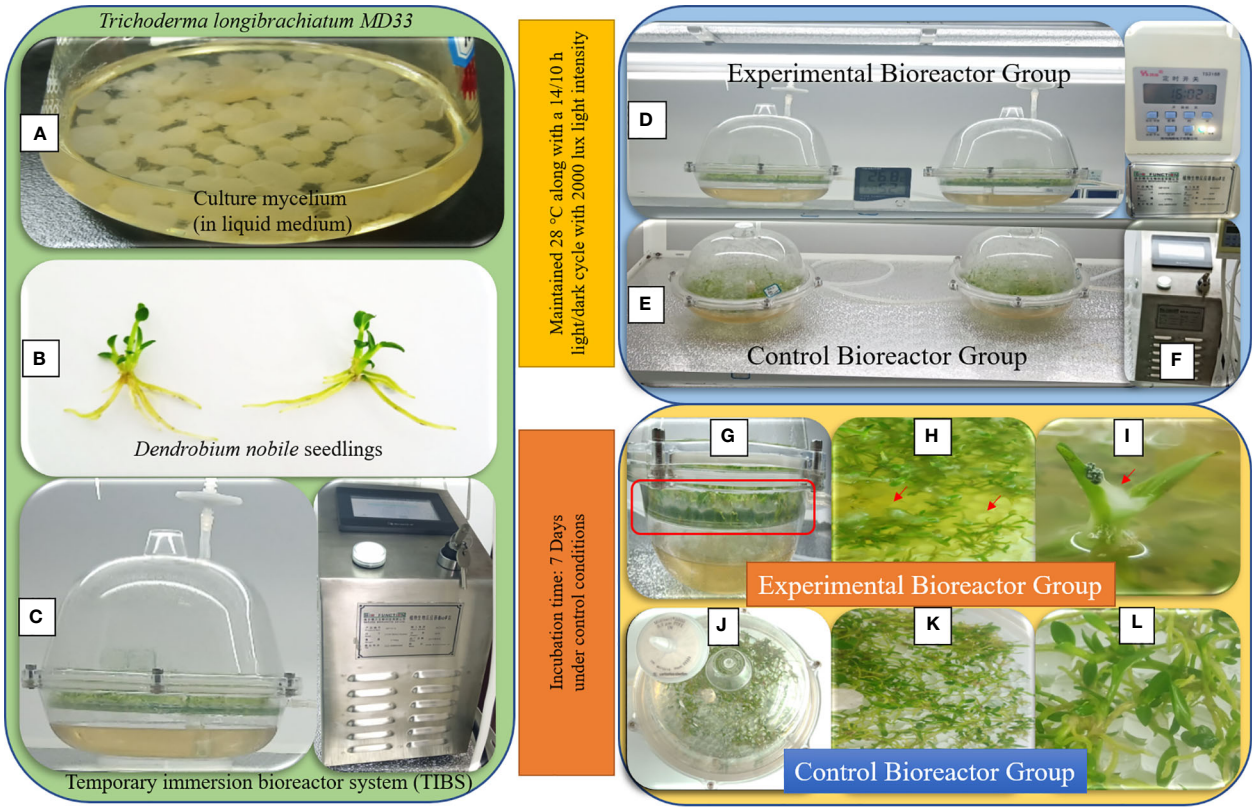


FIGURE 4
Temporary immersion bioreactor system (TIBS) co-culturing for high yield of *D nobile* seedlings dendrobine. **(A)** *Trichoderma longibrachiatum* MD33 mycelial bolls in the liquid medium; **(B)** Young healthy *D nobile* seedlings; **(C)** Temporary immersion bioreactor system (TIBS) connected to the BioF II machine for supply constant air with constant time; **(D)** Light growth chamber rack for TIBS; G: Side view of EGTIBS bioreactor with clear visible of fungal growth; **(H, I)** *D nobile* seedlings and *Trichoderma longibrachiatum* MD33 mycelium interaction; **(J)** Top view of CGTIBS; **(K, L)** Inside view of CGTIBS.

Structural elucidation of dendrobine

Different proton resonances were observed in the compound's ^{1}H NMR spectrum, which was recorded at 400 MHz using CD_3OD. The provided ^{1}H NMR spectrum reveals distinct chemical shifts and coupling patterns for various proton environments in the organic compound. Peaks A through K exhibit diverse patterns, including singlets (s), doublets (d), triplets (t), and multiplets (m). Notably, peak E appears as a sharp singlet at 2.49 ppm with a high integration value of 2.94, suggesting the presence of three equivalent protons in a unique chemical environment. Peaks G and K, both identified as multiplets, display complex coupling patterns, with peak G showing three protons and an integration value of 3.04, while peak K involves five protons with a substantial integration value of 4.51. The coupling constants (*J*) associated with certain peaks, such as peaks D, I, and K, provide additional information about the coupling interactions between neighboring protons. The provided ^{13}C NMR spectrum illustrates distinct carbon chemical shifts and environments within the organic compound. Peaks A to N display sharp singlets (s) with varying chemical shifts, ranging from 80.68 to 20.61 ppm. These peaks correspond to individual carbon atoms in the molecular structure, each characterized by a unique chemical shift and integral value. Notably, peak F stands out with a multiplet (m) at 49.01 ppm, suggesting a carbon environment associated with thirteen protons and an integration value of 13.23. This multiplet indicates a complex coupling pattern, possibly arising from the interaction with nearby protons. Peaks G and O, appearing as singlets at 45.15 ppm and 44.53 ppm, respectively, represent distinct carbon atoms with integral values of 1.27 and 1.32. These findings, combined with the ^{1}H NMR results, contribute to a comprehensive understanding of the molecular structure of the compound and serve as a valuable foundation for further analyses in this research context. The ^{1}H NMR spectrum in Figure 6A shows a chemical shift of the CD_3OD solvent peak to 3.31, and the ^{13}C NMR spectrum shows a chemical shift of the CD_3OD solvent peak to 49.22 (Figure 6B).

Discussion

Endophytic fungi residing within various tissues and organs of healthy plants maintain a unique association with their host plants without inciting them to stimulate natural compound biosynthesis (Ding et al., 2018; Wang et al., 2023). However, high commercial demand for these fungi has led to their extraction from natural habitats, placing them at risk of extinction. To safeguard and revitalize threatened and rare orchid plants, a reintroduction strategy involving fungal co-culturing becomes imperative

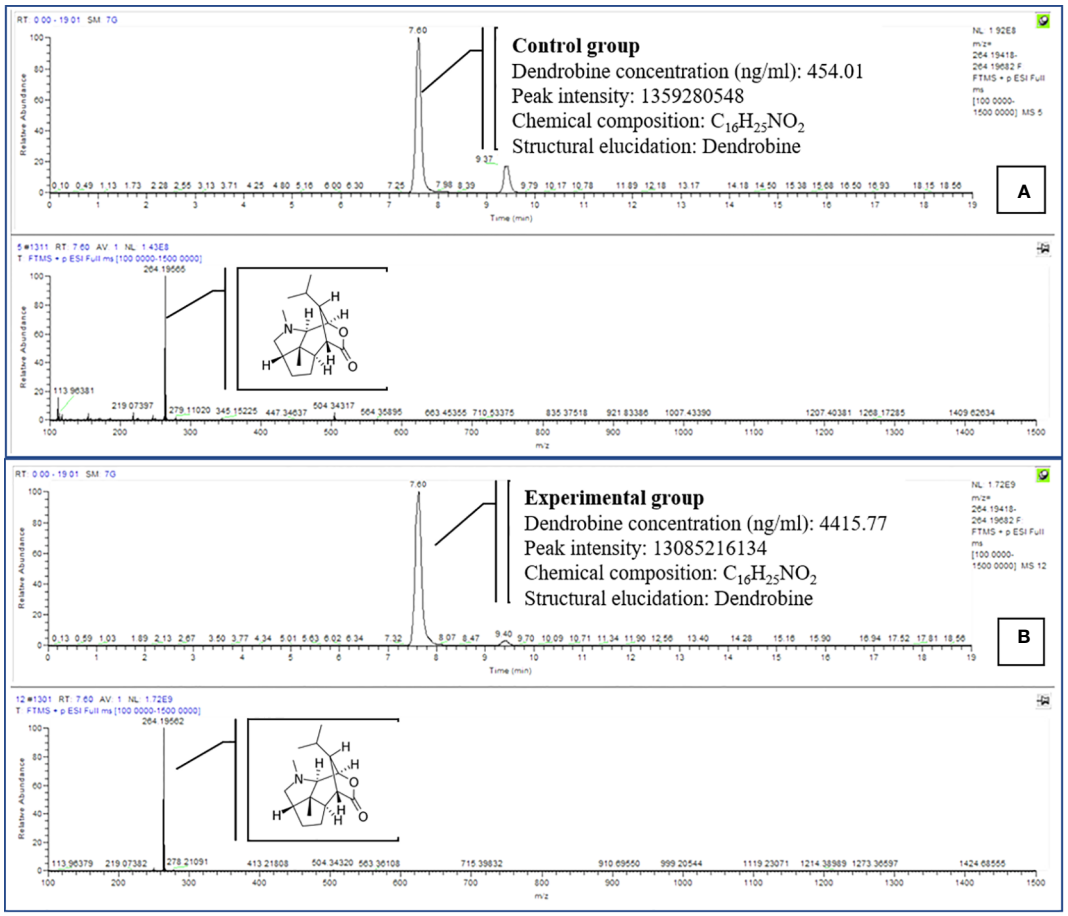


FIGURE 5
Comparison of dendrobine profiles produced by CGTIBS **(A)** and EGTIBS **(B)** *D. nobile* seedlings dendrobine. Tentative LCMS identification of the dendrobine is highlighted in the chromatogram (Molecular weight 264.195).

(Sarsaiya et al., 2020a). Therefore, it is crucial to identify pure endophytic fungal strains that can stimulate or inhibit the growth of *Dendrobium* seedlings *in vitro*. Research on the endophytic fungi associated with medicinal plants has revealed their potential. Some endophytic fungi can promote the accumulation of bioactive secondary metabolites in their host plants, whereas others can synthesize secondary metabolites similar to those found in their hosts (Sarsaiya et al., 2020b). These characteristics of endophytic fungi present a novel avenue for improving the production of active compounds through co-culturing approaches. Such methods offer new perspectives and techniques for enhancing the accumulation of bioactive components in medicinal plants during co-cultivation while ensuring the sustainable development of traditional Chinese medicine resources (Wen et al., 2022).

The present study explored the mechanism underlying increased dendrobine production, involving intricate interactions between the endophytic fungus and the host plant, especially when cultivated in TIBS. This finding aligns with previous research, suggesting the potential of endophytic fungi to enhance the production of secondary metabolites in medicinal plants (Jia

TABLE 3 Comparative analysis of EGTB and EGTIBS for *D. nobile* seedlings dendrobine concentration.

Co-cultured group	Host Plant	Test strain	Detection Technique	Dendrobine Yield (ng/ml)	Incubation time (days)	Fold increased *	Fold increased#
EGTB	*D. nobile* seedlings	*Trichoderma longibrachiatum* MD33	UHPLC	1804.23	7	2.6	–
EFTIBS	*D. nobile* seedlings	*Trichoderma longibrachiatum* MD33	UHPLC	4415.77	7	9.7	3.7

*as compared to CGTB and CGTIBS; #as compared to EGTB dendrobine concentration; ng/ml, nanogram per millilitre.

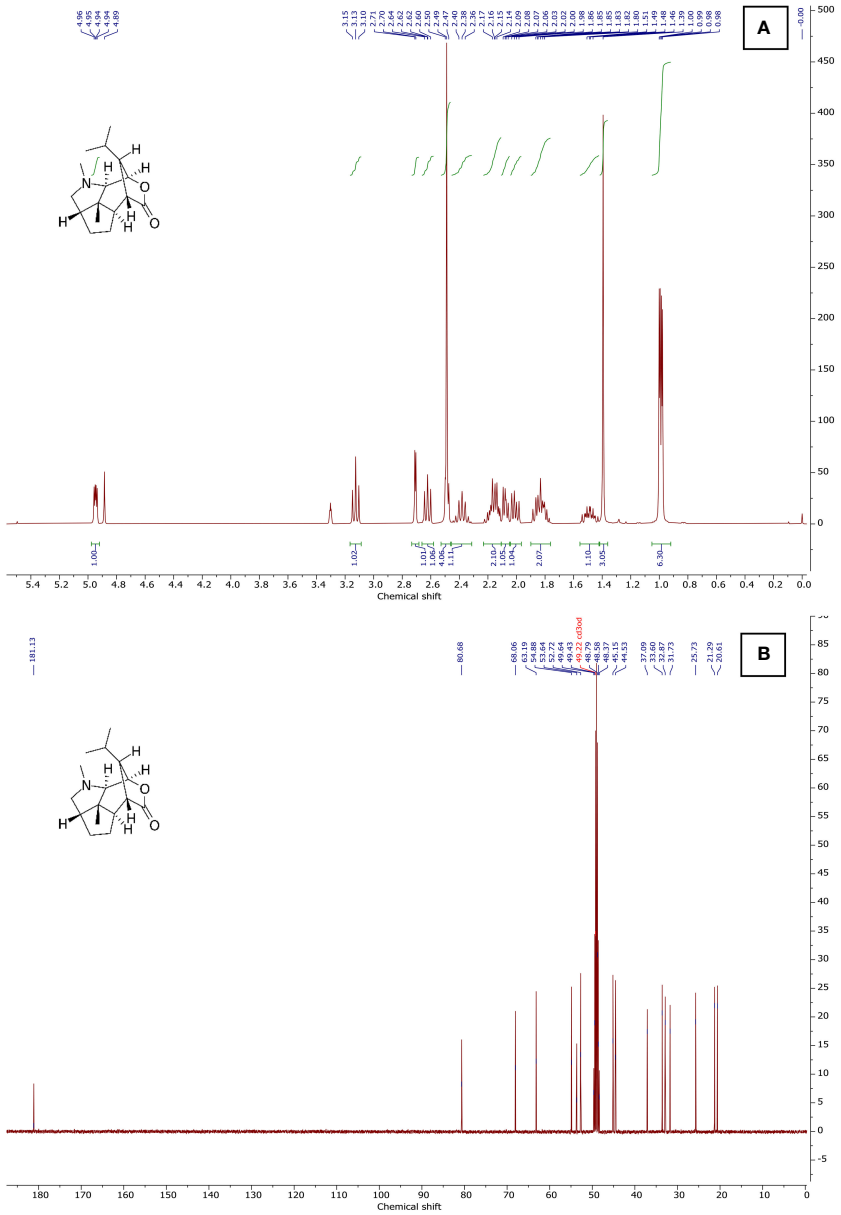


FIGURE 6
NMR spectra of *Dendrobium nobile* seedling dendrobine using ^{1}H and ^{13}C NMR. **(A)** ^{1}H NMR (400 MHz, CD_3OD), **(B)** ^{13}C NMR (101 MHz, CD_3OD).

et al., 2016; Ye et al., 2021). The use of the TIBS system proved to be even more promising, resulting in a remarkable 9.7-fold increase in dendrobine concentration after just seven days of incubation. This finding highlights the efficiency of the TIBS system in optimizing dendrobine production. Endophytic fungi form symbiotic relationships with their host plants and influence various physiological and biochemical processes, including the production of bioactive compounds (Verma et al., 2022). *Trichoderma* spp. have been reported to enhance plant growth, induce systemic resistance, and modulate the biosynthesis of secondary metabolites (Dutta et al., 2023). The UHPLC-MS analysis is a robust method for dendrobine identification and quantification. Confirmation of the presence of dendrobine by comparison with reference standards and the chemical reference dendrobine adds credibility to the results. The controlled and optimized environment provided by TIBS seemingly influenced the production and expression of secondary metabolites in MD33 cells compared to the control group. This alteration in secondary metabolite composition could be attributed to specific conditions and nutrient availability within the TIBS system, suggesting that the TIBS approach has a notable effect on the metabolic activity of endophytic fungi and plants (Cao et al., 2024).

Trichoderma longibrachiatum (MD33) establishes itself as an endophyte on *D. nobile* by entering the plant through natural openings. This initiated a symbiotic relationship, creating a conducive environment for both organisms. Endophytic fungi

contribute to plant well-being by providing nutrients and growth-promoting substances. *D. nobile* serves as a habitat offering nutrients to MD33. The fungus is known for producing secondary metabolites, including dendrobine, which enhances plant biosynthetic pathways, leading to increased dendrobine production (Sarsaiya et al., 2020b). A Temporary Immersion Bioreactor System (TIBS) provides a controlled and optimized environment for co-cultivation with a formulated liquid medium supporting growth. Periodic immersion and perfusion ensure uniform nutrient and oxygen distribution, promoting optimal growth and metabolic activity (Zhang et al., 2018; Hwang et al., 2022). Employing UHPLC-MS analysis, the study quantifies dendrobine, revealing a substantial increase in concentration in the TIBS approach compared to traditional test bottles. The researchers found that the TIBS is effective in quickly multiplying a significant number of *Dendrobium* seedlings. This makes it well-suited for large-scale production of dendrobine-type total alkaloids (DTTAs) and supports robust plant growth (Zhang et al., 2022; Cao et al., 2024). This highlights the scalability and commercial potential of TIBS for dendrobine production.

The utilization of LC-MS analyses has revealed notable variations in metabolite expression during dynamic interactions between the phytopathogen *Fusarium verticillioides* and the endophyte *Ustilago maydis*. Similarly, co-culturing *Paraconiothyrium variable* with *F. oxysporum* led to the induction of 12 metabolites, as reported by Serrano et al. (2017). These findings emphasize the significance of investigating the potential of Orchidaceae fungal endophytes (OFEs) as valuable sources of bioactive products, with implications for both agriculture and pharmaceuticals (Gupta et al., 2023). However, it is essential to acknowledge that previous research on OFEs-based products is relatively limited, necessitating further investigations into various facets, including screening procedures, extraction methods, separation techniques, and structural elucidation. The continuous symbiotic growth between plants and fungal endophytes makes them promising candidates for mediating bioproduct opportunities through enzymatic processes that alter their biochemical composition. Fungal endophytes can exert either stimulatory or inhibitory effects on host Orchidaceae metabolism. Additionally, the diversity within the host varieties may also play a pivotal role in influencing the profile of secondary metabolic bioproducts in OFEs, as underscored by studies conducted by Sarsaiya et al. (2019) and Adeleke and Babalola (2021).

Fungal-plant interactions involve complex signaling pathways. *Trichoderma* spp. produce signaling molecules and enzymes that modulate plant gene expression. The activation of specific genes involved in alkaloid biosynthesis could be a result of these signaling interactions, leading to increased dendrobine production (Druzhinina et al., 2011). Such symbiosis may result in a metabolic priming effect, in which the plant reallocates resources for secondary metabolite production, including alkaloids. The presence of endophytes such as MD33 may contribute to an altered metabolic profile in *Dendrobium*, favoring metabolite synthesis (Chen et al., 2022). The mechanism by which fungi enhance alkaloid production in plants is likely multifaceted and involves a combination of growth promotion, defense induction, signaling modulation, nutrient exchange, biosynthetic pathway activation, and epigenetic modifications. Further research is needed to unravel the specific molecular interactions between *D. nobile* and *T. longibrachiatum* (MD33) and to identify the key factors influencing the observed increase in dendrobine production. While acknowledging the known capabilities of *Trichoderma* species in enhancing plant growth and modulating biosynthetic pathways, the findings suggest that the effects of MD33 on dendrobine metabolic pathways should be explored. This included elucidating the specific molecular interactions between *D. nobile* and MD33 and identifying key factors influencing the observed increase in dendrobine production. Such insights are crucial for optimizing dendrobine biosynthesis and may have broader implications for cultivating medicinal plants with enhanced bioactive compound yields, emphasizing the multifaceted nature of symbiosis between endophytic fungi and host plants in influencing secondary metabolite production.

The intricate orchestration of dendrobine production involves the integration of various biogenetic routes, with the mevalonate (MVA) and MEP pathways playing pivotal roles. Qian et al. (2021) explored fungal molecular adaptation and revealed a complex genetic landscape with 1,024 differentially expressed genes, suggesting a metabolic symphony controlled by backbone post-modification enzymes and transcription factors. Furthermore, this study highlights the potential regulatory role of miRNAs in modulating dendrobine biosynthesis by targeting key enzymes such as AACT, MK, DXR, and HDS. This co-cultivation strategy encourages a broader perspective, extending beyond genetic expression, to elucidate the intricate relationships between mutagenesis, genetic networks, and metabolite dynamics. Additionally, the findings of Mou et al. (2021) challenged the conventional primacy of the MVA and MEP pathways, emphasizing the significance of AACT, PMK, and MVD genes in the MVA pathway and the upregulation of DXR and DXS genes in the MEP pathway after methyl jasmonate treatment. A comprehensive approach integrating genetic, miRNA, and pathway studies advances our understanding of dendrobine biosynthesis and opens avenues for biotechnological applications and targeted metabolic engineering to enhance dendrobine yield. Gong et al. (2022) identified enzyme genes that contribute to the groundwork for a comprehensive investigation of the dynamic interactions among key enzymes, shedding light on the complex regulatory mechanisms governing dendrobine biosynthesis pathways in a co-cultured system. This study significantly enhances our knowledge of dendrobine biosynthesis within the TIA family, offering potential insights for optimizing production and applications in the field of plant biology.

The structural elucidation of dendrobine through the combined analysis of ^{1}H and ^{13}C NMR spectra provided valuable insights into the molecular composition and bonding patterns of this natural product. The diverse array of proton environments revealed in the ^{1}H NMR spectrum reflects the complexity of the dendrobine chemical structure

(Liu et al., 2020). Dendrobine exhibits structural similarity to the potent convulsant picrotoxinin, as elucidated by Connolly and Heathcock (1985) using proton and ^{13}C nuclear magnetic resonance (NMR) spectroscopy. This analysis revealed that dendrobine formed a single stereoisomer at the newly created stereocenter. The ^{13}C NMR data showed 16 carbon signals, including four methyls (one N-CH_3), three methylenes, four methines, two quaternary carbons, two carbonyl carbons, and an oxygenated olefinic tertiary carbon. The ^{13}C NMR spectrum recorded at 101 MHz in CD_3OD supported these results. It showed clear carbon resonances with chemical shifts at 80.68 to 20.61 ppm, creating a singlet. Yang et al. (2018) provided additional insights into the NMR structure of dendrobine compared to the known dendrobine-type alkaloid dendrobine. The correlations from H3-16 (δ_H 2.96) to C-11 (δ_C 52.1, t) and C-15 (δ_C 171.6, s), and from H2-11 (δ_H 3.42, 2.93) to C-15, indicated a linkage between C-11 and C-15 via a nitrogen atom, resulting in the formation of a seven-membered lactam moiety. Analyzing the ^1H-NMR spectrum, four dendrobine methyl signals were identified. Peaks A to K exhibit diverse patterns, including singlets, doublets, triplets, and multiplets, each indicative of unique chemical environments. The ^{13}C-NMR spectrum displayed all 16 carbons, with C-12 readily identified in the carbonyl region. The quaternary carbon signal in the spectrum was assigned to C-11. Additionally, an oxygen-bearing CH signal was attributed to C-9, and two deshielded CH signals were assigned to C-7 and C-10, respectively (Wang et al., 1985). These combined NMR results provide a solid foundation for ongoing research, offering valuable insights into the intricate molecular architecture of the compound and paving the way for further in-depth analyses and applications.

Conclusion

In this study, we embarked on a journey to harness the potential of endophytic fungi, specifically *T. longibrachiatum* (MD33), for the enhanced production of dendrobine, a valuable alkaloid found in *D. nobile*. Our findings illuminate the transformative impact of mono-culturing these endophytic fungi within the confines of a Temporary Immersion Bioreactor System (TIBS) and traditional test bottles. In the test bottle experiment (EGTB), the co-cultivation of *D. nobile* seedlings with *T. longibrachiatum* (MD33) led to a commendable 2.6-fold increase in dendrobine concentration. This result underscores the potential of *T. longibrachiatum* (MD33) as a facilitator for dendrobine production within a controlled environment. However, the real breakthrough came with the application of the Temporary Immersion Bioreactor System (TIBS) in the EGTIBS experiment. Here, co-culturing *D. nobile* seedlings with *T. longibrachiatum* (MD33) resulted in a remarkable 9.7-fold increase in dendrobine concentration. The dendrobine content in the TIBS experimental group soared to an impressive 4415.77 ng/ml, far surpassing the control group's yield. This striking difference underscores the immense potential of the TIBS system for optimizing dendrobine production. Our study exemplifies the power of endophytic fungi to enhance the accumulation of valuable metabolites in medicinal plants like *D. nobile*. This newfound knowledge opens doors to potential applications in pharmaceuticals and other industries. Additionally, our research suggests that scaling up the TIBS approach for commercial dendrobine production holds great promise. As we conclude this endeavor, we look forward to future research directions with optimism. Further investigations could delve into refining the TIBS system, optimizing co-culturing conditions, and exploring the pharmacological potential of the increased dendrobine yield. Additionally, understanding the underlying mechanisms responsible for enhanced dendrobine production within the TIBS system could unlock even greater possibilities for the cultivation of *D. nobile* and other medicinal plants. The comprehensive analysis of the ^1H and ^{13}C NMR spectra of dendrobine has provided essential structural information, allowing for the identification of various proton and carbon environments within the molecule. The elucidated structural features contribute to a deeper understanding of dendrobine molecular architecture, laying the foundation for further investigations into its biological activities and potential synthetic applications. Our study marks a significant step forward in the sustainable utilization of natural resources and the advancement of biotechnological applications in the realm of medicinal plant metabolites.

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

SS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. AJ: Data curation, Formal analysis, Investigation, Resources, Visualization, Writing – review & editing. FS: Formal analysis, Writing – review & editing. MY: Resources, Writing – review & editing. MP: Resources, Writing – review & editing. QJ: Formal analysis, Resources, Visualization, Writing – review & editing. QG: Project administration, Supervision, Writing – review & editing. QW: Project administration, Supervision, Writing – review & editing. XQ: Resources, Visualization, Writing – review & editing. JS: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. JC: Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors are grateful for the financial support under the Distinguished High-Level Talents Research Grant from the Guizhou Science and Technology Corporation Platform Talents Fund (Grant No.: [2017] 5733-001 & CK-1130-002), the National Natural Science Foundation of China (82373981 and U1812403), the Jiangsu Synergetic Innovation Centre for Advanced Bio-Manufacture (No. XTC2208).

Acknowledgments

The authors are grateful for Zunyi Medical University for their advanced research facilities. We are also thanks to Dr. Yun Zhang from Department of Medicinal Chemistry for their help.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1302817/full#supplementary-material

References

Adeleke, B. S., and Babalola, O. O. (2021). Pharmacological potential of fungal endophytes associated with medicinal plants: A review. *J. Fungi.* 7 (2), 147. doi: 10.3390/jof7020147

Akone, S. H., Mándi, A., Kurtán, T., Hartmann, R., Lin, W., Daletos, G., et al. (2016). Inducing secondary metabolite production by the endophytic fungus Chaetomium sp. through fungal–bacterial co-culture and epigenetic modification. *Tetrahedron* 72 (41), 6340–6347. doi: 10.1016/j.tet.2016.08.022

Cao, J., Fan, X., Sarsaiya, S., Pan, X., Yang, N., Jin, L., et al. (2018). Accumulative and component analysis of polysaccharide in protocorm of *Dendrobium Candidum*. *J. Biobased Mater. Bioenergy* 12, 348–355. doi: 10.1166/jbmb.2018.1783

Cao, Y., Qian, X., Yu, T., Jia, Q., Sarsaiya, S., and Chen, J. (2024). Improving biomass and dendrobine-type total alkaloids (DTTAs) production of Dendrobium nobile through combining Temporary Immersion Bioreactor System (TIBS) with endophyte MD33 elicitation. *Plant Cell Tiss. Organ Cult.* 156, 9. doi: 10.1007/s11240-023-02653-w

Chen, X. L., Sun, M. C., Chong, S. L., Si, J. P., and Wu, L. S. (2022). Transcriptomic and metabolomic approaches deepen our knowledge of plant–endophyte interactions. *Front. Plant Sci.* 12, 700200. doi: 10.3389/fpls.2021.700200

Chen, W., Wang, J., Song, J., Sun, Q., Zhu, B., and Qin, L. (2023a). Exogenous and endophytic fungal communities of *Dendrobium nobile* Lindl. across different habitats and their enhancement of host plants' dendrobine content and biomass accumulation. *ACS Omega* 8 (13), 12489–12500. doi: 10.1021/acsomega.3c00608

Chen, X., Wang, S., Farag, M. A., Han, Z., Chen, D., Zhang, X., et al. (2023b). Interactions between endophytic fungus Pestalotiopsis sp. DO14 and *Dendrobium catenatum*: Deciphering plant polysaccharide and flavonoid accumulation and underlying mechanisms by comparative transcriptome and metabolome analyses. *Plant Physiol. Biochem.* 202, 107942. doi: 10.1016/j.plaphy.2023.107942

Connolly, P. J., and Heathcock, C. H. (1985). An approach to the total synthesis of dendrobine. *J. Org. Chem.* 50, 4135–4144. doi: 10.1021/jo00221a033

Ding, C., Wang, Q.-B., Guo, S., and Wang, Z. (2018). The improvement of bioactive secondary metabolites accumulation in *Rumex gmelini* Turcz through co-culture with endophytic fungi. *Braz. J. Microbiol.* 49 (2), 362–369. doi: 10.1016/j.bjm.2017.04.013

Druzhinina, I. S., Seidl-Seiboth, V., Herrera-Estrella, A., Horwitz, B. A., Kenerley, C. M., Monte, E., et al. (2011). *Trichoderma*: the genomics of opportunistic success. *Nat. Rev. Microbiol.* 9 (10), 749–759. doi: 10.1038/nrmicro2637

Dutta, P., Mahanta, M., Singh, S. B., Thakuria, D., Deb, L., Kumari, A., et al. (2023). Molecular interaction between plants and *Trichoderma* species against soil-borne plant pathogens. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1145715

Gegenbauer, C., Bellaire, A., Schintlmeister, A., Schmid, M. C., Kubicek, M., Voglmayr, H., et al. (2023). Exo- and endophytic fungi enable rapid transfer of nutrients from ant waste to orchid tissue. *New Phytol.* 238 (5), 2210–2223. doi: 10.1111/nph.18761

Gong, D., Wu, B., Qin, H., Fu, D., Guo, S., Wang, B., et al. (2022). Functional characterization of a farnesyl diphosphate synthase from *Dendrobium nobile* Lindl. *AMB Express.* 12 (1), 129. doi: 10.1186/s13568-022-01470-2

Gupta, A., Meshram, V., Gupta, M., Goyal, S., Qureshi, K. A., Jaremko, M., et al. (2023). Fungal endophytes: Microfactories of novel bioactive compounds with therapeutic interventions; A comprehensive review on the biotechnological developments in the field of fungal endophytic biology over the last decade. *Biomolecules* 13 (7), 1038. doi: 10.3390/biom13071038

Hajong, S., and Kapoor, R. (2020). An amalgam of pathogenic and beneficial endophytic fungi colonizing four *Dendrobium* species from Meghalaya, India. *J. Basic Microbiol.* 60 (5), 415–423. doi: 10.1002/jobm.201900631

Hernández-Ramírez, F., Damon, A., Fernández Pavía, S. P., Guillén-Navarro, K., Iracheta-Donjuan, L., Zarza, E., et al. (2023). Community richness and diversity of endophytic fungi associated with the orchid *Guarianthe skinneri* Infested with "Black Blotch" in the Soconusco Region, Chiapas, Mexico. *Diversity* 15 (7), 807. doi: 10.3390/d15070807

Hu, Y., Zhang, B., Jia, M., and Chen, J. (2016). Tissue culture of *dendrobium candidum* by batch immersion bioreactor. *China J. Agric. Sci. Technol.* 18, 190–194. doi: 10.13304/j.nykjdb.2016.022

Hwang, H. D., Kwon, S. H., Murthy, H. N., Yun, S. W., Pyo, S. S., and Park, S. Y. (2022). Temporary immersion bioreactor system as an efficient method for mass production of in *vitro* plants in horticulture and medicinal plants. *Agron* 12, 346. doi: 10.3390/agronomy12020346

Jia, M., Chen, L., Xin, H. L., Zheng, C. J., Rahman, K., Han, T., et al. (2016). A friendly relationship between endophytic fungi and medicinal plants: a systematic review. *Front. Microbiol.* 7, 906. doi: 10.3389/fmicb.2016.00906

Liu, G. Y., Tan, L., Cheng, L., Ding, L. S., Zhou, Y., Deng, Y., et al. (2020). Dendrobine-type alkaloids and bibenzyl derivatives from *Dendrobium findlayanum*. *Fitoterapia* 142, 104497. doi: 10.1016/j.fitote.2020.104497

Mou, Z., Zhao, Y., Ye, F., Shi, Y., Kennelly, E. J., Chen, S., et al. (2021). Identification, biological activities and biosynthetic pathway of dendrobium alkaloids. *Front. Pharmacol.* 12. doi: 10.3389/fphar.2021.605994

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

Qian, X., Zhu, J., Yuan, Q., Jia, Q., Jin, H., Han, J., et al. (2021). Illumina sequencing reveals conserved and novel MicroRNAs of *Dendrobium nobile* protocorm involved in synthesizing dendrobine, a potential nanodrug. *J. Biomed. Nanotech.* 17 (3), 416–425. doi: 10.1166/jbn.2021.3036

Salazar, J. M., Pomavilla, M., Pollard, A. T., Chica, E. J., and Peña, D. F. (2020). Endophytic fungi associated with roots of epiphytic orchids in two Andean forests in Southern Ecuador and their role in germination. *Lankesteriana* 20 (1), 37–47. doi: 10.15517/lank.v20i1.41157

Sarsaiya, S., Jain, A., Fan, X., Jia, Q., Xu, Q., Shu, F., et al. (2020b). New insights into detection of a dendrobine compound from a novel endophytic *Trichoderma longibrachiatum* strain and its toxicity against phytopathogenic bacteria. *Front. Microbiol.* 11, 337. doi: 10.3389/fmicb.2020.00337

Sarsaiya, S., Jain, A., Jia, Q., Fan, X., Shu, F., Chen, Z., et al. (2020a). Molecular identification of endophytic fungi and their pathogenicity evaluation against *Dendrobium nobile* and *Dendrobium officinale*. *J. Mol. Sci.* 21 (1), 316. doi: 10.3390/ijms21010316

Sarsaiya, S., Shi, J., and Chen, J. (2019). A comprehensive review on fungal endophytes and its dynamics on Orchidaceae plants: current research, challenges, and future possibilities. *Bioengineered* 10 (1), 316–334. doi: 10.1080/21655979.2019.1644854

Serrano, R., González-Menéndez, V., Rodríguez, L., Martín, J., Tormo, J. R., and Genilloud, O. (2017). Co-culturing of fungal strains against botrytis cinerea as a model

for the induction of chemical diversity and therapeutic agents. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00649

Verma, A., Shameem, N., Jatav, H. S., Sathyanarayana, E., Parray, J. A., Poczai, P., et al. (2022). Fungal endophytes to combat biotic and abiotic stresses for climate-smart and sustainable agriculture. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.953836

Vinale, F., Nicoletti, R., Borrelli, F., Mangoni, A., Parisi, O. A., Marra, R., et al. (2017). Co-Culture of plant beneficial microbes as source of bioactive metabolites. *Sci. Rep.* 7, 14330. doi: 10.1038/s41598-017-14569-5

Wang, Y., Zhang, Y., Cong, H., Li, C., Wu, J., Li, L., et al. (2023). Cultivable endophyte resources in medicinal plants and effects on hosts. *Life* 13, 1695. doi: 10.20944/preprints202306.1851.v1

Wang, H., Zhao, T., and Che, C. T. (1985). Dendrobine and 3-hydroxy-2-oxodendrobine from *dendrobium nobile*. *J. Nat. Prod.* 48 (5), 796–801. doi: 10.1021/np50041a014

Wen, J., Okyere, S. K., Wang, S., Wang, J., Xie, L., Ran, Y., et al. (2022). Endophytic fungi: An effective alternative source of plant-derived bioactive compounds for pharmacological studies. *J. Fungi (Basel)* 8 (2), 205. doi: 10.3390/jof8020205

Yang, D., Cheng, Z. Q., Yang, L., Hou, B., Yang, J., Li, X. N., et al. (2018). Seco-Dendrobine-Type alkaloids and bioactive phenolics from *Dendrobium findlayanum*. *J. Nat. Prod.* 81 (2), 227–235. doi: 10.1021/acs.jnatprod.7b00150

Ye, H. T., Luo, S. Q., Yang, Z. N., Wang, Y. S., Ding, Q., Wang, K. F., et al. (2021). Endophytic fungi stimulate the concentration of medicinal secondary metabolites in houttuynia cordata thumb. *Plant Signal. Behav.* 16 (9), 1929731. doi: 10.1080/15592324.2021.1929731

Yu, J., Jiang, H., Zhang, B., Jin, L., Xu, D., Zhang, B., et al. (2017). Effects of two endophytic fungi from *Dendrobium candidum* on the growth of plantlets and protocorms. *Acta Plant Pathol.* 47, 541–550. doi: 10.13926/j.cnki.apps.000062

Zhang, B., Hu, Y., Jin, L., and Chen, J. (2015). Comparison of tissue culture medicinal and nutrient content of *Dendrobium candidum*. *Jiangsu Agric. Sci.* 43, 324–327. doi: 10.15889/j.issn.1002-1302.2015.10.106

Zhang, B., Niu, Z., Li, C., Hou, Z., Xue, Q., Liu, W., et al. (2022). Improving large-scale biomass and total alkaloid production of *Dendrobium nobile* Lindl. using a temporary immersion bioreactor system and MeJA elicitation. *Plant Methods* 18 (1), 10. doi: 10.1186/s13007-022-00843-9

Zhang, B., Sarsaiya, S., Pan, X., Jin, L., Xu, D., Zhang, B., et al. (2018). Optimization of nutritional conditions using a temporary immersion bioreactor system for the growth of Bletilla striata pseudobulbs and accumulation of polysaccharides. *Sci. Hortic.* 240, 155–161. doi: 10.1016/j.scienta.2018.06.010

Zhao, Y., Qin, L., Tan, D., Wu, D., Wu, X., Fan, Q., et al. (2023). Fatty acid metabolites of *Dendrobium nobile* were positively correlated with representative endophytic fungi at altitude. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1128956

ZhiPing, L., Yuan, L., FeiFei, S., RuoNan, Z., and QinGeng, H. (2023). Isolation and characterization of *Paenibacillus peoriae* JC-3jx: An endophytic bacterium from *Dendrobium nobile* with antibacterial and growth-promoting potential. *Res. Square* [Preprint]. doi: 10.21203/rs.3.rs-2940079/v1 (Accessed July 15, 2023).

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 February 2024
DOI 10.3389/fpls.2024.1284125

OPEN ACCESS

EDITED BY
Naoki Kitaoka,
Hokkaido University, Japan

REVIEWED BY
Hexin Tan,
Second Military Medical University, China
Gerasimos Daras,
Agricultural University of Athens, Greece

*CORRESPONDENCE
Rajani Sarojam
rajanis@tll.org.sg

RECEIVED 28 August 2023
ACCEPTED 23 January 2024
PUBLISHED 05 February 2024

CITATION
Reddy VA, Saju JM, Nadimuthu K and Sarojam R (2024) A non-canonical *Aux/IAA* gene *MsIAA32* regulates peltate glandular trichome development in spearmint. *Front. Plant Sci.* 15:1284125. doi: 10.3389/fpls.2024.1284125

A non-canonical *Aux/IAA* gene *MsIAA32* regulates peltate glandular trichome development in spearmint

Vaishnavi Amarr Reddy, Jolly Madathiparambil Saju, Kumar Nadimuthu and Rajani Sarojam*

Temasek Life Sciences Laboratory, National University of Singapore, Singapore, Singapore

Phytohormone auxin controls various aspects of plant growth and development. The typical auxin signalling involves the degradation of canonical Aux/IAA proteins upon auxin perception releasing the auxin response factors (ARF) to activate auxin-regulated gene expression. Extensive research has been pursued in deciphering the role of canonical Aux/IAAs, however, the function of non-canonical *Aux/IAA* genes remains elusive. Here we identified a non-canonical *Aux/IAA* gene, *MsIAA32* from spearmint (*Mentha spicata*), which lacks the TIR1-binding domain and shows its involvement in the development of peltate glandular trichomes (PGT), which are the sites for production and storage of commercially important essential oils. Using yeast two-hybrid studies, two canonical Aux/IAAs, MsIAA3, MsIAA4 and an ARF, MsARF3 were identified as the preferred binding partners of MsIAA32. Expression of a R2R3-MYB gene *MsMYB36* and a cyclin gene *MsCycB2-4* was altered in *MsIAA32* suppressed plants indicating that these genes are possible downstream targets of MsIAA32 mediated signalling. Ectopic expression of *MsIAA32* in *Arabidopsis* affected non-glandular trichome formation along with other auxin related developmental traits. Our findings establish the role of non-canonical Aux/IAA mediated auxin signalling in PGT development and reveal species-specific functionalization of Aux/IAAs.

KEYWORDS

RNAi, peltate glandular trichomes, Aux/IAA, ARF, auxin, spearmint

1 Introduction

Auxin plays a critical role in regulating multiple aspects of plant development and growth, leading to significant interest in understanding its signalling and response mechanism. The key players involved in auxin responsive transcription are transport inhibitor-resistant 1/auxin signalling F-box (TIR1/AFB) receptors, auxin/indole-3-acetic acid (Aux/IAA) repressors and auxin response factors (ARF) (Nemhauser, 2018).

Canonical auxin signalling is based on TIR1/AFB mediated Aux/IAA protein ubiquitination and degradation upon auxin perception leading to ARF's derepression and the activation of auxin-responsive gene regulation. All canonical Aux/IAAs possess four conserved domains (I–IV). The smallest Domain I has the "LxLxL" motif, known as the ethylene response factor (ERF)-associated amphiphilic repression (EAR) motif (Luo et al., 2018) which is important for conferring repression. Domain II has the "GWPPV" degron motif and controls the Aux/IAA proteins' turnover by interacting with the E3 protein-ubiquitin ligase, SCF^{TIR1} (SKP1, Cullin, and TIR1 F-box-containing proteins) (Ramos et al., 2001). SCFs ubiquitinate their targets, marking them for degradation by the 26S proteasome (Deshaies, 1999). Domains III and IV together form type Bem1p (PB1) and I/II Phox domains that mediate hetero- and homodimerization of Aux/IAA proteins and heterodimerization of ARF and Aux/IAA proteins (Reed, 2001; Luo et al., 2018). Non-canonical Aux/IAAs are characterized by the absence of either full or partial domain I, II and/or III and are presumed to function differently from canonical Aux/IAAs. Two nuclear localization signals (NLSs) are found in canonical Aux/IAAs compared to only one NLS in non-canonical Aux/IAAs (Wu et al., 2017). Both canonical and non-canonical Aux/IAAs regulate gene expression by binding with other Aux/IAAs or ARFs. Similar to domains III and IV of Aux/IAA proteins, ARF proteins have a C-terminal dimerization domain (CTD (Li et al., 2016). They also have a well-preserved N-terminal region that binds DNA and specifically targets auxin response elements (AuxRE) in the promoter regions of genes regulated by auxin (Tiwari et al., 2003). These AuxRE sites contain variations of the TGTCNN motif, which are involved in the response to auxin in *Arabidopsis* (Zemlyanskaya et al., 2016). ARFs have a middle region that functions as a repression domain (RD) or an activation domain (AD) depending on the amino acid sequences present, and accordingly, they are called as activator ARFs or repressor ARFs (Li et al., 2016). Canonical auxin signalling is well studied and documented but with the emergence of non-canonical Aux/IAAs, information regarding their functions and signalling mechanisms remains limited. Only a few non-canonical *Aux/IAA* genes have been characterized till date. Study on two non-canonical *Aux/IAA* genes *IAA30* and *IAA20* in *Arabidopsis*, highlighted their importance in vascular pattering (Müller et al., 2016). Another study in rice showed the involvement of a non-canonical *Aux/IAA*, *OsIAA26* in ethylene regulated root growth (Chen et al., 2018). Recently the role of *IAA32* and *IAA34* in regulating differential growth of apical hook (Cao et al., 2019) and *IAA33* in maintaining root distal stem identity in *Arabidopsis* were established (Lv et al., 2020).

Glandular trichomes are specialized structures that produce and store a diverse array of specialized metabolites, which are critical for plants' fitness and their interaction with the surrounding environment. The model plant *Arabidopsis* lacks such structures and possesses non-glandular trichomes. Spearmint possesses specialized structures called peltate glandular trichomes where valuable essential oils are produced (Reddy et al., 2017). These "green bio-factories" are located in the aerial tissues of aromatic plants, producing and storing large quantities of volatile metabolites (Turner et al., 2000; Champagne and Boutry, 2013; Lange and Turner, 2013). Since these metabolites have high commercial value, understanding the development of PGTs is of great interest. Studies in tomato and cotton have established the role of auxin as an important regulator of trichome development (Li et al., 2021) but minimal knowledge is available on the role of Aux/IAAs in glandular or non-glandular trichome development (Deng et al., 2012b; Zhang et al., 2015). The only well characterized *Aux/IAA* gene involved in trichome development is a canonical *Aux/IAA*, *SlIAA15* from tomato. Downregulation of *SlIAA15* which shows ubiquitous expression across many tissues resulted in multiple phenotypes along with reduced trichome number (Deng et al., 2012b). Later trichome enriched ARFs, *SlARF4* and *SlARF3* were identified from tomato which regulates auxin induced glandular and non-glandular trichome development (Zhang et al., 2015; Yuan et al., 2021). *SlARF4* was shown to mediate trichome development through MYBs and cyclin genes (Yuan et al., 2021). Recently *ARF1* was characterized from *Artemisia annua* as positive regulator of glandular trichome formation and artemisinin biosynthesis (Guo et al., 2023). But no information is available on the role of non-canonical Aux/IAAs in trichome development.

In this study, we isolated and characterized a novel non-canonical *Aux/IAA* gene, *MsIAA32*, from spearmint which regulates PGT development. Phylogenetic analysis showed that MsIAA32 belongs to the same clade of non-canonical Aux/IAA proteins as *Arabidopsis* IAA32 and IAA34 that lack domain II. Down-regulation of *MsIAA32* in spearmint lead to a decrease in the number of PGTs. Using yeast two-hybrid assay, MsARF3, MsIAA3 and MsIAA4 were identified as the binding partners of MsIAA32. The possible role of *MsIAA32* in non-glandular trichome formation was confirmed by ectopic expression in *Arabidopsis*. Along with an increase in trichome density, typical auxin-related phenotypes like reduced leaf size, decreased lateral roots and curled leaves, were also observed. Our results demonstrate the involvement of a non-canonical Aux/IAA protein in glandular and non-glandular trichome formation, and it also shows species-specific functionalization of Aux/IAAs. Further, identifying new gene targets controlling glandular trichome numbers in aromatic plants will provide new ways to increase secondary metabolite production and improve plant stress response.

2 Methods

2.1 Plant material, transformation, and selection of transgenics

Spearmint (*M. spicata*) and tobacco (*N. benthamiana)* were grown in a greenhouse under natural conditions. Columbia ecotype of *Arabidopsis thaliana* was used as the wild-type (WT). Plants were grown in a growth chamber maintained at 20°C with light conditions of 16 h light and 8 h dark. *Agrobacterium*-mediated transformation of spearmint was carried out using a previously published protocol (Reddy et al., 2017). Spearmint transgenics were screened for the fluorescence of the visual marker, GFP. Transformation of tobacco was pursued as described previously (Gallois and Marinho, 1995). Tobacco transgenics were screened for

the fluorescence of their visual marker, mCherry. Genomic DNA (gDNA) was then extracted from GFP/mCherry-positive plants and used as a template for genotyping the transgenic lines. Southern blot analysis was pursued on the DNA-positive lines. The floral dip method was used for transforming *Arabidopsis* as described previously (Zhang et al., 2006). *Arabidopsis* lines were screened by basta selection (10 mg/L) and GFP fluorescence. T3 generation plants were used for analysis.

2.2 RNA isolation and quantitative real-time PCR

PGTs were isolated from 2-3 cm leaves as previously described (Champagne and Boutry, 2013). Later, total RNA was isolated from PGTs and other tissues. cDNA was synthesized from 500 ng of RNA and the gene expression levels in various tissues were analyzed by qRT-PCR. The expression was normalized using the GAPDH for spearmint and Tubulin for *Arabidopsis* and analyzed through the comparative $\Delta\Delta^{CT}$ method (Livak and Schmittgen, 2001).

2.3 Gene amplification and plasmid construction

The sequences were amplified and cloned into pENTR™/D-TOPO® gateway vector (Invitrogen, Germany). LR recombination transferred the cloned target gene into destination vectors, and the destination plasmids were transformed into Agrobacterium EHA105 by heat shock which was used for further experiments. The primer sequences are shown in Supplementary Table S1.

2.4 Subcellular localization and BiFC

The *MsIAA32* ORF was cloned into the pENTR/D-TOPO gateway vector (Invitrogen, Germany). *MsIAA32*/pENTR vector was then transferred into the pBA-DC-YFP destination vector (Zhang et al., 2005), which contains the C-terminal in frame with YFP and CaMV 35S promoter. For BiFC, donor vectors were transferred into pBA-^{C}YFP-DC or pBA-^{N}YFP-DC. *Agrobacterium* strain EHA105 was transformed with the constructs by heat shock. The EHA105 cells transformed with the constructs were cultured overnight at 28°C and 200 rpm, then resuspended in a solution containing 10 mM MES pH 5.6, 100 μM acetosyringone and 10 mM $MgCl_2$ to 1 OD_{600}. After a 3 h incubation, the solutions were infiltrated in combinations of 1:1 ratio into one-month-old *N. benthamiana* leaves and fluorescence signals were viewed using confocal scanning laser microscope two days later.

2.5 Auxin treatment

For auxin-mediated PGT induction study, 2,4-D (10 mg/L, 0.045mM) was sprayed every two days for ten days on the top leaves of one-month-old WT spearmint plants. Newly grown leaves were subjected to SEM analysis. For auxin-mediated gene expression study, 100 mM IAA was sprayed onto leaves of one-month-old spearmint plants and the top two 2 cm leaves were collected at 0, 2, 5, 10, 30 and 60 min time points for RNA isolation and expression analysis.

2.6 Southern blot analysis

Southern blot was performed on DNA-positive lines. 15 µg of genomic DNA was cut with NdeI overnight and then run on a 1% (w/v) agarose gel at 40 V for 5 h. The gel was then washed with 250 mM HCL for 15 min with shaking at RT, followed by a wash with distilled water and soaked twice in denaturation solution (0.5 M NaOH and 1.5 M NaCl) for 15 min each at RT with gentle shaking. The gel was again washed with water and submerged twice in neutralization solution (0.5 M Tris-HCL and 1.5 M NaCl) for 15 min each at RT with gentle shaking. The gel was then transferred to nylon membrane by capillary transfer method. Next day, the wet membrane was UV-crosslinked using a crosslinker followed by a brief water rinse. The membrane was then hybridized with CaMV 35S promoter probe. Next day, the membrane was rinsed with 2x wash solution (2x saline-sodium citrate (SSC) and 0.1% SDS) for 5 min each at RT with gentle shaking. Later, the membrane was rinsed with 0.5x wash solution (0.5x SSC and 0.1% SDS) for 15 min each at 68°C with gentle shaking. The membrane was then submerged in washing buffer for 1 min and blocked by incubating in blocking buffer for 3 h at RT with gentle shaking. Later the membrane was soaked in fresh blocking solution containing anti-DIG antibody (150 U/ml) and incubated for 30 min at RT with gentle shaking. The membrane was then washed twice with washing buffer for 15 min each at RT with gentle shaking followed by equilibration in detection buffer for 2 min. The membrane was then placed in a clean, transparent sheet, and diluted CDP-star (1:300) was added on top of it. Finally, the number of T-DNA insertions were analyzed using a ChemiDoc imaging system (BioRad).

2.7 GC-MS analysis

For GC-MS analysis, 2-3 cm leaves (4-6) were ground to powder with liquid nitrogen and mixed with 500 µl ethyl acetate containing camphor (50 µg/ml) as an internal control. The samples were shaken for 10 minutes at room temperature, then centrifuged for 10 minutes at 13,000 rpm. The top organic layer was dehydrated with anhydrous Na_2SO_4 and analyzed using GC-MS (Agilent Technologies, USA). 2 µl of sample was injected and separated using a HP-5 MS column with a temperature program of 50°C for 1 min increasing at 8°C/min to 300°C and held for 5 min.

2.8 Promoter cloning and GUS staining

GenomeWalker™ Universal kit was used to amplify the flanking sequences of *MsIAA32*. The identified 729-bp promoter

region of *MsIAA32*, was amplified and cloned into pENTR™/D-TOPO® which was then transferred into pBGWFS7 by LR recombination and later used to transform *Agrobacterium* EHA105. For *MsIAA32* promoter analysis, transgenic tobacco lines were generated using the recombinant EHA105 Agrobacterium strain. The transgenic lines were tested for GUS staining by immersing the tissue in a solution (100mM sodium phosphate buffer (pH 7), 0.1% Triton X-100, 2 mM potassium ferrocyanide, 1 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-glucuronide, 2 mM potassium ferricyanide and10 mM EDTA) and incubating in in the dark overnight at 37°C. Later, the tissue was cleared of chlorophyll by treating with 70% ethanol, and the stained GUS was captured using a Zeiss Whitefield microscope.

2.9 Yeast two-hybrid assay

Matchmaker Gold yeast two-hybrid library screening system was used to identify and confirm the binding partners of MsIAA32. ORF of *MsIAA32* was cloned into pGBKT7 vector and transformed into Y2H gold yeast, which was then used to screen a library generated by cloning spearmint leaf cDNA transcripts into pGADT7 vector. Positive clones were selected in SD/–Ade/–His/–Leu/–Trp/X-a-Gal/AbA plates and confirmed by colony PCR. ORFs of identified targets were individually cloned into pGADT7 vector and subjected to two-hybrid screening to confirm interaction. p53 bait and prey vectors provided in the kit were used as positive controls. Yeast transformation was implemented as mentioned in the user manual of Matchmaker Gold Yeast two-hybrid system.

2.10 Scanning electron microscope

For Scanning Electron Microscopy (SEM), similar sized leaves from WT and selected RNAi lines were collected from the same node and observed under a scanning electron microscope (Joel). One leaf each from three different plants of each line was selected and eight identical spots of same diameter were imaged for each selected leaf of WT and RNAi lines to avoid variation. All 24 images from each line were used for counting the number of trichomes. For epidermal cells and stomatal counting, identical spots from three similar sized leaves were imaged. For *Arabidopsis* overexpression lines and WT, eight identical spots of same diameter from leaf number six were chosen (n=3, 24 images per line). The number of trichomes, epidermal cells and stomata were measured using ImageJ. Trichome density assessment was done as described previously with slight modifications (Watts and Kariyat, 2021). Images were captured at 50x magnification for spearmint which contains ~ 4.78 mm^2 leaf area and 25x for *Arabidopsis* which contains ~19.38 mm^2 leaf area as measured by Image J. Trichome density per mm^2 was calculated by dividing the number of trichomes in 50x magnified image by 4.78 for spearmint and by dividing the number of trichomes in 25x magnified image by 19.38. Similarly, epidermal cells and stomatal density per mm^2 was calculated by dividing the number of epidermal cells and stomata in 900x and 350x magnified images respectively by their corresponding leaf areas of ~ 0.014 mm^2 and 0.104 mm^2.

2.11 Phylogenetic analysis

MEGA11 was used to construct the phylogenetic tree using Maximum Likelihood method with bootstrap values of 1000 replicates. *A. thaliana* and *Solanum lycopersicum* sequences were taken from NCBI database.

2.12 Statistical analysis

The data is shown as "mean ± standard deviation (SD)" based on 3 to 6 biological replicates, each with 3 duplicates. Significance between transgenic plants and WT was evaluated using a two-tailed Student's t-test and is indicated by asterisks. Asterisk notation: * signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$.

3 Results

3.1 Identification and characterization of a PGT-specific non-canonical *MsIAA32*

From the previously generated transcriptome data of four different tissues of spearmint, namely leaves devoid of PGTs, PGTs, leaves and roots (Jin et al., 2014), 22 *Aux/IAAs* were annotated (Figure 1A). Out of these *Aux/IAAs*, *MsIAA32*, showed PGT-specific expression and was selected for further characterization. This was validated by qRT-PCR across various tissues (Figure 1B). The full-length ORF of *MsIAA32*, consisting of 540 bp and encoding a 180 amino acid polypeptide, was amplified from PGT cDNA. MsIAA32 showed the highest sequence similarity (~76%) to *Sesamum indicum* IAA32 in NCBI protein BLAST analysis. The amino acid sequences of known *Solanum lycopersicum* (Sl) IAAs and *Arabidopsis thaliana* (At) IAAs were used to construct a phylogenetic tree (Supplementary Figure S1). Previous phylogenetic analysis of AtIAAs and SlIAAs has led the IAAs to be grouped into 11 distinct clades (A-K), of which Clade H, Clade I and Clade K members are considered as the non-canonical IAAs due to the absence of one or two of the conserved domains (Audran-Delalande et al., 2012). MsIAA32 fell under clade I along with AtIAA32, AtIAA34 and SlIAA32. A clustal alignment of MsIAA32 with IAA32 from other plants revealed the absence of domain II which is also observed in the members of Clade I, SlIAA32 (Audran-Delalande et al., 2012), AtIAA32 and AtIAA34 (Shimizu-Mitao and Kakimoto, 2014) (Supplementary Figure S2). Sequence of MsIAA32 was ~ 62%, 46% and 45% identical to the sequences of members of clade I, SlIAA32, AtIAA32 and AtIAA34 respectively. Additionally, a SV40-like nucleus localization signal (NLS) was present in domain IV (Supplementary Figure S2).

The subcellular localization of MsIAA32 was examined by fusing its ORF with a yellow fluorescent protein (YFP) and expressing it in *Nicotiana benthamiana* leaves using Agrobacterium infiltration under control of the CaMV 35S promoter. The recombinant

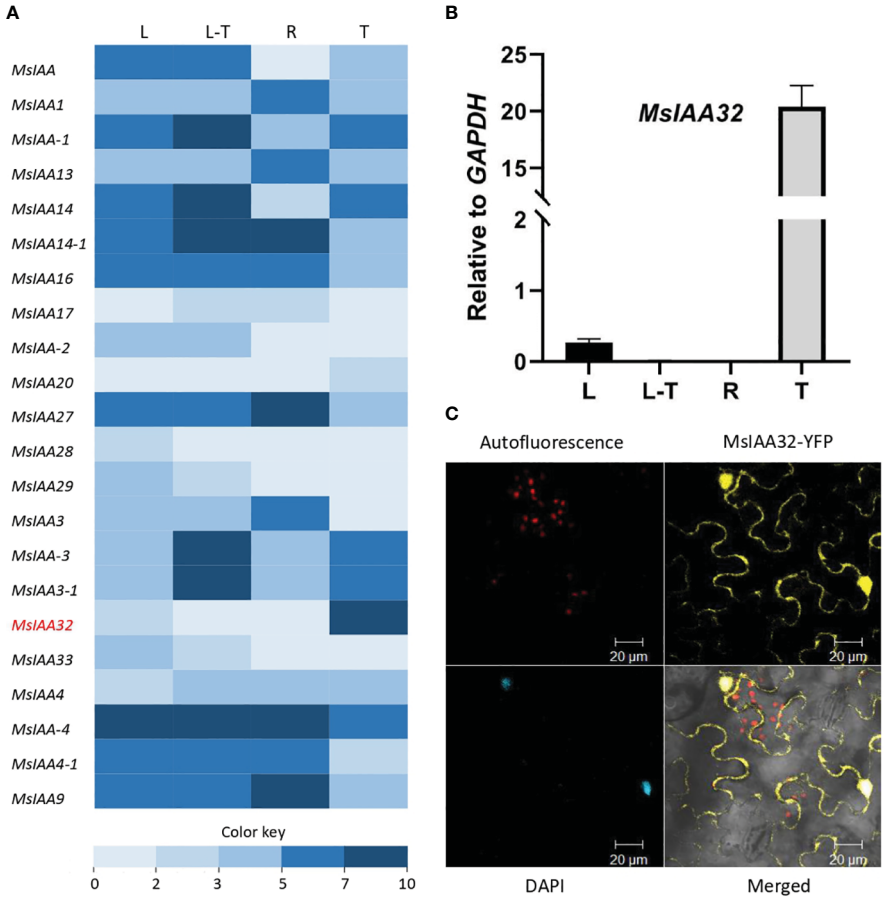


FIGURE 1
Expression analysis of *MsIAA32*. **(A)** A heat map of spearmint *Aux/IAAs* expression in different tissues: PGTs (T), leaf stripped of PGTs (L-T), leaf (L) and roots (R). *MsIAA32* is shown in red and demonstrates high expression in PGTs. Supplementary Table S2 has expression values and sequence data. **(B)** *MsIAA32* shows specific expression in PGTs. **(C)** *N. benthamiana* leaf cells showing nucleus and cytoplasm localization of MsIAA32.

protein was found to be localized in both nucleus and cytoplasm, as seen in Figure 1C. Such extranuclear localization is also observed for SlIAA32 (Audran-Delalande et al., 2012).

3.2 *MsIAA32* promoter shows trichome specific expression

A 729-bp upstream DNA fragment of the translation start site was obtained via genome walking. PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) identified cis-acting regulatory elements in the promoter, including an AuxRE-motif, an auxin-responsive element (Ulmasov et al., 1995). Two variants of the core AuxRE (TGTC) sites with a 5-bp spacer (TGTCTT—–TGTCCC) were present (Figure 2A). ARFs are known to bind as dimers on palindromic AuxREs, which are spaced by 5 to 9 nucleotides (Michal et al., 2019) suggesting a probable regulation of *MsIAA32* by ARFs. Additionally, several TFs and light-responsive elements were also present. The expression pattern of the promoter was studied by pairing it with a β-glucuronidase (GUS) reporter gene and transforming it into *N. benthamiana* plants. Transgenic tobacco leaves displayed trichome-specific staining (Figures 2B, C).

3.3 MsIAA32 expression is induced by auxin

Presence of AuxRE sites in the promoter of *MsIAA32* indicates its possible regulation by auxin. Many *Aux/IAA* gene members are known to be induced by auxin and show the presence of AuxRE sites in its promoter. To test whether *MsIAA32* expression is induced by auxin, 100 mM of 3-Indoleacetic acid was sprayed to one-month-old WT spearmint plants and the expression of *MsIAA32* was analyzed at various time points as previously described (Wu et al., 2012). Increased expression of *MsIAA32* was noted 30 min after auxin treatment (Figure 2D).

3.4 Silencing of *MsIAA32* reduces the number of PGTs formed and monoterpene production in spearmint

To study the function of *MsIAA32*, we created transgenic spearmint lines with an RNAi construct controlled by the CaMV 35S promoter and targeting *MsIAA32*. Thirteen transgenic lines were developed, which were confirmed by the southern blot, of which we selected eight independent single-copy lines for further

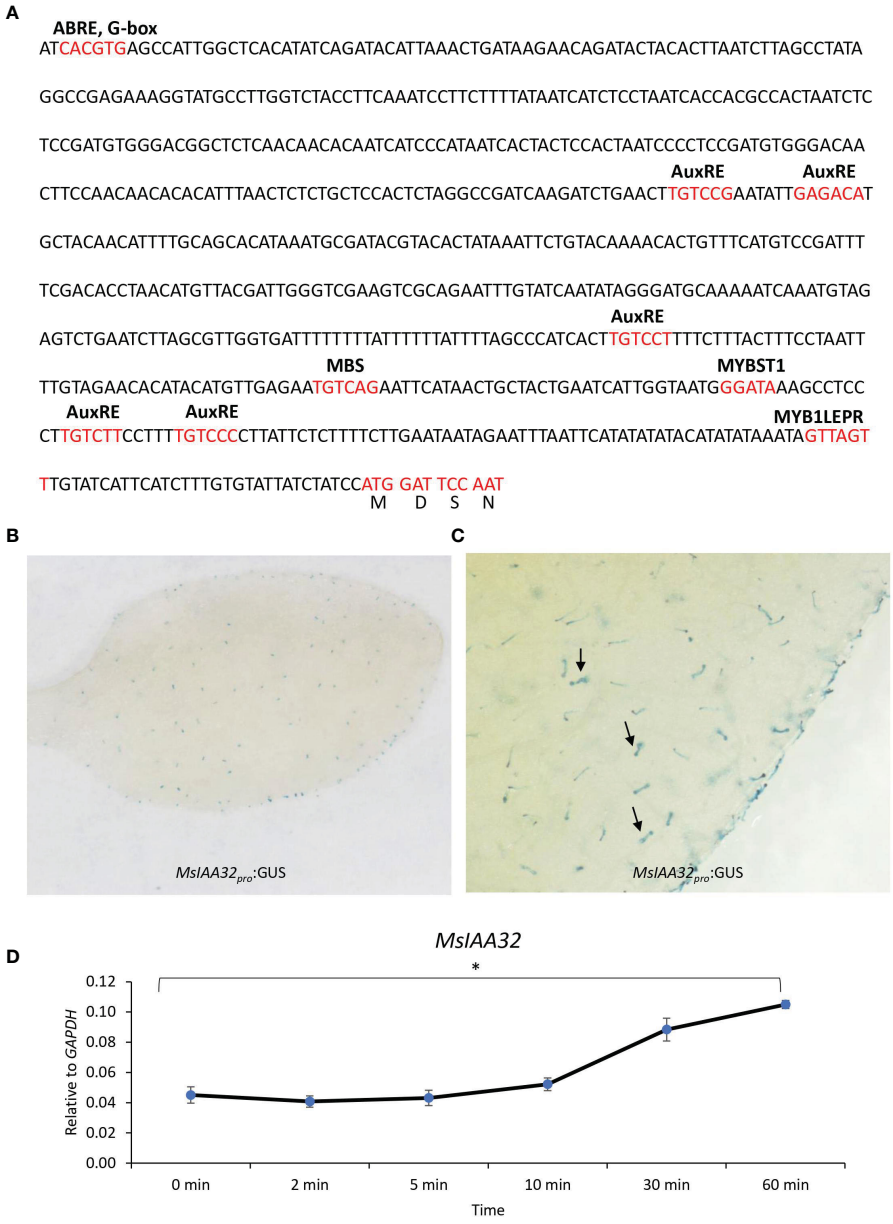


FIGURE 2
MsIAA32 promoter analysis. **(A)** Promoter region of *MsIAA32* contains regulatory elements. **(B, C)** *MsIAA32pro*:GUS exhibit trichome-specific expression in *N. benthamiana* leaves, as indicated by black arrows pointing to GUS-stained trichomes. **(D)** qRT-PCR showing enhanced expression of *MsIAA32* after auxin treatment. *P < 0.05.

analysis (Supplementary Figure S3A). qRT-PCR analysis showed that of the eight chosen lines, six lines displayed reduced levels of *MsIAA32* transcripts when compared to wild type (WT) (Figure 3A). Three best lines (RNAi-6, RNAi-12 and RNAi-13) were chosen for further characterization. The *MsIAA32* RNAi lines had a similar appearance to WT (Supplementary Figure S3B). No change in leaf size or shape was observed (Supplementary Figure S3C). A closer examination of leaf cells and PGTs in these plants was conducted using scanning electron microscopy. The epidermal cells and stomata did not show any significant changes in morphology or number (Figure 3B, Supplementary Figures S3H, I). Compared to WT plants *MsIAA32* RNAi plants had a reduced number of PGTs, but PGT morphology was not affected (Supplementary Figures S3D-G). In WT plants the density of PGTs ranges from 8 to 10 per mm^2 while density of PGTs in *MsIAA32* RNAi was 6 to 7.5 per mm^2. Overall ~ 22-35% reduction in trichome density was observed in *MsIAA32* RNAi lines compared to WT (Figures 3C, D).

In spearmint, PGTs are the site of production and storage of secondary metabolites, primarily monoterpenes (limonene and carvone) (Reddy et al., 2017). To assess any changes in volatile production in *MsIAA32*-RNAi transgenics, a gas chromatography-mass spectrometry (GC-MS) analysis was done, which revealed ~30-40% reduction in limonene and carvone content (Figure 3E). To investigate if downregulation of *MsIAA32* has any effect on the expression levels of the biosynthetic enzymes of the 2-C-methyl-D-

10.3389/fpls.2024.1284125

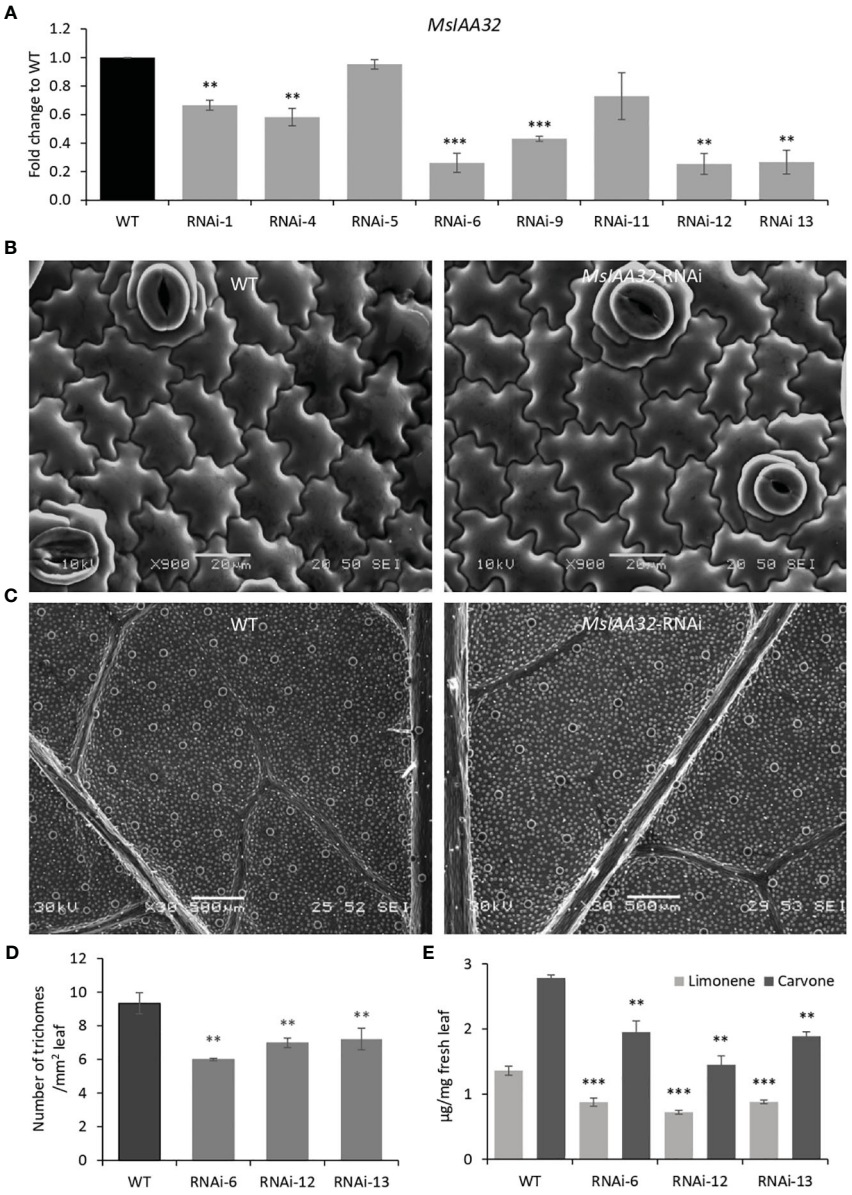


FIGURE 3
Characterization of *MsIAA32*-RNAi lines. **(A)** Reduced levels of *MsIAA32* in *MsIAA32*-RNAi lines. **(B)** SEM images of spearmint leaf showing similar epidermal cell size and number in WT and *MsIAA32*-RNAi line. **(C)** SEM images of spearmint leaf showing lesser PGTs (globular structures) in *MsIAA32*-RNAi line than the WT. Reduced number of PGTs **(D)** and reduced levels of carvone and limonene **(E)** in transgenic spearmint *MsIAA32*-RNAi lines compared to WT. **P < 0.01; ***P < 0.001

erythritol 4-phosphate (MEP) pathway, we did a qRT-PCR of major biosynthetic enzymes of the MEP pathway. As shown in Supplementary Figure S4, the expression of enzymes remains unaltered. This suggests that the observed reduction in secondary metabolites in *MsIAA32* RNAi transgenics is mainly due to a decrease in the number of PGTs.

3.5 MsIAA32 binds with MsIAA3, MsIAA4 and MsARF3

Aux/IAA proteins are known to form homodimers or heterodimers with other Aux/IAAs or ARFs to regulate downstream genes (Piya et al., 2014). To identify the binding partners of MsIAA32, we did a yeast two-hybrid (Y2H) screen where MsIAA32 was used as a bait against the library of clones from spearmint leaf cDNA. Matchmaker Gold Yeast Two-Hybrid System was used for Y2H screen where bait and prey interactions activate transcription of four independent reporter genes (AUR1-C, ADE2, HIS3, and MEL1). MEL-1 encodes a-galactosidase, AUR1-C expression confers strong resistance to antibiotic Aureobasidin A, ADE2 and HIS3 provides the strain the ability to synthesis adenine and histidine respectively. From the list of partners, we identified two Aux/IAAs and one ARF as interacting partners of MsIAA32. Through NCBI blast, we identified and named the two IAAs as MsIAA3 and MsIAA4 and the ARF as MsARF3. All the three

identified partners had expressions in PGTs as seen by their FPKM values within the various tissues of spearmint in RNA-Seq data (Supplementary Figure S5). To further confirm positive interaction between the MsIAA32 and its identified partners, individual Y2H assays between MsIAA32 and its partners were pursued which showed strong interaction between MsIAA32 and MsIAA3, MsIAA32 and MsIAA4 and weak interaction between MsIAA32 and MsARF3 (Figure 4A). MsIAA3 and MsIAA4 did not interact with MsARF3. All the positive interactions were further confirmed using Biomolecular fluorescence complementation (BiFC) assay. MsIAA32 fused to ^{N}YFP and MsIAA3/MsIAA4/MsARF3 fused to ^{C}YFP showed a strong fluorescence signal in the nucleus (Figure 4B). However, no signal was observed in the empty vectors.

MsIAA3 and MsIAA4 sequences had all the conserved domains (I-IV) of a canonical IAA (Supplementary Figure S6). A phylogenetic analysis of MsIAA3 and MsIAA4 with all other IAAs from *Arabidopsis thaliana* and *Solanum lycopersicum* placed MsIAA3 and MsIAA4 under clade A (Supplementary Figure S1). Gain-of-function mutations in AtIAA3 is known to cause short hypocotyls, altered auxin-regulated root development, and enlarged cotyledons (Tian et al., 2002). Recently it was shown that AtIAA3-mediated repression of PHYTOCHROME-INTERACTING FACTORs coordinates light and auxin signalling (Xi et al., 2021). Mutant studies of SlIAA4 has not been pursued, however expression analysis shows it has high expression in fruits and leaves (Audran-Delalande et al., 2012). This highlights that the function of MsIAA3 and MsIAA4 in glandular trichomes development needs to be deciphered.

A phylogenetic tree of MsARF3 with other AtARFs and SlARFs placed MsARF3 with a clade containing SlARF9, SlARF12, SlARF18, AtARF9, AtARF11 and AtARF18 (Supplementary Figure S7A), which are members of class 1a ARFs. This set of ARFs is majorly known to act as negative regulators (Xu et al., 2016). MsARF3 contains all the characteristic domains of an ARF, which are the B3 domain, ARF domain and Aux/IAA domains (Supplementary Figure S7B). In addition, it has a serine-rich middle region which is characteristic of a repressor ARF (Tiwari et al., 2003; Roosjen et al., 2018).

3.6 Auxin spray can induce PGT formation

To investigate the effect of auxin on PGT development, 2,4-Dichlorophenoxyacetic acid (2,4-D) (10 mg/L, 0.045mM) was sprayed every two days for ten days on the apical regions of one-month-old WT spearmint plants. Newly grown leaves were then

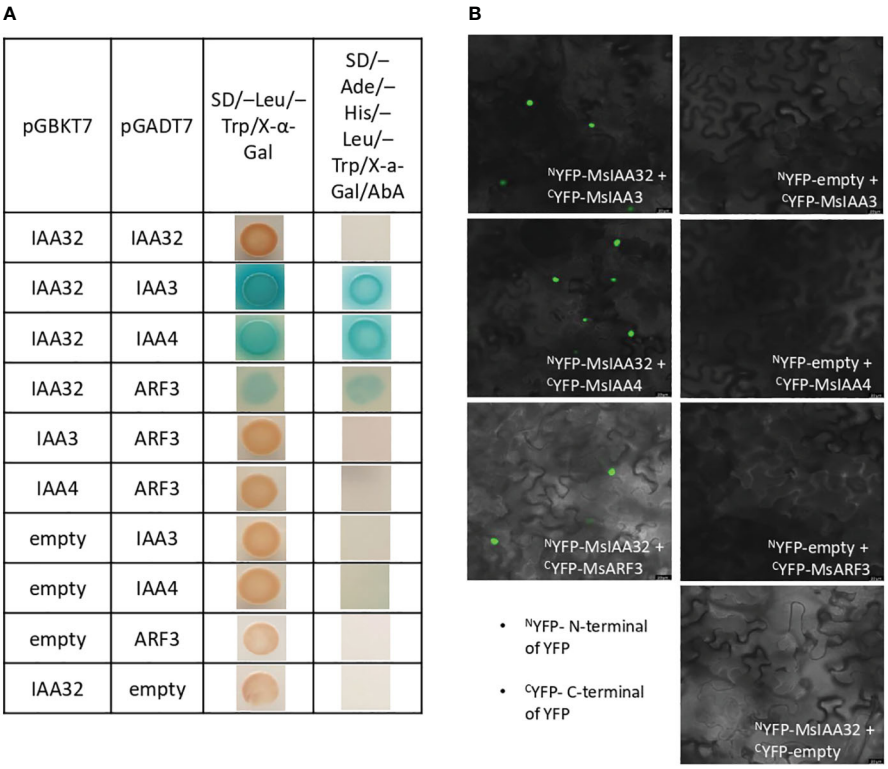


FIGURE 4
Interaction study between MsIAA32 and its partners. **(A)** Yeast two-hybrid assay between MsIAA32 and its partners. First low stringent screening was performed on SD/–Trp/–Leu/X-α-Gal selection medium where positive interactions should result in blue colonies. The positive interactions were further confirmed by growing on SD/–His/–Ade/–Trp/–Leu/AbA/X- α -Gal selection medium for more stringent screening. SD/–Trp/–Leu/X-α-Gal, SD medium without Leu and Trp but containing X-α-Gal (40 µg/ml); SD/–His/–Ade/–Trp/–Leu/AbA/X- α -Gal, SD medium without His, Ade, Trp and Leu but containing aureobasidin A (800 ng/ml) and X-α-Gal (40 µg/ml). **(B)** BiFC analysis of the interaction between MsIAA32 and its partners. The BiFC assays were conducted using the split YFP system. Tobacco epidermal cells were transformed with the different combinations of plasmids, and the YFP signals in the nucleus were observed with a confocal microscope. Scale bars, 20 µm. ^{C}YFP, C-terminal of YFP; ^{N}YFP, N-terminal of YFP.

subjected to SEM analysis and trichome density analysis. We found ~ 60% increase in the number of PGTs in auxin treated leaves when compared to solvent treated leaves (Figures 5A–C). A similar result was also observed previously in tomato plants (Yuan et al., 2021). To examine if auxin spray can rescue the *MsIAA32* RNAi phenotype, a similar spray experiment was performed as mentioned above. Auxin treatment led to the restoration of PGT density to wild type levels in the *MsIAA32* RNAi plants (Figure 5D).

3.7 Expression of a *R2R3-MYB* and a B2-type cyclin gene is altered in *MsIAA32* down-regulated lines

MYB genes are known to regulate trichome development in various plants, and in tomato, trichome enriched MYBs mediate auxin dependent formation of glandular and non-glandular trichomes (Yuan et al., 2021). To elucidate how *MsIAA32* regulates PGT development in spearmint, we mined a list of potential MYB candidates from our RNA-Seq data which had high expression in PGTs. Expression of these MYB genes were analyzed in *MsIAA32* RNAi lines, using qRT-PCR. One R2R3-MYB, *MsMYB36*, was significantly reduced in RNAi plants compared to WT plants (Figure 6A). Sequence analysis showed the presence of conserved features of R2R3-MYBs, a R2, R3 repeat and five tryptophan residues within the R2 and R3 repeats, which together forms a helix-turn-helix motif at the N-terminus (Dubos et al., 2010) (Supplementary Figure S8A). *Arabidopsis* MYBs were used to construct a phylogenetic tree (Supplementary Figure S8B). MsMYB36 fell under the S14 subfamily, members of which have been reported to function in the regulation of axillary meristems and root development (Keller et al., 2006; Müller et al., 2006).

In addition, expression of B2-type cyclins genes was also investigated as recently, a tomato cyclin, *SlCycB2*, and tobacco cyclin, *NtCycB2* were identified to function as inhibitors of glandular and non-glandular trichome formation (Yuan et al., 2021; Zhang et al., 2021; Wang et al., 2022). Additionally, tomato cyclin, *SlCycB2* was also shown to be regulated by MYBs involved in trichome formation. We screened the B2-type cyclins from our RNA-Seq data and found one cyclin, *MsCycB2-4*, which showed expression in PGTs, to have enhanced expression in the *MsIAA32*-RNAi line compared to WT by qRT-PCR (Figure 6B). The overexpression of *MsCycB2-4* might be the reason for the decrease in PGT numbers in *MsIAA32* RNAi lines.

3.8 Ectopic expression of *MsIAA32* causes diverse phenotypes in *Arabidopsis*

To observe if *MsIAA32* has an influence on non-glandular trichome development, it was ectopically expressed using CaMV35S promoter in *Arabidopsis*. We generated multiple transgenic lines and selected the three with the strongest phenotype for further study. *MsIAA32* expression was confirmed by qRT-PCR in these lines (Figure 7A). The transgenic plants exhibited several auxin-related phenotypes. The leaves exhibited

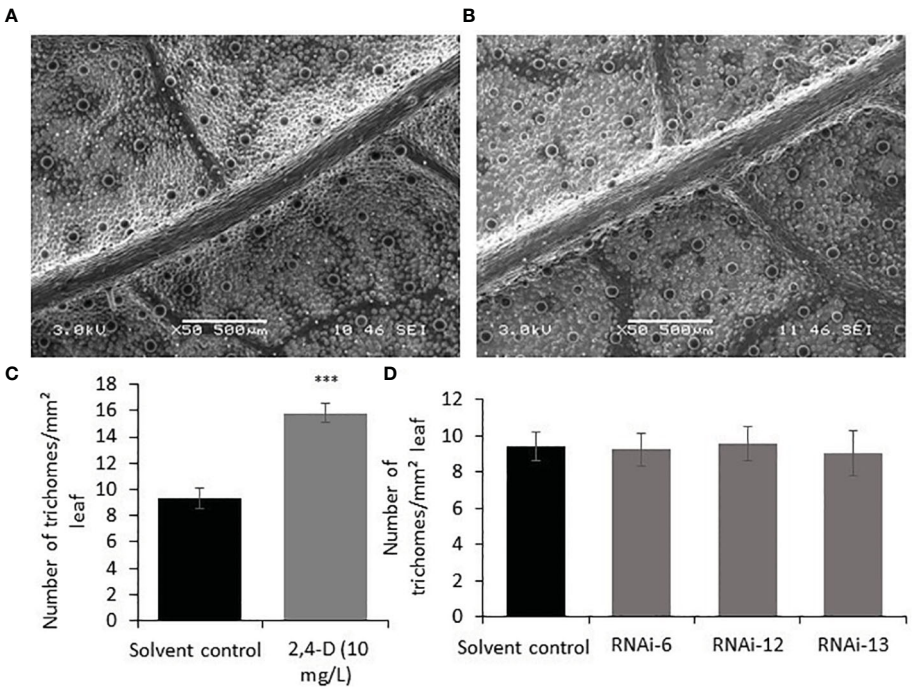


FIGURE 5
Effect of auxin on PGT development and *MsIAA32* expression. **(A)** SEM image of spearmint leaf treated with water **(B)** SEM image of spearmint leaf treated with auxin (10 mg/L 2,4-D, 0.045mM) showing more PGTs. **(C)** Graph depicting the increased number of PGTs in auxin treated leaves compared to solvent treated leaves. **(D)** Graph depicting the recovery of PGT number in *MsIAA32* RNAi lines treated with auxin (10 mg/L 2,4-D, 0.045mM) to levels observed in solvent treated WT leaves. ***P < 0.001.

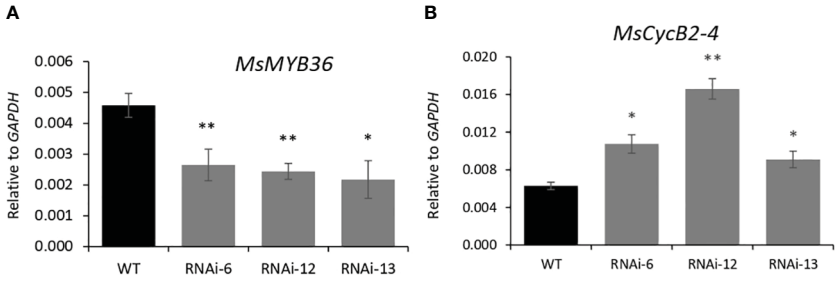


FIGURE 6
Characterization of *MsMYB36* and *MsCycB2-4*. Reduced levels of *MsMYB36* **(A)** and increased levels of *MsCycB2-4* **(B)** in *MsIAA32*-RNAi lines. *P < 0.05; **P < 0.01.

inward curling (Figure 7B). The leaves and the rosette diameter were smaller (Figures 7C, D).

The seedlings were smaller with reduced lateral roots (Figures 8A, B). Phenotypically adult transgenic plants looked similar in stature to WT plants. A closer examination of leaf cells and PGTs in these plants was conducted using scanning electron microscopy. The epidermal cells had no change in size or number (Figure 9A), however, the trichome density was increased by ~ 60% in the transgenic lines, with no effect on the size and shape of trichomes (Figures 9B–D). To test whether any B-type cyclin is involved in the trichome phenotype observed in *Arabidopsis* transgenics, the ortholog of *MsCycB2-4* was identified, and the expression analysis was performed using q-RT PCR. As expected, *AtCycB2-4* (NM_106281.6) had reduced expression in *Arabidopsis* transgenics when compared to WT, which might be a possible reason for the increase in the number of trichomes in transgenics since CycB2-4 appears to be a negative regulator of trichome initiation (Figure 9E) (Yuan et al., 2021; Zhang et al., 2021; Wang et al., 2022).

4 Discussion

The role of canonical Aux/IAAs in plant development and growth through auxin-regulated gene expression is well studied. However, with the identification of non-canonical Aux/IAAs that lack one or more of the four conserved domains, the presence of an alternate non-canonical auxin signalling pathway is now established. The majority of our knowledge in understanding the functions of *Aux/IAA* genes in plant growth and development comes from research on *Arabidopsis*, revealing extensive overlap in function among its *Aux/IAA* gene family. However, studies in Solanaceae demonstrate specialized functions of these genes, highlighting the need to study them in diverse plants for a deeper understanding of their significance. *Arabidopsis* contain six non-canonical *Aux/IAA* genes (*IAA34*, *IAA33*, *IAA32*, *IAA31*, *IAA30*, and *IAA20*), while tomato has only two non-canonical *Aux/IAA* genes (*Sl-IAA33* and *Sl-IAA32*), which lack typical domains I and II (Audran-Delalande et al., 2012). In this study, we identified a non-canonical *Aux/IAA* gene *MsIAA32* from spearmint involved in PGT development. Phylogenetic analysis puts MsIAA32 in the same clade as Sl-IAA32 and AtIAA32, with which it exhibits 62% and 46% sequence similarity. *SlIAA32* mutants have not been characterized, but its expression is limited to fruits and is shown to function as a repressor of auxin signalling (Audran-Delalande et al., 2012). Single mutants of *AtIAA32* show no obvious phenotype, but along with *At1AA34*, it redundantly controls apical hook maintenance (Cao et al., 2019). Compared to canonical auxin signalling, insights into the mechanisms of non-canonical Aux/IAA mediated auxin signalling remains limited.

In typical canonical auxin signaling, auxin induces degradation of Aux/IAAs protein, thus releasing the inhibition of ARF transcription factors, leading to transcriptional reprogramming of auxin responsive genes. Interestingly many Aux/IAAs possess AuxRE binding sites to which ARFs bind, hence their transcription is concomitantly upregulated by auxin, thus establishing a negative feedback loop to generate an appropriate auxin response (Guilfoyle and Hagen, 2007). In contrast, studies in non-canonical Aux/IAAs, AtIAA32, AtIAA33, and AtIAA34 have shown that auxin promotes their protein accumulation through phosphorylation which increases their stability (Cao et al., 2019). It was also shown that auxin treatment had no effect on the transcription of *AtIAA33* (Lv et al., 2020). However expression of non-canonical *Aux/IAAs* like *AtIAA20* and *AtIAA30* were shown to be induced by auxin (Sato and Yamamoto, 2008). But increased levels of auxin did not have any effect on IAA20 protein longevity (Dreher et al., 2006). In summary, unlike canonical Aux/IAAs where auxin mediates the destruction of their proteins and generally upregulates their transcription, non-canonical Aux/IAAs response to auxin can vary at transcriptional and protein level. Auxin treatment was able to increase the expression of *MsIAA32* but the influence of auxin on MsIAA32 protein needs to be investigated. Lack of domain II should render MsIAA32 protein insensitive to auxin mediated degradation hence their levels should probably remain unchanged under high auxin. In our study, auxin treatment was able to rescue the *MsIAA32* RNAi phenotype. This might be probably due to the increased transcription resulting in increased protein levels in RNAi plants leading to PGT number restoration. Further studies need to be performed with regards to MsIAA32 steady state levels and influence of auxin on its protein stability to better understand the role of *MsIAA32* in auxin mediated PGT development.

Generally, Aux/IAAs are known to possess two kinds of putative nuclear localization signals, with a basic residue-rich

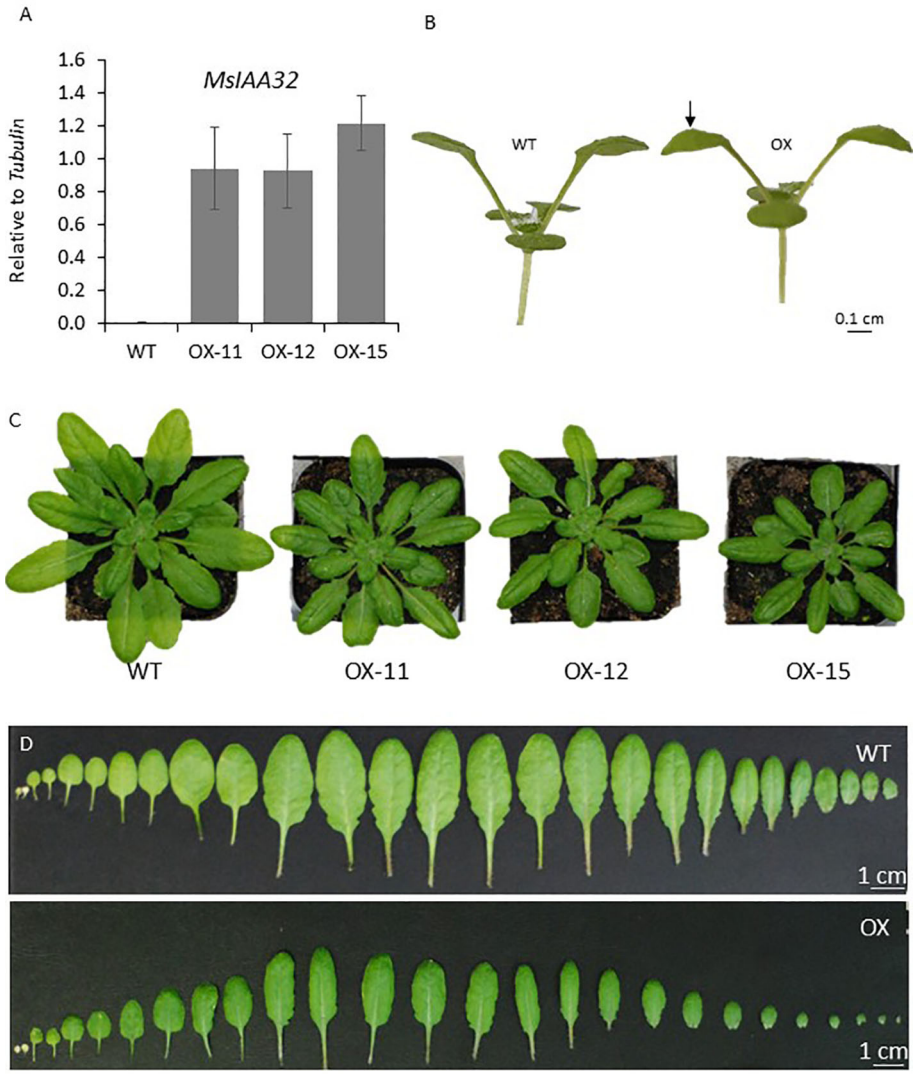


FIGURE 7
Characterization of *Arabidopsis* plants expressing *MsIAA32*. **(A)** Ectopic expression of *MsIAA32* in *Arabidopsis* plants. **(B)** *MsIAA32* expressing *Arabidopsis* plant showing epinasty. The black arrow denotes the downward curling of the leaf. **(C)** *MsIAA32* expressing *Arabidopsis* plants are visibly smaller in size when compared to WT. **(D)** Cotyledons and leaves excised from WT and *MsIAA32* expressing *Arabidopsis*. The two cotyledons are shown on the left, followed by the vegetative leaves in the order they initiated. WT, Wildtype; OX, *MsIAA32* expressing *Arabidopsis* line.

SV40-type NLS in domain IV and a bipartite structure between domains I and II with a conserved basic doublet "KR" (Ke et al., 2019). However, MsIAA32 only has the SV40-type NLS and lacks the bipartite NLS, which might be the reason for its extended localization in the cytosol in addition to the nucleus. The lack of Domain II, which is involved in protein degradation, may also contribute towards the extended localization. Similar subcellular localization to nucleus and cytoplasm was also observed for SlIAA32, AtIAA32, and AtIAA34 (Audran-Delalande et al., 2012; Cao et al., 2019). The nuclear localization indicates a transcriptional regulatory function of MsIAA32, and the presence of MsIAA32 in cytoplasm suggests probable novel interactions with binding factors specifically available in the cytoplasm.

From yeast two-hybrid studies, it was evident that MsIAA32 prefers binding with MsIAA3 or MsIAA4 (Figure 4A), both of which formed a clade with canonical AtIAA3 in phylogenetic analysis (Supplementary Figure S1). AtIAA3 was previously identified to bind with activator ARFs (Piya et al., 2014) and is involved in light and auxin signalling in *Arabidopsis* (Xi et al., 2021). Aux/IAA proteins are known to interact with other Aux/IAA proteins but the biological significance of these interactions in auxin signalling has not been well elucidated. In *Arabidopsis*, an interactome map of all the 29 Aux/IAA proteins was constructed, which shows that the 29 *Aux/IAA* genes could interact with each other via 253 interactions (Luo et al., 2018). Given that MsIAA32 is a non-canonical Aux/IAA hence more stable, these types of interactions among canonical and non-canonical Aux/IAA proteins probably add more flexibility in fine tuning the activity of the auxin signalling network. Further research is required to explore the biological significance of interactions between non-canonical MsIAA32 and canonical MsIAA3/MsIAA4 towards PGT development. MsIAA32 was also able to bind with MsARF3. ARFs

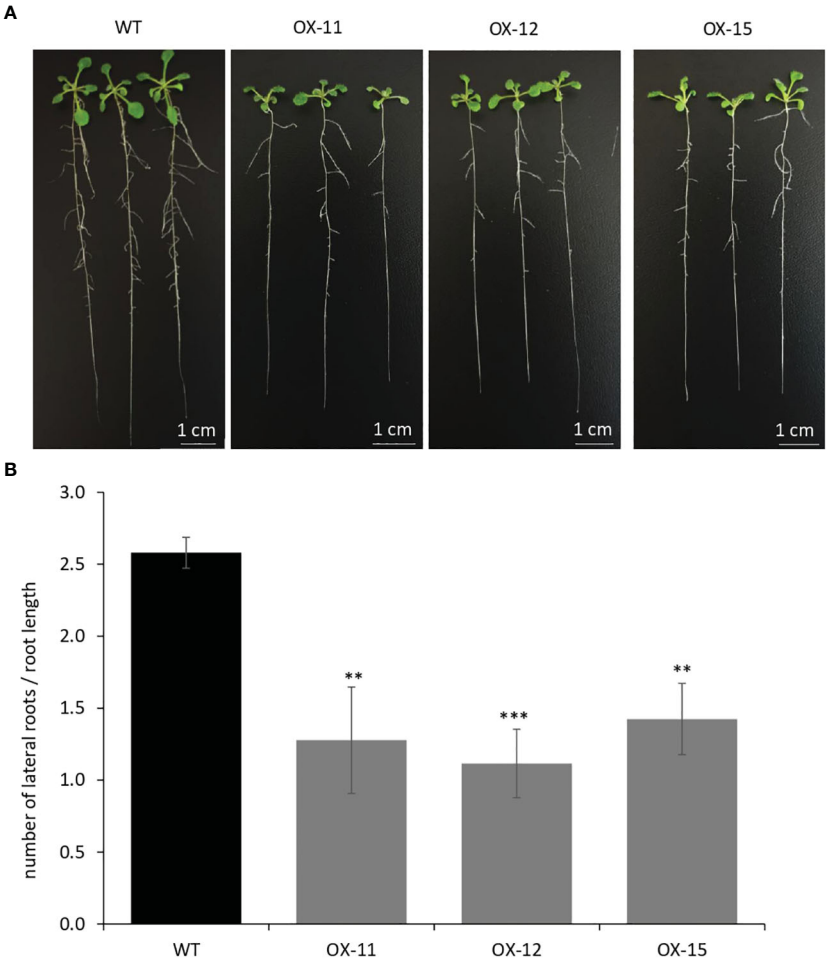


FIGURE 8
Quantification of lateral roots in *MsIAA32* expressing *Arabidopsis* plants. **(A, B)** *MsIAA32* expressing *Arabidopsis* plants showing reduced lateral roots when compared to plants 12 days after germination. WT, Wildtype; OX, *MsIAA32* expressing *Arabidopsis* line. **P < 0.01; ***P < 0.001

play a crucial role in defining auxin response specificity by regulating the expression of transcription factors that depend on auxin. ARFs are divided into three conserved classes: A, B, and C. Class A ARFs act as transcriptional activators, while class B and C ARFs function as transcriptional repressors (Chung et al., 2019). MsARF3 is a class B repressor ARF which is supported by its serine-rich middle region, a characteristic feature of repressor ARFs (Roosjen et al., 2018) (Supplementary Figure S7B).

Studies in Tomato have revealed the importance of auxin mediated signalling in trichome formation. Tomato leaves possess eight different types of non-glandular and glandular trichomes (McDowell et al., 2010). In tomato, two ARFs have been characterized which are involved in trichome formation in leaves, SlARF3 was found to positively regulate the density of glandular type I and VI and non-glandular type V trichomes (Zhang et al., 2015), and SlARF4 was shown to positively regulate the auxin-induced formation of II, V and VI type trichomes. It was determined that SlARF4 regulates trichome development through repressing two R2R3-MYB genes, *SlTHM1* and *SlMYB52*. These MYB genes, in turn, controlled trichome formation through the regulation of the B-type cyclin gene *SlCycB2* (Yuan et al., 2021). Studies in cell cycle genes, including cyclins in tomato and *Arabidopsis*, have shown that the transition from the mitotic phase to the endoreduplication cycle is key for the development of both unicellular and multicellular trichomes (Fambrini and Pugliesi, 2019). In tomato *SlCycB2*, is highly expressed in the trichomes and negatively regulates the formation of unicellular and multicellular trichomes. Recently SlMYB75 was identified to regulate trichome formation in tomato where SlMYB75 directly targeted the promoters of *SlCycB2*, *SlTHM1* and *SlMYB52* (Gong et al., 2021). Further, it was found that SlARF4 interacts with *SlMYB75* promoter and inhibits its expression. Taken together, it was proposed that in tomato, a *SlARF/SlMYB/SlCycB* transcriptional cascade controls the formation of type II, V, and VI trichomes. Regulation of these ARFs by Aux/IAAs is not known in tomato. The only other characterized *Aux/IAA* gene involved in trichome development is the canonical *Aux/IAA* gene *SlIAA15* from tomato. The down regulation of *SlIAA15* affected the development of both non-glandular and glandular types of trichomes. Given its ubiquitous expression pattern, downregulation of *SlIAA15* also showed diverse phenotypes like altered stem xylem development, thickened leaves, decreased fruit set, increased lateral root and reduced apical dominance revealing its multiple functions in plant

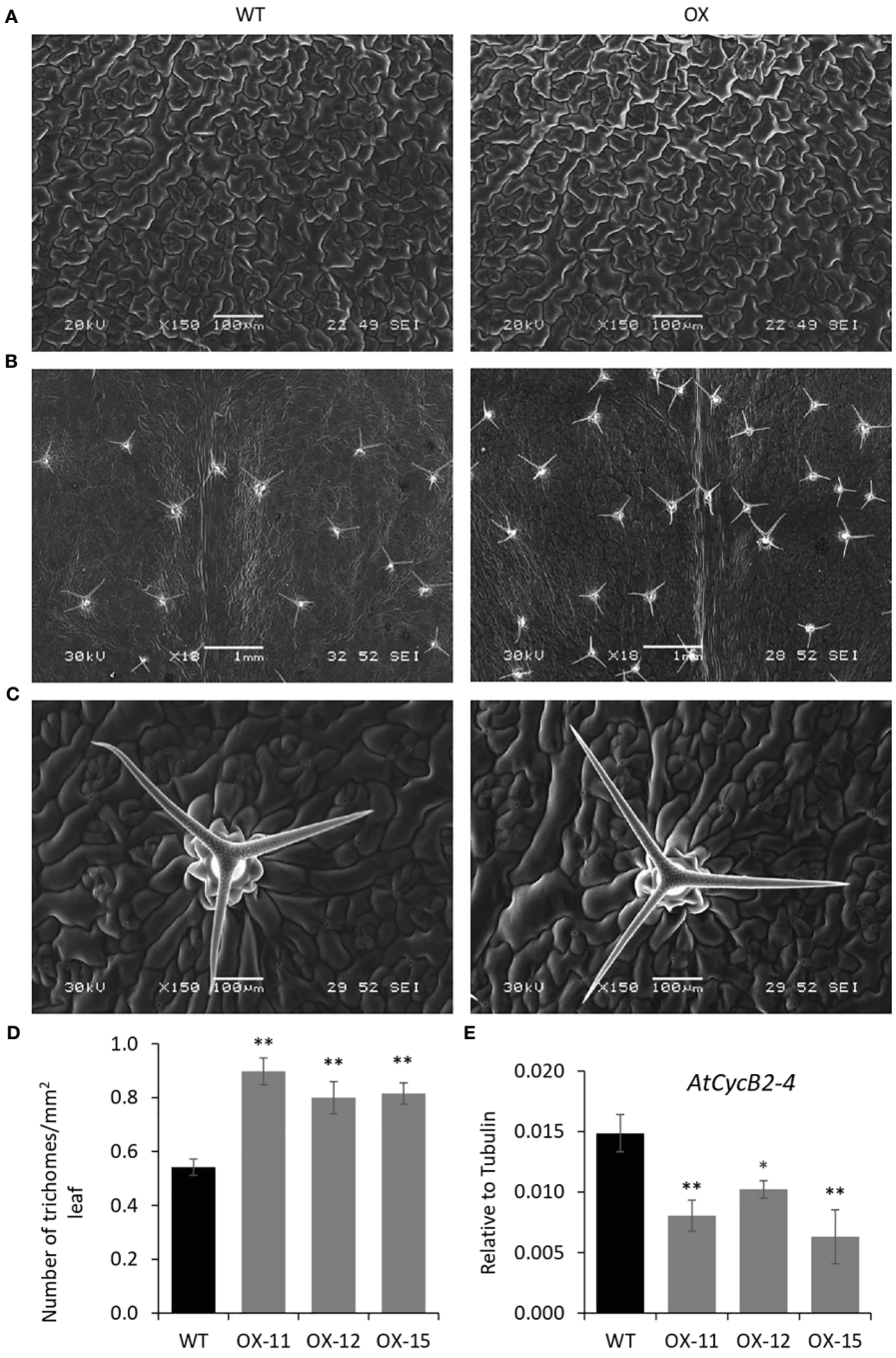


FIGURE 9
Quantification of trichomes in *MsIAA32* expressing *Arabidopsis* plants. **(A)** SEM images of *Arabidopsis* leaves showing similar size and number of epidermal cells. **(B)** SEM images of *Arabidopsis* leaves showing an increased number of trichomes. **(C)** SEM image of WT leaf and *MsIAA32* expressing leaf showing similar PGT shape and size. **(D)** Increased number of trichomes in *MsIAA32* expressing *Arabidopsis* plants compared to WT. **(E)** Reduced expression of *AtCycB2-4* in *MsIAA32* expressing *Arabidopsis* plants. WT, Wildtype; OX, *MsIAA32* expressing *Arabidopsis* line. *P < 0.05; **P < 0.01.

development (Deng et al., 2012a; Deng et al., 2012b). Interestingly, the down regulation of *SlIAA15* resulted in the reduction in expression of few R2R3-MYB genes including *SlTHM1*gene. It was proposed that *SlIAA15* regulates trichome formation via direct or indirect control of MYB genes. Later in a different study looking into tomato fruit development showed that SlARF4, can interact with SlIAA15 in a BiFC assay (Shinozaki et al., 2018). Similar to the studies in tomato, MsIAA32 downregulation results in misexpression of a R2R3-MYB gene *MsMYB36* and B-type cyclin gene *MsCycB2-4* suggesting their involvement in PGT development. The exact mechanism of MsIAA32 mediated regulation of the possible *ARF/MYB/CycB* transcriptional cascade in spearmint towards PGT development through its binding partners needs further exploration. Role of *MsIAA32* in glandular trichome development highlights species-specific distinct and novel functions of Aux/IAA family members sustained by cell/tissue-specific expression patterns.

Apart from increased trichome number, overexpression of *MsIAA32* in Arabidopsis resulted in developmental phenotypes like downward leaf curling, reduced leaf size, smaller rosette diameter and decreased lateral root growth. These phenotypes have been associated with alteration in auxin homeostasis and signaling including misexpression of *Aux/IAA* genes (Rogg et al., 2001; Sato and Yamamoto, 2008; Arase et al., 2012; Rinaldi et al., 2012; Guseman et al., 2015; Liu et al., 2015; Sandalio et al., 2016; Xiong and Jiao, 2019). Recently overexpression of a non-canonical *Aux/IAA* gene *EgrIAA20* from *Eucalyptus* in *Arabidopsis* showed leaf curling, reduced leaf size and reduced diameter of rosette similar to *MsIAA32* along with other phenotypes (Yu et al., 2022). *Arabidopsis Aux/IAA* genes show high functional redundancy, and it has been suggested that *Aux/IAA* genes can have both similar and specific functions. The interactome map of Aux/IAA proteins in *Arabidopsis* shows that 29 *Aux/IAA* genes can interact with each other via 253 interactions and with 20 ARF proteins via 544 interactions (Luo et al., 2018). When *MsIAA32* is constitutively expressed in different tissues, the pleiotropic effects observed in *Arabidopsis* might be due to novel interactions with different sets of tissue-specific Aux/IAA or ARF proteins leading to changes in auxin perception. In *Arabidopsis*, *MsIAA32* could regulate non-glandular trichome formation, possibly through cyclin *AtCycB2-4*. Although the mutant phenotype of the *AtCycB2-4* gene in *Arabidopsis* is not available, other cyclin genes have been characterized which are involved in trichome formation (Schnittger et al., 2002). The effect of *MsIAA32* expression on other cyclin genes involved in trichome formation in *Arabidopsis* remains to be investigated. The ability of *MsIAA32* to control both non-glandular and glandular trichome formation indicates shared non-canonical Aux/IAA regulated auxin signaling mechanism among them. In contrast, *SlIAA15* studies indicate the involvement of canonical Aux/IAA mediated auxin signal pathways in the development of non-glandular and glandular trichomes. It is possible that both canonical and noncanonical Aux/IAA signaling pathways are connected and interact to fine tune auxin mediated development of trichomes in plants. Further research is required to understand such interactions and molecular mechanism of auxin triggered trichome development.

Our finding uncovers a new role for non-canonical *Aux/IAA* genes in glandular trichomes' development in a non-model plant, spearmint, emphasizing the need to study these genes in different plant species for a deeper understanding of their functions in plant development. Secondary metabolites produced by glandular trichomes are important for plants' response to abiotic and biotic stress and are economically important. Elucidating the genetic machinery underlying their formation will help to regulate their density and productivity.

Data availability statement

The datasets presented in this study can be found in online repositories. Supplementary Table S2 contains sequence data for all *MsIAAs* in Figure 1A. Sequences of *A. thaliana* and *Solanum lycopersicum* Aux/IAAs used in Supplementary Figure S1 can be retrieved by the accession numbers shown in Supplementary Table S3 from the NCBI database. *A. thaliana* and *Solanum lycopersicum* ARFs in Supplementary Figure S7A can be accessed in NCBI with accession numbers shown in Supplementary Table S4. *A. thaliana* R2R3-MYBs in Supplementary Figure S8B can be accessed in NCBI with accession numbers shown in Supplementary Table S5. The sequences of *MsIAA32, MsIAA3, MsIAA4, MsARF3* and *MsMYB36* are available in NCBI under sequence IDs MZ971168, MZ971169, MZ971170, MZ971171 and MZ971172 respectively.

Author contributions

VR: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. JS: Data curation, Investigation, Writing – review & editing. KN: Methodology, Writing – review & editing. RS: Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Temasek Life Sciences Laboratory.

Acknowledgments

We are grateful to Temasek Life Sciences Laboratory for their central facility's support on confocal and scanning electron microscopy.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1284125/full#supplementary-material

References

Arase, F., Nishitani, H., Egusa, M., Nishimoto, N., Sakurai, S., Sakamoto, N., et al. (2012). IAA8 involved in lateral root formation interacts with the TIR1 auxin receptor and ARF transcription factors in Arabidopsis. *PloS One* 7, e43414. doi: 10.1371/journal.pone.0043414

Audran-Delalande, C., Bassa, C., Mila, I., Regad, F., Zouine, M., and Bouzayen, M. (2012). Genome-wide identification, functional analysis and expression profiling of the aux/IAA gene family in tomato. *Plant Cell Physiol.* 53, 659–672. doi: 10.1093/pcp/pcs022

Cao, M., Chen, R., Li, P., Yu, Y., Zheng, R., Ge, D., et al. (2019). TMK1-mediated auxin signalling regulates differential growth of the apical hook. *Nature* 568, 240–243. doi: 10.1038/s41586-019-1069-7

Champagne, A., and Boutry, M. (2013). Proteomic snapshot of spearmint (Mentha spicata L.) leaf trichomes: A genuine terpenoid factory. *PROTEOMICS* 13, 3327–3332. doi: 10.1002/pmic.201300280

Chen, H., Ma, B., Zhou, Y., He, S. J., Tang, S. Y., Lu, X., et al. (2018). E3 ubiquitin ligase SOR1 regulates ethylene response in rice root by modulating stability of Aux/IAA protein. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4513–4518. doi: 10.1073/pnas.1719387115

Chung, Y., Zhu, Y., Wu, M.-F., Simonini, S., Kuhn, A., Armenta-Medina, A., et al. (2019). Auxin Response Factors promote organogenesis by chromatin-mediated repression of the pluripotency gene SHOOTMERISTEMLESS. *Nat. Commun.* 10, 886. doi: 10.1038/s41467-019-08861-3

Deng, W., Yan, F., Liu, M., Wang, X., and Li, Z. (2012a). Down-regulation of SlIAA15 in tomato altered stem xylem development and production of volatile compounds in leaf exudates. *Plant Signaling Behav.* 7, 911–913. doi: 10.4161/psb.20723

Deng, W., Yang, Y., Ren, Z., Audran-Delalande, C., Mila, I., Wang, X., et al. (2012b). The tomato SlIAA15 is involved in trichome formation and axillary shoot development. *New Phytol.* 194, 379–390. doi: 10.1111/j.1469-8137.2012.04053.x

Deshaies, R. J. (1999). SCF and cullin/RING H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* 15, 435–467. doi: 10.1146/annurev.cellbio.15.1.435

Dreher, K. A., Brown, J., Saw, R. E., and Callis, J. (2006). The Arabidopsis Aux/IAA protein family has diversified in degradation and auxin responsiveness. *Plant Cell* 18, 699–714. doi: 10.1105/tpc.105.039172

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

Fambrini, M., and Pugliesi, C. (2019). The dynamic genetic-hormonal regulatory network controlling the trichome development in leaves. *Plants (Basel Switzerland)* 8, 253. doi: 10.3390/plants8080253

Gallois, P., and Marinho, P. (1995). Leaf disk transformation using Agrobacterium tumefaciens-expression of heterologous genes in tobacco. *Methods Mol. Biol.* 49, 39–48. doi: 10.1385/0-89603-321-X:39

Gong, Z., Luo, Y., Zhang, W., Jian, W., Zhang, L., Gao, X., et al. (2021). A SlMYB75-centred transcriptional cascade regulates trichome formation and sesquiterpene accumulation in tomato. *J. Exp. Bot.* 72, 3806–3820. doi: 10.1093/jxb/erab086

Guilfoyle, T. J., and Hagen, G. (2007). Auxin response factors. *Curr. Opin. Plant Biol.* 10, 453–460. doi: 10.1016/j.pbi.2007.08.014

Guo, Z., Hao, K., Lv, Z., Yu, L., Bu, Q., Ren, J., et al. (2023). Profiling of phytohormone-specific microRNAs and characterization of the miR160-ARF1 module involved in glandular trichome development and artemisinin biosynthesis in Artemisia annua. *Plant Biotechnol. J.* 21, 591–605. doi: 10.1111/pbi.13974

Guseman, J. M., Hellmuth, A., Lanctot, A., Feldman, T. P., Moss, B. L., Klavins, E., et al. (2015). Auxin-induced degradation dynamics set the pace for lateral root development. *Development* 142, 905–909. doi: 10.1242/dev.117234

Jin, J., Panicker, D., Wang, Q., Kim, M. J., Liu, J., Yin, J.-L., et al. (2014). Next generation sequencing unravels the biosynthetic ability of Spearmint (Mentha spicata) peltate glandular trichomes through comparative transcriptomics. *BMC Plant Biol.* 14, 292. doi: 10.1186/s12870-014-0292-5

Ke, Y., Abbas, F., Zhou, Y., Yu, R., Yue, Y., Li, X., et al. (2019). Genome-wide analysis and characterization of the aux/IAA family genes related to floral scent formation in hedychium coronarium. *Int. J. Mol. Sci.* 20, 3235. doi: 10.3390/ijms20133235

Keller, T., Abbott, J., Moritz, T., and Doerner, P. (2006). Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 controls a leaf axil stem cell niche and modulates vegetative development. *Plant Cell* 18, 598–611. doi: 10.1105/tpc.105.038588

Lange, B. M., and Turner, G. W. (2013). Terpenoid biosynthesis in trichomes-current status and future opportunities. *Plant Biotechnol. J.* 11, 2–22. doi: 10.1111/j.1467-7652.2012.00737.x

Li, J., Wang, X., Jiang, R., Dong, B., Fang, S., Li, Q., et al. (2021). Phytohormone-based regulation of trichome development. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.734776

Li, S.-B., Xie, Z.-Z., Hu, C.-G., and Zhang, J.-Z. (2016). A review of auxin response factors (ARFs) in plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00047

Liu, S., Hu, Q., Luo, S., Li, Q., Yang, X., Wang, X., et al. (2015). Expression of wild-type PtrIAA14.1, a poplar Aux/IAA gene causes morphological changes in Arabidopsis. *Front. Plant Sci.* 6, 388. doi: 10.3389/fpls.2015.00388

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

Luo, J., Zhou, J.-J., and Zhang, J.-Z. (2018). Aux/IAA gene family in plants: molecular structure, regulation, and function. *Int. J. Mol. Sci.* 19, 259. doi: 10.3390/ijms19010259

Lv, B., Yu, Q., Liu, J., Wen, X., Yan, Z., Hu, K., et al. (2020). Non-canonical AUX/IAA protein IAA33 competes with canonical AUX/IAA repressor IAA5 to negatively regulate auxin signaling. *EMBO J.* 39, e101515. doi: 10.15252/embj.2019101515

McDowell, E. T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.-H., Descour, A., et al. (2010). Comparative functional genomic analysis of solanum glandular trichome types. *Plant Physiol.* 155, 524–539. doi: 10.1104/pp.110.167114

Michal, L.-L., Chen, Y., Alon, I., and Idan, E. (2019). Deep conservation of response element variants regulating plant hormonal responses. *Plant Cell.* 31 (11), 2559–2572. doi: 10.1105/tpc.19.00129

Müller, C. J., Valdés, A. E., Wang, G., Ramachandran, P., Beste, L., Uddenberg, D., et al. (2016). PHABULOSA mediates an auxin signaling loop to regulate vascular patterning in arabidopsis. *Plant Physiol.* 170, 956–970. doi: 10.1104/pp.15.01204

Müller, D., Schmitz, G., and Theres, K. (2006). Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in Arabidopsis. *Plant Cell* 18, 586–597. doi: 10.1105/tpc.105.038745

Nemhauser, J. L. (2018). Back to basics: what is the function of an Aux/IAA in auxin response? *New Phytol.* 218, 1295–1297. doi: 10.1111/nph.15172

Piya, S., Shrestha, S. K., Binder, B., Stewart, C. N. Jr., and Hewezi, T. (2014). Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. *Front. Plant Sci.* 5, 744. doi: 10.3389/fpls.2014.00744

Ramos, J. A., Zenser, N., Leyser, O., and Callis, J. (2001). Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. *Plant Cell* 13, 2349. doi: 10.1105/tpc.010244

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

Reed, J. W. (2001). Roles and activities of Aux/IAA proteins in Arabidopsis. *Trends Plant Sci.* 6, 420–425. doi: 10.1016/S1360-1385(01)02042-8

Rinaldi, M. A., Liu, J., Enders, T. A., Bartel, B., and Strader, L. C. (2012). A gain-of-function mutation in IAA16 confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Mol. Biol.* 79, 359–373. doi: 10.1007/s11103-012-9917-y

Rogg, L. E., Lasswell, J., and Bartel, B. (2001). A gain-of-function mutation in IAA28 suppresses lateral root development. *Plant Cell* 13, 465–480. doi: 10.1105/tpc.13.3.465

Roosjen, M., Paque, S., and Weijers, D. (2018). Auxin Response Factors: output control in auxin biology. *J. Exp. Bot.* 69, 179–188. doi: 10.1093/jxb/erx237

Sandalio, L. M., Rodriguez-Serrano, M., and Romero-Puertas, M. C. (2016). Leaf epinasty and auxin: A biochemical and molecular overview. *Plant Sci.* 253, 187–193. doi: 10.1016/j.plantsci.2016.10.002

Sato, A., and Yamamoto, K. T. (2008). Overexpression of the non-canonical Aux/IAA genes causes auxin-related aberrant phenotypes in Arabidopsis. *Physiol. Plant.* 133, 397–405. doi: 10.1111/j.1399-3054.2008.01055.x

Schnittger, A., Schöbinger, U., Stierhof, Y.-D., and Hülskamp, M. (2002). Ectopic B-type cyclin expression induces mitotic cycles in endoreduplicating arabidopsis trichomes. *Curr. Biol.* 12, 415–420. doi: 10.1016/S0960-9822(02)00693-0

Shimizu-Mitao, Y., and Kakimoto, T. (2014). Auxin sensitivities of all arabidopsis aux/IAAs for degradation in the presence of every TIR1/AFB. *Plant Cell Physiol.* 55, 1450–1459. doi: 10.1093/pcp/pcu077

Shinozaki, Y., Nicolas, P., Fernandez-Pozo, N., Ma, Q., Evanich, D. J., Shi, Y., et al. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nat. Commun.* 9, 364. doi: 10.1038/s41467-017-02782-9

Tian, Q., Uhlir, N. J., and Reed, J. W. (2002). Arabidopsis SHY2/IAA3 inhibits auxin-regulated gene expression. *Plant Cell* 14, 301–319. doi: 10.1105/tpc.010283

Tiwari, S. B., Hagen, G., and Guilfoyle, T. (2003). The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* 15, 533–543. doi: 10.1105/tpc.008417

Turner, G. W., Gershenzon, J., and Croteau, R. B. (2000). Distribution of peltate glandular trichomes on developing leaves of peppermint. *Plant Physiol.* 124, 655–664. doi: 10.1104/pp.124.2.655

Ulmasov, T., Liu, Z. B., Hagen, G., and Guilfoyle, T. J. (1995). Composite structure of auxin response elements. *Plant Cell* 7, 1611–1623. doi: 10.1105/tpc.7.10.1611

Wang, Z., Yan, X., Zhang, H., Meng, Y., Pan, Y., and Cui, H. (2022). NtCycB2 negatively regulates tobacco glandular trichome formation, exudate accumulation, and aphid resistance. *Plant Mol. Biol.* 108, 65–76. doi: 10.1007/s11103-021-01222-z

Watts, S., and Kariyat, R. (2021). Morphological characterization of trichomes shows enormous variation in shape, density and dimensions across the leaves of 14 Solanum species. *AoB Plants* 13, plab071. doi: 10.1093/aobpla/plab071

Wu, J., Liu, S., He, Y., Guan, X., Zhu, X., Cheng, L., et al. (2012). Genome-wide analysis of SAUR gene family in Solanaceae species. *Gene* 509, 38–50. doi: 10.1016/j.gene.2012.08.002

Wu, W., Liu, Y., Wang, Y., Li, H., Liu, J., Tan, J., et al. (2017). Evolution analysis of the aux/IAA gene family in plants shows dual origins and variable nuclear localization signals. *Int. J. Mol. Sci.* 18, 2107. doi: 10.3390/ijms18102107

Xi, Y., Yang, Y., Yang, J., Zhang, X., Pan, Y., and Guo, H. (2021). IAA3-mediated repression of PIF proteins coordinates light and auxin signaling in Arabidopsis. *PloS Genet.* 17, e1009384. doi: 10.1371/journal.pgen.1009384

Xiong, Y., and Jiao, Y. (2019). The diverse roles of auxin in regulating leaf development. *Plants (Basel)* 8, 243. doi: 10.3390/plants8070243

Xu, Z., Ji, A., Song, J., and Chen, S. (2016). Genome-wide analysis of auxin response factor gene family members in medicinal model plant Salvia miltiorrhiza. *Biol. Open* 5, 848–857. doi: 10.1242/bio.017178

Yu, H., Liu, M., Zhu, Z., Wu, A., Mounet, F., Pesquet, E., et al. (2022). Overexpression of egrIAA20 from eucalyptus grandis, a non-canonical aux/IAA gene, specifically decouples lignification of the different cell-types in arabidopsis secondary xylem. *Int. J. Mol. Sci.* 23, 5068. doi: 10.3390/ijms23095068

Yuan, Y., Xu, X., Luo, Y., Gong, Z., Hu, X., Wu, M., et al. (2021). R2R3 MYB-dependent auxin signalling regulates trichome formation, and increased trichome density confers spider mite tolerance on tomato. *Plant Biotechnol. J.* 19, 138–152. doi: 10.1111/pbi.13448

Zemlyanskaya, E. V., Wiebe, D. S., Omelyanchuk, N. A., Levitsky, V. G., and Mironova, V. V. (2016). Meta-analysis of transcriptome data identified TGTCNN motif variants associated with the response to plant hormone auxin in Arabidopsis thaliana L. *J. Bioinform. Comput. Biol.* 14, 1641009. doi: 10.1142/S0219720016410092

Zhang, H., Lu, X., Wang, Z., Yan, X., and Cui, H. (2021). Excretion from long glandular trichomes contributes to alleviation of cadmium toxicity in Nicotiana tabacum. *Environ. pollut.* 285, 117184. doi: 10.1016/j.envpol.2021.117184

Zhang, X., Garreton, V., and Chua, N.-H. (2005). The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev.* 19, 1532–1543. doi: 10.1101/gad.1318705

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat. Protoc.* 1, 641–646. doi: 10.1038/nprot.2006.97

Zhang, X., Yan, F., Tang, Y., Yuan, Y., Deng, W., and Li, Z. (2015). Auxin response gene slARF3 plays multiple roles in tomato development and is involved in the formation of epidermal cells and trichomes. *Plant Cell Physiol.* 56, 2110–2124. doi: 10.1093/pcp/pcv136

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 February 2024
DOI 10.3389/fpls.2024.1349009

OPEN ACCESS

EDITED BY
Naoki Kitaoka,
Hokkaido University, Japan

REVIEWED BY
Naoki Ube,
Toyama Prefectural University, Japan
Lemeng Dong,
University of Amsterdam, Netherlands

*CORRESPONDENCE
Sibongile Mafu
smafu@umass.edu

RECEIVED 04 December 2023
ACCEPTED 19 January 2024
PUBLISHED 15 February 2024

CITATION
Hendrickson H, Islam M, Wabo GF and Mafu S (2024) Biochemical analysis of the TPS-a subfamily in *Medicago truncatula*. *Front. Plant Sci.* 15:1349009. doi: 10.3389/fpls.2024.1349009

Biochemical analysis of the TPS-a subfamily in *Medicago truncatula*

Hannah Hendrickson[1], Monirul Islam[2], Ghislain Fotso Wabo[3] and Sibongile Mafu[1,2]*

[1]Plant Biology Graduate Program, University of Massachusetts-Amherst, Amherst, MA, United States, [2]Department of Biochemistry and Molecular Biology, University of Massachusetts-Amherst, Life Science Laboratories, Amherst, MA, United States, [3]Department of Organic Chemistry, University of Yaoundé 1, Yaounde, Cameroon

Terpenes are important mediators of plant chemical response to environmental cues. Here, we describe the genome-wide identification and biochemical characterization of TPS-a members in *Medicago truncatula*, a model legume crop. Genome mining identified thirty-nine full-length terpene synthases with a significant number predicted to produce monoterpenes and sesquiterpenes. Biochemical characterization of the TPS-a subfamily associated with sesquiterpene biosynthesis revealed such compounds, that exhibit substantial biological activity in other plants. Gene expression analysis using qPCR and the *Medicago* gene atlas illustrated distinct tissue and time-based variation in expression in leaves and roots. Together our work establishes the gene-to-metabolite relationships for sesquiterpene synthases in *M. truncatula*. Understanding the biosynthetic capacity is a foundational step to defining the ecological roles of this important family of compounds.

KEYWORDS

***Medicago truncatula*, terpene synthase, sesquiterpenes, biochemical analysis, chemical diversity**

Introduction

Plants produce an array of specialized metabolites in response to environmental stressors. Terpenoids are one such class of natural products with unprecedented chemical diversity. Biosynthesis proceeds from simple core steps which initiate from five-carbon units to make varying lengths of backbones (Figure 1). Terpene synthases (TPS) increase chemical diversity by conversion of precursors into hydrocarbons through complex lysis and allylic diphosphate ester bond cleavage (Christianson, 2017). The resulting compounds are structurally related forming variably sized and species-specific gene families (Pichersky and Raguso, 2018).

Phylogenetic analysis of TPSs results in the clustering of genes into seven subfamilies based on sequence homology, gene architecture and functional studies in various plants (Chen et al., 2011). Generally, sesquiterpene synthases (C15) fall primarily into TPS-a,

monoterpene synthases (C10) in TPS-b and TPS-g, and diterpene synthases (C20) in TPS-c and TPS-e/f whereas TPS-d and TPS-h are specific to gymnosperms and lycophytes respectively. This vast chemical diversity results in species-specific complex mixtures of terpenes with broadly varying biological activity. Terpenoids mediate interactions between plants and their environment and are critical in coordinating responses to (a)biotic stress (Pichersky and Raguso, 2018). Induced volatile terpene blends composed mainly of lower molecular weight monoterpenes and sesquiterpenes have additional functions in inter and intra-plant communication (He et al., 2022).

Legumes are important crops providing a significant portion of protein in human and animal diets and play an important role in agricultural sustainability. Early investigations in *M. truncatula*, have demonstrated that differential terpene-enriched volatile blends are induced in response to herbivores (Leitner et al., 2005; Arimura et al., 2008). Subsequent studies identified biosynthetic genes underlying these observed metabolites summarized in Figure 1 (Gomez et al., 2005; Arimura et al., 2008; Navia-Giné et al., 2009; Garms et al., 2010). *Mt*TPS1 produces a sesquiterpene, (*E*)-β caryophyllene, which in maize is instrumental in indirect defense by attracting parasitoids of herbivores and entomopathogenic nematodes (Rasmann et al., 2005; Degenhardt et al., 2009). *Mt*TPS3 forms nerolidol and geranyllinalool common to many plants with defense capabilities. In *M. truncatula*, *Mt*TPS4 which encodes for a monoterpene β-ocimene, was upregulated in response to methyl jasmonate and feeding insects (Navia-Giné et al., 2009). β-ocimene has a priming effect that activates a defense response in Chinese cabbage and cucumber (Navia-Giné et al., 2009; Kang et al., 2018; He et al., 2022). *Mt*TPS5 is a multiproduct TPS, producing up to twenty-seven different terpene products (Garms et al., 2010). More recently, Yadav et al. (2019) showed that *Mt*TPS10, which catalyzes the formation of a sesquiterpene alcohol, himachalol, was upregulated in the root upon infection by an oomycete pathogen, *Aphanomyces euteiches* (Yadav et al., 2019). The combined literature illustrates that *M. truncatula* produces a blend of terpene-based compounds upon interactions with biotic stressors as part of its chemical response strategy.

Soybean terpene synthases *Gm*TPS3, *Gm*TPS18 and *Gm*TPS21 have been characterized and shown to produce β-ocimene, geraniol and α-farnesene respectively, each of which have demonstrated protective roles in plants (Liu et al., 2014; Lin et al., 2017; Han et al., 2023). For example, *Gm*TPS21 (which encodes for α-farnesene) provided resistance against soybean cyst nematode belowground and aphids aboveground (Lin et al., 2017; Constantino et al., 2021). In cowpea, terpenes were identified as part of the volatiles produced in response to herbivory whereas biochemical characterization of the genes responsible for floral scent was carried out in sweet pea (Bao et al., 2020; Steinbrenner et al., 2022).

In grasses, such as maize, rice and wheat, terpenes play an important role in the interactions of plants with their environment (Rasmann et al., 2005; Schnee et al., 2006; Schmelz et al., 2014). Although terpenes are widespread there still exist distinct chemotypes in species with new chemistry and biological roles which are still being uncovered (Pichersky and Raguso, 2018; Ding et al., 2020; Murphy and Zerbe, 2020; Zhang et al., 2021). Comparatively less is known about the composition of the terpenome in legumes.

Here, we define the gene-to-metabolite relationships underlying terpene biosynthesis in *M. truncatula*. Through genome-wide identification of terpene synthases and biochemical characterization of the TPS-a subfamily, we have identified sesquiterpene synthases essential in generating the sesquiterpene-based chemical diversity in *M. truncatula*. Understanding the terpenoid biosynthetic capacity is a foundational step to defining the ecological roles of this important family of compounds.

Materials and methods

*Mt*TPS phylogenetic and sequence analysis

Putative terpene synthases (protein models) from the *Medicago truncatula* cv. Jemalong A17 were collated from the Mt4.0v1 genome in Phytozome (https://phytozome-next.jgi.doe.gov) through an annotation-based search for TPSs. The putative TPSs were further manually curated for the presence of catalytic domains namely the Class I site with characteristic DDxxD and secondary NSE/DTE metal binding motif involved in ionization-dependent cyclization. Class II synthases were identified by the DxDD motif involved in protonation-initiated cyclization characteristic of copalyl diphosphate synthases (CPPs) as shown in Supplementary Figure S1 (Christianson, 2017). For phylogenetic analysis, previously reported and biochemically characterized genes were collated from *S. lycopersicum* ITAG2.4 and *A. thaliana* TAIR10 genomes (Zhou and Pichersky, 2020), aligned with *M. truncatula* TPSs using maximum likelihood analysis with 1000 bootstrap repetitions using CLC software and visualized using iTOL (Letunic and Bork, 2021).

Combinatorial expression in *Escherichia coli*

Synthetic *Mt*TPSs codon-optimized for expression in *Escherichia coli* were ordered through ThermoFisher and subcloned into pET28b using primers listed in Supplementary Table S1. Functional analysis of genes was carried out using a previously described modular metabolic engineering system in *E. coli* enhanced for sesquiterpene production (Kitaoka et al., 2015). Briefly, TPS-a family genes in pET28b were co-transformed with an (*E*, *E*)-FPP synthase from *Zea mays* (*Zm*FPPs) in *E. coli* BL21 DE3-C41 cells (Lucigen). Transformed cultures were grown in 45 mL of Terrific Broth medium to an OD_{600} of ~0.6 at 37°C Cultures were cooled to 16°C before induction with 1 mM isopropyl-thiogalactopyranoside and supplemented with 1 mM $MgCl_2$ and 25 mM sodium pyruvate incubation for 72 hours. Dodecane (20 mL) was added to ensure the capture of volatile compounds. In a second flask, enzyme products were extracted with 50 mL of 100% hexane and concentrated under an N_2 stream. Samples were resuspended in 1 mL of *n*-hexane for analysis by gas chromatography–mass spectrometry (GC-MS).

GC-MS analysis

Samples were analyzed on an Agilent 8890B gas chromatograph (GC) coupled to a mass spectrometer (MS) 5977B extractor electron ionization detector at 70 eV. Samples (1 µL) were injected in pulsed spitless mode with the inlet temperature set to 250°C. Separation was achieved on an HP5MS column (30 m, 0.25 mm i.d., 0.25 µm film) using helium as the carrier gas at a flow rate of 1.2 mL/min. The initial oven temperature of 70°C was increased after 1 min to 260°C at a rate of 10°C/min and held for 2 min at 260°C. MS data was collected from 60 to 500 mass-to-charge ratio(m/z). Compounds were identified by comparison of mass spectra and retention times with those of the authentic standards, when available, or with the Wiley, National Institute of Standards and Technology (NIST) mass spectral library.

Product purification for NMR analysis

To obtain sufficient compound for structural elucidation by NMR, cultures were scaled up to 500 mL batches to a total volume of 5 Liters and expressed using the combinatorial expression system described above (Kitaoka et al., 2015). The resultant products were extracted with an equal volume of hexanes. The organic extract was recovered using a separatory funnel and then dried by rotary evaporation. The resulting residue was passed through a silica column and eluted using a hexane: ethyl acetate gradient. Fractions of interest were further purified by HPLC using an Agilent 1260 series instrument equipped with an autosampler, fraction collector and diode array UV detector over a ZORBAX Eclipse XDB C18 (4.6 x 150 mm, 5 mm) at a 0.5 mL/min flow rate using a water and acetonitrile gradient as the mobile phase. Purified compounds were resuspended in $CDCl_3$. Structural analysis was performed using 1D (^{1}H, ^{13}C) and 2D (TOCSY, ROESYPHPR, HMBCGP, DQFCOSY, and 13CHSQC) NMR experiments. Spectra were acquired on a Bruker Avance 500-MHz TopSpin NMR spectrometer.

Plant material, growth conditions and elicitors treatment method

Germination of *M. truncatula* cv. Jemalong A17 seeds were carried out by scarification using concentrated anhydrous sulfuric acid for 5 min followed by intensive washing steps with Milli-Q water. Seeds were then placed in water in a 50 ml falcon tube at 4°C for 3 days to break dormancy. Seeds were transferred to moist blotting paper and placed on a petri dish incubated at 26°C. Single seedlings were transferred to pots (9 cm diameter) filled with a perlite and sand mixture (5:1), then grown in a growth chamber with 16 h light/8 h dark for 6 weeks. After 6 weeks of growth, plants were treated by foliar application with solutions containing 1 mM methyl jasmonate (MeJA) in 0.01% Triton-X-100 (Sigma-Aldrich, St Louis, MO, USA) or 1.5 mM salicylic acid solution dissolved in 1% ethanol (SA) (Sigma-Aldrich, St Louis, MO, USA). Plants were treated with 0.01% Triton-X-100 for MeJA control and 1% ethanol for SA. For each plant, 20 ml solution was applied (either MeJA or SA or controls) and at least three biological replicates were conducted for each treatment. The leaf and root tissues were separately collected in time courses from 0h, 2 h, 6 h, 12 h, 24 h, 48 h, 72 h and 96 h and immediately frozen in liquid nitrogen and stored at -80°C for further RNA extraction.

RNA extraction and cDNA synthesis from root and leaf tissues

Sampling was performed on three biological replicates for both elicitors and controls. The root and leaf tissues were separately ground in liquid nitrogen and 100 mg powder was used for total RNA isolation with an RNeasy Plant Mini Kit (Qiagen), and RNase-Free DNase to remove genomic DNA following the manufacturer's protocol. RNA purity and yield were checked using a NanoDrop 8000 Spectrophotometer (ThermoFisher Scientific), while RNA integrity and quantification were evaluated with an Agilent 2100 bioanalyzer (Agilent Technologies). First-strand cDNA was synthesized using a Superscript IV Reverse transcriptase kit (ThermoFisher Scientific, USA).

Quantitative real time-PCR

Primers were designed using coding sequences from the genome of *M. truncatula* (Mt4.0v1) in Phytozome13 (https://phytozome-next.jgi.doe.gov/) (Supplementary Table S2). A quantitative real-time PCR system was carried out using SYBR green with ROX (pre-mixed) as an internal loading standard performed on an Eppendorf Mastercycler. The reaction mixture was 10 µl and comprised of 5 µl of 2× SYBR™ Green qPCR Master Mix (Thermofisher Scientific, USA), 100 nM primers (Thermo Fisher Scientific) and 0.5 µl of (1:10 dilution) cDNA. PCR protocol was followed to initiate polymerase activation: 10 min at 95°C; 40 cycles of 30 s at 95°C, 60 s at 55°C and 30 s at 72°C. Each run of qPCR was followed by a melting curve analysis from 55 to 95 °C. PCR conditions were determined by comparing the threshold values of the RT product (cycle threshold (Ct) value). Relative RNA levels were calibrated and normalized with the level of actin and histone-3 reference genes based on the $2^{-\Delta\Delta Ct}$ method (Rey et al., 2016; Zhang et al., 2020). For each gene, three technical replicates were run for three biological samples. qPCR data were expressed as a fold change with respect to the equivalent time-point in the control groups of MeJA (in 0.01% Triton-X-100 and SA (in 1% ethanol).

Results

Genome-wide analysis of *Mt*TPSs

We assessed for the genetic potential of *M. truncatula* cv. Jemalong A17 to produce terpenes using the Mt4.0v1 genome in Phytozome. We identified 54 putative *Mt*TPS gene models using BLAST and subsequent analysis of characteristic motifs

(Supplementary Table S3 and Supplementary Figure S1). Further curation through sequence alignments confirmed thirty- nine full-length genes. Twenty-two of the reported fifty-four potential TPSs in *M. truncatula* were previously assigned numbers by Parker et al., 2014 (*Mt*TPS16 is no longer classified as a putative terpene synthase) (Parker et al., 2014). Supplementary Table S3 is a summary of the previous and newly assigned TPS numbers. A majority of the genes had the Class I DDxxD and secondary NSE/DTE metal binding motif involved in ionization-dependent cyclization (Christianson, 2017). Three Class II synthases were identified by DxDD motifs involved in protonation-initiated cyclization characteristic of copalyl diphosphate synthases (CPPs) involved in diterpene scaffold formation. Phylogenetic analysis of the terpene synthases from *M. truncatula* and the comprehensively biochemically characterized TPS collections of *A. thaliana* and *S. lycopersicum* shows plant species clustering more than enzymes with similar functions (Figure 2) (Chen et al., 2011; Falara et al., 2011; Parker et al., 2014; Zhou and Pichersky, 2020). The genes are distributed into the different clades with fourteen members in TPS-a; five members in TPS-b; three members in TPS-c and four members in TPS e/f. Interestingly there is an expanded TPS-g (thirteen members) subfamily as compared to *A. thaliana* and *S. lycopersicum*.

The biochemical analysis described here focuses on the TPS-a subfamily only, guided by previous studies that showed sesquiterpene compounds are a predominant component of volatile collections of *M. truncatula* in response to herbivory (Leitner et al., 2005; Arimura et al., 2008). In addition, subsequent biochemical characterization assigned metabolites for three of fourteen sesquiterpene synthases - namely *Mt*TPS1, 5, and 10 - which are in the TPS-a subfamily (Figure 1) (Gomez et al., 2005; Arimura et al., 2008; Garms et al., 2010). MtTPS2 showed increased gene expression upon elicitation and was putatively described as germacrene-D synthase based on phylogenetic analysis but there was no corresponding biochemical confirmation (Gomez et al., 2005). Figure 2 shows little overlap of the members of TPS-a clade amongst plants as they tend to group by species. In *S. lycopersicum* the fifteen TPSs encode for sesquiterpenes whereas *Arabidopsis* TPS-a members are not limited to sesquiterpene biosynthetic genes (Zhou and Pichersky, 2020). Combined, these observations guided our investigations to determine the functional landscape of TPS-a subfamily members in *M. truncatula* through heterologous expression.

Biochemical analysis of TPS-a subfamily

To correlate metabolites produced to the identified biosynthetic genes, we assessed for TPS activity and resultant metabolite

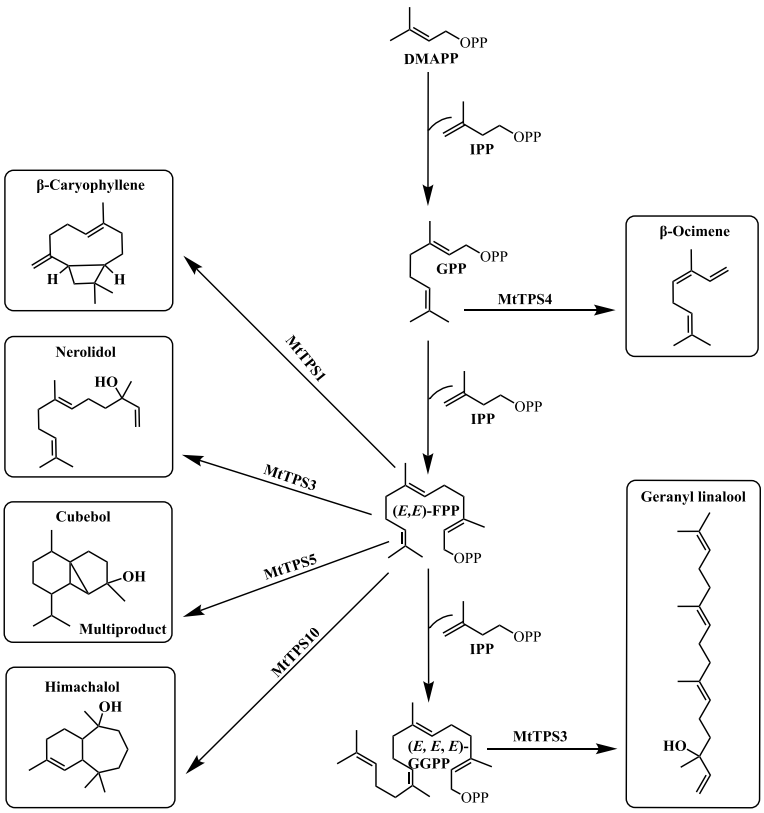


FIGURE 1
Schematic of terpene biosynthesis. Three major substrates are used in the production of most terpene products by TPSs: GPP (C10), (*E, E*)-FPP (C15), and (*E, E, E*)-GGPP (C20). Resultant terpene products of previously characterized *Mt*TPSs: *Mt*TPS1 produces β-caryophyllene, *Mt*TPS3 produces both the sesquiterpene product nerolidol and the diterpene product geranyl linalool, *Mt*TPS4 produces β-ocimene, *Mt*TPS5 produces cubebol as a major product but is a multiproduct enzyme and *Mt*TPS10 produces mainly himachalol.

formation using a modular metabolic engineering system in *E. coli* (Kitaoka et al., 2015). Briefly, heterologous expression was performed through co-expression of an (*E*, *E*)-FPP synthase with uncharacterized full-length TPS-a subfamily genes (*Mt*TPS2, 7, 8, 9, 11, 12, 13, 24, 25, 26 and 27). Metabolite production was analyzed using GC-MS (Figure 3). Compounds were compared to authentic standards when available, the NIST library database, or verified through *de novo* structural elucidation by NMR.

Analysis of resultant metabolites produced by TPS-a subfamily members reveals the production of an assortment of sesquiterpenes. Figure 3 highlights the most abundant products formed by TPSs and products in smaller quantities are described in the supplementary information (Suppplementary Figures S3-S8). *Mt*TPS7 is a multiproduct enzyme that produces α-copaene (2), cyclosativene (3) and a hydroxylated compound which matches a compound produced by *Mt*TPS26 (Figure 3B; Supplementary Figure S3). *Mt*TPS8 is a multiproduct synthase that produces farnesene (4) as the major compound and exhibits relatively significant turnover to amorphadiene, bergamotene and nerolidol, and production of up to fourteen products in smaller amounts (Figure 3C; Supplementary Figure S4), *Mt*TPS11, produces longicyclene (5) as the main product and a smaller amount of longifolene and longipinene (Figure 3D; Supplementary Figure S5).

Although terpene synthases more commonly form hydrocarbons, the addition of water before terminating deprotonation generates hydroxylated products (Christianson, 2017). *M. truncatula* produces a set of TPSs capable of producing such hydroxylated compounds as supported by a molecular weight of 222. *Mt*TPS13 produces germacrene-D-4-ol (6) as compared to NIST (Figure 3E; Supplementary Figure S6). The genes, *Mt*TPS24 and *Mt*TPS26 are present in the *Mt*4.0v1 genome and not identified in earlier genomes (Parker et al., 2014). *Mt*TPS24 produces a compound with a distinct base peak at 151 (7) (Figure 3F; Supplementary Figure S7). However, we were unable to collect NMR data as the compound aggregates in tested deuterated solvents (chloroform, methanol and DMSO) and no further analysis could be performed. Upon expression, *Mt*TPS26 produces two predominant compounds that required structural verification by NMR due to the unavailability of authentic standards. Structural elucidation by NMR and comparison to literature values showed that compound 8 was eudesm4(14)-en-6β-ol (Figure 3G) with chemical shift data closely matching previously reported data for 10-*epi*-junenol except for the hydroxyl in a 6β-position in 8 (Supplementary Table S5) (F. et al., 1976). Compound 9 was confirmed to be guai-6-en-10β-ol, matching previously reported chemical shifts as shown in Supplementary Table S6 (Lago et al., 2000). No activity was detected with *Mt*TPS2, 9, 12, 25 and 27 upon co-expression with (*E*, *E*)-FPP synthase using the combinatorial expression system.

The ability of TPSs to make several products is not uncommon, for example, *Mt*TPS5 was demonstrated to produce up to twenty-seven products (Garms et al., 2010; Vattekkatte et al., 2017). Here, *Mt*TPS8 is the most promiscuous forming up to twenty compounds (Supplementary Figure S4). Multiple products produced in smaller

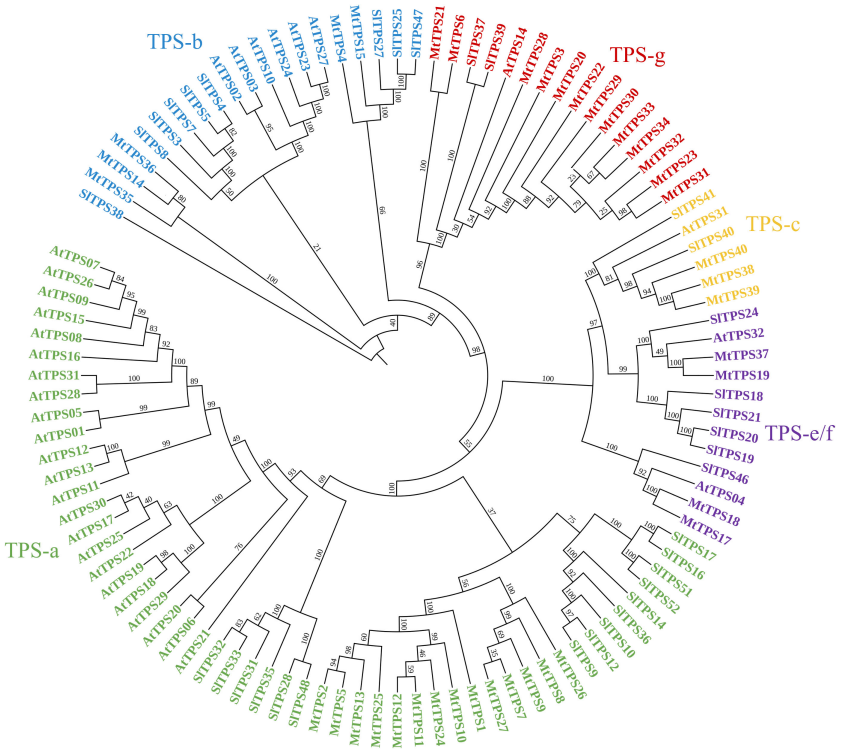


FIGURE 2
Phylogeny-based analysis of terpene synthases in *Medicago truncatula*, *Arabidopsis thaliana*, and *Solanum lycopersicum*. The sequences were aligned using maximum likelihood analysis with 1000 bootstrap repetitions using CLC software and visualized using iTOLTPS subfamilies are grouped as follows: TPS- a (green), monoterpenes in TPS-b (blue) and TPS-g (red), and diterpenes in TPS-c (orange) and TPS-e/f (purple).

10.3389/fpls.2024.1349009

quantities are highlighted for all the genes (Supplementary Figures S3-S8). Similarly, because of the related mechanism of terpene synthases, it is common to observe that some genes will also produce the same compounds in varying amounts *in vitro*. This is observed for *Mt*TPS7 and *Mt*TPS26 which both produce 9; a similar case is noted for *Mt*TPS11 which produces longicyclene (5) as the main product, but profiles of *Mt*TPS5 & 10 also produce longicyclene in small amounts (Garms et al., 2010; Yadav et al., 2019).

In summary, this biochemical analysis defines gene-to-metabolite relationships in the TPS-a subfamily of *M. truncatula*. We defined the genetic basis for the biosynthesis of farnesene,

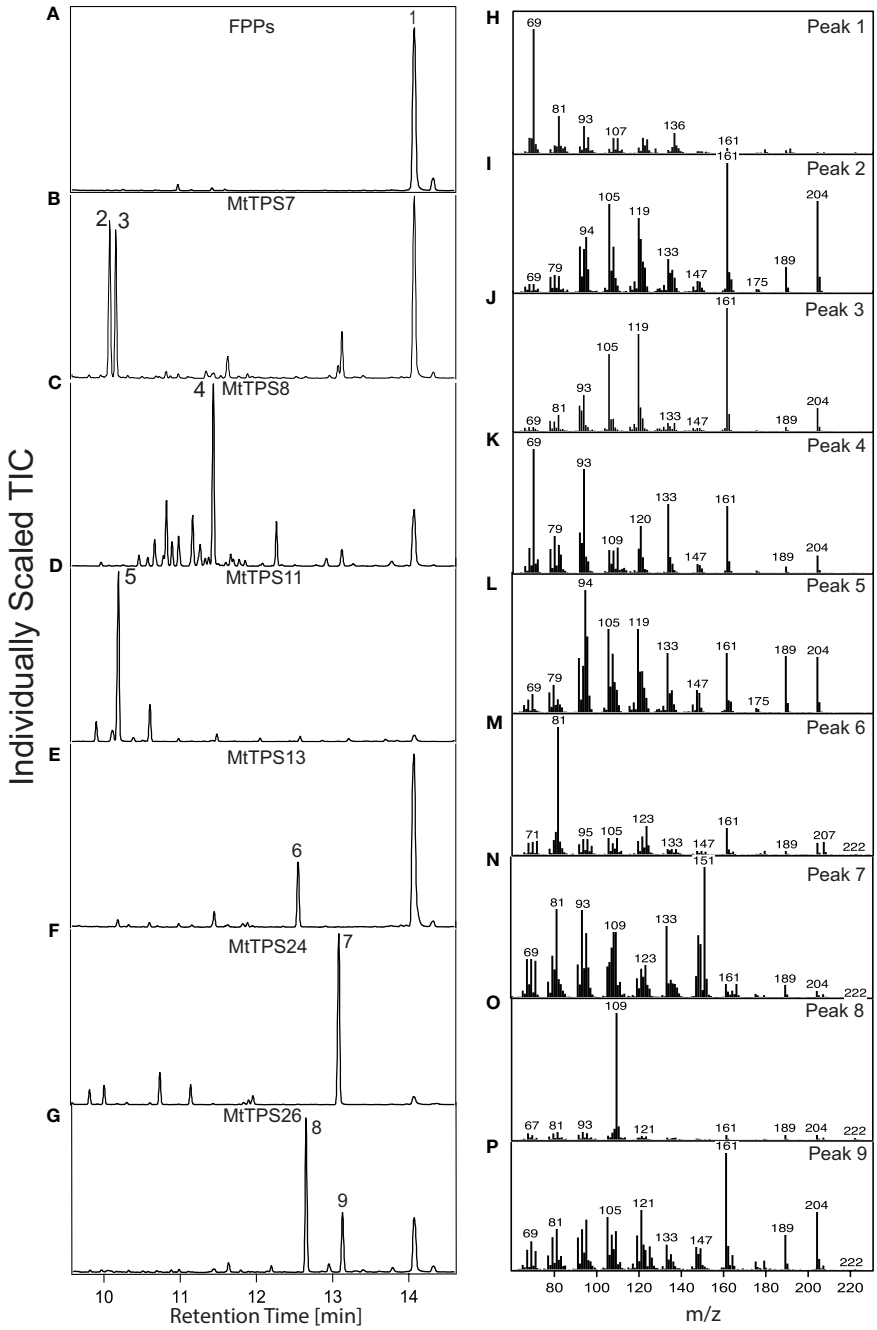


FIGURE 3
Functional of the TPS-a subfamily in *Medicago truncatula*. Total ion chromatograms of products resulting from combinatorial expression in *E. coli* of (*E*, *E*)-FPP synthase and corresponding genes from the TPS-a subfamily. The chromatograms focus on the main products. Detailed analysis of the smaller peaks is depicted in Supplementary Figures S3-8. Compound identification was through similarity to library matches (NIST, Wiley) for B-E and NMR verification for G. **(A)** pET28b containing FPP synthase produces **(H)** dephosphorylated FPP (farnesol) (1) **(B)** *Mt*TPS7 produces **(I, J)** cyclosativene (2), α-copaene (3). **(C)** *Mt*TPS8 produces **(K)** farnesene-like product (4). **(D)** *Mt*TPS11 produces **(L)** longicyclene (5). **(E)** *Mt*TPS13 makes a **(M)** germacrene-D-4-ol product (6), **(F)** *Mt*TPS24 makes a **(N)** unknown product (7) and **(G)** *Mt*TPS26 makes a **(O, P)** eudesm4(14)-en-6β-ol (8) and a guai-6-en-10β-ol (9).

nerolidol, cyclosativene, copaene and longicyclene, enabling the association of genes responsible for the biosynthesis of several compounds observed in terpene metabolite response to herbivore feeding (Leitner et al., 2005) and identified a set of sesquiterpene alcohols (Figure 4).

Gene expression analysis of TPS-a members in response to treatment with methyl jasmonate and salicylic acid

Gene expression of the TPS-a subfamily members in response to elicitors was carried out through exogenous application of phytohormones MeJA and SA and analysis by qPCR (Figure 5). Overall, TPS-a genes showed tissue-specific and temporal differences in expression in response to elicitation with MeJA and SA. There was higher expression of TPS-a genes in leaf tissue compared to the root upon elicitation with MeJA. Specifically, our time course evaluations revealed that MeJA application resulted in increased expression (16-20 fold) of *Mt*TPS1 after hours of exposure in leaves whereas expression of *Mt*TPS2 and *Mt*TPS5 was highest after 12 hours. A 4-8 fold increase in expression was observed amongst the *Mt*TPS8, *Mt*TPS10, *Mt*TPS13 and *Mt*TPS26 genes in leaf tissues after 6 and 12 hours of post-treatment of MeJA. *Mt*TPS11, 12, 24, 25 & 27 have low expression in both the leaves and roots. There was a consistently low expression for all the TPS-a members in roots under MeJA elicitation, with MtTPS25 & 27 showing down-regulation.

In contrast, SA induced gene expression occurred mainly in roots. The genes *Mt*TPS10 and *Mt*TPS11 showed the highest expression at 24 hours compared to control with a 20-fold increase whereas a 16-fold increase in expression was observed for genes *Mt*TPS13, 24 & 26 (Figure 5). However, comparatively lower expression was observed in both leaf and root tissue for the remaining eight genes (*Mt*TPS1, 2, 5, 7, 8, 12, 25 & 27). Overall, the TPS-a genes exhibited differential expression to the phytohormones where MeJA largely affected the expression of TPS genes in leaf tissue and SA in root tissue upon foliar treatment.

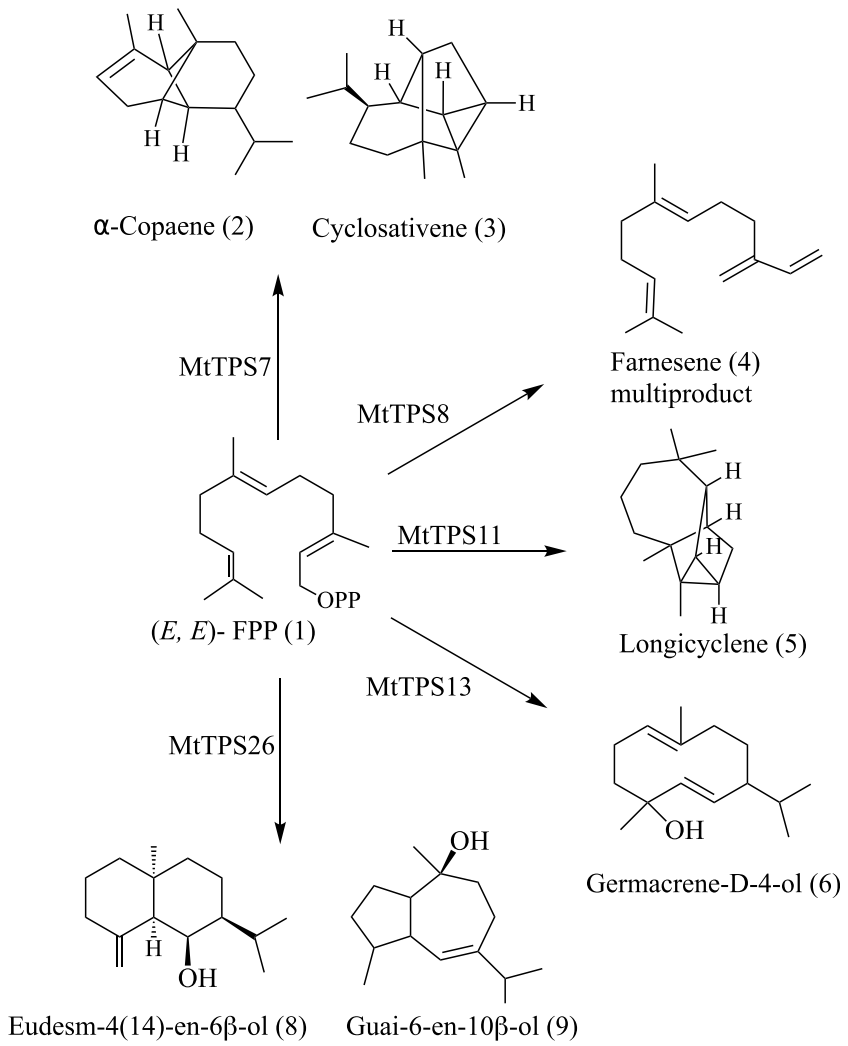


FIGURE 4
Summary of the main products from biochemical characterization of TPS-a subfamily members in *M. truncatula*.

Expression patterns of TPS in different tissues and under (a)biotic stress

We used a publicly available expression data set of *M. truncatula* from the Noble Research Institute legacy Gene Atlas V3 to survey the effect of (a)biotic stress on the TPS-a subfamily (Benedito et al., 2008; Carrere et al., 2021). Mining of the reference data set, which is a collation of gene expression in different tissues, shows that terpene synthases *Mt*TPS1, 2 & 5 which produce volatile compounds exhibited increased expression in the shoots and leaves. Transcript levels of the *Mt*TPS13 were increased during root and nodule development, AM symbiosis, and fungal colonization (Supplementary Figure S23A).

Evaluation of the *TPS*-a subfamily genes response to abiotic stress highlights increased transcript abundance of *Mt*TPS1, 2, 5 and *Mt*TPS13 under limited nitrogen (N_2), ammonium (NH_4), nitrate (NO_3), drought, and salt stress in shoot and root tissues (Supplementary Figure S23B). An assessment of varied biotic stressors showed that *Mt*TPS1 and 2 had increased expression in leaf tissue after aphid infestation whereas *Mt*TPS13, *Mt*TPS24 and *Mt*TPS25 displayed increased transcript abundance in root tissues under pathogenic fungi, *Rhizoctonia solani* and *Fusarium oxysporum* (Supplementary Figure S23C). Notably, datasets obtained from Gene Atlas do not capture the expression of six terpene synthases (*Mt*TPS7,8, 9, 25, 26 & 27). Combined this data supports the variable expression of select terpene synthases dependent on tissue, time and stressor applied.

Discussion

Biochemical analysis reveals functional divergence of TPS-a subfamily

The catalytic mechanism of sesquiterpenes proceeds via a succession of carbocationic intermediates initiated by dephosphorylation and subsequent deprotonation or water capture (Christianson, 2017). The metabolites of the encoded enzyme of *Mt*TPS1, *Mt*TPS7 and *Mt*TPS8 have characteristic profiles to previously described enzymes, for example, *Mt*TPS8 produces (*E*)-β-farnesene with bergamotene similar to what is observed for *Zm*TPS10 (Schnee et al., 2002; Rasmann et al., 2005; Schnee et al., 2006). Notably, *Mt*TPS11 produces longicyclene, a compound commonly found in pine trees (Martin et al., 2004). Previous studies have identified a longifolene synthase with smaller amounts of longicyclene and longipinene - here the enzyme *Mt*TPS11 forms longicyclene as the major product. Olefins (compounds 2-5) produced by *M. truncatula* are characteristic in volatile blends of several plants upon herbivore feeding and associated interactions (Schnee et al., 2002; Schnee et al., 2006). Here, some of the TPS-a genes quenched the reaction through water capture resulting in sesquiterpene alcohols (compounds 6-9) that have been identified in other plants, but little is known of their biological activity. Our biochemical analysis combined with the previously characterized genes establishes the use of terpenes common to plants generating a species-specific sesquiterpene blend.

A summary of the known gene-to-metabolite relationships in the TPS-a subfamily of *M. truncatula* shows functional divergence from the

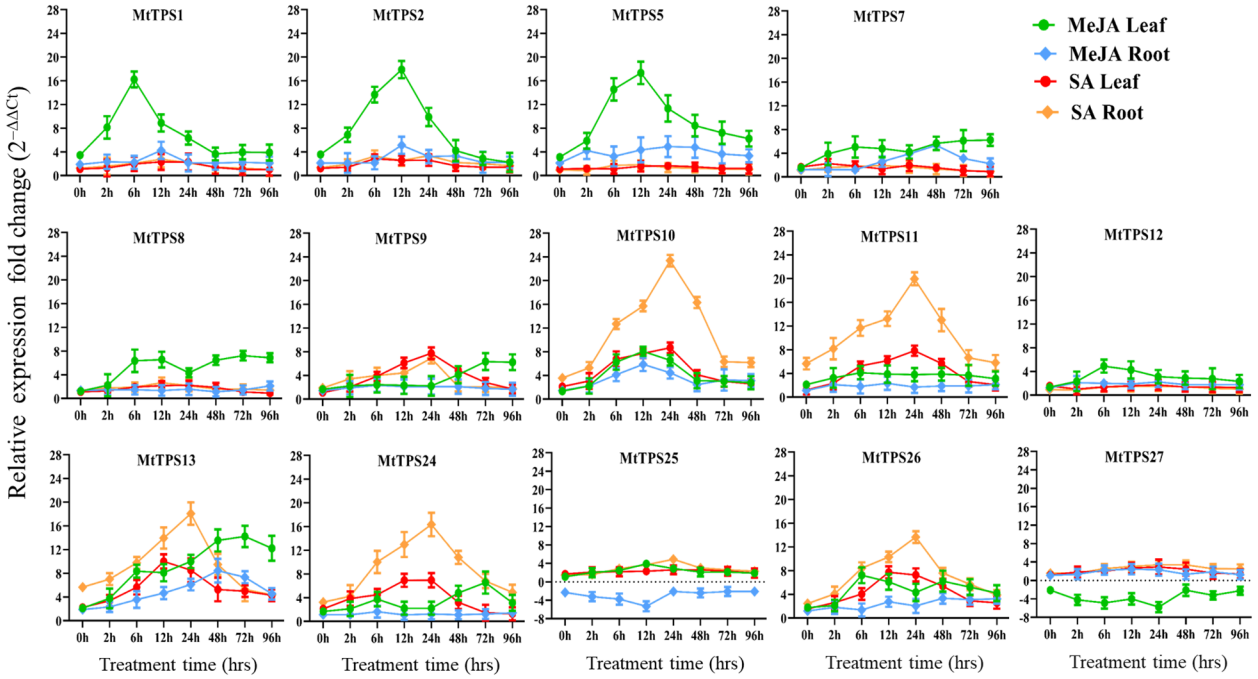


FIGURE 5
Expression profile of TPS-a terpene synthase genes in leaf and root tissues in response to exogenous application of MeJA and SA in a time course exposure by qRT-PCR. Plants were exposed for 0h, 2 h, 6 h, 12 h, 24 h, 48 h, 72 h and 96 h after foliar application of MeJA (1 mM) or SA (1.5 mM). The data are expressed as the relative level of expression of fold change relative to controls (0.01% Triton-100 for MeJA, or 1% ethanol for SA control) plants, based on the $2^{-\Delta\Delta ct}$ method. Error bars represent the mean ± standard deviation (SD) levels of the relative abundance of three biological replicates.

TPS profiles of the two comprehensively biochemically characterized eudicots (Figure 4). Previously characterized genes from *M. truncatula* mainly fall in the TPS-a clade (Figures 1 & 2). Comparison of the currently defined *M. truncatula* TPS-a subfamily members compared to those of *Arabidopsis* and tomato show a difference in metabolite profiles. In tomato, the fifteen TPSs encode for sesquiterpenes whereas the product outcome of the TPS-a subfamily in *A. thaliana* is not limited to sesquiterpenes (Supplementary Table S4) (Zhou and Pichersky, 2020). Biochemical analysis here revealed that a subset of the metabolites synthesized by *M. truncatula* sesquiterpene synthases was specific to the species, revealing functional divergence.

Differential expression of TPS-a subfamily in response to phytohormone elicitation

In this study, we elicited TPS expression by exogenous application of phytohormones. The genes that produce volatile compounds had increased expression on the early onset of foliar application of MeJA (Figures 1, 3). Genes corresponding to the production of volatile compounds β-caryophyllene (MtTPS1), putative germacrene-D (*Mt*TPS2) and *Mt*TPS5 a multiproduct synthase with the demonstrated antiherbivore response were upregulated by 16 to 20-fold (Figure 5). These observations correlate in part to a study by Leitner et al. (2005), which showed that herbivore interactions and artificial wounding induced the emission of up to 23 compounds (Leitner et al., 2005; Arimura et al., 2008). This observation is supported by data from gene atlas that also showed increased expression of the volatile terpenes predominantly in the leaves under various stresses (Supplementary Figure S23). However, ten out of fourteen TPS-a subfamily members exhibited moderate upregulation by 4-6-fold using MeJA.

In contrast, elicitation using SA increased the transcript abundance in root tissue of select genes; *Mt*TPS10, *Mt*TPS11, *Mt*TPS13, *Mt*TPS24 and *Mt*TPS26 (Figure 5). In general, SA mediates response to biotrophic and hemibiotrophic pathogens and triggers systemic acquired resistance (Anand et al., 2008). Yadav et al. (2019) demonstrated that MtTPS10, which produces himalachol, reduced susceptibility to the oomycete pathogen *Aphanomyces euteiches* (Yadav et al., 2019). The terpene scaffolds may be elaborated by downstream enzymes resulting in further oxygenated non-volatile sesquiterpenoids. The resultant compounds may have various bioactivity similar to previously identified antifungals and insect deterrents such as germacrene-A based costunolides in lettuce; costic acids derived from selinene and β-macrocarpene based zealexins produced in maize under pathogen attack (Nguyen et al., 2010; Ding et al., 2017; Ding et al., 2019).

In *conclusion*, the results reported here establish gene-metabolite relationships of the TPS-a subfamily that encodes for sesquiterpene synthases in *M. truncatula*. Biochemical analysis reveals a blend of sesquiterpenes that highlight the evolution of species-specific chemotypes. Gene expression analysis shows a temporal tissue-specific and stress induced variation. While determining the function of these compounds is beyond the scope of this work, our study further defines the sesquiterpene biosynthetic capacity and provides a foundation to understand the biological roles of the compounds.

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

HH: Investigation, Methodology, Writing – original draft, Writing – review & editing. MI: Investigation, Writing – original draft. GW: Formal analysis, Writing – review & editing. SM: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Financial support was provided by the Department of Biochemistry and Molecular Biology at the University of Massachusetts – Amherst; Richard and Susan Smith Family Foundation and National Science Award 2300060 to SM. HH was supported by the National Research Service Award T32 GM135096 from the National Institutes of Health.

Acknowledgments

Mass spectral data were obtained at the University of Massachusetts Mass Spectrometry Core Facility, RRID: SCR_019063. NMR data were obtained at Rensselaer Polytechnic Institute (RPI) NMR Core Faculties: NIST Award # 60NANB22D167.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1349009/full#supplementary-material

References

Anand, A., Uppalapati, S. R., Ryu, C. M., Allen, S. N., Kang, L., Tang, Y., et al. (2008). Salicylic acid and systemic acquired resistance play a role in attenuating crown gall disease caused by Agrobacterium tumefaciens. *Plant Physiol.* 146 (2), 703–715. doi: 10.1104/pp.107.111302

Arimura, G., Garms, S., Maffei, M., Bossi, S., Schulze, B., Leitner, M., et al. (2008). Herbivore-induced terpenoid emission in Medicago truncatula: concerted action of jasmonate, ethylene and calcium signaling. *Planta* 227 (2), 453–464. doi: 10.1007/s00425-007-0631-y

Bao, T., Shadrack, K., Yang, S., Xue, X., Li, S., Wang, N., et al. (2020). Functional characterization of terpene synthases accounting for the volatilized-terpene heterogeneity in Lathyrus odoratus cultivar flowers. *Plant Cell Physiol.* 61 (10), 1733–1749. doi: 10.1093/pcp/pcaa100

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume Medicago truncatula. *Plant J.* 55 (3), 504–513. doi: 10.1111/j.1365-313X.2008.03519.x

Carrere, S., Verdier, J., and Gamas, P. (2021). MtExpress, a comprehensive and curated RNAseq-based gene expression atlas for the model legume Medicago truncatula. *Plant Cell Physiol.* 62 (9), 1494–1500. doi: 10.1093/pcp/pcab110

Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E.. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66 (1), 212–229. doi: 10.1111/j.1365-313X.2011.04520.x

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117 (17), 11570–11648. doi: 10.1021/acs.chemrev.7b00287

Constantino, N., Oh, Y., Şennik, E., Andersen, B., Warden, M., Oralkan, Ö, et al. (2021). Soybean cyst nematodes influence aboveground plant volatile signals prior to symptom development. *Front. Plant Sci.* 12, 749014. doi: 10.3389/fpls.2021.749014

Degenhardt, J., Hiltpold, I., Köllner, T. G., Frey, M., Gierl, A., Gershenzon, J., et al. (2009). Restoring a maize root signal that attracts insect-killing nematodes to control a major pest. *Proc. Natl. Acad. Sci. U.S.A.* 106 (32), 13213–13218. doi: 10.1073/pnas.0906365106

Ding, Y., Huffaker, A., Köllner, T. G., Weckwerth, P., Robert, C.A.M., Spencer, J.L., et al. (2017). Selinene volatiles are essential precursors for maize defense promoting fungal pathogen resistance. *Plant Physiol.* 175 (3), 1455–1468. doi: 10.1104/pp.17.00879

Ding, Y., Murphy, K. M., Poretsky, E., Mafu, S., Yang, B., Char, S. N., et al. (2019). Multiple genes recruited from hormone pathways partition maize diterpenoid defences. *Nat. Plants* 5 (10), 1043–1056. doi: 10.1038/s41477-019-0509-6

Ding, Y., Weckwerth, P. R., Poretsky, E., Murphy, K. M., Sims, J., Saldivar, E., et al. (2020). Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nat. Plants* 6 (11), 1375–1388. doi: 10.1038/s41477-020-00787-9

F., Thomas, Ozainne, M., Decorzant, R., and Näf, F. (1976). 10-epijunenol, a new cis-eudesmane sesquiterpenoid. *Tetrahedron* 32, 2261–2264. doi: 10.1016/0040-4020(76)85143-5

Falara, V., Akhtar, T. A., Nguyen, T. T., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., et al. (2011). The tomato terpene synthase gene family. *Plant Physiol.* 157 (2), 770–789. doi: 10.1104/pp.111.179648

Garms, S., Köllner, T. G., and Boland, W. (2010). A multiproduct terpene synthase from Medicago truncatula generates cadalane sesquiterpenes via two different mechanisms. *J. Org. Chem.* 75 (16), 5590–5600. doi: 10.1021/jo100917c

Gomez, S. K., Cox, M. M., Bede, J. C., Inoue, K., Alborn, H. T., Tumlinson, J. H., et al. (2005). Lepidopteran herbivory and oral factors induce transcripts encoding novel terpene synthases in Medicago truncatula. *Arch. Insect Biochem. Physiol.* 58 (2), 114–127. doi: 10.1002/arch.20037

Han, T., Y., Shao, R., Gao, J., Gao, Y., Jiang, Y., Yang, Y., et al. (2023). Functional Characterization of a (E)-B ocimene synthase gene contributing to the defense against Spordoptera litura. *Int. J. Mol. Sci.* 24 (8), 7182. doi: 10.3390/ijms24087182.

He, J., Verstappen, F., Jiao, A., Dicke, M., Bouwmeester, H. J., and Kappers, I. F. (2022). Terpene synthases in cucumber (Cucumis sativus) and their contribution to herbivore-induced volatile terpenoid emission. *New Phytol.* 233 (2), 862–877. doi: 10.1111/nph.17814

Kang, Z. W., Liu, F.H., Zhang, Z. F., Tian, H.G., and Liu, T.X. (2018). Volatile β-ocimene can regulate developmental performance of peach aphid. *Front. Plant Sci.* 9, 708. doi: 10.3389/fpls.2018.00708

Kitaoka, N., Wu, Y., Xu, M., and Peters, R. J. (2015). Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. *Appl. Microbiol. Biotechnol.* 99 (18), 7549–7558. doi: 10.1007/s00253-015-6496-2

Lago, J. H., Brochini, C. B., and Roque, N. F. (2000). Terpenes from leaves of Guarea macrophylla (Meliaceae). *Phytochemistry* 55 (7), 727–731. doi: 10.1016/S0031-9422(00)00302-2

Leitner, M., Boland, W., and Mithöfer, A. (2005). Direct and indirect defences induced by piercing-sucking and chewing herbivores in Medicago truncatula. *New Phytol.* 167 (2), 597–606. doi: 10.1111/j.1469-8137.2005.01426.x

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 (W1), W293–W296. doi: 10.1093/nar/gkab301

Lin, J., Wang, D., Chen, X., Köllner, T. G., Mazarei, M., Guo, H., et al. (2017). An (E, E)-α-farnesene synthase gene of soybean has a role in defence against nematodes and is involved in synthesizing insect-induced volatiles. *Plant Biotechnol. J.* 15 (4), 510–519. doi: 10.1111/pbi.12649

Liu, J., Huang, F., Wang, X., Zhang, M., Zheng, R., Wang, J., et al. (2014). Genome-wide analysis of terpene synthases in soybean: functional characterization of GmTPS3. *Gene* 544 (1), 83–92. doi: 10.1016/j.gene.2014.04.046

Martin, D. M., Fäldt, J., and Bohlmann, J. (2004). Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.* 135 (4), 1908–1927. doi: 10.1104/pp.104.042028

Murphy, K. M., and Zerbe, P. (2020). Specialized diterpenoid metabolism in monocot crops: Biosynthesis and chemical diversity. *Phytochemistry* 172, 112289. doi: 10.1016/j.phytochem.2020.112289

Navia-Giné, W. G., Yuan, J. S., Mauromoustakos, A., Murphy, J. B., Chen, F., and Korth, K.L. (2009). Medicago truncatula (E)-beta-ocimene synthase is induced by insect herbivory with corresponding increases in emission of volatile ocimene. *Plant Physiol. Biochem.* 47 (5), 416–425. doi: 10.1016/j.plaphy.2009.01.008

Nguyen, D. T., Göpfert, J.C., Ikezawa, N., Macnevin, G., Kathiresan, M., Conrad, J., et al. (2010). Biochemical conservation and evolution of germacrene A oxidase in asteraceae. *J. Biol. Chem.* 285 (22), 16588–16598. doi: 10.1074/jbc.M110.111757

Parker, M. T., Zhong, Y., Dai, X., Wang, S., and Zhao, P.. (2014). Comparative genomic and transcriptomic analysis of terpene synthases in Arabidopsis and Medicago. *IET Syst. Biol.* 8 (4), 146–153. doi: 10.1049/iet-syb.2013.0032

Pichersky, E., and Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? *New Phytol.* 220 (3), 692–702. doi: 10.1111/nph.14178

Rasmann, S., Köllner, T. G., Degenhardt, J., Hiltpold, I., Toepfer, S., Kuhlmann, U., et al. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature* 434 (7034), 732–737. doi: 10.1038/nature03451

Rey, T., Laporte, P., Bonhomme, M., Jardinaud, M. F., Huguet, S., Balzergue, S., et al. (2016). A central transcriptional regulator of symbiotic nodule development, is also a determinant of. *Front. Plant Sci.* 7, 1837. doi: 10.3389/fpls.2016.01837

Schmelz, E. A., Huffaker, A., Sims, J.W., Christensen, S.A., Lu, X., Okada, K., et al. (2014). Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* 79 (4), 659–678. doi: 10.1111/tpj.12436

Schnee, C., Köllner, T. G., Gershenzon, J., and Degenhardt, J. (2002). The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (E)-beta-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. *Plant Physiol.* 130 (4), 2049–2060. doi: 10.1104/pp.008326

Schnee, C., Köllner, T. G., Held, M., Turlings, T. C., Gershenzon, J., and Degenhardt, J. (2006). The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl. Acad. Sci. U.S.A.* 103 (4), 1129–1134. doi: 10.1073/pnas.0508027103

Steinbrenner, A. D., Saldivar, E., Hodges, N., Guayazán-Palacios, N., Chaparro, A. F., and Schmelz, E. A. (2022). Signatures of plant defense response specificity mediated by herbivore-associated molecular patterns in legumes. *Plant J.* 110 (5), 1255–1270. doi: 10.1111/tpj.15732

Vattekkatte, A., Garms, S., and Boland, W. (2017). Alternate cyclization cascade initiated by substrate isomer in multiproduct Terpene synthase from Medicago truncatula. *J. Org. Chem.* 82 (6), 2855–2861. doi: 10.1021/acs.joc.6b02696

Yadav, H., Dreher, D., Athmer, B., Porzel, A., Gavrin, A., Baldermann, S., et al. (2019). Medicago TERPENE SYNTHASE 10 is involved in defense against an oomycete root pathogen. *Plant Physiol.* 180 (3), 1598–1613. doi: 10.1104/pp.19.00278

Zhang, H., Cao, Y., Xu, Y., Zhou, C., Liu, W., Zhu, R., et al. (2020). Efficient generation of CRISPR/Cas9-mediated homozygous/Biallelic. *Front. Plant Sci.* 11, 294. doi: 10.3389/fpls.2020.00294

Zhang, J. R., Li, M., Xu, R. I., Hoffmann, Y., Zhang, B., Liu, M., et al. (2021). A (conditional) role for labdane-related diterpenoid natural products in rice stomatal closure. *New Phytol.* 230 (2), 698–709. doi: 10.1111/nph.17196

Zhou, F., and Pichersky, E. (2020). The complete functional characterisation of the terpene synthase family in tomato. *New Phytol.* 226 (5), 1341–1360. doi: 10.1111/nph.16431

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 February 2024
DOI 10.3389/fpls.2024.1344095

OPEN ACCESS

EDITED BY
Xiang Pu,
Sichuan Agricultural University, China

REVIEWED BY
Surendra Sarsaiya,
Zunyi Medical University, China
Enrico Doria,
University of Pavia, Italy
Xinbo Guo,
South China University of Technology, China

*CORRESPONDENCE
Chunyu Zhang
zhchy@mail.hzau.edu.cn
Peng Chen
chenpeng@mail.hzau.edu.cn

RECEIVED 25 November 2023
ACCEPTED 31 January 2024
PUBLISHED 26 February 2024

CITATION
Qin P, Chen P, Zhou Y, Zhang W, Zhang Y, Xu J, Gan L, Liu Y, Romer J, Dörmann P, Cahoon EB and Zhang C (2024) Vitamin E biofortification: enhancement of seed tocopherol concentrations by altered chlorophyll metabolism.
Front. Plant Sci. 15:1344095.
doi: 10.3389/fpls.2024.1344095

Vitamin E biofortification: enhancement of seed tocopherol concentrations by altered chlorophyll metabolism

Ping Qin[1], Peng Chen[1]*, Yuanwei Zhou[2], Wei Zhang[1], Yunyun Zhang[3], Jingjing Xu[1], Lu Gan[4], Yingnan Liu[5], Jill Romer[6], Peter Dörmann[6], Edgar B. Cahoon[4] and Chunyu Zhang[1]*

[1]National Key Laboratory of Crop Genetic Improvement and College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]Yichang Academy of Agricultural Science, Ministry of Agriculture and rural areas, Yichang, Hubei, China, [3]Industrial Crops Institute of Yunnan Academy of Agricultural Sciences, Ministry of Agriculture and rural areas, Kunming, China, [4]Department of Biochemistry and Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, United States, [5]Lincang Agricultural Technology Extension Center, Lincang, Yunnan, China, [6]Institute of Molecular Physiology and Biotechnology of Plants (IMBIO), University of Bonn, Bonn, Germany

Homogentisate Phytyltransferase (*HPT*) catalyzes condensation of homogentisate (HGA) and phytyl diphosphate (PDP) to produce tocopherols, but can also synthesize tocotrienols using geranylgeranyl diphosphate (GGDP) in plants engineered for deregulated HGA synthesis. In contrast to prior tocotrienol biofortification efforts, engineering enhanced tocopherol concentrations in green oilseeds has proven more challenging due to the integral role of chlorophyll metabolism in supplying the PDP substrate. This study show that RNAi suppression of *CHLSYN* coupled with *HPT* overexpression increases tocopherol concentrations by >two-fold in Arabidopsis seeds. We obtained additional increases in seed tocopherol concentrations by engineering increased HGA production via overexpression of bacterial *TyrA* that encodes chorismate mutase/prephenate dehydrogenase activities. In overexpression lines, seed tocopherol concentrations increased nearly three-fold, and resulted in modest tocotrienol accumulation. We further increased total tocochromanol concentrations by enhancing production of HGA and GGDP by overexpression of the gene for hydroxyphenylpyruvate dioxygenase (HPPD). This shifted metabolism towards increased amounts of tocotrienols relative to tocopherols, which was reflected in corresponding increases in ratios of GGDP/PDP in these seeds. Overall, our results provide a theoretical basis for genetic improvement of total tocopherol concentrations in green oilseeds (e.g., rapeseed, soybean) through strategies that include seed-suppression of *CHLSYN* coupled with increased HGA production.

KEYWORDS

chlorophyll synthase, homogentisate phytyltransferase, homogentisate, phytyl diphosphate, geranylgeranyl diphosphate, tocopherol

1 Introduction

Vitamin E tocochromanols are a class of fat-soluble antioxidants that contain a homogentisate (HGA)-derived aromatic head group linked to an isoprenoid-derived hydrocarbon tail. Vitamin E tocochromanols are comprised of tocopherols and tocotrienols that differ based on the saturation of the hydrocarbon tail: tocopherols are saturated and tocotrienols are tri-unsaturated (Hunter and Cahoon, 2007). Additionally, α, β, γ and δ forms of tocopherols and tocotrienols occur that have differing numbers or arrangements of methylation of the head group that affect their bioavailability (Netscher, 2007; Yang et al., 2020). Vitamin E biosynthesis occurs in photosynthetic organisms such as higher plants, algae, and cyanobacteria (Lichtenthaler, 1968; Sattler et al., 2004) and is present in a number of plant organs (Hunter and Cahoon, 2007).

Vitamin E is a required nutrient in human and animal diets and functions as an antioxidant that quenches free radicals derived from processes such as unsaturated fatty acid peroxidation (Warner et al., 2003). Vitamin E tocochromanols typically accumulate in seeds are components of seed oils that contribute to their oxidative stability by quenching free radicals arising from unsaturated fatty acid peroxidation (Grusak and DellaPenna, 1999; Kanwischer et al., 2005; Muñoz and Munné-Bosch, 2019). Vitamin E has also been widely incorporated in the food and cosmetic industries as a supplement for prolonging food stability and preventing UV and ozone skin damage (Thiele and Ekanayake-Mudiyanselage, 2007; Kmiecik et al., 2019).

Given the nutritional and economic importance of vitamin E tocochromanols, considerable efforts have been directed toward their biofortification in oilseeds (Shintani and DellaPenna, 1998). Homogentisate geranylgeranyl transferase (HGGT) and homogentisate phytyltransferase (HPT or VTE2) are rate-limiting enzymes involved in the biosynthesis of tocotrienols and tocopherols, respectively, and have been the targets of research focused on tocochromanol production (Savidge et al., 2002; Cahoon et al., 2003; Rippert et al., 2004). *HGGT* was originally identified in seeds of monocots, including barley, wheat, and rice and is most active with geranylgeranyl diphosphate (GGDP) as its substrate (Cahoon et al., 2003; Yang et al., 2011; Zhang et al., 2013; Chen et al., 2017). Previous work has shown that heterologous expression of barley *HGGT* in soybean and corn leads to a six to tenfold increase in seed tocochromanols, principally in the form of tocotrienols (Cahoon et al., 2003; Konda et al., 2020). *HPT* primarily uses phytyl diphosphate (PDP) for tocopherol synthesis, but may also appropriate GGDP as a substrate for tocotrienol synthesis when the homogentisate level is high (Figure 1) (Cahoon et al., 2003; Collakova and DellaPenna, 2003; Yang et al., 2011; Zhang et al., 2013).

HGA, produced from the shikimic acid pathway, is substrate for both HGGT and HPT in the initial step of tocochromanol synthesis (Cahoon et al., 2003; Zhang et al., 2013) and is considered a limiting precursor for vitamin E production (Rippert et al., 2004; Karunanandaa et al., 2005; Yang et al., 2011; Stacey et al., 2016). Experiments aimed at generating large increases in HGA concentrations to enhance tocopherol production have resulted, instead, to the unexpected production of tocotrienols, with only small increases in tocopherol concentrations. This result is presumably due to limiting PDP pools for tocopherol biosynthesis that shifts the relative PDP : GGDP ratios to promote HPT-mediated tocotrienol biosynthesis. For example, the deregulated, enhanced HGA production by co-expression of transgenes for the

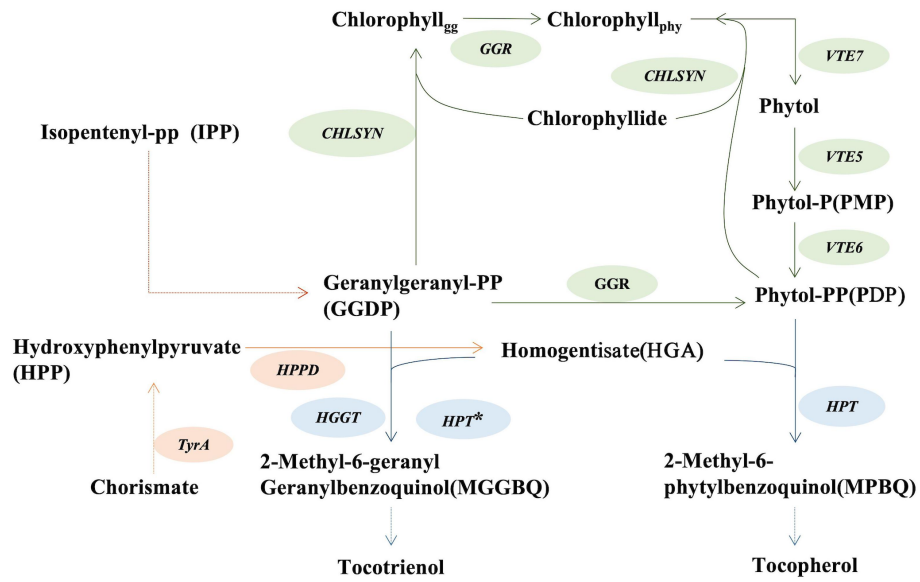


FIGURE 1
Schematic biosynthetic pathways of tocopherol and tocotrienol synthesis. *CHLSYN*, chlorophyll synthetase; *HPPD*, p-hydroxyphenylpyruvate dioxygenase; *HPT*, homogentisate phytyltransferase; *GGR*, geranylgeranyl reductase; *HGGT*, homogentisate geranylgeranyl transferase; *VTE5*, phytol kinase; *TyrA*, prephenate dehydrogenase. *HPT**: When the endogenous level of homogentisate is high, *HPT* can also catalyze the step from GGDP to MGGBQ towards tocotrienol synthesis.

yeast prephenate dehydrogenase and hydroxyphenylpyruvate dioxygenase (HPPD) in tobacco leaves yielded nearly no increase in tocopherol concentrations. This, instead, conferred production of tocotrienols and their accumulation to amounts ~ten-fold higher than tocopherols (Rippert et al., 2004). Similarly, overexpression of the *E. coli tyrA* gene and Arabidopsis *HPPD* gene generated large increases homogentisate concentrations in Arabidopsis leaves, which was accompanied by accumulation of tocotrienols, which are not normally present in Arabidopsis (Figure 1) (Zhang et al., 2013). Furthermore, co-overexpression of *HPPD*, *TyrA*, and *HPT* transgenes in transgenic canola and soybean seeds led to a two- to three-fold increase in total tocochromanol concentrations, primarily as tocotrienols rather than tocopherols (Karunanandaa et al., 2005).

The findings above highlight the feasibility of generating large increases in vitamin E tocochromanols, but these increases are largely the result of enhanced tocotrienol production. By contrast, biofortification of similar large increases in tocopherol content of plant organs has proven more elusive. It has succeeded in only more modest enhancement in tocopherol levels in plants. Findings from the Arabidopsis *vte5* mutant that is impaired in phytol kinase activity have indicated that PDP biosynthesis is a major limitation for tocopherol production in leaves and green seeds. These studies showed that nearly 80% of PDP required for tocopherol synthesis is derived from cycling and reduction of geranylgeraniol through chlorophyll (Valentin et al., 2005; Ischebeck et al., 2006; Zhang et al., 2015). Phytol formed from geranylgeranyl reductase activity using geranygeraniol on chlorophyll is released and converted to PDP by two sequential phosphorylation steps catalyzed by VTE5 and VTE6 kinases (Valentin et al., 2005; Gutbrod et al., 2019; Gutbrod et al., 2021). The resulting PDP is available for use in tocopherol biosynthesis. In Arabidopsis *vte5* mutants, seed tocopherol levels were reduced to 20% of those in wild type plants, suggesting that a large fraction of phytol generated through the chlorophyll turnover is used for PDP and subsequently tocopherol biosynthesis (Valentin et al., 2005; Ischebeck et al., 2006). The Arabidopsis *CHLSYN* knockout mutant had only 20-26% tocopherol in leaves with severe photosynthetic defects, suggesting a major portion of PDP for tocopherol synthesis comes from the chlorophyll salvage pathway (Vom Dorp et al., 2015; Zhang et al., 2015). Downregulation of *CHLSYN* by RNAi in Arabidopsis resulted in reduced chlorophyll content and higher levels of tocopherols in leaves, though tocochromanol levels in seeds were not determined (Zhang et al., 2015). This result was consistent with a negative correlation between *CHLSYN* expression levels and tocopherol content, in accordance with the competition for PDP between the chlorophyll and tocopherol biosynthetic pathways. downregulation of *CHLSYN* expression on tocopherol content in seeds may provide a route for enhancement of overall seed tocopherol concentrations.

In this study, we first found that the tocopherol content in Arabidopsis seeds is also negatively correlated with *CHLSYN* expression levels, as observed in *CHLSYN* overexpression and down-regulation lines. In a seed-specific *CHLSYN* RNAi background, we overexpressed *TyrA* and *HPT* with seed-specific promoters to stimulate tocopherol synthesis with high HGA input. We obtained Arabidopsis seeds displaying high vitamin E content primarily composed of tocopherols. We also found that elevated HGA production shifts the relative amount of PDP and GGDP and determines metabolic flow into tocopherol and tocotrienol biosynthesis. Overall, our findings provides a strategy for vitamin E biofortification of green oilseeds for enhanced tocopherol concentrations.

2 Materials and methods

2.1 Plant materials and growth conditions

Wild-type and transgenic *Arabidopsis thaliana* lines used for this study were of the Columbia-0 ecotype. Arabidopsis *vte5-2* seeds were previously described (Vom Dorp et al., 2015). Homozygous seeds carrying the 35S:*TyrA* and 35S:*AtHPPD*-35S:*TyrA* overexpression constructs were previously generated by our laboratory (Zhang et al., 2013). Plants were grown on plates with ½ MS agar supplemented with 2% (w/w) sucrose. Pot growth was performed in a growth chamber at 22°C under a 16 h day, with light intensity at 100 μmol m^{-2} sec^{-1}.

2.2 Arabidopsis transformation

Arabidopsis plants were grown in a growth chamber in long day conditions. 4-5 weeks old healthy plants were chosen for Agrobacteria transformation using floral dip method (Clough and Bent, 1998; Zhang et al., 2015). The transformed plants were grown and seeds were harvested. Transgenic seeds were selected based on mCherry marker, followed by genotyping using gene-specific primers.

2.3 Vector construction and selection of transgenic plants

The vector pBinGlyRed3 containing a DsRed fluorescent protein marker under the control of the 35S promoter was used in this studies (Jach et al., 2001; Nguyen et al., 2013).

CHLSYN (AT3G51820) was amplified from Arabidopsis cDNA using primers as following: CHLSYN-F: 5'- GCTCTAGA CCGTCGGTTCTATGACTTCGAT-3' and CHLSYN-R: 5'-CCCCTCGAGTCAAAATACGCCTTTTTCAGT-3' (restriction sites underlined). All PCR reactions were performed using Phusion polymerase (Vazyme, Wuhan, China). The PCR product was cloned into pBinGlyRed3 using *Xba*I/*Xho*I sites. The resulting plasmid with *AtCHLSYN* flanked by glycinin promoter and glycinin terminator was designated as SYN-OE.

A specific *AtCHLSYN* 430bp fragment was chosen for constructing the *CHLSYN* RNAi (RNA interference) vector. The forward segment Si01 was amplified using the following primers: Si01-F, 5'-CCGCTCGAGGACGCAATTAATGAGCCATATCG-3' and Si01-R, 5'-GGACTAGTTGCCAAAAGCTA

CTGGGAGAGAC-3'. A second fragment Si02 was amplified using Si02-F: 5'-GCTCTAGAGACGCAATTAATGAGCCATATCG-3' and Si02-R: 5'-CCCAAGCTTTGCCAAAAAG CTACTGGGAGAGA-3'. The Si01 segment was cloned into the pINTRON vector using *Xho*I/*Bcu*I sites. Into the resulting vector pIN-Si01 the second Si02 segment was ligated using *Hind*III/*Xba*I sites (Nguyen et al., 2013). The resulting vector pIN-Si containing a *Not*I fragment with Si01-intron-Si02 was subcloned into vector pBetaCon with glycinin promoter (seed-specific promoter) and phaseolin terminator. Finally, the complete expression cassette (glycinin promoter:Si01-intron-Si02::phaseolin terminator) was sub-cloned into pBinGlyRed3 using the *Sgs*I site. The final interference vector pBinGlyRed3-*CHLSYN*: RNAi was abbreviated as SYN-RNAi.

For *HPT* overexpression, *HPT* (AT2G18950) cDNA was amplified from Arabidopsis cDNA using the following primers: At*HPT-Eco*RI-LP, 5'-CCGAATTCTCACTTCAAAA AAGGTAACAG-3'; At*HPT-Sma*I-RP, 5'-TTCCCGGGATGG AGTCTCTGCTCTCTAGT-3' (restriction sites underlined). The PCR product was cloned into pBinGlyRed3 with *Eco*RI/*Sma*I sites, the resulting vector pBinGlyRed3-35S:At*HPT* was designated 35S:At*HPT*.

The SgsI fragment from 35S:At*HPT* containing a complete cassette was inserted into the *Sgs*I site of pBinGlyRed3 and the *Mlu*I site of Red3-Si to construct pBinGlyRed3 –Ole : At*HPT* (OleAtHPT) and pBinGlyRed3-*CHLSYN*: RNAi + Ole-HPT. Primer Red3-BamHI-F GCGTATGGATTATGGAACTATCA and AtHPT-YZ-R AAAGGAGATATATCAGAAACCTTCTC were used to confirm the transcription orientation of the two cassettes.

2.4 HPLC analysis of tocochromanol content and composition

An HPLC with fluorescence detector was used for quantification of seed tocochromanol contents, using 5,7-dimethyltocol as standard (Yang et al., 2011). ~5 mg of dried seeds, 1 ml methanol/dichloromethane (9:1 v/v) and 5 µl internal standard 5,7-dimethyltocol (Matreya, www.matreya.com) were added to a 2 ml centrifuge tube, the sample was ground with steel balls, then incubate at room temperature for 3 h and centrifuged. Tocochromanols were separated on an Agilent Eclipse XDB-C18 reversed-phase column (4.6 × 150 mm, 5 µm particle size, www.agilent.com), with isocratic conditions of methanol/water (95:5 v/v) at flow rate of 1.5 ml/min. The abundance of each compound was monitored by excitation at 292 nm and emission at 330 nm.

2.5 Semi-quantitative PCR analysis

Total RNA was extracted from seeds harvested 10 days after flowering, with the TRIzol Reagent Kit (Ambion) according to the manufacture's protocol. RNA was reverse transcribed to cDNA using the HiScript II 1st Strand cDNA Synthesis Kit (Vazyme). qRT-PCRs was performed with the SYBRgreen qPCR Master Mix (Vazyme) using the CFX Connect™ real-time PCR detection system (BIO-RAD, Hercules, CA, USA). The following qRT primers were used: qAtHPT-F1-1, 5'-TCGCAAAACCGA AGTTTAGGAAC-3'; qAtHPT-R1-1, 5'-TGTTTGCTATTCGAGTCGAAAGC-3' for AtCHLSYN. Actin7 (AT5G09810) was chosen as reference gene, qRT primers: βActin7-F, 5'-GATATTCAGCCACTTGTCTGTGAC-3'; and βActin7-R: 5'-CATGTTCGATTGGATACTTCAGAG-3'.

2.6 Quantification of GGDP, PDP and HGA contents in seeds

Mature Arabidopsis seeds were used for GGDP and PDP determination (Valentin et al., 2005). Prenyl-transferase assays were based on a previously described method (Collakova and DellaPenna, 2003; Yang et al., 2011). UPLC LC-MS system linked to a QTRAP4500 mass spectrometer was used for the determination of HGA, the method is described by Karunanandaa et al. (2005).

2.7 Statistical analysis

We have quantified the seed VitE contents in transgenic materials with single-copy T-DNA insertion based 3:1 segregation ratio of the mCherry marker in T2 generation to avoid gene dosage effect. All transgenic lines have their corresponding non-transgenic control. Due to the fluctuation of seed VitE contents by environmental factors, the tocopherol contents were normalized to 540 µg/g as "wild type" value based on literature (Karunanandaa et al., 2005) and our empirical Col.0 seed records under conditions used in this study. Positive transgenic seeds (Red) and negative segregant seeds of Ole : HPT/Col-0, (SYN-RNAi+Ole : HPT)/Col-0 and (SYN-RNAi+Ole : HPT)/35sTyrA transgenic lines were separated according to their segregation ratio of 3:1 in the T_2 generation. According to this, we calculated the seed tocochromanol contents in T2 generation based on seeds with red fluorescence accounting for 75% of the total T2 seeds and non-transgenic seeds accouting for 25% from a heterozygous T1 plant.

$$\text{The tocochromanols of T2 seeds} = \text{levels from red (transgenic) seeds} \times 75\% + \text{levels from non} - \text{transgenic seeds} \times 25\%$$

3 Results

3.1 Chlorophyll synthase expression levels and seed tocopherol content are negatively correlated

RNAi lines of *CHLSYN* were generated and annotated as "SYN-RNAi" (Figures 2A, C). The T-DNA in the expression vector for these lines contained a DsRed selection marker to facilitate the screening procedure for transgenic events. From 68 T_1 positive lines, we isolated 3 heterozygous lines from the T_2 generation with a single copy of the

10.3389/fpls.2024.1344095

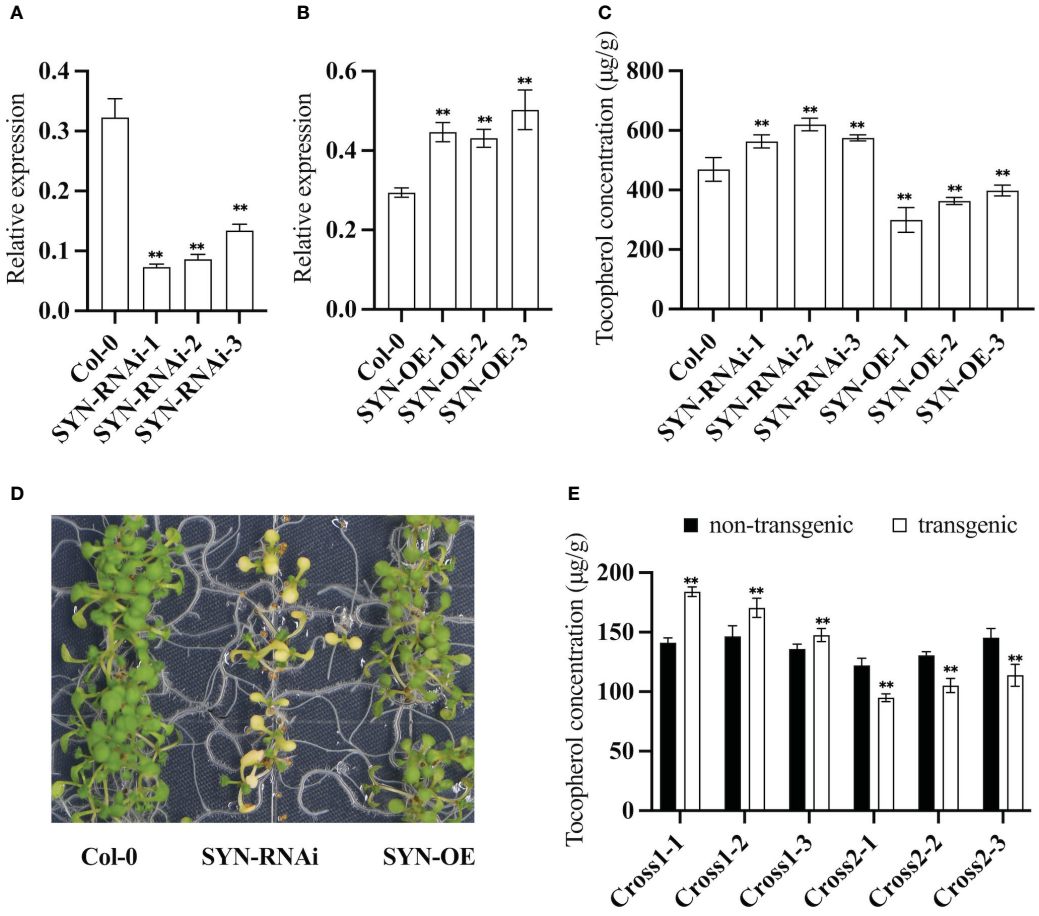


FIGURE 2
Expression pattern of *AtCHLSYN* and quantification of seed tocochromanol contents in Arabidopsis transgenic lines. **(A, B)**, qRT-PCR of *AtCHLSYN* expression in SYN−RNAi (RNAi line of *AtCHLSYN*) or SYN (overexpression line of *AtCHLSYN*) transgenic seeds (10 days after flowering). **(C)** Tocopherol contents in T2 mature seeds of SYN-RNAi and SYN transgenic plants. **(D)** Growth phenotype of young seedlings of SYN-RNAi transgenic plants, SYN and Col-0. **(E)** *vte5-2* homozygous mutants crossed with SYN-OE and SYN-RNAi transgenic lines, cross1 and cross2 represent F3 plants of *vte5* carrying SYN-OE and SYN-RNAi constructs respectively. Tocopherol contents of seeds with DsRed fluorescence (transgenic) and seeds lacking DsRed fluorescence (non-transgenic) from F3 progeny. Seed tocopherol contents were calculated from three biological replicates. ** represents significance level of $p < 0.01$ by student t-test.

SYN-RNAi insertion. qRT-PCR results indicated that RNAi lines had reduced transcript levels of the *CHLSYN* gene (Figure 2A). The SYN-RNAi seeds upon maturation had higher tocochromanol contents, with up to 34% increases of tocopherol concentrations compared to amounts in non-transgenic segregant seeds (Figure 2C). *AtCHLSYN* was also overexpressed under the control of the strong seed-specific glycinin promoter. Positive transgenic lines were selected and annotated as "SYN-OE" (Zhang et al., 2015). We obtained a total of 54 SYN positive lines in the T_1 generation, and picked 3 single-copy insertion T_2 plants in the heterozygous state based on their 3:1 segregation ratio through DsRed marker selection. qRT-PCR results indicated that SYN-OE lines had higher transcript levels of the *CHLSYN* gene than control (Figure 2B). In the T_2 generation, lines with seed-specific over-expression of *CHLSYN* showed large reduction in tocopherol content, with reductions of ≤42% compared to levels in isogenic wild type segregant seeds (Figure 2C). Young seedlings of SYN-RNAi were yellow but gradually turned green during later development, whereas the SYN-OE seedlings were morphologically similar to wild type plants (Figure 2D). Collectively, results from the SYN-OE and SYN-RNAi transgenic lines suggest that transcript levels of *CHLSYN* are negatively correlated with seed tocopherol contents in Arabidopsis transgenic plants.

VTE5 catalyzes the conversion of phytol to PMP during chlorophyll degradation (Figure 1). We investigated whether a negative correlation also exists when the level of PDP is low. For these studies, the *vte5* mutant, which has altered PDP accumulation was used. We crossed the *vte5* mutant with the SYN-RNAi or SYN-OE transgenic plants and measured the seed tocopherol concentrations of the progeny. Measurements were conducted with the homozygous state for *vte5* (knock-out background) and a heterozygous state for SYN-OE or SYN-RNAi to investigate the effect of up- or down- regulation of *CHLSYN* (Figure 2E). As shown in Figure 2, average tocopherol concentrations in non-transgenic *vte5* seeds was 140 μg/g, a reduction in tocopherol concentrations compared to wild type seeds (400~500 μg/g, Figure 2C). Knocking out *VTE5* dramatically decreased the tocopherol content to ~20% of that in wild type. Despite this, a

10.3389/fpls.2024.1344095

negative correlation could still be observed between *CHLSYN* transcript level and tocopherol content in engineered *vte5* seeds (Figure 2E).

3.2 Modest increase in tocopherol content in seeds of transgenic Ole : AtHPT and AtCHLSYN RNAi plants

HPT is a rate-limiting enzyme for tocopherol biosynthesis (Figure 1). Under most conditions, this enzyme only catalyzes the condensation of PDP with HGA during tocopherol synthesis, but in the presence of high levels of homogentisate, the HPT enzyme may also resort to using GGDP for condensation with HGA during tocotrienol synthesis (Yang et al., 2011). Utilizing the seed-specific oleosin promoter to drive *AtHPT* expression, seed tocochromanol contents in transgenic plants were measured (Figure 3). After two generations of selfing from the T_1 positive transgenic lines, 13 T_2 lines with a single copy insertion of Ole : *AtHPT* (Ole : HPT/Col-0) were obtained (Figure 3). These seeds had ~1.8 times the seed tocopherol concentration relative to wild type segregants (Figure 3, Supplementary Table 1). On the basis of this result, we further combined SYN-RNAi with Ole : HPT/Col-0 in a single construct that was transferred into transgenic plants (Figure 3). As shown in Figure 3, the seed tocopherol concentration in (SYN-RNAi+Ole : HPT)/Col-0 plants was ~2.1 times higher than that in non-transgenic segregant plants (Figure 3, Supplementary Table 2). The seeds of (SYN-RNAi+Ole : HPT)/Col-0 plants revealed a higher seed tocopherol content than seeds overexpressing the *HPT* gene alone. These data suggest that the SYN-RNAi transgene may dramatically reduce the *CHLSYN* activity for PDP recycling, and allow more PDP to be available for tocopherol biosynthesis.

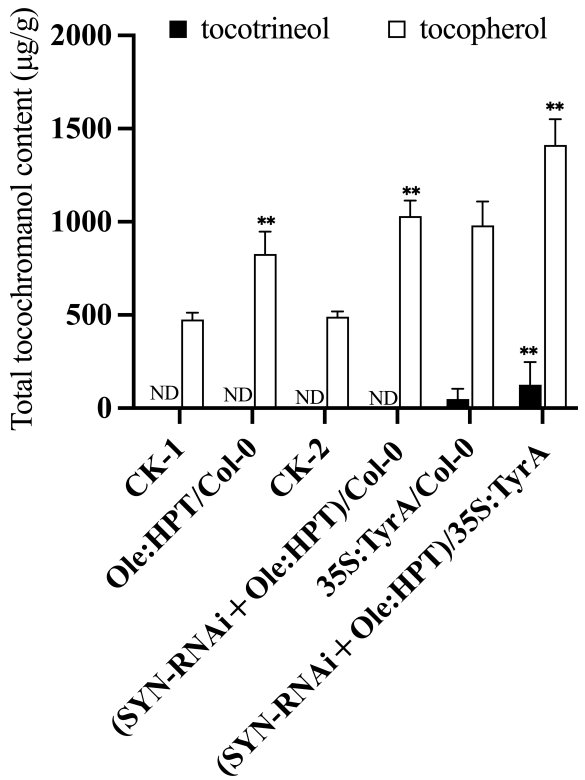


FIGURE 3
Synergistic role of HPT and TyrA overexpression in combination with SYN-RNAi on seed tocopherol or tocotrienol accumulation. Tocochromanol contents in mature T2 seeds from Col-0 transformed with Ole : HPT construct (Ole : HPT/Col-0) and SYN-RNAi+Ole : HPT construct ((SYN-RNAi+Ole : HPT)/Col-0) respectively. CK-1 and CK-2 represent the separated seeds lacking DsRed fluorescence from Ole : HPT/Col-0 and (SYN-RNAi+Ole : HPT)/Col-0 transgenic plants. Tocochromanol contents of seeds from Col-0 transformed with 35S:TyrA construct (35S:TyrA/Col-0) and 35S:TyrA/Col-0 transformed with SYN-RNAi+Ole : HPT construct. Seed tocochromanol contents were calculated from three biological replicates. ND, below the limit of detection. ** represents significance level of $p < 0.01$ by student t-test.

3.3 Seed tocopherol content was increased in seeds with CHLSYN RNAi background carrying Ole : AtHPT and 35S: TyrA constructs

HGA along with PDP is substrate for tocopherol synthesis and is formed from the shikimic acid pathway (Raclaru et al., 2006). Overexpression of *TyrA* has been shown to effectively increase HGA content in Arabidopsis seeds (Karunanandaa et al., 2005). We introduced the (SYN-RNAi+Ole : HPT)/Col-0 construct into a plant homozygous for 35S:TyrA/Col-0. We annotated these lines as "(SYN-RNAi+Ole : HPT)/35S:TyrA". Transgenic T_2 plants with single copy insertions were used for seed tocochromanol content determination (Figure 3). We found that the tocopherol content of non-red seeds (carrying 35S:TyrA only) were 1.7 times of wild type seeds on average (Figure 3, Supplementary Table 3), and tocopherol content of red seeds (carrying (SYN-RNAi+Ole : HPT)/35S:TyrA) were ≤2.7 times those of wild type seeds (Figure 3, Supplementary Table 3). The maximum tocopherol content of (SYN-RNAi+Ole : HPT)/35S:TyrA-9 seeds was or ~1,600 μg/g seed wt or 2.7 times those of the wild type seed concentrations. In our knowledge, this is the highest reported accumulation of tocopherols in Arabidopsis seeds. In contrast, low amounts of tocotrienols were detected in red seeds (carrying (SYN-RNAi+Ole : HPT)/35S:TyrA) and non-red seeds (carrying 35S:TyrA only), with the average tocotrienol content being 8% and 5% of total tocochromanol, respectively (Figure 3, Supplementary Table 3). The RNA interference of the *CHLSYN* gene did not alter the fatty acid content and composition of Arabidopsis seeds in the wild type, 35S:TyrA/Col-0 and (SYN-RNAi+Ole : HPT)/Col-0 backgrounds (Supplementary Table 4).

TyrA-encoded bifunctional chorismate mutase/prephenate dehydrogenase catalyzes the conversion of chorismate to HPP. HPPD catalyzes the conversion of HPP to HGA (Figure 1). A transgenic Arabidopsis plant with *HPPD* and *TyrA* over-expression constructs contains more available HGA than lines with 35S:TyrA/Col-0 only (Zhang et al., 2013). Reciprocal crossing was performed using (35S:TyrA+35S:HPPD)/Col-0 plants with the above (SYN-RNAi+Ole : HPT)/Col-0 plants (Figure 4). In crosses with (35S:TyrA+35S:HPPD)/Col-0-2-4 as maternal parent but different paternal (SYN-RNAi+Ole : HPT)/Col-0 lines, we observed altered tocopherol and tocotrienol concentrations (Figure 4), suggesting that the expression levels of

10.3389/fpls.2024.1344095

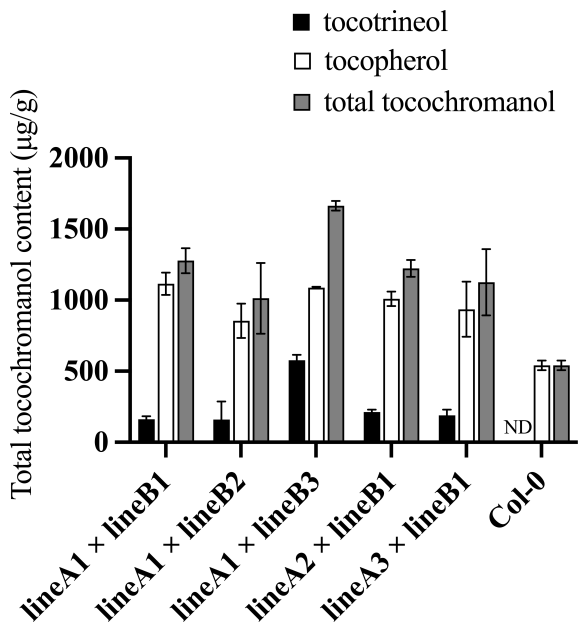


FIGURE 4
Measurements of total tocochromanol contents in seeds harboring the combination of (35S:TyrA+35S:HPPD)/Col-0 and (SYN-RNAi +Ole : HPT)/Col-0 constructs. Total tocochromanol contents from F2 seeds from crosses between lineA: (35S:TyrA+35S:HPPD)/Col-0 and lineB: (SYN-RNAi+Ole : HPT)/Col-0. ND, below the limit of detection.

CHLSYN and *HPT* influence seed vitamin E synthesis. On the other hand, when (SYN-RNAi+Ole : HPT)/Col-0 was fixed but different (35S:TyrA+35S:HPPD)/Col-0 lines used as paternal parent, no major difference was observed (Figure 4), suggesting that *HPT* and *CHLSYN* are dominating factors over *HPPD* and *TyrA* for seed tocopherol and tocotrienol synthesis.

We further measured tocochromanol contents in the seeds of the F_1 progeny. In the crosses, using (35S:TyrA+35S:HPPD)/Col-0 as paternal parent and (SYN-RNAi+Ole : HPT)/Col-0 as maternal recipient, a maximum of 1664 µg/g total tocochromanol was observed in F_1 seeds of (35S:TyrA+35S:HPPD)/Col-0-2-4 x (SYN-RNAi+Ole : HPT)/Col-0-3. The total tocochromanol, which includes considerable amount of tocotrienols (35%) is equivalent to 3.1 times that of the wild type control (Figure 4, Supplementary Table 5). Another reciprocal combination (SYN-RNAi+Ole : HPT)/Col-0-1 ×(35S:TyrA+35S:HPPD)/Col-0-2-4, resulted in slightly lower total tocochromanol but considerably less tocotrienol content (11%, Supplementary Table 5). In the F_2 population of (TyrA+HPPD)-OE-2-4 x (SYN-RNAi+Ole : HPT)/Col-0-3 and (SYN-RNAi+Ole : HPT)/Col-0-1 ×(TyrA+HPPD)-OE-2-4, the maximum seed tocochromanol concentration was 1953 µg/g, which is 3.4 times that of wild type control (Supplementary Table 6). Seed tocopherol concentrations ranging from 1.9 to 2.4 those that of wild type seeds Tocotrienols accounted for 10% to 30% of the total seed tocochromanol (Supplementary Table 6). Overexpression of the *HPPD* gene in the (SYN-RNAi+Ole : HPT)/Col-0-1×35S:TyrA background further increased the total content of tocochromanols, but the content of tocopherol was slightly reduced.

The tocochromanol content of red fluorescent seeds and non-red seeds were separately measured in the T_2 population. In order to compare our results to those of the Karunanandaa et al., 2005 studies, the values in this study were adjusted to the same level (Table 1). The converted results show that the average tocopherol content of (SYN-RNAi+Ole : HPT)/35S:TyrA line was 2.2 times as high as those of the wild type, providing a significant improvement compared to the Napin : HPT+Napin : TyrA line and the Napin : HPPD+Napin: TyrA+Napin : HPT line (Karunanandaa et al., 2005). The average seed tocotrienol of these prior studies was further increased to 54% of the total tocochromanols upon introducing Napin : HPPD due to a higher level of HGA (Savidge et al., 2002; Karunanandaa et al., 2005). In our study, the proportion of tocotrienols was significantly lower than results from

TABLE 1 Comparison of seed tocopherol content between this study and prior results.

Transgenic lines	Tocopherols				Tocotrienols			
	Maximum (µg/g)	Fold	Average (µg/g)	Fold	Maximum (µg/g)	Maximum percentage	Average (µg/g)	Average percentage
Current study								
Ole : HPT/Col-0 [a]	929	1.7x	848	1.6x	0	0%	0	0%
(SYN-RNAi+Ole : HPT)/Col-0 [b]	1195	2.2x	989	1.8x	0	0%	0	0%
(Ole : HPT+SYN-RNAi)/35S:TyrA [c]	1344	2.5x	1186	2.2x	244	16%	97	8%
(SYN-RNAi+Ole : HPT)/Col-0 x (35S: TyrA+35S:HPPD)/Col-0 [d]	1319	2.4x	1158	2.1x	498	27%	271	19%
Karunanandaa et al. (2005)								
Napin : HPT+Napin : TyrA [e]	1179	2.2x	860	1.6x	279	23%	120	12%
Napin : HPPD+Napin : TyrA +Napin : HPT [f]	1022	1.9x	702	1.3x	1688	62%	834	54%

a, b, c: Data from T_2 population.
d: Data from F_2 population.
e,f: Data from Karunanandaa et al. (2005).

10.3389/fpls.2024.1344095

Karunanandaa et al., 2005. By inhibiting the expression of the *CHLSYN* gene, the average seeds tocotrienol contents vary from 8% to 19% in (SYN-RNAi+Ole : HPT)/35S:TyrA line and (SYN-RNAi +Ole : HPT)/Col-0 x (35S:TyrA+35S:HPPD)/Col-0 line (Table 1).

3.4 High HGA and GGDP led to the tocotrienol biosynthesis in Arabidopsis seeds

Based on the results above, it is observed that genes from different metabolic pathways can be used together to boost seed vitamin E content. When comparing the vitamin E contents and compositions between (SYN-RNAi+Ole : HPT)/35S:TyrA and (SYN-RNAi+Ole : HPT)/Col-0 x (35S:TyrA+35S:HPPD)/Col-0, a substantial amount of tocotrienol was observed in the latter case. In order to verify the synthesis of tocotrienol, we quantified PDP, GGDP and HGA contents in mature seeds of 35S:TyrA/Col-0 and (35S:TyrA+35S:HPPD)/Col-0 (Figure 5). High HGA contents were detected in two transgenic seeds with similar level (Figure 5A). We also found that GGDP levels in transgenic seeds were increased in various degrees, the improvement of GGDP contents in (35S:TyrA+35S:HPPD)/Col-0 seeds were ten-fold than those in wild type seeds, it is far greater than the 35S: TyrA/Col-0 seeds (Figure 5B). Our results from seeds confirm the conclusion that a high level of HGA and GGDP concentration lead to the tocotrienol synthesis in seeds. PDP levels of (35S:TyrA +35S:HPPD)/Col-0 seeds were doubled increased than wild type, rather than 35S:TyrA/Col-0 seeds (Figure 5C). We hypothesize that high GGDP contents might be the reason for the improvement of PDP, but high GGDP/PDP ratios is unfavorable to the synthesis of tocopherols.

4 Discussion

In this work, we used biotechnological approaches to increase total Arabidopsis seed tocochromanol concentrations, and in particular, tocopherol concentrations. First, our results confirmed that the negative correlation between *CHLSYN* expression and tocochromanol synthesis, previously observed in Arabidopsis leaves, also occurs in seeds. Second, we achieved the highest report tocopherol concentrations in mature Arabidopsis seeds via genetic combination of Ole : HPT, 35S:TyrA/Col-0 and SYN-RNAi (Table 1). In the transgenic lines with *TyrA* and *HPPD* overexpression, the resulting high HGA and GGDP input triggered HPT activity to use GGDP as a substrate for tocotrienol synthesis (Figure 6). In combination with SYN-RNAi, which increases the pool size of PDP available for tocopherol synthesis, HPT-OE and 35S:TyrA/Col-0 create a genetic background favorable for high accumulation of tocopherol, resulting in a maximum of 2.5 times the seed tocopherol elevation compared to WT (Table 1).

A key conclusion from this study is the negative correlation between chlorophyll synthase (*CHLSYN*) expression levels and tocopherol concentrations in Arabidopsis seeds. Indeed, we propose that the metabolic flow from either GGDP or PDP to tocotrienol or tocopherol synthesis, respectively, is critical for the final proportion of tocochromanols in mature seeds (Figure 6). Our data suggests that starting from the background of *CHLSYN* downregulation and *HPT* overexpression, simultaneous overexpression of *HPPD* and *TyrA* resulted in an optimal accumulation of total tocochromanol and tocopherols (Figure 6B). The overexpression of *HPPD* and *TyrA* boosts HGA supply and therefore increases the pool size of GGDP and PDP (Figure 5, Figures 6A, B). Furthermore, at high levels of HGA, *HPT*

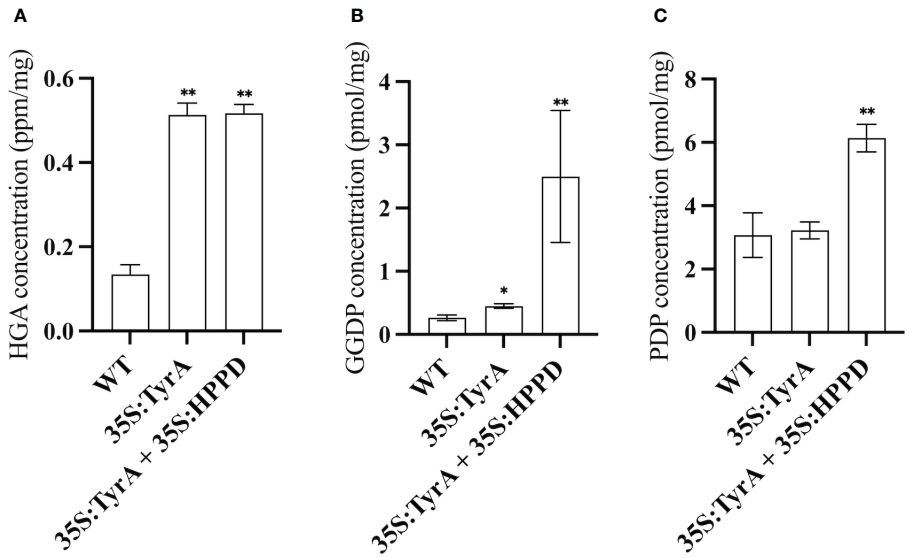


FIGURE 5
Measurement of HGA **(A)**, GGDP **(B)**, PDP **(C)** contents in seeds of 35S:TyrA/Col-0, 35S:TyrA+35S:HPPD plants. Data calculated from three biological replicates, * and ** indicate significant differences at $p < 0.05$ and $p < 0.01$ respectively, by Student's t-test.

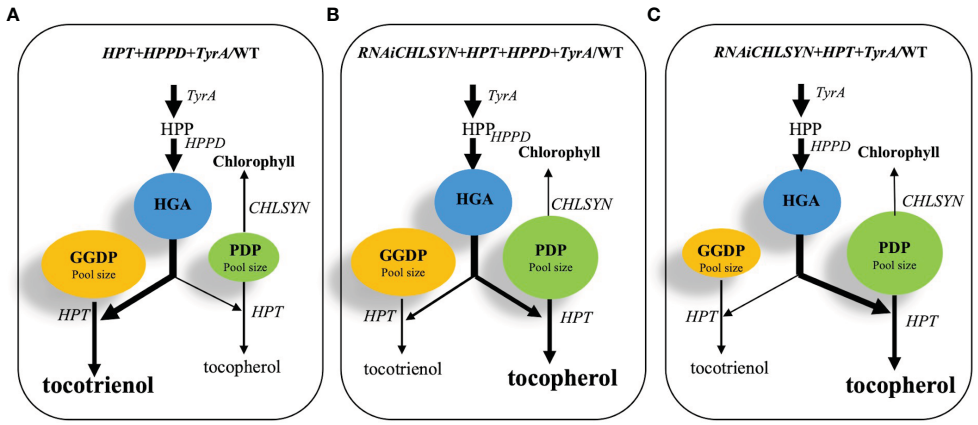


FIGURE 6
Schematic model for the metabolic engineering of *CHLSYN*, *TyrA*, *HPPD* and *HPT* genes to achieve high seed tocochromanol contents in Arabidopsis. **(A–C)**, three scenarios of combinations of different genetic backgrounds for the chlorophyll salvage pathway, the *TyrA* and *HPPD* pathway regulating HGA content, and the *HPT* expression level alter the balance between GGDP and PDP, resulting in tocotrienol or tocopherol synthesis, respectively. The thickness of the arrows represents flux through the metabolic pathway; the sizes of the blue, green and yellow circles showing HGA, PDP and GGDP contents were adjusted according to the metabolite pool size under each scenario. The dominating product of the pathway, i.e. tocopherol or tocotrienol, were shown in bold letters. According to the results, **(B)** scenario would be the best option, resulting in elevated PDP and HGA levels, as well as a favored condition for the tocopherol biosynthesis.

reveals a tendency toward using both GGDP and PDP for tocotrienol and tocopherol synthesis, respectively. Introduction of the *CHLSYN* RNAi construct results in the increased availability of PDP for tocopherol synthesis. Together with the elevated pool sizes of both GGDP and PDP, the balance shifts towards tocopherol production resulting in high seed tocopherol accumulation (Figure 6B). As seed tocopherol accumulation is a highly dynamic process, it would be useful to quantify the level of PDP and GGDP at different stages of seed development. However, the technique of PDP and GGDP measurement in minute amounts of seed material is not currently available. Although we could not observe a significant change of the PDP and GGDP pools in CHLSYN-RNAi or CHLSYN-OE seeds, the effect from (35S:TyrA+35S:HPPD)/Col-0 is apparent (Figure 5). Pool sizes at seed maturation reflect biosynthetic ability during seed development. Accordingly, we propose that the content and ratio of GGDP and PDP are critical for metabolic flow toward the synthesis of tocotrienols or tocopherols, as previously suggested (Yang et al., 2011; Zhang et al., 2013).

The genetic manipulations reported in this study can be theoretically used for the improvement of tocopherol concentrations in any photosynthetic "green" oilseed crop, including canola and soybean. While most studies on oilseeds have achieved significant increases in tocotrienol concentrations, enhancing tocopherol accumulation has been more elusive. Our current work demonstrates that RNAi of *CHLSYN* in the chlorophyll salvage pathway can reduce the final proportion of tocotrienols in total seed tocochromanols. We proposed a scenario in which overexpression of *TyrA* and *HPT* combined with *CHLSYN* suppression created a condition that favors tocopherol biosynthesis (Figure 6B). These data will provide important information to guide future genetic engineering for increasing seed tocopherol contents in oil seed crops. Since many crop species, e.g. *Brassica napus*, are polyploid, in contrast to *Arabidopsis*, the numbers of *CHLSYN* homologs are higher. Therefore, the use of CRISPR technology would be instrumental in generating mutants for disrupting some of the multiple *CHLSYN* loci for increasing PDP input. Based on these mutant backgrounds, suitable genetic material accumulating elevated amounts of tocopherols in mature seeds could be obtained by overexpressing *TyrA* and *HPT*.

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

CZ: Conceptualization, Funding acquisition, Resources, Writing – review & editing. PQ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing. PC: Conceptualization, Formal Analysis, Investigation, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. YWZ: Writing – review & editing. WZ: Conceptualization, Data curation, Formal Analysis, Writing – review & editing. YYZ: Formal Analysis, Investigation, Methodology, Writing – review & editing. JX: Writing – review & editing. LG: Funding acquisition, Resources, Visualization, Writing – review & editing. YL: Investigation, Validation, Visualization, Resources. JR: Investigation, Writing – review & editing, Data

curation, Formal Analysis, Methodology. PD: Data curation, Formal Analysis, Investigation, Methodology, Writing – review & editing. EC: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study is supported by the National Key R&D plan (2022YFD1200400). This research was financially supported by the National Natural Science Foundation of China (Grant No: 31671723 to CZ and Grant No: 31801392 to LG). EC was funded by the United States Department of Agriculture-National Institute of Food and Agriculture, Grant number 2015-67013-22839.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1344095/full#supplementary-material

References

Cahoon, E. B., Hall, S. E., Ripp, K. G., Ganzke, T. S., Hitz, W. D., and Coughlan, S. J. (2003). Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. *Nat. Biotechnol.* 21, 1082–1087. doi: 10.1038/nbt853

Chen, J., Liu, C., Shi, B., Chai, Y., Han, N., Zhu, M., et al. (2017). Overexpression of *HvHGGT* enhances tocotrienol levels and antioxidant activity in barley. *J. Agric. Food Chem.* 65, 5181–5187. doi: 10.1021/acs.jafc.7b00439

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

Collakova, E., and DellaPenna, D. (2003). Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiol.* 131, 632–642. doi: 10.1104/pp.015222

Grusak, M. A., and DellaPenna, D. (1999). Improving the nutrient composition of plants to enhance human nutrition and health. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 133–161. doi: 10.1146/annurev.arplant.50.1.133

Gutbrod, K., Romer, J., and Dörmann, P. (2019). Phytol metabolism in plants. *Prog. Lipid Res.* 74, 1–17. doi: 10.1016/j.plipres.2019.01.002

Gutbrod, P., Yang, W., Grujicic, G. V., Peisker, H., Gutbrod, K., Du, L. F., et al. (2021). Phytol derived from chlorophyll hydrolysis in plants is metabolized *via* phytenal. *J. Biol. Chem.* 296, 100530. doi: 10.1016/j.jbc.2021.100530

Hunter, S. C., and Cahoon, E. B. (2007). Enhancing vitamin E in oilseeds: unraveling tocopherol and tocotrienol biosynthesis. *Lipids.* 42, 97–108. doi: 10.1007/s11745-007-3028-6

Ischebeck, T., Zbierzak, A. M., Kanwischer, M., and Dörmann, P. (2006). A salvage pathway for phytol metabolism in Arabidopsis. *J. Biol. Chem.* 281, 2470–2477. doi: 10.1074/jbc.M509222200

Jach, G., Binot, E., Frings, S., Luxa, K., and Schell, J. (2001). Use of red fluorescent protein from Discosoma sp. (dsRED) as a reporter for plant gene expression. *Plant J* 28, 483–491. doi: 10.1046/j.1365-313X.2001.01153.x

Kanwischer, M., Porfirova, S., Bergmüller, E., and Dörmann, P. (2005). Alterations in tocopherol cyclase activity in transgenic and mutant plants of *Arabidopsis* affect tocopherol content, tocopherol composition, and oxidative stress. *Plant Physiol* 137, 713–723. doi: 10.1104/pp.104.054908

Karunanandaa, B., Qi, Q., Hao, M., Baszis, S. R., Jensen, P. K., Wong, Y.-H. H., et al. (2005). Metabolically engineered oilseed crops with enhanced seed tocopherol. *Metab. Eng.* 7, 384–400. doi: 10.1016/j.ymben.2005.05.005

Kmiecik, D., Fedko, M., Siger, A., and Kulczyński, B. (2019). Degradation of tocopherol molecules and its impact on the polymerization of triacylglycerols during heat treatment of oil. *Molecules.* 24, 4555. doi: 10.3390/molecules24244555

Konda, A. R., Nazarenus, T. J., Nguyen, H., Yang, J., Gelli, M., Swenson, S., et al. (2020). Metabolic engineering of soybean seeds for enhanced vitamin E tocochromanol content and effects on oil antioxidant properties in polyunsaturated fatty acid-rich germplasm. *Metab. Eng.* 57, 63–73. doi: 10.1016/j.ymben.2019.10.005

Lichtenthaler, H. K. (1968). Distribution and relative concentrations of lipophilic plastid quinones in green plants. *Planta.* 81, 140–152. doi: 10.1007/BF00417443

Muñoz, P., and Munné-Bosch, S. (2019). Vitamin E in plants: biosynthesis, transport, and function. *Trends Plant Science.* 24, 1040–1051. doi: 10.1016/j.tplants.2019.08.006

Netscher, T. (2007). Synthesis of vitamin E. *Vitamins Hormones.* 76, 155–202. doi: 10.1016/S0083-6729(07)76007-7

Nguyen, H. T., Silva, J. E., Podicheti, R., Macrander, J., Yang, W., Nazarenus, T. J., et al. (2013). Camelina seed transcriptome: a tool for meal and oil improvement and translational research. *Plant Biotech. J.* 11, 759–769. doi: 10.1111/pbi.12068

Raclaru, M., Gruber, J., Kumar, R., Sadre, R., Lühs, W., Zarhloul, M. K., et al. (2006). Increase of the tocochromanol content in transgenic *Brassica napus* seeds by overexpression of key enzymes involved in prenylquinone biosynthesis. *Mol. Breeding.* 18, 93–107. doi: 10.1007/s11032-006-9014-5

Rippert, P., Scimemi, C., Dubald, M., and Matringe, M. (2004). Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiol.* 134, 92–100. doi: 10.1104/pp.103.032441

Sattler, S. E., Cheng, Z., and DellaPenna, D. (2004). From *Arabidopsis* to agriculture: engineering improved vitamin E content in soybean. *Trends Plant Sci.* 9, 365–367. doi: 10.1016/j.tplants.2004.06.002

Savidge, B., Weiss, J. D., Wong, Y.-H. H., Lassner, M. W., Mitsky, T. A., Shewmaker, C. K., et al. (2002). Isolation and characterization of homogentisate phytyltransferase genes from Synechocystis sp. PCC 6803 and *Arabidopsis*. *Plant Physiol.* 129, 321–332. doi: 10.1104/pp.010747

Shintani, D., and DellaPenna, D. (1998). Elevating the vitamin E content of plants through metabolic engineering. *Science.* 282, 2098–2100. doi: 10.1126/science.282.5396.2098

Stacey, M. G., Cahoon, R. E., Nguyen, H. T., Cui, Y., Sato, S., Nguyen, C. T., et al. (2016). Identification of homogentisate dioxygenase as a target for vitamin E biofortification in oilseeds. *Plant Physiol.* 172, 1506–1518. doi: 10.1104/pp.16.00941

Thiele, J. J., and Ekanayake-Mudiyanselage, S. (2007). Vitamin E in human skin: organ-specific physiology and considerations for its use in dermatology. *Mol. Aspects Med.* 28, 646–667. doi: 10.1016/j.mam.2007.06.001

Valentin, H. E., Lincoln, K., Moshiri, F., Jensen, P. K., Qi, Q., Venkatesh, T. V., et al. (2005). The *Arabidopsis* vitamin E pathway gene*5-1* mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. *Plant Cell.* 18, 212–224. doi: 10.1105/tpc.105.037077

Vom Dorp, K., Hölzl, G., Plohmann, C., Eisenhut, M., Abraham, M., Weber, A. P. M., et al. (2015). Remobilization of phytol from chlorophyll degradation is essential for tocopherol synthesis and growth of *Arabidopsis*. *Plant Cell.* 27, 2846–2859. doi: 10.1105/tpc.15.00395

Warner, K., Neff, W. E., and Eller, F. J. (2003). Enhancing quality and oxidative stability of aged fried food with γ-tocopherol. *J. Agri Food Chem.* 51, 623–627. doi: 10.1021/jf020937e

Yang, C. S., Luo, P., Zeng, Z., Wang, H., Malafa, M., and Suh, N. (2020). Vitamin E and cancer prevention: Studies with different forms of tocopherols and tocotrienols. *Mol. Carcinog.* 59, 365–389. doi: 10.1002/mc.23160

Yang, W., Cahoon, R. E., Hunter, S. C., Zhang, C., Han, J., Borgschulte, T., et al. (2011). Vitamin E biosynthesis: functional characterization of the monocot homogentisate geranylgeranyl transferase. *Plant J.* 65, 206–217. doi: 10.1111/j.1365-313X.2010.04417.x

Zhang, C., Cahoon, R. E., Hunter, S. C., Chen, M., Han, J., and Cahoon, E. B. (2013). Genetic and biochemical basis for alternative routes of tocotrienol biosynthesis for enhanced vitamin E antioxidant production. *Plant J.* 73, 628–639. doi: 10.1111/tpj.12067

Zhang, C., Zhang, W., Ren, G., Li, D., Cahoon, R. E., Chen, M., et al. (2015). Chlorophyll synthase under epigenetic surveillance is critical for vitamin E synthesis, and altered expression affects tocopherol levels in *Arabidopsis*. *Plant Physiol.* 168, 1503–1511. doi: 10.1104/pp.15.00594

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 March 2024
DOI 10.3389/fpls.2024.1368894

OPEN ACCESS

EDITED BY
Naoki Kitaoka,
Hokkaido University, Japan

REVIEWED BY
Bruno Esteves,
Instituto Politecnico de Viseu, Portugal
Kengo Shigetomi,
Hokkaido University, Japan

*CORRESPONDENCE
Yunjie Gu
guyunjie7901@163.com
Xueqin Wan
wanxueqin@sicau.edu.cn

†These authors have contributed equally to this work

RECEIVED 11 January 2024
ACCEPTED 11 March 2024
PUBLISHED 26 March 2024

CITATION
Yang H, Zhang S, Gu Y, Peng J, Huang X, Guo H, Chen L, Jiang Y, Liu M, Luo X, Xie J and Wan X (2024) Identification and variation analysis of the composition and content of essential oil and fragrance compounds in *Phoebe zhennan* wood at different tree ages. *Front. Plant Sci.* 15:1368894. doi: 10.3389/fpls.2024.1368894

Identification and variation analysis of the composition and content of essential oil and fragrance compounds in *Phoebe zhennan* wood at different tree ages

Hanbo Yang[1†], Shuaiying Zhang[1†], Yunjie Gu[2*], Jian Peng[2], Xin Huang[1], Hongying Guo[3], Lianghua Chen[1], Yongze Jiang[1], Minhao Liu[2], Xiandan Luo[1], Jiaxin Xie[1] and Xueqin Wan[1*]

[1]Forestry Ecological Engineering in the Upper Reaches of the Yangtze River Key Laboratory of Sichuan Province, National Forestry and Grassland Administration Key Laboratory of Forest Resource Conservation and Ecological Safety on the Upper Reaches of the Yangtze River, College of Forestry, Sichuan Agricultural University, Chengdu, China, [2]Sichuan Academy of Forestry, Sichuan Key Laboratory of Ecological Restoration and Conservation for Forest and Wetland, Chengdu, China, [3]Sichuan Academy of Grassland Sciences, Chengdu, China

Wood essential oil and wood products with special fragrances are high value-added forest products. Despite the availability of essential oil and volatile organic compounds (VOCs) from *Phoebe zhennan* wood, their variation and dependence on tree age have not been examined. After essential oil extraction and wood processing, the yields and compositions of essential oils and VOCs in wood from *P. zhennan* trees of different ages (10a, 30a, and 80a) were determined. The yield of essential oil from 30a wood was significantly greater than that from 10a and 80a wood. Liquid chromatography–mass spectrometry (LC–MS) and gas chromatography–mass spectrometry (GC–MS) revealed 672 and 41 volatile compounds, respectively, in the essential oil and wood, the majority of which exhibited large fluctuations in relative content and composition depending on tree age. Sesquiterpenoids, fatty acids and conjugates may greatly contribute to the main components of essential oil from wood. Almost all major sesquiterpenoid compounds, such as caryophyllene α-oxide, eudesmo, and cubebene, were identified in the essential oils from the 30a and 80a wood, and their relative contents were much greater than those in the 10a wood. The main components of the wood fragrance were sesquiterpenoids. The types and relative contents of sesquiterpenoids from wood increased with tree age. These results suggest that choosing wood from trees of a suitable age will significantly improve the efficiency of wood utilization.

KEYWORDS

volatile organic compounds, essential oil, wood, *Phoebe zhennan*, GC-MS, LC-MS

1 Introduction

Plant-synthesized volatile organic compounds (VOCs) are metabolites that play crucial roles in the processes enabling plant survival under challenging conditions, such as defense against natural enemies (Wang et al., 2023). The VOCs of plants are composed of more than 1700 volatile substances, including terpenes (especially sesquiterpenoids and monoterpenes), phenols, benzoic acid derivatives, and aliphatic compounds (Shen et al., 2023). The unique fragrance of woody plants is one of the most distinctive characteristics of wood products (Wang et al., 2006). It has been suggested that fragrance directly stimulates the limbic lobe and hypothalamus, thus having a profound effect on the mind and body (Kodis et al., 1998). Therefore, a pleasant fragrance is one of the key drivers of the high economic value of wood. The unique fragrance of wood is due to its extractives, which are secondary metabolites of VOCs. Wang et al. (2006) analyzed the fragrance compositions of six precious coniferous wood and reported that β-elemene, myrtenol, β-cedrene, β-cedrene, 3-carene, and *p*-cymene were the compounds with the strongest fragrance in *Chamaecyparis obtusa*, *Chamaecyparis formosensis*, *Taiwania cryptomerioides*, *Cunninghamia lanceolata*, *Cryptomeria japonica*, and *Calocedrus macrolepis*, respectively. For example, α-cadinol, cedrol, and α-terpineol contribute to the fragrance of *T. cryptomerioides*, *C. lanceolata*, and *C. obtusa*, respectively, and the fragrance of these compounds is similar to that of the wood from which they were isolated (Chang et al., 1998). In agarwood, more than 150 compounds, mostly sesquiterpenoids, chromones, and volatile aromatic compounds, have been identified in the fragrant heartwood (Das et al., 2022). Distinct sesquiterpenes and gurjunene are the main components that contribute to the uniqueness of the fragrance of Dipterocarpaceae wood (Wang et al., 2022). In addition to the genetic characteristics of tree species, tree age also has a significant effect on the composition of VOCs. In *Platycladus orientalis*, the compositions of leaf VOCs varied among trees of different ages (Cui et al., 2023). However, the fragrance compounds of only a few plants have been studied, and the composition and formation of fragrant molecules are still not fully understood, especially in the wood products of woody perennials.

Essential oils (EOs) are extracted from aromatic plants and are lipid-soluble, volatile natural compounds with a rich aroma (Shao et al., 2020). EOs are predominantly used in perfumes, cosmetics, and food flavoring due to the presence of these strong aromas (Wani et al., 2021). In addition, continuous research has demonstrated the immense potential of EOs and their constituent chemical species in the management, protection, and treatment of several human diseases (Caputi and Aprea, 2011; Wani et al., 2021). Several factors influence the yield and composition of EOs. Genotype, environment, and genotype × environment interaction have important effects on the yield and composition of plant EOs (Srivastava et al., 2021; Kumar and Lal, 2022; Kumar et al., 2022). Some authors have shown that the composition of conifer oils is not influenced by environmental conditions but determined only at the genetic level (Koukos et al., 2001). For woody perennials, the yield and composition of EOs also differed among parts (stem bark, branch bark, wood, and leaves) of the tree. For instance, there were 17 compounds in EOs extracted from the bark of *Phoebe zhennan* but 24 compounds in leaf oil (Shao et al., 2020). A variety of EO compounds also exist in *Cinnamomum camphora*; for example, the EO yields of leaves and branches are more than two times greater than those of wood, and the oxygenated monoterpene camphor is the major component in all tissues, except for safrole in the roots (Poudel et al., 2021). The yield of EOs from *C. japonica* decreased in the order of leaf>bark>heartwood>sapwood, and the compounds of EOs from the four tissues showed obvious differences (Cheng et al., 2005). Nevertheless, the results of Han et al. (2022) for 297 *Phoebe bournei* genotypes from 33 provenances showed no significant differences in the contents of monoterpenoids and sesquiterpenoids between leaves and stems. The optimal yield and composition will vary with differences in the physiological age of the tree. In their reports, Gathara et al. (2022) determined that tree age but not soil nutrients and agro-ecological factors significantly influenced the oil yield in *Osyris lanceolata*. Tree age affects the components of EOs in the leaves, branch bark, and stem bark of *Cinnamomum burmannii* (Fajar et al., 2019). Understanding the optimal tree age for the extraction of EOs will provide a useful reference for optimizing harvest strategies and increasing productivity on a larger production scale.

Phoebe zhennan S. Lee et F.N. Wei (nanmu) belongs to the family Lauraceae and genus *Phoebe* Nees and is mainly distributed in the subtropical evergreen broad-leaved forests (EBLFs) of China; it is widely cultivated in China, especially in Sichuan Province (Zhu et al., 2022). Nanmu wood is the main source of "golden thread nanmu", which has high economic value, and harbors EOs. The tree has an attractive visible golden thread pattern, a rich fragrance, and high durability (Xiao et al., 2020; Yang et al., 2022a). Due to these qualities, *P. zhennan* wood has become popular for use in furniture, various high-class furnishing materials, cosmetics, health products, among others. The price of nanmu wood in China is greater than \$700/m^3, and the price can reach \$1400/m^3 if the diameter at breast height (DBH) of the tree exceeds 30 cm (Cheng et al., 2023). In their excellent review, Han et al. (2022) reported that Lauraceae species produce VOCs with high economic value in the spice and perfume industries. Trees of *Phoebe*, a genus in the Lauraceae, emit a scent dominated by sesquiterpenoids (Joshi et al., 2009; Xie et al., 2015; Ding et al., 2020). In our previous study, the main fragrance-related metabolites in *P. zhennan* heartwood were cadinene, *p*-cymene, and 1,3,5-triisopropylbenzene, which are sesquiterpenoids, monoterpenoids, and aromatic hydrocarbons, respectively (Yang et al., 2023). Differences in the contents and major chemical compounds between the bark and leaf oils were observed in *P. zhennan* (Shao et al., 2020). The determination of the main chemical components of EOs and VOCs affecting the quality of the wood and the establishment of quantitative methods can effectively help determine the quality of the wood in *P. zhennan*; however, these main components and tree age effects are currently unknown. In this study, *P. zhennan* wood of different ages were used as research materials. The aim of this study was to determine the influence of tree age on the yield and composition of EOs extracted from wood and the VOCs of wood. The results provide a reference for the utilization of EOs from *P. zhennan* and wood

improvement of *P. zhennan*. Furthermore, this study also provides a reference for wood improvement and utilization in other tree species.

2 Materials and methods

2.1 Wood samples

Wood samples (10a, 30a, and 80a) were collected from five *P. zhennan* trees of each age (five biological replicates) planted in Yibin city, Sichuan Province, China (E104.599400°, N28.208611°). The trunk at DBH (1.3 m) was harvested and air-dried for further use. The wood was ground to powder and then sifted through sieves (0.5 mm and 0.25 mm). The powder was collected for wood EO and VOC extraction.

2.2 Wood EO extract

The steam distillation method was used to extract EOs from *P. zhennan* wood (Shao et al., 2020). Fifty grams of wood powder from a tree of each age was soaked in distilled water until moistened and then put in a Soxhlet drawer for reflux extraction for 6 h. After stratification, the EOs were collected for liquid chromatography –mass spectrometry (LC–MS) and gas chromatography–mass spectrometry (GC–MS) analysis.

2.3 VOC extraction

The solid-phase microextraction (SPME) method was used to extract VOCs from *P. zhennan* wood. Five grams of wood powder was placed in an extraction bottle that was preaged in the sample injection port of a gas chromatograph (250°C for 120 min) and preheated (30°C for 10 min). Then, the extraction fibers with 100 μm polydimethylsiloxane (PDMS) were inserted into the extraction bottle to a distance from the wood powder of 0.5 cm and allowed to adsorb for 30 min at a constant temperature of 30°C. The extraction fibers were subjected to GC–MS desorption for 3 min (at 250°C) to perform GC–MS analysis.

2.4 GC–MS analysis

The EOs and VOCs were analyzed using an Agilent GC8890 plus MS5977 instrument (Agilent, Shanghai, China). The chromatographic column was an Agilent 19091S-433UI (30 m × 250 μm × 0.25 μm, -60°C-325°C). The injection volume was 1 μL. The temperature program employed for EO detection had the following settings: an initial temperature of 50°C, an increase to 120°C at 5°C/min, an increase to 180°C at 4°C/min, and then an increase to 300°C at 10°C/min. The ion source and quadrupole were 200 and 150°C, respectively. The temperature program employed for VOC detection had the following settings: an initial temperature of 40°C, an increase to 140°C at 10°C/min, an increase to 200°C at 5°C/min, and then an increase to 300°C at 8°C/min. The ion source and quadrupole were 230 and 150°C, respectively.

2.5 LC–MS analysis

The EOs were analyzed using an ACQUITY UPLC System (Water, Milford, MA, USA) with an ACQUITY UPLC® HSS T3 (150 × 2.1 mm, 108 μm) (Water, Milford, MA, USA). LC-ESI (+)-MS and LC-ESI (-)-MS were separated using the method described by Zelena et al. (2009). Mass spectrometric detection of metabolites was performed on a Q Exactive instrument (Thermo Fisher Scientific, USA) with an ESI ion source. Simultaneous MS1 and MS/MS (full MS-ddMS2 mode, data-dependent MS/MS) acquisition was used (Want et al., 2013).

2.6 Identification and quantification of metabolites

The metabolites were identified with GC–MS was performed using Agilent Mass Hunter software and the public databases HMDB (Wishart et al., 2007), Massbank (Horai et al., 2010), LipidMaps (Sud et al., 2007), and mzCloud (Abdelrazig et al., 2020). The metabolites detected with LC–MS were identified using the public HMDB (Wishart et al., 2007), massbank (Horai et al., 2010), LipidMaps (Sud et al., 2007), mzCloud (Abdelrazig et al., 2020), and KEGG (Kanehisa and Goto, 2000) databases and the self-built database of BioNovoGene (Chengdu, Sichuan, China) (http://query.biodeep.cn/) with the following parameters: retention time, ppm (<30 ppm), and fragmentation model. Finally, the relative contents of the corresponding metabolites are presented as the percentage of the chromatographic peak area integral relative to the total identified peak area integrals.

2.7 Data analysis

To discriminate the metabolite variation among 10a, 30a, and 80a wood, principal component analysis (PCA) and supervised orthogonal partial least squares discriminant analysis (OPLS-DA) were performed after Pareto-scaling normalization. Furthermore, hierarchical cluster analysis (HCA) was conducted to analyze the accumulation modes of metabolites among different samples. DAMs were screened under the following filtering conditions: variable importance in projection (VIP)≥1 and absolute log2(fold change)≥1 (P<0.05). Finally, metabolic pathway enrichment analysis was performed based on the KEGG database in MetaboAnalyst 5.0 (Xia and Wishart, 2011). DAMs annotated against the KEGG database were enriched in different KEGG pathways, and pathways with a KEGG enrichment ratio with P<0.05 were considered significantly enriched and analyzed.

3 Results

3.1 Metabolite identification of EOs

The color of the EOs from 10a and 30a wood was yellow–green but duller than that of the EOs from 80a wood (Supplementary Figure 1A). The yield of EOs from wood of differently aged trees was 2.29-7.59 μL/g (Supplementary Figure 1B). There was a significant difference in the EO content among wood from trees of different ages, and the EO yield in 30a wood was significantly greater than that in 10a and 80a wood. To evaluate the composition and variation of EOs and VOCs among 10a, 30a, and 80a wood from *P. zhennan*, an untargeted metabolomics approach was carried out using LC–MS and GC–MS to analyze the metabolic compounds of each sample. After data preprocessing, a total of 594 and 75 metabolites were obtained from the EOs by LC–MS and GC–MS, respectively. The 594 metabolites obtained with LC–MS were identified and classified into 14 superclasses, including 109 lipids and lipid-like molecules, 94 benzenoids, and 69 organic acids and derivatives (Figure 1A and Supplementary Table 1). Of these, the 109 lipids and lipid-like molecules included 36 fatty acids and conjugates, 15 sesquiterpenoids, and 12 monoterpenoids (Figure 1B). In general, only a few metabolites in EOs with low relative contents showed small differences in composition and relative content between trees of the same age (Supplementary Tables 2, 4). Cubebene was the major metabolite identified in the EOs from the 10a and 80a wood; in contrast, carvone was the major compound in the EOs from 30a wood (Table 1 and Supplementary Table 2). Forty-four of 75 metabolites from GC–MS were annotated to lipids and lipid-like molecules, which included 32 sesquiterpenoids and 12 monoterpenoids (Figures 1C, D and Supplementary Table 3). According to the GC–MS analysis results, cadinene (15.11%) and agarospirol (21.95%) had the highest relative contents of the identified compounds in the EO from the 10a and 30a wood, respectively, and agarospirol (17.90%) and guaiol (17.90%) had the highest relative contents in the 80a wood (Table 1 and Supplementary Table 4). The main metabolites in the EOs were sesquiterpenoids, monoterpenoids, and fatty acids, and the relative contents of sesquiterpenoids in the 80a wood were greater than those in the 10a and 30a wood (Figures 2A, B). Thus, sesquiterpenoids may greatly contribute to the formation of EOs.

3.2 Metabolite identification of VOCs

A total of 40 metabolites obtained from VOCs by GC–MS were matched and classified into two superclasses, including 19 lipids and lipid-like molecules and one organoheterocyclic compound

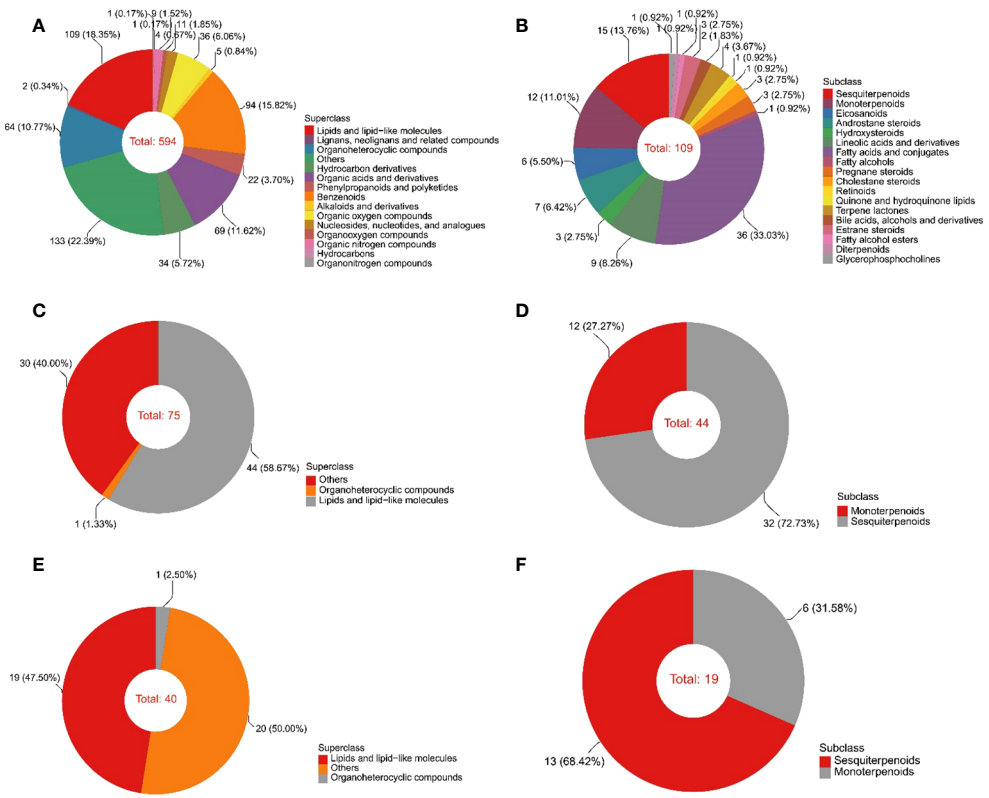


FIGURE 1
Classification of the identified metabolites by LC–MS and GC–MS. **(A)** The superclass of metabolites in EO according to LC–MS analysis. **(B)** The subclass of lipids and lipid-like molecules in EO determined by LC–MS analysis. **(C)** The superclass of metabolites in EO according to GC–MS. **(D)** The subclass of lipids and lipid-like molecules in EO determined by GC–MS. **(E)** The superclass of metabolites in VOCs determined by GC–MS. **(F)** The subclass of lipids and lipid-like molecules in VOCs determined by GC–MS.

10.3389/fpls.2024.1368894

TABLE 1 The main compounds in essential oils and wood volatile organic compounds in trees of each age.

		Assignments	Subclass	10a (%)	30a (%)	80a (%)
EOs	GC-MS	(3R,3aR,3bR,4S,7R,7aR)-4-Isopropyl-3,7-dimethyloctahydro-1H-cyclopenta[1,3]cyclopropa[1,2]benzen-3-ol	Sesquiterpenoids	14.07±27.75	0.77±0.14	1.68±0.21
		Cadinene	Sesquiterpenoids	15.11±11.91	1.24±2.38	1.40±1.71
		(1R,7S,E)-7-Isopropyl-4,10-dimethylenecyclodec-5-enol	Sesquiterpenoids	4.78±6.66	–	–
		2-((2S,4aR)-4a,8-Dimethyl-1,2,3,4,4a,5,6,7-octahydronaphthalen-2-yl)propan-2-ol	Sesquiterpenoids	10.93±6.48	1.02±2.28	5.73±1.47
		Guaiol	Sesquiterpenoids	4.59±4.45	4.51±0.38	17.90±2.28
		Eudesmol	Sesquiterpenoids	4.94±4.02	13.80±0.86	15.86±1.90
		Cadalene	Sesquiterpenoids	10.1±6.00	–	–
		Mustakone	Sesquiterpenoids	4.58±2.70	–	–
		Agarospirol	Sesquiterpenoids	5.57±4.00	21.95±1.74	17.90±2.28
		Viridflorol	Sesquiterpenoids	0.52±0.55	3.49±1.96	–
		Guaiac acetate	Sesquiterpenoids	4.94±4.02	13.8±0.86	12.46±7.18
		Uncineol	Sesquiterpenoids	0.58±1.30	–	5.73±1.47
		Caryophyllene oxide	Sesquiterpenoids	–	3.05±0.36	0.33±0.04
		(E)-3-((4S,7R,7aR)-3,7-Dimethyl-2,4,5,6,7,7a-hexahydro-1H-inden-4-yl)-2-methylacrylaldehyde	Sesquiterpenoids	–	3.04±1.71	–
	LC-MS	Carvone	Monoterpenoids	1.83±0.03	8.22±0.38	0.01±0.00
		5-Aminopentanoic acid	Amino acids, peptides and analogs	29.77±0.39	18.9±0.68	12.8±0.04
		Cubebene	Sesquiterpenoids	8.46±0.11	8.08±0.4	21.56±0.09
		Caryophyllene oxide	Sesquiterpenoids	2.55±0.03	4.27±0.21	0.74±0.01
		Diethylpropion	Carbonyl compounds	1.51±0.02	4.91±0.24	0.06±0.02
		Gentisic acid	Benzoic acids and derivatives	3.75±0.04	3.12±0.11	5.37±0.06
		Guanidoacetic acid	Amino acids, peptides and analogs	9.94±0.11	8.31±0.29	14.55±0.03
		Nootkatol	Alcohols and polyols	3.78±0.16	5.31±0.28	4.3±0.02
VOCs	GC-MS	Elemol	Sesquiterpenoids	0.27±0.15	2.12±0.22	1.37±0.19
		Guaiol	Sesquiterpenoids	4.59±4.45	4.51±0.38	17.9±2.28
		Uncineol	Sesquiterpenoids	0.58±1.30	–	5.73±1.47
		Agarospirol	Sesquiterpenoids	5.57±4.00	21.95±1.74	17.9±2.28
		Eudesmol	Sesquiterpenoids	4.94±4.02	13.8±0.86	15.86±1.9
		Longiverbenone	Sesquiterpenoids	6.57±0.57	2.27±1.28	0.24±0.53
		Caryophyllene oxide	Sesquiterpenoids	–	3.05±0.36	0.33±0.04

Only compounds accounting for more than 3.00% of the total area are listed. Other compounds (compounds accounting for more than 0.2%) are listed in Supplementary Tables 2, 4, 6. The data in the table are presented as the mean ± standard deviation, indicating that the metabolite was not identified in the wood sample.

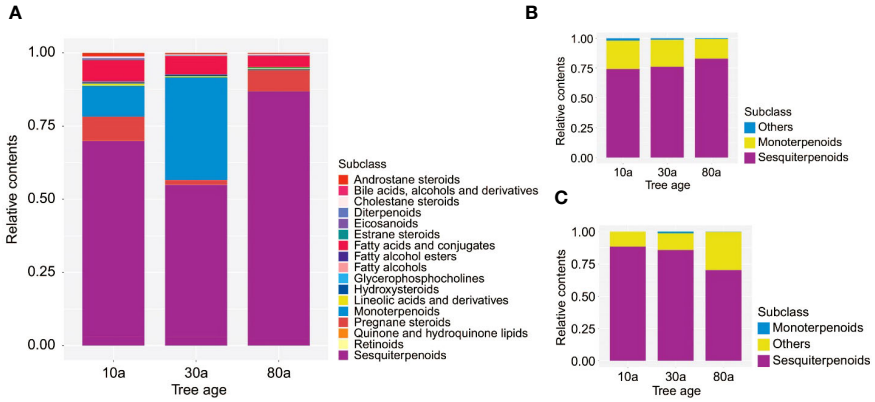


FIGURE 2
The relative contents of identified metabolites in EOs and VOCs. **(A)** The relative contents of identified metabolites in EOs determined by LC–MS. **(B)** The relative contents of identified metabolites in EOs determined by GC–MS. **(C)** The relative contents of identified metabolites in VOCs determined by GC–MS.

(Figure 1E and Supplementary Table 5). Of these, the 19 lipids and lipid-like molecules included 13 sesquiterpenoids and six monoterpenoids (Figure 1F). Similar to the results for EOs, only a few metabolites in VOCs with low relative contents showed small differences in composition and relative content between trees of the same age (Supplementary Table 6). According to the GC–MS analysis results, the major metabolite identified in the VOCs from 10a and 30a wood was agarospirol (Table 1 and Supplementary Table 6). For 80a wood, the major identified VOCs were agarospirol and eudesmol. The main compounds of *P. zhennan* wood VOCs were sesquiterpenoids (70.22%-88.45%), and the relative contents of sesquiterpenoids in 10a and 30a wood were greater than those in 80a wood (Figure 2C). Hence, sesquiterpenoids are important metabolites that contribute to the fragrance of *P. zhennan* wood.

3.3 PCA, OPLS-DA, and clustering analysis

To obtain preliminary knowledge of the metabolite variation in EOs and VOCs among 10a, 30a, and 80a wood, the metabolic data matrix for wood from trees of three ages was analyzed using PCA. The principal components of the PCA plot explained 88.6%, 55.7%, and 72.8% of the total variance in the EOs according to LC–MS and GC–MS and VOCs according to GC–MS, respectively (Figures 3A, C, E). According to the PCA plot, the three groups were well separated, indicating significant differences in metabolism among wood samples from trees of different ages. The clustering heatmap also revealed obvious variation in metabolite accumulation patterns among the 10a, 30a, and 80a wood (Figures 3B, D, F). In the EOs analyzed by LC–MS, the accumulation of 94 metabolites (C3) increased in 80a wood (carbohydrates and carbohydrate conjugates, amino acids, peptides and analogs, fatty acids and conjugates, etc.), 138 metabolites (C5) increased in 10a wood (amino acids, peptides and analogs, alcohols and polyols, fatty acids and conjugates, etc.), 156 metabolites (C4) increased in 30a wood (fatty acids and conjugates, amino acids, peptides and analogs, and monoterpenoids, etc.), 157 metabolites (C2) increased in both 10a and 30a wood (amino acids, peptides and analogs, carbohydrates and carbohydrate conjugates, alcohols and polyols, etc.), and 51 metabolites (C1) increased in both 30a and 80a wood (amino acids, peptides and analogs, fatty acids and conjugates, and derivatives, etc.) (Figure 3B). According to GC–MS analysis of the EOs, 20 metabolites (C2) were up-accumulated in 30a wood (five sesquiterpenoids and two monoterpenoids), 28 metabolites (C3) were up-accumulated in 10a wood (nine sesquiterpenoids and four monoterpenoids), and 28 metabolites (C1) were up-accumulated in 80a wood (10 sesquiterpenoids and six monoterpenoids) (Figure 3D). The 41 VOCs were separated into three clusters; of these, 18 metabolites (C1) were up-accumulated in 80a wood (five sesquiterpenoids and four monoterpenoids), 17 metabolites (C2) were up-accumulated in 30a wood (five sesquiterpenoids and two monoterpenoids), and six metabolites (C3) were up-accumulated in 10a wood (one sesquiterpenoid) (Figure 3F).

To obtain better discrimination among groups of 10a, 30a, and 80a wood, OPLSA-DA was employed to maximize the differences between the wood samples. The 10a wood and the 30a and 80a wood from trees of different ages were obviously separated into two blocks on the basis of their metabolic profiles in the OPLS-DA model, indicating that tree age had an impact on the metabolic profiles of EOs and VOCs (Supplementary Figure 2). The simulated values of R2 and Q2 on the left side were smaller than the original values in the upper right corner, suggesting that the original models were effective and reliable. These results indicated that the OPLS-DA models have good predictive ability and could be applied to further metabolite variance analysis. The S-plot displayed the crucial differences among the comparison groups of three tree ages (Supplementary Figure 3). For the EOs, a total of 53 (9), 51 (9), and 50 (6) potential differentially abundant metabolites (DAMs) identified via LC–MS (GC–MS) were screened in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively (Supplementary Figures 3A–F; Supplementary Table 7). For fragrance-related metabolites, a total of 25, 22, and 21 potentially DAMs were screened in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively (Supplementary Figures 3G–I; Supplementary Table 8). There were high numbers of sesquiterpenoids, amino acids, and peptides and analogs in the EOs,

10.3389/fpls.2024.1368894

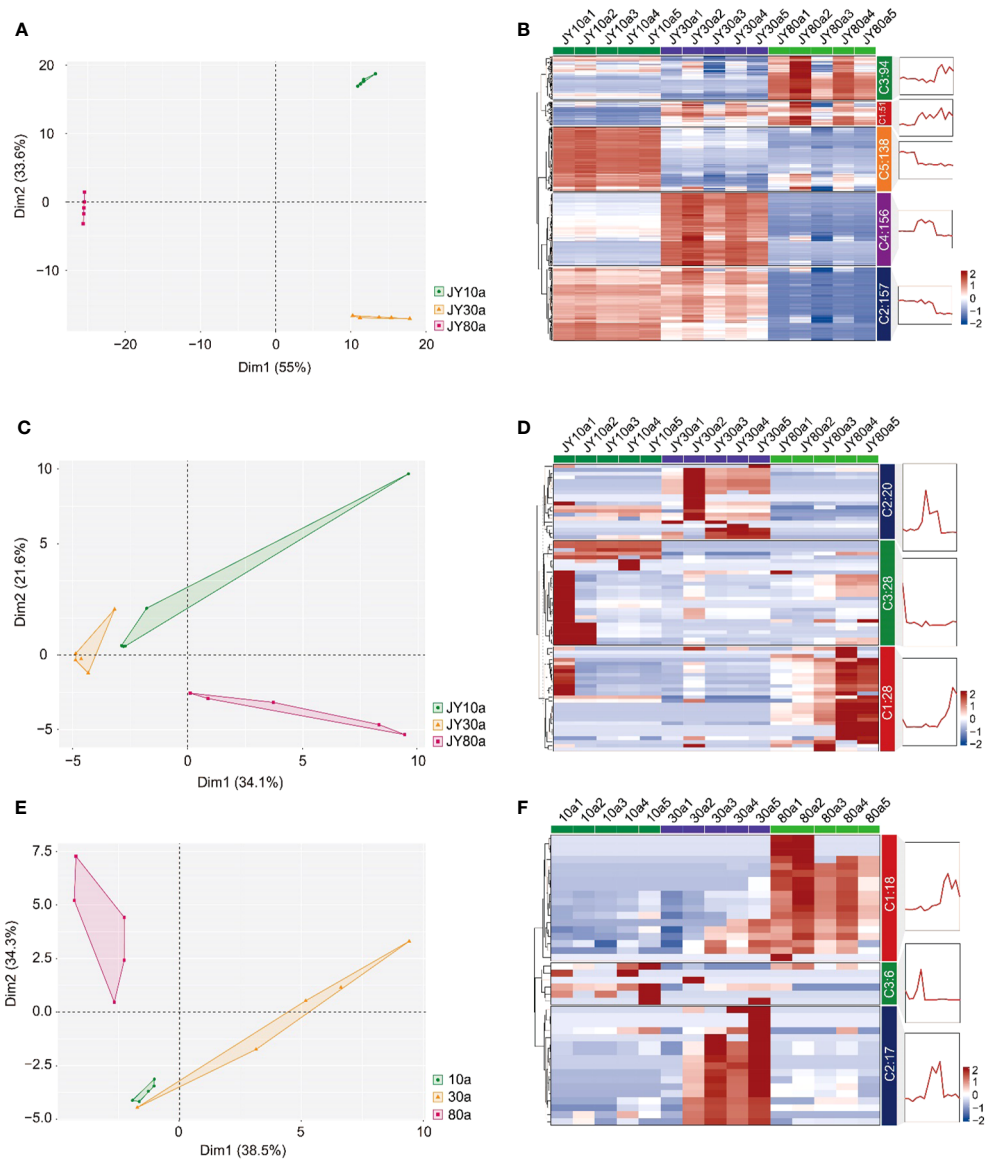


FIGURE 3
Principal component analysis (PCA) and hierarchical cluster analysis (HCA) of the metabolite distribution among different tree ages. **(A, B)** PCA and HCA map of the metabolites of EO determined by LC–MS distribution among trees of three ages. **(C, D)** PCA and HCA maps of the metabolites of EO determined by GC–MS distribution among trees of three ages. **(E, F)** PCA and HCA map of the metabolites of VOCs determined by GC–MS distribution among trees of the three ages. JY10a1-JY10a5, JY30a1-JY30a5, and JY80a1-JY80a5 represent the EOs from five trees (five biological replicates) at the ages of 10a, 30a, and 80a, respectively. 10a1-10a5, 30a1-30a5, and 80a1-80a5 represent VOCs from five trees (five biological replicates) at the tree ages of 10a, 30a, and 80a, respectively. C1, C2, C3, and C4 in **(B, D, F)** represent cluster 1, cluster 2, cluster 3, and cluster 4, respectively, according to hierarchical cluster analysis.

and the VOCs significantly differed among the 10a, 30a, and 80a wood, especially in the 10a vs. 30a and 10a vs. 80a comparison groups (Figure 4).

3.4 Identification and analysis of metabolites in EOs

To further screen the DAMs of EOs and VOCs among wood from trees of three different ages, a volcano plot was used to visualize the significantly DAMs (Figure 5A). LC–MS revealed a total of 209 (108 up- and 101 down-accumulated), 338 (51 up- and 287 down-accumulated), and 330 (53 up- and 277 were down-accumulated) significantly DAMs in the EOs for the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively. For GC–MS analysis, there were 45 (17 up- and 28 down-accumulated), 49 (25 up- and 24 down-accumulated), and 57 (34 up- and 23 down-accumulated) significantly DAMs of EOs in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively. The final significantly DAMs were defined as those with a relative log2FC≥1.0

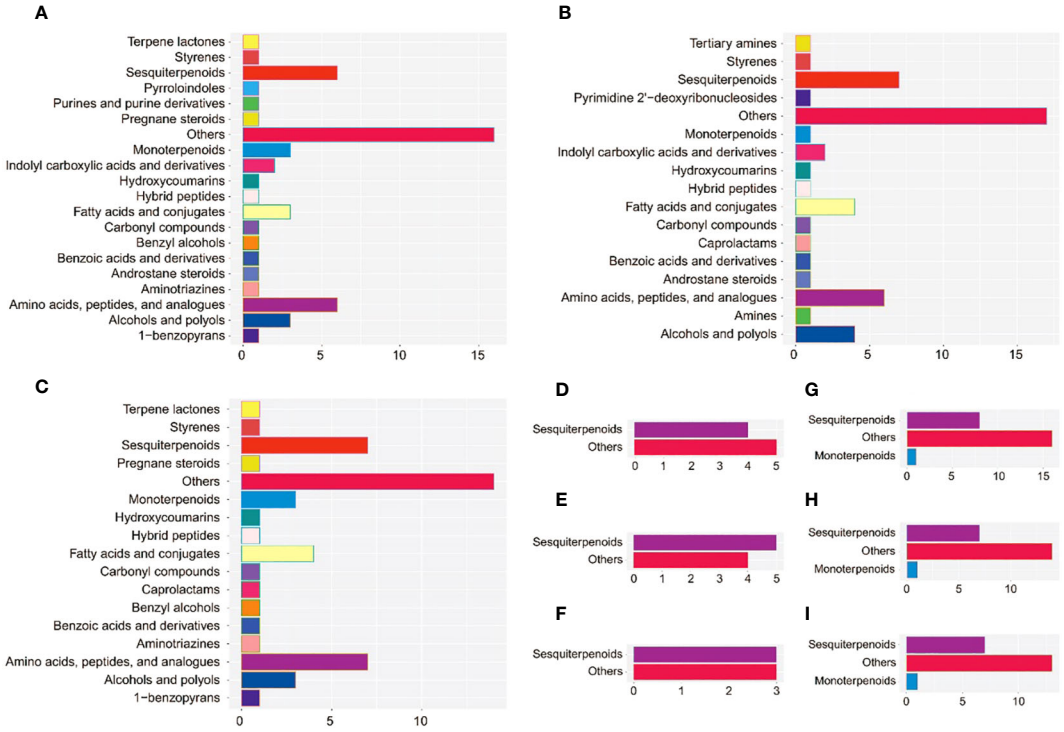


FIGURE 4
The subclass and number of potentially DAMs with VIP>1 between the comparison groups. **(A–C)** Potentially DAMs of EO identified by LC–MS analysis in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups. **(D–F)** Potentially DAMs of EO identified by GC–MS analysis in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups. **(G–I)** Potentially DAMs of VOCs in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups.

and ≤1.0, P<0.05, and VIP ≥1. There were 34 (22 up- and 12 down-accumulated), 45 (8 up- and 37 down-accumulated), and 43 (5 up- and 38 down-accumulated) significant DAMs of EO according to LC–MS analysis for the comparison groups of 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a, respectively (Figure 5B and Supplementary Table 9). The Venn diagram showed that nine DAMs were common among the three comparison groups, and three (two up- and one down-accumulated), eight (three up- and five down-accumulated), and two (down-accumulated) DAMs were unique in the comparison groups of 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a, respectively (Figure 5B). A total of nine (down-accumulated), eight (three up- and five down-accumulated), and six (up-accumulated) DAMs of EOs were identified by GC–MS analysis in the comparison groups of 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a, respectively (Figure 5C and Supplementary Table 9). The Venn diagram showed that one DAM was common among the three comparison groups, and two DAMs (one up- and one down-accumulated) were unique to the 10a vs. 30a and 30a vs. 80a comparison groups, respectively (Figure 5C). As the main compounds of EOs, three monoterpenoids (cuminaldehyde, carvone, and perillic acid) and three sesquiterpenoids (capsidiol, caryophyllene α-oxide, and abscisic acid) accumulated significantly more in 30a wood than in 10a and 80a wood (Figure 6A and Supplementary Table 10). Compared with those from 10a wood, the EOs from 80a wood had three sesquiterpenoids (caryophyllene, guaiol, and eudesmol) that were significantly accumulated.

3.5 Identification and analysis of metabolites in VOCs

For wood VOCs, a total of 25 (22 up- and three down-accumulated), 29 (23 up- and six down-accumulated), and 34 (15 up- and 19 down-accumulated) significantly DAMs were detected in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively (Figure 5A). For VOCs, there were seven (six up- and one down-accumulated), three (up-accumulated), and three (up-accumulated) DAMs in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups, respectively (Figure 5D and Supplementary Table 9). A Venn diagram showed no common DAMs detected among all three comparison groups, and seven (six up- and one down-accumulated) DAMs were unique to the 10a vs. 30a comparison group (Figure 5D). Among the wood VOCs, one monoterpenoid (verbenol) was detected only in the 30a and 80a wood (Figure 6B and Supplementary Table 11). Caryophyllene oxide was found only in the 30a wood. The relative content of -eudesmol was significantly greater in the 80a wood than in the 10a and 30a wood.

3.6 Metabolic pathway enrichment analysis

To understand the metabolic pathway variation of EOs and VOCs in wood of different ages, metabolic pathway enrichment analysis was

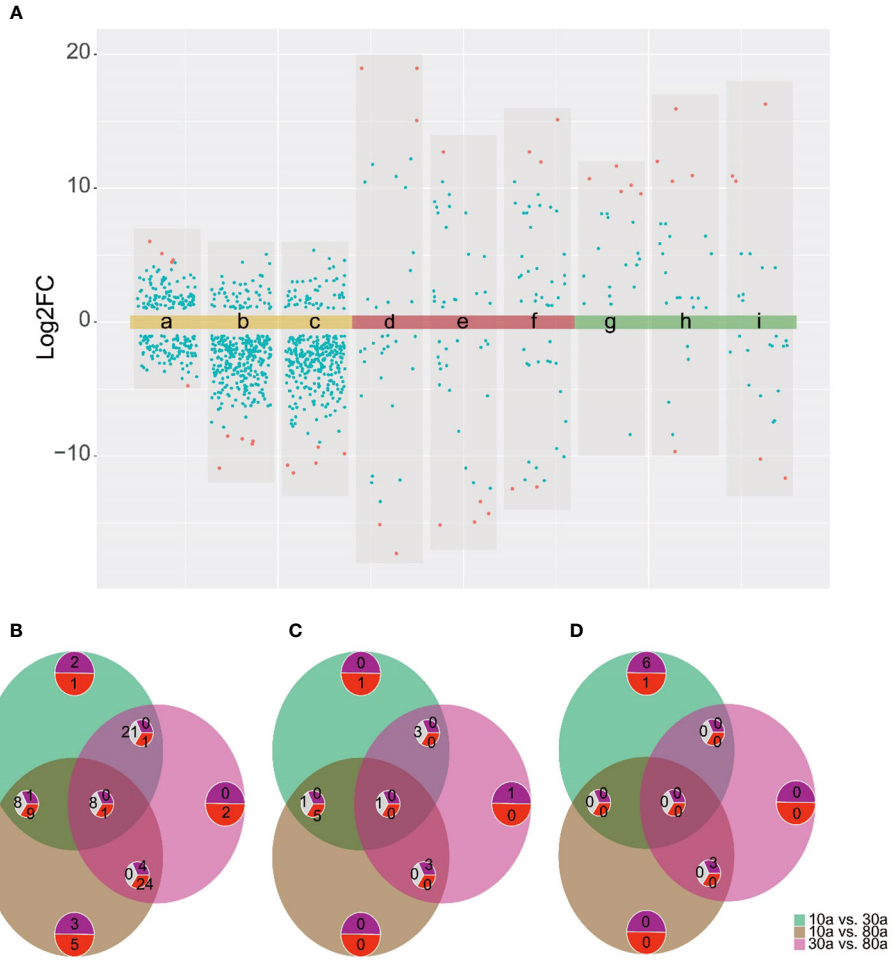


FIGURE 5
DAM analysis for the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups. **(A)** Volcano plots of the comparison group of 10a vs. 30a (a, d, and g show the EO determined by LC–MS and GC–MS and VOCs, respectively), 10a vs. 80a (b, e, and h show the EO determined by LC–MS and GC–MS and VOCs, respectively), and 30a vs. 80a (c, f, and i show the EO determined by LC–MS and GC–MS and VOCs, respectively). Red dots represent the top five metabolites according to the log2FC values. **(B)** Venn diagram of the significantly DAMs of EO identified by LC–MS analysis in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups. The purple semicircle with numbers represents the number of increased metabolites, the red semicircle with numbers represents the number of decreased metabolites, and the gray semicircle with numbers represents the intersection of metabolites with inconsistent increasing and decreasing trends in different groups. The same below. **(C)** Venn diagram of the DAMs of EO according to GC–MS analysis in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups. **(D)** Venn diagram of the DAMs of VOCs in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups.

applied according to the KEGG database. The DAMs of EO in the 10a vs. 30a, 10a vs. 80a, and 30a vs. 80a comparison groups were enriched mainly in 52, 58, and 60 pathways, respectively. The significant pathways were mainly involved in tyrosine metabolism, aminoacyl-tRNA biosynthesis, phenylpropanoid biosynthesis, etc.,,, indicating that tree age also affects the composition and contents of other bioactive metabolites, such as flavonoids, in EOs (Figure 7). The sesquiterpenoid and triterpenoid biosynthesis pathways changed significantly at $P<0.05$ in the comparisons of 10a vs. 30a and 30a vs. 80a, implying that the sesquiterpenoid and triterpenoid metabolites screened in this study affect the EOs of wood from trees of different ages. The biosynthesis of unclassified secondary metabolites and the biosynthesis of unsaturated fatty acids pathways also changed significantly at $P<0.05$ in the comparisons of 10a vs. 30a and 30a vs. 80a, indicating that the secondary metabolites and unsaturated fatty acids screened in this study may also affect the EO of wood among trees of three different ages. Unfortunately, no DAMs of VOCs detected by GC–MS analysis were enriched in KEGG pathways.

4 Discussion

Trees grow and develop at the physiological, morphological, and molecular levels as they age. Similar to previous findings on the EOs of *O. lanceolata* (Fajar et al., 2019) and *C. burmannii* (Gathara et al., 2022), tree age had a significant effect on the content of EOs in *P. zhennan* wood. As the tree ages, the width and volume of its timber will also increase, which will increase the accumulation of EOs within the wood. The main constituents of EOs from the leaves of *L. cubeba* and *C. kanehirae* are monoterpenoids (Cheng et al., 2015; Chen et al., 2020). In comparison, the results of the EO analysis of *P. bournei* were similar to those of our study (Han et al., 2022), in that sesquiterpenoids

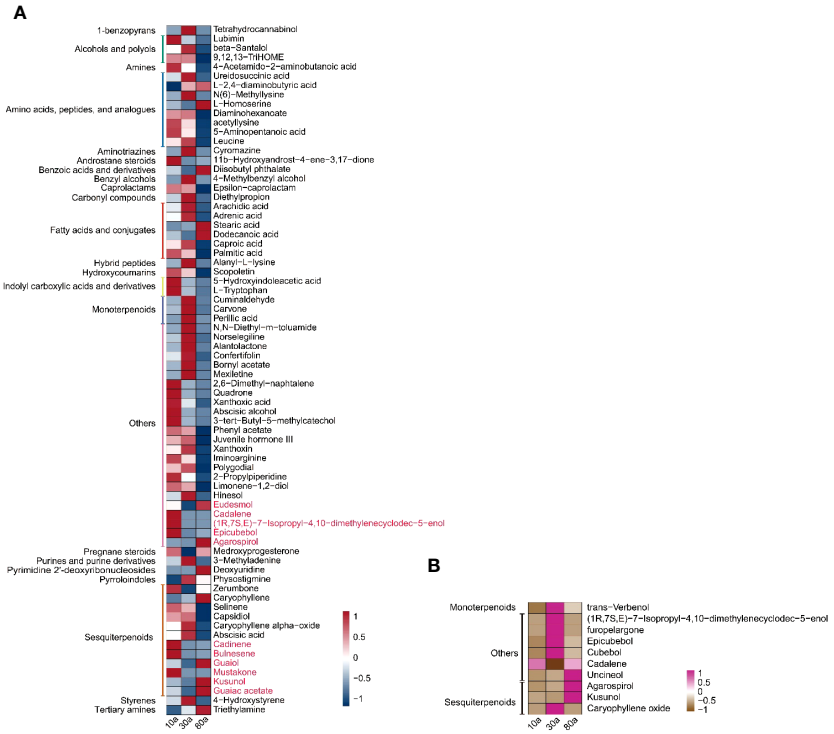


FIGURE 6
Trends of the final significantly DAMs in wood from differently aged trees. **(A)** Trends of the final significantly DAMs of EO in wood from trees of different ages. The red words represent the metabolites identified by GC–MS. The right of the heatmap shows the names of the metabolites, and the left of the heatmap shows the subclasses of those metabolites. **(B)** Trends of the final significantly DAMs of VOCs in wood of different ages. The right of the heatmap shows the names of the metabolites, and the left of the heatmap shows the subclasses of those metabolites.

such as cubebene, guaiol, eudesmol, and cadinene were dominant in the EOs from *P. zhennan* wood. The major compounds in the bark oil of *P. zhennan* included calarene and cadinene, while the major compounds in the leaf oil were cadiene, copaene, and eudesmol (Shao et al., 2020). The major compounds in the EOs from the old wood of *P. zhennan* were 1-(3-methylbutyl)-2,3,4,5-tetramet hylbenzene and caroyllene oxide, while the major compounds in the newer wood were γ-cadinene and 1H-cycloprop[e] azulene, decahydro-1,1,7-trimethyl-4-methylene-, [1aR- (1a, α.,4a. β.,7. α.,7b.α.)] (Xie et al., 2015). In contrast with the findings of previous studies, 5-aminopentanoic acid, cubebene, and guanidine acetic acid, which were identified via LC–MS analysis, and agarospirol, guaiol, and naphthalene,1,2,3,5,6,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, (1s-cis)- which were identified via GC–MS, were the major compounds in the EOs from *P. zhennan* wood. A high content of the sesquiterpenoid cubebene has also been reported in *Melissa officinalis* EOs (Rădulescu et al., 2021). This phytochemical polymorphism is significantly determined by genetic factors and environmental variations (Aghaei et al., 2013; Kittler et al., 2018). The composition of EOs in plants can also be affected by age (Geng et al., 2011; Shiferaw et al., 2019). Our findings suggested that the main differences among the EOs from wood of differently aged trees in *P. zhennan* were their constituent compounds and their concentrations. Our results showed that the contents and types of sesquiterpenoids and monoterpenoids in EOs were significantly influenced by tree age. The analysis of wood EO composition showed that most identified compounds had high fluctuations in percentage composition across tree ages. These significant differences in components support the previous argument that the function of EOs is closely related to plant defense against pests or weeds (Mishra and Srivastava, 2020). The studied EOs from wood of differently aged trees displayed different chemical profiles. For instance, the proportions of two sesquiterpenoids (capsidiol and caryophyllene) exhibited contrasting trends among the three tree ages. Additionally, some components were not present in any of the samples. These phenomena were also investigated in *Cinnamomum cassia* (Kittler et al., 2018). The variations can be partly explained by the fact that plants with external secretory structures can release secretions with organ maturation due to trichome cuticle disruption, whereas plants with internal secretory structures more often maintain a more stable yield and composition (Figueiredo et al., 2008; Geng et al., 2011).

A pleasant fragrance endows wood with excellent durability and high economic value (Wang et al., 2006). Accumulated secondary metabolites in wood are the main substances contributing to wood fragrance. Terpenoids, especially sesquiterpenoids and monoterpenoids, are the main compounds related to the fragrance of wood in many woody plants, such as *Chamaecyparis formosensis*, *Taiwania cryptomerioides*, *Cryptomeria* species, and *P. hui* (Wang et al., 2006; Xie et al., 2011; Yang et al., 2022b). Similar to our previous findings (Yang et al., 2023), sesquiterpenoids and monoterpenoids were the main metabolites contributing to the unique fragrance of *P. zhennan* wood. However, the types of sesquiterpenoids identified in this study are different from those identified in that previous study (Yang et al., 2023).

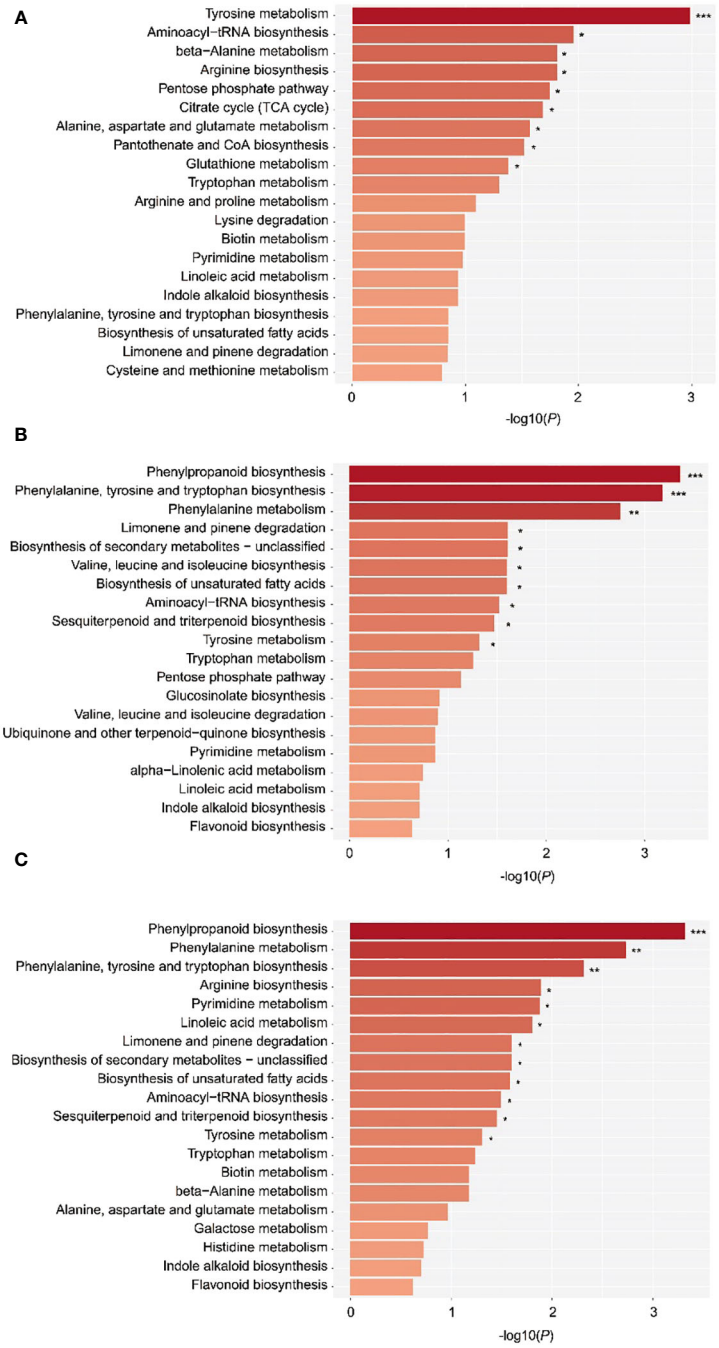


FIGURE 7
The top 20 enriched KEGG pathways of the DAMs. **(A–C)** KEGG pathways associated with the DAMs detected in the 10a vs. 80a, 10a vs. 30a, and 30a vs. 80a comparison groups. The −log10(*P*) value of the hypergeometric test shows the reliability and statistical significance levels of the test, and the larger the value is, the greater the degree of enrichment of metabolites in this pathway. ***, **, and * represent P values <0.001, 0.01, and 0.05, respectively.

For instance, the major sesquiterpenoids epi-γ-eudesmol and agarospirol could not be found in the heartwood of *P. zhennan*. This may be due to differences in the extraction methods used for VOCs, cultivation environments, and tree ages of wood samples. Agarospirol has been identified in agarwood oil in many reports (Nakanishi et al., 1984; Pornpunyapat et al., 2011; Gogoi et al., 2023). In the present investigation, agarospirol was the major compound of VOCs and EOs from *P. zhennan* wood, similar to the findings in the above report. This result also indicated a close correlation between EOs and VOCs. Previous studies reported that with increasing tree age, the contents of the main components of VOCs from leaves increased gradually (Cui et al., 2023). In this study, the relative content and composition of VOCs from wood were significantly affected by tree age. With increasing tree age, the relative content of major VOC components, such as agarospirol, increased gradually, and more new metabolites, such as two sesquiterpenoids (himachalene and 3-cyclohexen-1-ol, 1-(1,5-

dimethyl-4-hexenyl)-4-methyl-), were formed in wood from older trees. This indicated that an increasing number of fragrance-related compounds accumulated in trunk wood with increasing tree age in *P. zhennan*. The above results are related to the formation of heartwood; more fragrance-related compounds were formed and accumulated during heartwood formation, and the proportion of heartwood increased with increasing tree age. Therefore, in the development and utilization of wood in *P. zhennan*, an appropriate processing technology and product type should be selected based on the variation trend of VOCs, and differentiated development and utilization should be carried out.

5 Conclusion

The yield and composition of EOs and VOCs in *Phoebe zhennan* wood greatly differed among trees of different ages. The contents of EOs in the wood varied depending on tree age, and the 30a wood had the highest EO yield. A total of 596 (LC–MS) and 76 (GC–MS) metabolites were identified in the EOs that showed high fluctuations in percentage composition between the different tree ages. Sesquiterpenoids were the main compounds in the EOs, and the relative contents of caryophyllene oxide, eudesmol, carvone, and cubebene in the 30a and 80a wood were greater than those in the 10a wood. The main compounds of wood VOCs were also sesquiterpenoids, of which the major compounds, such as agarospirol and eudesmol, were more highly accumulated in 30a and 80a wood than in 10a wood. In general, the wood of an intermediate age (30a) is suitable for EO production, but older wood of *P. zhennan* is more suitable for making high-grade wooden furniture because of its remarkable and pleasant fragrance.

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

HY: Conceptualization, Data curation, Funding acquisition, Software, Writing – original draft, Writing – review & editing. SZ: Methodology, Writing – original draft, Writing – review & editing. YG: Conceptualization, Funding acquisition, Writing – review & editing. JP: Conceptualization, Data curation, Methodology, Writing – review & editing. XH: Conceptualization, Methodology, Software, Writing – review & editing. HG: Data curation, Writing – review & editing. LC: Methodology, Writing – review & editing. YJ: Data curation, Methodology, Writing – review & editing, Software. ML: Data curation, Writing – review & editing, Methodology. XL: Data curation, Software, Writing – review & editing, Conceptualization. JX: Data curation, Methodology, Software, Writing – review & editing. XW: Data curation, Methodology, Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Science and Technology Innovation Team Project of Sichuan Forestry and Grassland [2024LCTD0213, 2023LCTD0203]; the Key Research and Development Project of Sichuan Province [2021YFYZ0032]; the National Science Foundation of Sichuan Province [2022NSFSC1062]; the Forest Ecosystem Improvement in the Upper Reaches of the Yangtze River Basin Program [510201202038467].

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1368894/full#supplementary-material

References

Abdelrazig, S., Safo, L., Rance, G. A., Fay, M. W., Theodosiou, E., Topham, P. D., et al. (2020). Metabolic characterisation of Magnetospirillum gryphiswaldense MSR-1 using LC-MS-based metabolite profiling. *RSC Adv.* 10, 32548–32560. doi: 10.1039/D0RA05326K

Aghaei, Y., Hossein Mirjalili, M., and Nazeri, V. (2013). Chemical Diversity among the Essential Oils of Wild Populations of Stachys lavandulifolia Vahl (Lamiaceae) from Iran. *Chem. Biodivers.* 10, 262–273. doi: 10.1002/cbdv.201200194

Caputi, L., and Aprea, E. (2011). Use of terpenoids as natural flavouring compounds in food industry. *Recent Patents Food Nutr. Agric.* 3, 9–16. doi: 10.2174/2212798411103010009

Chang, S.-T., Wang, S.-Y., Wu, C.-L., and Chang, H.-T. (1998). Isolation and identification of fragrant constituents from three coniferous woods grown in Taiwan. *For. Prod. Industries* 17, 737–750.

Chen, Y.-C., Li, Z., Zhao, Y.-X., Gao, M., Wang, J.-Y., Liu, K.-W., et al. (2020). The Litsea genome and the evolution of the laurel family. *Nat. Commun.* 11, 1675. doi: 10.1038/s41467-020-15493-5

Cheng, S.-S., Lin, C.-Y., Yang, C.-K., Chen, Y.-J., Chung, M.-J., and Chang, S.-T. (2015). Chemical polymorphism and composition of leaf essential oils of *Cinnamomum kanehirae* using gas chromatography/mass spectrometry, cluster analysis, and principal component analysis. *J. Wood Chem. Technol.* 35, 207–219. doi: 10.1080/02773813.2014.924967

Cheng, S.-S., Lin, H.-Y., and Chang, S.-T. (2005). Chemical composition and antifungal activity of essential oils from different tissues of Japanese cedar (*Cryptomeria japonica*). *J. Agric. Food Chem.* 53, 614–619. doi: 10.1021/jf0484529

Cheng, Y., Peng, J., Gu, Y., Guo, H., Jiang, T., and Yang, H. (2023). The early effect of plant density on soil physicochemical attributes and bacterial and understory plant diversity in *Phoebe zhennan* plantations. *Forests* 14, 1612. doi: 10.3390/f14081612

Cui, B., Zheng, T., Deng, P., Zhang, S., and Zhao, Z. (2023). Chemotaxonomic variation in volatile component contents in ancient *Platycladus orientalis* Leaves with different tree ages in huangdi mausoleum. *Molecules* 28, 2043. doi: 10.3390/molecules28052043

Das, R., Naziz, P. S., and Sen, S. (2022). The fragrance of agarwood. *Sci. Rep.*, 46–47.

Ding, W., Liping, N., Xing, H., Wei, Z., Zhoua, Q., Nong, R., et al. (2020). Essential oil extracted from leaf of *Phoebe bournei* (Hemsl.) yang: chemical constituents, antitumor, antibacterial, hypoglycemic activities. *Nat. Prod. Res.* 34, 2524–2527. doi: 10.1080/14786419.2018.1542393

Fajar, A., Ammar, G. A., Hamzah, M., Manurung, R., and Abduh, M. Y. (2019). Effect of tree age on the yield, productivity, and chemical composition of essential oil from Cinnamomum burmannii. *Curr. Res. Biosci.* 1, 17–22. doi: 10.5614/crbb

Figueiredo, A. C., Barroso, J. G., Pedro, L. G., and Scheffer, J. J. C. (2008). Factors affecting secondary metabolite production in plants: volatile components and essential oils. *Flavour Fragrance J.* 23, 213–226. doi: 10.1002/ffj.1875

Gathara, M., Kamondo, B., Mwadalu, R., Omondi, S., Kivuti, M., Wairagu, N., et al. (2022). Influence of soil nutrients, tree age, and sandalwood provenances on sandalwood oil yield and quality. *Int. J. Forest. Res.* 2022, 1577979. doi: 10.1155/2022/1577979

Geng, S., Cui, Z., Huang, X., Chen, Y., Xu, D., and Xiong, P. (2011). Variations in essential oil yield and composition during Cinnamomum cassia bark growth. *Ind. Crops Prod.* 33, 248–252. doi: 10.1016/j.indcrop.2010.10.018

Gogoi, R., Sarma, N., Begum, T., Chanda, S. K., Lekhak, H., Sastry, G. N., et al. (2023). Agarwood (Aquilaria malaccensis L.) a quality fragrant and medicinally significant plant based essential oil with pharmacological potentials and genotoxicity. *Ind. Crops Prod.* 197, 116535. doi: 10.1016/j.indcrop.2023.116535

Han, X., Zhang, J., Han, S., Chong, S. L., Meng, G., Song, M., et al. (2022). The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Commun.* 3, 100410. doi: 10.1016/j.xplc.2022.100410

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectr.* 45, 703–714. doi: 10.1002/jms.1777

Joshi, S. C., Padalia, R. C., Bisht, D. S., and Mathela, C. S. (2009). Terpenoid diversity in the leaf essential oils of himalayan Lauraceae species. *Chem. Biodivers.* 6, 1364–1373. doi: 10.1002/cbdv.200800181

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

Kittler, J., Krüger, H., Lohwasser, U., Ulrich, D., Zeiger, B., Schütze, W., et al. (2018). Evaluation of 28 balm and lemon balm (Melissa officinalis) accessions for content and composition of essential oil and content of rosmarinic acid. *Genet. Resour. Crop Evol.* 65, 745–757. doi: 10.1007/s10722-017-0568-3

Kodis, M., Moran, D. T., and Houy, D. (1998). *Love Scents* (New York: EP Dutton).

Koukos, P. K., Papadopoulou, K. I., Papagiannopoulos, A. D., and Patiaka, D. T. (2001). Essential oils of the twigs of some conifers grown in Greece. *Holz als Roh- und Werkstoff* 58, 437–438. doi: 10.1007/s001070050457

Kumar, A., Jnanesha, A. C., Kumar, V. M., and Lal, R. K. (2022). GGE biplot vs. AMMI analysis of genotype-by-environment data on essential oil yield in lemongrass [Cymbopogon flexuosus (nees ex. Steud) wats.] grown in semi-arid tropical regions of southern India under different agro-climatic conditions. *Biochem. Syst. Ecol.* 103, 104439. doi: 10.1016/j.bse.2022.104439

Kumar, A., and Lal, R. K. (2022). The consequence of genotype × environment interaction on high essential oil yield and its composition in clove basil (Ocimum gratissimum L.). *Acta Ecol. Sin.* 42, 633–640. doi: 10.1016/j.chnaes.2021.06.002

Mishra, N., and Srivastava, R. (2020). "Therapeutic and pharmaceutical potential of cinnamon," in *Ethnopharmacological Investigation of Indian Spices*. Ed. N. Mishra (IGI Global, Hershey, USA), 124–136.

Nakanishi, T., Yamagata, E., Yoneda, K., Nagashima, T., Kawasaki, I., Yoshida, T., et al. (1984). Three fragrant sesquiterpenes of agarwood. *Phytochemistry* 23, 2066–2067. doi: 10.1016/S0031-9422(00)84975-4

Pornpunyapat, J., Chetpattananondh, P., and Tongurai, C. (2011). Mathematical modeling for extraction of essential oil from Aquilaria crassna by hydrodistillation and quality of agarwood oil. *angladesh J. Pharmacol.* 6, 18–24. doi: 10.3329/bjp.v6i1.7902

Poudel, D. K., Rokaya, A., Ojha, P. K., Timsina, S., Satyal, R., Dosoky, N. S., et al. (2021). The chemical profiling of essential oils from different tissues of *Cinnamomum camphora* L. and their antimicrobial activities. *Molecules* 26, 5132. doi: 10.3390/molecules26175132

Rădulescu, M., Jianu, C., Lukinich-Gruia, A. T., Mioc, M., Mioc, A., Şoica, C., et al. (2021). Chemical composition, in *vitro* and in silico antioxidant potential of *Melissa officinalis* subsp. officinalis essential oil. *Antioxidants* 10, 1081. doi: 10.3390/antiox10071081

Shao, H., Jiang, Y., Pan, F., Xie, J., Qi, J., Xiao, H., et al. (2020). Chemical composition, UV/vis absorptivity, and antioxidant activity of essential oils from bark and leaf of phoebe zhennan S. K. Lee & F. N. Wei. *Nat. Prod. Res.* 34, 876–879. doi: 10.1080/14786419.2018.1504047

Shen, L., Ding, C., Zhang, W., Zhang, T., Li, Z., Zhang, J., et al. (2023). The Populus koreana genome provides insights into the biosynthesis of plant aroma. *Ind. Crops Prod.* 197, 116453. doi: 10.1016/j.indcrop.2023.116453

Shiferaw, Y., Kassahun, A., Tedla, A., Feleke, G., and Abebe, A. A. (2019). Investigation of essential oil composition variation with age of Eucalyptus globulus growing in Ethiopia. *Nat. Prod. Chem. Res.* 7, 360. doi: 10.35248/2329-6836

Srivastava, S., Lal, R. K., Maurya, R., Mishra, A., Yadav, A. K., Pandey, G., et al. (2021). Chemical diversity of essential oil among basil genotypes (Ocimum viride Willd.) across the years. *Ind. Crops Prod.* 173, 114153. doi: 10.1016/j.indcrop.2021.114153

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., et al. (2007). LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* 35, D527–D532. doi: 10.1093/nar/gkl838

Wang, S., Liang, H., Wang, H., Li, L., Xu, Y., Liu, Y., et al. (2022). The chromosome-scale genomes of Dipterocarpus turbinatus and Hopea hainanensis (Dipterocarpaceae) provide insights into fragrant oleoresin biosynthesis and hardwood formation. *Plant Biotechnol. J.* 20, 538–553. doi: 10.1111/pbi.13735

Wang, S.-Y., Wang, Y.-S., Tseng, Y.-H., Lin, C.-T., and Liu, C.-P. (2006). Analysis of fragrance compositions of precious coniferous woods grown in Taiwan. *Holzforschung* 60, 528–532. doi: 10.1515/HF.2006.087

Wang, Y., Ma, Q., Lin, L., Zhang, H., Luo, X., Wang, J., et al. (2023). Integrated cytological, physiological, and transcriptome analysis of the bud mutant of jujube (Ziziphus jujuba Mill.) with non-lethal albino phenotype. *Ind. Crops Prod.* 201, 116964. doi: 10.1016/j.indcrop.2023.116964

Wani, A. R., Yadav, K., Khursheed, A., and Rather, M. A. (2021). An updated and comprehensive review of the antiviral potential of essential oils and their chemical constituents with special focus on their mechanism of action against various influenza and coronaviruses. *Microb. Pathogen.* 152, 104620. doi: 10.1016/j.micpath.2020.104620

Want, E. J., Masson, P., Michopoulos, F., Wilson, I. D., Theodoridis, G., Plumb, R. S., et al. (2013). Global metabolic profiling of animal and human tissues *via* UPLC-MS. *Nat. Protoc.* 8, 17–32. doi: 10.1038/nprot.2012.135

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). HMDB: the human metabolome database. *Nucleic Acids Res.* 35, D521–D526. doi: 10.1093/nar/gkl923

Xia, J., and Wishart, D. S. (2011). Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* 6, 743–760. doi: 10.1038/nprot.2011.319

Xiao, J.-H., Ding, X., Li, L., Ma, H., Ci, X.-Q., van der Merwe, M., et al. (2020). Miocene diversification of a golden-thread nanmu tree species (*Phoebe zhennan*, Lauraceae) around the Sichuan Basin shaped by the East Asian monsoon. *Ecol. Evol.* 10, 10543–10557. doi: 10.1002/ece3.6710

Xie, J., Qi, J., Huang, X., Zhou, N., and Hu, Y. (2015). Comparative analysis of modern and ancient buried Phoebe zhennan wood: surface color, chemical components, infrared spectroscopy, and essential oil composition. *J. Forest. Res.* 26, 501–507. doi: 10.1007/s11676-015-0034-z

Xie, Y., Wang, J., Yang, F., and Lei, C. (2011). Comparative analysis of essential oil components of two *Cryptomeria* species from China. *Ind. Crops Prod.* 34, 1226–1230. doi: 10.1016/j.indcrop.2011.04.016

Yang, H., An, W., Gu, Y., Peng, J., Jiang, Y., Li, J., et al. (2022a). Integrative metabolomic and transcriptomic analysis reveals the mechanism of specific color formation in *Phoebe zhennan* heartwood. *Int. J. Mol. Sci.* 23, 13569. doi: 10.3390/ijms232113569

Yang, H., An, W., Wang, F., Gu, Y., Guo, H., Jiang, Y., et al. (2022b). Integrated Transcriptomic, Metabolomic, and Physiological Analyses Reveal New Insights into Fragrance Formation in the Heartwood of *Phoebe hui*. *Int. J. Mol. Sci.* 23, 14044. doi: 10.3390/ijms232214044

Yang, H., Wang, F., An, W., Gu, Y., Jiang, Y., Guo, H., et al. (2023). Comparative metabolomics and transcriptome analysis reveal the fragrance-related metabolite formation in Phoebe zhennan wood. *Molecules* 28, 7047. doi: 10.3390/molecules28207047

Zelena, E., Dunn, W. B., Broadhurst, D., Francis-Mcintyre, S., Carroll, K. M., Begley, P., et al. (2009). Development of a robust and repeatable UPLC–MS method for the long-term metabolomic study of human serum. *Anal. Chem.* 81, 1357–1364. doi: 10.1021/ac8019366

Zhu, Y., An, W., Peng, J., Li, J., Gu, Y., Jiang, B., et al. (2022). Genetic diversity of Nanmu (*Phoebe zhennan* S. Lee. et F. N. Wei) breeding population and extraction of core collection using nSSR, cpSSR and phenotypic markers. *Forests* 13, 1320. doi: 10.3390/f13081320

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 April 2024
DOI 10.3389/fpls.2024.1392212

OPEN ACCESS

EDITED BY
Xiang Pu,
Sichuan Agricultural University, China

REVIEWED BY
Peng Di,
Jilin Agriculture University, China
Tian Xie,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Yukihiro Sugimoto
yukihiro@kobe-u.ac.jp

RECEIVED 27 February 2024
ACCEPTED 03 April 2024
PUBLISHED 18 April 2024

CITATION
Homma M, Uchida K, Wakabayashi T, Mizutani M, Takikawa H and Sugimoto Y (2024) 2-oxoglutarate-dependent dioxygenases and BAHD acyltransferases drive the structural diversification of orobanchol in Fabaceae plants. *Front. Plant Sci.* 15:1392212. doi: 10.3389/fpls.2024.1392212

2-oxoglutarate-dependent dioxygenases and BAHD acyltransferases drive the structural diversification of orobanchol in Fabaceae plants

Masato Homma[1], Kiyono Uchida[2], Takatoshi Wakabayashi[1,2], Masaharu Mizutani[1], Hirosato Takikawa[2] and Yukihiro Sugimoto[1]*

[1]Department of Agrobioscience, Graduate School of Agricultural Science, Kobe University, Kobe, Japan, [2]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Strigolactones (SLs), a class of plant apocarotenoids, serve dual roles as rhizosphere-signaling molecules and plant hormones. Orobanchol, a major naturally occurring SL, along with its various derivatives, has been detected in the root exudates of plants of the Fabaceae family. Medicaol, fabacyl acetate, and orobanchyl acetate were identified in the root exudates of barrel medic (*Medicago truncatula*), pea (*Pisum sativum)*, and cowpea (*Vigna unguiculata*), respectively. Although the biosynthetic pathway leading to orobanchol production has been elucidated, the biosynthetic pathways of the orobanchol derivatives have not yet been fully elucidated. Here, we report the identification of 2-oxoglutarate-dependent dioxygenases (DOXs) and BAHD acyltransferases responsible for converting orobanchol to these derivatives in Fabaceae plants. First, the metabolic pathways downstream of orobanchol were analyzed using substrate feeding experiments. Prohexadione, an inhibitor of DOX inhibits the conversion of orobanchol to medicaol in barrel medic. The DOX inhibitor also reduced the formation of fabacyl acetate and fabacol, a precursor of fabacyl acetate, in pea. Subsequently, we utilized a dataset based on comparative transcriptome analysis to select a candidate gene encoding DOX for medicaol synthase in barrel medic. Recombinant proteins of the gene converted orobanchol to medicaol. The candidate genes encoding DOX and BAHD acyltransferase for fabacol synthase and fabacol acetyltransferase, respectively, were selected by co-expression analysis in pea. The recombinant proteins of the candidate genes converted orobanchol to fabacol and acetylated fabacol. Furthermore, fabacol acetyltransferase and its homolog in cowpea acetylated orobanchol. The kinetics and substrate specificity analyses revealed high affinity and strict recognition of the substrates of the identified enzymes. These findings shed light on the molecular mechanisms underlying the structural diversity of SLs.

KEYWORDS

strigolactone, biosynthesis, orobanchol diversification, 2-oxoglutarate-dependent dioxygenase, BAHD acyltransferase

1 Introduction

Strigolactones (SLs) are a type of plant apo-carotenoid that originates from all-*trans*-β-carotene. They serve as rhizosphere-signaling molecules and plant hormones (Aquino et al., 2021). These SLs, when exuded from plant roots, trigger the germination of seeds in root parasitic weeds belonging to the Orobanchaceae family, which includes the *Striga*, *Alectra*, *Orobanche*, and *Phelipanche* genera. This results in significant yield losses in key cereals, legumes, and vegetables across various regions of the world (Samejima and Sugimoto, 2018). Additionally, SLs induce hyphal branching in arbuscular mycorrhizal fungi, thereby fostering a beneficial symbiosis for plants (Akiyama et al., 2005). Internally, SLs act as regulators of plant architecture and development, functioning as a class of plant hormones (Gomez-Roldan et al., 2008; Umehara et al., 2008).

In terms of their chemical structure, SLs are categorized into canonical SLs, which consist of a tricyclic lactone (ABC ring) and a butenolide (D ring) connected with an enol ether bridge, and non-canonical SLs, where the ABC-ring system is incomplete (Supplementary Figure S1A). Canonical SLs are further subdivided into two types based on the configuration of the C-ring: strigol-type (Supplementary Figure S1B) with β-oriented C-rings, and orobanchol-type (Supplementary Figure S1C) with α-oriented C-rings. The structural variation is further increased by modifications of the A and B rings, such as hydroxylation, epoxidation, oxidation, and acetylation. To date, over 30 different SLs have been identified in the root exudates of various plants (Chesterfield et al., 2020).

Orobanchol, a major naturally occurring SL, has been detected in the root exudate of numerous plants, including monocots like rice (*Oryza sativa*) (Zhang et al., 2014) and dicots such as Sonalaceae (Wakabayashi et al., 2022), Fabaceae (Ueno et al., 2011), Cucurbitaceae (Khetkam et al., 2014), and Linaceae (Xie et al., 2009a). Regarding SL biosynthesis, enzyme reactions catalyzed by DWARF27 (D27), carotenoid cleavage dioxygenase 7 (CCD7), and CCD8 are known to convert all-*trans*-β-carotene to carlactone (Alder et al., 2012). The CYP711A subfamily is involved in the subsequent conversion of carlactone to carlactonoic acid (CLA) (Abe et al., 2014). Two distinct biosynthetic pathways of orobanchol from CLA have been elucidated. In rice, OsCYP711A2 and OsCYP711A3 convert CLA to 4-deoxyorobanchol (4DO) and then to orobanchol, respectively (Zhang et al., 2014). However, this biosynthetic pathway is presumed to be limited to certain monocot plants (Yoneyama et al., 2018). In dicot plants like tomato (*Solanum lycopersicum*) and cowpea (*Vigna unguiculat*a), an alternative pathway bypassing 4DO has been identified. In this pathway, CLA is converted to 18-oxo-CLA by CYP722C and subsequently to orobanchol by stereoselective BC-ring-forming factor (SRF) (Homma et al., 2023) (Supplementary Figure S1D).

Additionally, putative derivatives of orobanchol have been identified in the root exudates of various plants. For instance, downstream canonical SLs of orobanchol, namely 6,7-didehydroorobanchol and phelipanchols, have been identified in tomato (Supplementary Figure S1C) (Wakabayashi et al., 2022). These orobanchol derivatives were collectively termed didehydroorobanchol isomers (DDHs), with CYP712G1 identified as the enzyme responsible for the conversion of orobanchol to DDHs in tomato (Wang et al., 2022). The oxidation at the C7 position is suggested to be crucial for the conversion of orobanchol to DDHs in tomato (Wakabayashi et al., 2022). Furthermore, other plants also produce orobanchol derivatives biosynthesized through oxidation at the C7 position, such as 7-α/β-hydroxyorobanchyl acetate and 7-oxoorobanchyl acetate identified in cucumber (*Cucumis sativum*) (Khetkam et al., 2014) and flax (*Linum usitatissimum*) (Xie et al., 2009a), respectively (Supplementary Figure S1C).

Fabaceae plants, which encompass numerous important crops, exhibit several unique orobanchol derivatives. Medicaol, identified in the root exudates of barrel medic (*Medicago truncatula*), features a seven-membered A-ring instead of the typical six-membered A-ring found in SLs (Tokunaga et al., 2015). Similarly, fabacyl acetate, identified in the root exudates of pea (*Pisum sativum*), possesses an epoxide group at the A-ring and an acetoxy group at the C4 position (Xie et al., 2009b). Furthermore, orobanchyl acetate, characterized by an acetoxy group at the C4 position, was identified in the root exudates of cowpea (Ueno et al., 2011). These findings suggest distinct diversification pathways downstream of orobanchol in Fabaceae plants (Figure 1). However, the enzymes responsible for these diversifications are yet to be elucidated.

Despite advancements in SL research, the biological significance of the structural diversity of SLs remains unclear. Recent studies suggest that canonical SLs primarily function as signaling molecules in the rhizosphere (Ito et al., 2022), and investigations into their effects on the rhizosphere microbe community have been initiated (Schlemper et al., 2017; Kim et al., 2022). In a previous study, nodulation signaling pathway-1 (NSP1) and NSP2, which are GRAS-type transcription factors essential for rhizobium Nod factor-induced nodulation, are suggested to control not only downstream targets after rhizobium-induced signaling but also SL biosynthesis in barrel medic (Liu et al., 2011). Accordingly, molecular information on SL diversification in Fabaceae plants will accelerate studies on the rhizosphere microbiome.

The present study aimed to comprehensively elucidate the molecular and biochemical aspects of orobanchol diversification in Fabaceae plants. We conducted feeding experiments with potential biosynthetic precursors and inhibitors using plant hydroponics. Candidate genes of enzymes responsible for orobanchol derivatization were selected based on transcriptome data and co-expression analysis. Recombinant proteins of the candidate genes were heterologously expressed, and their enzyme activities were characterized. We report the identification of 2-oxoglutarate-dependent dioxygenases and BAHD acyltransferases responsible for the diversification of orobanchol to medicaol, fabacol, fabacyl acetate, and orobanchyl acetate. These findings will enable the manipulation of SL composition and shed light on the biological significance of SL diversification.

10.3389/fpls.2024.1392212

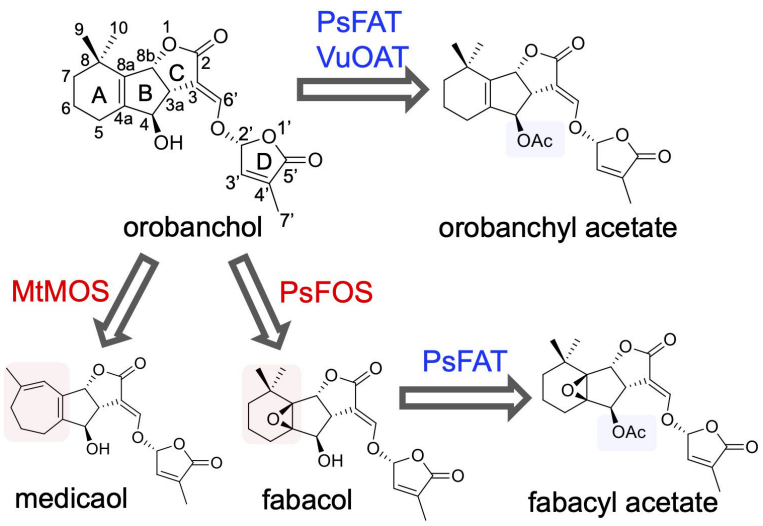


FIGURE 1
The diversification pathway of orobanchol in Fabaceae plants identified in this work (blue and red arrows). In barrel medic (*Medicago truncatula*), MtMOS converts orobanchol to medicaol. In pea (*Pisum sativum*), PsFOS converts orobanchol to fabacol. PsFAT acetylated fabacol and orobanchol, resulting in fabacyl acetate and orobanchyl acetate, respectively. In cowpea (*Vigna unguiculata*), VuOAT acetylated orobanchol.

2 Materials and methods

2.1 General

The ^{1}H NMR spectra were captured using a JNM-ECZ400S/L1 spectrometer (JEOL, Tokyo, Japan) in C_6D_6 at 400 MHz. The UV spectra were recorded on a V-630 spectrophotometer (JASCO, Tokyo, Japan). An LC-MS/MS system (Waters, Milford, MA, USA), which includes an Acquity UPLC H-Class and an Acquity TQD tandem mass spectrometer, was used to examine the SLs. Chromatographic separation was achieved using an ODS column (COSMOSIL 2.5C_{18}-MS-II, 100 × 2.0 mm i.d., 2.5 µm) (Nacalai Tesque, Kyoto, Japan) maintained at a temperature of 30°C. The elution process was carried out in a linear gradient system of MeOH-H_2O with 0.1% formic acid (50:50–100:0 in 20 min) at a flow rate of 0.2 ml/min. The TQD tandem mass spectrometer was operated under the following analytical conditions: in positive electrospray ionization mode, the capillary voltage was set to 3 kV, the source temperature was 120°C, the desolvation gas temperature was 350°C, the cone voltage was 30 V, and the collision energy was 20 eV. The nebulizer and desolvation N_2 gas flow rates were set to 50 and 550 L/h, respectively. Data acquisition and analysis were conducted using MassLynx 4.2 software (Waters).

2.2 Chemicals

Rac-orobanchol was synthesized as described in a previous study (Uchida et al., 2023). *Rac*-2′-*epi*-orobanchol was procured from StrigoLab (Turin, Italy). The racemates were separated chiral-wise using HPLC on a CHIRALPAK IC (250 × 10 mm, 5 µm) (Daicel Corporation, Osaka, Japan), as reported by Nomura et al. (2013). Fabacol stereoisomers were synthesized by treating the corresponding orobanchol stereoisomer with mCPBA in CH_2CL_2, as outlined in a previous study (Xie et al., 2009b). This was followed by HPLC separation on a COSMOSIL 5C_{18}-MS-II column (250 × 10 mm, 5 µm) (Nacalai Tesque) using MeOH-H_2O with 0.1% formic acid (50:50) as the mobile phase at a flow rate of 5.0 ml/min. Subsequently, HPLC separation on a CHIRALPAK IC (250 × 10 mm, 5 µm) was used with EtOH containing 0.1% formic acid as the mobile phase at a flow rate of 0.8 ml/min. Orobanchyl acetate stereoisomers were synthesized by acetylating the corresponding orobanchol stereoisomer, following the method described previously (Ueno et al., 2011). Fabacyl acetate diastereomers were synthesized by epoxidizing the corresponding orobanchol acetate stereoisomer with mCPBA. Both orobanchyl acetate and fabacyl acetate stereoisomers were purified using HPLC, following the same procedures used for the purification of fabacol stereoisomers. The enzyme-reaction product of orobanchol by MtMOS was characterized as follows: A large-scale enzymatic reaction was performed using 50 ml of the reaction mixture with the same composition as described in the section of the *in vitro* enzyme assay. The enzyme reaction was conducted for 90 min using the crude enzyme prepared from 500 ml of *E. coli* culture medium and approximately 0.5 mg of *rac*-orobanchol as substrate. The reaction product was extracted with EtOAc (50 ml × 3), and the organic layer was dried over Na_2SO_4 and concentrated in vacuo. The large-scale reaction was conducted twice, and the reaction product was purified using HPLC, following the same procedure used for the purification of fabacol stereoisomers, yielding approximately 240 µg of the product. Its ^{1}H-NMR spectrum was recorded in C_6D_6 (Supplementary Figure S2), which matched that reported for medicaol isolated from barrel medic (Tokunaga et al., 2015). 4DO was prepared as previously reported (Wakabayashi et al., 2023). The concentration of these chemicals was calculated using an absorption coefficient reported by Akiyama et al. (2010).

2.3 Plant material and hydroponics condition

Pea seeds (cv. *Akabana-Tsuruari*) and cowpea seeds (cv. *Blackeye*) were sourced from local suppliers, while barrel medic seeds (*Medicago truncatula* accession PI670016) were provided by the USDA. The pea and cowpea seeds were germinated on moistened filter paper in the dark for a period of 3 days. The barrel medic seeds were germinated on a 5 cm layer of vermiculite in a plastic strainer for a duration of 5 days. The seedlings of pea, cowpea, and barrel medic were then transferred to conical tubes and grown hydroponically in 50 mL of half-strength Hoagland nutrient solution at 23°C with a 16-h light/8-h dark photoperiod for 1 week, 2 weeks, and 4 weeks, respectively. For the feeding experiment, pea and barrel medic plants were grown in a half-strength Hoagland nutrient solution without phosphate. After 7 days and 11 days, respectively, the culture medium was replaced with fresh medium under phosphate-deficient conditions supplemented with 1 µM fluridone. After an additional 3 days, the culture medium was replaced again with fresh medium supplemented with 1 µM fluridone, the feeding substrate 40 nM *rac*-orobanchol or *rac*-orobanchyl acetate, and oxidation enzyme inhibitors 50 µM uniconazole-P or prohexadione dissolved in acetone. In the case of cowpea, plants grown in half-strength Hoagland were transferred to tap water with 1 µM fluridone. After 3 days, the hydroponic solution was replaced with fresh tap water supplemented with 1 µM fluridone and 40 nM *rac*-orobanchol. Twenty-four hours after feeding the substrate, each of the hydroponic solutions was subjected to Oasis HLB Vac Cartridge (Waters) to extract SLs as previously described (Moriyama et al., 2022), followed by LC-MS/MS analysis. For SL analysis in root exudates, plants were grown under the same conditions described above, but without fluridone and the oxidation enzyme inhibitors.

2.4 Gene expression analysis by reverse transcription-quantitative PCR (RT-qPCR)

cDNA templates were generated from RNAs isolated from the roots using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and ReverTra Ace qPCR RT Mix with genomic DNA Remover (TOYOBO, Osaka, Japan). The RNAs were extracted from the roots of three independent lines of hydroponically grown cowpea, barrel medic, and pea plants under both normal and phosphate-deficient conditions, as described in the previous section. These cDNA templates were then amplified using a StepOne (Applied Biosystems, Foster City, CA, USA) with PowerTrack SYBR Green Master Mix (Applied Biosystems). The primer sets, which are listed in Supplementary Table S1, were utilized to perform the qPCR analysis. The gene expression levels were normalized against the values of reference genes, and the relative gene expression levels were calculated based on the $2^{-\Delta\Delta Ct}$ method. The actin gene (ACT) (Wakabayashi et al., 2019), the ubiquitin gene (UBQ) (Liu et al., 2011), and the tubulin gene (TUB) (Guercio et al., 2022) served as reference genes for cowpea, barrel medic, and pea, respectively. The data acquisition and analysis were conducted using StepOne software v2.3 (Applied Biosystems).

2.5 Co-expression analysis by Confeito algorithm

Using the dataset from the Plant Bioinformatics Facility (https://urgi.versailles.inra.fr), genopea_counts_on_Genome_TPM.csv, which links the pea genome (Kreplak et al., 2019) to the pea transcriptome data (Alves-Carvalho et al., 2015), and the Confeito algorithm (Ogata et al., 2009) was used to perform gene co-expression analysis. This analysis revealed a local *PsSRF* (*Psat0ss8330g0240*) gene module; the genes within this module are detailed in Supplementary Table S2.

2.6 The expression of recombinant proteins in *E. coli*

The complete coding sequences (CDS) were sourced from the Phytozome and Ensemble Plants databases. The CDS of the candidate genes were amplified from root cDNA, which was prepared from barrel medic, pea, and cowpea under phosphate-deficient conditions. This amplification used primers with 15-bp overlap sequences for in-fusion cloning, as listed in Supplementary Table S1. The amplified DNA fragments were then inserted into the NdeI and KpnI restriction sites of the modified pCold III (Takara Bio, Shiga, Japan) expression vector using the In-Fusion HD cloning kit, as described by Wakabayashi et al. (2023). The constructed vectors were sequenced and introduced into the *E. coli* strain BL21 (DE3), which harbored the pGro12 plasmid (Nishihara et al., 1998), for the co-expression of the molecular chaperonins.

The transformed *E. coli* was cultured overnight at 37°C in LB medium supplemented with 100 µg/ml ampicillin and 50 µg/ml kanamycin. The overnight cultures were then diluted into LB medium containing 100 µg/ml ampicillin, 50 µg/ml kanamycin, and 0.5 mg/ml L-arabinose (for the induction of GroES/GroEL expression). The culture medium was incubated at 37°C until the OD600 reached 0.7. After chilling the cultures on ice for 30 min, protein expression was induced by adding 0.1 mM isopropyl β-D-thiogalactoside, and the cells were incubated for an additional 24 h at 15°C. The *E. coli* cells were collected by centrifugation and washed with phosphate-buffered saline. The crude enzymes were prepared as follows: the cells were disrupted by sonication in buffer A (50 mM potassium phosphate buffer (pH 7.4) and 20% (v/v) glycerol). After centrifugation at 10,000 g for 30 min at 4°C, the supernatant was used as the crude enzymes. A negative control was prepared from the cell harboring the empty vector.

2.7 Protein purification

E. coli cells, which expressed the recombinant proteins, were disrupted by sonication in a lysis buffer composed of 50 mM potassium phosphate buffer (pH 7.4), 500 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 20% (v/v) glycerol, 1.0% (w/v) CHAPS, 5 mM ATP, and 25 mM imidazole. Following centrifugation at 10,000 g for 30 min at 4°C, the supernatant was

applied to a TALON spin column (Takara Bio) that had been equilibrated with the lysis buffer. The proteins bound to the column were washed twice with 600 µl of the lysis buffer and then eluted twice with 200 µl of an elution buffer. This elution buffer was composed of 50 mM potassium phosphate buffer (pH 7.4), 500 mM KCl, 0.1 mM EDTA, 0.1 mM DTT, 20% (v/v) glycerol, 1.0% (w/v) CHAPS, and 500 mM imidazole. The eluent was then desalted using buffer A (50 mM potassium phosphate buffer (pH 7.4) and 20% (v/v) glycerol) with a 10 kDa cut-off Amicon Ultra concentrator (Merck Millipore, Burlington, MA, USA). Protein purification was confirmed by SDS-PAGE (Supplementary Figure S3A). The purified proteins were stored at −80°C. The protein concentration was estimated using the Pierce BCA Protein Assay Kit, following the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA).

2.8 *In vitro* enzyme assay

The DOX enzyme assay was conducted with a reaction mixture (50 µl) that included 50 mM potassium phosphate buffer (pH 7.4), 5 mM 2-oxoglutaric acid, 10 mM sodium ascorbate, 200 µM $FeSO_4$, and either the crude enzymes or the purified enzymes. The BAHD acyltransferase enzyme assay was performed with a reaction mixture (50 µl) that consisted of 50 mM citrate buffer (pH 6.0), 200 µM acetyl-CoA, and either the crude enzymes or the purified enzymes. The identification assay for MtMOS, PsFOS, PsFAT, and VuOAT was conducted using the crude enzymes and 1 µM orobanchol or fabacol as substrates. For the kinetics analysis of the purified enzymes of MtMOS, PsFOS, and VuOAT, the concentration of orobanchol ranged from 0.05 to 1 µM. For the kinetics analysis of the PsFAT crude enzyme, the concentration of fabacol and orobanchol ranged from 0.25 to 10 µM. The enzymatic reaction was initiated by the addition of a substrate and conducted at 30°C for 10 min. The reaction products were extracted twice using 100 µl of EtOAc. The collected organic phase was evaporated, and the residues dissolved in acetonitrile were analyzed by LC-MS/MS. The standard curves for quantification were plotted using fabacol, fabacyl acetate, orobanchyl acetate, and medicaol standards as described in the section on chemicals. K_m and K_{cat} values were obtained using GraphPad Prism version 10 (Supplementary Figure S3B). For the analysis of substrate specificity of the purified enzymes (MtMOS, PsFOS, and VuOAT) and the crude enzyme of PsFAT, the enzymatic reaction using 0.2 µM substrates (orobanchol stereoisomers, fabacol stereoisomers, orobanchyl acetate, and 4DO) was conducted at 30°C for 20 min.

2.9 Phylogenetic tree analysis

The amino acid sequences utilized for the phylogenetic analysis were sourced from three databases: the National Center for Biotechnology Information, Phytozome, and Ensemble Plants. These sequences were aligned using the ClustalW tool. A phylogenetic tree was then constructed using the maximum-likelihood method, specifically employing the LG model with gamma-distributed rate variation among sites. Bootstrap statistics were computed using 1,000 replicates to ensure the robustness of the tree. All these phylogenetic analyses were conducted using the MEGA software, version 11 (Tamura et al., 2021).

3 Results

3.1 Metabolic pathway downstream of orobanchol in cowpea, barrel medic, and pea

Cowpea, barrel medic, and pea plants were cultivated hydroponically under phosphate-deficient conditions to stimulate the production of SLs (Yoneyama et al., 2007; López-Ráez et al., 2008). SLs extracted from root exudates were analyzed using LC-MS/MS. Orobanchyl acetate, medicaol, and fabacyl acetate were detected in the respective culture filtrates, along with orobanchol, as previously reported (Supplementary Figure S4) (Xie et al., 2009b; Ueno et al., 2011; Tokunaga et al., 2015). Each hydroponic system was treated with fluridone (Iseki et al., 2018), a phytoene desaturase inhibitor, which reduced SL accumulation to negligible levels. Feeding orobanchol to the hydroponics with fluridone revealed the conversion of the exogenously applied orobanchol to orobanchyl acetate, medicaol, and fabacyl acetate, respectively (Figures 2A–C).

In the pea root exudate, orobanchyl acetate was detected, but fabacol was not (Supplementary Figure S4C), as previously reported (Xie et al., 2009b). The exogenous application of orobanchyl acetate to the pea hydroponics did not significantly increase the accumulation of fabacyl acetate (Figure 2C). Notably, a significant amount of fabacol, which was below the detection limit in pea root exudate (Supplementary Figure S4C), accumulated in the hydroponics supplemented with orobanchol (Figure 2C; Supplementary Figure S5B). Therefore, the biosynthetic pathway from orobanchol to fabacyl acetate is presumed to proceed via fabacol (Supplementary Figure S5A).

The effects of uniconazole-P and prohexadione, inhibitors of cytochrome P450 (CYP) and 2-oxoglutarate and iron (II)-dependent dioxygenase (DOX) family enzymes, respectively, on the diversification of orobanchol were investigated. In barrel medic, the conversion of orobanchol to medicaol was significantly inhibited by prohexadione but not by uniconazole-P, suggesting the involvement of a DOX enzyme in the conversion (Figure 2B). In pea, the amount of fabacyl acetate and fabacol derived from orobanchol significantly decreased due to prohexadione (Figure 2C). The conversion of orobanchol to orobanchyl acetate and fabacyl acetate was inhibited by uniconazole-P (Figure 2C), with an increase in fabacol (Figure 2C). The uniconazole-P treatment seemed to inhibit the acetylation of fabacol and orobanchol by unknown mechanisms, although the inhibition of acyltransferase by uniconazole-P has not been reported. These results indicated that a DOX enzyme is responsible for the epoxidation of orobanchol in pea.

10.3389/fpls.2024.1392212

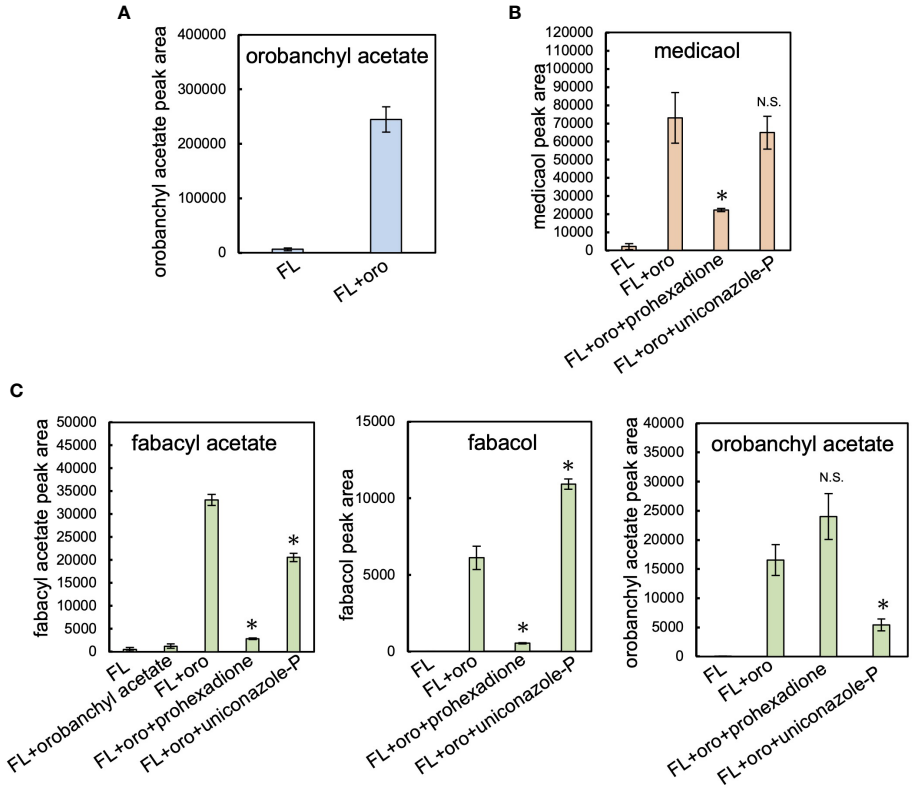


FIGURE 2
Feeding experiment in cowpea, barrel medic, and pea. **(A)** The level of orobanchyl acetate in cowpea root exudate supplemented with orobanchol. **(B)** The conversion of orobanchol to medicaol in barrel medic treated with oxidation enzyme inhibitors. Uniconazole-P and prohexadione are inhibitors of CYP and DOX, respectively. **(C)** The conversion of feeding substrates by pea treated with oxidation enzyme inhibitors to fabacyl acetate, fabacol, and orobanchyl acetate. Data are the means of three replicates ± SE. FL and oro denote fluridone and orobanchol, respectively. An asterisk indicates a significant difference (*$p < 0.05$, Student's t test): N.S. means not significant between orobanchol feeding experiments with and without oxidation enzyme inhibitors. The peak area of each SL estimated by LC-MS/MS analysis was used for quantification.

3.2 Identification of medicaol synthase in barrel medic

We expanded our research on orobanchol diversification to identify medicaol synthase (MOS), the DOX enzyme responsible for converting orobanchol to medicaol in barrel medic (Figure 3A). Previous studies on barrel medic mutants of NSP2, a GRAS-type transcription factor essential for rhizobium Nod factor-induced nodulation, reported a drastic decrease in the level of medicaol and an increase in orobanchol in the root exudate of the *nsp2* mutant (Liu et al., 2011; Tokunaga et al., 2015). These findings suggested that NSP2 regulates the expression of MOS, and impairment of NSP2 leads to the downregulation of *MOS* (Figure 3A). In a previous study, downregulated probes in *nsp2* roots compared to the wild type were listed based on microarray data. The top 100 downregulated probes were converted into gene IDs based on the barrel medic genome assembly (Mt4.0 v1), revealing that *Medtr7g063730* is the only gene encoding DOX (Supplementary Table S3). SL biosynthesis genes, including *MtD27* (*Medtr1g471050*), are known to be upregulated under phosphate-deficient conditions (Yoneyama et al., 2007; Liu et al., 2011). RT-qPCR analysis on the roots of barrel medic plants showed the upregulation of *Medtr7g063730* under phosphate deficiency (Figure 3B). These results suggest the involvement of this gene in SL biosynthesis as a MOS. We cloned *Medtr7g063730* from the cDNA of barrel medic roots, expressed the recombinant protein in *E. coli*, and performed enzyme assays using orobanchol as a substrate. The recombinant protein catalyzed the conversion of orobanchol to medicaol (Figure 3C), confirming that *Medtr7g063730* is the gene encoding MOS in barrel medic (Figure 3A).

DOX enzymes constitute the second-largest enzyme family in plants and are phylogenetically classified into DOXA, DOXB, and DOXC classes (Kawai et al., 2014). DOXC comprises 57 clades and is involved in the biosynthesis of plant secondary metabolites and plant hormones (Kawai et al., 2014). To understand the phylogeny of MtMOS among diverse DOX enzymes, a BLASTP search using the amino acid sequences of MtMOS as a query was conducted. Only Fabaceae plants, such as barrel medic, pea, soybean (*Glycine max*), red clover (*Trifolium pratense*), chickpea (*Cicer arietinum*), and birdsfoot trefoil (*Lotus japonicus*), have DOX enzymes homologous to MtMOS, sharing an amino acid sequence identity of more than 65%. DOX enzymes with the highest identities to MtMOS and its homologs are found in the DOXC54 clade, which contains lateral branching oxidoreductase (LBO) involved in SL biosynthesis (Kawai et al., 2014; Brewer et al., 2016). However, the

10.3389/fpls.2024.1392212

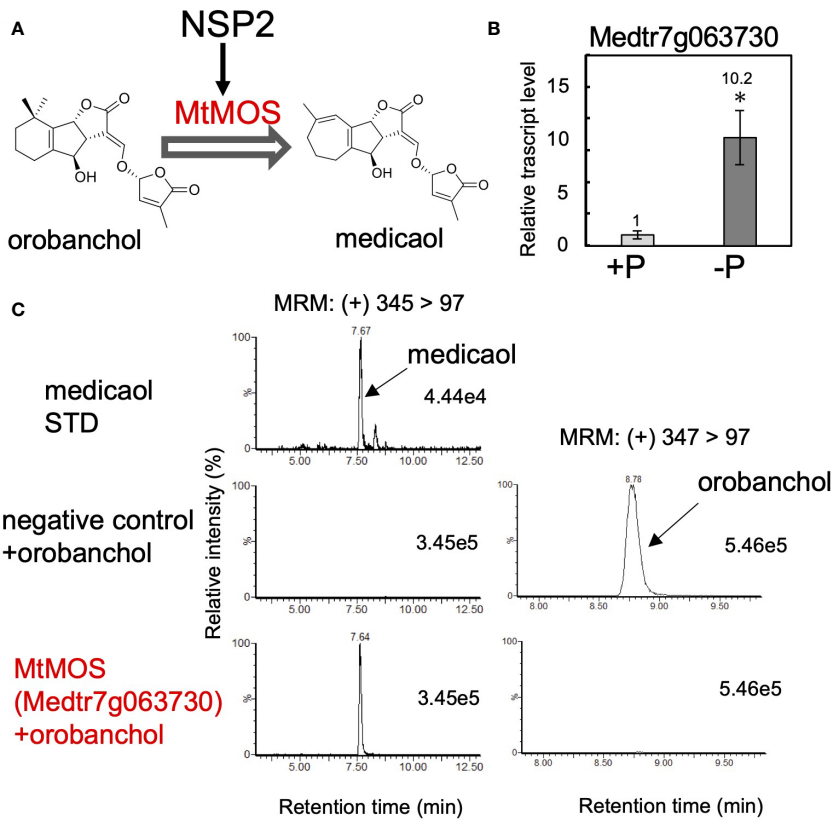


FIGURE 3
Identification of MtMOS. **(A)** The expression of MtMOS responsible for the conversion of orobanchol to medicaol is mediated by NSP2 in barrel medic. **(B)** The relative gene expression level of Medtr7g063730 under conditions with or without phosphate. Data are the means of three replicates ± SE. An asterisk indicates a significant difference (*$p < 0.05$, Student's t test). **(C)** Enzyme activity of the recombinant protein of the candidate gene *Medtr7g063730* toward orobanchol. An enzyme assay was conducted using crude protein prepared from *E. coli* expressing candidate enzymes. The crude enzyme prepared from *E. coli* expressing a non-inserted empty vector was used as a negative control.

amino acid identity is less than 45%. For instance, MtMOS shares amino acid identities of 43% and 44% with AtLBO (AT3G21420.1) from *Arabidopsis* (*Arabidopsis thaliana*) and its homolog from rice (LOC_Os01g70930.1), respectively. The phylogenetic tree shows that these Fabaceae DOX enzymes form a monophyletic clade, which is significantly different from the DOXC54 clade (Supplementary Figure S6). Thus, we designate these DOX enzymes represented by MtMOS as DOXC54-like 1.

3.3 Identification of fabacol synthase in pea

Following the results obtained from the feeding experiments conducted in pea plants, our research expanded to the identification of fabacol synthase (FOS), the DOX enzyme responsible for converting orobanchol to fabacol (Figure 4A). The biosynthetic pathway of orobanchol in pea has not yet been elucidated. However, homologs of VuCYP722C and VuSRF, which are responsible for the biosynthesis of orobanchol in cowpea (Homma et al., 2023), are found in the pea genome assembly (*Pisum sativum* Cameor genome v1a) (Kreplak et al., 2019). PsCYP722C (Psat4g095720) and PsSRF (Psat0ss8330g0240) exhibited high amino acid identities of 86% and 83%, respectively, with VuCYP722C and VuSRF, respectively. Thus, the same biosynthetic pathway of orobanchol as that of cowpea is suggested to operate in pea. Typically, biosynthetic genes of plant secondary metabolites exhibit similar expression patterns. Given that orobanchol serves as the substrate for PsFOS, we hypothesized that PsFOS would exhibit co-expression with PsSRF, responsible for orobanchol formation. The Plant Bioinformatics Facility (https://urgi.versailles.inra.fr) provided the dataset that links the pea genome (Kreplak et al., 2019) to the pea transcriptome, which was generated for a comprehensive gene expression atlas using 20 cDNA libraries prepared from various tissues collected at different developmental stages of pea plants grown under normal and low nitrate conditions (Alves-Carvalho et al., 2015). Nitrate deficiency was reported to induce SL biosynthesis in pea plants (Foo et al., 2013). Analysis of the dataset revealed exclusive expression of *PsSRF* (*Psat0ss8330g0240*) in roots (Supplementary Figure S7). Subsequently, co-expression analysis using the Confeito algorithm (Ogata et al., 2009) was conducted based on the dataset, resulting in the identification of the co-expression gene module of *PsSRF* (Supplementary Table S2). This gene module contained *PsD27* (*Psat6g058200*) and *PsCYP711A* (*Psat5g036360*), homologous genes to known SL biosynthesis genes, validating its efficacy in

10.3389/fpls.2024.1392212

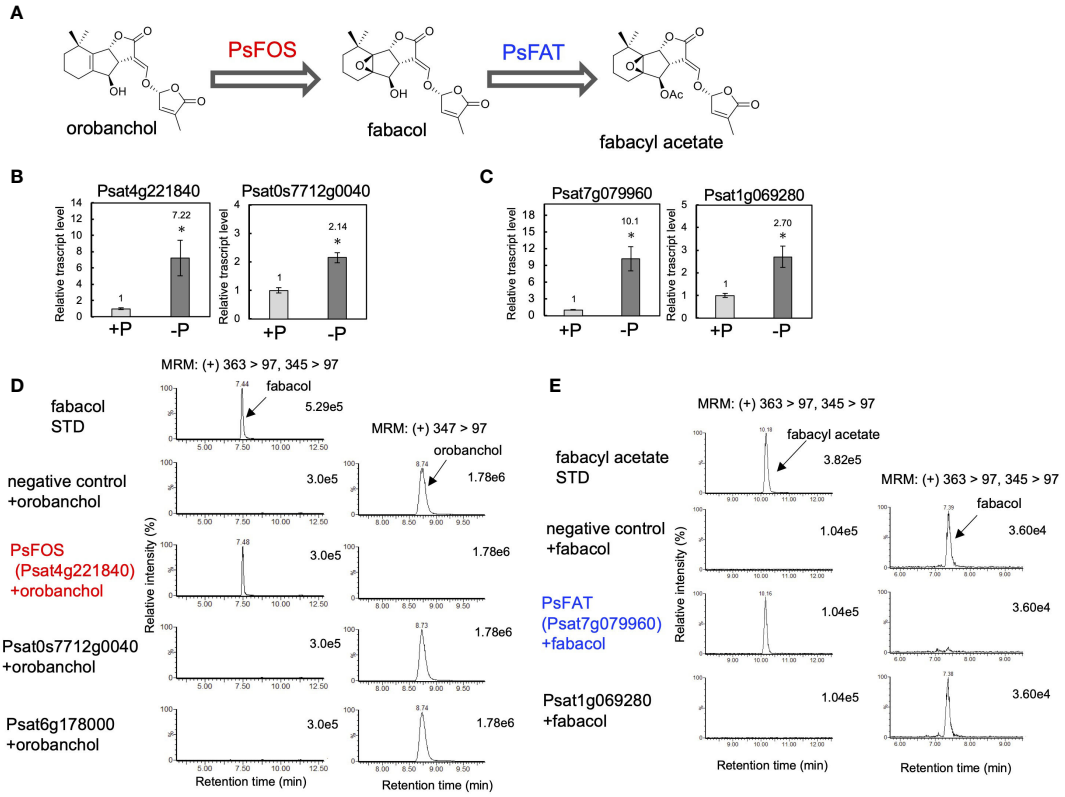


FIGURE 4
Identification of PsFOS and PsFAT. **(A)** The putative biosynthetic pathway of fabacyl acetate. PsFOS converts orobanchol to fabacol, which is acetylated by PsFAT. **(B)** The relative gene expression level of the candidate genes for PsFOS with or without phosphate. **(C)** The relative gene expression level of the candidate genes for PsFAT with or without phosphate. Data are the means of three replicates ± SE. An asterisk indicates a significant difference (*$p < 0.05$, Student's t test). **(D)** Enzyme activity of the recombinant protein of the candidate gene for PsFOS. **(E)** Enzyme activity of the recombinant protein of the candidate gene for PsFAS. An enzyme assay was conducted using crude protein prepared from *E. coli* expressing candidate enzymes. The crude enzyme prepared from *E. coli* expressing a non-inserted empty vector was used as a negative control.

exploring candidate genes encoding PsFOS. Two DOX-encoding genes, *Psat4g221840* and *Psat0s7712g0040*, were included in the co-expression gene module (Supplementary Table S2, Supplementary Figure S7), with their expression upregulated by phosphate deficiency in roots (Figure 4B). Consequently, these two genes were selected as candidate genes encoding FOS. *Psat4g221840* and *Psat0s7712g0040* were cloned from cDNA synthesized from mRNA extracted from pea roots. Recombinant proteins of both genes were expressed in *E. coli*, and enzyme assays were conducted using orobanchol as the substrate. The recombinant protein of *Psat4g221840* facilitated the conversion of orobanchol to fabacol, while the other candidate gene did not yield any reaction product (Figure 4D), confirming *Psat4g221840* as the gene encoding FOS in pea (Figure 4A).

According to the pea genome, a homolog of PsFOS (Psat6g178000) shares a high amino acid identity of 83%. However, the recombinant protein of *Psat6g178000* did not exhibit enzyme activity toward orobanchol (Figure 4D). Other Fabaceae plants, including soybean, chickpea, barrel medic, red clover, and birdsfoot trefoil, possess homologs of PsFOS with amino acid identities exceeding 65%. DOX enzymes with the highest identities to PsFOS and its homologs are found in DOXC31, which encompasses functionally diverse enzymes identified from various plant species (Kawai et al., 2014), with amino acid identities reaching approximately 60%. The phylogenetic tree, including previously identified DOXC31 enzymes, indicated that these Fabaceae DOX enzymes are part of the DOXC31 clade (Supplementary Figure S8).

3.4 Identification of fabacol acetyltransferase in pea and orobanchol acetyltransferase in cowpea

The negligible conversion of orobanchyl acetate to fabacyl acetate and the successful conversion of orobanchol to fabacol (Figure 2C), combined with the identification of PsFOS in pea (Figure 4D), suggest that fabacyl acetate is biosynthesized from fabacol through the acetylation of the hydroxy group at the C4 position (Figure 4A). To identify fabacol acetyltransferase (FAT) in pea, we selected genes encoding BAHD acyltransferases, which constitute the enzyme family involved in the acylation of plant secondary metabolites, from the same gene module used for identifying PsFOS. This gene module contained two genes

10.3389/fpls.2024.1392212

encoding BAHD acyltransferase (*Psat7g079960* and *Psat1g069280*) (Supplementary Table S2, Supplementary Figure S7), both of which were upregulated by phosphate deficiency in pea roots (Figure 4C). The candidate genes, *Psat7g079960* and *Psat1g069280*, were cloned from pea root cDNA. Recombinant proteins of the two genes were expressed in *E. coli*, and the enzyme assay was conducted with acetyl-CoA and fabacol as the acyl donor and acceptor, respectively. The recombinant protein of *Psat7g079960* catalyzed the conversion of fabacol to fabacyl acetate, while that of *Psat1g069280* did not (Figure 4E), indicating that *Psat7g079960* is the gene encoding FAT in pea (Figure 4A).

In accordance with the detection of orobanchyl acetate in pea root exudates (Supplementary Figure S4C), PsFAT also acetylated orobanchol, while the recombinant protein of *Psat1g069280* did not (Figure 5C). This result suggests that the homolog of PsFAT in cowpea could acetylate orobanchol because both orobanchol and orobanchyl acetate are detected in cowpea root exudates (Supplementary Figure S4A). *Vigun03g171500*, the only gene encoding the homolog of PsFAT with 59% amino acid sequence identity, was upregulated by phosphate deficiency in cowpea roots (Figure 5B). The recombinant protein of this gene exhibited acetyltransferase activity toward orobanchol (Figure 5C), indicating that *Vigun03g171500* is the gene encoding orobanchol acetyltransferase (OAT) in pea (Figure 5A).

Previous studies have indicated that BAHD acyltransferases are phylogenetically divided into eight clades (Tuominen et al., 2011; Yuan et al., 2022). Phylogenetic analysis of PsFAT and VuOAT, the first identified acyltransferases involved in SL biosynthesis, revealed that only Fabaceae plants, including soybean, barrel medic, red clover, and birdsfoot trefoil, have BAHD acyltransferases sharing more than 60% amino acid identity with PsFAT or VuOAT. The phylogenetic tree, including previously identified BAHD acyltransferases from each clade, showed that PsFAT, VuOAT, and their homologs in Fabaceae plants belong to the IIIa clade of BAHD acyltransferases (Supplementary Figure S9), many of which are known to be involved in the acylation of alcohol acceptors using acetyl-CoA as the primary acyl donor (D'Auria, 2006; Yuan et al., 2022).

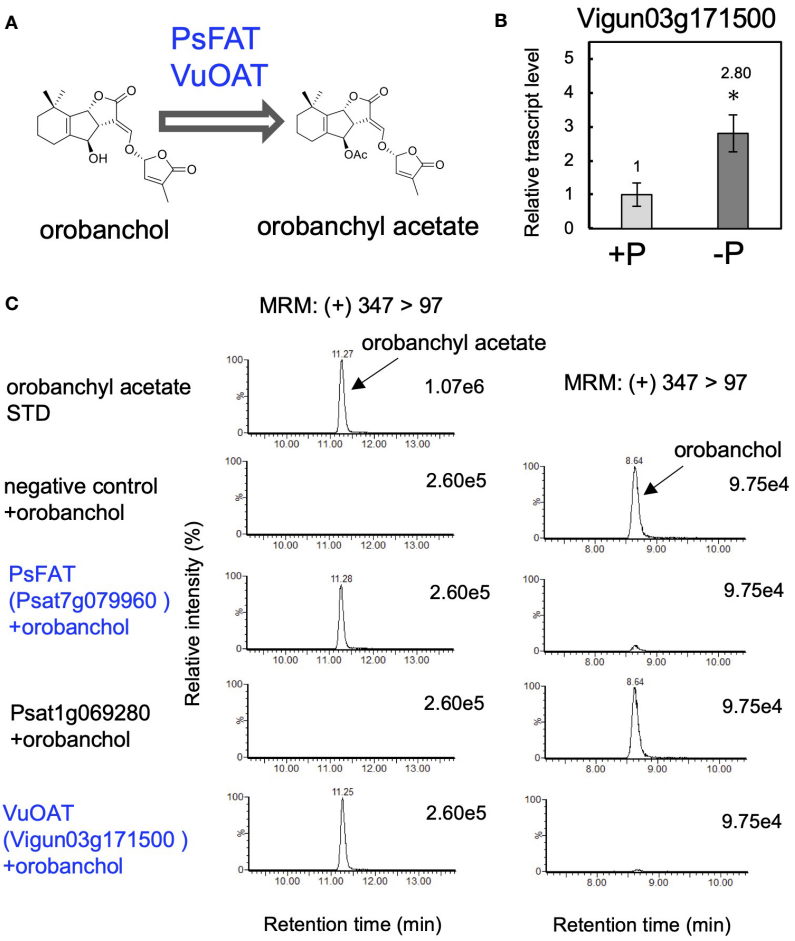


FIGURE 5
Identification of orobanchol acetyltransferases in pea and cowpea. **(A)** The acetylation of orobanchol by PsFAT and VuOAT was elucidated in this study. **(B)** The relative gene expression level of *Vigun03g171500* under conditions with or without phosphate. Data are the means of three replicates ± SE. An asterisk indicates a significant difference (*$p < 0.05$, Student's t test). **(C)** Acetylation activity toward orobanchol of candidate enzymes. Enzyme assays were conducted using crude protein prepared from *E. coli* expressing PsFAT, Psat1g069280, and VuOAT, a homolog of PsFAT from cowpea. The crude enzyme prepared from *E. coli* expressing a non-inserted empty vector was used as a negative control.

10.3389/fpls.2024.1392212

3.5 The biochemical characterization of MtMOS, PsFOS, PsFAT, and VuPAT

Finally, we conducted biochemical characterization of the enzymes identified in this research. The recombinant proteins of DOX enzymes, MtMOS, and PsFOS were successfully purified (Supplementary Figure S3A), and their kinetic parameters for orobanchol were meticulously investigated. Notably, MtMOS and PsFOS displayed small apparent affinities (K_m) for orobanchol, measuring at 0.070 and 0.162 μM, respectively (Table 1A). Due to the exceptionally high affinity of MtMOS toward orobanchol, its catalytic efficiency (K_{cat}/K_m) surpassed that of PsFOS, although their K_{cat} values were comparable (Table 1A). To explore substrate specificity, an enzyme assay was conducted using orobanchol stereoisomers (orobanchol, *ent*-orobanchol, 2′-*epi*-orobanchol, and *ent*-2′-*epi*-orobanchol) as substrates (Figure 6A). Remarkably, MtMOS and PsFOS exclusively exhibited enzyme activity toward orobanchol (Figures 6B, Supplementary Figure S10A). Furthermore, the conversion of 4DO and orobanchyl acetate by MtMOS and PsFOS was not observed (Figures 6B, Supplementary Figures S10B, C), indicating strict stereochemical recognition of orobanchol and the hydroxy group at the C4 position by these enzymes.

Moving on to the biochemical characterization of PsFAT, due to its low expression level in *E. coli*, crude proteins were utilized. In contrast, the recombinant VuOAT enzyme was successfully purified (Supplementary Figure S3A). The apparent affinities (K_m) of PsFAT for fabacol and orobanchol were determined to be 0.287 and 2.277 μM, respectively (Table 1B). PsFAT exhibited a higher affinity for fabacol than for orobanchol, with the V_{max} value toward fabacol approximately double that toward orobanchol (Table 1B), indicating fabacol as the more preferred substrate for PsFAT. The K_m and K_{cat}/K_m values of VuOAT for orobanchol were 0.340 μM and 151×10^{-3} ($s^{-1} \cdot \mu M^{-1}$), respectively (Table 1A). The K_m value of VuOAT for orobanchol was comparable to that of PsFAT for fabacol and much smaller than that of PsFAT for orobanchol (Tables 1A, B).

To further probe substrate specificities, PsFAT and VuOAT were incubated with acetyl-CoA and four stereoisomers of fabacol and orobanchol, respectively (Figure 6A). Stereoisomers with β-oriented C-rings (*ent*-fabacol, *ent*-2′-*epi*-fabacol, *ent*-orobanchol, and *ent*-2′-*epi*-orobanchol) were not acetylated (Figures 6C, Supplementary Figures S11A, B). Weak acetylation activities toward 2′-*epi*-fabacol and 2′-*epi*-orobanchol were observed (Supplementary Figures S11A, B), approximately one twenty-fifth of the activity toward fabacol and orobanchol, respectively (Figure 6C). These results suggest that the C-ring configuration plays a crucial role in substrate recognition for these acyltransferases, while 2′-epimers are unlikely to occur in plants since naturally occurring SLs exclusively possess the *R* configuration at C2′ in the D ring.

4 Discussion

More than 30 SLs have been identified in the root exudates of various plants (Chesterfield et al., 2020), yet the evolutionary trajectory of enzymes involved in SL biosynthesis and its biological significance remain areas of ongoing investigation. The diverse oxidative modifications resulting from the evolution of CYP and DOX enzymes are major drivers of the structural diversity observed in secondary metabolites. Previous research has primarily focused on CYP family enzymes responsible for modifications on the A-ring of SLs, such as CYP728B35 (Wakabayashi et al., 2021), CYP706C37 (Li et al., 2023), CYP71AH (Wakabayashi et al., 2023), and CYP712G1 (Wang et al., 2022), in sorghum (*Sorghum bicolor*), maize (*Zea mays*), cotton (*Gossypium hirsutum*) and tomato, respectively. However, the identification of MtMOS and PsFOS in this work underscores the crucial role of DOX enzymes in the structural diversification of orobanchol within Fabaceae plants. To date, several DOX enzymes implicated in SL biosynthesis have been reported, although their precise functions remain uncertain. For example, LBO, classified into the DOXC54 clade, oxidizes methyl carlactonoate and is speculated to be involved in the biosynthesis of an unidentified SL that acts as a plant hormone inhibiting shoot branching (Brewer et al., 2016). Similarly, lotuslactone defective (LLD) (Mori et al., 2020) and Sb3500 (Yoda et al., 2023), classified

TABLE 1 The kinetics analysis of the enzymes identified in this study.

(A)				
enzyme	substrate	K_m (μM)	K_{cat} (/s)	K_{cat}/K_m (/s·μM)
MtMOS	orobanchol	0.070 ± 0.013	$2.66 \times 10^{-3} \pm 0.12 \times 10^{-3}$	$38.0 \times 10^{-3} \pm 7.2 \times 10^{-3}$
PsFOS	orobanchol	0.162 ± 0.022	$2.66 \times 10^{-3} \pm 0.12 \times 10^{-3}$	$16.4 \times 10^{-3} \pm 2.3 \times 10^{-3}$
VuOAT	orobanchol	0.340 ± 0.094	$51.4 \times 10^{-3} \pm 6 \times 10^{-3}$	$151 \times 10^{-3} \pm 45 \times 10^{-3}$

(B)			
enzyme	substrate	K_m (μM)	V_{max} (pM/s)
PsFAT	fabacol	0.287 ± 0.076	124.232 ± 6.500
PsFAT	orobanchol	2.277 ± 0.429	57.222 ± 3.588

(A) The kinetics parameters of three enzymes: MtMOS, PsFOS, and VuOAT. (B) The kinetics parameters of PsFOS. The data are the means of three replicates ± SE.

10.3389/fpls.2024.1392212

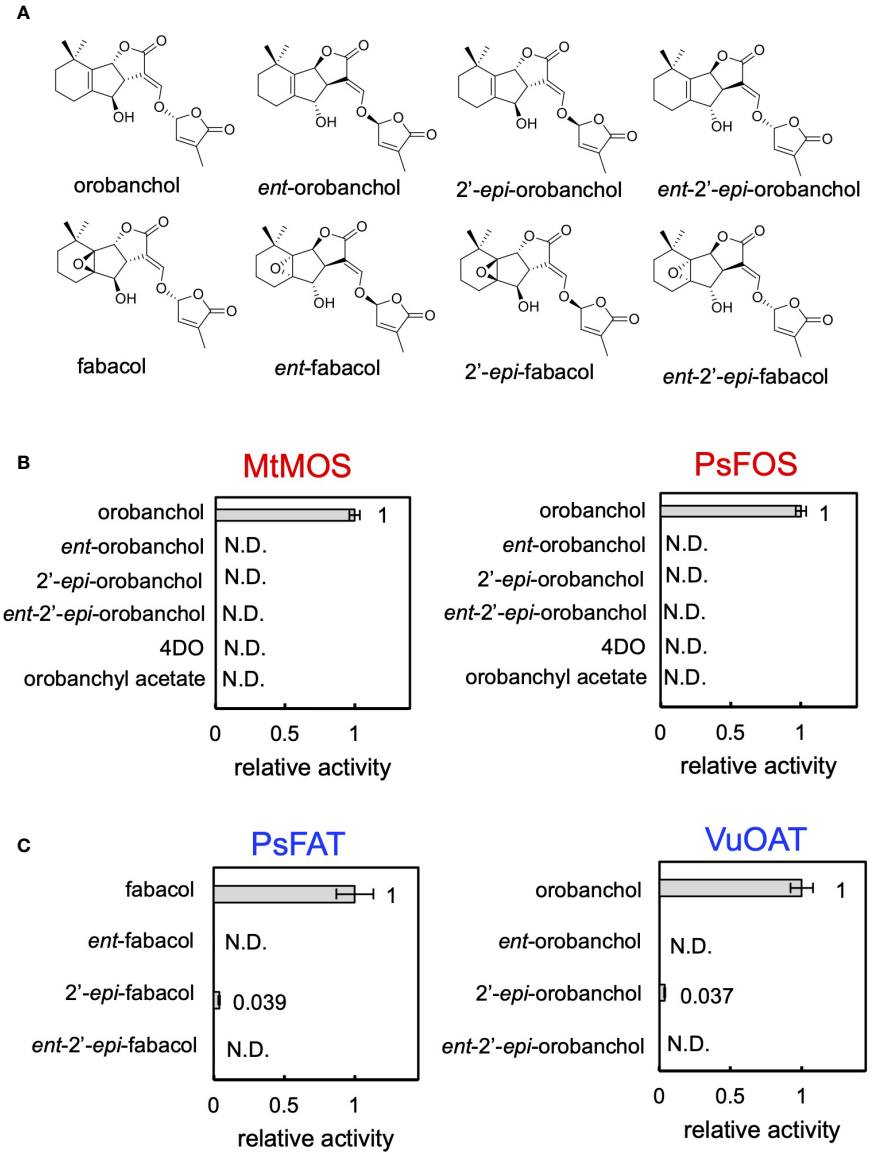


FIGURE 6
The substrate specificity of MtMOS, PsFOS, PsFAT, and VuOAT. **(A)** The chemical structure of the substrate used for the analysis, the stereoisomers of orobanchol and fabacol. **(B)** The enzyme activity of MtMOS and PsFOS toward the substrates, including orobanchol diastereomers, 4DO, and orobanchyl acetate. **(C)** The substrate specificity of PsFAT and VuOAT toward the diastereomers of fabacol and orobanchol, respectively. Relative enzyme activity of PsFAT toward the fabacol diastereomers normalized to the activity toward fabacol. The relative acetylation activity of VuOAT toward orobanchol diastereomers was normalized to the activity toward orobanchol. Data are the means of three replicates ± SE. N.D. means not detected.

into the DOXC55 clade, are involved in the biosynthesis of lotuslactone and 5-deoxystrigol in birdsfoot trefoil and sorghum, respectively. It is anticipated that further exploration of DOXC-class enzymes, particularly those phylogenetically related to the DOXC54 and DOXC55 clades, will lead to the identification of additional SL biosynthetic enzymes.

In the phylogenetic analysis, MtMOS is classified within the phylogenetic clade DOXC54-like 1, which is suggested to be phylogenetically linked to the DOXC54 clade (Supplementary Figure S6). While LBO, classified as DOXC54, remains highly conserved among seed plant species (Brewer et al., 2016), non-Fabaceae plants, such as Solanaceae, Cucurbitaceae, and Linaceae, known for producing orobanchol derivatives (Xie et al., 2009a; Kohlen et al., 2013; Wang and Bouwmeester, 2018), lack DOX enzymes homologous to DOXC54-like 1. DOX enzymes in these plants share only about 40% amino acid identity with DOXC54-like 1 enzymes. Consequently, DOXC54-like 1 enzymes are presumed to have evolved from DOXC54 clade enzymes in Fabaceae plants and are likely implicated in SL diversification specific to this plant family. Several Fabaceae plants reportedly produce putative DDHs, including medicaol (Yoneyama et al., 2008; Tokunaga et al., 2015; Galili et al., 2021). DOXC54-like 1 enzymes may have a role in the conversion of orobanchol to DDHs in this plant family. Furthermore, birdsfoot trefoil produces lotuslactone (Xie et al., 2019), a non-canonical SL sharing the same seven-membered A-ring as medicaol (Supplementary Figure S1A). In birdsfoot trefoil, a

DOXC54-like 1 enzyme is probably involved in the A-ring expansion reaction of an unidentified substrate leading to lotuslactone. Therefore, further investigation into DOXC54-like 1 can shed more light on SL diversification in Fabaceae plants.

On the contrary, PsFOS is categorized under DOXC31, phylogenetically distinct from the adjacent clades of DOXC54 and DOXC55 (Kawai et al., 2014). The DOXC31 clade is one of the largest DOXC clades, encompassing functionally diverse enzymes responsible for lineage-specific metabolites (Kawai et al., 2014). For example, ZmBX6 (Jonczyk et al., 2008), CrD4H (Vazquez-Flota et al., 1997), AtGSLOH (Hansen et al., 2008), and E8/Sl27DOX (Akiyama et al., 2021), from maize, rosy periwinkle (*Catharanthus roseus*), *Arabidopsis*, and tomato, respectively, catalyze the hydroxylation of DIBOA glucoside, desacetoxyvindoline, 3-butenyl glucosinolate, and lycoperoside C. Additionally, SmTIIAS (Song et al., 2021) from redroot sage (*Salvia miltiorrhiza*), catalyzing the dehydrogenation of dihydrofuran to furan, is also included. Fabacyl acetate has solely been identified in pea, whereas DDHs including medicaol exhibit a broader distribution within Fabaceae plants (Yoneyama et al., 2008; Tokunaga et al., 2015; Galili et al., 2021). Thus, the unique diversification pathway catalyzed by PsFOS is suggested to have emerged through the evolution of DOXC31 in different lineages of DOXC55 and DOXC54, which are presumed to be involved in SL biosynthesis.

In Solanaceae and Fabaceae plants, orobanchol-type SLs are produced, but the rearrangement of the A-ring proceeds differently using different oxidation enzymes. As previously described, SlCYP712G1 has been identified as the enzyme responsible for converting orobanchol to DDHs in tomato. The DDHs catalyzed by SlCYP712G1 likely include 6,7-didehydroorobanchol, phelipanchol, and epiphelipanchol, as indicated by their chromatographic behavior compared to authentic SLs we identified (Wakabayashi et al., 2022; Wang et al., 2022). The elimination of the hydroxy group at the C7 position of 7-hydroxyorobanchol is presumed to be the key reaction, supported by the detection of putative 7-hydroxyorobanchol as a reaction product of SlCYP712G1 (Wakabayashi et al., 2022; Wang et al., 2022). Similarly, detailed analysis of a reaction product by MtMOS indicated a trace amount of putative monohydroxyorobanchol (Supplementary Figure S12A). The product could be an intermediate hydroxylated at the C9 or C10 position of orobanchol in the following reaction mechanism. A primary carbocation formed by the elimination of the hydroxy group could yield medicaol with the seven-membered A-ring through Wagner-Meerwein rearrangement followed by deprotonation, as previously proposed (Tokunaga et al., 2015) (Supplementary Figure S12B). However, recent studies have indicated radical-based ring expansion in the biosynthesis of tropolone (Doyon et al., 2020) and brevione (Yan and Matsuda, 2022), both catalyzed by DOX enzymes. Radical formation at the C9 or C10 position of orobanchol catalyzed by MtMOS may initiate the A-ring expansion (Supplementary Figure S12C). Based on this mechanism, the putative monohydroxyorobanchol detected in the enzyme reaction (Supplementary Figure S12A) may be a byproduct resulting from the rebound hydroxylation at the C9 or C10 position (Supplementary Figure S12C). Further studies are needed to gain more insight into this intriguing ring expansion mechanism.

In pea plants, orobanchol is presumed to be metabolized by two pathways: one involves the conversion to fabacyl acetate via fabacol, and the other involves the conversion to orobanchyl acetate (Supplementary Figure S5A). Considering the high affinity of PsFOS for orobanchol (Figure 6B, Table 1A) and the substrate preference of PsFAT for fabacol over orobanchol (Table 1B), the conversion of orobanchol to fabacyl acetate via fabacol is presumed to be the primary pathway of orobanchol derivatization in pea. This is consistent with the composition of SLs in pea root exudates supplemented with orobanchol, where the level of fabacyl acetate was much higher than orobanchyl acetate (Supplementary Figure S5C). The accumulation of a significant amount of fabacol in hydroponics supplemented with orobanchol (Supplementary Figure S5B) may be due to the exceedance of PsFAT's conversion capacity toward an excess amount of fabacol formed by PsFOS from exogenously applied orobanchol. It is noteworthy that fabacol was identified for the first time in this work, whereas its occurrence was expected when fabacyl acetate was identified (Xie et al., 2009b), demonstrating the importance of substrate feeding experiments not only in elucidating the biosynthetic pathway but also in identifying biosynthetic precursors that are below the detection limit under normal cultural conditions.

Orobanchyl acetate has been identified in numerous Fabaceae plants (Yoneyama et al., 2008). Additionally, other SLs possessing an acetoxy group at the C4 position, such as fabacyl acetate and lotuslactone, have been isolated in pea and birdsfoot trefoil, respectively. Indeed, homologs of PsFAT and VuOAT are widely conserved in Fabaceae plants, and they can be potentially involved in the acetylation of the hydroxy group at the C4 position of SLs in Fabaceae plants. Alternatively, orobanchol-type SLs with an acetoxy group at the C4 position have been detected in plants outside the Fabaceae family. In Solanaceae plants, orobancyl acetate was detected in tomato and tobacco (*Nicotiana tabacum*) (Kohlen et al., 2013; Wang and Bouwmeester, 2018). Furthermore, 7-α/β-hydroxyorobanchyl acetate and 7-oxoorobanchyl acetate were isolated in cucumber (Khetkam et al., 2014) and flax (Xie et al., 2009a), respectively. However, tomato, tobacco, cucumber, and flax do not have homologs of PsFAT and VuOAT; BAHD acyltransferases in these plants share only ≤40% amino acid identity. Further studies are needed to identify the enzymes responsible for acetylation of the hydroxy group at the C4 position of orobanchol and its derivatives in plant species other than Fabaceae.

In conclusion, this study provides a comprehensive view of the biosynthesis pathway of canonical SLs downstream of orobanchol in Fabaceae plants (Figure 1) and offers a molecular basis to understand the structural diversity of SLs. Canonical SLs are suggested to primarily function in the rhizosphere as signaling molecules (Ito et al., 2022), and the structural diversification of orobanchol may play roles in plant-plant and plant-microbe interactions in the rhizosphere. The diversified orobanchol-type SLs may modulate the rhizosphere microbiome by selectively enriching specific bacteria beneficial to the secreting plants (Schlemper et al., 2017; Kim et al., 2022). The identification of genes responsible for orobanchol derivatization will enable genetic modification of SL composition in plants belonging to not only Fabaceae but also other plant families due to the wide distribution of orobanchol in plants. Studies on the rhizosphere microorganism

profile using such plants with unique SL spectra will provide insight into the physiological significance of each of the diversified SLs. Results obtained from such studies can lead to agricultural applications for optimizing the rhizosphere environment.

Data availability statement

The data in the study are included in the main article and supplementary material. The accession numbers for CDS of MtMOS, PsFOS, PsFAT, and VuOAT are LC800624, LC800625, LC800626, and LC800627, respectively.

Author contributions

MH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. KU: Formal Analysis, Investigation, Validation, Writing – review & editing. TW: Formal analysis, Writing – review & editing. MM: Writing – review & editing. HT: Project administration, Supervision, Writing – review & editing. YS: Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the JSPS KAKENHI (22K05463).

Acknowledgments

MH acknowledges Dr. Nanami Shiotani, The University of Tokyo, for useful discussion on epoxidation of orobanchol and orobanchyl acetate.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1392212/full#supplementary-material

References

Abe, S., Sado, A., Tanaka, K., Kisugi, T., Asami, K., Ota, S., et al. (2014). Carlactone is converted to carlactonoic acid by MAX1 in *Arabidopsis* and its methyl ester can directly interact with AtD14 in *vitro*. *Proc. Natl. Acad. Sci.* 111, 18084–18089. doi: 10.1073/pnas.1410801111

Akiyama, K., Matsuzaki, K., and Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435, 824–827. doi: 10.1038/nature03608

Akiyama, K., Ogasawara, S., Ito, S., and Hayashi, H. (2010). Structural requirements of strigolactones for hyphal branching in AM fungi. *Plant Cell Physiol.* 51, 1104–1117. doi: 10.1093/pcp/pcq058

Akiyama, R., Nakayasu, M., Umemoto, N., Kato, J., Kobayashi, M., Lee, H. J., et al. (2021). Tomato E8 encodes a C-27 hydroxylase in metabolic detoxification of α-tomatine during fruit ripening. *Plant Cell Physiol.* 62, 775–783. doi: 10.1093/pcp/pcab080

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

Alves-Carvalho, S., Aubert, G., Carrere, S., Cruaud, C., Brochot, A. L., Jacquin, F., et al. (2015). Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 84, 1–19. doi: 10.1111/tpj.12967

Aquino, B., Bradley, J. M., and Lumba, S. (2021). On the outside looking in: roles of endogenous and exogenous strigolactones. *Plant J.* 105, 322–334. doi: 10.1111/tpj.15087

Brewer, P. B., Yoneyama, K., Filardo, F., Meyers, E., Scaffidi, A., Frickey, T., et al. (2016). Lateral branching oxidoreductase acts in the final stages of strigolactone biosynthesis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A* 113, 6301–6306. doi: 10.1073/pnas.1601729113

Chesterfield, R. J., Vickers, C. E., and Beveridge, C. A. (2020). Translation of strigolactones from plant hormone to agriculture: achievements, future perspectives, and challenges. *Trends Plant Sci.* 25, 1087–1106. doi: 10.1016/j.tplants.2020.06.005

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

Doyon, T. J., Skinner, K., Yang, D., Mallik, L., Wymore, T., Koutmos, M., et al. (2020). Radical tropolone biosynthesis. *ChemRxiv*. doi: 10.26434/chemrxiv.12780044.v1

Foo, E., Yoneyama, K., Hugill, C. J., Quittenden, L. J., and Reid, J. B. (2013). Strigolactones and the regulation of pea symbioses in response to nitrate and phosphate deficiency. *Mol. Plant* 6, 76–87. doi: 10.1093/mp/sss115

Galili, S., Hershenhorn, J., Smirnov, E., Yoneyama, K., Xie, X., Amir-Segev, O., et al. (2021). Characterization of a chickpea mutant resistant to *Phelipanche aEgyptiaca* Pers. and *Orobanche crenata* Forsk. *Plants* 10, 2552. doi: 10.3390/plants10122552

Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pagès, V., Dun, E. A., Pillot, J.-P., et al. (2008). Strigolactone inhibition of shoot branching. *Nature* 455, 189–194. doi: 10.1038/nature07271

Guercio, A. M., Torabi, S., Cornu, D., Dalmais, M., Bendahmane, A., Le Signor, C., et al. (2022). Structural and functional analyses explain pea KAI2 receptor diversity and reveal stereoselective catalysis during signal perception. *Commun. Biol.* 5, 126. doi: 10.1038/s42003-022-03085-6

Hansen, B. G., Kerwin, R. E., Ober, J. A., Lambrix, V. M., Mitchell-Olds, T., Gershenzon, J., et al. (2008). A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. *Plant Physiol.* 148, 2096–2108. doi: 10.1104/pp.108.129981

Homma, M., Wakabayashi, T., Moriwaki, Y., Shiotani, N., Shigeta, T., Isobe, K., et al. (2023). Insights into stereoselective ring formation in canonical strigolactone:

Discovery of a dirigent domain-containing enzyme catalyzing orobanchol synthesis. *bioRxiv*. doi: 10.1101/2023.08.07.552212

Iseki, M., Shida, K., Kuwabara, K., Wakabayashi, T., Mizutani, M., Takikawa, H., et al. (2018). Evidence for species-dependent biosynthetic pathways for converting carlactone to strigolactones in plants. *J. Exp. Bot.* 69, 2305–2318. doi: 10.1093/jxb/erx428

Ito, S., Braguy, J., Wang, J. Y., Yoda, A., Fiorilli, V., Takahashi, I., et al. (2022). Canonical strigolactones are not the major determinant of tillering but important rhizospheric signals in rice. *Sci. Adv.* 8, eadd1278. doi: 10.1126/sciadv.add1278

Jonczyk, R., Schmidt, H., Osterrieder, A., Fiesselmann, A., Schullehner, K., Haslbeck, M., et al. (2008). Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. *Plant Physiol.* 146, 1053–1063. doi: 10.1104/pp.107.111237

Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479

Khetkam, P., Xie, X., Kisugi, T., Kim, H. I., Yoneyama, K., Uchida, K., et al. (2014). 7α- and 7β-Hydroxyorobanchyl acetate as germination stimulants for root parasitic weeds produced by cucumber. *J. Pestic. Sci.* 39, 121–126. doi: 10.1584/jpestics.D14-038

Kim, B., Westerhuis, J. A., Smilde, A. K., Floková, K., Suleiman, A. K. A., Kuramae, E. E., et al. (2022). Effect of strigolactones on recruitment of the rice root-associated microbiome. *FEMS Microbiol. Ecol.* 98, 1–12. doi: 10.1093/femsec/fiac010

Kohlen, W., Charnikhova, T., Bours, R., López-Ráez, J. A., and Bouwmeester, H. (2013). Tomato strigolactones: A more detailed look. *Plant Signal. Behav.* 8, e22785. doi: 10.4161/psb.22785

Kreplak, J., Madoui, M. A., Cápal, P., Novák, P., Labadie, K., Aubert, G., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 51, 1411–1422. doi: 10.1038/s41588-019-0480-1

Li, C., Dong, L., Durairaj, J., Guan, J. C., Yoshimura, M., Quinodoz, P., et al. (2023). Maize resistance to witchweed through changes in strigolactone biosynthesis. *Science* 379, 94–99. doi: 10.1126/science.abq4775

Liu, W., Kohlen, W., Lillo, A., Op den Camp, R., Ivanov, S., Hartog, M., et al. (2011). Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. *Plant Cell* 23, 3853–3865. doi: 10.1105/tpc.111.089771

López-Ráez, J. A., Charnikhova, T., Gómez-Roldán, V., Matusova, R., Kohlen, W., de Vos, R., et al. (2008). Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. *New Phytol.* 178, 863–874. doi: 10.1111/j.1469-8137.2008.02406.x

Mori, N., Nomura, T., and Akiyama, K. (2020). Identification of two oxygenase genes involved in the respective biosynthetic pathways of canonical and non-canonical strigolactones in *Lotus japonicus*. *Planta* 251, 40. doi: 10.1007/s00425-019-03332-x

Moriyama, D., Wakabayashi, T., Shiotani, N., Yamamoto, S., Furusato, Y., Yabe, K., et al. (2022). Identification of 6-*epi*-heliolactone as a biosynthetic precursor of avenaol in *Avena strigosa*. *Biosci. Biotechnol. Biochem.* 86, 998–1003. doi: 10.1093/bbb/zbac069

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998). Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 1694–1699. doi: 10.1128/AEM.64.5.1694-1699.1998

Nomura, S., Nakashima, H., Mizutani, M., Takikawa, H., and Sugimoto, Y. (2013). Structural requirements of strigolactones for germination induction and inhibition of *Striga gesnerioides* seeds. *Plant Cell Rep.* 32, 829–838. doi: 10.1007/s00299-013-1429-y

Ogata, Y., Sakurai, N., Suzuki, H., Aoki, K., Saito, K., and Shibata, D. (2009). The prediction of local modular structures in a co-expression network based on gene expression datasets. *Genome Inform.* 23, 117–127. doi: 10.1142/9781848165632_0011

Samejima, H., and Sugimoto, Y. (2018). Recent research progress in combatting root parasitic weeds. *Biotechnol. Biotechnol. Equip.* 32, 221–240. doi: 10.1080/13102818.2017.1420427

Schlemper, T. R., Leite, M. F. A., Lucheta, A. R., Shimels, M., Bouwmeester, H. J., van Veen, J. A., et al. (2017). Rhizobacterial community structure differences among sorghum cultivars in different growth stages and soils. *FEMS Microbiol. Ecol.* 93, 1–11. doi: 10.1093/FEMSEC/FIX096

Song, J. J., Fang, X., Li, C. Y., Jiang, Y., Li, J. X., Wu, S., et al. (2021). A 2-oxoglutarate dependent dioxygenase converts dihydrofuran to furan in *Salvia* diterpenoids. *Plant Physiol.* 188, 1496–1506. doi: 10.1093/plphys/kiab567

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

Tokunaga, T., Hayashi, H., and Akiyama, K. (2015). Medicaol, a strigolactone identified as a putative didehydro-orobanchol isomer, from *Medicago truncatula*. *Phytochemistry* 111, 91–97. doi: 10.1016/j.phytochem.2014.12.024

Tuominen, L. K., Johnson, V. E., and Tsai, C. J. (2011). Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics* 12, 236. doi: 10.1186/1471-2164-12-236

Uchida, K., Ogura, Y., Okamura, H., Sugimoto, Y., and Takikawa, H. (2023). Novel and efficient stereoselective synthesis of (±)-orobanchol, a representative canonical strigolactone, based on acid-mediated cascade cyclization. *Tetrahedron Lett.* 120, 154454. doi: 10.1016/j.tetlet.2023.154454

Ueno, K., Nomura, S., Muranaka, S., Mizutani, M., Takikawa, H., and Sugimoto, Y. (2011). *Ent-2′-epi*-orobanchol and its acetate, as germination stimulants for *Striga gesnerioides* seeds isolated from cowpea and red clover. *J. Agric. Food Chem.* 59, 10485–10490. doi: 10.1021/jf2024193

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455, 195–200. doi: 10.1038/nature07272

Vazquez-Flota, F., De Carolis, E., Alarco, A.-M., and De Luca, V. (1997). Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in *Catharanthus roseus* (L.) G. Don. *Plant Mol. Biol.* 34, 935–948. doi: 10.1023/A:1005894001516

Wakabayashi, T., Hamana, M., Mori, A., Akiyama, R., Ueno, K., Osakabe, K., et al. (2019). Direct conversion of carlactonoic acid to orobanchol by cytochrome P450 CYP722C in strigolactone biosynthesis. *Sci. Adv.* 5, eaax9067. doi: 10.1126/sciadv.aax9067

Wakabayashi, T., Ishiwa, S., Shida, K., Motonami, N., Suzuki, H., Takikawa, H., et al. (2021). Identification and characterization of sorgomol synthase in sorghum strigolactone biosynthesis. *Plant Physiol.* 85, 902–913. doi: 10.1093/plphys/kiaa113

Wakabayashi, T., Moriyama, D., Miyamoto, A., Okamura, H., Shiotani, N., Shimizu, N., et al. (2022). Identification of novel canonical strigolactones produced by tomato. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1064378

Wakabayashi, T., Nakayama, M., Kitano, Y., Homma, M., Miura, K., Takikawa, H., et al. (2023). Discovery of strigol synthase from cotton (*Gossypium hirsutum*): The enzyme behind the first identified germination stimulant for *Striga*. *Plants People Planet.*, 1–6. doi: 10.1002/ppp3.10420,2023/08/02

Wang, Y., and Bouwmeester, H. J. (2018). Structural diversity in the strigolactones. *J. Exp. Bot.* 69, 2219–2230. doi: 10.1093/jxb/ery091

Wang, Y., Durairaj, J., Suárez Duran, H. G., van Velzen, R., Flokova, K., Liao, C.-Y., et al. (2022). The tomato cytochrome P450 CYP712G1 catalyses the double oxidation of orobanchol en route to the rhizosphere signalling strigolactone, solanacol. *New Phytol.* 235, 1884–1899. doi: 10.1111/nph.18272

Xie, X., Mori, N., Yoneyama, K., Nomura, T., Uchida, K., Yoneyama, K., et al. (2019). Lotuslactone, a non-canonical strigolactone from *Lotus japonicus*. *Phytochemistry.* 157, 200–205. doi: 10.1016/j.phytochem.2018.10.034

Xie, X., Yoneyama, K., Harada, Y., Fusegi, N., Yamada, Y., Ito, S., et al. (2009b). Fabacyl acetate, a germination stimulant for root parasitic plants from *Pisum sativum*. *Phytochemistry.* 70, 211–215. doi: 10.1016/j.phytochem.2008.12.013

Xie, X., Yoneyama, K., Kurita, J., Harada, Y., Yamada, Y., Takeuchi, Y., et al. (2009a). 7-Oxoorobanchyl acetate and 7-oxoorobanchol as germination stimulants for root parasitic plants from flax (*Linum usitatissimum*). *Biosci. Biotechnol. Biochem.* 73, 1367–1370. doi: 10.1271/bbb.90021

Yan, D., and Matsuda, Y. (2022). Biosynthetic elucidation and structural revision of brevione E: Characterization of the key dioxygenase for pathway branching from setosusin biosynthesis. *Angew. Chem. Int. Ed* 61, e202210938. doi: 10.1002/anie.202210938

Yoda, A., Xie, X., Yoneyama, K., Miura, K., McErlean, C. S. P., and Nomura, T. (2023). A stereoselective strigolactone biosynthesis catalyzed by a 2-oxoglutarate-dependent dioxygenase in sorghum. *Plant Cell Physiol.* 64, 1034–1045, pcad060. doi: 10.1093/pcp/pcad060

Yoneyama, K., Mori, N., Sato, T., Yoda, A., Xie, X., Okamoto, M., et al. (2018). Conversion of carlactone to carlactonoic acid is a conserved function of MAX1 homologs in strigolactone biosynthesis. *New Phytol.* 218, 1522–1533. doi: 10.1111/nph.15055

Yoneyama, K., Xie, X., Sekimoto, H., Takeuchi, Y., Ogasawara, S., Akiyama, K., et al. (2008). Strigolactones, host recognition signals for root parasitic plants and arbuscular mycorrhizal fungi, from Fabaceae plants. *New Phytol.* 179, 484–494. doi: 10.1111/j.1469-8137.2008.02462.x

Yoneyama, K., Yoneyama, K., Takeuchi, Y., and Sekimoto, H. (2007). Phosphorus deficiency in red clover promotes exudation of orobanchol, the signal for mycorrhizal symbionts and germination stimulant for root parasites. *Planta* 225, 1031–1038. doi: 10.1007/s00425-006-0410-1

Yuan, Z., Yang, H., Pan, L., Zhao, W., Liang, L., Gatera, A., et al. (2022). Systematic identification and expression profiles of the BAHD superfamily acyltransferases in barley (*Hordeum vulgare*). *Sci. Rep.* 12, 5063. doi: 10.1038/s41598-022-08983-7

Zhang, Y., van Dijk, A. D. J., Scaffidi, A., Flematti, G. R., Hofmann, M., Charnikhova, T., et al. (2014). Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. *Nat. Chem. Biol.* 10, 1028–1033. doi: 10.1038/nchembio.1660

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 May 2024
DOI 10.3389/fpls.2024.1374912

OPEN ACCESS

EDITED BY
Naoki Kitaoka,
Hokkaido University, Japan

REVIEWED BY
Takafumi Shimizu,
Research Institute of Innovative Technology for the Earth (RITE), Japan
Sibongile Mafu,
University of Massachusetts Amherst, United States

*CORRESPONDENCE
Shanfa Lu
sflu@implad.ac.cn
Yuhang Chen
chenyuhang221@126.com

†These authors have contributed equally to this work

RECEIVED 23 January 2024
ACCEPTED 15 April 2024
PUBLISHED 01 May 2024

CITATION
Yan C, Li C, Jiang M, Xu Y, Zhang S, Hu X, Chen Y and Lu S (2024) Systematic characterization of gene families and functional analysis of *PvRAS3* and *PvRAS4* involved in rosmarinic acid biosynthesis in *Prunella vulgaris*. *Front. Plant Sci.* 15:1374912. doi: 10.3389/fpls.2024.1374912

Systematic characterization of gene families and functional analysis of *PvRAS3* and *PvRAS4* involved in rosmarinic acid biosynthesis in *Prunella vulgaris*

Chao Yan[1,2,3†], Caili Li[1,2†], Maochang Jiang[1,2], Yayun Xu[1,2], Sixuan Zhang[1,2], Xiangling Hu[1,2,3], Yuhang Chen[3*] and Shanfa Lu[1,2*]

[1]State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, [2]Engineering Research Center of Chinese Medicine Resource, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, [3]College of Pharmaceutical Sciences, Chengdu Medical College, Chengdu, China

Prunella vulgaris is an important material for Chinese medicines with rosmarinic acid (RA) as its index component. Based on the chromosome-level genome assembly we obtained recently, 51 RA biosynthesis-related genes were identified. Sequence feature, gene expression pattern and phylogenetic relationship analyses showed that 17 of them could be involved in RA biosynthesis. *In vitro* enzymatic assay showed that PvRAS3 catalyzed the condensation of *p*-coumaroyl-CoA and caffeoyl-CoA with pHPL and DHPL. Its affinity toward *p*-coumaroyl-CoA was higher than caffeoyl-CoA. PvRAS4 catalyzed the condensation of *p*-coumaroyl-CoA with pHPL and DHPL. Its affinity toward *p*-coumaroyl-CoA was lower than PvRAS3. UPLC and LC-MS/MS analyses showed the existence of RA, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid in *P. vulgaris*. Generation and analysis of *pvras3* homozygous mutants showed significant decrease of RA, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid and significant increase of DHPL and pHPL. It suggests that PvRAS3 is the main enzyme catalyzing the condensation of acyl donors and acceptors during RA biosynthesis. The role of PvRAS4 appears minor. The results provide significant information for quality control of *P. vulgaris* medicinal materials.

KEYWORDS

biosynthetic pathway, CRISPR/Cas9, *in vitro* enzymatic activity assay, *Prunella vulgaris*, PvRAS3, PvRAS4, rosmarinic acid, rosmarinic acid synthase

Introduction

Prunella vulgaris L. is a perennial medicinal plant of Lamiaceae, which is widely distributed in Asia, North America, Europe and North Africa (National Pharmacopoeia Committee, 2020; Hu et al., 2023). The whole plants and spikes of *P. vulgaris* are commonly used to treat thyroiditis, mastitis, tuberculosis, infectious hepatitis and hypertension in East Asia, the Middle East, and Europe (Tang et al., 2023). In addition, *P. vulgaris* spikes are used as the main raw materials of functional herbal tea in the southern provinces of China. Its fresh leaves are used as seasonal vegetables in southeastern China. The whole plants are often used as urban landscape plants for urban greening (Chen et al., 2019). The demand for *P. vulgaris* in the production of Chinese patented medicines and functional herbal tea is approximately 60 million kilograms per year (Chen et al., 2022; Li et al., 2022).

P. vulgaris is rich in polyphenols, of which rosmarinic acid (RA) is an index component in evaluating the quality of *P. vulgaris* medicinal materials and Chinese patented medicines. As the main polyphenol component produced in *P. vulgaris*, RA has a variety of pharmacological activities, such as antioxidant, anti-inflammatory, anti-tumor, anti-allergy, anti-depression, and anti-anxiety (Taguchi et al., 2017). It also has unique pharmacological effects in improving sleep, neurological prevention, reducing testicular injury and inhibiting elastin degradation (Kwon et al., 2017), and has obvious inhibitory effect on liver tumor cells, lung tumor cells and stomach tumor cells (Radziejewska et al., 2018). In addition, RA is easily absorbed and no toxic side effects on blood cells, kidney, and liver (Noguchi-Shinohara et al., 2015).

RA is a depside condensed from two single phenolic acids (Scarpati and Oriente, 1958). One of them is derived from the general phenylpropanoid pathway. It serves as the acyl donor during condensation. The other one is come from the tyrosine-derived pathway. It serves as the acyl acceptor (Figure 1). RA is present in some hornworts, ferns, and multiple taxa of flowering plants and its biosynthetic pathways are probably evolved independently in differently species (Petersen et al., 2009; Levsh et al., 2019). Analysis of RA biosynthesis in *Coleus blumei*, *Phacelia campanularia* and *Salvia miltiorrhiza* showed the existence of three proposed RA biosynthetic routes in different plants (Figure 1), which include the biosynthesis of 4-coumaroyl-4'-hydroxyphenyllactic acid from *p*-coumaroyl-CoA and *p*-hydroxyphenyllactic acid (pHPL) in *C. blumei* and *P. campanularia* (route 1), the biosynthesis of 4-coumaroyl-3',4'-dihydroxyphenyllactic acid from *p*-coumaroyl-CoA and 3,4-dihydroxyphenyllactic acid (danshensu, DHPL) in *S. miltiorrhiza* (route 2), and the biosynthesis of caffeoyl-4'-hydroxyphenyllactic acid from caffeoyl-CoA and pHPL in *S. miltiorrhiza* (route 3) (Eberle et al., 2009; Petersen et al., 2009; Di et al., 2013; Levsh et al., 2019; Liu et al., 2019; Lu, 2021). The biosynthetic routes of RA in *P. vulgaris* are largely unknown.

Recently, we sequenced and assembled the genome of *P. vulgaris*, which provide a solid foundation for analyzing RA biosynthetic routes in *P. vulgaris* (Zhang et al., 2024). In this study, a total of 51 P. *vulgaris* genes belonging to seven RA biosynthesis-related gene families were systematically studied through genome-wide identification, feature analysis, expression analysis, and phylogenetic analysis. Among them, seventeen were identified as candidate genes for RA biosynthesis. *In vitro* enzymatic assay of PvRAS3 and PvRAS4, *in vivo* phenolic acid compound determination and *PvRAS3* transgenic analysis showed that PvRAS3 was the main enzyme catalyzing the condensation of acyl donors and acceptors during RA biosynthesis, whereas PvRAS4 played a minor role.

Materials and methods

Plant materials and growth conditions

A wild and whole genome sequenced *Prunella vulgaris* L. line, named Bangshan-XKC, was transplanted from Bangshan village, Shunchang county of Fujian Province of China and grown in a greenhouse at the Institute of Medicinal Plant Development in Beijing of China. Shoots were cut from the plant and surface-sterilized using 75% ethanol for 1 min and 5% sodium hypochlorite for 20 min. Subsequently, the shoots were rinsed three times with sterile water and inserted into MS medium supplemented with 30 g L^{-1} sucrose with pH value adjusted to 5.8. After two weeks, the resulting sterile plantlets were transferred to a fresh MS medium. To induce rooting, the apical and axillary buds were cut and placed on 1/2 MS medium containing 0.1 mg L^{-1} indole-3-butyric acid (IBA). The sterile plantlets were sub-cultivated in a tissue culture room on 1/2 MS medium supplemented with 30 g L^{-1} sucrose under a 16/8 h light/dark photoperiod at 25°C.

Sequence retrieval and gene prediction

The deduced amino acid sequences of RA biosynthesis-related genes from *S. miltiorrhiza* were downloaded from NCBI GenBank (https://www.ncbi.nlm.nih.gov/protein) (Wang et al., 2015). BLAST analysis of the downloaded proteins against the chromosome-level assembly of *P. vulgaris* was carried out using the tBLASTn algorithm (Altschul et al., 1997; Zhang et al., 2024). An *E*-value cut-off of 10^{-5} was applied to the homologue recognition. Gene models were predicted from the retrieved *P. vulgaris* genomic DNA sequences based on the downloaded *S. miltiorrhiza* genes and through BLASTx analysis of retrieved sequences against the NR database using the default parameters (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The predicted gene models were further examined and corrected manually through BLASTn analysis against *P. vulgaris* transcriptome sequencing data (Altschul et al., 1997; Zhang et al., 2024).

Gene and protein feature analysis

The theoretical isoelectric point (p*I*) and molecular weight (Mw) were calculated using the Compute pI/MW tool on the ExPASy server (https://web.expasy.org/compute_pi/). Protein subcellular

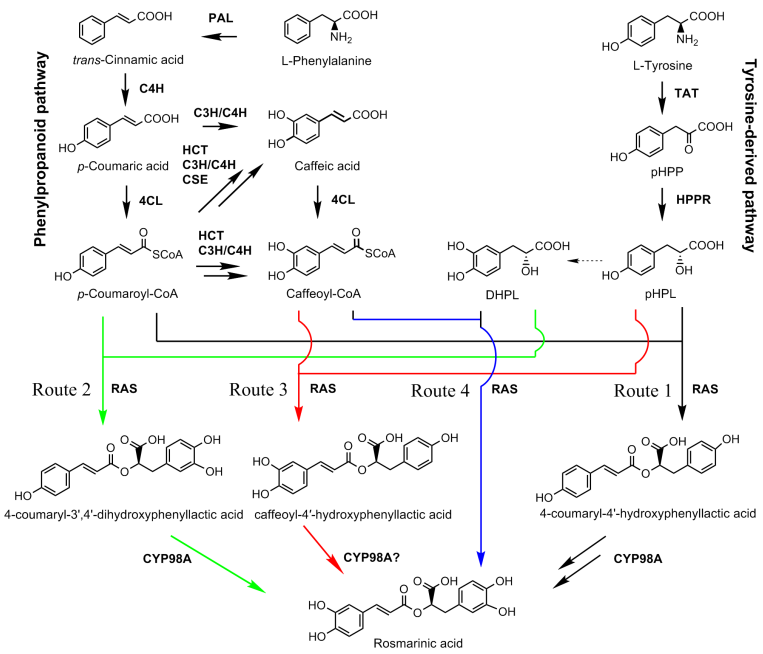


FIGURE 1
The proposed RA biosynthetic pathways (Updated from Lu, 2021). Solid arrow represents single biosynthetic step. Two arrows represent two or more steps. Dashed arrow indicates the enzyme involved in the reaction is unknown. Four proposed biosynthetic routes of RA are shown in black, green, red, and blue, respectively. Route 1 was found in *C. blumei* and *P. campanularia* (Eberle et al., 2009; Petersen et al., 2009; Levsh et al., 2019). Routes 2 and 3 were found in *S. miltiorrhiza* (Di et al., 2013; Liu et al., 2019). Route 4 was proposed in this study. 4CL, 4-coumaroyl CoA ligase; C3H, *p*-coumaroyl shikimate 3'-hydroxylase/coumarate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; CSE, caffeoyl shikimate esterase; DHPL, 3,4-dihydroxyphenyllactate (danshensu); HCT, *p*-hydroxycinnamoyl-CoA: shikimate *p*-hydroxycinnamoyltransferase; HPPR, p-hydroxyphenylpyruvate reductase; PAL, phenylalanine ammonia lyase; pHPL, *p*-hydroxyphenyllactic acid; pHPP, *p*-hydroxyphenylpyruvic acid; RAS, RA synthase; TAT, tyrosine aminotransferase.

localization was predicted using Plant-mPLoc version 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#). The number of transmembrane regions was predicted using DeepTMHMM version 1.0.24 (https://dtu.biolib.com/DeepTMHMM). Distribution of genes on the chromosomes of *P. vulgaris* was visualized using TBtools (Chen et al., 2020). Intron/exon structures were predicted using GSDS2.0 on the Gene Structure Display Server (http://gsds.gao-lab.org/).

Phylogenetic analysis

RA biosynthesis-related protein sequences from various plant species were downloaded from NCBI GenBank (https://www.ncbi.nlm.nih.gov/protein). Sequence alignment was carried out using the ClustalW algorithms in MEGA version 7.0.26 (Kumar et al., 2016). Neighbor-joining trees were constructed for amino acid sequences using MEGA versopm 7.0.26 with default parameters (Kumar et al., 2016). The number of bootstrap replications was 1000.

Quantitative real-time PCR analysis of gene expression

Total RNA was extracted from roots, stems, leaves and spikes using the EASYspin Plus Complex Plant RNA kit (Aidlab, China) as described previously (Cui et al., 2022). Genomic DNA contamination was eliminated by treating with RNase-free DNase (Aidlab, China). RNA integrity was evaluated on a 1% argarose gel. RNA quantity was determined using a NanoDrop 2000C Spectrophotometer (Thermo Scientific, USA). RNA was reverse-transcribed into single-stranded cDNA using Superscript III Reverse Transcriptase (Invitrogen, USA). qRT-PCR was carried out using TB Green Premix Ex Taq II (Takara, Japan) on a Bio-Rad CFX96 Real-Time system. Primers used for qRT-PCR were designed using Primer Premier 5 (Lalitha, 2000) and are shown in Supplementary Table 1. Gene amplification efficiency of each primer pair was evaluated using the standard curves. Primer pairs with an appropriate PCR amplification efficiency were used for subsequent analysis (Supplementary Figure 1). *PveIF-2* was selected as the reference gene as described before (Zheng et al., 2022). The specificity of amplification was assessed by dissociation curve analysis. Relative abundance of transcripts was determined using the $2^{-\Delta\Delta Ct}$ method. Standard deviations were calculated from three biological replicates and three PCR replicates per biological replicates.

Analysis of gene expression using RNA-seq data

Gene expression was analyzed using the published transcriptome data from roots, stems, leaves, and spikes (Zhang et al., 2024).

Salmon software (v1.10.3) (Sahraeian et al., 2017) was used to quantify the level of gene expression. Heat maps were constructed using the TBtools software (Chen et al., 2020).

PvRAS3 and *PvRAS4* gene cloning and expression vector construction

Total RNA extracted from leaves of *P. vulgaris* was reverse-transcribed into cDNA using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, USA). *PvRAS3* and *PvRAS4* were amplified by nested PCR using cDNA from *P. vulgaris* leaves as the template. The nesting and nested primers used for PCR are listed in Supplementary Table 1. PCR products were inserted into pGEX-4T-1 and verified by Sanger sequencing.

Heterologous expression of PvRAS3 and PvRAS4 proteins in *E. coli*

The pGEX-4T-1 vector with *PvRAS3* or *PvRAS4* was introduced into *E. coli* strain BL21 (DE3). Heterologous expression of PvRAS3 and PvRAS4 proteins were induced with 0.5 mmol L^{-1} IPTG at 16 °C for 20–24 h. Cells were collected through centrifugation at 6,000 rpm for 10 min at 4 °C. After resuspension in 10 mM PBS buffer (pH 7.2), the cells were sonicated on ice. Purification of soluble proteins was carried out using the PurKine™ GST-Tag Protein Purification kit (Glutathione) (Abbkine, China). Concentration of the purified proteins was determined using the BCA Protein Assay kit (Takara Biomedical Technology, Beijing).

In vitro enzymatic activity assay of PvRAS3 and PvRAS4 recombinant proteins

The enzymatic activity assay was carried out in a 500 µl reaction system comprising 100 µg purified proteins, 1 mM caffeoyl-CoA or *p*-coumaroyl-CoA as the acyl donors, 1 mM pHPL or DHPL as the acyl receptors. The reactions were incubated at 25 °C for 60 min and terminated by adding 10 µl of 10 M acetic acid. Controls were carried out using total proteins from *E. coli* transformed with the empty pGEX-4T-1 vector. Reaction products were collected and analyzed using ACQUITY UPLC system (Waters, Milford, MA, USA). MS/MS data were recorded on a Xevo G2-XS Q-ToF Mass Spectrometer (Waters, Milford, MA, USA) coupled to a Waters Acquity I-Class UPLC system (Waters, Milford, MA, USA). MS/MS analyses were conducted in negative-ion mode. The samples were separated on an ACQUITY UPLC BEH C18 column (1.7 μm, 100×2.1 mm) at 25°C. The mobile phase A was 0.1% (v/v) formic acid-acetonitrile. The mobile phase B was 0.1% (v/v) formic acid in water. The flow rate was 0.3 mL min^{-1}. The mobile phases changed with the following gradient: 0–6 min, 5% A and 95% B; 6–8 min, 20% A and 80% B; 8–14 min, 21% A and 79% B; 14–18 min, 95% A and 5% B. MS was analyzed using electrospray ionization (ESI) at negative ion mode. MS-MS data were analyzed using the MssLynx V4.1 software (Waters) as described previously (Pan et al., 2023).

Kinetic analysis of PvRAS3 and PvRAS4 recombinant proteins

Kinetic analysis of PvRAS3 and PvRAS4 was carried out in a 200 µL reaction system consisting of Tris-HCl buffer (100 mM Tris-HCl, pH 7.0, 2 mM DTT, 4 mM $MgCl_2$, 10% glycerol), 100 µg recombinant protein, and different concentrations of substrates. The reactions were incubated at 25 °C for 30 min and terminated by adding 10 μl of 10 M acetic acid. The reaction products were analyzed using UPLC system as described as *in vitro* enzymatic activity assay of recombinant proteins. Enzyme activity was determined by measuring the variation of substrate contents. To determine kinetic parameters, PvRAS3 or PvRAS4 was incubated with different concentrations of acyl donor and acyl acceptor. The saturation concentration of one substrate was set at 2 mM, while the concentration of another substrate was varied at different levels, including 10 µM, 50 µM, 100 µM, 250 µM, 350 µM, 500 µM, and 1000 µM, respectively. The kinetic constants of the donor substrates were calculated based on contents of the product. The kinetic constants of the acceptor substrates were determined through monitoring the consumption of the acceptor substrates. Enzyme assays were performed in triplicate at each concentration of substrate. *V*max and *K*m values were calculated using Origin 8.0 software with nonlinear regression analysis.

UPLC and LC-MS/MS analyses of phenolic acids

Roots, stems, leaves and spikes of two-year-old *P. vulgaris* were ground in liquid nitrogen. The ground samples (0.5 g) were dissolved in 10 ml of 80% ethanol and sonicated for 60 min. The extracts were collected by centrifugation and filtered using a 0.22 µm filter (Merk Millipore, USA). UPLC and LC-MS/MS analyses were performed using the ACQUITY UPLC I-Class system (Waters) as described as *in vitro* enzymatic activity assay of recombinant proteins. Three biological and three technological replicates were carried out for analysis of each tissue.

Generation and analysis of *pvras3* mutants

Pvras3 mutants of *P. vulgaris* hairy roots were generated using the CRISPR/Cas9 system described previously (Wang et al., 2022a). Briefly, PCR amplification was carried out using two pairs of primers containing two dividual guide RNAs (sgRNAs) sequences of *PvRAS3*. pDT1T2 vector was used as a template. The products were purified, digested with *Bsa* I, and ligated into the binary vector pHEE401E. The resulting constructs were transferred into *Agrobacterium* strain ATCC15834.

Leaf discs from thirty-day-old sterile plantlets were cultivated on 1/2 MS medium in dark for two days, immersed for 10 min in the suspension of *Agrobacterium* cells with or without the constructs, and co-cultivated on MS medium for 2 days. The leaf discs were then transferred onto 1/2 MS medium supplemented with 30 mg L^{-1} of hygromycin and 400 mg L^{-1} of cefotaxime for

generation of hairy roots. Leaf discs were subcultured every two weeks. Hairy roots generated were transferred to 1/2 MS medium supplemented with 200 mg L^{-1} of cefotaxime and cultivated for about two weeks. Newly generated hairy roots were then transferred to 1/2 MS medium supplemented with 100 mg L^{-1} of cefotaxime and cultivated for about two weeks. Finally, newly generated hairy roots from medium with 100 mg L^{-1} of cefotaxime were transferred to 1/2 MS medium without cefotaxime and cultivated for two weeks. Root tips with 3-4 cm in length were cut, transferred to 100 ml of 1/2 MS medium in 250 ml-flasks, and cultivated at 25°C in dark with 100 rpm shaking.

To analyze the mutations of *PvRAS3* in transgenic hairy roots, genomic DNA was extracted. DNA fragments around the target site were PCR-amplified using gene-specific primers, Mut-F: GTCGTTTGCTCCCTTACAAAT, and Mut-R: GATCGAAGTGA AGGAGTCGACG. PCR products were sequenced using the primer Mut-F. Hairy roots generated from leaf discs through inoculation with *Agrobactrium* without the constructs were used as a control. UPLC analysis of chemical compounds was performed using the ACQUITY UPLC I-Class system (Waters). Three biological and three technological replicates were carried out for analysis of each transgenic hairy root line.

Results and discussion

Genome-wide identification of genes associated with RA biosynthesis in *P. vulgaris*

RA is synthesized through the general phenylpropanoid pathway and the tyrosine-derived pathway, involving at least nine enzymes encoded by seven gene families (Figure 1) (Deng and Lu, 2017; Lu, 2021). In order to identify *P. vulgaris* genes involved in RA biosynthesis, tblastn analysis of the deduced protein sequences of RA biosynthesis-related genes in *Salvia miltiorrhiza* against the whole-genome assembly of *P. vulgaris* (2n=28) was carried out (Altschul et al., 1997; Wang et al., 2015; Zhang et al., 2024). It resulted in the identification of 51 full-length candidate genes, including four putative *PvPALs*, three putative *PvC4Hs*, seventeen putative *Pv4CLs*, seven putative *PvTATs*, four putative *PvHPPRs*, three putative *PvHCTs*, eight putative *PvRASs*, and five putative *PvCYP98As* (Table 1). These genes were designated as *PvPAL1–PvPAL4*, *PvC4H1–PvC4H3*, *Pv4CL1–Pv4CL17*, *PvTAT1–PvTAT7*, *PvHPPR1–PvHPPR4*, *PvHCT1–PvHCT3*, *PvRAS1–PvRAS8*, and *PvCYP98A-1–PvCYP98A-5*, respectively. Among them, *PvPAL1* and *PvC4H1* showed 99% identities at the amino acid level with the reported *PvPAL* (KJ010815) and *PvC4H* (KJ010816), respectively (Kim et al., 2014). *Pv4CL3* showed 97% identities at the amino acid level with the reported *Pv4CL1* (KJ010814), whereas *Pv4CL8* showed 99% amino acid identities with the reported 5' truncated *Pv4CL2* (KJ010817) (Kim et al., 2014). *PvTAT3*, PvRAS3 and PvCYP98A-1 showed 99% identities at the amino acid level with the reported *PvTAT* (M053278), *PvRAS* (KM053280) and *PvCYP98A101* (AJW87635), respectively (Ru et al., 2017a; Ru et al., 2017b). *PvHPPR3* was identical to the sequence under the GenBank accession number KM053279. The other 43 genes have not been reported previously.

The identified genes are distributed on the 14 chromosomes of the whole genome assembly of *P. vulgaris* (Supplementary Figure 2) (Zhang et al., 2024). The deduced proteins have length varying from 309 to 709 amino acid residues, p*I* varying from 5.29 to 9.22, and molecular weight varying from 34.00 to 77.01 (Table 1). All of them do not contain transmembrane regions and were predicted to be localized in the cytoplasm, endoplasmic reticulum, peroxisome, chloroplast and mitochondrion, respectively (Table 1). The predicted localization of PvPALs, PvHCTs and PvRASs in the cytoplasm is consistent with the experimental results from tobacco, *S. miltiorrhiza* (Achnine et al., 2004; Di et al., 2013). The predicted localization of PvC4Hs and PvCTP98As in the endoplasmic reticulum is consistent with the experimental results from *Populus*, *S. miltiorrhiza*, and *P. vulgaris* (Ro et al., 2001; Di et al., 2013; Ru et al., 2017b). Pv4CLs, PvTATs and PvHPPRs were predicted to be localized in the peroxisome, chloroplast and mitochondrion, respectively. However, *Peucedanum praeruptorum* 4CL, *P. vulgaris* TAT and *S. miltiorrhiza* HPPR were previously found to be located in the cytoplasm (Liu et al., 2017; Wang et al., 2017; Ru et al., 2017a). Thus, the actual subcellular localization of Pv4CLs, PvTATs and PvHPPRs remain to be experimentally analyzed.

Characterization and expression analysis of genes involved in the general phenylpropanoid pathway

The general phenylpropanoid pathway involves three enzymes, including phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumaroyl CoA ligase (4CL) (Figure 1). PAL catalyzes the conversion of L-phenylalanine to *trans*-cinnamic acid through deamination of L-phenylalamine. It is the first reaction of the general phenylpropanoid pathway and a rate limiting step mediating the influx from primary metabolism into the general phenylpropanoid pathway (Raes et al., 2003). In a plant, PAL is usually encoded by a small gene family. For instance, there are two *PAL* genes in tobacco, three in *S. miltiorrhiza*, and four in *Arabidopsis* (Raes et al., 2003; Achnine et al., 2004; Wang et al., 2015). Genome-wide analysis showed that there were four putative *PvPAL* genes in *P. vulgaris* (Table 1), all of which contained an intron and had similar gene structures (Figure 2A). qRT-PCR analysis showed that the four *PvPAL* genes were differentially expressed (Figure 2B). RNA-seq analysis showed that *PvPAL1* and *PvPAL2* were expressed relatively higher than the other two *PvPALs* (Figure 3A). The expression patterns revealed between qRT-PCR and RNA-seq were largely consistent for *PvPALs* and other genes analyzed hereinafter. However, discrepancy was also observed, which could be results from the difference of detection technologies, plant tissues, data analysis method, or other unknown factors. Analysis of the deduced PvPAL proteins showed that all of them contained the conserved "GTITASGDLVPLSYIA" motif involved in substrate binding and catalytic activity and the conserved "FL" residues impartent for substrate specificity (Poppe and Rétey, 2005; Watts et al., 2006; Ma et al., 2013). In addition,

10.3389/fpls.2024.1374912

TABLE 1 Sequence features of RA biosynthesis-related genes in *P. vulgaris*.

Gene name	Gene Length (bp)	ORF length (bp)[1]	Protein length (aa)	p*I* [2]	Mw (kDa)[3]	Loc[4]	TMR[5]
PvPAL1	2,699	2,130	710	6.06	76.62	Cyt	0
PvPAL2	2,676	2,127	709	5.84	77.01	Cyt	0
PvPAL3	3,224	2,121	707	6.04	76.87	Cyt	0
PvPAL4	2,732	2,127	709	5.9	76.83	Cyt	0
PvC4H1	3,081	1,515	505	9.22	57.94	ER	0
PvC4H2	3,062	1,515	505	9.17	57.84	ER	0
PvC4H3	2,200	1,509	503	9.21	57.02	ER	0
Pv4CL1	2,866	1,620	540	5.5	58.62	Per	0
Pv4CL2	3,719	1,620	540	6.62	58.44	Per	0
Pv4CL3	3,115	1,695	565	5.48	60.78	Per	0
Pv4CL4	4,608	1,671	557	5.95	60.04	Per	0
Pv4CL5	4,069	1,680	560	6.33	60.3	Per	0
Pv4CL6	3,650	1,695	565	5.42	60.7	Per	0
Pv4CL7	4,143	1,647	549	5.95	59.73	Per	0
Pv4CL8	3,820	1,632	544	5.84	59.02	Per	0
Pv4CL9	1,962	1,695	565	5.87	61.55	Per	0
Pv4CL10	3,338	1,635	545	8.75	59.46	Per	0
Pv4CL11	3,773	1,662	554	8.39	59.99	Per	0
Pv4CL12	2,488	1,722	574	7.28	62.33	Per	0
Pv4CL13	4,674	1,605	535	8.59	58.67	Per	0
Pv4CL14	1,930	1,692	564	5.82	61.8	Per	0
Pv4CL15	2,806	1,722	574	6.77	62.51	Per	0
Pv4CL16	3,290	1,620	540	7.24	59.16	Per	0
Pv4CL17	3,430	1,608	536	6.97	58.73	Per	0
PvTAT1	1,759	1,206	402	7.06	44.33	Chl	0
PvTAT2	2,707	1,263	421	5.88	45.95	Chl	0
PvTAT3	2,627	1,233	411	5.8	45.11	Chl	0
PvTAT4	2,421	1,233	411	5.8	45.05	Chl	0
PvTAT5	1,997	1,308	435	6.23	48.49	Chl	0
PvTAT6	1,904	1,251	417	7.61	46	Chl	0
PvTAT7	2,498	1,281	427	5.77	47.18	Chl	0
PvHPPR1	1,818	927	309	5.29	34.04	M	0
PvHPPR2	2,026	939	313	5.99	33.99	M	0
PvHPPR3	1,716	939	313	5.66	34	M	0
PvHPPR4	1,459	957	319	7.74	35.09	M	0
PvHCT1	1,640	1,281	427	5.63	47.12	Cyt	0
PvHCT2	3,296	1,281	427	5.73	47.33	Cyt	0
PvHCT3	6,909	1,281	427	6.47	47.28	Cyt	0
PvRAS1	1,528	1,299	433	6.51	48.14	Cyt	0

(Continued)

TABLE 1 Continued

Gene name	Gene Length (bp)	ORF length (bp)[1]	Protein length (aa)	pI [2]	Mw (kDa)[3]	Loc[4]	TMR[5]
PvRAS2	1,411	1,326	442	6.64	49.22	Cyt	0
PvRAS3	2,635	1,305	435	5.97	48.17	Cyt	0
PvRAS4	1,281	1,278	426	6.14	47.41	Cyt	0
PvRAS5	1,599	1,305	435	6.94	48.37	Cyt	0
PvRAS6	3,031	1,293	431	6.47	48.08	Cyt	0
PvRAS7	4,937	1,293	431	6.15	48.13	Cyt	0
PvRAS8	3,331	1,308	436	6.08	48.71	Cyt	0
PvCYP98A-1	2,357	1,530	510	8.62	57.76	ER	0
PvCYP98A-2	2,304	1,533	511	9.12	58.23	ER	0
PvCYP98A-3	1,811	1,533	511	8.61	57.86	ER	0
PvCYP98A-4	2,490	1,530	510	8.13	57.68	ER	0
PvCYP98A-5	2,703	1,530	510	8.13	57.84	ER	0

[1]ORF stands for the open reading frame of a gene;
[2,3]pI and molecular weight (Mw) were calculated using the Compute pI/MW tool on the ExPASy server (https://web.expasy.org/compute_pi/);
[4]Loc represents protein subcellular localization predicted using Plant-mPLoc version 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#).' Cyt', 'ER', 'Per', 'Chl' and 'M'stand for cytoplasm, endoplasmic reticulum, peroxisome, chloroplast and mitochondrion, respectively.
[5]TMR represents the number of transmembrane regions predicted using DeepTMHMM version 1.0.24 (https://dtu.biolib.com/DeepTMHMM).

other three conserved catalytic active sites, including "GLALVNG", "NDN" and "HNQD", were also found (Supplementary Figure 3) (He et al., 2020). It indicates that all of the four identified PvPALs have catalytic activity. Phylogenetic analysis of PALs from *P. vulgaris*, *S. miltiorrhiza*, *Arabidospsis*, *Populus trichocarpa* and various other plant showed that PvPAL1, PvPAL2 and PvPAL4 were grouped with SmPAL1, SmPAL3 and MoPAL involved in RA biosynthesis (Figure 2C) (Weitzel and Petersen, 2010; Song and Wang, 2011; Hou et al., 2013). Taken together with previous results for PvPAL1 (Kim et al., 2014), the presence of RA in roots, stems, leaves and flowers (Kim et al., 2014), and the results from gene expression analysis, we speculated that PvPAL1 and PvPAL2 could be the main PvPALs for RA biosynthesis.

C4H catalyzes the hydroxylation of *trans*-cinnamic acid to *p*-coumaric acid (Figure 1). It is encoded by the members of CYP73A gene subfamily. Genome-wide analysis showed that there were three putative *PvC4H* genes in *P. pulgaris*. All of them contained two introns (Figure 2D). *PvC4H1* and *PvC4H2* showed similar expression patterns with the highest expression in roots, followed by spikes, stems and leaves (Figures 2E, 3A). High expression of *PvC4H1* and *PvC4H2* is consistent with high content of RA in roots, stems, leaves and flowers (Kim et al., 2014). The expression of *PvC4H3* was very low in four tissues analyzed (Figure 3A). Analysis of the deduced proteins showed that all of the three PvC4Hs contained five conserved P450 motifs, including the proline-rich motif "PPGP", the oxygen binding motif "AAIETT", the "ETLR" motif, the "PERF" motif, and the heme-binding motif "FGVGRRSCPG" (Supplementary Figure 4) (Khatri et al., 2023). Phylogenetic analysis of C4Hs from *P. vulgaris*, *S. miltiorrhiza*, *Arabidospsis*, *P. trichocarpa* and various other plants showed that PvC4H1 and PvC4H2 were grouped with SmC4H1 involved in RA biosynthesis (Figure 2F), indicating the involvement of PvC4H1 and PvC4H2 in RA biosynthesis (Xiao et al., 2011; Kim et al., 2014; Wang et al., 2015).

4CL is the third and the last enzyme of the general phenylpropanoid pathway. It catalyzes the thioesterification of *p*-coumaric acid (Figure 1). The product, *p*-coumaroyl-CoA, can be funneled into downstream branch pathways for lingnins, flavonoids, coumarins, lignans, and RA (Deng and Lu, 2017). 4CL is encoded by a multiple gene family. For instance, there are seventeen *Pt4CLs* in *P. trichocarpa*, ten *Sm4CLs* in *S. miltiorrhiza*, and thirteen *At4CLs* and *At4CL-likes* in *Arabidopsis* (Raes et al., 2003; Shi et al., 2010; Wang et al., 2015). Genome-wide analysis showed that there were seventeen putative *Pv4CL* genes with 3–5 introns in *P. pulgaris* (Figure 4A). Gene expression analysis showed that the seventeen *Pv4CLs* had differential expression patterns (Figure 4B). RNA-seq analysis showed that the levels of *Pv4CL1*, *Pv4CL3*, *Pv4CL8*, *Pv4CL10* and *Pv4CL11* were relatively high, whereas the levels of *Pv4CL9* and *Pv4CL12–Pv4CL17* were very low in the four tissues analyzed (Figure 3B).

It is generally known that the 4CL proteins contain two conserved motifs, including Box I with the representative sequence "SSGTTGLPKGV" and Box II with the representative sequence "GEICIRG" (Uhlmann and Ebel, 1993). Box I is conserved in adenylate-forming enzymes and involved in adenosine monophosphate (AMP)-binding. Box II is conserved in 4CL and related to the spatial conformation of the enzyme (Wang et al., 2022b). Sequence alignment of the seventeen Pv4CL proteins showed that Pv4CLs also had the two conserved motifs. However, their sequences were divergent (Supplementary Figure 5). It indicates that the identified seventeen Pv4CLs could be functionally diverse. Phylogenetic analysis of 4CLs from *P. vulgaris*, *S. miltiorrhiza*, *Arabidospsis*, rice and various other plants showed that Pv4CL1, Pv4CL3, Pv4CL6, Pv4CL7 and Pv4CL8 were grouped with Mo4CL1,

10.3389/fpls.2024.1374912

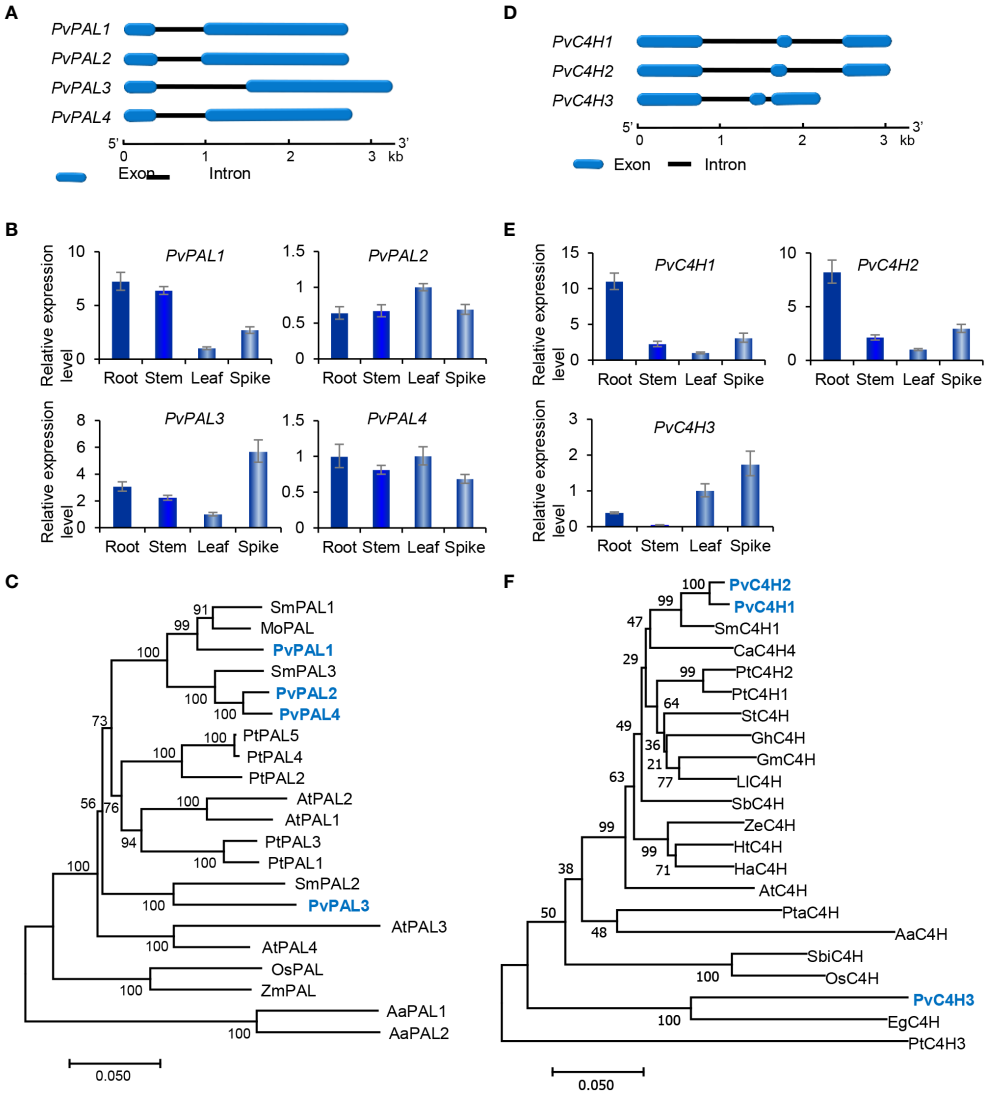


FIGURE 2
Gene structures, expression patterns and phylogenetic analysis of *PvPAL* and *PvC4H* genes and their deduced proteins. **(A, D)** The intron-exon structures of *PvPAL* **(A)** and *PvC4H* **(D)** genes. **(B, E)** Fold changes of *PvPAL* **(B)** and *PvC4H* **(E)** gene expression in roots, stems, leaves and spikes of *P. vulgaris* plants. The expression level in leaves was arbitrarily set to 1, respectively. **(C)** Phylogenetic analysis of PAL proteins. The rooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. AaPAL1 (QPI70499.1) and AaPAL2 (SPO49995.1) from *Anthoceros agrestis* were used as outgroup. Ingroup consists of four PvPALs and the PALs from *S. miltiorrhiza* (Sm), *Arabidopsis* (At), *Populus trichocarpa* (Pt), *Melissa officinalis* (Mo), and maize (Zm) (Supplementary Table 2). **(F)** Phylogenetic analysis of C4H proteins. C4Hs included are three PvC4Hs and the C4Hs from *S. miltiorrhiza* (Sm), *Coffea Arabica* (Ca), *P. trichocarpa* (Pt), *Solanum tuberosum* (St), *Gossypium hirsutum* (Gh), *Glycine max* (Gm), *Leucaena leucocephala* (Ll), *Scutellaria baicalensis* (Sb), *Zinnia elegans* (Ze), *Helianthus tuberosus* (Ht), *Helichrysum aureonitens* (Ha), *Arabidopsis* (At), *Pinus taeda* (Pt), *A. agrestis* (Aa), *Sorghum bicolor* (Sb), rice (Os), and *Erythranthe guttata* (Eg) (Supplementary Table 2).

Sm4CL2 and Sm4CL3 involved in RA biosynthesis (Figure 4C) (Zhao et al., 2006; Weitzel and Petersen, 2010; Wang et al., 2015). These Pv4CLs could be associated with RA biosynthesis in *P. vulgaris* (Kim et al., 2014).

Characterization and expression analysis of genes involved in the tyrosine-derived pathway

The tyrosine-derived pathway involves two enzymes, including tyrosine aminotransferase (TAT) and *p*-hydroxyphenylpyruvate reductase (HPPR) (Figure 1). TAT catalyzes the conversion of L-tyrosine to pHPP, a precursor for the biosynthesis of various secondary metabolites, such as plastoquinone, tocopheros, benzylisoquinoline alkaloids, and RA. In plants, TAT is encoded by multiple gene family with three members in *S. miltiorrhiza*, four in apple, and eight in *Arabidopsis* (Wang et al., 2015, 2018). Through genone-wide mining, we identified seven putative *PvTAT* genes that had five or six introns (Figure 5A). Among them, *PvTAT2–PvTAT3* showed relatively high expression (Figures 3C, 5B). The expression of other four *PvTATs*, including *PvTAT1* and *PvTAT5–PvTAT7*, was mainly in spikes and the level was very low in the tissues analyzed, in comparison with *PvTAT2–*

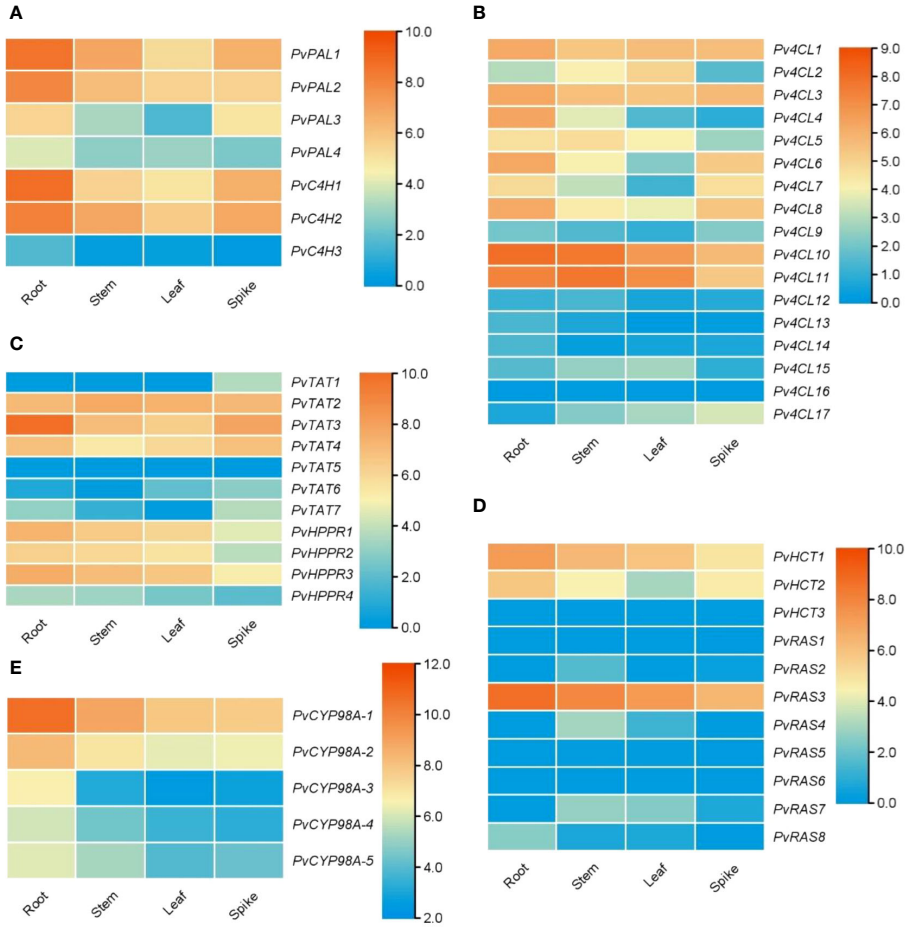


FIGURE 3
Expression of RA biosynthesis-related genes in roots, stems, leaves and spikes of *P. vulgaris*. **(A–E)** Hierarchical clustering of the expression levels of RA biosynthesis-related genes analyzed using RNA-seq clean data.

PvTAT3 (Figures 3C, 5B). Sequence alignment of the seven PvTAT proteins showed that all of them contained the conserved Motif 1 for aminotransferase family-I pyridoxal phosphate binding site and Motif 2 with the highly conserved residue Arg (Supplementary Figure 6) (Lu et al., 2013; Wang et al., 2018). Phylogenetic analysis of seven PvTATs and TATs from *S. miltiorrhiza*, *Arabidopsis* and other plants showed that PvTAT3 and PvTAT4 were clustered with SmTAT1 and PfTAT involved in RA biosynthesis (Figure 5C) (Xiao et al., 2011; Lu et al., 2013). PvTAT3 was previously shown to participate in the biosynthesis of RA in *P. vulgaris* and its high expression in four tissues analyzed is consistent with the accumulation of RA (Figure 5B) (Kim et al., 2014; Ru et al., 2017a). PvTAT4 could be a novel PvTAT playing a redundant role with PvTAT3 in RA biosynthesis.

HPPR, belonging to the family of D-isomer-specific 2-hydroxyacid dehydrogenases, is the other enzyme involved in the tyrosine-derived pathway. It catalyzes the conversion of pHPP to pHPL (Kim et al., 2004) (Figure 1). HPPR is encoded by a small gene family, such ast there are three *SmHPPR* genes in *S. miltiorrhiza* and four in *Arabidopsis* (Wang et al., 2015; Xu et al., 2018). Its involvement in RA biosynthesis has been verified through functional analysis of *MoHPPR* from *Melissa offcianalis* (Mansouri and Mohammadi, 2021), *CsHPPR* from *Coleus scutellarioides* (Kim et al., 2004), and *SmHPPR1* from *S. miltiorrhiza* (Xiao et al., 2011; Wang et al., 2017). Genome-wide analysis showed that there were four *PvHPPR* genes in *P. vulgaris*, all of which contained one intron and had similar structures (Figure 5D). All of them showed differential expression patterns, and the overall expression level of *PvHPPR1–PvHPPR3* was higher than *PvHPPR4* (Figures 3C,5E). Amino acid sequence alignment showed that all four PvHPPR proteins contained the NAD(P)H binding motif with the representative sequence "GLGRIG" and the putative myristylation site with the representative sequences "GTVETR" and "GNLEA" (Supplementary Figure 7) (Wang et al., 2017). Phylogenetic analysis of four PvHPPRs and HPPRs from *S. miltiorrhiza*, *Arabidopsis* and other plants showed that PvHPPR1 and PvHPPR3 were grouped with SmHPPR1, MoHPPR and CsHPPR involved in RA biosynthesis (Figure 5F). It suggests that PvHPPR1 and PvHPPR3 could be involved in the biosynthesis of RA in *P. vulgaris*. The function of PvHPPR2 and PvHPPR4 remain to be elucidated.

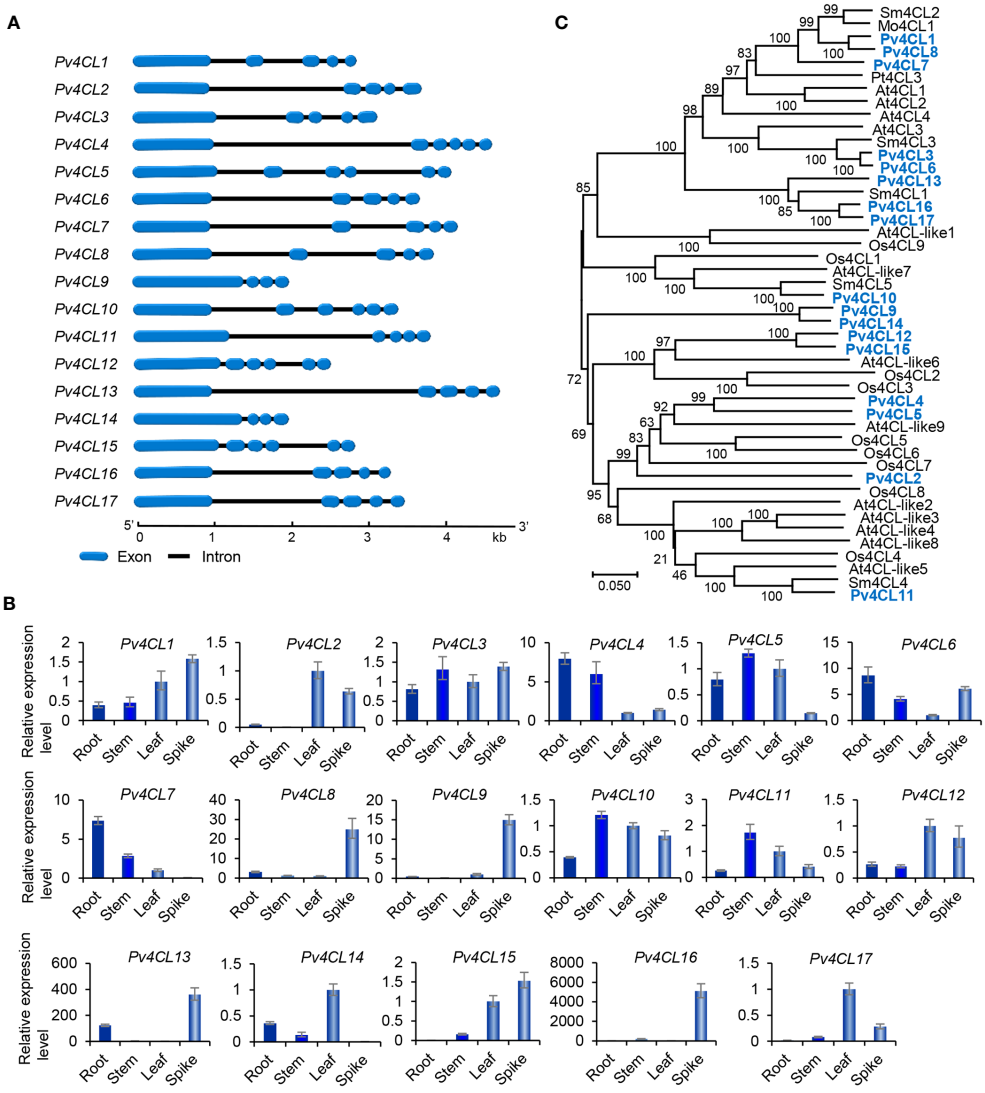


FIGURE 4
Gene structures, expression patterns and phylogenetic analysis of *Pv4CL* genes and their deduced proteins. **(A)** The intron-exon structures of *Pv4CL* genes. **(B)** Fold changes of *Pv4CL* gene expression in roots, stems, leaves and spikes of *P. vulgaris* plants. The expression level in leaves was arbitrarily set to 1, respectively. **(C)** Phylogenetic analysis of 4CL proteins. The unrooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. 4CLs included are seventeen Pv4CLs and other 4CLs from *S. miltiorrhiza* (Sm), *Arabidopsis* (At), rice (Os), *M. officinalis* (Mo), and *P. trichocarpa* (Pt) (Supplementary Table 2).

Characterization and expression analysis of genes involved in downstream of the RA biosynthetic pathway

The downstream of RA biosynthetic pathway involves four known enzymes, including *p*-hydroxycinnamoyl-CoA: shikimate *p*-hydroxycinnamoyltransferase (HCT), rosmarinic acid synthase (RAS), *p*-coumaroyl shikimate 3'-hydroxylase/coumarate 3-hydroxylase (C3H), and the enzyme catalyzing the final step of RA biosynthetic pathway (Figure 1). Enzymes involved in DHPL biosynthesis are currently unknown (Figure 1). Among the four known enzymes, HCT catalyzes the coupling of *p*-coumaroyl-CoA with shikimate to form *p*-coumaroyl shikimic acid. It also catalyzes the reverse reaction converting caffeoyl shikimate ester to caffeoyl-CoA (Hoffmann et al., 2003). Differently, RAS catalyzes the coupling of *p*-coumaroyl-CoA/caffeoyl-CoA and pHPL/DHPL to form 4-coumaroyl-4'-hydroxyphenyllactic acid, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, and/or caffeoyl-4'-hydroxyphenyllactic acid in different plant species (Figure 1) (Eberle et al., 2009; Petersen et al., 2009; Di et al., 2013; Levsh et al., 2019; Liu et al., 2019; Lu, 2021). Both HCT and RAS belong to the BAHD acyltransferase family and are known as CoA-ester-dependent BAHD hydroxycinnamoyltransferases (Petersen et al., 2009).

Genome-wide analysis showed that there were three *PvHCT* and eight *PvRAS* genes in *P. vulgaris* (Table 1). Except that *PvRAS4* had no intron and *PvRAS8* had two introns, other three *PvHCTs* and six *PvRASs* contained an intron and shared similar gene structures (Figure 6A). qRT-PCR analysis showed that the three *PvHCTs* showed differential expression (Figure 6B). RNA-seq analysis showed that *PvHCT1* had the highest expression,

10.3389/fpls.2024.1374912

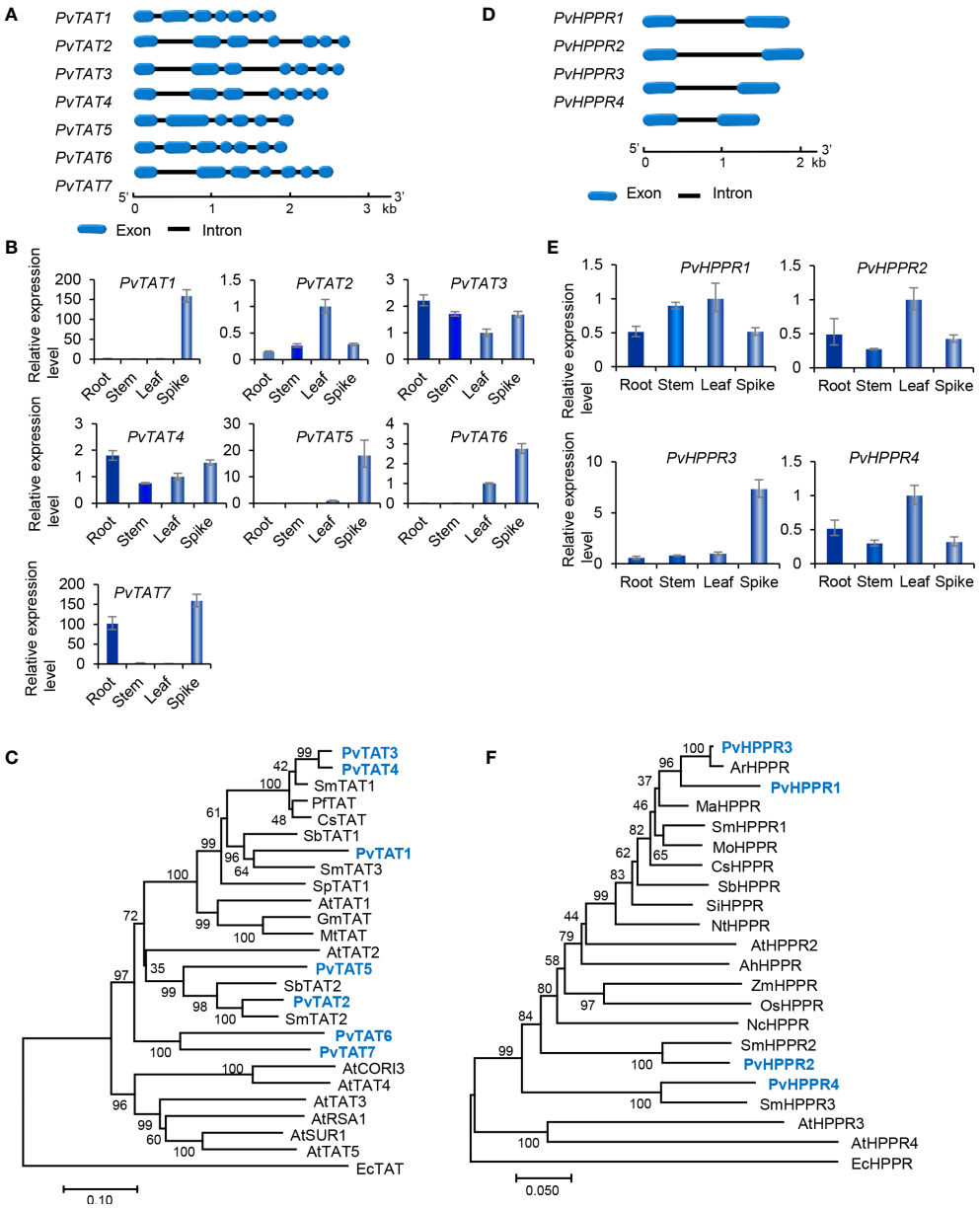


FIGURE 5
Gene structures, expression patterns and phylogenetic analysis of *PvTAT* and *PvHPPR* genes and their deduced proteins. **(A, D)** The intron-exon structures of *PvTAT* **(A)** and *PvHPPR* **(D)** genes. **(B, E)** Fold changes of *PvTAT* **(B)** and *PvHPPR* **(E)** gene expression in roots, stems, leaves and spikes of *P. vulgaris* plants. The expression level in leaves was arbitrarily set to 1, respectively. **(C)** Phylogenetic analysis of TAT proteins. The rooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. EcTAT (NP_418478.1) from *Escherichia coli* was used as outgroup. Ingroup TATs include PvTAT1–PvTAT7 and others from *S. miltiorrhiza* (Sm), *Arabidopsis* (At), *Perilla frutescens* (Pf), *Coleus scutellarioides* (Cs), *S. baicalensis* (Sb), *Solanum pennellii* (Sp), *G. max* (Gm), and *Medicago truncatula* (Mt) (Supplementary Table 2). **(F)** Phylogenetic analysis of HPPR proteins. The rooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. EcHPPR (WP_000811015.1) from *Escherichia coli* was used as outgroup. Ingroup HPPRs include PvHPPR1–PvHPPR4 and others from *S. miltiorrhiza* (Sm), *Arabidopsis* (At), *Agastache rugosa* (Ar), *Mentha aquatica* (Ma), *M. officinalis* (Mo), *C. scutellarioides* (Cs), *S. baicalensis* (Sb), *Sesamum indicum* (Si), *Nicotiana tomentosiformis* (Nt), *Arachis hypogaea* (Ah), *Zea mays* (Zm), *Oryza sativa* (Os), and *Nymphaea colorata* (Nc) (Supplementary Table 2).

followed by *PvHCT2* (Figure 3D). The expression of *PvHCT3* was very low (Figure 3D). It indicates that, among the three *PvHCTs*, *PvHCT1* could be most likely to be involved in RA biosynthesis in *P. vulgaris*. Among the eight *PvRASs*, *PvRAS3* showed the highest expression and was highly expressed in spikes, followed by stems, leaves, and roots (Figures 3D, 6B). The expression of other seven *PvRASs* was relatively low in the tissues analyzed (Figure 3D).

Sequence alignment of PvHCT and PvRAS proteins showed that all of them contained the conserved "HXXXD" and "DFGWG" motifs (Supplementary Figure 8) (Berger et al., 2006). Phylogenetic analysis of three PvHCTs, eight PvRASs, and HCTs and RASs from various other plants showed that RASs and HCTs separated into two clades (Figure 6C). All HCTs were clustered in one clade, whereas all RASs were clustered in the other one. In addition, the RAS clade

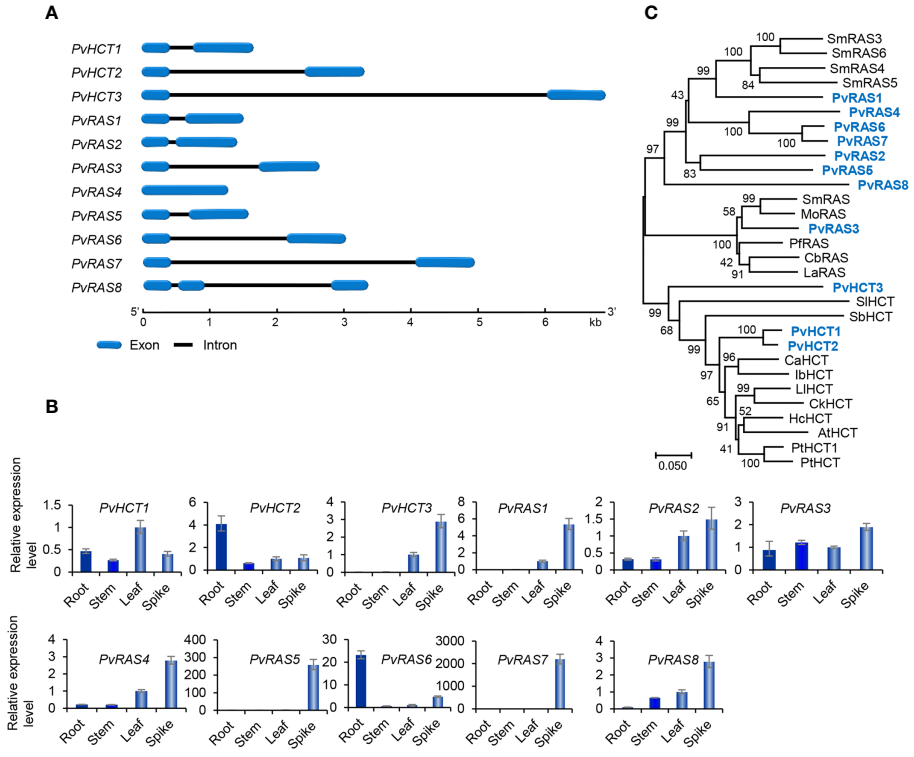


FIGURE 6
Gene structures, expression patterns and phylogenetic analysis of *PvHCT* and *PvRAS* genes and their deduced proteins. **(A)** The intron-exon structures of three *PvHCT* and *eight PvRAS* genes. **(B)** Fold changes of *PvHCT* and *PvRAS* gene expression in roots, stems, leaves and spikes of *P. vulgaris* plants. The expression level in leaves was arbitrarily set to 1, respectively. **(C)** Phylogenetic analysis of PvHCT and PvRAS proteins. The unrooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. Proteins included are PvHCT1–PvHCT3, PvRAS1–PvRAS8, and the HCTs and RASs from *S. miltiorrhiza* (Sm), *M. officinalis* (Mo), *P. frutescens* (Pf), *C. blumei* (Cb), *L. angustifolia* (La), *Solanum lycopersicum* (Sl), *S. bicolor* (Sb), *C. arabica* (Ca), *Ipomoea batatas* (Ib), *L. leucocephala* (Ll), *Caragana korshinskii* (Ck), *Hibiscus cannabinus* (Hc), *Arabidopsis* (At), *P. trichocarpa* (Pt) (Supplementary Table 2).

could be divided into two sub-clades. PvRAS3 was clustered with functionally known RASs from *C. blumei*, *Melissa officinalis*, *Lavandula Angustifolia* and *S. miltiorrhiza* in a sub-clade (Berger et al., 2006; Landmann et al., 2011; Sander and Petersen, 2011; Weitzel and Petersen, 2011; Di et al., 2013; Zhou et al., 2018; Fu et al., 2020). Taken together with the high expression of *PvRAS3* gene (Figure 6B), the results suggest the involvement of PvRAS3 in RA biosynthesis. The other RAS sub-clade included PvRAS1, PvRAS2, PvRAS4–PvRAS8, and four putative SmRASs (Figure 5C). The function of these RASs is currently unknown. Among them, the expression of *PvRAS2*, *PvRAS4* and *PvRAS8* showed similar patterns with RA distribution (Kim et al., 2014) (Figure 6B). It indicates that these *PvRASs* could also be associated with RA biosynthesis.

C3H is the other enzyme involved in downstream of the RA biosynthetic pathway (Figure 1). It catalyzes the hydroxylation of *p*-coumaroyl shikimic acid, a shikimate ester of *p*-coumarate generated from *p*-coumaroyl-CoA and shikimate under the catalysis of HCT, into caffeoyl shikimic acid, a shikimate ester of caffeic acid (Schoch et al., 2001). C3H is a cytochrome P450 encoded by members of the *CYP98* gene family (Schoch et al., 2001; Franke et al., 2002; Ralph et al., 2006). Similarily, the enzyme involved in the final step of RA biosynthetic pathway is also a cytochrome P450 encoded by members of the *CYP98* gene family. It introduces the hydroxyl group(s) to the products of RAS (Eberle et al., 2009; Di et al., 2013; Levsh et al., 2019; Liu et al., 2019; Fu et al., 2020) (Figure 1).

Genome-wide analysis showed that there were five *PvCYP98* genes in *P. vulgaris* (Table 1). All of them contain two introns and share similar gene structures (Figure 7A). They showed differential expression patterns and *PvCYP98A-1* and *PvCYP98A-2* had relatively high expression among the five *PvCYP98s* (Figures 3E, 7B). Similar to PvC4Hs, the other family of cytochrome P450 proteins involved in RA biosynthesis (Supplementary Figure 4), PvCYP98 proteins also contain the five conserved P450 motifs, including "PPGP", "(A/G)(A/G)X(D/E)T(T/S)", "EXLR", "PERF", and "FGXGRRXCXG" (Supplementary Figure 9) (Eberle et al., 2009; Khatri et al., 2023). Phylogenetic analysis showed that PvCYP98A-4 and PvCYP98A-5 were clustered with the functionally known C3Hs (Figure 7C), such as AtC3H1 (AtCYP98A3) and PtC3H3 (Schoch et al., 2001; Chen et al., 2011). PvCYP98A-1 and PvCYP98A-2 were clustered with CsCYP98A14 and SmCYP98A78 involved in the last step of RA biosynthetic pathway (Figure 7C) (Eberle et al., 2009; Di et al., 2013; Liu et al., 2019; Fu et al., 2020). PvCYP98A-3 was clustered with the putative *Sesamum indicum* SiC3H and *P. campanularia* PcCYP98A113 involved in the last step of RA biosynthetic pathway (Figure 6C) (Anterola et al., 2002; Levsh et al., 2019). Taken together with the expression patterns of *PvCYP98As*

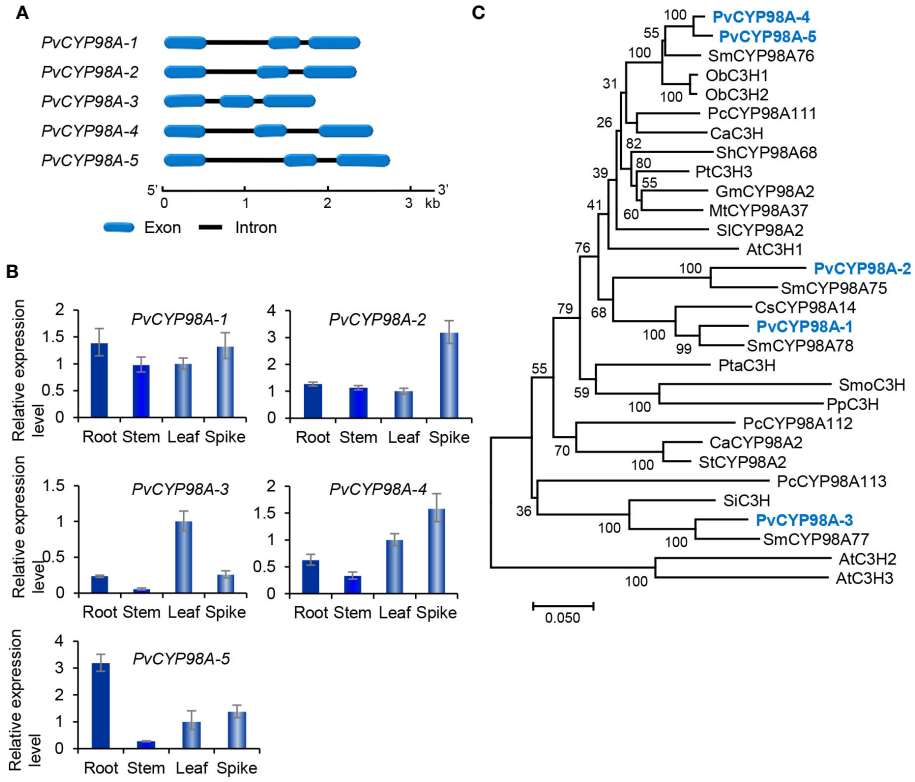


FIGURE 7
Gene structures, expression patterns and phylogenetic analysis of *PvCYP98* genes and their deduced proteins. **(A)** The intron-exon structures of five *PvCYP98* genes. **(B)** Fold changes of *PvCYP98* gene expression in roots, stems, leaves and spikes of *P. vulgaris* plants. The expression level in leaves was arbitrarily set to 1, respectively. **(C)** Phylogenetic analysis of PvCYP98 proteins. The unrooted Neighbor-Joining tree was constructed using the MEGA program (version 7.0) with default parameters. Proteins included are PvCYP98A-1–PvCYP98A-5 and the C3Hs and CYP98As from *Arabidopsis* (At), *S. miltiorrhiza* (Sm), *Ocimum basilicum* (Ob), *P. campanularia* (Pc), *C. arabica* (Ca), *Sinopodophyllum hexandrum* (Sh), *P. trichocarpa* (Pt), **(G)** *max* (Gm), *M. truncatula* (Mt), *S. lycopersicum* (Sl), *C. scutellarioides* (Cs), *P. taeda* (Pta), *Selaginella moellendorffii* (Smo), *Physcomitrium patens* (Pp), *Capsicum annuum* (Ca), *S. tuberosum* (St), and *S. indicum* (Si) (Supplementary Table 2).

(Figures 3E, 7B) and the distribution patterns of RAs in *P. vulgaris* (Kim et al., 2014), the results indicated that PvCYP98A-1 and PvCYP98A-2 could be involved in the hydroxylation of RAS products, PvCYP98A-4 and PvCYP98A-5 could be PvC3Hs, whereas the function of PvCYP98A-3 remained to be analyzed.

PvRAS3 and PvRAS4 were involved in RA biosynthesis *in vitro*

Based on gene expression patterns, phylogenetic relationships and the significance of RAS in RA biosynthesis, *PvRAS3* and *PvRAS4* were selected for functional analysis using experimental approaches. Among them, *PvRAS3* showed the highest expression among the eight *PvRAS* genes identified and was clustered with functionally known RASs in the phylogenetic tree constructed (Figures 6B, C). *PvRAS4* was one of the three *PvRAS* genes with expression patterns similar to RA distribution patterns (Kim et al., 2014) (Figure 6B). *PvRAS3* and *PvRAS4* cDNAs were cloned and introduced into *E. coli* competent cells. Recombinant proteins were induced and purified (Supplementary Figure 10). To assay the activity of recombinant PvRAS3 enzyme, *p*-coumaroyl-CoA or caffeoyl-CoA was used as the acyl donor, and pHPL or DHPL was used as the acyl acceptor. Negative controls were performed with *E. coli* BL21 (DE3) cells transformed with the empty pGEX-4T-1 vector (Supplementary Figure 11). LC-MS/MS analysis showed that PvRAS3 could catalyze the condensation of acyl donors and acceptors to generate four compounds, respectively (Figures 8A–D). Compound 1, generated through the condensation of caffeoyl-CoA and DHPL, was identified as RA based on UPLC and LC-MS/MS analyses (Figures 8A, E) and previous publication (Landmann et al., 2011; Di et al., 2013). Compounds 2 (Figure 8B) and 3 (Figure 8C) had the formula $C_{18}H_{16}O_7$ according to the MS spectra in negative mode $[M\text{-}H]^-$ (m/z = 343) (Figures 8F, G). They corresponded to an ester of caffeoyl-CoA and DHPL (caffeoyl-4'-hydroxyphenyllactic acid) or *p*-coumaroyl-CoA and pHPL (4-coumaroyl-3',4'-dihydroxyphenyllactic acid), respectively, as described previously (Landmann et al., 2011). Compound 4 was determined to be the ester of *p*-coumaroyl-CoA and pHPL based on the negative ion spectra (m/z = 327, $C_{18}H_{15}O_6$) (Figures 8D, H) and previous publication (Landmann et al., 2011).

For kinetic analysis of PvRAS3, *p*-coumaroyl-CoA or caffeoyl-CoA was used as the acyl donor and pHPL or DHPL was used as the acyl acceptor. To test the acceptor specificity, the concentration of the donor substrates remained saturated, while the levels of the acceptor substrates were varied (Supplementary Figure 12). When

10.3389/fpls.2024.1374912

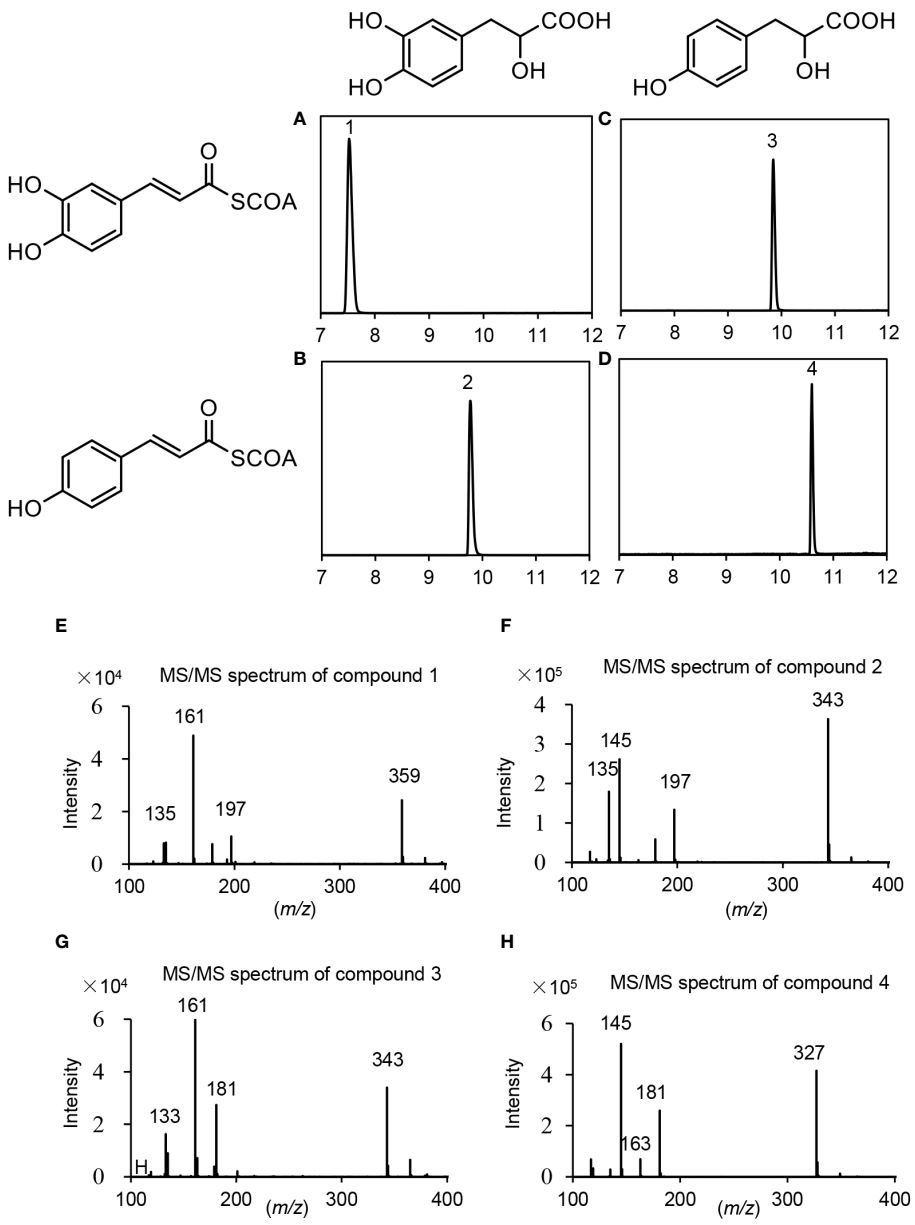


FIGURE 8
UPLC and LC-MS/MS detection of *in vitro* reaction products catalyzed by PvRAS3. **(A–D)** UPLC chromatograms of the reaction products catalyzed by PvRAS3 using caffeoyl-CoA and DHPL **(A)**, *p*-coumaroyl-CoA and DHPL **(B)**, caffeoyl-CoA and pHPL **(C)**, and *p*-coumaroyl-CoA and pHPL **(D)** as acyl donors and acceptors, respectively. **(E–H)** MS/MS spectra of compounds 1, 2, 3, and 4.

caffeoyl-CoA was used as the donor, the *K*m values of PvRAS3 with DHPL and pHPL were 197.6 and 166.8 μM, respectively (Table 2). Using *p*-coumaroyl-CoA as the donor, the *K*m values of DHPL and pHPL were 41.7 and 32.1 μM, respectively. The results suggested that the *K*m values of DHPL and pHPL with *p*-coumaroyl-CoA were lower than those with caffeoyl-CoA. To assess the donor affinity, the concentrations of the acceptors (DHPL and pHPL) were kept constant while the levels of caffeoyl-CoA and *p*-coumaroyl-CoA were varied. The results showed that the *K*m value of PvRAS3 for Caffeoyl-CoA was approximately 5.7-fold higher than that for *p*-coumaroyl-CoA (Table 2). Overall, PvRAS3 exhibited a high affinity toward DHPL and pHPL when *p*-coumaroyl-CoA was used as the acyl donor.

Similarly, *p*-coumaroyl-CoA, caffeoyl-CoA, pHPL and DHPL were also used as substrates for the analysis of PvRAS4 (Supplementary Figure 13). LC-MS/MS analysis showed that products could be detected when *p*-coumaroyl-CoA was used as acyl donor and pHPL or DHPL were used as acyl acceptor (Figures 9B, D). Based on MS spectra, the products were identical to compounds 2 and 4 catalyzed by PvRAS3 (Figures 9E, F). The K_{cat}/K_m values of PvRAS4 for *p*-coumaroyl-CoA and pHPL or DHPL were smaller than those of PvRAS3 (Table 3). No product was found when caffeoyl-CoA was used as acyl donor (Figures 9A, B). The results indicated that PvRAS4 could use *p*-coumaroyl-CoA, but not caffeoyl-CoA, as acyl donor. However, its affinity toward *p*-coumaroyl-CoA was lower than PvRAS3.

10.3389/fpls.2024.1374912

TABLE 2 Kinetic parameters of recombinant PvRAS3 toward different substrates.

Saturating subatrate	Varying substrate	V_{max} (nkat mg^{-1})	K_m (μM)	K_{cat} (S^{-1})	K_{cat}/K_m (M^{-1} S^{-1})
Caffeoyl-CoA	DHPL	187.6 ± 1.22	197.6 ± 3.4	14.0 ± 1.2	0.7×10^5
Caffeoyl-CoA	pHPL	189.7 ± 9.3	166.8 ± 14.4	13.2± 1.4	0.8×10^5
p-coumaroyl-CoA	DHPL	303.0 ± 5.0	41.7 ± 1.2	30.5 ± 2.1	7.0×10^5
p-coumaroyl-CoA+	pHPL	323.4 ± 5.4	32.1 ± 0.7	33.3 ± 2.3	10.0×10^5
DHPL	Caffeoyl-CoA	165.8 ± 6.2	173.2 ± 9.0	14.1 ± 1.0	0.8×10^5
DHPL	*p*-coumaroyl-CoA	307.5 ± 4.0	28.9 ± 7.5	13.1 ± 1.1	5.0×10^5
pHPL	Caffeoyl-CoA	155.6 ± 6.3	175.0 ± 9.1	32.8 ± 2.1	1.9×10^5
pHPL	*p*-coumaroyl-CoA	311.5 ± 3.2	30.6 ± 1.2	34.3 ± 2.2	11.0×10^5

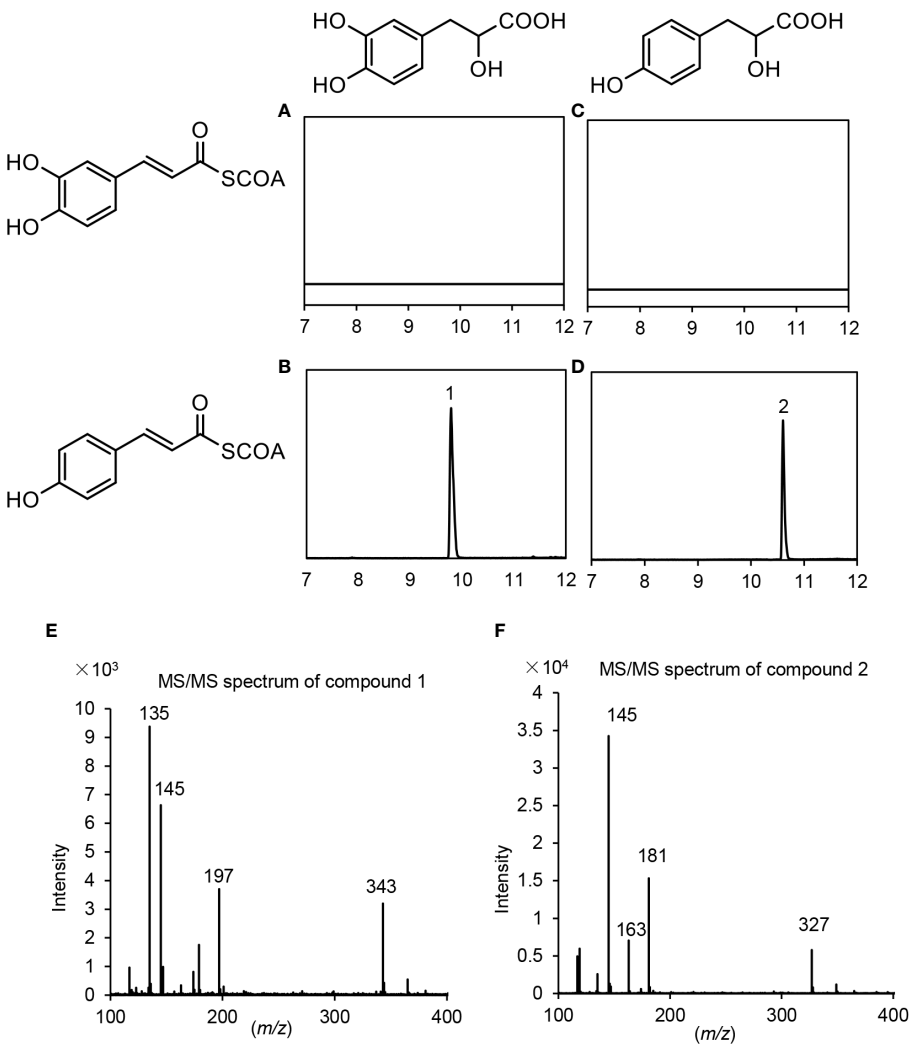


FIGURE 9
UPLC and LC-MS/MS detection of *in vitro* reaction products catalyzed by PvRAS4. **(A–D)** UPLC chromatograms of the reaction products catalyzed by PvRAS4 using caffeoyl-CoA and DHPL **(A)**, *p*-coumaroyl-CoA and DHPL **(B)**, caffeoyl-CoA and pHPL **(C)**, and *p*-coumaroyl-CoA and pHPL **(D)** as acyl donors and acceptors, respectively. **(E, F)** MS/MS spectra of compounds 1 and 2.

10.3389/fpls.2024.1374912

TABLE 3 Kinetic parameters of recombinant PvRAS4 toward different substrates.

Saturating subatrate	Varying substrate	V_{max} (nkat mg^{-1})	K_m (μM)	K_{cat} (S^{-1})	K_{cat}/K_m (M^{-1} S^{-1})
Caffeoyl-CoA	DHPL	N.D.	N.D.	N.D.	N.D.
Caffeoyl-CoA	pHPL	N.D.	N.D.	N.D.	N.D.
p-coumaroyl-CoA	DHPL	122.6± 5.7	226.9 ± 17.5	4.3± 1.2	0.2×10^5
p-coumaroyl-CoA+	pHPL	125.5 ± 3.6	228.5 ± 16.6	5.1 ± 2.1	0.2×10^5
DHPL	Caffeoyl-CoA	N.D.	N.D.	N.D.	N.D.
DHPL	*p*-coumaroyl-CoA	132.5 ± 9.2	217.5 ± 10.3	4.3 ± 1.6	0.2×10^5
pHPL	Caffeoyl-CoA	N.D.	N.D.	N.D.	N.D.
pHPL	*p*-coumaroyl-CoA	127.0 ± 5.9	216.8 ± 15.2	5.4± 1.3	0.2×10^5

N.D., not detected.

Existence of RA, 4-coumaroyl-3′,4′-dihydroxyphenyllactic acid, 4-coumaroyl-4′-hydroxyphenyllactic acid and caffeoyl-4′-hydroxyphenyllactic acid in *P. vulgaris* plants

Enzyme activity assay showed that PvRAS3 could condense *p*-coumaroyl-CoA and caffeoyl-CoA with pHPL and DHPL *in vitro*. To analyze whether the products exist in *P. vulgaris* plants, phenolic acid compounds were extracted from roots, stems, leaves and flowers and analyzed using UPLC and LC-MS/MS as described (Kim et al., 2014). The results showed that all of the four products, including RA, 4-coumaroyl-3′,4′-dihydroxyphenyllactic acid, 4-coumaroyl-4′-hydroxyphenyllactic acid and caffeoyl-4′-hydroxyphenyllactic acid, could be detected in the tissues analyzed (Figure 10; Supplementary Figure 14). RA was highly accumulated at the level of mg g^{-1} fresh weight (FW) with the highest level of 6.1 mg g^{-1} FW in flowers and the lowest level of 3.1 mg g^{-1} FW in stems (Figure 10A). The contents of 4-coumaroyl-4′-hydroxyphenyllactic acid ranged from 51.9 μg g^{-1} FW in roots to 140.3 μg g^{-1} FW in leaves (Figure 10B). The contents of 4-coumaroyl-4′-hydroxyphenyllactic acid were relatively low, which

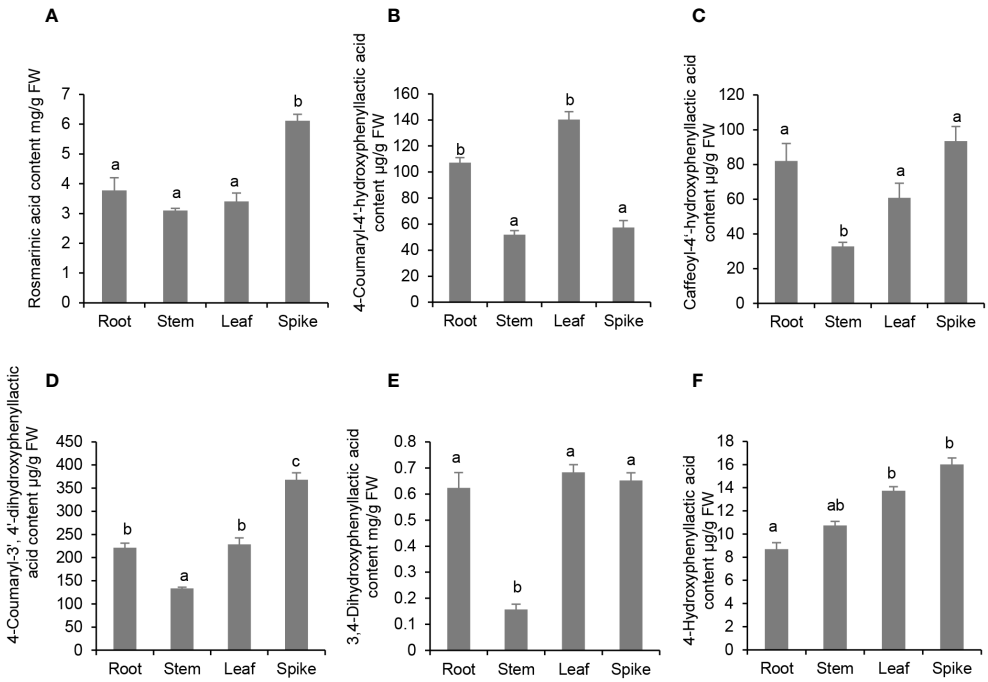


FIGURE 10
Analysis of phenolic acid contents in roots, stems, leaves and spikes of *P. vulgaris*. **(A)** Rosmarinic acid; **(B)** 4-Coumaroyl-4′-hydroxyphenyllactic acid; **(C)** Caffeoyl-4′-hydroxyphenyllactic acid; **(D)** 4-Coumaroyl-3′, 4′-dihydroxyphenyllactic acid; **(E)** 3,4-Dihydroxyphenyllactic acid; **(F)** 4-Hydroxyphenyllactic acid. Error bars represent standard deviations of mean value from three biological replicates. ANOVA (analysis of variance) was calculated using SPSS. $P < 0.05$ was considered statistically significant and was represented by different letters above the bars. $P \geq 0.05$ was considered statistically non-significant and was represented by same letters above the bars.

10.3389/fpls.2024.1374912

ranged from 32.9 μg g^{-1} FW in roots to 93.5 μg g^{-1} FW in flowers (Figure 10C). The contents of 4-coumaroyl-3',4'-dihydroxyphenyllactic acid ranged from 133.3 μg g^{-1} FW in roots to 368.0 μg g^{-1} FW in flowers (Figure 10D). The contents of RA and 4-coumaroyl-3',4'-dihydroxyphenyllactic acid in *P. vulgaris* were higher but comparable to those in lavender flowers, which were 2 mg g^{-1} FW and 150 μg g^{-1} FW, respectively (Landmann et al., 2011). To our best knowledge, this is the first report to detect 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid in plants. In addition, we also analyzed the contents of 3,4-dihydroxyphenyllactic acid and 4-hydroxyphenyllactic acid. Among them, 3,4-dihydroxyphenyllactic acid was highly accumulated in roots, leaves and flowers (Figure 10E), whereas the content of 4-hydroxyphenyllactic acid was relatively higher in leaves and flowers than roots and stems (Figure 10F). The results suggest that both the substrates and products of the four reactions catalyzed by PvRAS3 and/or PvRAS4 *in vitro* exist in *P. vulgaris* plants.

CRISPR/Cas9-mediated functional analysis of PvRAS3 in *P. vulgaris* hairy roots

Gene expression profiling and *In vitro* enzyme activity assay showed that PvRAS3 could be the main RAS catalyzing RA biosynthesis in *P. vulgaris*. To gain further insight into the involvement of PvRAS3 in RA biosynthesis, we designed two guide RNAs (gRNAs) targeting the first coding exon of *PvRAS3* for the CRISPR/Cas9 gene-editing tool. Five lines of transgenic hairy roots with the same insertion and deletion patterns in *PvRAS3* gene were obtained (Figure 11A). It indicated that these transgenic lines were homozygous mutants of *PvRAS3*, hereinafter referred to as *pvras3-1*, *pvras3-2*, *pvras3-3*, *pvras3-4* and *pvras3-5*, respectively. Analysis of RA, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid showed that the contents of these compounds decreased significantly in *pvras3* mutants with the contents of RA and caffeoyl-4'-hydroxyphenyllactic acid almost below the detection limit

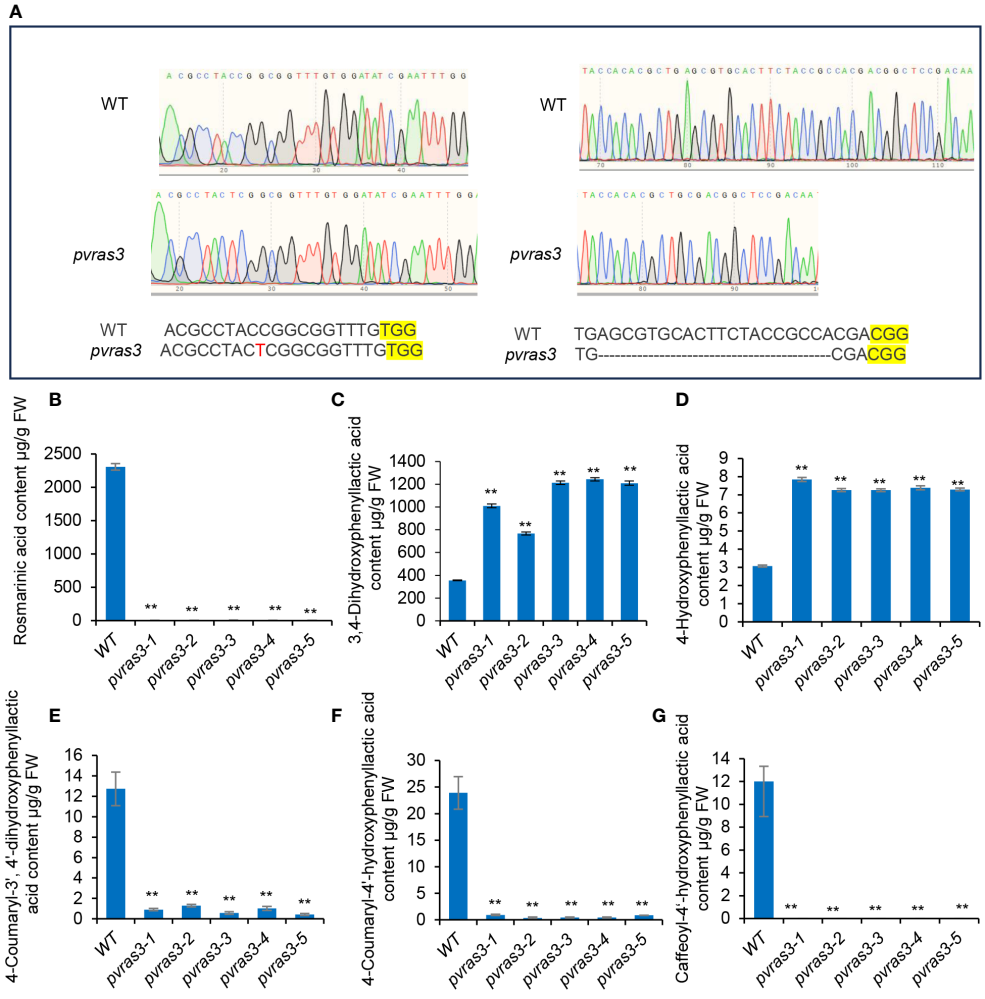


FIGURE 11
Analysis of phenolic acid contents in hairy roots of wild type and *pvras3* mutants. **(A)** The sgRNA-edited nucleotide sequences in *pvras3* mutants. WT, wild type. **(B–G)** Contents of rosmarinic acid **(B)**, 3,4-dihydroxyphenyllactic acid **(C)**, 4-hydroxyphenyllactic acid **(D)**, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid **(E)**, 4-coumaroyl-4'-hydroxyphenyllactic acid **(F)** and caffeoyl-4'-hydroxyphenyllactic acid **(G)** in wild type (WT) and five *pvras3* mutant lines. $P < 0.05$ (*) and $P < 0.01$ (**) were considered statistically significant and highly significant, respectively.

(Figures 11B, E–G). On the contrary, the contents of DHPL and pHPL increased significantly in the mutants (Figures 10C, D). It confirmed the catalytical function of PvRAS3 and its significance in RA biosynthesis.

Conclusions

P. vulgaris is a species of the Lamiaceae family with significant medicinal value. RA is one of the major bioactive components in *P. vulgaris* medicinal materials. Through genome-wide analysis, a total of 51 P. *vulgaris* genes belonging to seven RA biosynthesis-related gene families were identified. Subsequent gene and protein feature analysis, gene expression analysis and phylogenetic relationship analysis showed that seventeen of them, including *PvPAL1*, *PvPAL2*, *PvC4H1*, *PvC4H2*, *Pv4CL1*, *Pv4CL3*, *Pv4CL6*, *Pv4CL7*, *Pv4CL8*, *PvTAT3*, *PvTAT4*, *PvHPPR1*, *PvHPPR3*, *PvRAS3*, *PvRAS4*, *PvCYP98A-1* and *PvCYP98A-2*, could be involved in RA biosynthesis. *In vitro* enzymatic assay showed that both of PvRAS3 and PvRAS4 were involved in RA biosynthesis. PvRAS3 could catalyze the condensation of *p*-coumaroyl-CoA and caffeoyl-CoA with pHPL and DHPL. The affinity of PvRAS3 toward *p*-coumaroyl-CoA was higher than caffeoyl-CoA. PvRAS4 only catalyzed the condensation of *p*-coumaroyl-CoA with pHPL and DHPL. The affinity of PvRAS4 toward *p*-coumaroyl-CoA was lower than PvRAS3. These results were consistent with *in vivo* phenolic acid compound determination and *PvRAS3* transgenic analysis. UPLC analysis of phenolic acid compounds showed the existence of RA, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid in roots, stems, leaves and flowers of *P. vulgaris*. Generation of *pvras3* homozygous mutants through CRISPR/Cas9 technology and subsequent chemical compound analysis showed that the contents of RA, 4-coumaroyl-3',4'-dihydroxyphenyllactic acid, 4-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid decreased significantly, whereas the contents of DHPL and pHPL increased significantly in *pvras3* mutants. These results indicate the existence of four possible RA biosynthetic routes in *P. vulgaris*, which remains to be further confirmed through the analysis of *PvCYP98A* genes (Figure 1). Among them, routes 1 and 2 could be the main routes. PvRAS3 was the main enzyme catalyzing the condensation of acyl donors and acyl acceptors during RA biosynthesis in *P. vulgaris*. PvRAS4 could play a minor role. Further functional analysis of other fourteen candidate genes, particularly *PvCYP98A-1* and *PvCYP98A-2*, may provide a more complete picture of RA biosynthetic pathway.

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

CY: Data curation, Writing – original draft. CL: Data curation, Methodology, Validation, Writing – review & editing. MJ: Data curation, Writing – review & editing. YX: Data curation, Writing – review & editing. SZ: Writing – review & editing. XH: Writing – review & editing. YC: Writing – review & editing. SL: Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Chinese Academy of Medical Sciences Medical and Health Science and Technology Innovation Project [grant No. 2022-I2M-2-001], the National Natural Science Foundation of China [grant No. 31500263], and the Sichuan Science and Technology Program [grant No. 19YJ0368].

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1374912/full#supplementary-material

References

Achnine, L., Blancaflor, E. B., Rasmussen, S., and Dixon, R. A. (2004). Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 16, 3098–3109. doi: 10.1105/tpc.104.024406

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Anterola, A. M., Jeon, J. H., Davin, L. B., and Lewis, N. G. (2002). Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *J. Biol. Chem.* 277, 18272–18280. doi: 10.1074/jbc.M112051200

Berger, A., Meinhard, J., and Petersen, M. (2006). Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. *Planta* 224, 1503–1510. doi: 10.1007/s00425-006-0393-y

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 bigbiological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, H. C., Li, Q., Shuford, C. M., Liu, J., Muddiman, D. C., Sederoff, R. R., et al. (2011). Membrane protein complexes catalyze both 4- and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. *Proc. Natl. Acad. Sci. United States America* 108, 21253–21258. doi: 10.1073/pnas

Chen, Y., Zhang, X., Guo, Q., Cao, L., Qin, Q., Li, C., et al. (2019). Plant morphology, physiological characteristics, accumulation of secondary metabolites and antioxidant activities of *Prunella vulgaris* L. under UV solar exclusion. *Biol. Res.* 52, 17. doi: 10.1186/s40659-019-0225-8

Chen, Y., Li, Z., Guo, Q., Wang, C., Cao, L., Tang, H., et al. (2022). Identification of key genes related to flowering by transcriptome of flowering and nonflowering *Prunella vulgaris* L. *Biol. Plantarum* 66, 1–13. doi: 10.32615/bp.2021.056

Cui, X., Meng, F., Pan, X., Qiu, X., Zhang, S., Li, C., et al. (2022). Chromosome-level genome assembly of Aristolochia contorta provides insights into the biosynthesis of benzylisoquinoline alkaloids and aristolochic acids. *Horticulture Res.* 9, uhac005. doi: 10.1093/hr/uhac005

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

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

Eberle, D., Ullmann, P., Werck-Reichhart, D., and Petersen, M. (2009). cDNA cloning and functional characterisation of CYP98A14 and NADPH: cytochrome P450 reductase from *Coleus blumei* involved in rosmarinic acid biosynthesis. *Plant Mol. Biol.* 69, 239–253. doi: 10.1007/s11103-008-9420-7

Franke, R., Hemm, M. R., Denault, J. W., Ruegger, M. O., Humphreys, J. M., and Chapple, C. (2002). Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis*. *Plant J.* 30, 47–59. doi: 10.1046/j.1365-313X.2002.01267.x

Fu, R., Shi, M., Deng, C., Zhang, Y., Zhang, X., Wang, Y., et al. (2020). Improved phenolic acid content and bioactivities of *Salvia miltiorrhiza* hairy roots by genetic manipulation of *RAS* and *CYP98A14*. *Food Chem.* 331, 127365. doi: 10.1016/j.foodchem.2020.127365

He, Y., Zhong, X., Jiang, X., Cong, H., Sun, H., and Qiao, F. (2020). Characterisation, expression and functional analysis of *PAL* gene family in *Cephalotaxus hainanensis*. *Plant Physiol. Biochem.* 156, 461–470. doi: 10.1016/j.plaphy.2020.09.030

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

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

Hu, J., Yan, C., Li, S., Tang, H., and Chen, Y. (2023). Comparative physiological responses and transcriptome analysis revealing the metabolic regulatory mechanism of *Prunella vulgaris* L. induced by exogenous application of hydrogen peroxide. *Ind. Crops Products* 192, 116065. doi: 10.1016/j.indcrop.2022.116065

Khatri, P., Chen, L., Rajcan, I., and Dhaubhadel, S. (2023). Functional characterization of cinnamate 4-hydroxylase gene family in soybean (*Glycine max*). *PloS One* 18, e0285698. doi: 10.1371/journal.pone.0285698

Kim, K. H., Janiak, V., and Petersen, M. (2004). Purification, cloning and functional expression of hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis in cell cultures of *Coleus blumei*. *Plant Mol. Biol.* 54, 311–323. doi: 10.1023/B:PLAN.0000036367.03056.b2

Kim, Y. B., Shin, Y., Tuan, P. A., Li, X., Park, Y., Park, N. I., et al. (2014). Molecular cloning and characterization of genes involved in rosmarinic acid biosynthesis from *Prunella vulgaris*. *Biol. Pharm. Bull.* 37, 1221–1227. doi: 10.1248/bpb.b14-00139

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

Kwon, Y. O., Hong, J. T., and Oh, K. W. (2017). Rosmarinic acid potentiates pentobarbital-induced sleep behaviors and non-rapid eye movement (NREM) sleep through the activation of GABAA-ergic systems. *Biomolecules Ther.* 25, 105–111. doi: 10.4062/biomolther.2016.035

Lalitha, S. (2000). Primer premier 5. *Biotech. Software Internet Rep.* 1, 270–272. doi: 10.1089/152791600459894

Landmann, C., Hücherig, S., Fink, B., Hoffmann, T., Dittlein, D., Coiner, H. A., et al. (2011). Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). *Planta* 234, 305–320. doi: 10.1007/s00425-011-1400-5

Levsh, O., Pluskal, T., Carballo, V., Mitchell, A. J., and Weng, J. K. (2019). Independent evolution of rosmarinic acid biosynthesis in two sister families under the Lamiids clade of flowering plants. *J. Biol. Chem.* 294, 15193–15205. doi: 10.1074/jbc.RA119.010454

Li, Z., Hu, J., Tang, H., Cao, L., Guo, Q., Wang, C., et al. (2022). Temperature and photoperiod change the flowering process in *Prunella vulgaris* by inducing changes in morphology, endogenous hormones, carbon/nitrogen metabolites and gene expression. *J. Am. Soc. Hortic. Sci.* 147, 73–81. doi: 10.21273/JASHS05144-21

Liu, Q., Yao, L., Xu, Y., Cheng, H., Wang, W., Liu, Z., et al. (2019). *In vitro* evaluation of hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase expression and regulation in *Taraxacum antungense* in relation to 5-caffeoylquinic acid production. *Phytochemistry* 162, 148–156. doi: 10.1016/j.phytochem.2019.02.014

Liu, T., Yao, R., Zhao, Y., Xu, S., Huang, C., Luo, J., et al. (2017). Cloning, functional characterization and site-directed mutagenesis of 4-coumarate: Coenzyme A ligase (4CL) involved in coumarin biosynthesis in *Peucedanum praeruptorum* Dunn. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00004

Lu, S. (2021). Biosynthesis and regulatory mechanisms of bioactive compounds in *Salvia miltiorrhiza*, a model system for medicinal plant biology. *Crit. Rev. Plant Sci.* 40, 243–283. doi: 10.1080/07352689.2021.1935719

Lu, X., Hao, L., Wang, F., Huang, C., and Wu, S. (2013). Molecular cloning and overexpression of the tyrosine aminotransferase (TAT) gene leads to increased rosmarinic acid yield in *Perilla frutescens*. *Plant Cell Tissue Organ Culture* 115, 69–83. doi: 10.1007/s11240-013-0341-z

Ma, W. L., Wu, M., Wu, Y., Ren, Z., and Zhong, Y. (2013). Cloning and characterization of a phenylalanine ammonia-lyase gene from *Rhus chinensis*. *Plant Cell Rep.* 32, 1179–1190. doi: 10.1007/s00299-013-1413-6

Mansouri, M., and Mohammadi, F. (2021). Transcriptome analysis to identify key genes involved in terpenoid and rosmarinic acid biosynthesis in lemon balm (*Melissa officinalis*). *Gene* 773, 145417. doi: 10.1016/j.gene.2021.145417

National Pharmacopoeia Committee (2020). *Pharmacopoeia of the people's Republic of China (part I)* (Beijing: China Medical Science and Technology Press).

Noguchi-Shinohara, M., Ono, K., Hamaguchi, T., Iwasa, K., Nagai, T., Kobayashi, S., et al. (2015). Pharmacokinetics, safety and tolerability of *Melissa officinalis* extract which contained rosmarinic acid in healthy individuals: a randomized controlled trial. *PloS One* 10, e0126422. doi: 10.1371/journal.pone.0126422

Pan, X., Chang, Y., Li, C., Qiu, X., Cui, X., Meng, F., et al. (2023). Chromosome-level genome assembly of Salvia miltiorrhiza with orange roots uncovers the role of Sm2OGD3 in catalyzing 15,16-dehydrogenation of tanshinones. *Horticulture Res.* 10, uhad069. doi: 10.1093/hr/uhad069

Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Hücherig, S., et al. (2009). Evolution of rosmarinic acid biosynthesis. *Phytochemistry* 70, 1663–1679. doi: 10.1016/j.phytochem.2009.05.010

Poppe, L., and Rétey, J. (2005). Friedel-crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine. *Angewandte Chemie (Internattional Ed. English)* 44, 3668–3688. doi: 10.1002/anie.200461377

Radziejewska, I., Supruniuk, K., Nazaruk, J., Karna, E., Poplawska, B., Bielawska, A., et al. (2018). Rosmarinic acid influences collagen, MMPs, TIMPs, glycosylation and MUC1 in CRL-1739 gastric cancer cell line. *Biomedicine Pharmacotherapy* 107, 397–407. doi: 10.1016/j.biopha.2018.07.123

Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133, 1051–1071. doi: 10.1104/pp.103.026484

Ralph, J., Akiyama, T., Kim, H., Lu, F., Schatz, P. F., Marita, J. M., et al. (2006). Effects of coumarate 3-hydroxylase down-regulation on lignin structure. *J. Biol. Chem.* 281, 8843–8853. doi: 10.1074/jbc.M511598200

Ro, D. K., Mah, N., Ellis, B. E., and Douglas, C. J. (2001). Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiol.* 126, 317–329. doi: 10.1104/pp.126.1.317

Ru, M., Wang, K., Bai, Z., Peng, L., He, S., Pei, T., et al. (2017b). Molecular cloning and characterisation of two enzymes involved in the rosmarinic acid biosynthesis pathway of *Prunella vulgaris* L. *Plant Cell Tissue Organ Culture* 128, 381–390. doi: 10.1007/s11240-016-1117-z

Ru, M., Wang, K., Bai, Z., Peng, L., He, S., Wang, Y., et al. (2017a). A tyrosine aminotransferase involved in rosmarinic acid biosynthesis in *Prunella vulgaris* L. *Sci. Rep.* 7, 4892. doi: 10.1038/s41598-017-05290-4

Sahraeian, S. M. E., Mohiyuddin, M., Sebra, R., Tilgner, H., Afshar, P. T., Au, K. F., et al. (2017). Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat. Communication* 8, 59. doi: 10.1038/s41467-017-00050-4

Sander, M., and Petersen, M. (2011). Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth. *Planta* 233, 1157–1171. doi: 10.1007/s00425-011-1367-2

Scarpati, M., and Oriente, G. (1958). Isolation and constitution of rosmarinic acid from *Rosmarinus officinalis*. *La Ricerca Scientifica* 28, 2329–2333.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., et al. (2001). CYP98A3 from *Arabidopsis thaliana* is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.* 276, 36566–36574. doi: 10.1074/jbc.M104047200

Shi, R., Sun, Y. H., Li, Q., Heber, S., Sederoff, R., and Chiang, V. L. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.* 51, 144–163. doi: 10.1093/pcp/pcp175

Song, J., and Wang, Z. (2011). RNAi-mediated suppression of the phenylalanine ammonia-lyase gene in *Salvia miltiorrhiza* causes abnormal phenotypes and a reduction in rosmarinic acid biosynthesis. *J. Plant Res.* 124, 183–192. doi: 10.1007/s10265-010-0350-5

Taguchi, R., Hatayama, K., Takahashi, T., Hayashi, T., Sato, Y., Sato, D., et al. (2017). Structure-activity relations of rosmarinic acid derivatives for the amyloid β aggregation inhibition and antioxidant properties. *Eur. J. Medicinal Chem.* 138, 1066–1075. doi: 10.1016/j.ejmech.2017.07.026

Tang, H., Hu, J., Zhao, M., Cao, L., and Chen, Y. (2023). Comparative study of the physiological responses, secondary metabolites, and gene expression of medicinal plant *Prunella vulgaris* L. treated with exogenous methyl jasmonate and salicylic acid. *Acta Physiologiae Plantarum* 45, 20. doi: 10.1007/s11738-022-03498-0

Uhlmann, A., and Ebel, J. (1993). Molecular cloning and expression of 4-coumarate-Coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. *Plant Physiol.* 102, 1147–1156. doi: 10.1104/pp.102.4.1147

Wang, G. Q., Chen, J. F., Yi, B., Tan, H. X., Zhang, L., and Chen, W. S. (2017). *HPPR* encodes the hydroxyphenylpyruvate reductase required for the biosynthesis of hydrophilic phenolic acids in *Salvia miltiorrhiza*. *Chin. J. Natural Medicines* 15, 917–927. doi: 10.1016/S1875-5364(18)30008-6

Wang, H., Dong, Q., Duan, D., Zhao, S., Li, M., van Nocker, S., et al. (2018). Comprehensive genomic analysis of the *TYROSINE AMINOTRANSFERASE* (*TAT*) genes in apple (*Malus domestica*) allows the identification of *MdTAT2* conferring tolerance to drought and osmotic stresses in plants. *Plant Physiol. Biochem.* 133, 81–91. doi: 10.1016/j.plaphy.2018.10.033

Wang, Y., Guo, L., Zhao, Y., Zhao, X., and Yuan, Z. (2022b). Systematic analysis and expression profiles of the 4-coumarate: CoA ligase (4CL) gene family in Pomegranate (*Punica granatum* L.). *Int. J. Mol. Sci.* 23, 3509. doi: 10.3390/ijms23073509

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, C., Yang, J., Song, P., Zhang, W., Lu, Q., Yu, Q., et al. (2022a). FIONA1 is an RNA N^6-methyladenosine methyltransferase affecting *Arabidopsis* photomorphogenesis and flowering. *Genome Biol.* 23, 40. doi: 10.1186/s13059-022-02612-2

Watts, K. T., Mijts, B. N., Lee, P. C., Manning, A. J., and Schmidt-Dannert, C. (2006). Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chem. Biol.* 13, 1317–1326. doi: 10.1016/j.chembiol.2006.10.008

Weitzel, C., and Petersen, M. (2010). Enzymes of phenylpropanoid metabolism in the important medicinal plant *Melissa officinalis* L. *Planta* 232, 731–742. doi: 10.1007/s00425-010-1206-x

Weitzel, C., and Petersen, M. (2011). Cloning and characterisation of rosmarinic acid synthase from *Melissa officinalis* L. *Phytochemistry* 72, 572–578. doi: 10.1016/j.phytochem.2011.01.039

Xiao, Y., Zhang, L., Gao, S., Saechao, S., Di, P., Chen, J., et al. (2011). The *c4h*, *tat*, *hppr* and *hppd* genes prompted engineering of rosmarinic acid biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *PloS One* 6, e29713. doi: 10.1371/journal.pone.0029713

Xu, J. J., Fang, X., Li, C. Y., Zhao, Q., Martin, C., Chen, X. Y., et al. (2018). Characterization of *Arabidopsis thaliana* hydroxyphenylpyruvate reductases in the tyrosine conversion pathway. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01305

Zhang, S., Meng, F., Pan, X., Qiu, X., Li, C., and Lu, S. (2024). Chromosome-level genome assembly of *Prunella vulgaris* L. provides insights into pentacyclic triterpenoid biosynthesis. *Plant J.* doi: 10.1111/tpj.16629

Zhao, S. J., Hu, Z. B., Liu, D., and Leung, F. C. (2006). Two divergent members of 4-coumarate: coenzyme A ligase from *Salvia miltiorrhiza* Bunge: cDNA cloning and functional study. *J. Integr. Plant Biol.* 48, 1355–1364. doi: 10.1111/j.1744-7909.2006.00302.x

Zheng, H., Zhao, H., Zhang, X., Liang, Z., and He, Q. (2022). Systematic identification and validation of suitable reference genes for the normalization of gene expression in *Prunella vulgaris* under different organs and spike development stages. *Genes* 13, 1947. doi: 10.3390/genes13111947

Zhou, Z., Tan, H., Li, Q., Chen, J., Gao, S., Wang, Y., et al. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis of *RAS* in *Salvia miltiorrhiza*. *Phytochemistry* 148, 63–70. doi: 10.1016/j.phytochem.2018.01.015

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 May 2024
DOI 10.3389/fpls.2024.1394587

OPEN ACCESS

EDITED BY
Naoki Kitaoka,
Hokkaido University, Japan

REVIEWED BY
Xiujun Zhang,
Chinese Academy of Sciences (CAS), China
Yuhua Li,
Northeast Forestry University, China

*CORRESPONDENCE
Chao Sun
csun@implad.ac.cn
Baolin Guo
blguo@implad.ac.cn
Ying Li
liying@implad.ac.cn

†These authors have contributed equally to this work

RECEIVED 01 March 2024
ACCEPTED 24 April 2024
PUBLISHED 08 May 2024

CITATION
Li R, Du K, Zhang C, Shen X, Yun L, Wang S, Li Z, Sun Z, Wei J, Li Y, Guo B and Sun C (2024) Single-cell transcriptome profiling reveals the spatiotemporal distribution of triterpenoid saponin biosynthesis and transposable element activity in *Gynostemma pentaphyllum* shoot apexes and leaves. *Front. Plant Sci.* 15:1394587. doi: 10.3389/fpls.2024.1394587

Single-cell transcriptome profiling reveals the spatiotemporal distribution of triterpenoid saponin biosynthesis and transposable element activity in *Gynostemma pentaphyllum* shoot apexes and leaves

Rucan Li[1†], Ke Du[1†], Chuyi Zhang[1], Xiaofeng Shen[1], Lingling Yun[1], Shu Wang[1], Ziqin Li[2], Zhiying Sun[2], Jianhe Wei[1], Ying Li[1*], Baolin Guo[1*] and Chao Sun[1*]

[1]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China

Gynostemma pentaphyllum (Thunb.) Makino is an important producer of dammarene-type triterpenoid saponins. These saponins (gypenosides) exhibit diverse pharmacological benefits such as anticancer, antidiabetic, and immunomodulatory effects, and have major potential in the pharmaceutical and health care industries. Here, we employed single-cell RNA sequencing (scRNA-seq) to profile the transcriptomes of more than 50,000 cells derived from *G. pentaphyllum* shoot apexes and leaves. Following cell clustering and annotation, we identified five major cell types in shoot apexes and four in leaves. Each cell type displayed substantial transcriptomic heterogeneity both within and between tissues. Examining gene expression patterns across various cell types revealed that gypenoside biosynthesis predominantly occurred in mesophyll cells, with heightened activity observed in shoot apexes compared to leaves. Furthermore, we explored the impact of transposable elements (TEs) on *G. pentaphyllum* transcriptomic landscapes. Our findings the highlighted the unbalanced expression of certain TE families across different cell types in shoot apexes and leaves, marking the first investigation of TE expression at the single-cell level in plants. Additionally, we observed dynamic expression of genes involved in gypenoside biosynthesis and specific TE families during epidermal and vascular cell development. The involvement of TE expression in regulating cell differentiation and gypenoside biosynthesis warrant further exploration. Overall, this study not only provides new insights into the spatiotemporal

organization of gypenoside biosynthesis and TE activity in *G. pentaphyllum* shoot apexes and leaves but also offers valuable cellular and genetic resources for a deeper understanding of developmental and physiological processes at single-cell resolution in this species.

KEYWORDS

single-cell, RNA-seq, *Gynostemma pentaphyllum*, triterpenoid saponin, transposable element

1 Introduction

Gynostemma pentaphyllum (Thunb.) Makino is a perennial vine native to East Asia within the Cucurbitaceae family (Zhang et al., 2023). For centuries, this plant has been extensively used as a dietary herbal medicine due to its diverse pharmacological benefits, including antitumor, antiaging, immunomodulatory, and neuroprotective activities (Wang et al., 2020). Currently, *G. pentaphyllum* is readily available in various commercial formulations, such as tea, tablets, capsules, and powders (Niu et al., 2013). Additionally, extracts of *G. pentaphyllum* are utilized as additives in a range of products, including beverages, biscuits, bread, and noodles (Ji et al., 2018; Su et al., 2021). The major bioactive constituents of *G. pentaphyllum* are saponins, specifically gypenosides, the majority of which exist as dammarane-type triterpenoid saponins (Nguyen et al., 2021). More than 200 gypenosides have been identified in *G. pentaphyllum* (Nguyen et al., 2021). Notably, some gypenosides share identical structures with ginsenosides from *Panax ginseng*, making *G. pentaphyllum* the first ginsenoside-producing plant outside of the Araliaceae family (Zhang et al., 2023). Consequently, *G. pentaphyllum* may serve as an alternative resource for ginsenosides. The biosynthetic pathway of dammarane-type saponins in the genus *Gynostemma* has been partially elucidated. For instance, the first committed enzyme in this pathway is dammarenediol-II synthase (DS), which catalyzes the cyclization of 2,3-oxidosqualene to dammarenediol-II. A DS has been characterized from *G. longipes* (Ye et al., 2022). Furthermore, five UDP-dependent glycosyltransferases (UGTs), which are responsible for catalyzing the final steps of gypenoside biosynthesis, have been identified in *G. pentaphyllum* (Le et al., 2021). However, while cytochrome P450s (CYP450s) are predicted to play crucial roles in multiple hydroxylation steps, these processes have not been fully elucidated.

Single-cell RNA sequencing (scRNA-seq) technology offers an unparalleled opportunity to investigate intricate cellular and molecular processes by improving the spatiotemporal resolution of transcriptomic analysis to the level of the individual cell (Seyfferth et al., 2021). This cutting-edge technology has been extensively employed for the identification of rare and novel cellular entities, as well as for the elucidation of cellular differentiation and development in plant biology (Lopez-Anido et al., 2021). Notably, the use of scRNA-seq has been expanded to enhance the understanding of plant specialized metabolism. Compared to bulk RNA sequencing, scRNA-seq offers several advantages. These include providing single-cell resolution for gene expression profiling to resolve cellular heterogeneity, elucidating cell states and trajectories relevant to developmental processes, and revealing the spatiotemporal distribution of specialized metabolic pathways. For instance, scRNA-seq was used to dissect the spatial distribution of the vinblastine biosynthetic pathway in *Catharanthus roseus* leaves, and the pathway was found to be compartmentalized into three cell types: starting in internal phloem-associated parenchyma (IPAP) cells, followed by the intermediate enzymatic steps predominantly occurring in epidermal cells (ECs), and concluding with the late steps in idioblast cells (ICs) (Li et al., 2023; Sun et al., 2023). These findings are consistent with prior reports using RNA *in situ* hybridization. The integration of scRNA-seq with mass spectrometry imaging technology can further enhance our ability to explore the spatial organization of specialized metabolism. A combination of two technologies revealed that the majority of taxol biosynthesis genes are predominantly expressed in leaf mesophyll cells (MCs), while phenolic acid and flavonoid biosynthesis genes are highly expressed in leaf ECs (Zhan et al., 2023). In addition, a single-cell transcriptome atlas of tea leaves was constructed and a novel catechin ester glycosyltransferase was characterized by a gene coexpression network in MCs, suggesting that scRNA-seq has great potential for the screening and identifying genes involved in the biosynthesis of plant specialized metabolites (Wang et al., 2022).

Transposable elements (TEs) are DNA fragments characterized by their ability to mobilize or replicate within a host genome. In plants, in addition to their intrinsic transcription, TE transcripts strongly shape transcriptomic profiles by regulating host gene expression and chromatin accessibility (Fueyo et al., 2022). Given the diverse roles of TE transcripts in molding the host transcriptome, several computational tools have been developed to precisely quantify TE expression (Lanciano and Cristofari, 2020). Notably, two pipelines, scTE and soloTE, have been utilized to investigate TE expression at the single-cell level. For instance, the analysis of TE expression in mouse embryonic stem cells and during

human cardiac differentiation using scTE revealed that specific TE types are expressed in subpopulations of embryonic stem cells and undergo dynamic regulation during pluripotency reprogramming, differentiation, and embryogenesis (He et al., 2021). Furthermore, soloTE was employed to determine the impact of TE expression on the cellular heterogeneity of early gastric cancer (Rodriguez-Quiroz and Valdebenito-Maturana, 2022). This investigation revealed that two TEs, L1PA7 and THE1D, exhibit higher expression in the cancer cells than in other cell types. Although investigations of TE expression have expanded to single-cell resolution in various animal systems, characterization of TE transcriptional dynamics in plants has previously been limited to tissue-level profiles.

Here, we constructed high-resolution single-cell transcriptome atlases of *G. pentaphyllum* shoot apexes and leaves. Most of the genes involved in the gypenoside biosynthetic pathway exhibited high expression in MCs, indicating that MCs serve as the primary site for gypenoside biosynthesis. Furthermore, these pathway genes exhibited distinct expression patterns during EC and VC development. Notably, this study marks the first exploration of TE expression at the single-cell resolution in the plant kingdom. TE activity exhibited an uneven distribution among different cell types and during cell differentiation and development. Overall, this work provides novel insights into the spatiotemporal organization of triterpenoid saponin biosynthesis and TE activity in *G. pentaphyllum* shoot apexes and leaves. Additionally, the abundant datasets generated herein also establish a foundation for further elucidating developmental and physiological processes at single-cell resolution in this species.

2 Materials and methods

2.1 Plant material and protoplast preparation

G. pentaphyllum (Thunb.) Makino plants were grown at 25°C in a greenhouse under a 16/8 h photoperiod. Leaves and shoot apexes of young healthy stolons were harvested for protoplasting. Two biological replicates were prepared for each tissue. Protoplasts were obtained by enzymatic hydrolysis. Approximately 0.15 g of tender leaves and shoot apexes of *G. pentaphyllum* were harvested and cut into small pieces. The samples were treated with enzymolysis solution (1% cellulase R-10, 0.15% macerozyme R-10, 0.1% pectinase Y-23, 0.45 M mannitol and 20 mM MES, pH 5.7 - 5.8). The samples were incubated in darkness for 3.0 - 3.5 h at 25°C to isolate the protoplasts. After reaching a certain number of protoplasts were obtained and passed through 70-μm and twice 40-μm strainers, the protoplasts were centrifuged at 100 × g for 5 min and washed once with protoplasting solution (0.45 M mannitol and 20 mM MES, pH 5.7 - 5.8) without enzymes. Protoplasts were placed on ice until further processing. Protoplasts were stained with trypan blue (0.2% final) and checked on a hematocytometer under a Leica M205FA microscope to determine cell viability and concentration. The final cell concentration was adjusted to a range from 1,500 to 1,800 cells μl^{-1}.

2.2 scRNA-seq library construction and sequencing

The scRNA-seq libraries were constructed using the Chromium Single-cell 3′ Gel Beads-in-emulsion (GEM) Library and Gel Bead Kit v.3 (16 rxns PN-1000268, 10x Genomics) according to the user's manual supplied with the kit. In this study, approximately 20000 cells were counted per sample. The concentration of the DNA library was measured by a Qubit3.0. Qualitative analysis of the DNA library was performed with an Agilent 2100 Bioanalyzer. Libraries were sequenced on an Illumina HiSeq 4000 according to the manufacturer's instructions.

2.3 Preprocessing of scRNA-seq data

The *G. pentaphyllum* telomere-to-telomere (T2T) genome was downloaded from the National Center for Biotechnology Information (NCBI) (BioProject: PRJNA1030183). The raw reads generated via high-throughput sequencing were in the FASTQ format. The Alevin pipeline (Salmon v.0.6.0) (Srivastava et al., 2019) was used to count the unique molecular identifiers (UMIs) and construct digital expression matrices, using the command 'salmon alevin' with the arguments 'expectCells=15000'. We loaded the matrices of two leaf and two shoot apex to create seurat objects using the R package Seurat (v.4.3.1) (Hao et al., 2021). To obtain quality control statistics, such as high-quality cell numbers, gene medians, and sequencing saturation, we first specified the quality control standard, Cells with <500 and >6,500 detected genes, UMI counts <500 and >25,000, mitochondrial reads with percentages >3% and chloroplast genes with percentages >40% were removed. Doublets were filtered out by DoubletFinder (v.2.0.3) (McGinnis et al., 2019) with default settings. The cell cycle score was calculated using the CellCycleScoring function in Seurat and using the genes in Supplementary Table 12. We then used regularized negative binomial regression to normalize UMI counts using the SCTransform (vst.flavor = "v2") function in Seurat, with the percentage of mitochondrial genes, chloroplast genes, UMI counts and the cell cycle score regressed out. Principal component analysis (PCA) was performed using the RunPCA function based on highly variable genes detected using the SelectIntegrationFeatures function and 3,000 features. Batch correction was eliminated with the RunHarmony function using the R package Harmony (v.0.1.1) (Korsunsky et al., 2019), with the assay (parameter'assay.use') set as 'SCT'.

2.4 Cell clustering and annotation

We performed nearest-neighbor graph construction using the FindNeighbors function with the 'reduction' parameter in the FindNeighbors set as 'harmony', 'dims=20'. The UMAP algorithm was used to perform nonlinear dimensionality reduction and visualization of all cells with the RunUMAP function, with

parameter 'dims=20'. We determined graph-based clustering using the FindClusters function, and used clustree package (v.0.5.1) to select the appropriate resolution by plotting the clustering results at 0-1 resolution. The different clusters were extracted using the FindAllMarkers function, setting pct.2 < 0.1 and logfc.threshold=0.25 as thresholds to identify cell type-specific genes in particular clusters. After that, we selected upregulated cell type-specific genes in different clusters for GO enrichment analysis using ClusterProfiler (v.3.18.1). The simultaneous visualization of cell distribution and marker gene expression in two-dimensional space was performed using the Feature plot function.

2.5 Pseudotime analysis

Pseudotime trajectory analysis was conducted utilizing the Monocle R package (v.2.8.0) (Qiu et al., 2017). The log-normalized data derived from the Seurat object were imported into Monocle through the application of the as.CellDataSet function. Subsequently, the cells were ordered along the trajectory and presented in a reduced dimensional space. The determination of the trajectory root was performed according to the cell subcluster identities. To discern genes exhibiting significant changes along pseudotime, the differentialGeneTest function was applied, with the stringent criterion of a q value < 0.1. The identified genes were subsequently subjected to clustering using the plot_pseudotime_heatmap function, adhering to default parameters to ensure robust and reliable outcomes. Specifically, genes demonstrating dynamic expression patterns throughout pseudotime were visually represented through the plot_genes_in_pseudotime function, enhancing the granularity of our analysis. Differentially expressed genes including cluster-enriched genes with statistical significance were submitted to TBtools for GO enrichment analysis (Chen et al., 2020). The top 30 GO terms with $-\log_{10}$p-values were represented.

2.6 Analysis of scRNA-seq data using scTE

The TEs of the reference genome were annotated by RepeatModeler (v2.0.5) (Flynn et al., 2020), TEclass (v2.1.3C) (Abrusán et al., 2009) and RepeatMasker (v4.1.5). The annotation files of the genes and TEs were used to determine indices through the commander 'scTE_build' of the scTE (v1.0.4) (He et al., 2021) pipeline. The BAM file required for the pipeline was quantitatively generated by CellRanger (v7.0.1) from the single cell data. Consequently, the gene/TE expression matrix was generated through the commander 'scTE' with the arguments '-x, -i, -o, –expect-cells' and was used in Seurat (v4.3.1) for further analysis via the same process as that used for the analysis based on gene expression alone.

2.7 RNA *in situ* hybridization

For *in situ* hybridization assays, leaves harvested during the same growth phase as those used for protoplast isolation were fixed with formaldehyde-acetic acid-ethanol fixative (50%) at 4°C for 24 h, followed by manual embedding. The unique fragments of selected genes were cloned into the pGEM-T easy vector (Promega, catalogue no. A1360). The primers used are listed in Supplementary Table 15. For RNA probe synthesis, linearized vectors were added as templates. And the subsequent *in vitro* transcription was performed with a DIG RNA Labeling Kit (Sp6/T7) (Roche, catalogue no. 11175025910). The leaves were paraffin-embedded and sectioned (10 µm) with a sliding microtome (Leica). The sections were then affixed onto slides coated with 3-aminopropyltriethoxysilane (AES)-coated slides (WHITE 12-550-15, Thermo Fisher Scientific) overnight at 42 °C, and paraffin was removed using xylene (twice for 15 min) before rehydration in an ethanol gradient up to diethypyrocarbonate (DEPC)-treated water. Following rehydration, the sections were subjected to Proteinase K (Sigma, catalogue no. P2308) digestion, dehydrated via a series of ethanol solutions and hybridized with RNA probes. After washing, sections were incubated with anti-digoxigenin-AP Fab fragments (Roche, catalogue no. 11093274910). For color development, sections were immersed in NBT/BCIP (Sigma, catalogue no. B5655) staining solution at room temperature until the target color was clear. Microscopy was carried out in bright-field mode using BioTek CYTATION/5 imaging reader.

3 Results

3.1 Construction of the *G. pentaphyllum* shoot apex and leaf cell atlases

Leaves and young stems serve as prominent tissues for gypenoside biosynthesis in *G. pentaphyllum* (Figures 1A–C). Anatomical analyses were conducted using light microscopy on both leaf cross-sections and longitudinal sections of shoot apexes. After staining with Safranin-O and Fast Green, the sections were observed under a visible light microscope. Leaf sections revealed distinct upper and lower epidermal cell layers surrounding mesophyll parenchyma tissue and vascular bundles. Both the adaxial and abaxial epidermal tissues exhibited abundant large trichome structures (Figures 1D, E). In contrast, shoot apex tissues predominantly consisted of saffron-stained susceptible parenchymal tissue along with epidermal cell layers bearing smaller trichomes than did leaf tissues (Figures 1F). Protocols for protoplast isolation and purification were optimized, yielding highly viable protoplasts derived from vegetative shoot apexes and young leaves (Figures 1G–J). These intact protoplasts from specific gypenoside biosynthetic tissues were subsequently used to generate single-cell transcriptomic datasets.

Utilizing Illumina high-throughput sequencing, a total of 580 Gb of scRNA-seq data were generated from four distinct single-cell complementary DNA libraries originating from young leaf protoplasts and vegetative shoot apex protoplasts of *G. pentaphyllum*, with two biological replicates for each tissue (Supplementary Table 1). To ensure accurate quantification of the single-cell transcriptome, we utilized the reference-level telomere-to-telomere (T2T) genome of *G. pentaphyllum* (BioProject:

10.3389/fpls.2024.1394587

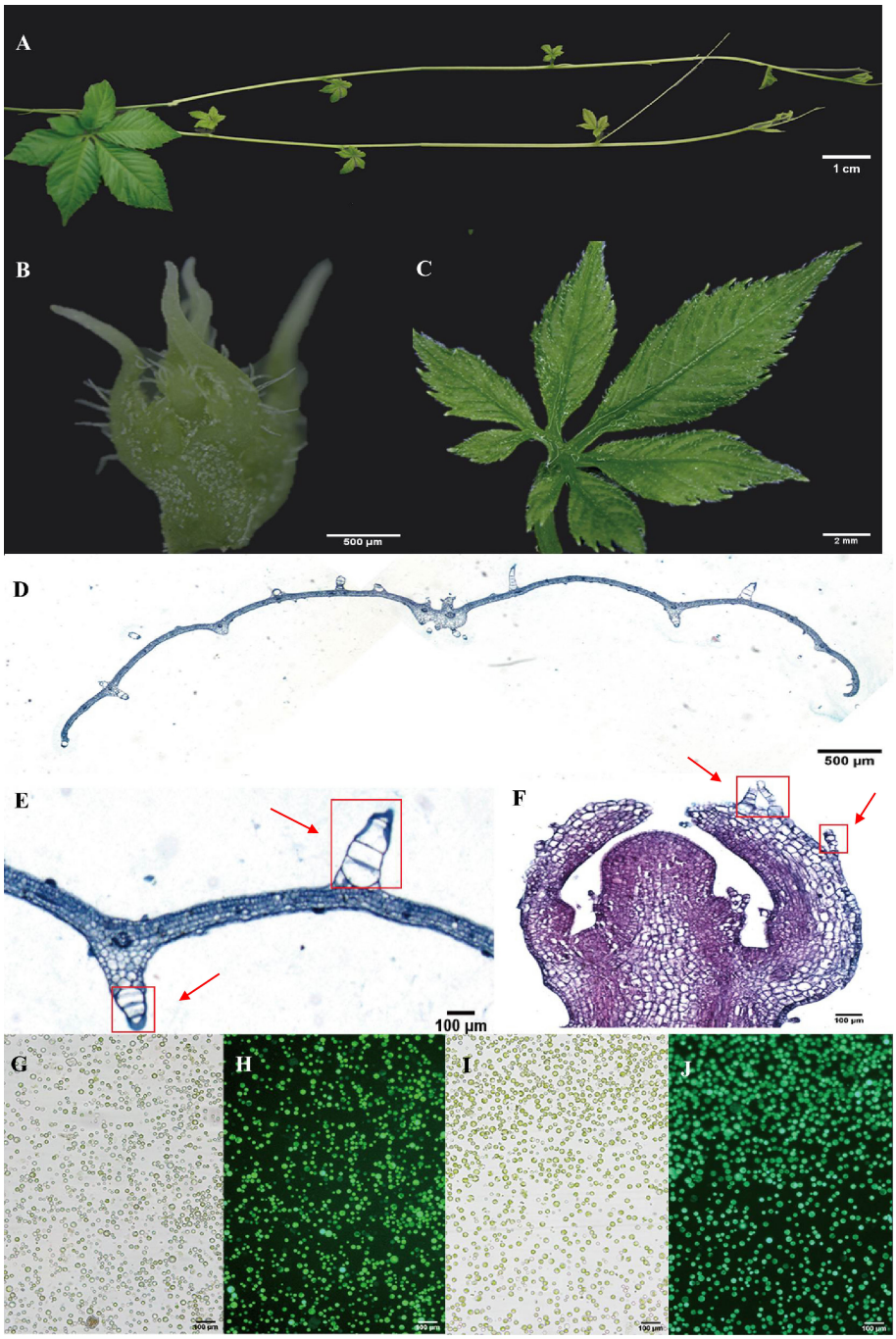


FIGURE 1
Anatomical features and protoplasts of *G. pentaphyllum* shoot apexes and leaves. **(A)** The stolon of *G. pentaphyllum*. **(B, C)** Shoot apex and young leaf for protoplast preparation and tissue section staining. **(D-F)** Leaf cross-sections and shoot apex longitudinal sections stained with Safranin-O and Fast Green. The red boxes represent the trichomes. **(G, I)** Protoplasts of shoot apexes and leaves under bright light. **(H, J)** Protoplasts of shoot apexes and leaves stained with fluorescein diacetate. Live cells can emit green fluorescence (488 nm) after staining with fluorescein diacetate.

PRJNA1030183). Following the acquisition of scRNA-seq datasets, Alevin was employed for quantification. After filtering, a high-quality dataset, comprising 21,103 and 29,395 cells from shoot apexes and leaves, respectively, was obtained. In the leaves, the median number of genes per cell was 2,249 and the median number of UMIs per cell was 4,510; in the shoot apexes, the median number of genes and UMIs per cell were 2,435 and 4,849, respectively (Supplementary Table 2). For each tissue, principal component analysis (PCA) of the 3000 most highly variable genes were performed for the dimensional reduction of the scRNA-seq data. Unsupervised clustering identified of 13 and 14 distinct cell clusters in shoot apexes and leaves, respectively. These clusters were visualized using two-dimensional uniform manifold approximation and projection (UMAP) plots (Figure 2).

10.3389/fpls.2024.1394587

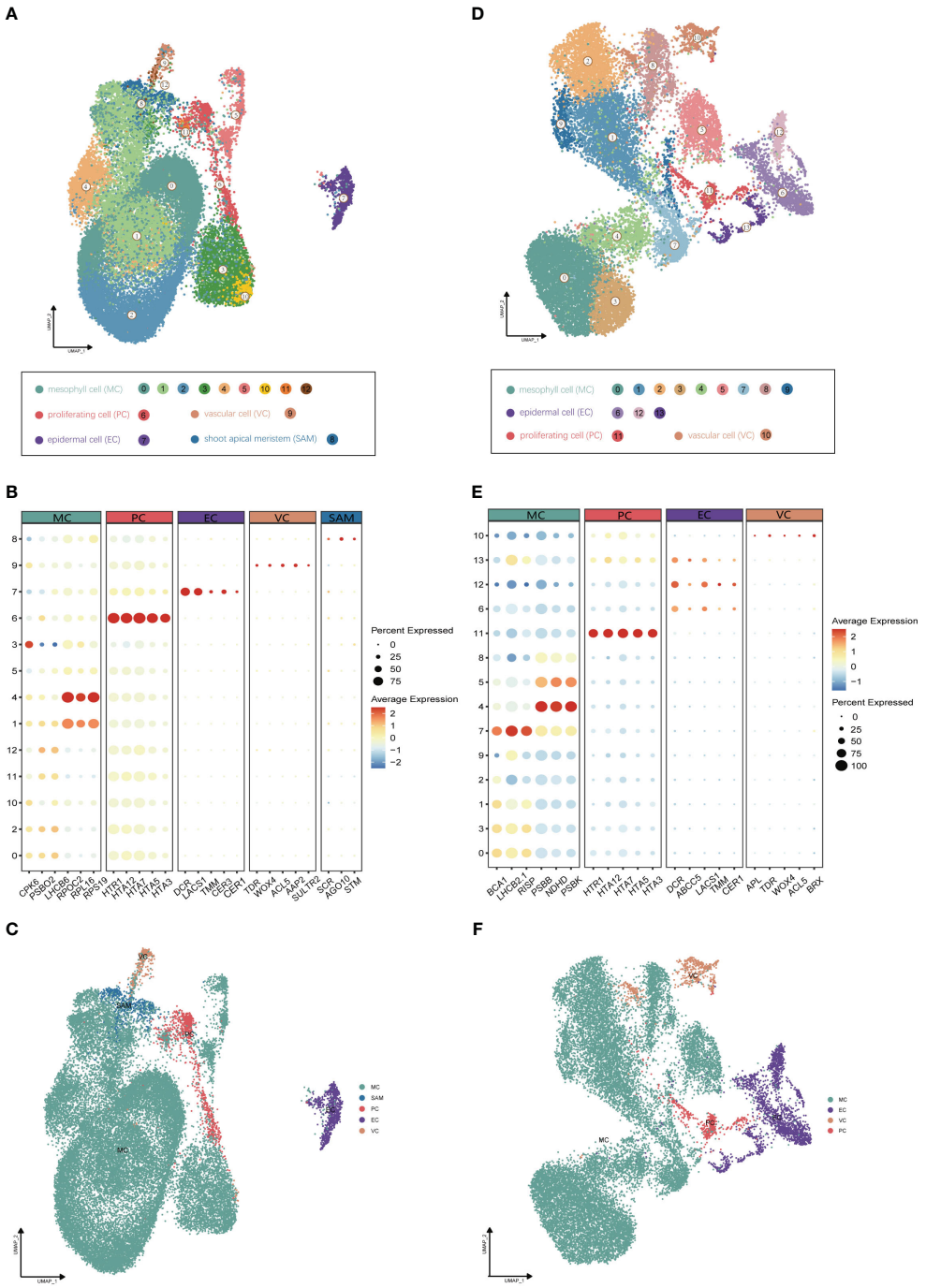


FIGURE 2
Identification of cell clusters from (G) *pentaphyllum* shoot apexes and leaves. **(A)** UMAP visualization of 13 cell clusters in shoot apexes. Each dot represents an individual cell; the total number of cells (n) is 29,395. Colors indicate different cell clusters. **(B)** Expression patterns of representative cluster-specific marker genes in shoot apexes. Dot diameter represents the proportion of cluster cells expressing a given gene. Full names of the selected genes are provided in Supplementary Table 3. **(C)** UMAP visualization of five broad populations of shoot apexes. Colors represent different population types. **(D)** UMAP visualization of 14 cell clusters in leaves. Each dot represents an individual cell; the total number of cells (n) is 21,103. Colors indicate different cell clusters. **(E)** Expression patterns of representative cluster-specific marker genes in leaves. Dot diameter represents the proportion of cluster cells expressing a given gene. Full names of the selected genes are provided in Supplementary Table 3. **(F)** UMAP visualization of four broad populations in leaves. Colors represent different population types.

The annotation of cell clusters was based on cell-type markers, whose functions have been well characterized in *Arabidopsis* (Supplementary Table 3). To ensure annotation accuracy, more than three cell-type markers were utilized for each cell type in *G. pentaphyllum* shoot apexes and leaves. For the shoot apexes, the gene markers specific to MCs, including Oxygen-evolving enhancer protein 1-2 (*PSBO2*, Gynpe.TU.chr03.1614) and Chlorophyll a-b binding protein (*LHCB6*, Gynpe.TU.chr10.262), exhibited a

concentrated expression pattern within Clusters 0-5 and 10-12 (Figures 2A–C) (Murakami et al., 2005; Chen et al., 2018). These clusters collectively delineated the population identified as MCs. Similarly, epidermal cell-specific genes, including *TOO MANY MOUTHS* (*TMM*, Gynpe.TU.chr04.1612) and *ECERIFERUM1* (*CER 1*, Gynpe.TU.chr06.3193) (Bhave et al., 2009; Bernard et al., 2012), were prominently expressed in Cluster 7, which was designated as epidermal cells (ECs) (Figure 2B). Cluster 9 was annotated as vascular cells (VCs), in which genes involved in governing vascular procambium differentiation and xylem specification, such as *WUSCHEL-related homeobox 4* (*WOX4*, Gynpe.TU.chr06.1767) and thermospermine synthase ACAULIS5 (*ACL5*, Gynpe.TU.chr06.1882), were predominantly expressed (Figure 2B) (Muniz et al., 2008; Ji et al., 2010). Moreover, Cluster 6 was designated as proliferating cells (PCs) due to the enrichment of genes associated with histone function, exemplified by histone H2A.Z (*HTA3*) (Figure 2B). Finally, Cluster 8 was designated as shoot apical meristem cells (SAMs) because *SCARECROW* (*SCR*, Gynpe.TU.chr07.931) and proteins such as the protein argonaute 10 (*AGO10*, Gynpe.TU.chr02.1132), which is essential for stem cell function, were markedly overrepresented in this cluster (Figure 2B) (Sabatini et al., 2003; Liu et al., 2009). In the leaves, the same method was used to annotate cell types. Finally, Clusters 0-5 and 7-9 were collectively identified as mesophyll cells (MCs), while Clusters 6, 12, and 13 were designated as ECs, and Clusters 10 and 11 were associated with VCs and PCs, respectively (Figures 2D–F). The cell identifies of ECs, VCs, and MCs were confirmed by RNA *in situ hybridization* (RIH) experiments using *G. pentaphyllum* leaves (Figure 3).

We observed significant cellular heterogeneity within the *G. pentaphyllum* shoot apexes and leaves, particularly for MCs (Figures 2A–D). For instance, MCs were segregated into 8 distinct clusters in leaves, exhibiting the highest heterogeneity among all the populations (Figure 2A), a phenomenon also reported in *Catharanthus roseus* leaves (Sun et al., 2023). Moreover, Gene Ontology (GO) enrichment analysis indicated substantial transcriptomic heterogeneity in the same cell type between shoot apexes and leaves (Supplementary Figure 1). Specifically, MCs from shoot apexes exhibited enrichment in processes associated with "ribosomal ribosome" and "mRNA binding", reflecting their involvement in cell proliferation and growth during the rapid development of shoot apexes (Supplementary Figure 1A). In contrast, signature genes in MCs from leaves were enriched for processes related to "photosynthesis" and "chloroplast thylakoid membrane", underscoring their role in photosynthesis (Supplementary Figure 1B). Additionally, shoot apex ECs displayed enrichment for processes related to "lipid catabolic process" and "lipid transport", indicating robust lipid metabolism and transport in the shoot apexes (Supplementary Figure 1C). On the other hand, the enriched gene signals in ECs from leaves were associated with "response to stimulus" and "response to stress", aligning with the role of leaves as sensory organs (Supplementary Figure 1D).

3.2 Spatiotemporal distribution of gypenoside biosynthesis in *G. pentaphyllum* shoot apexes and leaves

Gypenosides originate from C5 isopentenyl building blocks produced through two primary pathways: the mevalonic acid pathway (MVA) and the methylerythritol phosphate pathway (MEP) (Figure 4A) (Vranova et al., 2013). Within both shoot apexes and leaves of *G. pentaphyllum*, MEP pathway genes demonstrated elevated expression in MCs compared to other cell types, while MVA pathway genes exhibited higher expression levels in both MCs and ECs (Figures 4B, C). Notably, MEP pathway genes exhibited varying expression levels in subtypes of MCs, with the highest expression observed in Cluster 7 of MCs in leaves. In contrast, this pattern was not observed in the shoot apexes (Supplementary Figure 2). This observation suggested that distinct MC subtypes possess unique capabilities for C5 isopentenyl unit biosynthesis. Furthermore, a tissue-level differential expression analysis of MEP and MVA pathway genes in shoot apexes and leaves was conducted. With the exception of the deoxyxylulose 5-phosphate synthase (DXS) and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) genes of the MEP pathway, the expression levels of the other genes in the shoot apex were greater than those in the leaves (Supplementary Figure 3). These findings suggest that C5 isopentenyl unit biosynthesis is potentially more active in the shoot apexes than in leaves of *G. pentaphyllum*.

The expression of the genes encoding isopentenyl diphosphate isomerase (IDI), which catalyzes the interconversion of the C5 units between isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), was also elevated in MCs (Figures 4B, C). IPP and DMAPP are sequentially converted to dammarenediol-II through reactions mediated by geranyl pyrophosphate synthase (GPPS), farnesyl pyrophosphate synthase (FPS), squalene synthase (SS), squalene epoxidase (SE), and dammarenediol-II synthase (GpDS) (Figure 4A). *SS*, *SE* and *GpDS* exhibited preferential expression in MCs of both shoot apexes and leaves (Figures 4D, E). Both *GPPS* and *FPS* showed elevated transcription in MCs and SAMs of shoot apexes, whereas in leaves, *GPPS* expression was elevated in MCs and VCs, while *FPS* transcripts were enriched in MCs and ECs (Figures 4B, C). The dammarenediol-II skeleton underwent subsequent modifications to yield various gypenosides, a process mediated by cytochrome P450s (CYP450s) and UDP-glucosyltransferases (UGTs) (Figure 4A). While no CYP450s involved in gypenoside biosynthesis have been identified, five UGTs have been characterized thus far. *UGT71AW1*and *UGT71V3* demonstrated relatively greater expression in MCs in both shoot apexes and leaves, while other *UGT* transcripts were enriched in distinct cell types (Figures 4B, C). In summary, the majority of genes involved in gypenoside biosynthesis exhibited increased expression in the MCs of both shoot apexes and leaves, suggesting that MCs serve as the principal cellular sites for gypenoside biosynthesis.

10.3389/fpls.2024.1394587

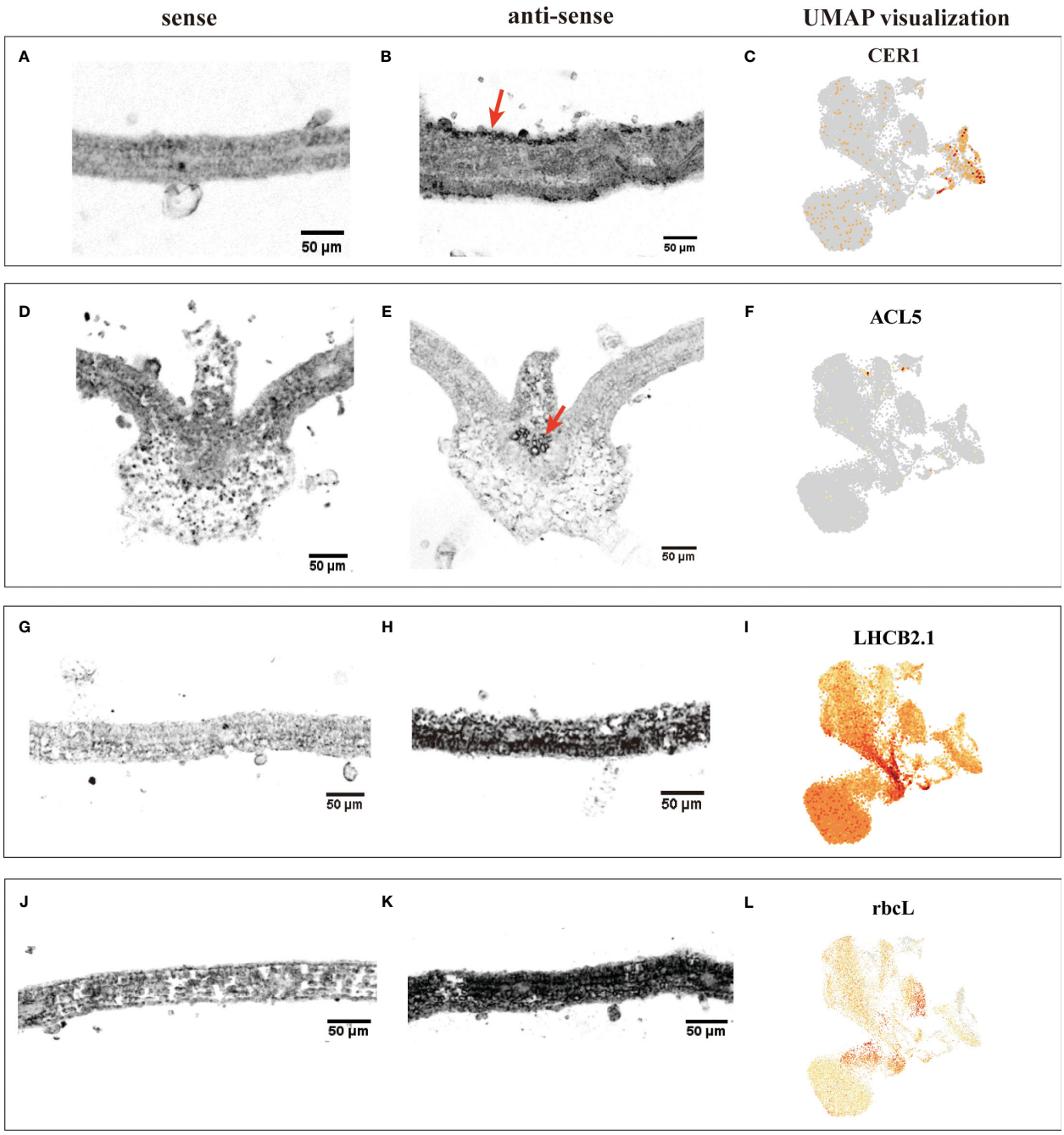


FIGURE 3
RIH validating the annotated cell identities in *G. pentaphyllum* leaves and UMAP visualization of the transcript accumulation of cell type-specific marker genes. **(A-C)** EC (*CER1*); **(D-F)**, VC (*ACL5*); **(G-I)**, MC (*LHCB2.1*); **(J-K)**, MC (*rbcL*). In the hybridized sections, the sense was the control, the black indicated the location of the signal in the anti-sense, the red arrows were the hybridization signal of ECs and VCs. On the UMAP plot, the color intensity represents the relative transcript expression level for the indicated gene in each cell.

3.3 Cell type-specific expression of TEs in shoot apexes and leaves

TE expression in single cells from shoot apexes and leaves was quantified using the scTE pipeline, generating gene/TE expression matrices for subsequent analysis. The resulting cell clusters in the shoot apexes and leaves were classified into four and five distinct cell types, respectively, consistent with the cell type outputs based solely on gene expression (Figures 5A, B; Supplementary Figure 4). The Sankey diagram illustrates that the identities of cells within each cell type were not significantly altered regarding both gene and TE expression (Supplementary Figure 5). In total, 17,376 expressed TE families were detected in shoot apexes, including 5,778 Class I TEs and 11,218 Class II TEs, while 15,404 TE expressed families were identified in leaves, including 5,019 Class I TEs and 10,037 Class II TEs (Supplementary Tables 5, 6). Among

10.3389/fpls.2024.1394587

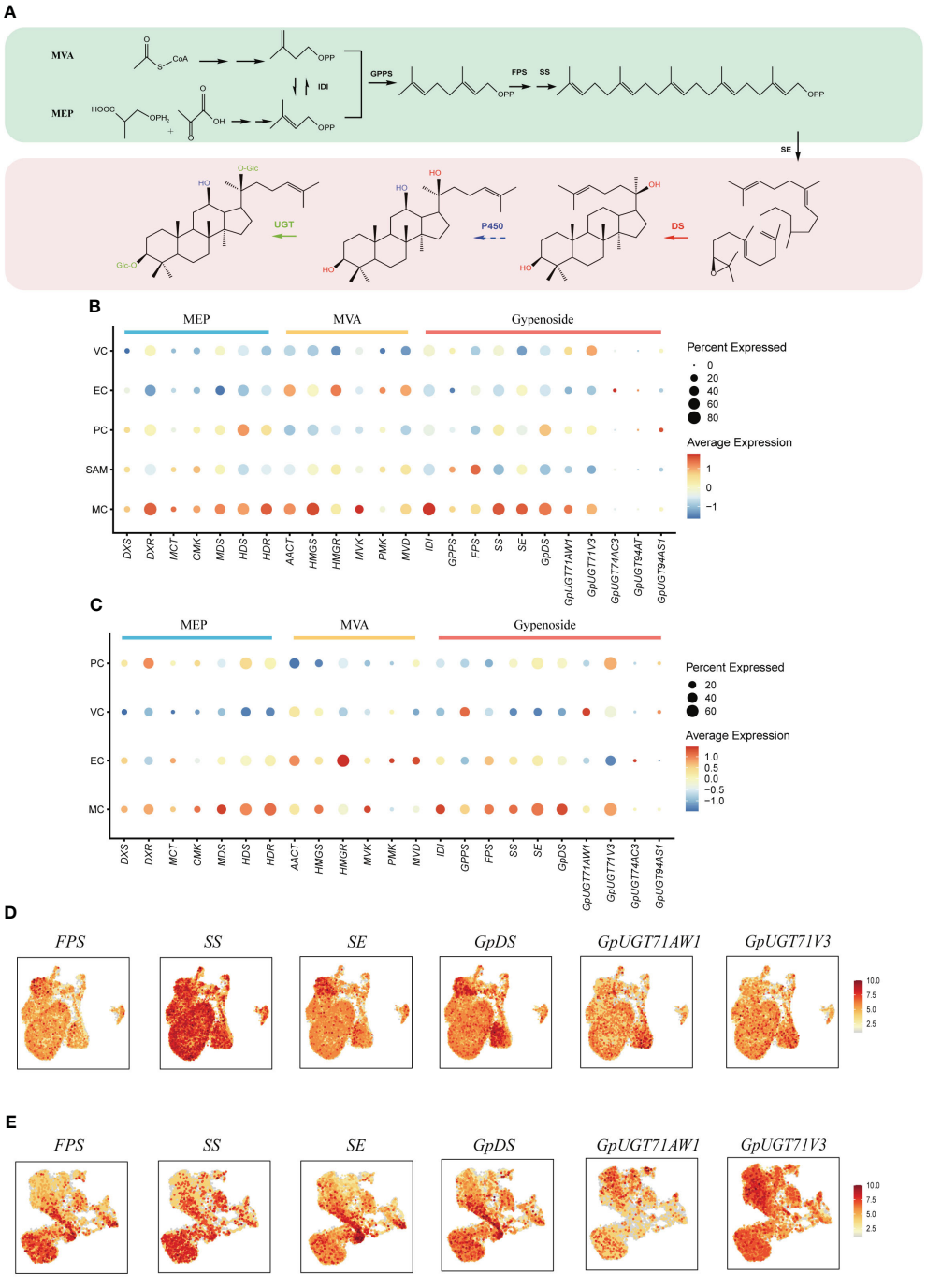


FIGURE 4
Spatial distribution of gypenoside biosynthesis in *G. pentaphyllum* shoot apexes and leaves. **(A)** The proposed gypenoside biosynthetic pathway. The full names of the selected genes are provided in Supplementary Table 4. **(B, C)** Cell type specificity of gypenoside biosynthesis in shoot apexes and leaves, respectively. **(D, E)** UMAP visualization of gene expression patterns involved in gypenoside biosynthesis in shoot apexes and leaves, respectively. Color intensity indicates the relative transcript level for the indicated gene in each cell.

the TEs expressed in the shoot apexes and leaves, the LTR subclass comprised the highest proportion (28%) of Class I TEs, while the MITE subclass accounted for the highest proportion of Class II TEs, 54% in the shoot apexes and 55% in the leaves, respectively. (Figures 5C, D). These results suggested that TE expression profiles at the class and subclass levels are roughly comparable between shoot apexes and leaves.

In the leaves, 21 TE subfamilies exhibited cell type-specific expression, whereas in the shoot apexes, 20 TE subfamilies exhibited cell type-specific expression (Figures 5E, F; Supplementary Tables 7, 8). Notably, we found that most of cell type-specific TE subfamilies differed even within the same cell type between shoot apexes and leaves. However, five cell type-specific TE subfamilies (mite-10328, TE1349-SO2-FAM2311, chr04.4539215.4543385-INT,

10.3389/fpls.2024.1394587

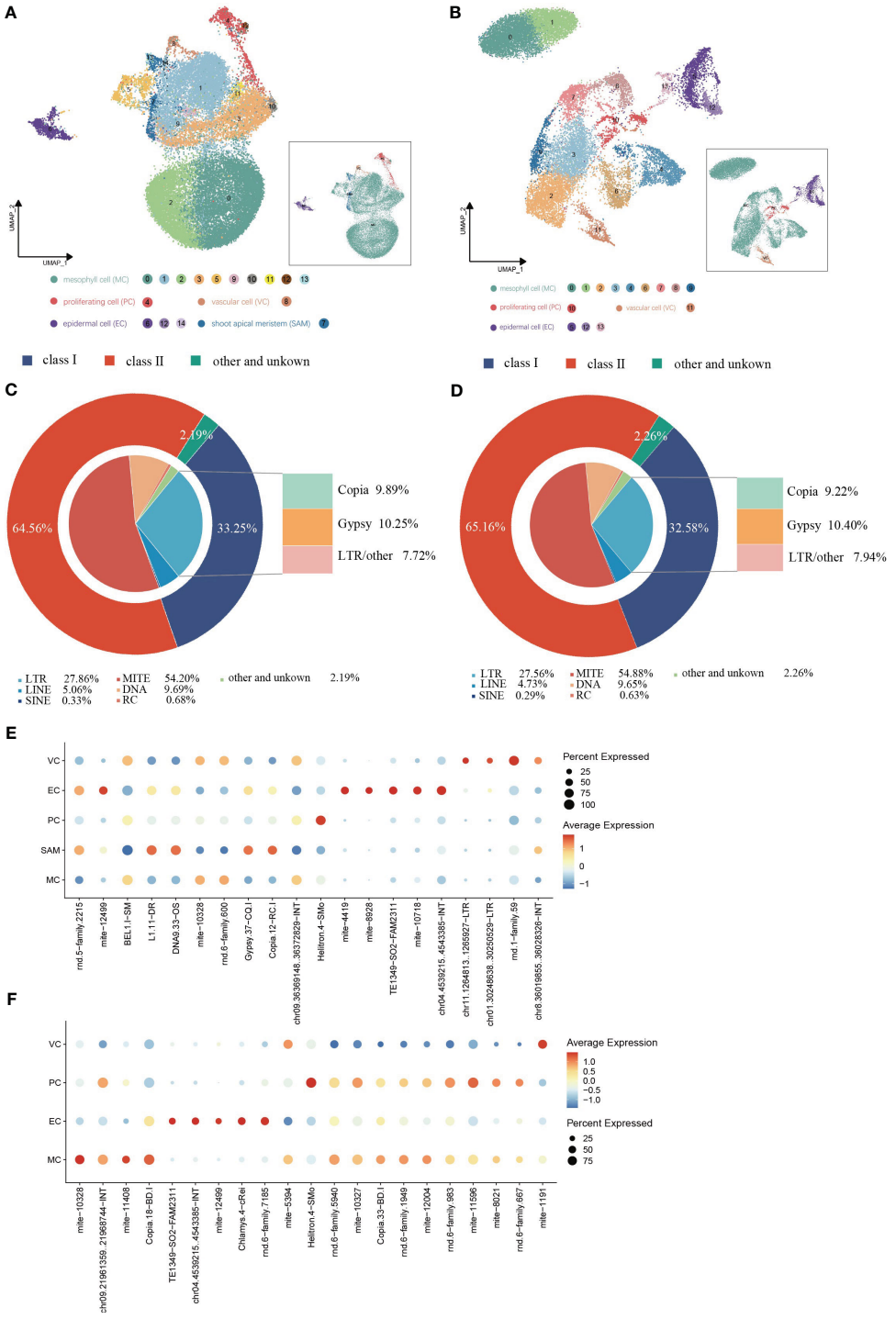


FIGURE 5
Spatial distribution of TE activity in (*G*) *pentaphyllum* shoot apexes and leaves. **(A, B)** UMAP visualization of cell clusters generated by the integrated gene/TE expression datasets from shoot apexes and leaves, respectively. Each dot represents an individual cell; colors indicate cell clusters. **(C, D)** Classification of the expressed TEs in shoot apexes and leaves, respectively. **(E, F)** Cell type-specific TE expression patterns in shoot apexes and leaves, respectively.

mite-12499, and Helitron.4-SMo) were present in both tissues, among which four TE subfamilies were specifically expressed in the same cell type across both leaves and shoot apexes. Interestingly, mite-12499 was specifically expressed in ECs and VCs in leaves, while in shoot apexes it was specifically expressed in ECs and SAMs (Figures 5E, F; Supplementary Tables 7, 8). These findings provide novel evidence for the spatiotemporal heterogeneity of TE expression within plants.

3.4 Stage-specific expression of TEs and genes involved in gypenoside biosynthesis during developmental trajectories

To establish coherent developmental trajectories, we initially integrated gene/TE expression datasets from shoot apexes and leaves, and the resultant dataset was used for cell clustering and cell identity annotation. Based on the distinctive expression patterns of cell-type markers, Clusters 10 and 11 were annotated as PCs, Cluster 14 as SAMs, cluster 15 as VCs, and Clusters 7,12, and 13 as ECs. Moreover, Clusters 0-6, 8, 9 and 16 were designated as MCs (Supplementary Figure 6).

Pseudotime trajectory analysis was applied to the EC populations (Figure 6A). We used the transcription factor ARABIDOPSIS THALIANA MERISTEM LAYER 1 (*ATML1*), which acts as a positive regulator of gibberellin (GA)-regulated epidermal gene expression, and *CER1*, which participates in the regulation of cuticle biosynthesis and wax accumulation, to distinguish between young and mature ECs (Bernard et al., 2012; Zuch et al., 2022). The developmental trajectory was initiated from Cluster EC_5, followed by EC_0, EC_4, EC_1, and EC_3, with EC_2 cells positioned at the terminus of the trajectory, according to the expression patterns of *AMTL1* and *CER3* along the developmental trajectory (Figure 6B; Supplementary Figures 7A, B). GO term analysis revealed that during the initial stages of the developmental trajectory (modules 1 and 3), there was an overrepresentation of genes associated with the regulation of meristem growth and cell growth. This observation aligned with the expected biological functions attributed to early ECs during development. Conversely, at the terminus of the trajectory, module 4 encapsulated the expression of genes associated with the abiotic stress response, aligning with the heightened sensitivity of mature epidermal cells to environmental stimuli (Figure 6C; Supplementary Table 9). The dynamic expression patterns of TEs with specifically high expression in ECs were examined throughout the course of EC development (Supplementary Figure 8). Specifically, DNA9.33.OS, Gypsy.37-CQ.1 and L1.11.DR exhibited downregulated expression in the late stages of EC development, while Chlamys.4-cRei, mite-2836 and Helitron.4-SMo exhibited upregulated expression in the late stages of EC development (Figures 6D, E; Supplementary Figure 8). These findings suggest that these TEs may regulate epidermal cell fate specification. We also analyzed the dynamic expression patterns of genes involved in the gypenoside biosynthetic pathway during epidermal development (Supplementary Figure 9). Intriguingly, the key genes (*FPS*, *SS*, *SE*, and *GpDS*) responsible for synthesizing the triterpene skeleton dammarenediol-II were expressed at higher levels in younger ECs, and progressively diminished along the developmental trajectory. Conversely, the expression of *GpUGT71V3* exhibited slight upregulation in the later stages of EC development (Figure 6F). Moreover, the expression of *GpUGT71AW1* exhibited slight upregulation in the early stages of EC development and then decreased in the middle stages of EC development but exhibited upregulation in the late stages of EC development.

Pseudotime analysis was also applied to the VC and SAM populations to determine the cell differentiation states during VC development (Figure 6G; Supplementary Figure 7C, D). First, the VCs were reassigned into two subclusters, among which Cluster 14_0 mainly expressed orthologues of *Arabidopsis* xylem cell markers, namely, *WOX4*, *TDR*, *ATHB8* and *VINI2* (Hirakawa et al., 2010; Fukuda and Ohashi-Ito, 2019) (Figure 6H; Supplementary Figure 10A). The SAM population (Cluster15) was reassigned into two subclusters, and the 15_1 subcluster was highly expressed SAM marker genes (Supplementary Figure 10B). We extracted subcluster 14_0 with subcluster 15_1 to construct a putative developmental trajectory from the SAM to the xylem (Figure 6G). GO enrichment of DEGs on the developmental trajectory was performed. The enriched gene signal at the beginning of the pseudotime axis (modules 3 and 4) was consistent with the functions of cell proliferation and nucleic acid replication in primary growth (Figure 6I; Supplementary Table 10). At the end of the trajectory, module 1 captured the expression of genes involved in material transport, the abiotic stress response and lignin catabolism, consistent with the function of mature xylem (Figure 6I; Supplementary Table 10). Interestingly, we found that the TE subfamilies Copia.12-RC.I DNA9.33-OS and L1.11-DR were highly expressed in the early stages of cell development, while seven TE subfamilies from mite subclass were high expressed in the later stage of development (Figures 6J, K; Supplementary Figure 11). In addition, we also examined the variation in the expression of genes involved in the biosynthesis of gypenosides (Supplementary Figure 12). The key genes *SS*, *SE* and *DS* involved in the synthesis of the triterpenoid skeleton were highly expressed in the early stage of development and then sharply decreased (Figure 6L). Only *GpUGT71V3* and *GPPS* were upregulated along the trajectory and were highly expressed in mature xylem cells (Supplementary Figure 12). Taken together, the results of pseudotime analysis provided novel insights into the dynamic process of xylem development and variations in the expression of specific TEs and genes involved in gypenoside biosynthesis during the transition of cell states.

4 Discussion

Over the past five years, plant science has been revolutionized by enabling transcriptome profiling at single-cell resolution via scRNA-seq, particularly in the exploration of cellular heterogeneity and developmental trajectories (Bawa et al., 2023). However, its applications in studying plant specialized metabolism and TE activities remain limited. In this study, by constructing single-cell atlases of *G. pentaphyllum* shoot apexes and leaves, we revealed the spatiotemporal distribution of gypenoside biosynthesis. More importantly, we provide the first landscape of TE activities in plant tissues at single-cell resolution. We found an imbalanced distribution of TE activities among different cell types and during leaf development.

Cell clustering and annotation constitute foundational steps in the analysis of scRNA-seq data. However, accurately annotating all cell types in the investigated tissues remains challenging, particularly for nonmodel plants. In this study, the absence of trichome cells limits the ability to construct a more comprehensive transcriptomic profile. A relatively large cell size (approximately

10.3389/fpls.2024.1394587

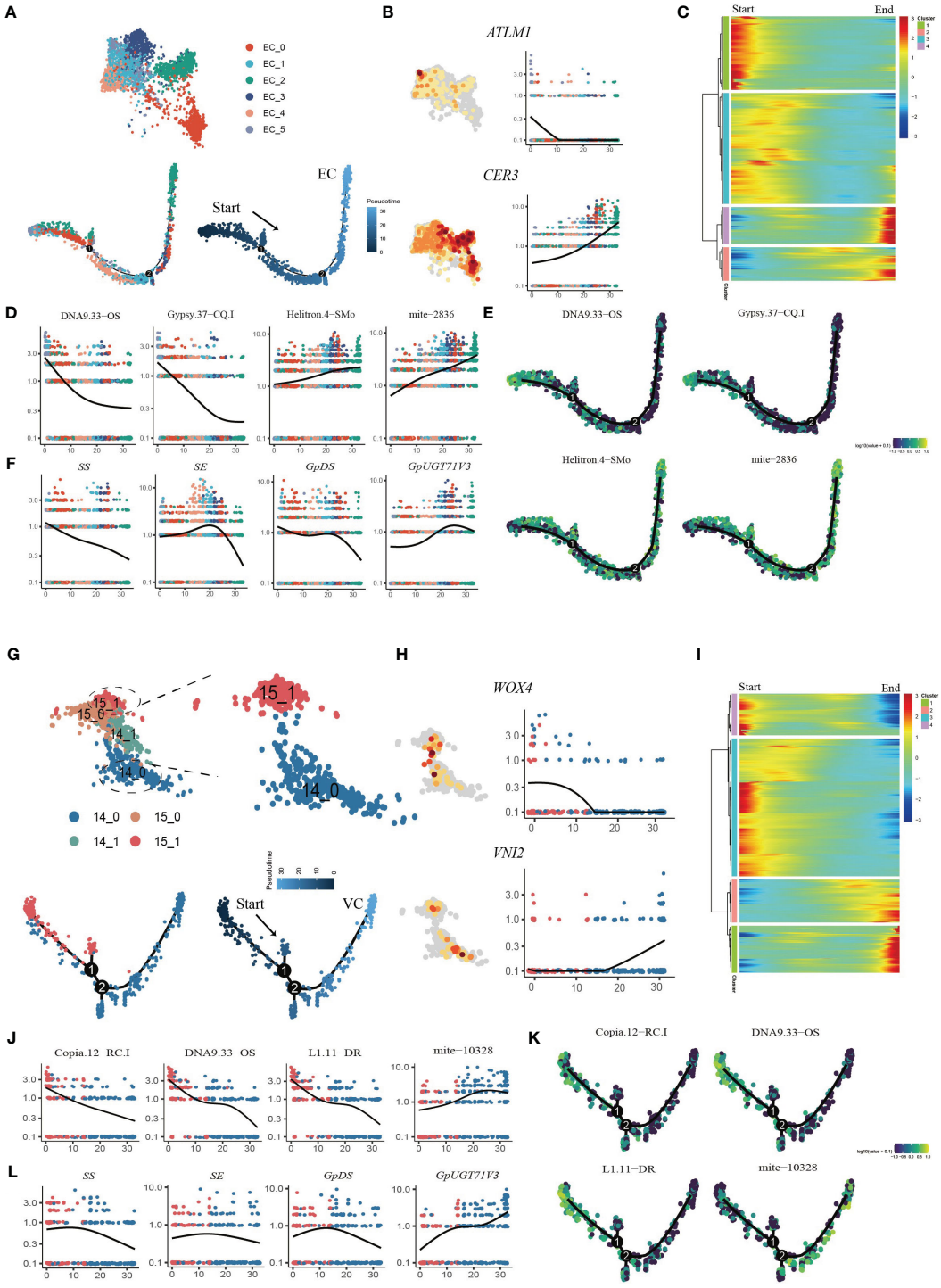


FIGURE 6
Reconstruction of the developmental trajectories of ECs and VCs. **(A)** UMAP visualization and cells distribution of ECs along the pseudotime trajectory. Each dot indicates a single cell, and the color of the cell indicates cell identities (below left). The color on the dot indicates the pseudotime score (below right). **(B)** UMAP plots of transcript accumulation and gene expression kinetics during EC pseudotime progression for marker genes, the full names of the selected genes are provided in Supplementary Table 11. **(C)** Heatmap showing the expression of pseudotime-dependent genes over the pseudotime trajectory of EC subclusters. **(D)** TE expression kinetics during EC pseudotime progression for representative TEs. **(E)** xpression patterns of representative TE are shown over the course of EC pseudo-time. **(F)** Gene expression kinetics during EC pseudotime progression for representative genes. **(G)** UMAP visualization and cells distribution of SAMs and VCs along the pseudotime trajectory. Each dot indicates a single cell, and the color of the cell indicates cell identities (below left). The color on the dot indicates the pseudotime score (below right). **(H)** UMAP plots of transcript accumulation and gene expression kinetics during VC pseudotime progression for marker genes, the full names of the selected genes are provided in Supplementary Table 11. **(I)** Heatmap showing the expression of pseudotime-dependent genes over the pseudotime trajectory of VC subclusters. **(J)** TE expression kinetics during VC pseudotime progression for representative TEs. **(K)** Expression patterns of representative TE are shown over the course of VC pseudotime. **(L)** Gene expression kinetics during VC pseudotime progression for representative genes.

150 μm) may lead to fragmentation of these cells, preventing them from passing through the cell strainer during protoplast preparation (Figure 1E). A similar mechanism has been proposed to explain the loss of laticifer cells during the protoplasting of C. roseus leaves (Sun et al., 2023). Despite the progress in plant science, the availability of specialized annotation tools and reference datasets for plants lags behind that of animals. Currently, multiple software tools exist for automated cell type annotation of scRNA-seq data, such as SingleR (Aran et al., 2019), scVI (Lopez et al., 2018), scDHA (Tran et al., 2021), and CIForm (Xu et al., 2023a). These tools primarily rely on reference data for animals like humans and mice, leaving a gap in annotation resources for plants. As plant science progresses, it is anticipated that more specialized annotation tools and reference datasets for plants will emerge, catering to the needs of plant research. In addition, the effective isolation of plant protoplasts is often influenced by various factors, including cell size, cell shape, the relative position of cells within tissues (surface or inner layer), the biochemical composition of the cell wall, and the developmental stage of the cells are situated (Shaw et al., 2021). Single-nucleus RNA-seq may provide an alternative approach to capture additional cell types through nucleus isolation rather than protoplasting. Nucleus preparation is simpler and less impacted by cellular characteristics or positional context within tissues. The lack of suitable cell marker genes remains a significant challenge in cell annotation for nonmodel plants. Spatial transcriptomics can provide additional insights into cell type annotation due to its ability to capture spatial information (Tian et al., 2023). Moreover, with the assistance of spatial information, subtypes such as palisade versus sponge mesophyll cells or upper versus lower epidermal cells could be resolved (Xia et al., 2022).With the rapid advancement of single-cell transcriptomics technology, emerging tools have opened new avenues for diversified data analysis and utilization. For instance, STGRNS effectively infers cell-specific gene regulatory networks through the analysis of single-cell transcriptomic data (Xu et al., 2023b), while scmFormer delves into the intricate relationships between gene expression and protein levels by integrating single-cell proteomic data (Xu et al., 2024). Looking ahead, the integration of single-cell transcriptomics with single-cell multi-omics technology will become a trend, providing us with more comprehensive and profound cellular analysis tools.

Compartmentalization of specialized metabolism is a widespread phenomenon across the plant kingdom that occurs at multiple hierarchical levels, ranging from the molecular to the organ level (Watkins and Facchini, 2022). This organizational strategy provides distinct advantages to plants, including the enhancement of metabolic output, alleviation of enzyme inhibition, and elimination of autotoxicity. Notably, among these levels of compartmentalization, the exploration of multicellular compartmentalization has been limited. ScRNA-seq has emerged as a potent tool for dissecting the multicellular compartmentalization of plant specialized metabolism. Utilizing this technology, it has been demonstrated that the vinblastine biosynthetic pathway in *C. roseus* leaves can be compartmentalized into distinct cell types, namely IPAP cells, ECs, and ICs (Li et al., 2023; Sun et al., 2023). However, we analyzed the expression patterns of the characterized genes involved in the gypenosides biosynthesis pathway at single-cell resolution to explore the spatial distribution of this pathway in *G. pentaphyllum* shoot apexes and leaves revealed that gypenoside biosynthesis is predominantly localized in MCs. Similarly, taxol biosynthesis was observed in MCs of *Taxus* leaves, as evidenced by the scRNA-seq data analysis (Zhan et al., 2023). These findings suggest that the biosynthesis of both gypenoside and taxol does not involve multicellular compartmentalization. Consequently, the extent to which specialized metabolic pathways are compartmentalized at the cellular level in the plant kingdom remains to be explored.

In this study, we conducted the first examination of single-cell TE expression in plants. TEs act as cis-regulatory elements and play pivotal roles in the regulation of gene expression (Marand et al., 2021). Furthermore, TE transcripts can also regulate host gene expression and reshape transcriptomes. Compared to those in animals, investigations into TE expression in plants have been scarce and confined to tissue-level analyses. Insertion time analysis unequivocally established the presence of a substantial population of young and transcriptionally active TEs within the *G. pentaphyllum* genome. Motivated by this discovery, we investigated TE expression at single-cell resolution in *G. pentaphyllum* shoot apexes and leaves. Our findings revealed that distinct TE families are expressed themselves in specific cell types, with certain TEs exhibiting dynamic expression patterns during the development of ECs and VCs. The precise mechanisms through which TE expression influences developmental processes warrant thorough investigation. Notably, we observed variation in the expression of genes involved in gypenoside biosynthesis during EC and VC development. However, whether TE expression participates in regulating pathway gene expression should be further explored. In conclusion, our study not only reveals the intricate landscape of TE expression at the single-cell level in plants but also opens avenues for investigating the roles of TE expression in regulating cell differentiation and development, as well as in specialized metabolism.

5 Conclusions

In this study, we generated single-cell transcriptome atlases of *G. pentaphyllum* leaf and shoot apexes. By analysis of the expression patterns of genes involved in gypenoside biosynthesis across distinct cell types, we observed that a majority of the genes associated with the gypenoside biosynthesis exhibited heightened expression in MCs. Additionally, we devised a novel single-cell transcriptome atlas of *G. pentaphyllum* by integrating gene and TEs expression data. Our analysis of TE expression profiling identified some TE families that are specific to cell types, tissues, or developmental stages, which indicated the potentially pivotal roles in cell differentiation and development. The scRNA-seq data presented in this study offers a valuable resource for investigating plant physiology and understanding the spatiotemporal distribution of specialized metabolism in this species.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA1064230, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1030183.

Author contributions

RL: Writing – original draft, Methodology, Validation. KD: Writing – original draft, Data curation, Formal Analysis, Visualization. CZ: Writing – original draft, Data curation, Formal Analysis, Visualization. XS: Writing – original draft. LY: Resources, Writing – original draft. SW: Resources, Writing – original draft. ZL: Investigation, Writing – original draft. ZS: Investigation, Writing – original draft. JW: Project administration, Writing – original draft. YL: Writing – original draft, Methodology, Software. BG: Project administration, Supervision, Writing – original draft. CS: Writing – original draft, Investigation, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the CAMS Innovation Fund for Medical Sciences (2022-I2M-2-001). The funders had no role in the design of the study; in the collection, analyses or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1394587/full#supplementary-material

References

Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass–a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 15, 1329–1330. doi: 10.1093/bioinformatics/btp084

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172. doi: 10.1038/s41590-018-0276-y

Bawa, G., Liu, Z., Yu, X., Tran, L. P., and Sun, X. (2023). Introducing single cell stereo-sequencing technology to transform the plant transcriptome landscape. *Trends Plant Sci.* 23, 1360–1385. doi: 10.1016/j.tplants.2023.10.002

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J. D., et al. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that arabidopsis eceriferum1 and eceriferum3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* 24, 3106–3118. doi: 10.1105/tpc.112.099796

Bhave, N. S., Veley, K. M., Nadeau, J. A., Lucas, J. R., Bhave, S. L., and Sack, F. D. (2009). Too many mouths promotes cell fate progression in stomatal development of *Arabidopsis* stems. *Planta* 229, 357–367. doi: 10.1007/s00425-008-0835-9

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

Chen, Y. E., Ma, J., Wu, N., Su, Y. Q., Zhang, Z. W., Yuan, M., et al. (2018). The roles of *Arabidopsis* proteins of lhcb4, lhcb5 and lhcb6 in oxidative stress under natural light conditions. *Plant Physiol. Biochem.* 130, 267–276. doi: 10.1016/j.plaphy.2018.07.014

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 28, 9451–9457. doi: 10.1073/pnas.1921046117

Fueyo, R., Judd, J., Feschotte, C., and Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* 23, 481–497. doi: 10.1038/s41580-022-00457-y

Fukuda, H., and Ohashi-Ito, K. (2019). Vascular tissue development in plants. *Curr. Topics Dev. Biol.* 131, 141–160. doi: 10.1016/bs.ctdb.2018.10.005

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M.3rd., Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587. doi: 10.1016/j.cell.2021.04.048

He, J., Babarinde, I. A., Sun, L., Xu, S., Chen, R., Shi, J., et al. (2021). Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scte. *Nat. Commun.* 12, 1456. doi: 10.1038/s41467-021-21808-x

Hirakawa, Y., Kondo, Y., and Fukuda, H. (2010). Tdif peptide signaling regulates vascular stem cell proliferation via the *WOX4* homeobox gene in *Arabidopsis*. *Plant Cell* 22, 2618–2629. doi: 10.1105/tpc.110.076083

Ji, X., Shen, Y., and Guo, X. (2018). Isolation, structures, and bioactivities of the polysaccharides from *Gynostemma pentaphyllum* (thunb.) Makino: a review. *Biomed. Res. Int.* 2018, 6285134. doi: 10.1155/2018/6285134

Ji, J., Strable, J., Shimizu, R., Koenig, D., Sinha, N., and Scanlon, M. J. (2010). *WOX4* promotes procambial development. *Plant Physiol.* 152, 1346–1356. doi: 10.1104/pp.109.149641

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. doi: 10.1038/s41592-019-0619-0

Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21, 721–736. doi: 10.1038/s41576-020-0251-y

Le, D. D., Kim, W., Lim, S., Kim, S. C., and Choi, G. (2021). Identification of three groups of ginsenoside biosynthetic udp-glycosyltransferases from *Gynostemma pentaphyllum*. *Plant Sci.* 313, 111069. doi: 10.1016/j.plantsci.2021.111069

Li, C., Wood, J. C., Vu, A. H., Hamilton, J. P., Rodriguez, L. C., Payne, R., et al. (2023). Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. *Nat. Chem. Biol.* 19, 1031–1041. doi: 10.1038/s41589-023-01327-0

Liu, Q., Yao, X., Pi, L., Wang, H., Cui, X., and Huang, H. (2009). The argonaute10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing mir165/166 in *Arabidopsis*. *Plant J.* 58, 27–40. doi: 10.1111/j.1365-313X.2008.03757.x

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15, 1053–1058. doi: 10.1038/s41592-018-0229-2

Lopez-Anido, C. B., Vaten, A., Smoot, N. K., Sharma, N., Guo, V., Gong, Y., et al. (2021). Single-cell resolution of lineage trajectories in the *Arabidopsis* stomatal lineage and developing leaf. *Dev. Cell* 56, 1043–1055. doi: 10.1016/j.devcel.2021.03.014

Marand, A. P., Chen, Z., Gallavotti, A., and Schmitz, R. J. (2021). A cis-regulatory atlas in *Maize* at single-cell resolution. *Cell* 184, 3041–3055. doi: 10.1016/j.cell.2021.04.014

McGinnis, C. S., Murrow, L. M., and Gartner, Z. J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8, 329–337. doi: 10.1016/j.cels.2019.03.003

Muniz, L., Minguet, E. G., Singh, S. K., Pesquet, E., Vera-Sirera, F., Moreau-Courtois, C. L., et al. (2008). Acaulis5 controls *Arabidopsis* xylem specification through the prevention of premature cell death. *Development* 135, 2573–2582. doi: 10.1242/dev.019349

Murakami, R., Ifuku, K., Takabayashi, A., Shikanai, T., Endo, T., and Sato, F. (2005). Functional dissection of two *Arabidopsis* psbo proteins: psbo1 and psbo2. *FEBS. J.* 272, 2165–2175. doi: 10.1111/j.1742-4658.2005.04636.x

Nguyen, N. H., Ha, T., Yang, J. L., Pham, H., and Oh, W. K. (2021). Triterpenoids from the genus gynostemma: chemistry and pharmacological activities. *J. Ethnopharmacol.* 268, 113574. doi: 10.1016/j.jep.2020.113574

Niu, Y., Yan, W., Lv, J., Yao, W., and Yu, L. L. (2013). Characterization of a novel polysaccharide from tetraploid *Gynostemma pentaphyllum* makino. *J. Agric. Food Chem.* 61, 4882–4889. doi: 10.1021/jf400236x

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402

Rodriguez-Quiroz, R., and Valdebenito-Maturana, B. (2022). Solote for improved analysis of transposable elements in single-cell RNA-seq data using locus-specific expression. *Commun. Biol.* 5, 1063. doi: 10.1038/s42003-022-04020-5

Sabatini, S., Heidstra, R., Wildwater, M., and Scheres, B. (2003). Scarecrow is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.* 17, 354–358. doi: 10.1101/gad.252503

Seyfferth, C., Renema, J., Wendrich, J. R., Eekhout, T., Seurinck, R., Vandamme, N., et al. (2021). Advances and opportunities in single-cell transcriptomics for plant research. *Annu. Rev. Plant Biol.* 72, 847–866. doi: 10.1146/annurev-arplant-081720-010120

Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. *Mol. Plant* 14, 115–126. doi: 10.1016/j.molp.2020.10.012

Srivastava, A., Malik, L., Smith, T., Sudbery, I., and Patro, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol.* 20, 65. doi: 10.1186/s13059-019-1670-y

Su, C., Li, N., Ren, R., Wang, Y., Su, X., Lu, F., et al. (2021). Progress in the medicinal value, bioactive compounds, and pharmacological activities of *Gynostemma pentaphyllum*. *Molecules* 20, 6249. doi: 10.3390/molecules26206249

Sun, S., Shen, X., Li, Y., Li, Y., Wang, S., Li, R., et al. (2023). Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism. *Nat. Plants* 9, 179–190. doi: 10.1038/s41477-022-01291-y

Tian, L., Chen, F., and Macosko, E. Z. (2023). The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* 41, 773–782. doi: 10.1038/s41587-022-01448-2

Tran, D., Nguyen, H., Tran, B., Vecchia, C., Luu, H. N., and Nguyen, T. (2021). Fast and precise single-cell data analysis using a hierarchical autoencoder. *Nat. Commun.* 15, 1029. doi: 10.1038/s41467-021-21312-2

Vranova, 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, M., Li, H., Liu, W., Cao, H., Hu, X., Gao, X., et al. (2020). Dammarane-type leads panaxadiol and protopanaxadiol for drug discovery: biological activity and structural modification. *Eur. J. Med. Chem.* 189, 112087. doi: 10.1016/j.ejmech.2020.112087

Wang, Q., Wu, Y., Peng, A., Cui, J., Zhao, M., Pan, Y., et al. (2022). Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves. *Plant Biotechnol. J.* 20, 2089–2106. doi: 10.1111/pbi.13891

Watkins, J. L., and Facchini, P. J. (2022). Compartmentalization at the interface of primary and alkaloid metabolism. *Curr. Opin. Plant Biol.* 66, 102186. doi: 10.1016/j.pbi.2022.102186

Xia, K., Sun, H. X., Li, J., Li, J., Zhao, Y., Chen, L., et al. (2022). The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves. *Dev. Cell* 57, 1299–1310. doi: 10.1016/j.devcel.2022.04.011

Xu, J., Huang, D. S., and Zhang, X. (2024). ScmFormer integrates large-scale single-cell proteomics and transcriptomics data by multi-task transformer. *Adv. Sci.* 14, e2307835. doi: 10.1002/advs.202307835

Xu, J., Zhang, A., Liu, F., Chen, L., and Zhang, X. (2023a). CIForm as a Transformer-based model for cell-type annotation of large-scale single-cell RNA-seq data. *Brief Bioinform.* 24, bbad195. doi: 10.1093/bib/bbad195

Xu, J., Zhang, A., Liu, F., and Zhang, X. (2023b). STGRNS: an interpretable transformer-based method for inferring gene regulatory networks from single-cell transcriptomic data. *Bioinformatics* 39, btad165. doi: 10.1093/bioinformatics/btad165

Ye, S., Feng, L., Zhang, S., Lu, Y., Xiang, G., Nian, B., et al. (2022). Integrated metabolomic and transcriptomic analysis and identification of dammarenediol-IIsynthase involved in saponin biosynthesis in *Gynostemma longipes*. *Front. Plant Sci.* 13, 852377. doi: 10.3389/fpls.2022.852377

Zhan, X., Qiu, T., Zhang, H., Hou, K., Liang, X., Chen, C., et al. (2023). Mass spectrometry imaging and single-cell transcriptional profiling reveal the tissue-specific regulation of bioactive ingredient biosynthesis in *Taxus* leaves. *Plant Commun.* 4, 100630. doi: 10.1016/j.xplc.2023.100630

Zhang, X., Zhao, Y., Kou, Y., Chen, X., Yang, J., Zhang, H., et al. (2023). Diploid chromosome-level reference genome and population genomic analyses provide insights into gypenoside biosynthesis and demographic evolution of *Gynostemma pentaphyllum* (cucurbitaceae). *Hortic. Res.* 10, uhac231. doi: 10.1093/hr/uhac231

Zuch, D. T., Doyle, S. M., Majda, M., Smith, R. S., Robert, S., and Torii, K. U. (2022). Cell biology of the leaf epidermis: fate specification, morphogenesis, and coordination. *Plant Cell* 34, 209–227. doi: 10.1093/plcell/koab250

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 June 2024
DOI 10.3389/fpls.2024.1380157

OPEN ACCESS

EDITED BY
Xiang Pu,
Sichuan Agricultural University, China

REVIEWED BY
David Jarvis,
Brigham Young University, United States
Fengli Zhao,
China National Rice Research Institute, China

*CORRESPONDENCE
Ji-yong Wang
wangjiyong75@163.com
Yan Zeng
zyzy1221@126.com

†These authors have contributed equally to this work

RECEIVED 01 February 2024
ACCEPTED 27 May 2024
PUBLISHED 11 June 2024

CITATION
Zheng S-h, Diao Y-c, Du J, Li J-t, Zhao S, Liu M-j, Lin H-c, Zeng Y and Wang J-y (2024) Genomics and resequencing of *Fagopyrum dibotrys* from different geographic regions reveals species evolution and genetic diversity.
Front. Plant Sci. 15:1380157.
doi: 10.3389/fpls.2024.1380157

Genomics and resequencing of *Fagopyrum dibotrys* from different geographic regions reveals species evolution and genetic diversity

Si-hao Zheng[1†], Yong-chao Diao[1†], Jie Du[1], Jin-tong Li[2], Sha Zhao[1], Mei-juan Liu[1], Hui-cai Lin[2], Yan Zeng[1*] and Ji-yong Wang[1*]

[1]China National Traditional Chinese Medicine Co., Ltd, Beijing, China, [2]China Traditional Chinese Medicine Seed&Seeding, Co., Ltd, Beijing, China

Fagopyrum dibotrys, belonging to the family *Polygonaceae* and genus *Fagopyrum*, is used in traditional Chinese medicine and is rich in beneficial components, such as flavonoids. As its abundant medicinal value has become increasingly recognized, its excessive development poses a considerable challenge to wild germplasm resources, necessitating artificial cultivation and domestication. Considering these factors, a high-quality genome of *F. dibotrys* was assembled and the evolutionary relationships within *Caryophyllales* were compared, based on which 58 individual samples of *F. dibotrys* were re-sequenced. We found that the samples could be categorized into three purebred populations and regions distributed at distinct elevations. Our varieties were cultivated from the parental populations of the subpopulation in central Yunnan. *F. dibotrys* is speculated to have originated in the high-altitude Tibetan Plateau region, and that its combination with flavonoids can protect plants against ultraviolet radiation; this infers a subpopulation with a high accumulation of flavonoids. This study assembled a high-quality genome and provided a theoretical foundation for the future introduction, domestication, and development of cultivated varieties of *F. dibotrys*.

KEYWORDS

genomics$_1$, resequencing$_2$, *Fagopyrum dibotrys*$_3$, evolution$_4$, genetic diversity$_5$

1 Introduction

Fagopyrum dibotrys, commonly known as golden buckwheat, belongs to the genus *Fagopyrum* and the *Polygonaceae* family. Genus *Fagopyrum* presently comprises 15 species, including *F. tataricum*, *F. esculentum*, *F. dibotrys*, etc. The compounds contained in the genus *Fagopyrum* can be divided into six classes: flavonoids, phenolics, fagopyritols,

triterpenoids, steroids, and fatty acids (Jing et al., 2016). Flavonoids represent a substantial accumulation in the genus *Fagopyrum*, especially six major classes: quercetin, rutin, orientin, homoorientin, vitexin, and isovitexin (Joshi et al., 2020). Flavonoids can also protect plants from environmental pressure and against biotic stressors (Shen et al., 2022). Among these, rutin and quercetin are highly accumulated in *F. dibotrys*. Rutin can effectively protect against cardiovascular and cerebrovascular diseases and liver damage (Lee et al., 2013), quercetin can prevent several ailments, including cardiovascular diseases, cancer, and tumors (Xu et al., 2019), and the rhizome of *F. dibotrys* is the most abundant and primary medicinal part of the plant. In *F. dibotrys*, although the types of flavonoid compounds are relatively clear, the pathways of flavonoid biosynthesis have not yet been fully elucidated. The biosynthesis of flavonoids in plants involves several enzymatic steps. Phenylalanine is produced by the shikimate pathway and then converted to cinnamic acid through the action of phenylalanine ammonialyase (PAL). Cinnamic acid is hydroxylated at the C4 position by cinnamic acid hydroxylase (C4H) to form p-coumaric acid, which is then catalyzed into p-coumaroyl CoA by coumarin CoA ligase (4CL). Chalcone synthase (CHS) catalyzes the condensation and isomerization of p-coumaroyl CoA to produce naringenin chalcone, which is then converted to chalcone by chalcone isomerase (CHI). Naringenin undergoes hydroxylation at various positions by flavanone 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), and flavonoid 3',5'-hydroxylase (F3'5'H) to produce dihydroquercetin. Dihydroquercetin can then be further converted into rutin, quercitrin, epicatechin, catechin, and kaempferol through a series of complex reactions (Huang et al., 2022; Shen et al., 2022).

In China, this plant is classified as a second-level protected plant. Throughout history, *Fagopyrum* is considered to have originated in China, particularly in the southwestern regions bordering the provinces of Yunnan, Sichuan, and Tibet (Hunt et al., 2018). *Fagopyrum* is distributed throughout countries around the Himalayan Mountains. Unlike *F. dibotrys*, *F. tataricum*, and *F. esculentum* are often used as staple crops for human consumption. *F. dibotrys* has long been used in China as a traditional medicinal herb because of its rich array of valuable bioactive constituents such as flavonoids and phenolics. The rhizome is a crucial component of Chinese herbal medicines. During the Qing Dynasty, Bencaoshiyi reported that the rhizome could be used to treat lung ailments, rheumatism, and abdominal pain. The extract from the rhizome of *F. dibotrys* inhibits the growth of lung, liver, colon, leukocyte, and bone cancer cells (Chan, 2003), and serves as the main active ingredient in lung cancer treatment drugs known as "WeiMaiNing" capsules (Ke et al., 2021). As research on *F. dibotrys* continues to progress, its value and various uses are gradually being discovered and applied. Simultaneously, the increasing demand has led to increased attention towards the exploitation of wild germplasm resources. This process not only exacerbates the degradation of the ecology and environment but also results in the overexploitation of wild germplasm resources, causing a rapid decline in population numbers in a short period and disrupting the local ecological balance and diversity of the species. Therefore, the rational and efficient utilization of wild germplasm resources for the domestication and breeding of cultivated varieties through artificial reproduction methods is crucial for addressing the current excessive demand. Obtaining a high-quality genome of *F. dibotrys* may not only provide insights into the evolution of the genus *Fagopyrum* but also establish foundational data for resequencing and mining genes involved in flavonol biosynthesis. This necessitates an understanding of the distinctions between *F. dibotrys* varieties in different regions. Therefore, understanding the evolutionary relationships and genetic backgrounds of different subpopulations is necessary. The purposeful selection of superior subpopulations for hybridization and cultivation of cultivated varieties is imperative, as it will guide breeding efforts in developing cultivated varieties to meet the demands of the vast market.

In the present study, a high-quality *F. dibotrys* genome is reported. Additionally, its evolutionary relationships within the *Caryophyllales* order were explored and 46 wild samples from different geographical locations and 12 cultivated varieties were re-sequenced. These re-sequenced samples can be divided into subpopulations based on the genetic differences identified through population evolutionary analysis and genetic foundation. The potential origin of *F. dibotrys* and the subpopulation with the highest flavonoid content was inferred. Furthermore, the limitations of the introduction and domestication were identified. The resequencing of samples from different regions can reveal their genetic backgrounds, facilitating targeted breeding efforts for cultivated varieties. This will facilitate the cultivation of superior and high-yielding *F. dibotrys* varieties to meet the demands of an extensive market. In addition, this study contributes to the conservation of wild germplasm resources.

2 Materials and methods

2.1 Plant material acquisition and DNA extraction

The representative *F. dibotrys* individuals used were obtained from the plantation located in Xiaoshao Village, Dabanqiao Town, Guandu District, Kunming City, Yunnan Province located at 25°11′ 8.2176″ N, 102°58′ 54.7032″ E and an altitude of 1982 meters above sea level, China. Fresh, healthy mature leaves were collected for high-quality PacBio HiFi and Hi-C library construction. In addition, re-sequenced individual samples were collected from various regions in Yunnan, Guizhou, Chongqing, and Sichuan provinces. We also collected one seed-propagated and several cultivated varieties (Supplementary Table S1). After harvesting, all materials were immediately frozen in liquid N_2. High-quality genomic DNA was extracted using the E.Z.N.A. Tissue DNA kit (Omega Bio-Tek) following the Tissue DNA-Spin Protocol. Subsequently, their quality was tested, and the DNA samples with yield of > 3 μg, concentration of >30 μg/μL, and OD_{260}/OD_{280} 1.80–2.00 were qualified for use in further study.

2.2 Genome survey and sequencing

Before whole-genome sequencing, a 450 bp paired-end read library was constructed, and *k*-mer frequency distribution was used to estimate the genome size of *F. dibotrys*. First, the DNA was fragmented to ~ 450 bp using a Covaris M220, and T4 DNA polymerase was applied to generate blunt ends. After adding an 'A' base to the 3' end of the blunt phosphorylated DNA fragments, adapters were ligated to the ends of the DNA fragments. Second, the desired fragments were purified by gel electrophoresis, selectively enriched, and amplified by PCR. The index tag was introduced into the adapter at the PCR stage, followed by library quality testing. Finally, paired-end libraries were quantified using the BGI MGISEQ-2000 platform. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) with default parameters was used to assess sequencing data quality. Trimmomatic (Bolger et al., 2014) with default parameters was used to filter the short reads. Then, *k*-mers were counted using KMC (https://github.com/tbenavi1/KMC) with the parameters k27, ci1, and cs50000. The output file was analyzed using Smudeplot (Ranallo-Benavidez et al., 2020) and GenomeScope2 (https://github.com/tbenavi1/genomescope2.0) with default parameters to estimate ploidy and genome size. Genomic DNA samples were subsequently fragmented using Megaruptor, and high-quality fragments in the 13–16 kb range were selected using Sage ELF. The DNA library was constructed using SMRTbell following the protocol of the PacBio Sequel II platform. Circular consensus sequencing (CCS) was performed to obtain HiFi reads, and the lengths of subreads were subsequently analyzed, the visualization was performed using the ggplot2 (https://ggplot2.tidyverse.org/) package.

The same part of the plant was used to construct the Hi-C library, and 1% formaldehyde was used to fix young leaves for crosslinking. A Dounce homogenizer was used to lyse the cells, and MboI was used to digest the DNA. The ends of the DNA fragments were incorporated into adapter sequences containing biotin-labeled bases and ligated using sticky ends. The products were purified, ultrasonicated, and re-fragmented; only biotin-labeled pieces were captured and used to construct Illumina sequencing libraries (Belton et al., 2012). The DNA library was interrupted by high-frequency ultrasonication for resequencing, resulting in fragments of approximately 450 bp in length, following the Illumina standard protocol. Paired-end reads (2 × 150 bp) were sequenced using an Illumina Hiseq platform.

2.3 Genome assembly and quality assessment

The corrected CCS HiFi reads were assembled using HiFiasm V0.15.1 (Cheng et al., 2021, Cheng et al, 2022) with the following parameters: -l 3 -i –h1 –h2. Redundant sequences of the primary contigs were eliminated using purge haplotigs (Roach et al., 2018) with default parameters. Minimap2 V2.17 (Li, 2021) was used to align the HiFi reads to the assembled genome using the default parameters. The output data were sorted using SAMtools (Danecek et al., 2021). Next, MarkDuplicates software (https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard) was used to remove duplicates with default parameters. Then BamDeal V0.24 (https://github.com/BGI-shenzhen/BamDeal) was used to calculate sequencing depth, GC ratio and depth, which were visualized using the gglot2 package. Finally, BUSCO v5.1.2 (Manni et al., 2021a, Manni et al., 2021b) was used to evaluate their completeness based on the embryophyta_odb10 library.

2.4 Hi-C-assisted genome assembly

Hi-C libraries were used to scaffold chromosomes and improve the accuracy of *F. dibotrys* genomic data. The Hi-C reads were filtered using SOAPnuke V1.6.5 (Chen et al., 2018) with parameters: filter -n 0.1 -l 15 -q 0.4 -A 0.25 –cutAdaptor -Q 2 -G –minLen 150. Juicer (Durand et al., 2016) was used to align the paired-end reads to the assembled genome and obtain effective Hi-C data (Hi-C Contacts). Thereafter, 3D-DNA (Dudchenko et al., 2017) was used to construct interaction matrices and obtain pseudochromosome-level scaffolds. The Hi-C data were divided into 300 Kb/bin, and a Hi-C interaction heatmap was generated. Completeness of the assemblies was evaluated using BUSCO (Manni et al., 2021b, Manni et al., 2021a) with the embryophyta_odb10 library. LAI (Ou et al., 2018) was used to determine genome completeness using LTR_retriever V1.9 (Ou & Jiang, 2018).

2.5 Genome annotation

A combination of homologous and *de novo* approaches was used to improve the identification of repeat sequences. First, RepeatModeler (https://www.repeatmasker.org/RepeatModeler/) and LTR_Finder (Xu & Wang, 2007) were used to build the *de novo* repetition library. RepeatMasker V4.0.7 (https://www.repeatmasker.org/RepeatMasker/) was used to identify repeat sequences based on the library and output data were used to generate four types of TE divergence plots using ggplot2. Next, the homolog method was then applied using the RepBase V21.12 (https://www.girinst.org/repbase/), RepeatMasker V4.0.7 (Benson, 1999), and RepeatProteinMask V4.0.7 (https://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest) databases with default parameters to search for homologous repeat sequences. Tandem Repeats Finder V4.10.0 (Benson, 1999) was used to identify the tandem repeat sequences. Finally, the outputs of all the software were integrated and used as the final results.

Both *de novo* and homology-based methods were used for gene structure annotation. In the initial round of homology-based prediction methods, protein sequences from four species *Arabidopsis thaliana* (Cheng et al., 2017), *Beta vulgaris* (McGrath et al., 2023), *Fagopyrum tataricum* (Zhang et al., 2017), and *Solanum tuberosum* (Pham et al., 2020) were downloaded from the Phytozome (https://phytozome-next.jgi.doe.gov/) database and the China National Central for Bioinformation (https://ngdc.cncb.ac.cn/). Subsequently, approximately 3000 candidate

genes obtained from this prediction were selected as gene models to train AUGUSTUS (Stanke et al., 2006). Subsequently, the model was trained for annotation using protein sequences from closely related species as input files in Maker (Holt & Yandell, 2011) with default parameters. Final consensus predictions were obtained through integration and deduplication using Maker (Holt & Yandell, 2011). BUSCO (Manni et al., 2021a, Manni et al., 2021b) was used to evaluate the completeness of the gene sets based on the embryophyta_odb10 library. For the functional annotation, BLAST V2.2.31 (Altschul et al., 1990) was chosen based on Non-Redundant Protein Database, InterPro (Apweiler et al., 2001) (https://www.ebi.ac.uk/interpro/), KEGG (Kanehisa & Goto, 2000) (https://www.genome.jp/kegg/), SwissProt (Boeckmann et al., 2003) (https://www.uniprot.org/), and KOG databases with default parameters.

For noncoding RNAs (ncRNAs), tRNAscan-SE V1.3.1 (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe & Eddy, 1997) with default parameters was utilized for tRNA annotation. The rRNA fragments were predicted using BLAST with default parameters. The Rfam model in INFEANAL software (http://eddylab.org/infernal/) with default parameters was used to predict miRNAs and snRNAs. Based on the annotation results, we divided the assembled chromosome data into 50 kb windows to calculate gene counts, GC counts, and the percentage of repetitive sequences. We also included the syntenic blocks of the eight pseudochromosomes and presented them alongside the images of the plant features in Figure 1.

2.6 Phylogenetic and gene family analyses

To elucidate the evolutionary history of *F. dibotrys*, we selected eight additional species from the *Caryophyllales* order: *F. tataricum* (He et al., 2023), *F. esculentum* (He et al., 2023), *Portulaca amilis* (Gilman et al., 2022), *Gypsophila paniculata* (Li et al., 2022), *Amaranthus hypochondriacus* (Sunil et al., 2014), *Suaeda glauca* (Yi et al., 2022), *Spinacia oleracea* (Cai et al., 2021), and *Beta vulgaris* (McGrath et al., 2023), while *Vitis vinifera* (Olivier et al., 2007) was selected as an outgroup. OrthoFinder (https://github.com/davidemms/OrthoFinder/releases) was used to classify the gene families and identify single-copy families with the parameters -M msa -S diamond -T fasttree. All single-copy homologous genes were extracted and aligned using Muscle (http://www.drive5.com/muscle/) with the default parameters. These genes were selected to construct a phylogenetic tree using IQtree (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017), treating *V. vinifera* as the outgroup, and the most appropriate model with the parameter (-m JTT+F+R4 -bb 1000) was selected. The divergence time was inferred using MCMCTree software (http://evolution.genetics.washington.edu/phylip.html) in the PAML (Mistry et al., 2021) package and calibrated using the TimeTree database (https://www.timetree.org). Subsequently, CAFE (Mendes et al., 2020) was used with default parameters to identify the contractions and expansions of gene families following the phylogenetic tree using a probabilistic graphical model of predicted divergence. Next, eggNOG-mapper (Hyatt et al., 2010; Eddy, 2011; Steinegger & Söding, 2017; Huerta-Cepas et al., 2019; Buchfink et al., 2021; Cantalapiedra et al., 2021) was used for the functional annotation of genes. Based on these results, KEGG pathway classification and enrichment analysis for significantly expanded gene families ($p < 0.05$) were conducted using clusterProfiler (Wu et al., 2021).

2.7 Whole-genome duplication and collinearity analysis

F. dibotrys, *F. esculentum* (He et al., 2023), *F. tataricum* (He et al., 2023), and *V. vinifera* (Olivier et al., 2007) were selected to identify recent whole genome duplication (WGD) events and collinearity relationships. WGD V1.2 (Chen & Zwaenepoel, 2023) (https://github.com/arzwa/wgd) was used to identify the most recent WGD events by analyzing the distribution of K_s within

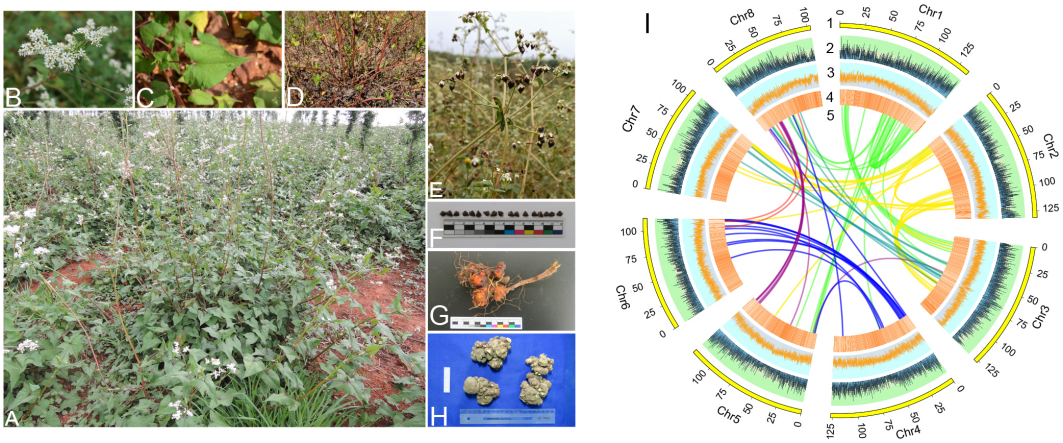


FIGURE 1
Morphology and genetic features of *F. dibotrys*. **(A)** Overall growth posture. **(B)** Inflorescence. **(C)** Leaves. **(D)** Stems. **(E, F)** Seeds. **(G, H)** Rhizomes. **(I)** Genomic landscape. (1–5): The lengths, gene count, GC count, percentage of repetitive sequences, and syntenic blocks of the eight pseudochromosomes.

paralogs with data visualization using the ggplot2 package. Subsequently, JCVI (https://github.com/tanghaibao/jcvi/wiki/MCscan-[Python-version]), with default parameters, was used to identify the collinearity blocks between these species. A collinearity block with at least five collinear gene pairs was constructed. Finally, CIRCOS V.0.69.8 (http://circos.ca/) software was used to visualize the gene count, gene synteny, percentage of repetitive sequences, and GC contents of individual pseudochromosomes.

2.8 RNA-seq data analysis

RNA-seq data was obtained from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP347690&o=acc_s%3Aa). During the withering stage and discontinuous removal of inflorescences in *F. dibotrys*, the following four tissue types of samples were selected: rhizome (RH), the top part of the stem (TS), the medium part of the stem (MS), and the basal part of the stem (BS); three biological replicates were set for each sample, totaling 12 transcriptome datasets.

The prefetch tool on the SRA Toolkit (https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit) was used to download the raw data from the SRA database. Then, the fastq-dump of the sra-tools software (https://github.com/ncbi/sra-tools) was used to convert the downloaded data into the fastq format. The quality of the output files was evaluated using the FastQC software. Based on the FastQC results, the adapter sequences were trimmed using the Trimmomatic tool (https://github.com/usadellab/Trimmomatic/). HISAT2 (Kim et al., 2019) (http://daehwankimlab.github.io/hisat2/) was used to align clean reads to the reference genome using the parameters –gzip -split3. Uniquely mapped reads were retained according to the NH:i:1 tag in the HISAT2 BAM file. The SAM file generated after the alignment was converted into a BAM file by SAMtools (Danecek et al., 2021). The featureCount tool in the Subread software (https://subread.sourceforge.net/) was used to calculate read counts, and the edgeR package (Robinson et al., 2010) was used to calculate the FPKM value.

BLAST with default parameters was used to identify homologous sequences in the *F. dibotrys* genome, based on genes related to flavonoid biosynthesis in the GenBank database. Additionally, HMMER (http://hmmer.org/) was used to identify *CYP* and *UGT* genes related to flavonoid biosynthesis in *F. dibotrys* genome based on the Pfam database (CYP: PF00067, UGT: PF00201). Subsequently, multiple sequence alignment was performed using MAFFT (https://mafft.cbrc.jp/alignment/software/linux.html) with the parameter: -maxiterate 1000 –auto, and the resulting alignment was trimmed using trimAl (https://vicfero.github.io/trimal/) with the parameters -automated1, which was then used as input for IQtree (Nguyen et al., 2015) to construct the phylogenetic tree for further screening of candidate genes with the parameter: -m MFP -bb 1000. Finally, heatmaps based on FPKM values were generated using the pheatmap package (https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap). *Rheum tanguticum* (Li et al., 2023) was used for the comparative analysis. We constructed *CYP* genes and *UGT* genes phylogenetic trees of the other nine species selected for the species phylogenetic tree by FastTree with default parameters, and then IQtree was used to construct the phylogenetic trees based on the identified *F3'H*, *F3'5'H*, *RT*, and *UGT78D* genes with parameters: -m MFP -bb 1000.

2.9 MADS-box gene family analysis

MADS-box genes play crucial roles in almost all aspects of plant development. They play important regulatory roles in plant growth and development. Arabidopsis thaliana genes were used as query sequences to identify MADS-box genes in *F. dibotrys*, *F. esculentum*, and *F. tataricum* (He et al., 2023); the *Arabidopsis thaliana* genes were used as the query sequence. HMMER was used to identify MADS-box-related genes based on the PlantTFDB database (https://planttfdb.gao-lab.org/aboutus.php) with default parameters. Phylogenetic trees were constructed using IQtree with the parameter: -m MFP -bb 1000 (Nguyen et al., 2015). At last, MG2C (http://mg2c.iask.in/mg2c_v2.0/) was used to annotate and visualize the chromosome distribution. For further analysis of MADS-box in *F. dibotrys* genome, the online program GSDS 2.0 (http://gsds.cbi.pku.edu.cn/index.php) was used to map the gene structures. Bedtools (https://bedtools.readthedocs.io/en/latest/index.html) was used to obtain the upstream 2000 bp sequences as promoter regions, and the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict cis-elements in their promoters. The MEME online programwas (https://meme-suite.org/meme/tools/meme) used to predict conserved motifs with a parameter of 10.

2.10 Resequencing library preparation and quality control

For Illumina pair-end sequencing, each sample containing at least 1 μg genomic DNA was used for sequencing library construction. Next, Covaris M220 was used to fragment the DNA into ~ 450 bp long pieces, following the standard genomic DNA library preparation protocol of the Illumina TruSeq™ Nano DNA Sample Prep Kit. Subsequently, a polyA tail was added to the 3' end of the blunt phosphorylated DNA fragments, and adapters were ligated at both ends. The desired fragments were purified by gel electrophoresis, selectively enriched, and amplified by PCR. An index tag was introduced into the adapter at the PCR stage as required, followed by a library quality test. Finally, the quantified paired-end libraries were sequenced using the Illumina HiSeq platform (2 × 150 bp), and FastQC was used to evaluate the quality of the sequence reads. Trimmomatic was used to trim the raw data, and Burrows-Wheeler Aligner software (BWA) with the following parameters: bwa mem -k 32 (https://bio-bwa.sourceforge.net/index.shtml) was used to map the clean reads to the assembled reference genome and obtain the sequencing depth of the resequenced samples. Duplicate reads produced by PCR were

removed using Picard tools (https://broadinstitute.github.io/picard/).

2.11 Variation detection and annotation

A valid BAM file was used to detect SNPs and short InDels (insertions and deletions) using the GATK HaplotypeCaller tool in the GATK toolkit (https://gatk.broadinstitute.org/hc/en-us) based on the Bayesian model, which can filter and refine the results for more accurate annotation. Variant call format (VCF) files were generated by quality filtering using VCF tools (https://vcftools.github.io/index.html) with the following parameters: QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 10.0. Subsequently, ANNOVAR (https://annovar.openbioinformatics.org/en/latest/user-guide/download/) was used to functionally annotate detected gene variants.

2.12 Population genetics

To explore genetic variations among different individuals, SNP annotation results from the resequenced individuals were used to construct an evolutionary tree using FastTree v2.1.10 (http://www.microbesonline.org/fasttree/), based on the neighbor-joining method. Thereafter, Principal Component Analysis (PCA) was used to classify the different subpopulations using GCTA v1.93.2 (https://yanglab.westlake.edu.cn/software/gcta/#Overview). Furthermore, population structure analysis was combined to ascertain the number of subpopulations among the resequenced individuals using fastSTRUCTURE (Raj et al., 2014) software with default parameters, and the different subpopulations were colored according to the clusters identified using fastSTRUCTURE software.

3 Result

3.1 Genome sequencing, assembly, and annotation

Genome size and ploidy were estimated using *k*-mer with short reads, revealing that *F. dibotrys* was a heterozygous diploid (Supplementary Figure S1). The estimated genome size was approximately 800 Mb with a heterozygosity of 1.4% (Supplementary Figure S2). Analyses of the GC ratio and depth (Supplementary Figure S3) showed that the sequenced samples were not polluted. Initially, a 314.36 Gb raw database was obtained using the PacBio Sequel II platform. Finally, we obtained 18.15 Gb of CCS data comprising 1.6 M HiFi reads, and ~ 21.81 × coverage of *F. dibotrys* (Supplementary Table S2), the sequencing depth was primarily between 8- to 27-fold (Supplementary Figure S4), and most reads were ~ 10 Kb in length (Supplementary Figure S5). After initial assembly and removal of duplicate sequences, the preliminary assembly results showed that the N50 value was 2.47 Mb and the N90 value was 735.20 kb (Supplementary Table S3).

To obtain a more accurate chromosome-level assembly, 142.77 Gb of high-quality Hi-C paired-end reads were generated, and the Q20 and Q30 values were 97.72% and 93.67%, respectively (Supplementary Table S4). Juicer results showed that 68 M Hi-C Contacts were obtained (Supplementary Table S5). After anchoring the scaffold with Hi-C sequencing data, we obtained 967.5 Mb of high-quality data spanning eight pseudochromosomes (superscaffolds). The anchored rate was 99.23%, and the N50 values of the contig and super-scaffold were 2.43 Mb and 127 Mb, respectively (Table 1 and Supplementary Table S6), and each chromosome was ~ 120 Mb in length (Table 1 and Supplementary Table S7). Subsequently, the Hi-C interaction heatmap showed that the interaction at the diagonal position was significantly higher than at other positions, demonstrating a better assembly effect (Supplementary Figure S6), which was also supported by the LAI evaluation results showing that most values were > 10 (Supplementary Figure S7). The BUSCO evaluation indicated that the contigs and scaffold assemblies covered 94.67% and 94.61% of the core conserved genes, respectively (Table 1 and Supplementary Table S8).

Repetitive sequences comprise most of the genome. Therefore, TRF, RepeatMasker, RepeatproteinMask, and *de novo* annotation data were integrated and redundancy was masked. The final results showed that repetitive sequences collectively occupy 0.58 Gb (62.47%) of the genome (Supplementary Table S9). Among these, LTRs were the most abundant, accounting for approximately 58%, followed by LINEs. The divergence rates of the four TE sequences indicated that most TE divergences were concentrated at ~ 30% (Supplementary Figure S8). Combining the *de novo* and homology-based methods ultimately yielded a gene set consisting of 40610 genes (Supplementary Table S10). Regarding genome annotation, 1523 (94.4%) of the 1614 Embryophyta single-copy orthologs from BUSCO were identified as complete genes in the *F. dibotrys* genome, indicating a high-quality genome assembly (Supplementary Table S8). Regarding ncRNAs, the annotation results indicated that rRNA had the highest proportion (0.066%), followed by snRNAs (0.010%). The lowest proportion was observed for tRNA (0.009%; Supplementary Table S10). Regarding gene functional annotation, after integrating the prediction results, 23037 genes were annotated using five databases, comprising 56.72% of the total (Supplementary Figure S9).

TABLE 1 Assembly and annotation statistics of the *F. dibotrys* genome.

Feature	Value
Assembled genome size (Mb)	967.5
Chromosome number	8
Average length per chromosome (Mb)	121
N50 of contig (Mb)	2.43
N50 of super-scaffold (Mb)	127
GC content (%)	0.42
Complete gene percentage of contig (%)	94.67
Complete gene percentage of super-scaffold (%)	94.61
Number of protein-coding genes	40610

3.2 Evolutionary history and whole-genome duplication

This study selected ten species, nine of *Caryophyllales*, and one of *V. vinifera* as an outgroup. These were clustered into 56424 gene families, including 328 single-copy genes, to infer the high-confidence phylogenies of these species. Using molecular clock analysis and calibrated with the TimeTree database, it was estimated that the genus *Fagopyrum* diverged from other genera in the *Caryophyllales* order 117.79 million years ago (Mya), *F. esculentum* divergence from the common ancestor of *F. dibotrys* and *F. tataricum* 21.00 Mya, and *F. dibotrys* diverged from *F. tataricum* 9.04 Mya (Figure 2A). The evolution of gene families plays a crucial role in the phenotypic variation in plants. Based on the analysis of the expansion and contraction of gene families in *F. dibotrys*, 1733 expanded gene families were detected, primarily enriched in sesquiterpenoid and triterpenoid biosynthesis, amino sugar and nucleotide sugar metabolism, and metabolic pathways (Figure 2B). Based on the phylogenetic tree, *F. tataricum, F. esculentum*, and *F. dibotrys* are closely related. Collinearity analysis revealed strong collinearity among *F. tataricum*, *F. esculentum* and *F. dibotrys* (Figure 2D).

WGD or polyploidy events, is common in plant genomes and contributes to variations in both size and content among plants (Soltis & Soltis, 2016). Considering the Log_{10} of the distribution of synonymous substitutions per synonymous site (K_s) value was used to increase the distance between different K_s peaks and better distinguish different WGD events (Figure 2C and Supplementary Figure S10). Unlike the K_s peak in the *V. vinifera* genome, the K_s values for paralogous genes in the collinear regions of *F. dibotrys*, *F. tataricum*, and *F. esculentum* appeared to show the same $K_s$peak, indicating that they shared the same WGD event before the divergence of the *Fagopyrum* species and that no WGD event occurred in *Fagopyrum* after the divergence event. Although they have a close evolutionary relationship, they exhibit significant differences in genome size, *F. esculentum* 1,219.3 Mb, *F. tataricum* ~ 453.9 Mb (He et al., 2023), and *F. dibotrys* 967.5 M (Table 2). In *Fagopyrum*, genomes with large sizes also have a higher proportion of repetitive sequences *F. esculentum* ~ 72.9%, 888.9 Mb, *F. tataricum* ~ 52.4%, 237.8 Mb, *F. dibotrys* ~ 62.5%, 604.7 Mb (Table 2), and the gene number is similar (*F. esculentum* 38078, *F. dibotrys* 40610, *F. tataricum* 34366). Compare genome size and the repetitive sequences among these three species, the repetitive sequences (*F. esculentum* - *F. dibotrys* 284.2 Mb, *F. esculentum* - *F. tataricum* 651.1 Mb, *F. dibotrys* - *F. tataricum* 366.9 Mb) seems the main reason for the different genome size, consistent with previous report (He et al., 2023).

3.3 Evolution of MADS-box genes family and their chromosomal localizations

A. thaliana MADS-box genes were used as query sequences. We identified 74 homologs in the *F. dibotrys* genome, in comparison to *F. tataricum* and *F. esculentum* within the same genus and 62 and 89 homologs, respectively (Supplementary Table S11 and

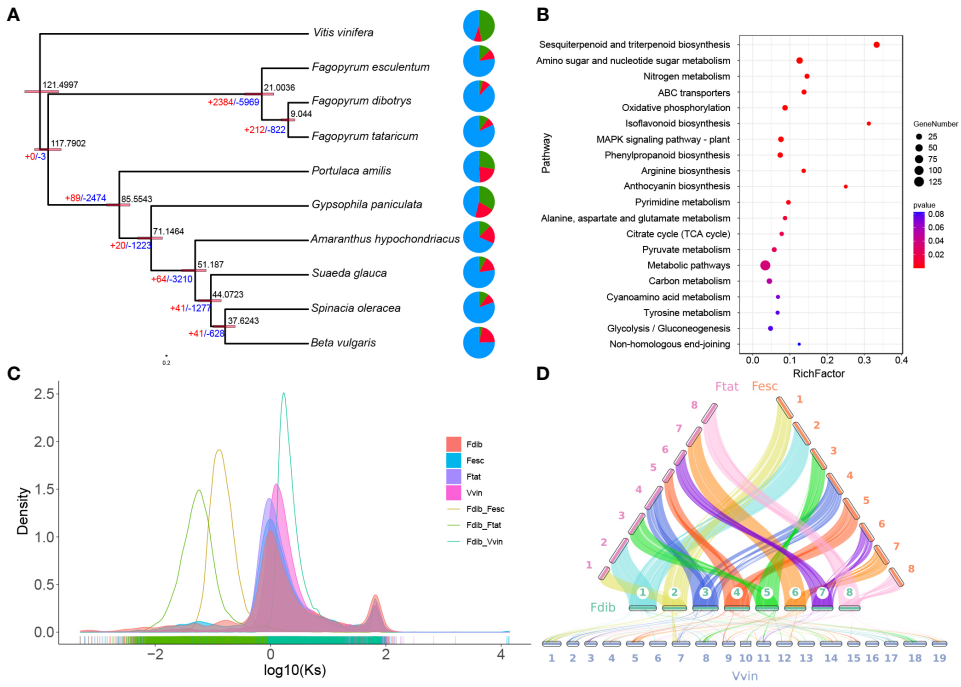


FIGURE 2
Evolution of the genome and enrichment of the expanded gene family. **(A)** Phylogenetic tree of Caryophyllales and the estimated divergence times are indicated at the internodes. The expanded and contracted gene families are represented in red and blue, respectively. **(B)** Significantly expanded gene families in *F. dibotrys* enriched using KEGG pathway analysis. **(C)** The frequency distribution of the K_s values of homologous gene pairs. **(D)** The collinearity of *F. dibotrys*, *F. tataricum*, *F. esculentum*, and *V. vinifera*.

10.3389/fpls.2024.1380157

TABLE 2 Assembly and annotation of the *F. dibotrys*, *F. esculentum*, and *F. tataricum* genome.

Feature	*F. dibotrys*	*F. esculentum*	*F. tataricum*
Genome size (Mb)	967.5	1,219.3	453.9
N50 of contig (Mb)	2.43	2.4	27.9
N50 of super-scaffold (Mb)	127	154.3	56.7
GC content (%)	0.42	0.39	0.38
Repeat region of Genome (%)	62.5	72.9	52.4
Predicted Gene number	38,078	40,610	34,366

Supplementary Figure S11). Among these, *F. esculentum* (54 copies) had more MIKC-type MADS-box genes than *F. dibotrys* (32 copies) and *F. tataricum* (40 copies) (Supplementary Figure S11). Moreover, analysis of whole-genome duplication revealed that they underwent the same WGD event before the divergence, and no WGD event occurred after the divergence event. From this we inferred that the additional copies of MIKC-type MADS-box might result from gene family expansion and the obvious differences of MADS-box members in genus *Fagopyrum* might be related to aspects of plant development including flower and embryo development (Schilling et al., 2020). To further analyze the structure of the MADS-box in the *F. dibotrys* genome, intron-exon analysis revealed that the vast majority of M-type (light blue and red parts) MADS-box genes have one or two introns, and most of them lack introns. In contrast, the number of exons and introns was high in MIKC-type (yellow part) subfamily; the number of exons from 1–11 and the number of introns from to 1–10 (Figure 3B), and two subgroups (Figure 3A, Mα: light blue part and Mγ: yellow part) had similar gene structures. Additionally, the same subgroup contained relatively conserved and similar motif compositions, indicating that these genes shared the basic general characteristics of the MADS-box gene family. Furthermore, motif 4 is unique in M-type subfamily, motif 9 and motif 5 are unique in Mα subgroup but Mα subgroups lack of motif 2 (Figure 3C). Motif 3 and motif 7 only identified in Mγ subgroup, this result is consistent with the result of subfamily classification. Based on the 2000 bp up-steam sequence analysis of the MADS-box in *F. dibotrys* genome, these cis-acting elements mainly contain light-responsive elements, hormone-responsive elements, and MYB-binding sites (Figure 3D). In the *F. dibotrys* genome, the majority of these were type I genes distributed across every chromosome, spanning nearly all subfamilies, and the Mα subfamily constituted the majority with 24 copies and was present on almost all chromosomes. Moreover, type I genes exhibited a tandem distribution, particularly the Mγ subfamily on chromosome 6. The *F. dibotrys* genome contained genes corresponding to the ABCE model, including four A-class genes from the SQUA subfamily, potentially determining sepal identity; three B-class genes from the DEF subfamily that, in combination with the A-class genes, might control petal identity; two C-class genes from the AG subfamily that were likely involved in stamen identification along with the B-class genes; and six E-class genes from the AGL2 subfamily, which have specific roles as mediators of higher-order complex formation (Supplementary Figure S12 and Supplementary Figure S13).

3.4 Mining genes involved in the biosynthesis of flavonoids

F. dibotrys accumulates numerous critical bioactive compounds, such as flavonoids, phenolics, fagopyritols, triterpenoids, steroids, and fatty acids (Jing et al., 2016). In *F. dibotrys*, various flavonoid compounds are initially derived from phenylalanine as substrates, catalyzed by multiple enzymes. In the genome of *F. dibotrys*, 4 *PAL*s (phenylalanine ammonia-lyase) and 4 *CHS*s (chalcone synthase), 3 *CHIs*: (chalcone isomerase), 9 *C4H*s (CYP73A: cinnamate 4-hydroxylase), 5 *FLS*s (flavonol synthase), 13 *UGT78D*s (UGT78D1: flavonol-3-O-rhamnosyltransferase; UGT78D2: flavonol 3-O-glucosyltransferase), 3 *LARs* (leucoanthocyanidin reductase), 3 *RTs* (UGT79A6: flavonol 3-O-glucoside 6-O-rhamnosyltransferase), 2 *4CLs* (4-coumarate–CoA ligase), 2 *F3'Hs* (CYP75B1: flavonoid 3'-hydroxylase), 2 *F3Hs* (flavanone 3-hydroxylase), 1 *ANR* (anthocyanidin reductase), 1 *ANS* (anthocyanin synthase), and 1 *F3'5'H* (flavonoid 3',5'-hydroxylase) were identified through homologous search (Figure 4).

From dihydrokaempferol to dihydroquercetin, flavonoid biosynthesis was catalyzed by F3'H and F3'5'H enzymes, which belong to the cytochrome P450 family. Based on the phylogenetic tree result of the *CYP* gene (Supplementary Figure S14), we identified 1 *F3'5'H* gene and 2 *F3'H* genes in the genome of *F. dibotrys*, *F. esculentum*, and *F. tataricum*; we identified 2 *F3'5'H* genes and 1 *F3'H* gene in *R. tanguticum* genome. However, no candidates for the *F3'5'H* and *F3'H* gene could be identified in the genome of *A. hypochondriacus* and *S. glauca*. In other *Caryophyllales* species, only one *F3'H* gene was identified, but no candidate for the *F3'5'H* gene was identified (Supplementary Table S12). The absence of *F3'5'H* and *F3'H* genes in certain species might impact flavonoid biosynthesis and diversity. We inferred that these species lacked these metabolic pathways during the speciation process. Additionally, the phylogenetic tree of *F3'H* indicated that the different number of *F3'H* genes was speculated to be produced by the expansion of the gene family (red branch) and this phenomenon only occurs in the genus *Fagopyrum* (Supplementary Figure S16B). UGTs convert quercetin to quercitrin and rutin. Based on the phylogenetic tree results of the *UGT* genes (Supplementary Figure S15), we identified 3 *RTs* genes in the *F. dibotrys*, *F. esculentum*, and *F. tataricum* genomes, 13 *UGT78D* genes in the *F. dibotrys* genome, 9 *UGT78D* genes in *F. tataricum* genome, and 12 *UGT78D* genes in *F. esculentum* genome. In the *R. tanguticum* genome, 2 *RT* genes and 4 *UGT78D* genes were identified (Supplementary Table S12), and the number of *UGT* genes was lower than that in the *Fagopyrum* species. Compared with other species of *Caryophyllales*, we also identified fewer *RT* and *UGT78D* genes in *Fagopyrum* species. The obvious differences in UGT78D members among various species might be related to the

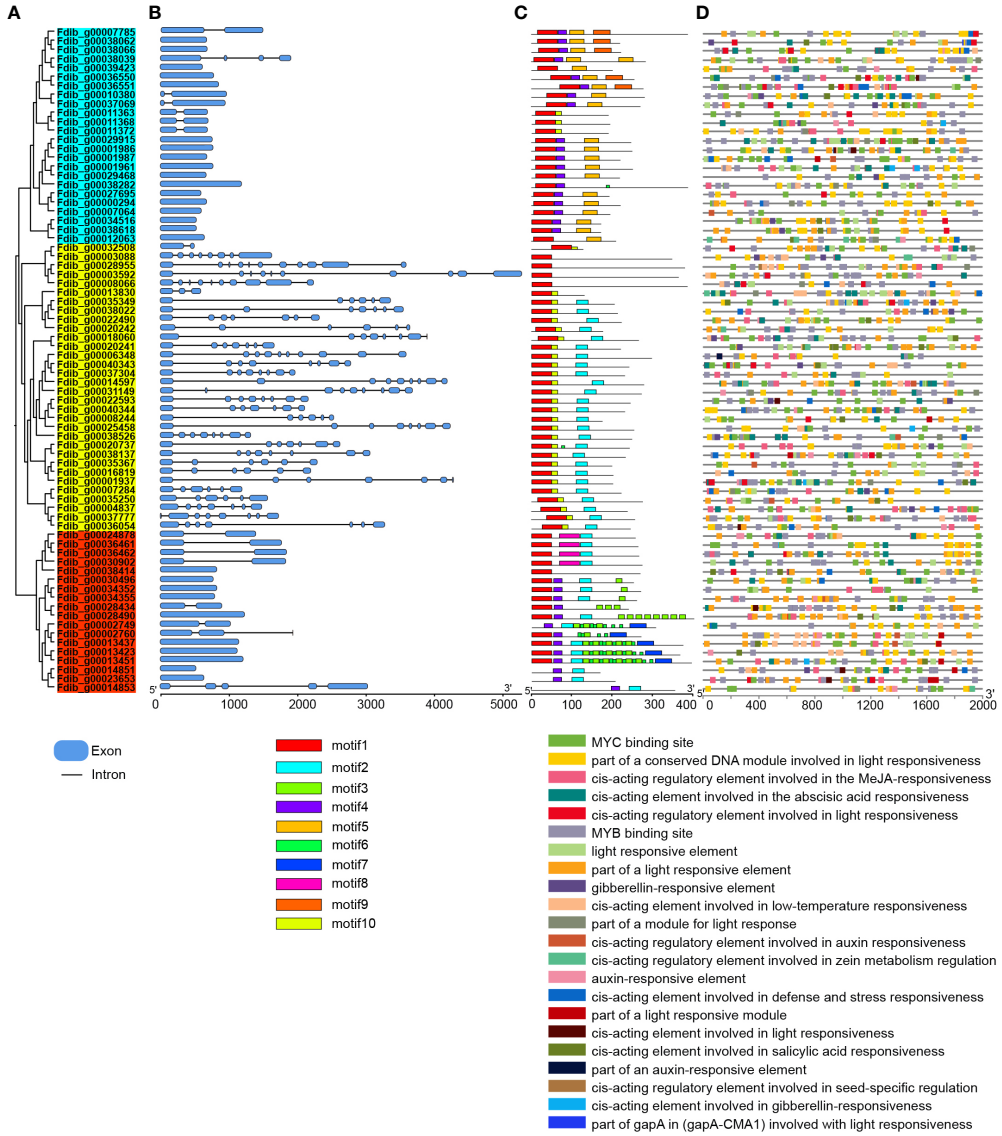


FIGURE 3
Phylogenetic tree, gene structures, conserved protein motifs and cis-elements analysis of MADS-box in *F. dibotrys*. **(A)**: Phylogenetic tree of MADS-box, light blue part represents Mα subgroup, red part represents Mγ subgroup, yellow part represents MIKC-type subfamily, **(B)**: gene structures of MADS-box, blue box represent exons, black line represent intron, **(C)**: conserved protein motifs MADS-box, different color represents different motif, d: cis-elements analysis of upstream 2000bp promoter sequence of MADS-box, different colors represent different elements.

accumulation and diversity of flavonoid glycosides. *Fagopyrum* had more gene copies of the *UGT78D* gene and *RT* gene compared with the other species of *Caryophyllales*, especially the *UGT78D* gene. We identified a substantial expansion in the *UGT78D* gene family (Supplementary Figure S17A red branch and Supplementary Table S12), and gene family expansion also occurred in the *RT* gene family (Supplementary Figure S17B red branch), which appears to be common among the Polygonaceae family. We inferred that the significant expansion of the *UGT78D* gene family might benefit flavonoid accumulation. Additionally, by calculating the gene expression levels related to flavonoid biosynthesis in different tissues of *F. dibotrys*, we found that during the withering stage, *ANR* and *ANS* were continuously expressed, which is beneficial for the accumulation of epicatechins and catechin. In addition, the expression levels in the rhizome were higher than those in the stem, suggesting that the rhizome accumulated more epicatechins (Figure 4).

3.5 Genetic structure of the populations of *F. dibotrys*

Sichuan, Guizhou, and Yunnan Provinces are located in the southwestern region of China, and the Hengduan Mountains are

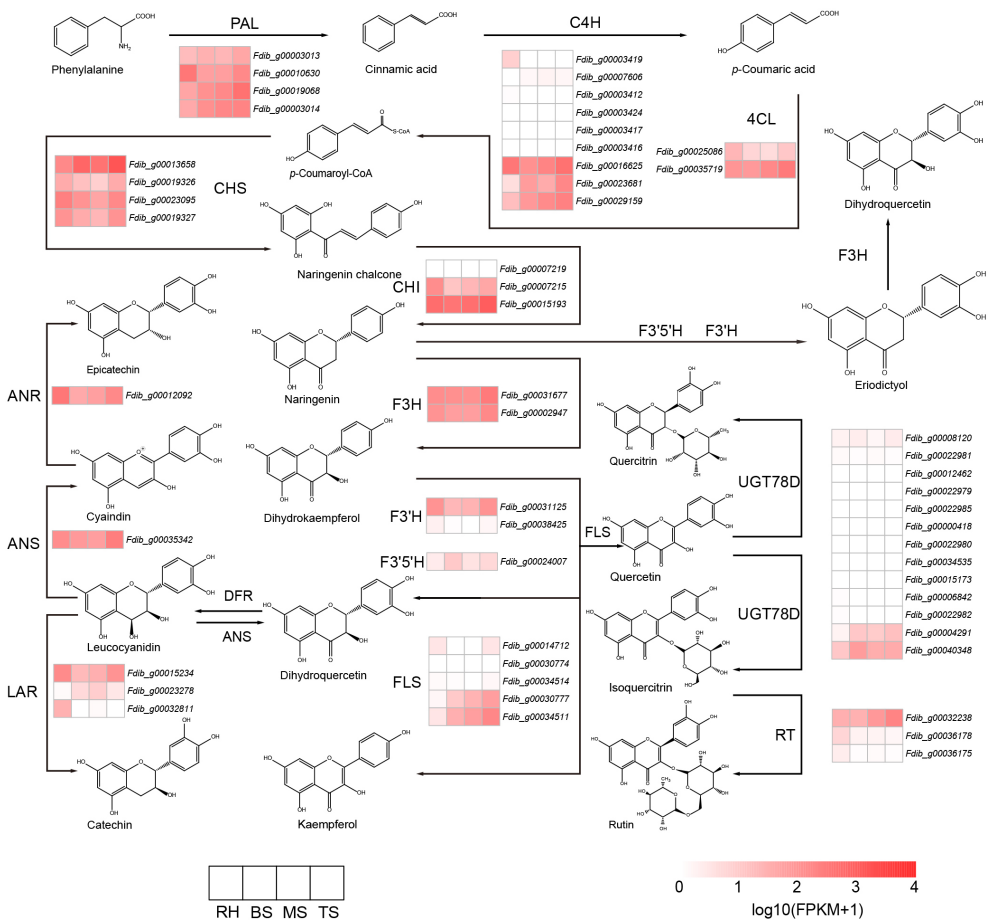


FIGURE 4
Flavonoid metabolic pathway and expression analysis of the genes involved. PAL: (phenylalanine ammonia-lyase, C4H: cinnamate 4-hydroxylase, 4CL: 4-coumarate–CoA ligase, CHS: chalcone synthase, CHI: chalcone isomerase, F3H: flavanone 3- hydroxylase, F3'H: flavonoid 3'-hydroxylase, F3'5'H: flavonoid 3',5'-hydroxylase, FLS: flavonol synthase, ANR: anthocyanidin reductase, ANS: anthocyanin synthase, LAR: leucoanthocyanidin reductase, RT: flavonol 3-O-glucoside 6-O-rhamnosyltransferase, UGT78D: UGT78D1: flavonol-3-O-rhamnosyltransferase; UGT78D2: flavonol 3-O-glucosyltransferase.

predominant in this area. This region benefits from its distinctive natural conditions, fostering rich flora and fauna resources, making it a crucial place for wild medicinal herbs in China; this region is also one of the main producing areas of *F. dibotrys*. After quality control, the average Q20 and Q30 values for the re-sequenced data from different sources were 97% and 89%, respectively, indicating that the data were sufficient for mapping to the reference genome (Supplementary Table S13). Except for the 6–3A dataset, the average alignment rate was 98% and the average coverage rate was 85% (Supplementary Table S13). A total of 609,024,689 SNPs and 62,964,288 indels were identified (Supplementary Table S14) and were primarily concentrated in the intergenic regions (Supplementary Table S15).

Individual phylogenetic trees, PCA, and population structure analyses were combined based on the resequencing data. The population structures of the wild and cultivated varieties of *F. dibotrys* were initially investigated to determine the genetic backgrounds of the re-sequenced individuals. In conjunction with the PCA results, the re-sequenced individuals were initially classified into four homozygous subpopulations and six pairwise hybridization subpopulations. Further integration with the phylogenetic tree constructed based on the re-sequenced individuals suggested that the red subpopulation evolved from the orange subpopulation, further dividing it into three distinct homozygous subpopulations (Figures 5A, B, Supplementary Figure S18, S19). Geographically, a light-blue subpopulation was found in the southwestern part of Yunnan Province at the highest altitude, an orange subpopulation in the central part of Yunnan Province, and a blue subpopulation in the lower-altitude plains of Guizhou (Figure 5 and Supplementary Table S1). The light blue subpopulation was located in the highest-altitude regions and differentiated first, followed by the orange subpopulation, with the blue subpopulation being the last to differentiate and was located in the lowest-altitude areas. Therefore, *F. dibotrys* is speculated to have

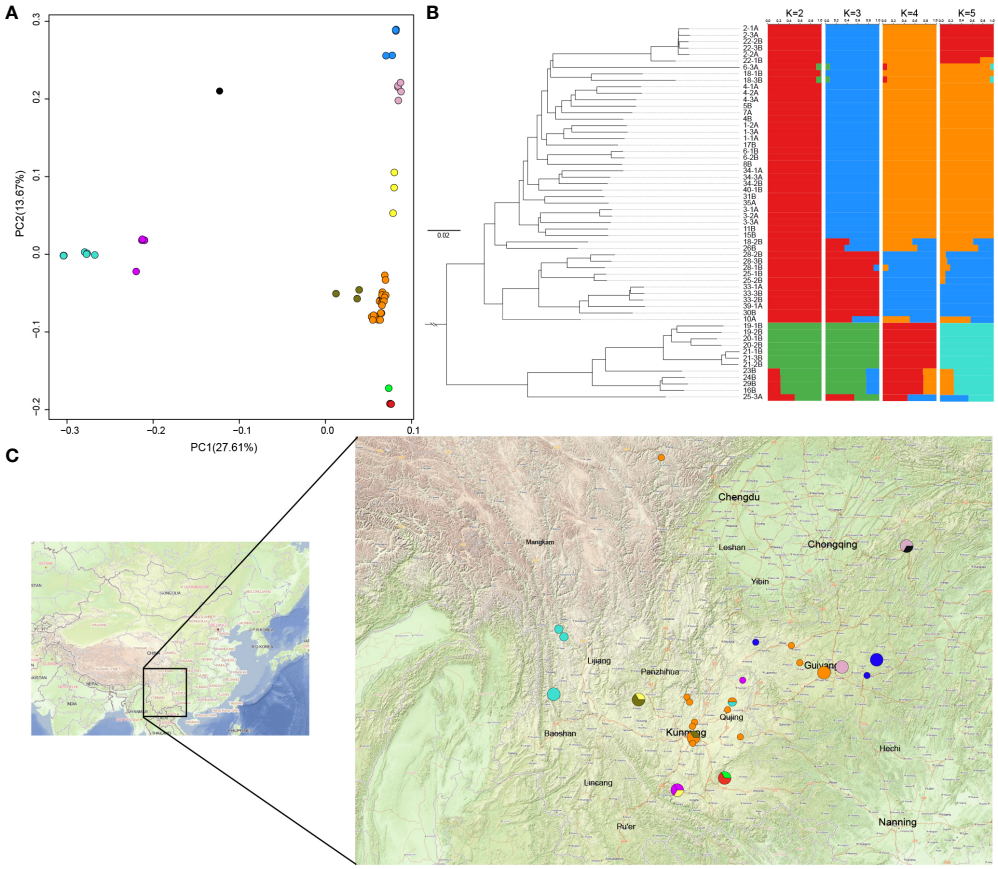


FIGURE 5
Genetic structure of the resequenced population. **(A)** PCA results of the resequenced individuals. Combining the results of population structure analysis, different colors represent the distinct subpopulations; Orange, blue, light blue, and red represent different homozygous subpopulations, while other colors represent heterozygous subpopulations, consistent with Supplementary Table S1. **(B)** Phylogenetic tree and population genetic structure of resequenced individuals. **(C)** Sampling site annotation, different-sized circles represent various sampling numbers (1–3), and different colors represent the different subpopulations.

originated from the high-altitude Tibetan Plateau. The cultivated varieties were bred using orange and red subpopulations from the central part of Yunnan Province as parental sources.

4 Discussion

F. dibotrys, *F. tataricum*, and *F. esculentum* belong to genus *Fagopyrum*, family *Polygonaceae*, and order *Caryophyllales*. *Fagopyrum* presently comprises 15 species (Jing et al., 2016). In contrast to *F. tataricum* and *F. esculentum*, which are staple food crops, *F. dibotrys* is a traditional Chinese medicinal herb with high therapeutic value. *F. dibotrys* was first recorded in "Bencaoshiyi". In traditional Chinese medicine, the rhizome of *F. dibotrys* is used to treat lung abscesses, dysentery, rheumatism, throat inflammation, and tumefaction (Chen & Li, 2016) and its extraction of rhizome can condensed Tannin compound, which is used to prepare a medicine "WeiMaiNing" (Ke et al., 2021). As research progresses, the medicinal value of *F. dibotrys* has been discovered and received rising attention. The overexploitation of wild germplasm resources can deplete environmental resources and reduce species diversity. Wild genetic resources cannot meet the growing demands and require exploration for industrial cultivation.

In the present study, we assembled and annotated the genome of *F. dibotrys*, which had a total length of 967.5 Mb, and the genome sequence of *F. dibotrys* provided a valuable genomic resource for *Fagopyrum*. Phylogenetic analysis revealed that *Fagopyrum* underwent a single WGD event before species divergence. Additionally, they have a one-to-one syntenic relationship at the whole-chromosome level. When combined with previous research, the recent large-scale LTR bursts contributed to most of the genome size of *F. esculentum* (He et al., 2023). Subsequently, we inferred that the amplification of the repetitive sequences was the primary factor driving the variation in genome size among genus *Fagopyrum*. In addition, MADS-box genes are prime candidates; they are involved in virtually all aspects of plant development, and especially have an important impact on flower development (Schilling et al., 2020). We identified 74 MADS-box genes in the genome of *F. dibotrys* and found that these genes corresponded to the ABCE model of flower structure formation. Genome mining can

considerably enhance the exploration of biosynthetic pathways for natural products. Flavonoids are one of the most important accumulations in the genus *Fagopyrum*, especially six major: quercetin, rutin, orientin, homoorientin, vitexin, and isovitexin (Joshi et al., 2020). While comparing gene copies number involved in flavonoid biosynthesis among different genera within *Caryophyllales*, we observed *Fagopyrum* possessed large number of copies of the *UGT78D*, *RT*, and *F3'H* genes compared with other species within *Caryophyllales*, particularly the abundance of copies within the *UGT78D* gene family relative to other genera. From this observation, we inferred that the *UGT78D* might have undergone expansion. Moreover, considering its role in catalyzing the biosynthesis of rutin and quercitrin, the expansion of the *UGT78D* family was crucial for the increased content of rutin and quercitrin in genus *Fagopyrum*. Additionally, we collected 46 wild samples and 12 cultivated varieties from Sichuan, Yunan, Guizhou, and Chongqing provinces, mainly located in the southwestern region of China, which is a critical area for producing Chinese medicinal herbs. By integrating the results of the population structure, PCA, and phylogenetic tree analyses, individuals from different regions were classified into four homozygous subpopulations based on topography, combining the evolutionary relationships that were further grouped into three homozygous subpopulations. Distributed across three distinct topographical areas, the light blue subpopulation was distributed at the highest altitude, followed by the orange subpopulation, whereas the blue subpopulation was distributed at the lowest altitude (Figure 5 and Supplementary Table S1). Additionally, the cultivated varieties were based on the genetic background of the orange subpopulation and its evolved red subpopulation in central Yunnan Province. In a previous study, the southwestern region of China was thought to be the origin of *F. esculentum* and *F. tataricum* (Hunt et al., 2018), and *F. dibotrys* to be widely distributed in China, India, Thailand, and Vietnam (Zhang et al., 2021); however, phylogenetic trees indicated that they were wild relatives. Combining the resequencing results of our individuals, the light blue subpopulation of *F. dibotrys* situated at the high altitude of the Tibetan Plateau region was the first to complete differentiation, suggesting that *F. dibotrys* also originated in western China, and the Tibetan Plateau region appeared to be the more accurate region of origin for the *F. dibotrys* species.

Buckwheat in high-altitude regions seems to accumulate more flavonoids than that in low-altitude regions (Cheng et al., 2019; He et al., 2022). In addition, flavonoids protect plants against the harmful effects of UV radiation (Shen et al., 2022), considering that the light-blue subpopulation in high-altitude regions receives more intense UV radiation. Therefore, we speculate that the light-blue subpopulation would likely have the highest accumulation of flavonoids. Additionally, the cultivation base of this study was located in Kunming City, and collection of germplasm resources from higher altitudes indicated the need to overcome additional challenges. The cultivated varieties were based only on the genetic backgrounds of the orange and red subpopulations, indicating that the wild genetic resources used were limited to the vicinity of the cultivation base for an extended period. This restricted the genetic diversity of the populations that were bred, as the genetic advantages of different subpopulations could not be harnessed completely. Based on these theories, it is necessary to ascertain the content of relevant compounds in the light-blue subpopulation, and if they meet these requirements, further domestication can be considered. Additionally, the collection of wild genetic resources should be expanded, extending beyond the vicinity of Yunnan Province, with particular emphasis on the Tibetan region. The quality of medicinal herbs can be improved by harnessing the advantages of different subpopulations to ultimately serve human health needs. Therefore, it is necessary to expand the collection range of germplasm resources in the future, increase the genetic diversity of the population, enhance the quality of *F. dibotrys*, and reduce the overexploitation of wild resources.

Data availability statement

The data presented in the study are deposited in the figshare repository, accession number: (https://doi.org/10.6084/m9.figshare.22240414.v1).

Author contributions

S-HZ: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. Y-CD: Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. JD: Conceptualization, Data curation, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. J-TL: Conceptualization, Formal analysis, Resources, Validation, Writing – original draft, Writing – review & editing. SZ: Conceptualization, Formal analysis, Validation, Writing – original draft, Writing – review & editing. M-JL: Conceptualization, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. H-CL: Formal analysis, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing. YZ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. J-YW: Conceptualization, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by China Agriculture Research System of MOF and MARA.

Conflict of interest

Authors S-HZ, Y-CD, JD, SZ, M-JL,YZ, J-YW were employed by the company China National Traditional Chinese Medicine Co., Ltd. Authors J-TL and H-CL were employed by the company China Traditional Chinese Medicine Seed&Seeding, Co., Ltd.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1380157/full#supplementary-material

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

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. doi: 10.1093/nar/29.1.37

Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi–C: a comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.ymeth.2012.05.001

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370. doi: 10.1093/nar/gkg095

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

Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x

Cai, X., Sun, X., Xu, C., Sun, H., Wang, X., Ge, C., et al. (2021). Genomic analyses provide insights into spinach domestication and the genetic basis of agronomic traits. *Nat. Commun.* 12, 7246. doi: 10.1038/s41467-021-27432-z

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

Chan, P. K. (2003). Inhibition of tumor growth in *vitro* by the extract of fagopyrum cymosum (fago-c). *Life Sci.* 72, 1851–1858. doi: 10.1016/S0024-3205(03)00013-4

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7, 1–6. doi: 10.1093/gigascience/gix120

Chen, C., and Li, A. (2016). Transcriptome analysis of differentially expressed genes involved in proanthocyanidin accumulation in the rhizomes of *Fagopyrum dibotrys* and an irradiation-induced mutant. *Front. Physiol.* 7. doi: 10.3389/fphys.2016.00100

Chen, H., and Zwaenepoel, A. (2023). Inference of ancient polyploidy from genomic data. *In Polyploidy: Methods Protoc.* 2545, 3–18. doi: 10.1007/978–1-0716–2561-3_1

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175. doi: 10.1038/s41592-020-01056-5

Cheng, H., Jarvis, E. D., Fedrigo, O., Koepfli, K.-P., Urban, L., Gemmell, N. J., et al. (2022). Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* 40, 1332–1335. doi: 10.1038/s41587-022-01261-x

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415

Cheng, C., Zhang, K., Tang, Y., Shao, J., Yan, M., and Zhou, M. (2019). Investigation on wild Fagopyrum cymosum resources in Yunnan and analysis of genetic diversity. *J. Plant Genet. Resour.* 20, 1438–1446. doi: 10.13430/j.cnki.jpgr.20190223001

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, O. M., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the Aedes aEgypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. doi: 10.1126/science.aal3327

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002

Eddy, S. R. (2011). Accelerated profile HMM searches. *PloS Comput. Biol.* 7, e1002195. doi: 10.1371/journal.pcbi.1002195

Gilman, I. S., Moreno-Villena, J. J., Lewis, Z. R., Goolsby, E. W., and Edwards, E. J. (2022). Gene co-expression reveals the modularity and integration of C4 and CAM in Portulaca. *Plant Physiol.* 189, 735–753. doi: 10.1093/plphys/kiac116

He, M., He, Y., Zhang, K., Lu, X., Zhang, X., Gao, B., et al. (2022). Comparison of buckwheat genomes reveals the genetic basis of metabolomic divergence and ecotype differentiation. *New Phytol.* 235, 1927–1943. doi: 10.1111/nph.18306

He, Q., Ma, D., Li, W., Xing, L., Zhang, H., Wang, Y., et al. (2023). High-quality Fagopyrum esculentum genome provides insights into the flavonoid accumulation among different tissues and self-incompatibility. *J. Integr. Plant Biol.* 65, 1423–1441. doi: 10.1111/jipb.13459

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12, 1–14. doi: 10.1186/1471-2105-12-491

Huang, J., Wang, L., Tang, B., Ren, R., Shi, T., Zhu, L., et al. (2022). Integrated transcriptomics and widely targeted metabolomics analyses provide insights into flavonoid biosynthesis in the rhizomes of golden buckwheat (*Fagopyrum cymosum*). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.803472

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hunt, H. V., Shang, X., and Jones, M. K. (2018). Buckwheat: a crop from outside the major Chinese domestication centres? A review of the archaeobotanical, palynological and genetic evidence. *Vegetation History Archaeobotany* 27, 493–506. doi: 10.1007/s00334–017-0649–4

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11, 1–11. doi: 10.1186/1471-2105-11-119

Jing, R., Li, H. Q., Hu, C. L., Jiang, Y. P., Qin, L. P., and Zheng, C. J. (2016). Phytochemical and pharmacological profiles of three fagopyrum buckwheats. *Int. J. Mol. Sci.* 17, 589. doi: 10.3390/ijms17040589

Joshi, D. C., Zhang, K., Wang, C., Chandora, R., Khurshid, M., Li, J. B., et al. (2020). Strategic enhancement of genetic gain for nutraceutical development in buckwheat: A genomics-driven perspective. *Biotechnol. Adv.* 39, 107479. doi: 10.1016/j.biotechadv.2019.107479

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

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

Ke, H. L., Wang, X. P., Zhou, Z., Ai, W., Wu, Z. Y., and Zhang, Y. W. (2021). Effect of weimaining on apoptosis and Caspase-3 expression in a breast cancer mouse model. *J. Ethnopharmacology* 264, 113363. doi: 10.1016/j.jep.2020.113363

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

Lee, C.-C., Shen, S.-R., Lai, Y.-J., and Wu, S.-C. (2013). Rutin and quercetin, bioactive compounds from tartary buckwheat, prevent liver inflammatory injury. *Food Funct.* 4, 794–802. doi: 10.1039/c3fo30389f

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37, 4572–4574. doi: 10.1093/bioinformatics/btab705

Li, F., Gao, Y., Jin, C., Wen, X., Geng, H., Cheng, Y., et al. (2022). The chromosome-level genome of Gypsophila paniculata reveals the molecular mechanism of floral development and ethylene insensitivity. *Horticulture Res.* 9, uhac176. doi: 10.1093/hr/uhac176

Li, Y., Wang, Z., Zhu, M., Niu, Z., Li, M., Zheng, Z., et al. (2023). A chromosome-scale Rhubarb (*Rheum tanguticum*) genome assembly provides insights into the evolution of anthraquinone biosynthesis. *Commun. Biol.* 6, 867. doi: 10.1038/s42003-023-05248-5

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021a). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021b). BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* 1, e323. doi: 10.1002/cpz1.323

McGrath, J. M., Funk, A., Galewski, P., Ou, S., Townsend, B., Davenport, K., et al. (2023). A contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.). *DNA Res.* 30, dsac033. doi: 10.1093/dnares/dsac033

Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

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

Olivier, J., Jean-Marc, A., Benjamin, N., Alberto, P., Christian, C., Alberto, C., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148

Ou, S. J., Chen, J. F., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 46, e126. doi: 10.1093/nar/gky730

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., et al. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* 9, giaa100. doi: 10.1093/gigascience/giaa100

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

Ranallo-Benavidez, T., Jaron, K., and Schatz, M. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11, 1432. doi: 10.1038/s41467–020-14998–3

Roach, M. J., Schmidt, S. A., and Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* 19, 1–10. doi: 10.1186/s12859-018-2485-7

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

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122

Shen, N., Wang, T., Gan, Q., Liu, S., Wang, L., and Jin, B. (2022). Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* 383, 132531. doi: 10.1016/j.foodchem.2022.132531

Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. doi: 10.1093/nar/gkl200

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi: 10.1038/nbt.3988

Sunil, M., Hariharan, A. K., Nayak, S., Gupta, S., Nambisan, S. R., Gupta, R. P., et al. (2014). The draft genome and transcriptome of Amaranthus hypochondriacus: a C4 dicot producing high-lysine edible pseudo-cereal. *DNA Res.* 21, 585–602. doi: 10.1093/dnares/dsu021

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141. doi: 10.1016/j.xinn.2021.100141

Xu, D., Hu, M.-J., Wang, Y.-Q., and Cui, Y.-L. (2019). Antioxidant activities of quercetin and its complexes for medicinal application. *Molecules* 24, 1123. doi: 10.3390/molecules24061123

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Yi, L., Sa, R., Zhao, S., Zhang, X., Lu, X., Mu, Y., et al. (2022). Chromosome-scale, haplotype-resolved genome assembly of Suaeda Glauca. *Front. Genet.* 13. doi: 10.3389/fgene.2022.884081

Zhang, L.-L., He, Y., Sheng, F., Hu, Y.-F., Song, Y., Li, Y., et al. (2021). Towards a better understanding of *Fagopyrum dibotrys*: a systematic review. *Chin. Med.* 16, 1–18. doi: 10.1186/s13020-021-00498-z

Zhang, L., Li, X., Ma, B., Gao, Q., Du, H., Han, Y., et al. (2017). The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol. Plant* 10, 1224–1237. doi: 10.1016/j.molp.2017.08.013

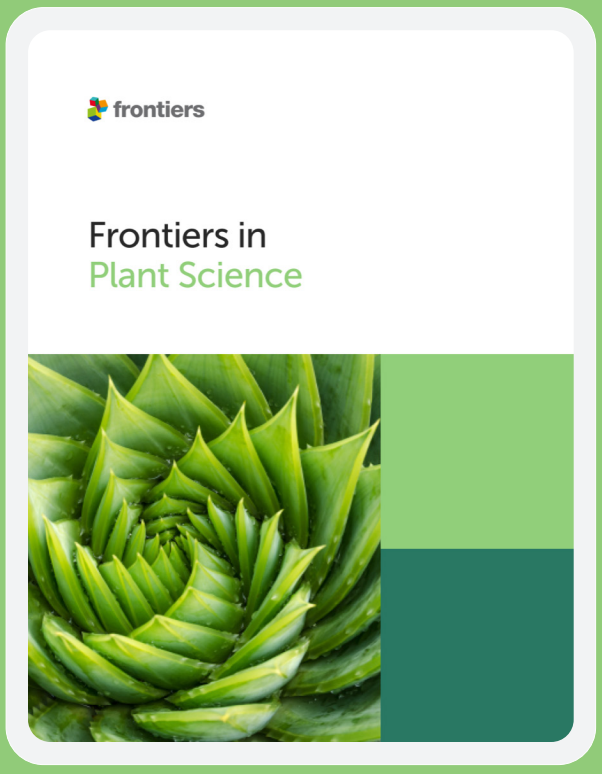
Frontiers in
Plant Science

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