EXPLORING AND ENGINEERING PLANT SPECIALIZED METABOLISM: LATEST ADVANCES AND NEW HORIZONS

EDITED BY: Thu Thuy Dang, Jakob Franke and Yang Zhang

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-900-6
DOI 10.3389/978-2-88971-900-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EXPLORING AND ENGINEERING PLANT SPECIALIZED METABOLISM: LATEST ADVANCES AND NEW HORIZONS

Topic Editors:
Thu Thuy Dang, University of British Columbia Okanagan, Canada
Jakob Franke, Leibniz University Hannover, Germany
Yang Zhang, Sichuan University, China

Citation: Dang, T. T., Franke, J., Zhang, Y., eds. (2021). Exploring and Engineering Plant Specialized Metabolism: Latest Advances and New Horizons. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-900-6

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 28 October 2021
doi: 10.3389/fpls.2021.783465

Editorial: Exploring and Engineering Plant Specialized Metabolism: Latest Advances and New Horizons

Jakob Franke[1,2]*, *Yang Zhang*[3]* and *Thu-Thuy T. Dang*[4]*

[1] Institute of Botany, Leibniz University Hannover, Hanover, Germany, [2] Centre of Biomolecular Drug Research, Leibniz University Hannover, Hanover, Germany, [3] Key Laboratory of Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China, [4] Department of Chemistry, Irving K. Barber Faculty of Science, The University of British Columbia, Kelowna, BC, Canada

Keywords: genomics, transcriptomics, plant biochemistry, transcription factors, metabolic engineering

Editorial on the Research Topic

Exploring and Engineering Plant Specialized Metabolism: Latest Advances and New Horizons

OPEN ACCESS

Edited by:
Reuben J. Peters, Iowa State University, United States

Reviewed by:
Bjoern Hamberger, Michigan State University, United States

***Correspondence:**
Jakob Franke
jakob.franke@botanik.uni-hannover.de
Yang Zhang
yang.zhang@scu.edu.cn
Thu-Thuy T. Dang
thuy.dang@ubc.ca

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

Received: *26 September 2021*
Accepted: *11 October 2021*
Published: *28 October 2021*

Citation:
Franke J, Zhang Y and Dang T-TT (2021) Editorial: Exploring and Engineering Plant Specialized Metabolism: Latest Advances and New Horizons. Front. Plant Sci. 12:783465. doi: 10.3389/fpls.2021.783465

Plants use specialized metabolic pathways to produce over 200,000 small molecules which often have potent biological activities. Many of these compounds have medicinal, nutritional or other applications. However, the natural supply from the producing plants is often strongly limited, for example because plants produce insufficient quantities or do not produce biomass fast enough. There is therefore an urgent need to improve our understanding of plant specialized metabolism, and to come up with strategies to engineer the underlying pathways. This is especially true in the light of climate change and the mandate to transition to a biobased economy. For that reason, this Research Topic aims to collate recent developments in the area of plant specialized metabolism, with a special focus on cutting-edge methods to explore as well as engineer biosynthetic pathways.

Probably the most important technical progress of the last decade has been achieved in the field of sequencing platforms. Nowadays, obtaining transcriptome and even genome data of non-model plant species is a realistic option even for smaller labs. For example, high-quality genome assemblies up to the chromosome level have been reported recently for medicinal plants such as *Senna tora* (Kang et al., 2020), *Camptotheca acuminata* (Zhao et al., 2017; Kang et al., 2021), *Ophiorrhiza pumila* (Rai et al., 2021), *Papaver somniferum* (opium poppy; Guo et al., 2018; Li et al., 2020), and *Taxus chinensis* var. *mairei* (Xiong et al., 2021) as well as for culinary herbs from the mint family (Lamiaceae; Bornowski et al., 2020; Lichman et al., 2020). This advance opens up numerous avenues to gain a better understanding of plant specialized metabolism on a transcriptome- or genome-wide level.

The importance of sequencing data for investigating the biochemistry of understudied plant species is underlined very well by the publications in this Research Topic: For example, Yamada et al. performed a genome-wide profiling of WRKY transcription factor genes in California poppy (*Eschscholzia californica*) to study their effect on benzylisoquinoline alkaloid biosynthesis. A similar approach was used by Cao et al. with a focus on MYB transcription factor genes in Chinese Bayberry (*Morella rubra*) to investigate flavonoid metabolism. Li et al. used genome data from red sage (*Salvia miltiorrhiza*) to identify TIFY transcription factors involved in regulation of specialized metabolism. Lastly, Zhang et al. combined multiple omics techniques to gain a better understanding of how lipid and fatty acid synthesis is regulated in sesame seeds (*Sesamum indicum*).

While all of these studies demonstrate well how current omics techniques can be applied to understand regulatory circuits of already known plant metabolic pathways, there remains a much larger number of pathways that have yet to be elucidated. The key challenge here is to

identify the genes and enzymes involved in these pathways, which is often a slow and tedious process and requires an efficient bioinformatic and biochemical pipeline. However, several breakthrough publications in the last years demonstrate that plant pathway elucidation is now becoming increasingly feasible (Caputi et al., 2018; Dang et al., 2018; Christ et al., 2019; Hodgson et al., 2019; Pluskal et al., 2019; Nett et al., 2020). In the course of these and numerous other projects, many unusual and powerful enzymes have been discovered, which are also attractive from a biocatalysis perspective. A particularly important class of enzymes are cytochromes P450, whose broad repertoire of catalytical function was reviewed by Nguyen and Dang. Again, analyzing genome and transcriptome data *via* state-of-the-art bioinformatics approaches has been key to discovering novel biosynthetic genes and enzymes from plants. Plant biosynthetic gene clusters are now reported more and more frequently, as plant genomic data can be obtained more readily. An overview over currently known plant biosynthetic gene clusters is provided by Bharadwaj et al..

Understanding how plant metabolic pathways work and are regulated is key to engineer them successfully. A strategy that is commonly used is to transfer these pathways into baker's yeast (*Saccharomyces cerevisiae*) as a versatile and easy-to-handle eukaryotic host system. As reviewed by Utomo et al., this success is particularly based on CRISPR/Cas9-based techniques for multiplex genome editing. However, not only microorganisms are attractive hosts for pathway engineering. Thanks to advancements in the fields of genome editing and plant biochemistry, original producer plants are now often engineered rationally as well. Examples from engineering terpenoid metabolism in glandular trichomes of Lamiaceae plants are reviewed by Mahmoud et al. How a better understanding of plant metabolism can translate into a relevant application is also demonstrated by the article of Koudounas et al.; in their work, they successfully silenced a gene involved in secoiridoid biosynthesis in olives (*Olea europaea*), which might be valuable to improve the effects of olive oil on human health.

As demonstrated by this Research Topic, the impact of modern sequencing techniques, bioinformatics analysis platforms, state-of-the-art biochemical approaches and new genetic engineering techniques to the field of plant specialized metabolism has been tremendous. This has enabled various biochemical discoveries and engineering applications, which would not have been possible only a few years ago. We are looking forward to seeing further progress in understanding plant specialized metabolism and additional real-world applications of pathway engineering in the years to come.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

YZ was supported by the Institutional Research Fund of Sichuan University (2020SCUNL106) and the Fundamental Research Funds for the Central Universities (SCU2021D006).

REFERENCES

Bornowski, N., Hamilton, J. P., Liao, P., Wood, J. C., Dudareva, N., and Buell, C. R. (2020). Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. *DNA Res.* 27:dsaa016. doi: 10.1093/dnares/dsaa016

Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T.-D., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in *Madagascar periwinkle*. *Science* 360, 1235–1239. doi: 10.1126/science.aat4100

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

Dang, T.-T. T., Franke, J., Carqueijeiro, I. S. T., Langley, C., Courdavault, V., and O'Connor, S. E. (2018). Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat. Chem. Biol.* 14, 760–763. doi: 10.1038/s41589-018-0078-4

Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., He, Z., et al. (2018). The opium poppy genome and morphinan production. *Science* 2018:eaat4096. doi: 10.1126/science.aat4096

Hodgson, H., Peña, R. D. L., Stephenson, M. J., Thimmappa, R., Vincent, J. L., Sattely, E. S., et al. (2019). Identification of key enzymes responsible for protolimonoid biosynthesis in plants: opening the door to azadirachtin production. *Proc. Natl. Acad. Sci. U. S. A.* 2019:201906083. doi: 10.1073/pnas.1906083116

Kang, M., Fu, R., Zhang, P., Lou, S., Yang, X., Chen, Y., et al. (2021). A chromosome-level Camptotheca acuminata genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* 12:3531. doi: 10.1038/s41467-021-23872-9

Kang, S.-H., Pandey, R. P., Lee, C.-M., Sim, J.-S., Jeong, J.-T., Choi, B.-S., et al. (2020). Genome-enabled discovery of anthraquinone biosynthesis in *Senna tora*. *Nat. Commun.* 11:5875. doi: 10.1038/s41467-020-19681-1

Li, Q., Ramasamy, S., Singh, P., Hagel, J. M., Dunemann, S. M., Chen, X., et al. (2020). Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-020-15040-2

Lichman, B. R., Godden, G. T., and Buell, C. R. (2020). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of Lamiaceae. *Curr. Opin. Plant Biol.* 55, 74–83. doi: 10.1016/j.pbi.2020.03.005

Nett, R. S., Lau, W., and Sattely, E. S. (2020). Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* 8, 1–6. doi: 10.1038/s41586-020-2546-8

Pluskal, T., Torrens-Spence, M. P., Fallon, T. R., Abreu, A. D., Shi, C. H., and Weng, J.-K. (2019). The biosynthetic origin of psychoactive kavalactones in kava. *Nat. Plants* 5, 867–878. doi: 10.1038/s41477-019-0474-0

Rai, A., Hirakawa, H., Nakabayashi, R., Kikuchi, S., Hayashi, K., Rai, M., et al. (2021). Chromosome-level genome assembly of *Ophiorrhiza pumila* reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* 12:405. doi: 10.1038/s41467-020-20508-2

Xiong, X., Gou, J., Liao, Q., Li, Y., Zhou, Q., Bi, G., et al. (2021). The Taxus genome provides insights into paclitaxel biosynthesis. *Nat. Plants* 2021, 1–11. doi: 10.1101/2021.04.29.441981

Zhao, D., Hamilton, J. P., Pham, G. M., Crisovan, E., Wiegert-Rininger, K., Vaillancourt, B., et al. (2017). *De novo* genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *GigaScience* 6, 1–7. doi: 10.1093/gigascience/gix065

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fpls.2021.630424

Genome-Wide Identification of the TIFY Family in *Salvia miltiorrhiza* Reveals That SmJAZ3 Interacts With SmWD40-170, a Relevant Protein That Modulates Secondary Metabolism and Development

Lin Li†, Yuanchu Liu†, Ying Huang†, Bin Li, Wen Ma, Donghao Wang, Xiaoyan Cao and Zhezhi Wang**

National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest of China, Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, College of Life Sciences, Shaanxi Normal University, Xi'an, China

OPEN ACCESS

Edited by:
Jakob Franke, Leibniz University Hannover, Germany

Reviewed by:
Aijia Ji, Guangzhou University of Chinese Medicine, China
Zhichao Xu, Chinese Academy of Medical Sciences and Peking Union Medical College, China
Carlos R. Figueroa, University of Talca, Chile

***Correspondence:**
Xiaoyan Cao
caoxiaoyan@snnu.edu.cn
Zhezhi Wang
zzwang@snnu.edu.cn

† These authors have contributed equally to this work

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

Received: *17 November 2020*
Accepted: *26 January 2021*
Published: *18 February 2021*

Citation:
Li L, Liu Y, Huang Y, Li B, Ma W, Wang D, Cao X and Wang Z (2021) Genome-Wide Identification of the TIFY Family in Salvia miltiorrhiza Reveals That SmJAZ3 Interacts With SmWD40-170, a Relevant Protein That Modulates Secondary Metabolism and Development. Front. Plant Sci. 12:630424. doi: 10.3389/fpls.2021.630424

Salvia miltiorrhiza Bunge (*S. miltiorrhiza*), a traditional Chinese medicinal herb, contains numerous bioactive components with broad range of pharmacological properties. By increasing the levels of endogenous jasmonate (JA) in plants or treating them with methyl jasmonate (MeJA), the level of tanshinones and salvianolic acids can be greatly enhanced. The jasmonate ZIM (JAZ) proteins belong to the TIFY family, and act as repressors, releasing targeted transcriptional factors in the JA signaling pathway. Herein, we identified and characterized 15 TIFY proteins present in *S. miltiorrhiza*. Quantitative reverse transcription PCR analysis indicated that the *JAZ* genes were all constitutively expressed in different tissues and were induced by MeJA treatments. SmJAZ3, which negatively regulates the tanshinones biosynthesis pathway in *S. miltiorrhiza* and the detailed molecular mechanism is poorly understood. SmJAZ3 acts as a bait protein to capture and identify a WD-repeat containing the protein SmWD40-170. Further molecular and genetic analysis revealed that SmWD40-170 is a positive regulator, promoting the accumulation of secondary metabolites in *S. miltiorrhiza*. Our study systematically analyzed the TIFY family and speculated a module of the JAZ-WD40 complex provides new insights into the mechanisms regulating the biosynthesis of secondary metabolites in *S. miltiorrhiza*.

Keywords: *Salvia miltiorrhiza*, jasmonate, TIFY proteins, SmJAZ3, SmWD40-170

Abbreviations: JA, jasmonate; MeJA, methyl jasmonate; JAZ, jasmonate ZIM; WD40 protein, WD40 repeat-containing protein; TF, transcription factor; CDS, coding sequence; RT-qPCR, quantitative reverse transcription PCR; LC/MS, liquid chromatography/mass spectrometry; CK, control check; OE, overexpression; RNAi, ribonucleic acid interference; YFP, yellow fluorescent protein; GFP, green fluorescent protein; Y2H, yeast two-hybrid; BiFC, bimolecular fluorescence complementation; EAR, ethylene-responsive element binding factor-associated amphiphilic repression; SA, salicylic acid; MBS, MYB binding site; HSE, heat shock element; DRE, dehydration responsive element; WUN, wound.

INTRODUCTION

The plant-specific TIFY family is characterized by a highly conserved motif (TIF[F/Y]XG) positioned within a TIFY domain of approximately 28-amino acids (aa) (Vanholme et al., 2007; Bai et al., 2011). According to phylogenetic and structural analyses, genes in that family can be assigned to four subgroups: TIFY, JAZ, PEAPOD (PPD), and ZIM-like (ZML). The TIFY subfamily proteins contain only a TIFY domain, whereas the ZML subfamily, including the ZIM and ZML proteins, contain a C2C2-GATA zinc-finger domain and a CCT domain (CONSTANS, CO-like, TOC1) (Staswick, 2008; Chung et al., 2009). In addition to the TIFY domain, the JAZ subfamily proteins are characterized by a conserved Jas motif of approximately 27 aa, near the C-terminal. These Jas sequences possess the characteristic SLX2FX2KRX2RX5PY motif and are similar to the N-terminal portion of the CCT domain (Staswick, 2008; Chung et al., 2009). In contrast, the PPD subfamily proteins possess a characteristic N-terminal PPD domain, and a modified Jas motif, that lacks the conserved PY (proline tyrosine) in the C-terminal region (Chung et al., 2009). Genes in the TIFY family have previously been systematically analyzed in several plant species, including 36 in *Brassica rapa*, 21 in *Brachypodium distachyon*, 34 in *Glycine soja*, 19 in grape (*Vitis vinifera*), 20 in rice (*Oryza sativa*), and 18 in *Arabidopsis* (Vanholme et al., 2007; Ye et al., 2009; Bai et al., 2011; Zhang et al., 2012, 2015; Zhu et al., 2013; Saha et al., 2016). The members of this family are involved in regulating diverse aspects of plant development, responses to abiotic stresses, and phytohormone treatments. For example, the *PvTIFY* gene plays a vital role in the adaptation of *Phaseolus vulgaris* to phosphorus (P) starvation by mediating JA signaling (Aparicio-Fabre et al., 2013). Certain *VvTIFY* genes in grapes can be induced by osmotic, low temperature, or drought, salinity conditions, as well as jasmonic acid (JA), or abscisic acid (ABA) treatments (Zhang et al., 2012). In rice, most *OsTIFY* genes are responsive to at least one type of abiotic stress, such as drought, salinity, or low temperatures (Ye et al., 2009). Perhaps the best-characterized members are the *JAZ* genes, which play a key role in the JA pathway (Chini et al., 2007; Thines et al., 2007; Yan et al., 2007; Garrido-Bigotes et al., 2019). Jasmonoyl-isoleucine (JA-Ile), the bioactive JA, is an important plant hormone that regulates various biological processes, including plant development, defense processes, and secondary metabolism (Creelman and Mullet, 1997; Liechti and Farmer, 2002; Farmer et al., 2003; De Geyter et al., 2012; Wasternack and Hause, 2013). Proteomic analysis of the JAZ interacting proteins under MeJA treatments in *Eleusine coracana*, illustrated that EcJAZ acts as a signaling hub for JA and other phytohormone signaling pathways (Sen et al., 2016). Moreover, JAZ expression, which regulates and fine-tunes the expression of downstream JA-responsive genes, is differential in various pathways and with certain stress responses (Demianski et al., 2011).

Salvia miltiorrhiza (*S. miltiorrhiza*), a model medicinal plant, is a well-known traditional Chinese herb (Ma et al., 2012; Xu et al., 2015). Its dried roots have been used to treat cardiovascular and cerebrovascular disorders, such as coronary heart disease, hyperlipidemia, and acute ischemic strokes (Su et al., 2015; Liu et al., 2016). *S. miltiorrhiza* has two primary active compounds namely water-soluble phenolic acids, including caffeic acid, rosmarinic acid, and salvianolic acid B (Liu et al., 2006), and lipid-soluble tanshinones containing dihydrotanshinone, cryptotanshinone, tanshinone I, and tanshinone IIA (Kai et al., 2011). These natural products can accumulate at low levels in plants over a long period (Ma et al., 2015). Both biotic and abiotic elicitors can induce the accumulation of secondary metabolites in *S. miltiorrhiza* (Zhao et al., 2010). Treatment with MeJA can lead to marked increase in tanshinones and salvianolic acids levels, as well as the expression of genes involved in their biosynthesis (Gu et al., 2012; Luo et al., 2015). However, it remains unclear how JA modulates the synthesis of these secondary metabolites in *S. miltiorrhiza*. The process of JA signal transduction occurs in three main stages: (1) The bioactive JA, JA-Ile is recognized by coronatine-insensitive 1 (COI1) protein forming Skp1/Cullin1/F-box protein COI1 (SCF^{COI1}) complexes. (2) JAZ proteins are ubiquitinated by SCF^{COI1}-type E3 ubiquitin ligase and degraded by the 26S proteasome. (3) MYC TFs are released thereby inducing the expression of downstream genes (Farmer, 2007). The JAs, SCF^{COI1} receptor complex, JAZ repressors, and TFs are all involved in JA signal transduction (Xie et al., 1998; Xu et al., 2002; Boter et al., 2004; Thines et al., 2007; Chini et al., 2010). JAZ is an important juncture that represses responses to JA by interacting with bHLH-TFs (MYC2, MYC3, MYC4, MYC5, GL3, EGL3, and TT8) and R2R3-MYB TFs (PAP, GL1, MYB21, and MYB24) (Fernández-Calvo et al., 2011; Song et al., 2011, 2013; Fonseca et al., 2014; Qi et al., 2014; Garrido-Bigotes et al., 2020). Certain JAZ proteins and their interaction partners have been identified in *S. miltiorrhiza*. SmMYC2a and SmMYC2b may interact with SmJAZ1 and SmJAZ2 to positively regulate tanshinones and salvianolic acid B production (Zhou et al., 2016; Yang et al., 2017). SmJAZ8 is a repressor involved in JA-induced biosynthesis of salvianolic acids and tanshinones via interactions with SmMYC2a (Ge et al., 2015; Pei et al., 2018). SmJAZ3 and SmJAZ9 negatively regulate tanshinones biosynthesis and JA signaling pathway in *S. miltiorrhiza* (Shi et al., 2016). SmJAZ9 can interact with AtMYC2, whereas SmJAZ3 cannot. In our study, we investigated the interaction partners of SmJAZ3 protein and molecular mechanism underlaying SmJAZ3 role in JA signaling pathway to regulate secondary metabolism.

WD40 repeat (WDR)-containing proteins feature a conserved sequence of approximately 40 amino acids, identified as the WD40 motif, which begins with glycine-histidine (Gly-His) and ends with tryptophan-aspartate (Trp-Asp) (Neer et al., 1994; Smith et al., 1999). Without a DNA-binding site, WD40 functions as a rigid scaffold for protein–protein and protein–DNA interactions, rather than directly regulating gene expression (Ramsay and Glover, 2005). Among the WD40 families, TTG1 has been the most studied regarding secondary metabolism. AtTTG1 and ortholog *TTG1* genes from *Zea mays*, *Medicago truncatula*, *Vitis vinifera,* and *Malus domestica* have been reported to be involved in the biosynthesis of anthocyanins or flavonoids (Walker et al., 1999; Carey et al., 2004; Pang et al., 2009; Matus et al., 2010;

An et al., 2012). We have previously reported that SmTTG1 increased salvianolic acid B accumulations by forming the SmTTG1-SmMYB111-SmbHLH51 ternary transcription complex (Li et al., 2018). We also identified 225 *SmWD40* genes and analyzed the evolutionary relationship, gene structure, and conserved protein motif, which provided prelamination data for studying the function of SmWD40 in secondary metabolism (Liu et al., 2020).

In this study, we aimed to identify and characterize TIFY proteins present in *S. miltiorrhiza*. Phylogenetic trees, gene structures, and conserved motif analyses were conducted on the SmTIFY proteins. Further, a Y2H screening assay was used to elucidate the molecular mechanisms of SmJAZ3 in mediating JA signaling and secondary metabolism. The functions of its interaction partner, SmWD40-170, were then also identified and analyzed. We hypothesize that the molecular mechanism underlaying JAZ3-mediated regulation of secondary metabolism is via the JAZ-WD40 regulatory module in *S. miltiorrhiza*.

MATERIALS AND METHODS

Plant Material

Salvia miltiorrhiza plants were acquired from Shangluo County, Shaanxi Province, China, and maintained in our laboratory at Shaanxi Normal University, Xi'an, China. Root, stem, leaf, and flower samples were collected from uniformly grown 2-year-old plants and immediately frozen in liquid nitrogen.

Tissue culture-derived plants of *S. miltiorrhiza* were used for the JA treatment experiments. After 30 days of culturing under normal laboratory conditions, the plants were assigned to two groups: (1) mock control, in which the leaves were sprayed with 10% ethanol; or (2) JA treatment, in which the leaves were sprayed with a solution of 100 μM MeJA plus 10% ethanol (Gu et al., 2012). Each experiment was performed three times (n = 3 per group) and whole plants were harvested from each group after 0.5 h, 1 h, 2 h, 4 h, 8 h, and 12 h. All samples were immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction.

Identification and Characterization of TIFY Family Genes and Phylogenetic Analyses

To identify all of the putative *SmTIFY* family members, we used the sequences for 18 *Arabidopsis TIFY* family genes, as well as 20 from rice, and 19 from grape, that were obtained from the TAIR databases[1], TIGR databases[2], and the Grape Genome Database[3], respectively. Local BLASTs were conducted with the *S. miltiorrhiza* genome (Xu et al., 2016). The protein sequences database for *Arabidopsis*, rice, and grape TIFY proteins were used as queries for BioEdit. The SmTIFYs were determined by screening for the conserved TIFY domains using an NCBI conserved domain search[4], Pfam[5], HM-MER[6], InterPro[7], and SMART[8] online tools. MEGA 6 software[9] was used to investigate the phylogenetic relationships among TIFY proteins in the four species, based on the neighbor-joining algorithm and the bootstrap method (1000 replicates).

Sequence Analysis of the SmTIFY Proteins

Relative molecular weights and isoelectric points (pI) of the TIFY family members were analyzed using ExPASy[10]. Subcellular localization of the proteins was determined according to WoLF PSORT[11]. The DNA and CDS sequences of the *SmTIFY* genes were submitted to the GSDS online tool[12] to analyze their gene structures. The SmTIFY protein sequences were submitted to the MEME web server[13] for analysis of the protein motifs. To explore potential *cis*-elements in the promoter sequences, 2000 bp of the SmTIFY genomic DNA upstream of the initiation codon (ATG) were downloaded, using the *S. miltiorrhiza* genome database. The promoter sequences were submitted to the Plant CARE database[14] to predict the *cis*-acting elements.

Expression Analysis of *SmTIFY* Genes in Different Tissues and Under MeJA Treatment

Total RNA was extracted from the *S. miltiorrhiza* root, stem, leaf, and flower samples. First-strand cDNA was reversed from the total mRNA, according to the instructions for the Prime Script® RT Master Mix (Takara). RT-qPCR primer sequences were designed using Primer Premier 5.0, and a housekeeping gene (β-*Actin*) was used as an internal control (**Supplementary Table 1**). Several key genes encoding enzymes such as phenylalanine ammonialyase (*PAL*), cinnamate 4-hydroxylase (*C4H*), hydroxycinnamate-CoA ligase (*4CL*), tyrosine aminotransferase (*TAT*), hydroxyl phenylpyruvate reductase (*HPPR*), rosmarinic acid synthase (*RAS*), 1-deoxy-D-xylulose-5-phosphate synthase (*DXS*), 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGR*), farnesyl diphosphate synthase (*FPPS*), geranylgeranyl diphosphate synthase (*GGPPS*), copalyl diphosphate synthase (*CPS*), and kaurene synthase-like synthase (*KSL*) were investigated using RT-qPCR (**Supplementary Table 1**). RT-qPCR analysis was conducted with SYBR Green (Takara Biotechnology) and a Roche LightCycler® 96 real-time PCR machine. All experiments were performed with three biological replicates. Relative expression levels were calculated based on the $2^{-\Delta\Delta Ct}$ method

[1]http://www.arabidopsis.org
[2]http://rice.plantbiology.msu.edu
[3]http://www.genoscope.cns.fr
[4]https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[5]http://pfam.xfam.org/
[6]http://www.ebi.ac.uk/Tools/hmmer/
[7]http://www.ebi.ac.uk/interpro/search/sequence-search
[8]http://smart.embl.de/
[9]http://www.megasoftware.net/
[10]http://web.expasy.org/protparam/
[11]http://www.genscript.com/wolf-psort.html
[12]http://gsds.gao-lab.org/
[13]http://meme-suite.org/
[14]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

(Vandesompele et al., 2002). Statistical significance was assessed using the Student's *t*-test.

Yeast Two-Hybrid (Y2H) Screening and Assay

The CDS and partial Jas domain of the SmJAZ3 were amplified with specific primers (**Supplementary Table 2**) and cloned into the pGBKT7 vector as bait, to search for interacting proteins. The autoactivation test of the SmJAZ3 in the yeast was conducted as previously described (Li et al., 2018). Based on the previously described protocols (Pan et al., 2017), the *S. miltiorrhiza* cDNA library expressed in the Y187 yeast cells was mated with the AH109 yeast cells expressing SmJAZ3, and then screened for the interacting partners of SmJAZ3.

The Y2H assay was conducted to confirm the interaction between SmJAZ3 and SmWD40-170. The CDS of SmWD40-170 was cloned into pGADT7 to fuse with the activation domain as the prey. According to the manufacturer's protocol for the Matchmaker Gold Yeast Two-Hybrid System (Clontech), BD-SmJAZ3 and AD-SmWD40-170 fusion constructs were co-transformed into yeast strain AH109, using the lithium acetate method (Gietz and Schiestl, 2007), and yeast cells were grown on SD/-Leu/-Trp medium. Positive clones were then selected on SD/-Ade/-His/-Leu/-Trp medium with X-α-gal. The sequences for the primers used are listed in **Supplementary Table 2.**

Bimolecular Fluorescence Complementation (BiFC) Assay

Based on the protocol from the Gateway technology manufacturer (Invitrogen, United States), the ORFs (without the termination codon) of *SmJAZ3* (*SmWD40-170*) were amplified from pMD19T-SmJAZ3 (pMD19T-SmWD40-170) with adaptor primers (**Supplementary Table 2**) and then cloned into the pDONR207 vector using a BP recombination reaction. For the BiFC assay, pDONR207-SmJAZ3 and pDONR207-WD40-170 were cloned into pEarleyGate202-YC and pEarleyGate201-YN, respectively, using the LR recombination reaction (Earley et al., 2010). Equal concentrations of the YN and YC recombinant plasmids were mixed before co-transformation. Co-transformed YC-SmJAZ3/YN and YN-SmWD40-170/YC served as negative controls. Then, the mixed plasmids were bombarded into onion epidermal cells, using particle bombardment with the Gene Gun PDS-1000, and then incubated at 28°C for 24 h. The fluorescence signals were detected using a Leica DM6000B microscope (Leica, Germany) at an excitation wavelength of 475 nm.

Determination of Salvianolic Acids and Tanshinones Concentrations by Liquid Chromatography/Mass Spectrometry (LC/MS) Analysis

Transgenic lines were obtained by *Agrobacterium*-mediated transformation method (Yan and Wang, 2007). Three OE lines (OE-3, OE-7, and OE-8) and three RNAi lines (i-11, i-14, and i-15) were acquired from our laboratory and used for further studies (Liu et al., 2020). Two-month-old culture seedlings were first cultivated in hydroponic cultures for 7 days and then transplanted into the soil medium (perlite: vermiculite: grass ash = 1:1:3). After 2 months of cultivation, the roots were removed, washed, dried in an oven at 30°C, and then ground to a powder. The salvianolic acids and tanshinones compounds were extracted, as previously described (Wang et al., 2018).

Separation of the lipophilic tanshinone and hydrophilic salvianolic acid was performed using a conventional Welch Ultimate XB-C18 column with two ion monitoring modes (2.1 × 150 mm, 3 μm, Agilent Corporation, MA, United States) and the following conditions: mobile phase A, acetonitrile; mobile phase B, 0.1% formic acid; injection volume, 5 μL; flow rate, 0.4 mL/min; gradient elution conditions, 0 min (25% A) → 5 min (10% B); ion source, AJS (Agilent jet) and ESI (electrospray ionization); quantitative detection, multiple reaction monitoring (MRM) mode. The negative ion mode detection of salvianolic acid B and rosmarinic acid had the following properties: salvianolic acid B detection range (m/z), 717 519; rosmarinic acid detection range (m/z), 359 → 161; fragment voltage, 130 V; collision energy, 20 eV. The ion pattern test for the tanshinone IIA had the following properties: tanshinone IIA detection range (m/z), 295.1 → 277.1, fragment voltage of 140 V, and collision energy of 32 eV. The ion pattern test for cryptotanshinone had the following properties: cryptotanshinone detection range (m/z), 297.1 → 254.1, fragment voltage of 140 V, and collision energy of 26 eV. The standards and samples were tested according to the above conditions, and standard curves were constructed. The peak area measured with the standard solution was the ordinate, while the standard concentration was the abscissa. Regression equations and linear coefficients were then calculated.

RESULTS

Identification and Characterization of the *TIFY* Family Genes in the *S. miltiorrhiza* Genome

In additions to the four *SmJAZ* genes that we have previously reported, a total of 15 *SmTIFY* genes (1 *TIFY*, 1 *PPD*, 3 *ZML*s, and 10 *JAZ*s) were identified in the *S. miltiorrhiza* genome (**Table 1**) (Ge et al., 2015; Xu et al., 2016). These were named according to the existing numbering system used for the phylogenetic tree of *Arabidopsis* (**Figure 1**). The lengths of the SmTIFY amino acids ranged from 122 to 455, which is higher than that of other species. Subcellular localization analysis indicated that 10 of the SmTIFY proteins were nuclear, while SmJAZ7, SmJAZ8, and SmJAZ10 were located in the chloroplasts, and SmJAZ1 and SmJAZ6 in the cytoplasm and mitochondria, respectively.

Phylogenetic Analysis of the TIFY Family Members

To investigate the evolutionary patterns and phylogenic relationships among the TIFYs in *S. miltiorrhiza* (15 proteins), grape (19 proteins), rice (20 proteins), and *Arabidopsis*

TABLE 1 | The basic information of *SmTIFY* family genes.

Gene ID	Accession no.	Genomic length (bp)	CDS length (bp)	Protein length (aa)	Molecular weight (Da)	pI	Sub-cellular localization
SmJAZ1	JQ936590	700	543	180	19815.67	9.14	Cyto
SmJAZ2	KX814385	855	642	213	22501.41	7.86	Nucl
SmJAZ3	KY225688	3367	1011	336	35515.09	9.16	Nucl
SmJAZ4	KY225684	2184	945	314	32463.59	9.38	Nucl
SmJAZ5	KY225689	4890	1368	455	49022.40	9.57	Nucl
SmJAZ6	KC864779	1968	705	234	25417.54	9.51	Mito
SmJAZ7	KX814384	962	372	123	14681.92	10.34	Chlo
SmJAZ8	JQ936591	557	369	122	13766.72	9.69	Chlo
SmJAZ9	KX814388	3129	921	306	32428.60	8.28	Nucl
SmJAZ10	KX814383	1902	540	179	19637.53	9.45	Chlo
SmTIFY8	KX814386	3427	1119	372	39330.10	8.82	Nucl
SmPPD	KX814387	4107	828	275	30085.00	5.79	Nucl
SmZML1	KY225685	2930	894	297	32564.05	6.06	Nucl
SmZML2	KY225686	2627	918	305	33295.74	5.54	Nucl
SmZML3	KY225687	4027	981	326	35776.06	5.07	Nucl

Cyto, cytoplasmic; Nucl, nuclear; Mito, mitochondrial; Chlo, Chloroplast.

(18 proteins), a neighbor-joining phylogenetic tree was constructed (**Figure 1**). All proteins fell into four major groups: TIFY, PPD, ZML, and JAZ. Among them, the ZML/ZIM proteins from all four species were clustered together. The TIFY proteins, which contain only a TIFY domain, were clustered into a single branch, except for OsTIFY. The PPD subfamily members comprised only AtPPD, VvPPD, and SmPPD. As expected, the JAZ subfamily members accounted for most proteins, including all the AtJAZ, VvJAZ, SmJAZ, and some OsTIFY (putative OsJAZ) proteins. SmJAZ proteins were clustered into four layers (JAZ I, JAZ II, JAZ III, and JAZ IV). SmJAZ1, SmJAZ2, SmJAZ5, and SmJAZ6 were assigned to JAZ I; SmJAZ7 and SmJAZ8, to JAZ II; and SmJAZ3, SmJAZ4, and SmJAZ9, to JAZ IV; JAZ III contained only one of these proteins, SmJAZ10.

Sequence Analysis of the SmTIFY Family

The full lengths of the CDSs with their corresponding genomic DNA sequences were compared to determine the number and positions of the exons and introns. Each *SmTIFY* gene has three to nine exons (**Figure 2B**). The structures in this family are diverse, especially within the SmJAZ subfamily. To identify the distribution of the conserved domains and the multiple sequence alignments among the SmTIFY proteins, we examined their sequences. The results from our MEME analysis showed that members of this family have six putative conserved domains, namely the TIFY domain, Jas domain, GATA zinc finger, CCT domain, PPD domain, and EAR-like motif (**Figure 2A**). In general, the TIFY domain contains 31 amino acids, with a highly conserved pattern of TIFYXG, T [L/I] SFXG, and SLSFQG (**Figure 2C**). All SmTIFY proteins include a TIFY domain, while all SmJAZ proteins have a Jas domain, with the conserved motif SLX2FX2KRX2RX5PY (**Figure 2D**). In addition, the N-terminals of SmJAZ1, SmJAZ2, SmJAZ5, and SmJAZ6 have EAR-like motifs. Three of the SmZML proteins (SmZML1, SmZML2, and SmZML3) contain a TIFY domain, Jas domain, CCT domain, and GATA zinc finger. The N-terminal of the SmPPD also has a PPD domain. In contrast to all other members, SmTIFY8 only has a TIFY domain.

Analysis of the *Cis*-Elements in the SmTIFY Family

A search of the PlantCARE database showed the promoter sequences of the SmTIFYs. Among them, were a series of *cis*-elements that are involved in responses to light, biotic and abiotic stresses, phytohormones (ABA, MeJA, auxin, SA, gibberellin, and ethylene), circadian rhythms, and fungal elicitors (**Supplementary Table 3**). We also identified TF binding sites, such as an MBS and G-box. Almost all members, except *SmPPD* and *SmZML1*, contained TC-rich repeats that are involved in plant defenses and stress responsiveness. All these genes, except *SmJAZ1*, *SmTIFY8*, and *SmZML2*, also have TCA-elements that are responsive to SA. *SmJAZ1*, *SmJAZ2*, *SmJAZ5*, *SmJAZ6*, *SmJAZ7*, and *SmZML2* have *cis*-elements involved in responses to MeJA. Except for *SmJAZ1* and *SmZML2*, all genes contain a HSE and MBS in their promoter sequences, and, except for *SmJAZ3,* have a G-box, which is a possible MYC2-binding motif (Yu et al., 2016). However, only *SmJAZ10* has C-repeat/DRE and WUN-motif elements that are involved in responses to cold, dehydration, and wounding. MBSI, an MBS that helps regulate genes for flavonoid biosynthesis, was only found in the *SmPPD* promoter sequence. All these *cis*-elements have essential roles in modulating gene expression, by controlling promoter efficiency. Therefore, these results provide vital information for further research into the functions of SmTIFY genes.

Expression Analysis of the *SmTIFY* Genes

To determine the function of the *SmTIFY* genes, we monitored the expression of these genes in four tissue types sampled from *S. miltiorrhiza* (**Figure 3**). The most highly expressed genes were

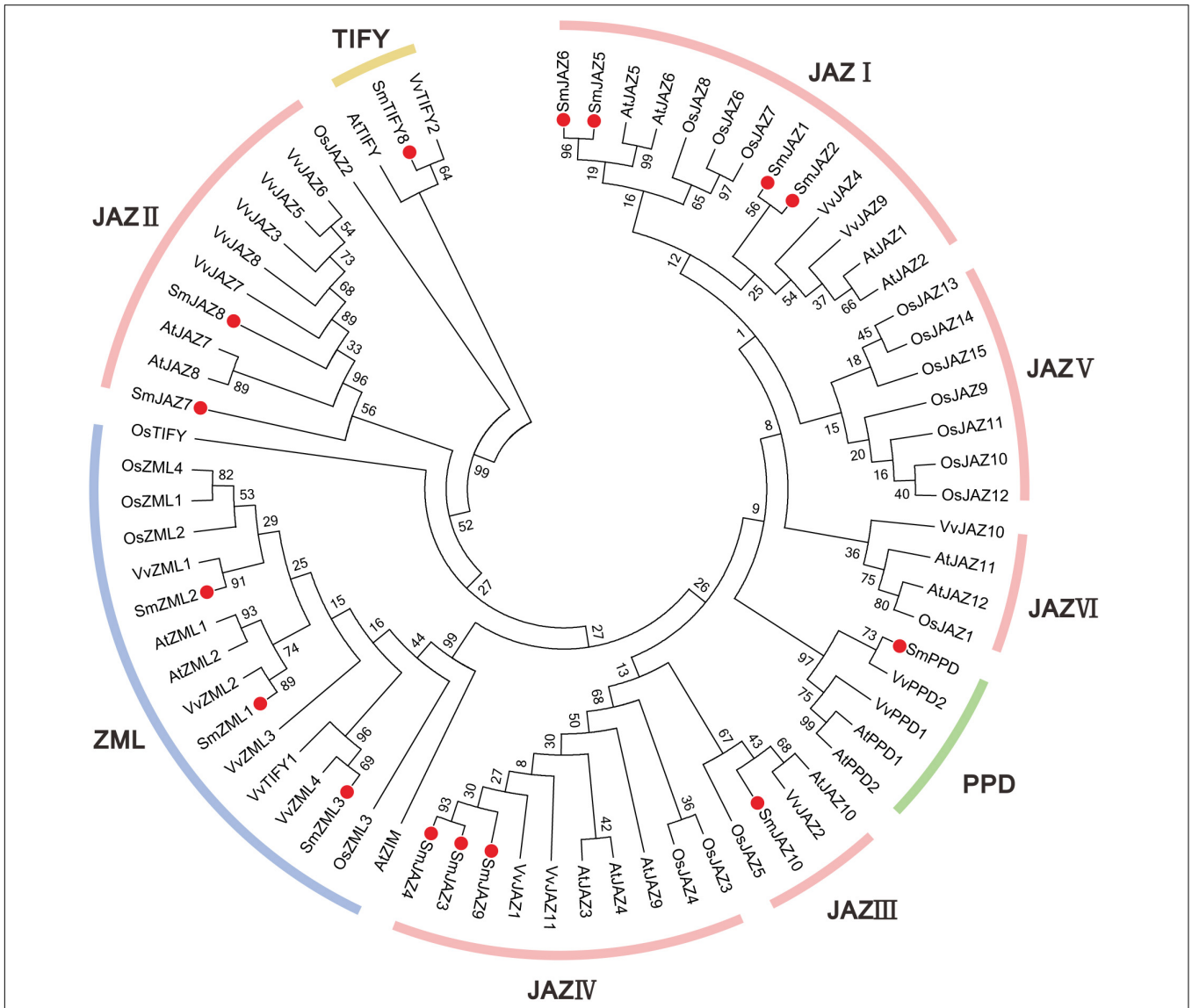


FIGURE 1 | Neighbor-joining phylogenetic tree of TIFY proteins from *Salvia miltiorrhiza*, *Arabidopsis*, *Vitis vinifera*, and *Oryza sativa*. The tree was developed using MEGA 6 software and the bootstrap method (1000 replicates).

as follows: *SmJAZ1*, *SmJAZ6*, *SmJAZ8*, and *SmJAZ9* in the leaf; *SmJAZ4*, *SmJAZ5*, *SmJAZ7*, *SmPPD*, *SmZML2* and *SmZML3*, in the flower; and *SmJAZ3* was predominantly expressed in the root. Expressions of *SmJAZ2*, *SmJAZ6*, *SmJAZ10*, *SmTIFY8*, and *SmZML1* were lower in the flower tissues, while the expressions of *SmJAZ4* and *SmJAZ7* were lower in the leaf tissues. Interestingly, the expression patterns of *SmJAZ1* and *SmJAZ8* were similar in different tissues. However, none of the genes in the same subfamily showed similar expression patterns, which indicates that each member plays an irreplaceable role. To investigate the role of the *SmTIFY* family genes in the JA signaling pathway, we monitored their expression in response to exogenous MeJA (**Figure 4**). Most genes were induced within 2 h of treatment and their expression continued to increase over time. In particular, the peak in expression of *SmJAZ3*, *SmJAZ4*, and *SmJAZ9* was delayed compared with that of the other *SmJAZ* genes. In contrast, MeJA inhibited the expression of *SmTIFY8*, *SmPPD*, and *SmZML1* at 2 h, but delayed the response of *SmZML2* and *SmZML3*.

Y2H Screening

Since the JA-responsive gene *SmJAZ3* negatively regulates tanshinones biosynthesis in *S. miltiorrhiza* (Shi et al., 2016), we selected *SmJAZ3* as a candidate gene, to explore the molecular mechanisms underlaying the regulation of JA-mediated secondary metabolism of *S. miltiorrhiza*. To determine whether SmJAZ3 affects potential target proteins or TFs, we screened for interacting proteins using a Y2H system. Initially, no autoactivation of the SmJAZ3 baits was detected (**Supplementary Figure 1C**). Since the results were not satisfactory when the full-length SmJAZ3 was used as the bait protein for the

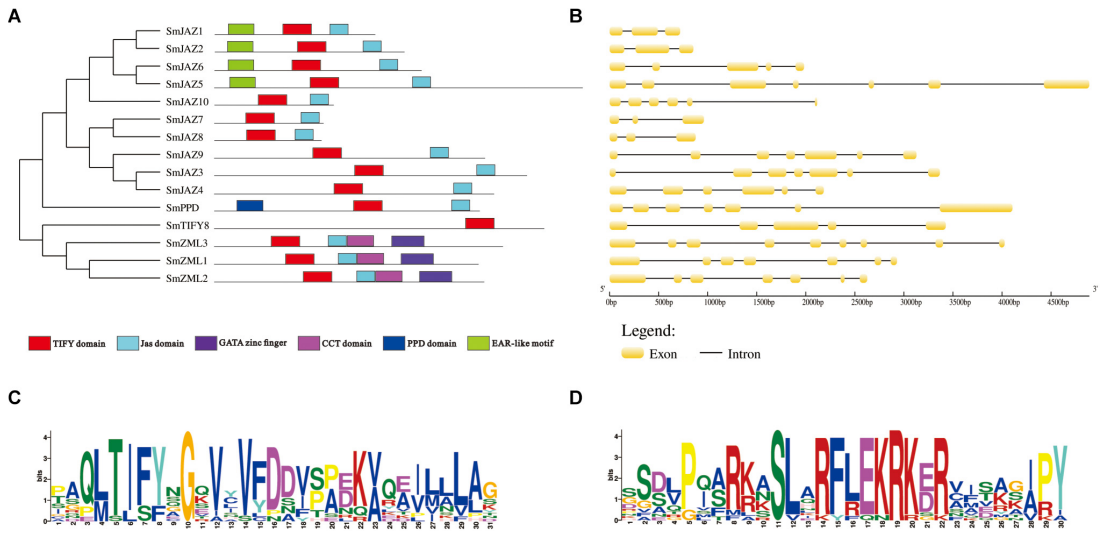


FIGURE 2 | (A) Distribution of conserved domains within SmTIFY, SmJAZ, SmPPD, and SmZML proteins. Relative positions of domains within each protein are shown in different colors. **(B)** Gene structures of *SmTIFY* gene family. Exon, yellow-filled boxes; intron, black single lines. **(C)** Sequence logos of TIFY domains from SmTIFY proteins. **(D)** Sequence logos of Jas domains from SmTIFY proteins.

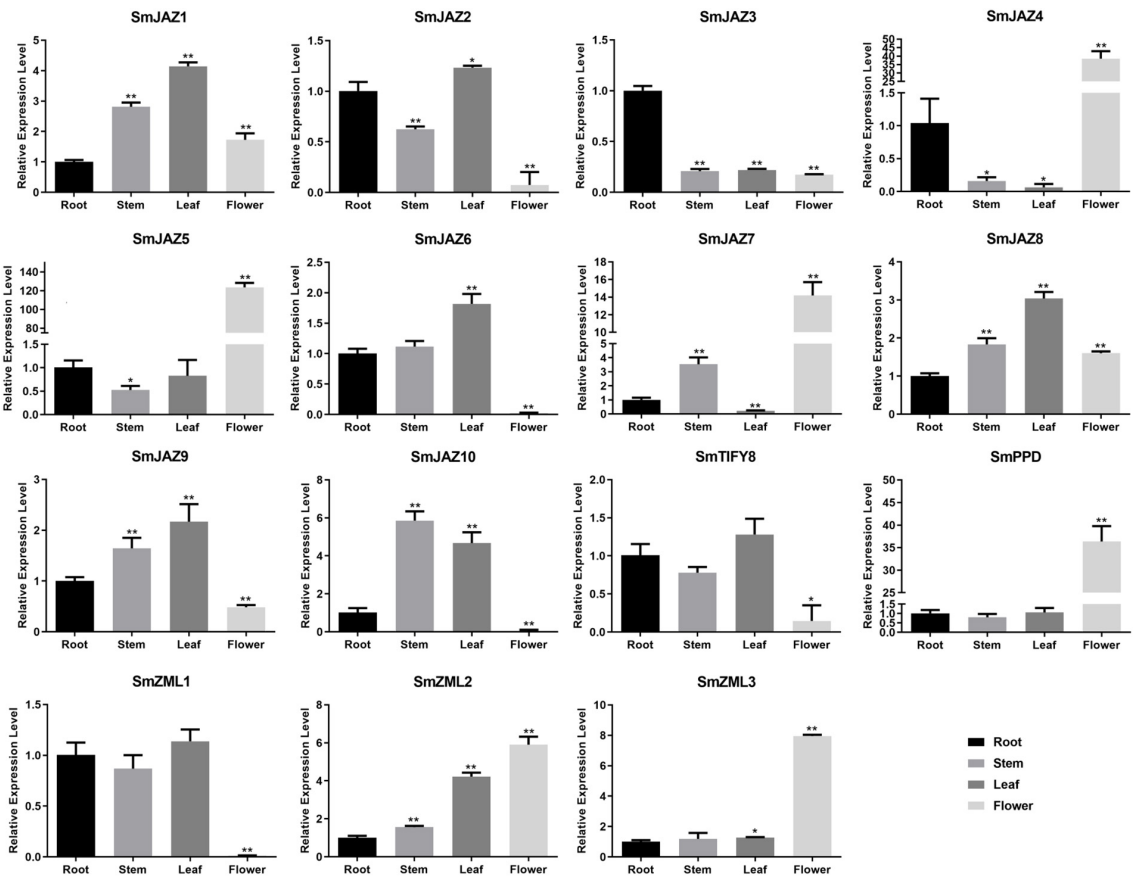


FIGURE 3 | Relative expression of *SmTIFY* genes in root, stem, leave, and flower. All data represent averages of three biological replicates, error bars indicate SD. Statistical significance was determined using the Student's *t*-test (*$p < 0.05$, **$p < 0.01$) between root, stem, leaf, and flower.

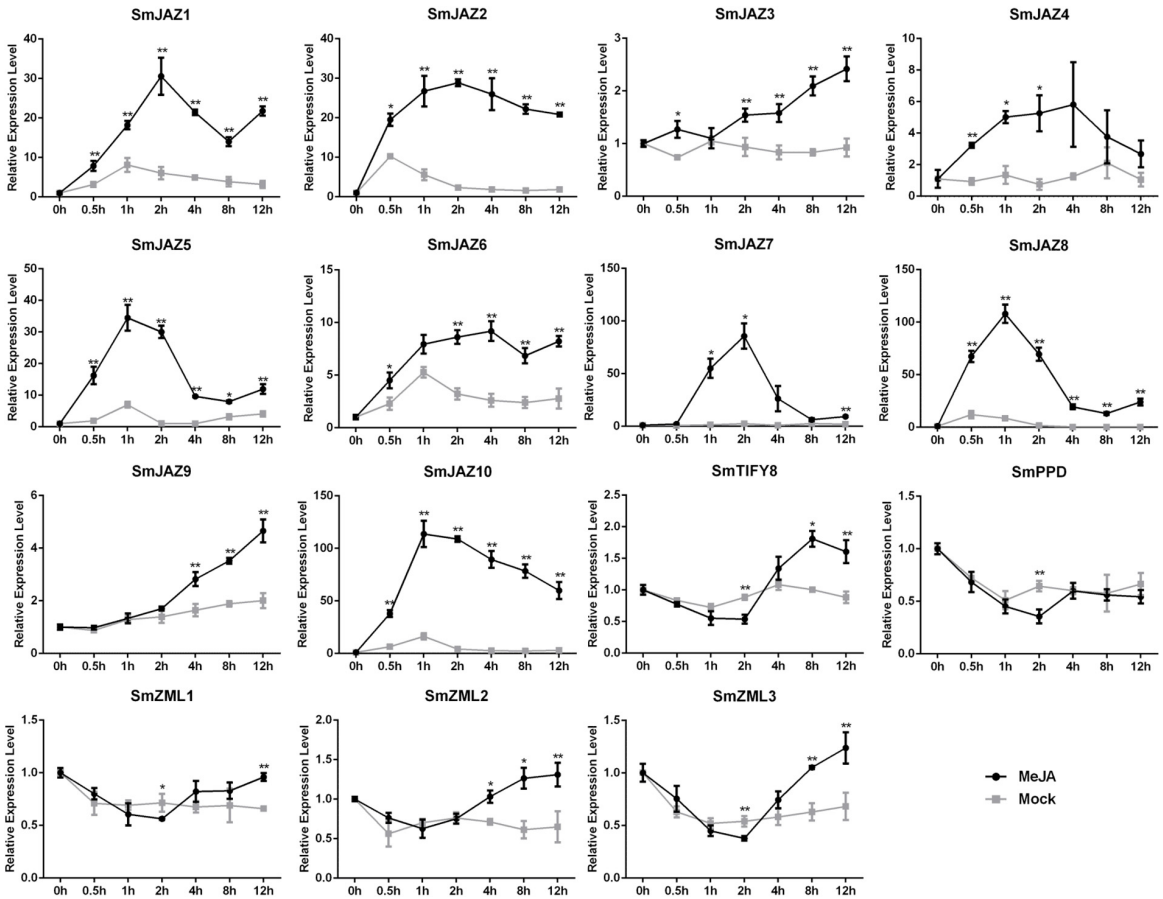


FIGURE 4 | Relative expression level of *SmTIFY* genes in *S. miltiorrhiza* plants treated with mock control and 100 μM MeJA. All data represent averages of three biological replicates, error bars indicate SD. Statistical significance was determined by the Student's *t*-test (*$p < 0.05$, **$p < 0.01$).

library screening, we used the Jas domain as the bait. The screening for the cDNA library in *S. miltiorrhiza*, resulted in 36 candidate proteins being identified, that interacted with SmJAZ3-Jas (**Supplementary Figure 1F** and **Supplementary Table 4**). Function annotation suggested that these candidate interaction proteins including JA signal member MYC2, as well as some enzymes are associated with biosynthesis process, metabolism process, and stress resistance (**Supplementary Table 4**). Interestingly, a WD40 protein, SmWD40-170 (Gene ID: ATA66299), was among them. The full-length CDS of the *SmWD40-170* was 972 bp, and it encoded a protein of 323 amino acids. It was located in the nucleus and the cytoplasm and is predicted to respond to MeJA-responsive elements (Liu et al., 2020). In addition, the SmWD40-170 responded to drought stress and regulated ABA- and H_2O_2-induced stomatal movements in the *S. miltiorrhiza* (Liu et al., 2020).

Interactions Between the SmJAZ3 and SmWD40-170 Proteins

We analyzed the interactions between the SmJAZ3 and SmWD40-170 proteins. For the Y2H, except for the full-length sequence, the SmJAZ3 was divided into three parts: N-terminal fragment (amino acids 1 to 275), Jas motif fragment (amino acids 276 to 304), and C-terminal fragment (amino acids 305 to 336), to examine whether other domains of the SmJAZ3 were responsible for the interaction with the SmWD40-170 protein (**Figure 5A**). The results showed that the full-length and Jas motif of the SmJAZ3 interacted with the SmWD40-170 in yeast (**Figure 5B**). A BiFC assay was then used to examine the Y2H results. SmWD40-170 and SmJAZ3 were fused with the N-terminal and C-terminal, respectively, of the YFP. As expected, a strong fluorescent signal was detected in the nucleus when SmWD40-170 and SmJAZ3 were co-transformed into onion epidermal cells, whereas no fluorescent signal was observed in the control groups (**Figure 5C**). Taken together, our results suggest that SmJAZ3 could interact with SmWD40-170.

SmWD40-170 Affects the Biosynthesis and Accumulation of Secondary Metabolites in *S. miltiorrhiza*

The levels of the total phenolic acids and total flavonoids in the OE-3, OE-7, and OE-8 lines, as well as in the i-11, i-14, and i-15 RNAi lines, were determined using the Folin–Ciocalteu method

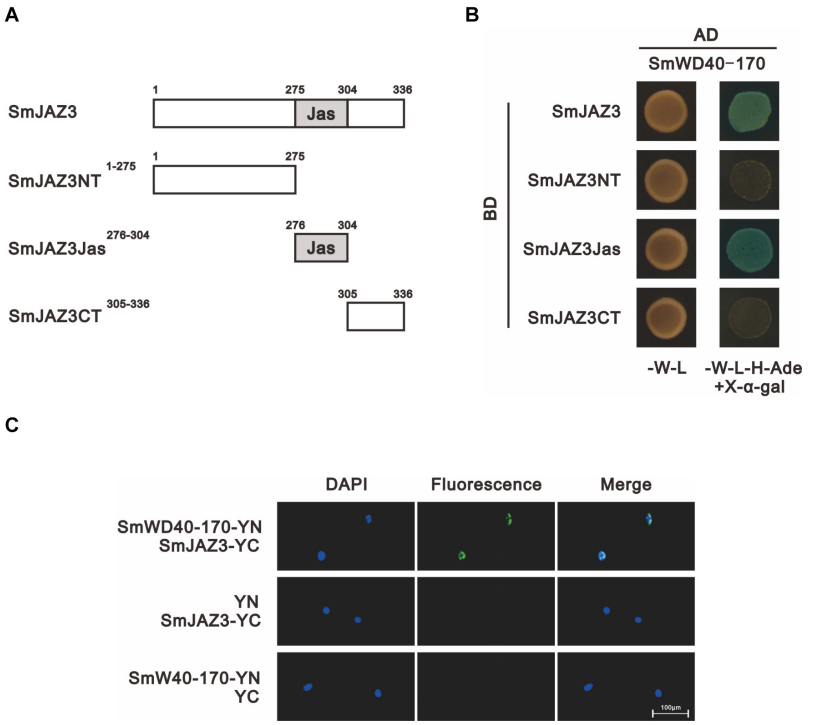


FIGURE 5 | SmJAZ3 interacts with SmWD40-170. **(A)** Schematic diagrams show domain constructs of SmJAZ3. **(B)** Y2H assays was used to test the interactions of SmWD40-170 with different domains of SmJAZ3. **(C)** BiFC assays was used to detect the interaction between SmJAZ3 and SmWD40-170.

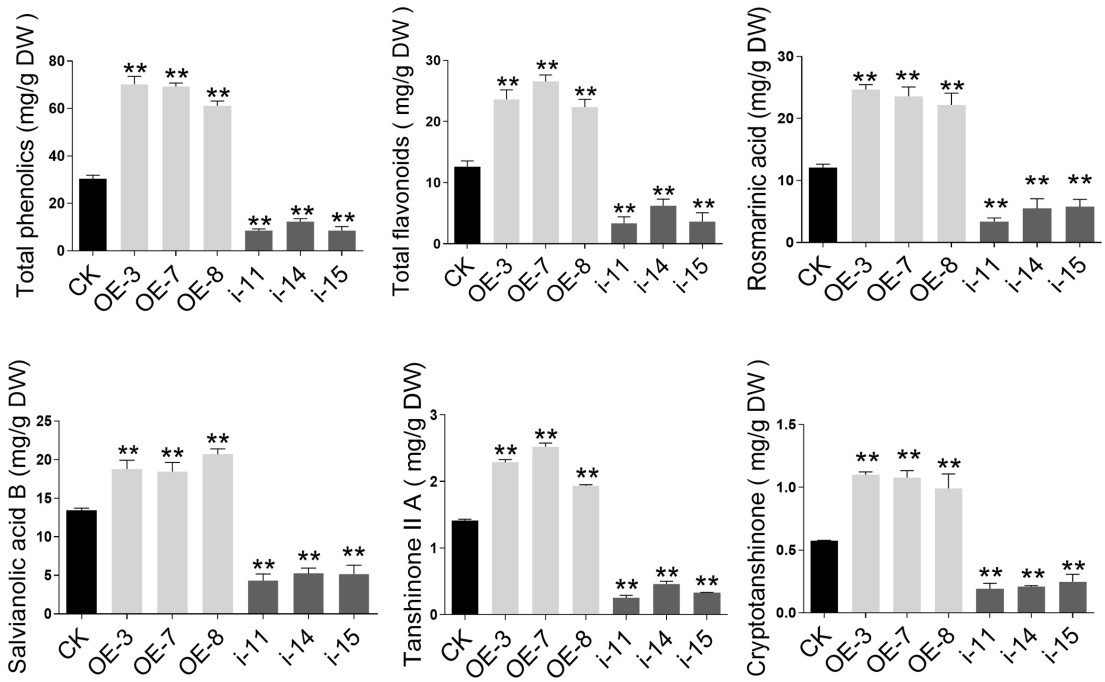


FIGURE 6 | Secondary metabolites contents in control and SmWD40-170 transgenic roots of *S. miltiorrhiza*. Comparisons of total phenolic acid, total flavonoids, rosmarinic acid, salvianolic acid B, tanshinone IIA, and cryptotanshinone concentration among transgenic and control lines. All data represent averages of three biological replicates, error bars indicate SD. Statistical significance was determined using the Student's *t*-test (**$p < 0.01$).

and the sodium nitrite-aluminum chloride colorimetric method (Dewanto et al., 2002; Chew et al., 2009) with 2-month-old roots of *S. miltiorrhiza*. Compared to the CK, the total phenolic acids and total flavonoids contents increased by more than 2 and 1.5 times, respectively, in the three OE lines. In the three RNAi lines, especially i-14, the contents of both total phenolic acids and total flavonoids were significantly decreased by 59.47% and 50.63%, respectively (**Figure 6**).

To further understand the effects of SmWD40-170 on the biosynthesis and accumulation of secondary metabolites in *S. miltiorrhiza*, the active components of salvianolic acids and tanshinones were more accurately determined and analyzed using LC/MS (**Supplementary Figure 2**). In comparison to CK, the concentrations of salvianolic acids and tanshinones were significantly increased in the OE-3, OE-7, and OE-8 lines (**Figure 6**). The rosmarinic acid content increased by 2.04 times, at most, and the salvianolic acid B content increased by approximately 1.54 times in the three OE lines, while the two tanshinones compounds increased by 1.37–1.78 times for tanshinone IIA and 1.72–1.91 times for cryptotanshinone, respectively. Compared to the CK, the levels of the three interference lines were the opposite, as was expected, and the contents of the four detected secondary metabolites were reduced by more than half. In particular, in the i-11 interference line, rosmarinic acid, salvianolic acid B, tanshinone II A, and cryptotanshinone were reduced by 71.93%, 67.95%, 82.23%, and 66.83%, respectively.

To determine whether metabolite accumulation was caused by changes in enzyme gene expression in the metabolic pathway, RT-qPCR was used to detect the expression of key enzyme genes involved in the salvianolic acids and tanshinones biosynthesis pathways in different lines. In the salvianolic acid biosynthesis pathway, the expression levels of *SmTAT*, *SmHPPR*, *SmPAL*, *SmC4H*, *Sm4CL*, *SmRAS*, and *SmCYP98A14* in the OE lines were significantly upregulated compared to the CK, but the activities of all the enzymes were repressed in the three interference lines (**Figures 7A,C**). In the tanshinones biosynthesis pathway, several enzyme genes, such as *SmDXS*, *SmHMGR*, *SmFPPS*, *SmGGPPS*, *SmCPS*, *SmKSL*, and *SmCYP76AH1* were markedly induced in the OE lines. In contrast, the expression levels of these genes were decreased in the interference lines compared to those in the CK (**Figures 7B,D**).

Morphological Differences Between Transgenic Lines of SmWD40-170 in *S. miltiorrhiza*

After obtaining the OE and interference lines of *S. miltiorrhiza*, the plantlet seedlings were cultured for 60 days, transplanted into soil media as part of a hydroponic cultivation system, and then cultivated for another 60 days. There were obvious morphological differences in the OE and interference SmWD40-170 transgenic lines. The results showed that the growth state of OE-3 line was better than the CK line and had

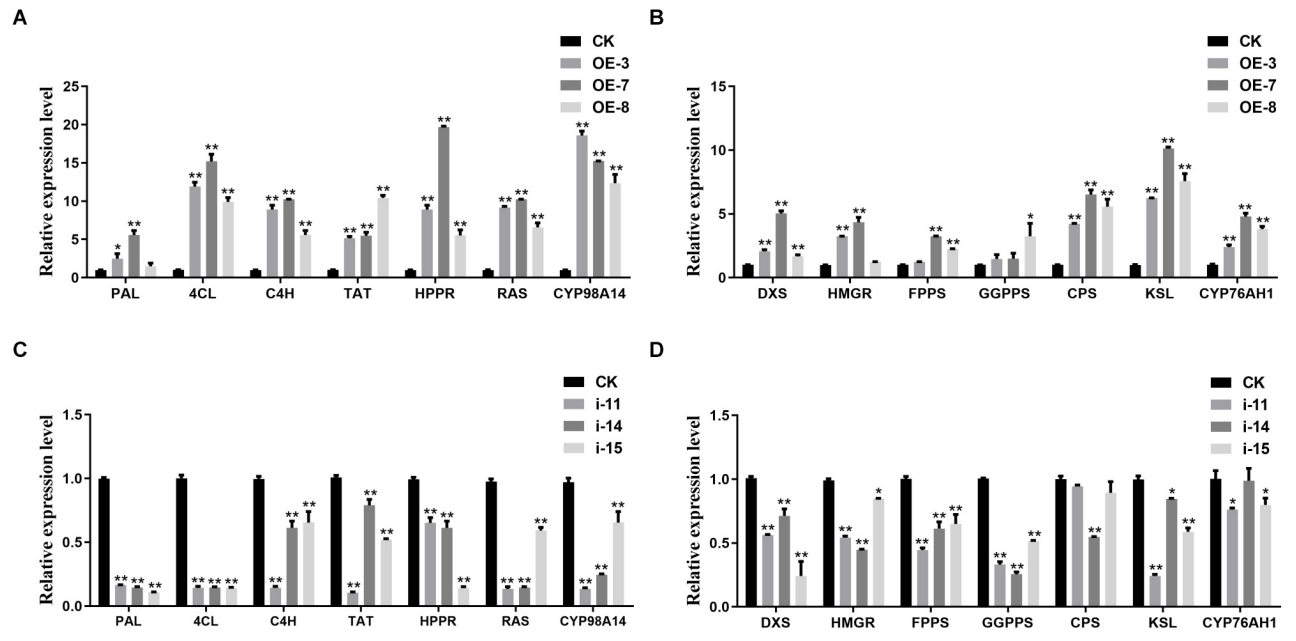


FIGURE 7 | Expression analysis of salvianolic acids and tanshinones biosynthesis genes. RT-qPCR analyses of the key enzyme genes of salvianolic acids biosynthetic pathway in OE **(A)** and RNAi **(C)** lines. RT-qPCR analyses of the key enzyme genes of tanshinones biosynthetic pathway in OE **(B)** and RNAi **(D)** lines. All data represent averages of three biological replicates, error bars indicate SD. Statistical significance was determined using the Student's *t*-test (*$p < 0.05$, **$p < 0.01$). *PAL*, phenylalanine ammonialyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, hydroxycinnamate-CoA ligase; *TAT*, tyrosine aminotransferase; *HPPR*, hydroxyl phenylpyruvate reductase; *RAS*, rosmarinic acid synthase; *DXS*, 1-deoxy-D-xylulose-5-phosphate synthase; *HMGR*, 3-hydroxy-3-methylglutaryl-CoA reductase; *FPPS*, farnesyl diphosphate synthase; *GGPPS*, geranylgeranyl diphosphate synthase; *CPS*, copalyl diphosphate synthase; *KSL*, kaurene synthase-like synthase.

higher biomass accumulation in both leaves and roots. In contrast, the RNAi line was shorter, and the growth state was poorer (**Figures 8A–F**). Statistical data was consistent with the phenotype. Compared with the CK, the OE lines had the advantages of root length. The feature was more obvious compared to those in the RNAi lines (**Figures 8G–I,M**). In addition, morphological changes were observed in the leaves. There was no significant difference in leaf size between OE with CK lines, however, the RNAi lines had smaller leaves with curled leaf edges (**Figures 8J–L,N**). These results indicate that the SmWD40-170 protein is necessary for the growth and development of *S. miltiorrhiza*.

DISCUSSION

The plant-specific TIFY family plays vital roles in the growth, development, and secondary metabolism of plants (Vanholme et al., 2007; Ye et al., 2009; Bai et al., 2011; Zhang et al., 2012, 2015; Zhu et al., 2013; Saha et al., 2016). Significant progress has been made toward identifying and characterizing the *TIFY* genes in various species. In the present study, we identified and analyzed the TIFY families in *S. miltiorrhiza*. This species contains 10 *SmJAZ*s, 3 *SmZML*s, 1 *SmTIFY*, and 1 *SmPPD* genes (**Table 1**). Similar to other species, the SmJAZ proteins have two typical domains, N-terminal TIFY and C-terminal Jas. The TIFY domain

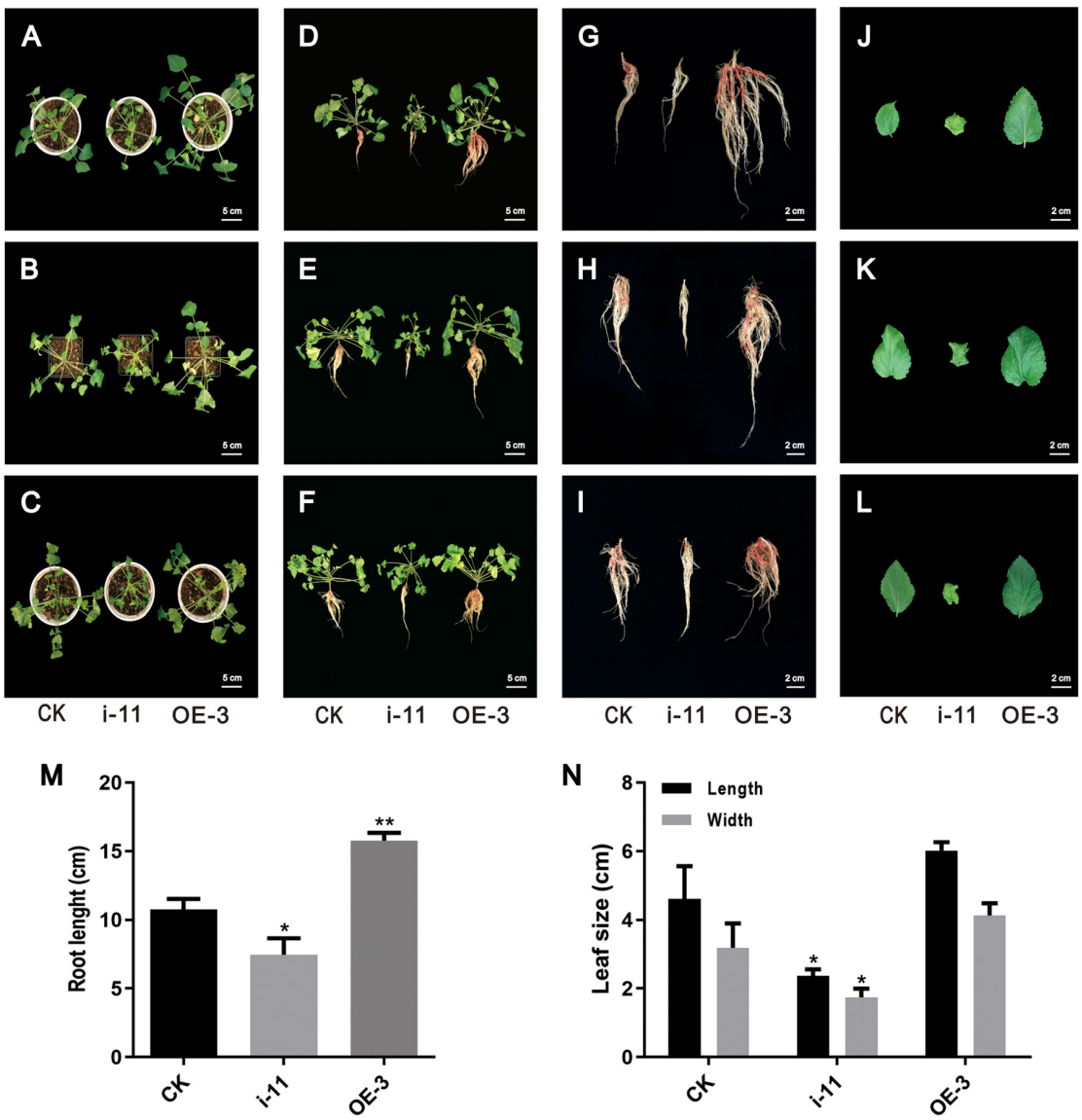


FIGURE 8 | Morphological differences in the control and SmWD40-170 transgenic lines. **(A/B/C,D/E/F,G/H/I,J/K/L)** represent three independent repeats. **(A–F)** Control and transgenic seedlings cultured in soil for 2 months. **(G–I)** Control and transgenic roots. **(J–L)** Control and transgenic leaves. **(M)** Comparison of the root length in control and transgenic lines. **(N)** Comparison of the leaf size in control and transgenic lines. All data represent averages of three biological replicates, error bars indicate SD. Statistical significance was determined using the Student's *t*-test (*$p < 0.05$, **$p < 0.01$).

participates in homomeric and heteromeric interactions, or in the interactions between JAZ proteins and MYC TFs (Bai et al., 2011). The Jas motif interacts with COI1, bHLH, or R2R3-MYB members (Chico et al., 2008; Browse, 2009). The SmZML proteins not only have these two domains, but also contain a CCT domain, similar to the Jas domain, and a GATA zinc finger. In contrast, SmPPD also has a PPD domain at its N-terminal (**Figure 2A**). SmJAZ1, SmJAZ2, SmJAZ5, and SmJAZ6 each carry an EAR-like motif (**Figure 2A**), which also exists in plant AUX/IAA proteins and functions as a binding motif for the regulator repressor TOPLESS. This suggests that these four proteins have different functions that are more critical than those of the other SmTIFY proteins (Szemenyei et al., 2008). Although all SmJAZ proteins could be clustered into the four groups previously described for other species (**Figure 1**), the high degree of variability among the sequences within this subfamily, suggests that these proteins have had a possible divergence in functions.

Members of the SmTIFY family are diverse in their exons and introns (**Figure 2B**), implying an important evolutionary role for their gene structure (Bai et al., 2011). Sequence alignments and phylogenetic analysis of the SmTIFY proteins indicated that almost all could be classified into the same four major groups with TIFY proteins from other species with the exception is for the rice members, from which OsPPD is missing. Screening the TIFY proteins from the different species has demonstrated that they only occur among Embryophyta (land plants) and that both group I (TIFY proteins with a C2C2-GATA domain) and group II (TIFY proteins without C2C2-GATA domain) of TIFY proteins are present in the liverworts (Marchantiophyta), which are considered primitive land plants (Vanholme et al., 2007). This suggests that the TIFY family was essential for the emergence of land plants during a series of evolutionary adaptations that increased the complexity of plant structure. This probably also enhanced the ability of those plants to respond to adverse environmental conditions. In contrast to many plant-specific gene families, obvious diversification has occurred between monocot and dicot species (Bai et al., 2011). Therefore, the plant *TIFY* genes may be derived from common ancestors that existed before the divergence of monocot and dicot species.

SmTTG1, a member of the WD40 protein family, has been shown to promote the accumulation of salvianolic acids in *S. miltiorrhiza* (Li et al., 2018). In the present study, we screened a WD40 protein, which participates in both salvianolic acids and tanshinones biosynthesis, except in drought stress responses (Liu et al., 2020). In the OE lines, the contents of the salvianolic acids and tanshinones were significantly increased, by approximately two fold, and the RNAi lines were expectedly decreased by at least 50%, which is consistent with the changes in the biosynthesis of the enzyme genes in the transgenic lines (**Figures 6**, **7**). Notably, all 14 key genes selected were induced by SmWD40-170. However, the WD40 protein is unable to directly regulate enzyme genes, and probably functions together with MYB and bHLH proteins to generate the MBW complex. In future research, we will focus on investigating the interaction partners of the SmWD40-170 protein. In addition, some obvious phenotypic changes were observed in the transgenic lines. Compared with the CK line, the RNAi lines presented underdeveloped roots, smaller leaves, and curled leaf edges, but the OE lines showed the opposite traits, with well-developed roots (**Figure 8**). In particular, the roots of OE lines were redder than those of the CK and RNAi lines. These results are consistent with the content determination of tanshinones in the transgenic and CK lines (**Figure 6**), and in accordance with the fact that more tanshinones were gathered in the redder roots of *S. miltiorrhiza* (Wang et al., 2014). It is great significance when aiming to improve the quality of *S. miltiorrhiza*. SmWD40-170 not only regulates secondary metabolism, but also affects growth and development, however, links between growth and secondary metabolite accumulation warrant further investigation.

Jasmonic acid has been shown to enhance the accumulation of secondary metabolites and promote growth and development in *S. miltiorrhiza* (Xiao et al., 2010; Gu et al., 2012; Ge et al., 2015; Shi et al., 2020), where SmCOI1 also plays a critical role (Chen et al., 2018). JAZ proteins are the targets of the SCF^{COI1} complex and function as key components of the JA signaling pathway (Shi et al., 2016; Pei et al., 2018), however, the mechanisms underlying JAZ-regulated repression events remain unclear for *S. miltiorrhiza*. Since JAZ proteins contain no DNA-binding domain, JAZs might affect gene expression and metabolite accumulation through their interactions with target genes (Cheng et al., 2011). It has been reported that a Jas motif of the JAZ proteins, participates in the protein–protein interactions with the MYB, bHLH, and other TFs (Withers et al., 2012). In the present study, we found that a WD40 protein interacts with the Jas motif of SmJAZ3 using Y2H screening, and determine the interaction

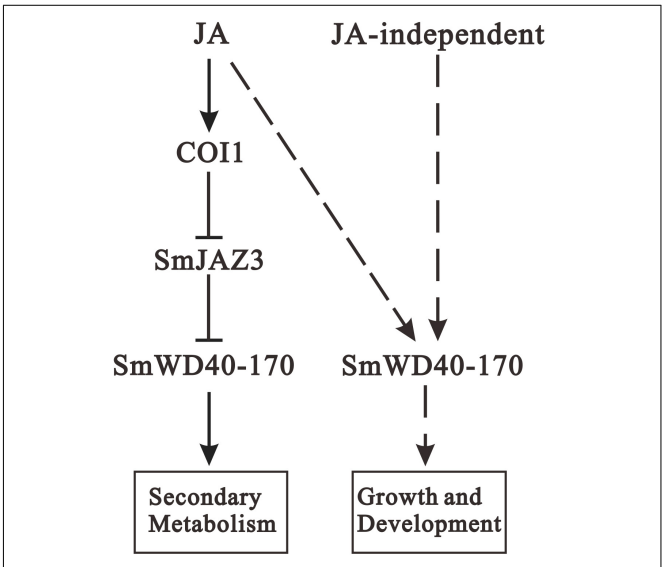


FIGURE 9 | Proposed module of the roles of JA in regulating secondary metabolites biosynthesis as well as growth and development. When treating with exogenous JA, the complex formation between the JA-Ile and COI1 promotes SmJAZ3 degradation via the 26S proteasome, and release the positive regulators such as WD40, then enhances the activities of enzymes to promote secondary metabolites biosynthesis. In addition, SmWD40-170 protein may regulate growth and development through JA-dependent or JA-independent pathway. Arrows, positive regulation; blunt ends, negative regulation; dotted line, uncertified process.

relationship between the SmJAZ3 and SmWD40-170. SmJAZ3 has been reported to act as a repressive transcriptional regulator in tanshinones biosynthesis regulation (Shi et al., 2016), while SmWD40-170 responds to drought stress by regulating ABA- and H_2O_2- induced stomal movement in *S. miltiorrhiza* (Liu et al., 2020). In the present study, we speculated that SmWD40-170 regulates the accumulation of secondary metabolites by interacting with SmJAZ3 in *S. miltiorrhiza*, which brings insights into the molecular mechanism of SmJAZ3 regulates tanshinone biosynthesis. We propose that SmJAZ3 protein interacts with SmWD40-170 in the JA signaling pathway and plays a vital role in the accumulation of secondary metabolites (**Figure 9**), which provides essential information for further exploring the mechanisms by which JA regulates the biosynthesis and accumulation of secondary metabolites in *S. miltiorrhiza*.

CONCLUSION

This study identified and analyzed 15 SmTIFY family members. Phylogenetic analysis suggested that the SmTIFY proteins could be clustered into four groups. RT-qPCR results showed that most of the *SmTIFY* genes responded to the MeJA treatments. Our analysis illustrated diversity in the *cis*-elements among the SmTIFY members, indicating that these genes have important roles in several hormone signaling pathways and stress responses, and may thus be applied to increase the production of valuable plant compounds. Furthermore, a novel interaction partner of SmJAZ3 was screened and physical interactions between SmJAZ3 and SmWD40-170 was also demonstrated in this study, which suggests a potential regulation mechanism for the SmJAZ involved in the JA signaling pathway. Subsequently, genetic assays showed that SmWD40-170 positively induced the accumulation of salvianolic acids and tanshinones, and promoted plant growth and development. Collectively, our research lays a foundation for future investigations into the mechanism of the hormone signal regulation network among the SmTIFY family members.

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

LL, XC, and ZW designed the experiments. LL, YH, and YL performed the experiments. YL, WM, and DW contributed analytical tools and provided technical support. LL and YH wrote the manuscript. BL, XC, and ZW promoted the manuscript. All authors revised and approved the manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (31670299, 31870276, 31900254, and 31800259), and the Project of the National Key Technologies R&D Program for Modernization of Traditional Chinese Medicine (2017YFC1701300 and 2019YFC1712602), and Fundamental Research Funds for the Central Universities (GK202003056).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.630424/full#supplementary-material

Supplementary Figure 1 | Screening of cDNA library using SmJAZ3-jas as bait. SmJAZ3 **(A)** and SmJAZ3-jas **(B)** bait vectors construction for screening yeast two-hybrid library. **(C)** Transcriptional activity analysis of BD-SmJAZ3 in yeast. **(D)** Bait bacteria with library bacteria forming yeast conjugates. **(E)** Part of the positive clones of yeast two-hybrid screening. **(F)** Agarose gel electrophoresis examine the PCR products for part of the yeast plaque.

Supplementary Figure 2 | MRM maps of rosmarinic acid, salvianolic acid B, tanshinone IIA, and cryptotanshinone standard and sample.

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

Supplementary Table 2 | Primers for PCR and vectors construction.

Supplementary Table 3 | Part of *cis*-elements in *SmTIFY* family genes. A series of *cis*-elements involved in light response, the biotic and abiotic stress response, phytohormone (abscisic acid, auxin, salicylic acid, gibberellin, ethylene, MeJA) response, circadian control, MYB binding site, G-box, endosperm and meristem expression, and fungal elicitor response were identified.

Supplementary Table 4 | Yeast library screening with SmJAZ3-Jas-BD as bait protein, NCBI BlastX results.

REFERENCES

An, X. H., Tian, Y., Chen, K. Q., Wang, X. F., and Hao, Y. J. (2012). The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. *J. Plant Physiol.* 169, 710–717. doi: 10.1016/j.jplph.2012.01.015

Aparicio-Fabre, R., Guillén, G., Loredo, M., Arellano, J., Valdés-López, O., Ramírez, M., et al. (2013). Common bean (*Phaseolus vulgaris* L.) PvTIFY orchestrates global changes in transcript profile response to jasmonate and phosphorus deficiency. *BMC Plant Biol.* 13:26. doi: 10.1186/1471-2229-13-26

Bai, Y. H., Meng, Y. J., Huang, D. L., Qi, Y. H., and Chen, M. (2011). Origin and evolutionary analysis of the plant-specific TIFY transcription factor family. *Genomics* 98, 128–136. doi: 10.1016/j.ygeno.2011.05.002

Boter, M., Ruiz-Rivero, O., Abdeen, A., and Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis*. *Genes Dev.* 18, 1577–1591. doi: 10.1101/gad.297704

Browse, J. (2009). Jasmonate passes muster: a receptor and targets for the defense hormone. *Annu. Rev. Plant Biol.* 60, 183–205.

Carey, C. C., Strahle, J. T., Selinger, D. A., and Chandler, V. L. (2004). Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT

TESTA GLABRA1 gene in *Arabidopsis thaliana*. *Plant Cell* 16, 450–464. doi: 10.1105/tpc.018796

Chen, C., Cao, X. Y., Hua, W. P., Huang, Y., Lv, T. T., Zhang, Y., et al. (2018). Roles of SmCOI1 in pest resistance and secondary metabolism regulation based on *Salvia miltiorrhiza* Bunge genome. *Sci. Sin.* 48, 399–411. doi: 10.1360/N052017-00189

Cheng, Z. W., Sun, L., Qi, T. C., Zhang, B. S., Peng, W., Liu, Y. L., et al. (2011). The bHLH transcription factor MYC3 interacts with the jasmonate ZIM-domain proteins to mediate jasmonate response in *Arabidopsis*. *Mol. Plant* 4, 279–288. doi: 10.1093/mp/ssq073

Chew, Y. L., Goh, J. K., and Lim, Y. Y. (2009). Assessment of in vitro antioxidant capacity and polyphenolic composition of selected medicinal herbs from Leguminosae family in Peninsular Malaysia. *Food Chem.* 116, 13–18. doi: 10.1016/j.foodchem.2009.01.091

Chico, J. M., Chini, A., Fonseca, S., and Solano, R. (2008). JAZ repressors set the rhythm in jasmonate signaling. *Curr. Opin. Plant Biol.* 11, 486–494. doi: 10.1016/j.pbi.2008.06.003

Chini, A., Boter, M., and Solano, R. (2010). Plant oxylipins: COI1/JAZs/MYC2 as the core jasmonic acid-signalling module. *FEBS J.* 276, 4682–4692. doi: 10.1111/j.1742-4658.2009.07194.x

Chini, A., Fonseca, S., Fernandez, G., Adie, B., Chico, J. M., Lorenzo, O., et al. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 448, 666–671. doi: 10.1038/nature06006

Chung, H. S., Niu, Y. J., Browse, J., and Howe, G. A. (2009). Top hits in contemporary JAZ: an update on jasmonate signaling. *Phytochemistry* 70, 1547–1559. doi: 10.1016/j.phytochem.2009.08.022

Creelman, R. A., and Mullet, J. E. (1997). Biosynthesis and action of jasmonates in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 48, 355–381. doi: 10.1146/annurev.arplant.48.1.355

De Geyter, N., Gholami, A., Goormachtig, S., and Goossens, A. (2012). Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends Plant Sci.* 17, 349–359. doi: 10.1016/j.tplants.2012.03.001

Demianski, A. J., Chung, K. M., and Kunkel, B. N. (2011). Analysis of *Arabidopsis* JAZ gene expression during *Pseudomonas* syringae pathogenesis. *Mol. Plant Pathol.* 13, 46–57. doi: 10.1111/j.1364-3703.2011.00727.x

Dewanto, V., Wu, X. Z., Adom, K. K., and Liu, R. H. (2002). Thermal processing enhances the nutritional value of tomatoes by increasing total antioxidant activity. *J. Agric. Food Chem.* 50, 3010–3014. doi: 10.1021/jf011 5589

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., et al. (2010). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45, 616–629. doi: 10.1111/j.1365-313x.2005.02617.x

Farmer, E. E. (2007). Jasmonate perception machines. *Nature* 448, 659–660. doi: 10.1038/448659a

Farmer, E. E., Alméras, E., and Krishnamurthy, V. (2003). Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr. Opin. Plant Biol.* 6, 372–378. doi: 10.1016/S1369-5266(03)00045-1

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J. M., Gimenez-Ibanez, S., Geerinck, J., et al. (2011). The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* 23, 701–715. doi: 10.1105/tpc.110.080788

Fonseca, S., Fernández-Calvo, P., Fernández, G. M., Díez-Díaz, M., Gimenez-Ibanez, S., López-Vidriero, I., et al. (2014). bHLH003, bHLH013 and bHLH017 are new targets of JAZ repressors negatively regulating JA responses. *PLoS One* 9:e86182. doi: 10.1371/journal.pone.0086182

Garrido-Bigotes, A., Torrejón, M., Solano, R., and Figueroa, C. R. (2020). Interactions of JAZ repressors with anthocyanin biosynthesis-related transcription factors of *Fragaria* × *ananassa*. *Agron. Basel* 10:1586. doi: 10.3390/agronomy10101586

Garrido-Bigotes, A., Valenzuela-Riffo, F., and Figueroa, C. R. (2019). Evolutionary analysis of JAZ proteins in plants: an approach in search of the ancestral sequence. *Intern. J. Mol. Sci.* 20:5060. doi: 10.3390/ijms20205060

Ge, Q., Zhang, Y. C., Hua, W. P., Wu, Y. C., Jin, X. X., Song, S. H., et al. (2015). Combination of transcriptomic and metabolomic analyses reveals a JAZ repressor in the jasmonate signaling pathway of *Salvia miltiorrhiza*. *Sci. Rep.* 5, 2809–2823. doi: 10.1038/srep14048

Gietz, R. D., and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34. doi: 10.1038/nprot.2007.15

Gu, X. C., Chen, J. F., Xiao, Y., Di, P., Xuan, H. J., Zhou, X., et al. (2012). Overexpression of allene oxide cyclase promoted tanshinone/phenolic acid production in *Salvia miltiorrhiza*. *Plant Cell Rep.* 31, 2247–2259. doi: 10.1007/s00299-012-1334-9

Kai, G. Y., Xu, H., Zhou, C. C., Liao, P., Xiao, J. B., Luo, X. Q., et al. (2011). Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metab. Eng.* 13, 319–327. doi: 10.1016/j.ymben.2011.02.003

Li, S. S., Wu, Y. C., Kuang, J., Wang, H. Q., Du, T. Z., Huang, Y. Y., et al. (2018). SmMYB111 is a key factor to phenolic acid biosynthesis and interacts with both SmTTG1 and SmbHLH51 in *Salvia miltiorrhiza*. *J. Agric. Food Chem.* 66, 8069–8078. doi: 10.1021/acs.jafc.8b02548

Liechti, R., and Farmer, E. E. (2002). The jasmonate pathway. *Science* 296, 1649–1650. doi: 10.1126/science.1071547

Liu, A. H., Li, L., Xu, M., Lin, Y. H., Guo, H. Z., and Guo, D. A. (2006). Simultaneous quantification of six major phenolic acids in the roots of *Salvia miltiorrhiza* and four related traditional Chinese medicinal preparations by HPLC-DAD method. *J. Pharm. Biomed. Analys.* 41, 48–56. doi: 10.1016/j.jpba.2005.10.021

Liu, B. Y., Du, Y. H., Cong, L. X., Jia, X., and Yang, G. (2016). Danshen (*Salvia miltiorrhiza*) compounds improve the biochemical indices of the patients with coronary heart disease. *Evid. Based Complement. Alternat. Med.* 2016:9781715. doi: 10.1155/2016/9781715

Liu, Y. C., Ma, W., Niu, J. F., Li, B., Zhou, W., Liu, S., et al. (2020). Systematic analysis of SmWD40s, and responding of SmWD40-170 to drought stress by regulation of ABA- and H_2O_2- induced stomal movement in *Salvia miltiorrhiza* bunge. *Plant Physiol. Biochem.* 153, 131–140. doi: 10.1016/j.plaphy.2020.05.017

Luo, H. M., Zhu, Y. J., Song, J. Y., Xu, L. J., Sun, C., Zhang, X., et al. (2015). Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiol. Plant.* 152, 241–255. doi: 10.1111/ppl.12193

Ma, X. H., Ma, Y., Tang, J. F., He, Y. L., Liu, Y. C., Ma, X. J., et al. (2015). The biosynthetic pathways of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Molecules* 20, 16235–16254. doi: 10.3390/molecules200916235

Ma, Y. M., Yuan, L. C., Wu, B., Li, X. E., Chen, S. L., and Lu, S. F. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exper. Bot.* 63, 2809–2823. doi: 10.1093/jxb/err466

Matus, J. T., Poupin, M. J., Caón, P., Bordeu, E., Alcalde, J. A., and Arce-Johnson, P. (2010). Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (*Vitis vinifera* L.). *Plant Mol. Biol.* 72, 607–620. doi: 10.1007/s11103-010-9597-4

Neer, E. J., Schmidt, C. J., Nambudripad, R., and Smith, T. F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371, 297–300. doi: 10.1038/371812b0

Pan, M., Zhou, Y. Q., Wang, Y. F., Li, L. J., Song, Y. L., Hou, L., et al. (2017). Screening and identification of the host proteins interacting with Toxoplasma gondii rhoptry protein ROP16. *Front. Microbiol.* 8:2408. doi: 10.3389/fmicb.2017.02408

Pang, Y. Z., Wenger, J. P., Saathoff, K., Peel, G. J., Wen, J. Q., Huhman, D., et al. (2009). A WD40 repeat protein from *Medicago truncatula* is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiol.* 151, 1114–1129. doi: 10.1104/pp.109.144022

Pei, T. L., Ma, P. D., Ding, K., Liu, S. J., Jia, Y. Y., Ru, M., et al. (2018). SmJAZ8 acts as a core repressor regulating JA-induced biosynthesis of salvianolic acids and tanshinones in *Salvia miltiorrhiza* hairy roots. *J. Exper. Bot.* 69, 1663–1678. doi: 10.1093/jxb/erx484

Qi, T. C., Huang, H., Wu, D. W., Yan, J. B., Qi, Y. J., Song, S. S., et al. (2014). *Arabidopsis* DELLA and JAZ proteins bind the WD-repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. *Plant Cell* 26, 1118–1133. doi: 10.1105/tpc.113.121731

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

Saha, G., Park, J., Kayum, M., and Nou, S. (2016). A genome-wide analysis reveals stress and hormone responsive patterns of TIFY family genes in *Brassica rapa*. *Front. Plant Sci.* 7:936. doi: 10.3389/fpls.2016.00936

Sen, S., Kundu, S., and Dutta, S. (2016). Proteomic analysis of JAZ interacting proteins under methyl jasmonate treatment in finger millet. *Plant Physiol. Biochem.* 108, 79–89. doi: 10.1016/j.plaphy.2016.05.033

Shi, M., Liao, P., Nile, S. H., Georgiev, M. I., and Kai, G. Y. (2020). Biotechnological exploration of transformed root culture for value-added products. *Trends Biotechnol.* 39, 137–149.

Shi, M., Zhou, W., Zhang, J. L., Huang, S. X., Wang, H. Z., and Kai, G. Y. (2016). Methyl jasmonate induction of tanshinone biosynthesis in *Salvia miltiorrhiza* hairy roots is mediated by JASMONATE ZIM-DOMAIN repressor proteins. *Sci. Rep.* 6:20919. doi: 10.1038/srep20919

Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* 24, 181–185. doi: 10.1016/S0968-0004(99)01384-5

Song, S. S., Qi, T. C., Fan, M., Zhang, X., Gao, H., Huang, H., et al. (2013). The bHLH subgroup IIId factors negatively regulate jasmonate-mediated plant defense and development. *PLoS Genet.* 9:e1003653. doi: 10.1371/journal.pgen.1003653

Song, S. S., Qi, T. C., Huang, H., Ren, Q. C., Wu, D. W., Chang, C. Q., et al. (2011). The jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect jasmonate-regulated stamen development in *Arabidopsis*. *Plant Cell* 23, 1000–1013. doi: 10.1105/tpc.111.083089

Staswick, P. E. (2008). JAZing up jasmonate signaling. *Trends Plant Sci.* 13, 66–71. doi: 10.1016/j.tplants.2007.11.011

Su, C. Y., Ming, Q. L., Rahman, K., Han, T., and Qin, L. P. (2015). *Salvia miltiorrhiza*: traditional medicinal uses, chemistry, and pharmacology. *Chin. J. Nat. Med.* 13, 163–182. doi: 10.1016/S1875-5364(15)30002-9

Szemenyei, H., Hannon, M., and Long, J. A. (2008). TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science* 319, 1384–1386. doi: 10.1126/science.1151461

Thines, B., Katsir, L., Melotto, M., Niu, Y. J., Mandaokar, A., Liu, G. H., et al. (2007). JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* 448, 661–665. doi: 10.1038/nature05960

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, 467–470. doi: 10.1186/gb-2002-3-7-research0034

Vanholme, B., Grunewald, W., Bateman, A., Kohchi, T., and Gheysen, G. (2007). The tify family previously known as ZIM. *Trends Plant Sci.* 12, 239–244. doi: 10.1016/j.tplants.2007.04.004

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., et al. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1349. doi: 10.2307/3870753

Wang, B., Niu, J. F., Li, B., Huang, Y. Y., Han, L. M., Liu, Y. C., et al. (2018). Molecular characterization and overexpression of SmJMT increases the production of phenolic acids in *Salvia miltiorrhiza*. *Intern. J. Mol. Sci.* 19:3788. doi: 10.3390/ijms19123788

Wang, H., Yan, Z. Y., Shen, Y. X., He, D. M., Lan, Y., and Wan, D. G. (2014). Correlation of color feature with effective constituents of radix *Salvia Miltiorrhiza*. *Tradit. Chin. Drug Res. Clin. Pharmacol.* 3, 333–338.

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

Withers, J., Yao, J., Mecey, C., Howe, G. A., Melotto, M., and He, S. Y. (2012). Transcription factor-dependent nuclear localization of a transcriptional repressor in jasmonate hormone signaling. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20148–20153. doi: 10.1073/pnas.1210054109

Xiao, Y., Gao, S. H., Di, P., Chen, J. F., Chen, W. S., and Zhang, L. (2010). Methyl jasmonate dramatically enhances the accumulation of phenolic acids in *Salvia miltiorrhiza* hairy root cultures. *Physiol. Plant.* 137, 1–9. doi: 10.1111/j.1399-3054.2009.01257.x

Xie, D., Feys, B. F., James, S., Nieto-Rostro, M., and Turner, J. G. (1998). COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* 280, 1091–1094. doi: 10.1126/science.280.5366.1091

Xu, H. B., Song, J. Y., Luo, H. M., Zhang, Y. J., Li, Q. S., Zhu, Y. J., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9, 949–952. doi: 10.1016/j.molp.2016.03.010

Xu, L., Liu, F., Lechner, E., Genschik, P., Crosby, W. L., Ma, H., et al. (2002). The SCF(COI1) ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. *Plant Cell* 14, 1919–1935. doi: 10.1105/tpc.003368

Xu, Z. C., Peters, R. J., Weirather, J., Luo, H. M., Liao, B. S., Zhang, X., et al. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* 82, 951–961. doi: 10.1111/tpj.12865

Yan, Y., Stolz, S., Chetelat, A., Reymond, P., Pagni, M., Dubugnon, L., et al. (2007). A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* 19, 2470–2483. doi: 10.1105/tpc.107.050708

Yan, Y. P., and Wang, Z. Z. (2007). Genetic transformation of the medicinal plant *Salvia miltiorrhiza* by *Agrobacterium tumefaciens*-mediated method. *Plant Cell Tissue Organ. Culture* 88, 175–184. doi: 10.1007/s11240-006-9187-y

Yang, N., Zhou, W. P., Su, J., Wang, X. F., Li, L., Wang, L. R., et al. (2017). Overexpression of SmMYC2 increases the production of phenolic acids in *Salvia miltiorrhiza*. *Front. Plant Sci.* 8:1804. doi: 10.3389/fpls.2017.01804

Ye, H. Y., Du, H., Tang, N., Li, X. H., and Xiong, L. Z. (2009). Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. *Plant Mol. Biol.* 71, 291–305. doi: 10.1007/s11103-009-9524-8

Yu, J., Zhang, Y. X., Di, C., Zhang, Q. L., Zhang, K., Wang, C. C., et al. (2016). JAZ7 negatively regulates dark-induced leaf senescence in *Arabidopsis*. *J. Exper. Bot.* 67, 751–762. doi: 10.1093/jxb/erv487

Zhang, L. H., You, J., and Chan, Z. L. (2015). Identification and characterization of TIFY family genes in *Brachypodium distachyon*. *J. Plant Res.* 128, 995–1005. doi: 10.1007/s10265-015-0755-2

Zhang, Y. C., Gao, M., Singer, S. D., Fei, Z. J., Wang, H., Wang, X. P., et al. (2012). Genome-wide identification and analysis of the TIFY gene family in grape. *PLoS One* 7:e44465. doi: 10.1371/journal.pone.0044465

Zhao, J. L., Zhou, L. G., and Wu, J. Y. (2010). Effects of biotic and abiotic elicitors on cell growth and tanshinone accumulation in *Salvia miltiorrhiza* cell cultures. *Appl. Microbiol. Biotechnol.* 87, 137–144. doi: 10.1007/s00253-010-2443-4

Zhou, Y. Y., Sun, W., Chen, J. F., Tan, H. X., Xiao, Y., Li, Q., et al. (2016). SmMYC2a and SmMYC2b played similar but irreplaceable roles in regulating the biosynthesis of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Sci. Rep.* 6, 1345–1359. doi: 10.1038/srep22852

Zhu, D., Bai, X., Luo, X., Chen, Q., Cai, H., Ji, W., et al. (2013). Identification of wild soybean (*Glycine soja*) TIFY family genes and their expression profiling analysis under bicarbonate stress. *Plant Cell Rep.* 32, 263–272. doi: 10.1007/s00299-012-1360-7

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fpls.2021.699326

Genome-Wide Profiling of *WRKY* Genes Involved in Benzylisoquinoline Alkaloid Biosynthesis in California Poppy (*Eschscholzia californica*)

Yasuyuki Yamada[1], Shohei Nishida[2], Nobukazu Shitan[1] and Fumihiko Sato[2,3]**

[1] Laboratory of Medicinal Cell Biology, Kobe Pharmaceutical University, Kobe, Japan, [2] Department of Plant Gene and Totipotency, Division of Integrated Life Science, Graduate School of Biostudies, Kyoto University, Kyoto, Japan, [3] Graduate School of Science, Osaka Prefecture University, Sakai, Japan

Transcription factors of the WRKY family play pivotal roles in plant defense responses, including the biosynthesis of specialized metabolites. Based on the previous findings of WRKY proteins regulating benzylisoquinoline alkaloid (BIA) biosynthesis, such as CjWRKY1—a regulator of berberine biosynthesis in *Coptis japonica*—and PsWRKY1—a regulator of morphine biosynthesis in *Papaver somniferum*—we performed genome-wide characterization of the WRKY gene family in *Eschscholzia californica* (California poppy), which produces various BIAs. Fifty *WRKY* genes were identified by homology search and classified into three groups based on phylogenetic, gene structure, and conserved motif analyses. RNA sequencing showed that several *EcWRKY* genes transiently responded to methyl jasmonate, a known alkaloid inducer, and the expression patterns of these *EcWRKY* genes were rather similar to those of BIA biosynthetic enzyme genes. Furthermore, tissue expression profiling suggested the involvement of a few subgroup IIc EcWRKYs in the regulation of BIA biosynthesis. Transactivation analysis using luciferase reporter genes harboring the promoters of biosynthetic enzyme genes indicated little activity of subgroup IIc EcWRKYs, suggesting that the transcriptional network of BIA biosynthesis constitutes multiple members. Finally, we investigated the coexpression patterns of *EcWRKY*s with some transporter genes and discussed the diversified functions of *WRKY* genes based on a previous finding that CjWRKY1 overexpression in California poppy cells enhanced BIA secretion into the medium.

Keywords: benzylisoquinoline alkaloid, *Eschscholzia californica*, California poppy, *WRKY*, methyl jasmonate, RNA sequencing, tissue expression

OPEN ACCESS

Edited by:
Yang Zhang,
Sichuan University, China

Reviewed by:
Thu Thuy Dang,
University of British Columbia Okanagan, Canada
Kazunori Okada,
The University of Tokyo, Japan

****Correspondence:***
Yasuyuki Yamada
yyamada@kobepharma-u.ac.jp
Fumihiko Sato
fsato@lif.kyoto-u.ac.jp

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

Received: *23 April 2021*
Accepted: *25 May 2021*
Published: *17 June 2021*

Citation:
Yamada Y, Nishida S, Shitan N and Sato F (2021) Genome-Wide Profiling of WRKY Genes Involved in Benzylisoquinoline Alkaloid Biosynthesis in California Poppy (Eschscholzia californica). Front. Plant Sci. 12:699326. doi: 10.3389/fpls.2021.699326

INTRODUCTION

Being sessile, plants have evolved a wide array of defense mechanisms to protect themselves from diverse environmental stresses. WRKY transcription factors (TFs), one of the most important transcriptional regulators, play pivotal roles in plant development, senescence, and defense responses (Eulgem and Somssich, 2007; Rushton et al., 2010). The WRKY family proteins harbor at least one highly conserved WRKY domain composed of 60 amino acid residues, which includes the conserved N-terminal WRKYGQK sequence followed by a C-terminal zinc finger motif

(Eulgem et al., 2000). The WRKY family can be divided into three groups (I–III). Group I proteins generally harbor two WRKY domains and a C2H2-type zinc finger motif. Group II proteins harbor a single WRKY domain and a C2H2-type zinc finger, and these can be further classified into five subgroups (IIa–IIe). Group III proteins also harbor a single WRKY domain and a C2HC-type zinc finger-like motif. The WRKY proteins modulate the expression of target genes by binding to the W-box DNA motif (C/TTGACC/T) in their promoter regions (Ulker and Somssich, 2004). The structures of several WRKY proteins indicate that the conserved WRKYGQK motif with a β-sheet structure binds to the major groove of the DNA strand of W-box sequence (Yamasaki et al., 2013). The RKYGQK residues are directly involved in DNA binding through extensive hydrophobic contacts with the methyl groups of thymine (Yamasaki et al., 2012).

In several species, WRKY family proteins regulate plant-specific (secondary) metabolism related to defense response against biotic and abiotic stresses (Yamada and Sato, 2013). For instance, GaWRKY1 regulates sesquiterpene biosynthesis in *Gossypium arboreum* (Xu et al., 2004). AaWRKT1 and GLANDULAR TRICHOME-SPECIFIC WRKY1 (AaGSW1) positively regulate antimalarial artemisinin biosynthesis in *Artemisia annua* (Ma et al., 2009; Chen et al., 2017). CrWRKY1 acts as an activator of monoterpenoid indole alkaloid biosynthesis via binding to the *tyrosine decarboxylase* (*TDC*) gene promoter in *Catharanthus roseus* (Suttipanta et al., 2011). The expression of these *WRKY* genes could be clearly induced by methyl jasmonate (MeJA)—a crucial phytohormone involved in plant defense and plant-specific metabolism—indicating that the WRKY TFs involved in the regulation of specialized metabolism in plants play important roles in the jasmonic acid (JA) signaling cascade.

Furthermore, the biosynthesis of benzylisoquinoline alkaloids (BIAs), which are pharmaceutically important and structurally divergent specialized chemicals (e.g., analgesics morphine and codeine are found in *Papaver somniferum*, and antimicrobial berberine in *Coptis japonica*), is also regulated by the WRKY TFs CjWRKY1 and PsWRKY (Kato et al., 2007; Mishra et al., 2013). In *C. japonica*, belonging to the Ranunculaceae family, CjWRKY1 specifically regulates the expression of berberine biosynthetic enzyme genes by binding to several W-boxes in their promoters (Kato et al., 2007; Yamada et al., 2016). In *P. somniferum*, belonging to the Papaveraceae family, PsWRKY plays an important role in wound-induced regulation of morphine biosynthesis (Mishra et al., 2013). Although the functions of both *WRKY* genes were induced by MeJA, they were classified into different groups of the WRKY family: CjWRKY1 in subgroup IIc and PsWRKY in group I.

Eschscholzia californica (California poppy), belonging to the Papaveraceae family, produces various BIAs, such as sanguinarine, chelerythrine, and escholtzine, which are different types of BIAs from berberine and morphine, while a common biosynthetic pathway from L-tyrosine to (*S*)-reticuline is shared. The main BIA found in *E. californica* is sanguinarine, which also produced by *P. somniferum* cultured cells. The biosynthetic pathways of sanguinarine and related BIAs have been intensively investigated at the molecular level (**Supplementary Figure 1**). Furthermore, the basic helix–loop–helix TFs EcbHLH1-1 and EcbHLH1-2 have been identified as the positive regulators of sanguinarine biosynthesis (Yamada et al., 2015). Recently, the draft genome sequence of California poppy was compiled and various gene families related to BIA biosynthesis in the genome of this plant were explored (Hori et al., 2018; Yamada et al., 2021). In fact, novel cytochrome P450 enzymes involved in macarpine biosynthesis and possible AP2/ERF TFs involved in the regulation of sanguinarine biosynthesis have been identified (Hori et al., 2018; Yamada et al., 2020).

Heterologous CjWRKY1 expression in *E. californica* cells strongly enhanced BIA biosynthesis, suggesting the involvement of WRKY protein(s) in the regulation of the BIA biosynthetic pathway in California poppy (Yamada et al., 2017). Although CjWRKY1 has been identified as a comprehensive regulator of almost all genes encoding berberine biosynthetic enzymes in *C. japonica*, ectopic CjWRKY1 expression in California poppy cells upregulated only a few BIA biosynthetic enzyme genes. Interestingly, CjWRKY1 overexpression in cultured California poppy cells enhanced BIA accumulation in the culture medium. Together, these findings suggest that the potential WRKY TF(s) involved in the regulation of the BIA biosynthetic pathway might be functionally diversified in BIA-producing plant species and gained additional functions associated with BIA production and accumulation in *E. californica*.

In this study, we investigated the WRKY family genes in the California poppy genome using gene annotation data and compared their expression profiles with those of *EcbHLH1* and some *EcAP2/ERF* genes involved in BIA biosynthesis. We classified the identified genes by phylogenetic analysis and performed gene structure and conserved motif analyses. The expression profiles of the *EcWRKY* genes in response to MeJA treatment were examined by RNA sequencing (RNA-Seq) and quantitative RT-PCR (qRT-PCR). Moreover, tissue-specific expression patterns of MeJA-responsive *EcWRKY* genes were investigated by qRT-PCR. We further searched for genes possibly involved in the efflux of BIAs and identified several transporter genes induced by MeJA, based on our previous finding of enhanced BIA secretion following CjWRKY1 overexpression in California poppy cells. The present characterization provides useful information on the physiological roles of *EcWRKY* genes and the transcriptional network of BIA biosynthesis in *E. caifornica*.

MATERIALS AND METHODS

Identification of *WRKY* Genes From *E. californica*

First, 76 putative *WRKY* genes were isolated from the *E. californica* draft genome based on annotated gene information in the Eschscholzia Genome Database.[1] Next, 20 genes that did not contain complete WRKY domain-encoding sequences were removed based on domain search using the SMART database,[2] and six genes were removed because they harbored

[1] http://eschscholzia.kazusa.or.jp

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

partial open reading frames or abnormal sequences, probably due to assembly errors. After sequence validation using the PhytoMetaSyn transcriptomic database[3] (Xiao et al., 2013) and the NCBI database[4] using BLAST (**Supplementary Table 1**), 50 *WRKY* genes were identified in the California poppy genome.

Phylogenetic Analysis of *E. californica WRKY* TFs

The WRKY domain sequences of WRKY TFs from *Arabidopsis thaliana* and *E. californica* were obtained using the SMART database. Multiple sequence alignment was performed with ClustalW using BioEdit.[5] An unrooted phylogenetic tree was created using MEGA 7.0.[6] The neighbor-joining (NJ) method with the Jones–Thornton–Taylor (JTT) model and 1,000 bootstrap replications was used (Kumar et al., 2016).

Genome Structure and Conserved Motif Analysis

The intron–exon organization of *E. californica WRKY* genes was visualized using the Gene Structure Display Server (GSDS)[7] based on the predicted coding sequences and their corresponding genomic sequences. Conserved motifs of the EcWRKY proteins were predicted using MEME Suite (version 5.1.0)[8] with the following parameters: maximum motif number of 15 and optimum motif width from ≥6 to ≤50 (Bailey et al., 2009). The topology of the phylogenetic tree was generated based on full-length WRKY protein sequences using MEGA 7.0.

Plant Material

California poppy seedlings ("Hitoezaki"; Takii Seed Co., Ltd.) were grown and treated with 0.1% dimethyl sulfoxide (DMSO) as a control or 100 μM MeJA, as previously described (Yamada et al., 2020). The California poppy plants for tissue expression and metabolite analyses were grown in flowerpots for 5–6 months.

RNA Sequencing and Expression Profiling Analyses

Total RNA extraction and sequencing were performed as described previously (Yamada et al., 2020) (Hokkaido Biosystem Science Co., Ltd., Hokkaido, Japan). The fragments per kilobase of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks to evaluate gene expression levels. Hierarchical clustering was performed and heat maps were constructed based on log2-transformed fold change (FC) values compared to the mock control (0 h) using R.[9]

qRT-PCR

Total RNA was extracted from six California poppy seedlings treated with 100 μM MeJA for 0, 0.5, 1, 2, 6, and 24 h, and tissues (leaf blade, petiole, root, flower bud, and flower) were obtained from nine plants using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Single-stranded cDNA was synthesized from to 500–1,000 ng of total RNA with the ReverTra Ace qPCR RT Master Mix using the gDNA Remover Kit (TOYOBO, Osaka, Japan). Real-time PCR was performed with specific primer pairs (**Supplementary Table 2**) using the THUNDERBIRD Next SYBR qPCR Mix (TOYOBO, Osaka, Japan) on the LightCycler 96 system (Roche, Basel, Switzerland). The PCR conditions were 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. Gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method to analyze MeJA response or generate a standard curve for tissue expression analysis. The relative expression levels were standardized to those of *actin* as the internal control.

LUC Reporter Assay

The promoter:*LUC* constructs of the *Ec6OMT* and *EcCYP719A5* promoters have been constructed previously (Yamada et al., 2020). The full-length cDNAs of subgroup IIc *EcWRKY* genes were fused to the CaMV 35S promoter in the pBI221 vector, which was used as the effector construct. A dual-LUC reporter assay was then performed using *C. japonica* protoplasts, as previously described (Yamada and Sato, 2016).

Metabolite Analysis

California poppy tissues were ground in liquid nitrogen and extracted overnight with 4 μL mg^{-1} fresh weight methanol containing 0.01 N HCl at room temperature (20°C). After filtration, the filtrate was prepared for metabolite analysis. Ultra-performance liquid chromatography (UPLC) equipped with QDa mass spectrometry was performed using the ACQUITY UPLC BEH C18 column (2.1 mm × 100 mm, 1.7 μm; Waters Corp.) operated at 40°C. Mobile phase A comprised an aqueous solution of 0.01% acetic acid, whereas mobile phase B comprised acetonitrile containing 0.01% acetic acid. Gradient elution was performed as follows: 0–1 min, 5% B; 1–13 min, 5-30% B; 13–17 min, 30–80% B; 17–18 min, 80–5% B; and 18–20 min, 5% B. The flow rate and injection volume were set at 0.3 mL min^{-1} and 2 μL, respectively. The QDa conditions were set as follows: cone voltage, 15 V; capillary voltage, 0.8 kV; and source temperature, 600°C. The predicted pavine-type BIAs were detected using total ion chromatography and mass spectrometry in the single-ion recording mode, and the fragmentation spectra (50 V cone voltage) were compared with previous data (Fabre et al., 2000).

RESULTS

Identification and Classification of WRKY Family Members in the California Poppy Genome

To identify the WRKY TF-encoding genes in California poppy, we searched the *E. californica* draft genome database with gene annotation information using the sequence of a typical WRKY domain. After the removal of incomplete and redundant

[3]https://bioinformatics.tugraz.at/phytometasyn/
[4]https://www.ncbi.nlm.nih.gov/
[5]https://bioedit.software.informer.com/
[6]http://www.megasoftware.net/
[7]http://gsds.cbi.pku.edu.cn/
[8]http://meme-suite.org/
[9]https://www.r-project.org/

sequences, a total of 50 putative *WRKY* genes were identified in the California poppy draft genome, which were designated as *EcWRKY1* to *EcWRKY50* (**Table 1**). Of the 50 putative EcWRKY proteins, eight proteins possessed two WRKY domains, while the remaining proteins possessed only a single WRKY domain.

TABLE 1 | Identified *WRKY* genes in the California poppy genome.

Gene name	Gene ID	Predicted ORF length	Subgroup
EcWRKY1	Eca_sc000058.1_g0310.1	1596	I
EcWRKY2	Eca_sc000993.1_g1210.1	1455	I
EcWRKY3	Eca_sc003413.1_g2610.1	1446	I
EcWRKY4	Eca_sc002150.1_g1640.1	1107	I
EcWRKY5	Eca_sc001139.1_g0510.1	2190	I
EcWRKY6	Eca_sc001139.1_g2480.1	828	III
EcWRKY7	Eca_sc000967.1_g1460.1	2118	I
EcWRKY8	Eca_sc016018.1_g0010.1	1893	I
EcWRKY9	Eca_sc000141.1_g0430.1	2364	I
EcWRKY10	Eca_sc194739.1_g0350.1	1362	I
EcWRKY11	Eca_sc194475.1_g0580.1	876	III
EcWRKY12	Eca_sc026098.1_g0870.1	933	III
EcWRKY13	Eca_sc194486.1_g1840.1	1635	IIb
EcWRKY14	Eca_sc000774.1_g0410.1	1005	III
EcWRKY15	Eca_sc002052.1_g0460.1	1062	I
EcWRKY16	Eca_sc001936.1_g0700.1	969	IIa
EcWRKY17	Eca_sc194540.1_g3890.1	1071	I
EcWRKY18	Eca_sc035472.1_g0020.1	915	IIe
EcWRKY19	Eca_sc014828.1_g0060.1	1017	IIe
EcWRKY20	Eca_sc194624.1_g0220.1	996	IIc
EcWRKY21	Eca_sc194624.1_g0620.1	1344	IIc
EcWRKY22	Eca_sc193975.1_g1270.1	1731	IIb
EcWRKY23	Eca_sc000153.1_g1770.1	816	IIc
EcWRKY24	Eca_sc000193.1_g1170.1	1293	IIb
EcWRKY25	Eca_sc188774.1_g0010.1	1836	IIb
EcWRKY26	Eca_sc001048.1_g0050.1	1383	IIe
EcWRKY27	Eca_sc014577.1_g1270.1	996	IId
EcWRKY28	Eca_sc194627.1_g0820.1	711	IIe
EcWRKY29	Eca_sc194627.1_g0450.1	1194	IIe
EcWRKY30	Eca_sc194718.1_g0150.1	525	IId
EcWRKY31	Eca_sc194541.1_g0680.1	777	IIc
EcWRKY32	Eca_sc194541.1_g0990.1	876	IIe
EcWRKY33	Eca_sc002191.1_g0270.1	1047	IIc
EcWRKY34	Eca_sc015821.1_g0220.1	954	IIc
EcWRKY35	Eca_sc003662.1_g0020.1	1125	IIe
EcWRKY36	Eca_sc001705.1_g0090.1	645	IIc
EcWRKY37	Eca_sc194693.1_g1320.1	1122	IIe
EcWRKY38	Eca_sc000585.1_g0210.1	1038	IIc
EcWRKY39	Eca_sc001875.1_g0400.1	1056	IIc
EcWRKY40	Eca_sc000725.1_g1020.1	978	IIc
EcWRKY41	Eca_sc000325.1_g1390.1	543	IIc
EcWRKY42	Eca_sc006961.1_g0400.1	687	IIc
EcWRKY43	Eca_sc013752.1_g0230.1	945	III
EcWRKY44	Eca_sc194480.1_g0450.1	990	III
EcWRKY45	Eca_sc194641.1_g0500.1	1035	IId
EcWRKY46	Eca_sc057080.1_g0140.1	744	IIc
EcWRKY47	Eca_sc001754.1_g0640.1	1041	IId
EcWRKY48	Eca_sc000537.1_g1450.1	738	IId
EcWRKY49	Eca_sc000399.1_g0520.1	1131	IId
EcWRKY50	Eca_sc000360.1_g0330.1	1035	III

To classify the 50 EcWRKY proteins, multiple sequence alignment using the WRKY domain of the 50 EcWRKY proteins and 72 AtWRKY proteins was performed, and an unrooted phylogenetic tree was constructed using the NJ method (**Figure 1** and **Supplementary Figure 2**). Based on the classification of AtWRKY proteins and the phylogenetic tree, 11, 32, and 7 proteins were classified into groups I, II, and III, respectively (**Table 1**). Of the 32, respectively 1, 4, 13, 6, and 8 group II EcWRKY proteins were further divided into subgroups IIa, IIb, IIc, IId, and IIe. While majority of the group I WRKY proteins harbored two WRKY domains, three EcWRKY proteins in this group, namely EcWRKY4, EcWRKY15, and EcWRKY17, harbored only a single WRKY domain. Since the presence of group I WRKY proteins with a single WRKY domain has been reported in other plant species (Wei et al., 2012, 2016), these three EcWRKY proteins were classified as the group I WRKY proteins. California poppy has a similar number of WRKY groups to other plant species, with a similar number of genes in each group (**Supplementary Table 3**).

Homology search using CjWRKY1 and PsWRKY amino acid sequences as queries in the Eschscholzia Genome Database revealed high similarity of CjWRKY1 with three subgroup IIc WRKY proteins, namely EcWRKY36, EcWRKY41, and EcWRKY42. PsWRKY showed the highest similarity to group I EcWRKY1. A phylogenetic tree constructed using the WRKY domain sequences of the 50 EcWRKY proteins, CjWRKY1, and PsWRKY also showed the same result as the homology search (**Supplementary Figure 3**).

Gene Structure and Conserved Motif Composition of the EcWRKY Family

To compare the genomic DNA sequences of 50 *EcWRKY* genes, we determined their intron–exon structures (**Figure 2**). All *EcWRKY* genes had at least two exons, with 43 of the 50 *EcWRKY* genes having more than three exons. The distributions of introns and exons in the genomic sequences were relatively similar in each group. Most of the group I genes had four to five exons, except *EcWRKY4* and *EcWRKY17*, which had two exons. Furthermore, all subgroup IIb, IId, and III genes had six, three, and three exons, respectively. The phylogenetic tree indicated that subgroup IIc genes were divided into several clades. Four subgroup IIc genes (*EcWRKY31*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42*) in one clade had two exons, whereas the remaining nine genes in the other clades had three exons.

To examine the potential motifs of EcWRKY proteins in each family, we analyzed their conserved sequences using MEME Suite, a motif-based sequence analysis tool (**Figure 3**). Motifs 1, 2, and 3, which are components of the WRKY domain, were found in all EcWRKY proteins, while motif 4, which also contains the WRKYGQK core sequence, was only found in group I proteins, suggesting that motif 4 corresponds to the second WRKY domain. In addition to the WRKY domain, several conserved motifs were found in each EcWRKY family member. For example, motifs 7, 9, 11, and 15 were found only

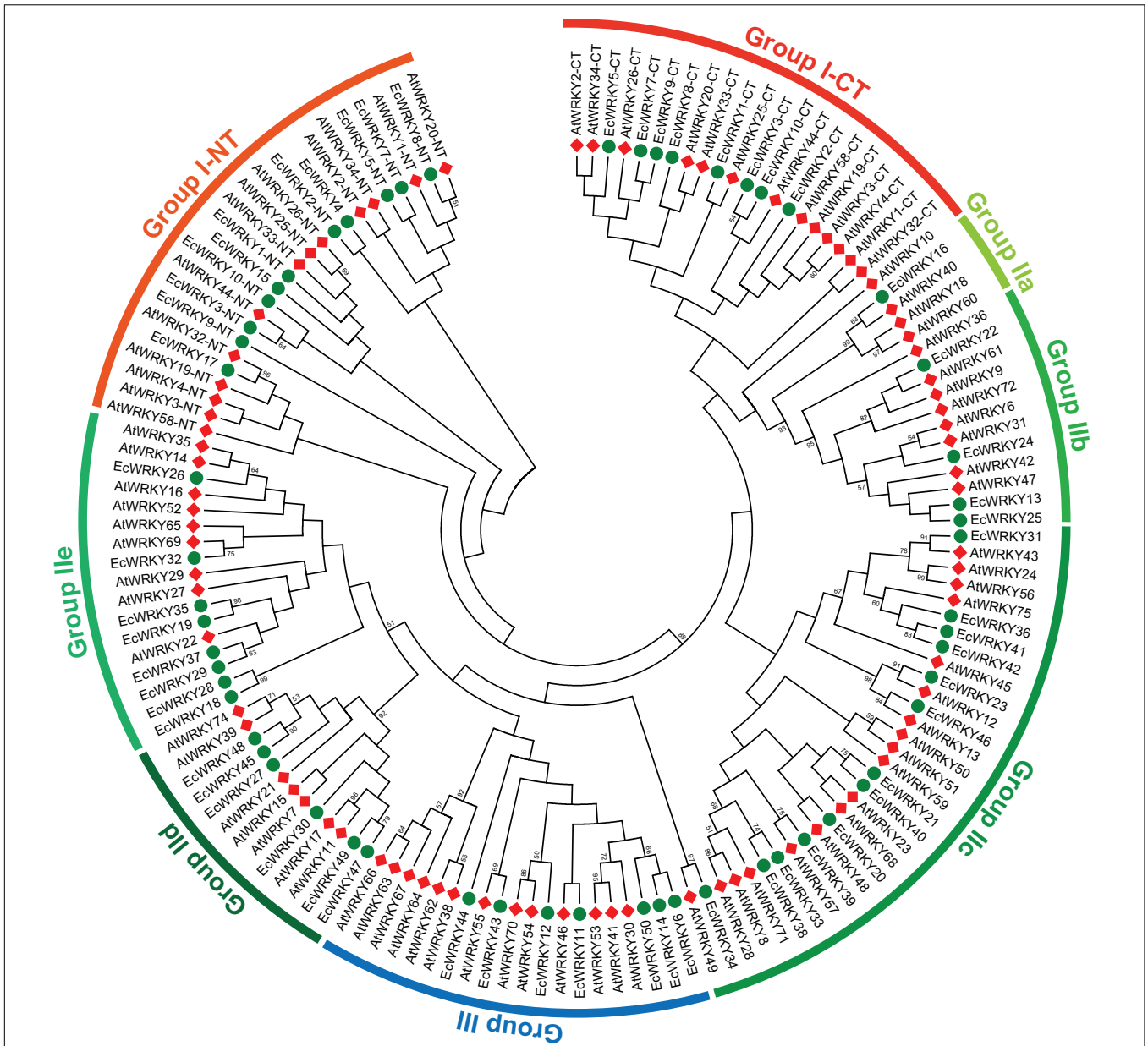


FIGURE 1 | Phylogenetic tree of WRKY proteins in *Eschscholzia californica* and *Arabidopsis thaliana*. A neighbor-joining tree was constructed based on the amino acid sequences of the WRKY domains of 50 EcWRKY (green circle) and 72 AtWRKY (red diamond) proteins using MEGA 7.0. Bootstrap confidence values from 1,000 replicates are indicated at each branch.

in (sub)group I, IIc, IIb, and IIe WRKY proteins, respectively. These conserved motifs might be important for the functional divergence of each protein group.

MeJA-Induced Expression Profiling of *EcWRKY* Genes

MeJA is an important phytohormone involved in defense response (Gundlach et al., 1992). Moreover, alkaloids play critical roles in protecting the plant body against pathogens and herbivores, and the expression of genes involved in the biosynthetic pathways of alkaloids, including BIAs, is strongly induced in response to MeJA (van der Fits and Memelink, 2000; Goossens et al., 2003; Ikezawa et al., 2007; Yamada et al., 2015). To investigate the MeJA responsiveness of *EcWRKY* genes, transcripts of California poppy seedlings treated with MeJA for 0, 0.5, 1, 3, 6, and 12 h were analyzed using RNA-Seq (**Figure 4**).

The expression of BIA biosynthetic enzyme genes (*Ec6OMT*, *Ec4'OMT*, *EcCYP80B1*, *EcCYP719A5*, *EcCYP719A2*, *EcMSH*, and *EcCYP719A9*) and TF genes (*EcbHLH1-2* and *EcAP2/ERFs*) were clearly induced in response to MeJA (**Figure 4**), as previously reported (Ikezawa et al., 2007, 2009; Yamada et al., 2015, 2020).

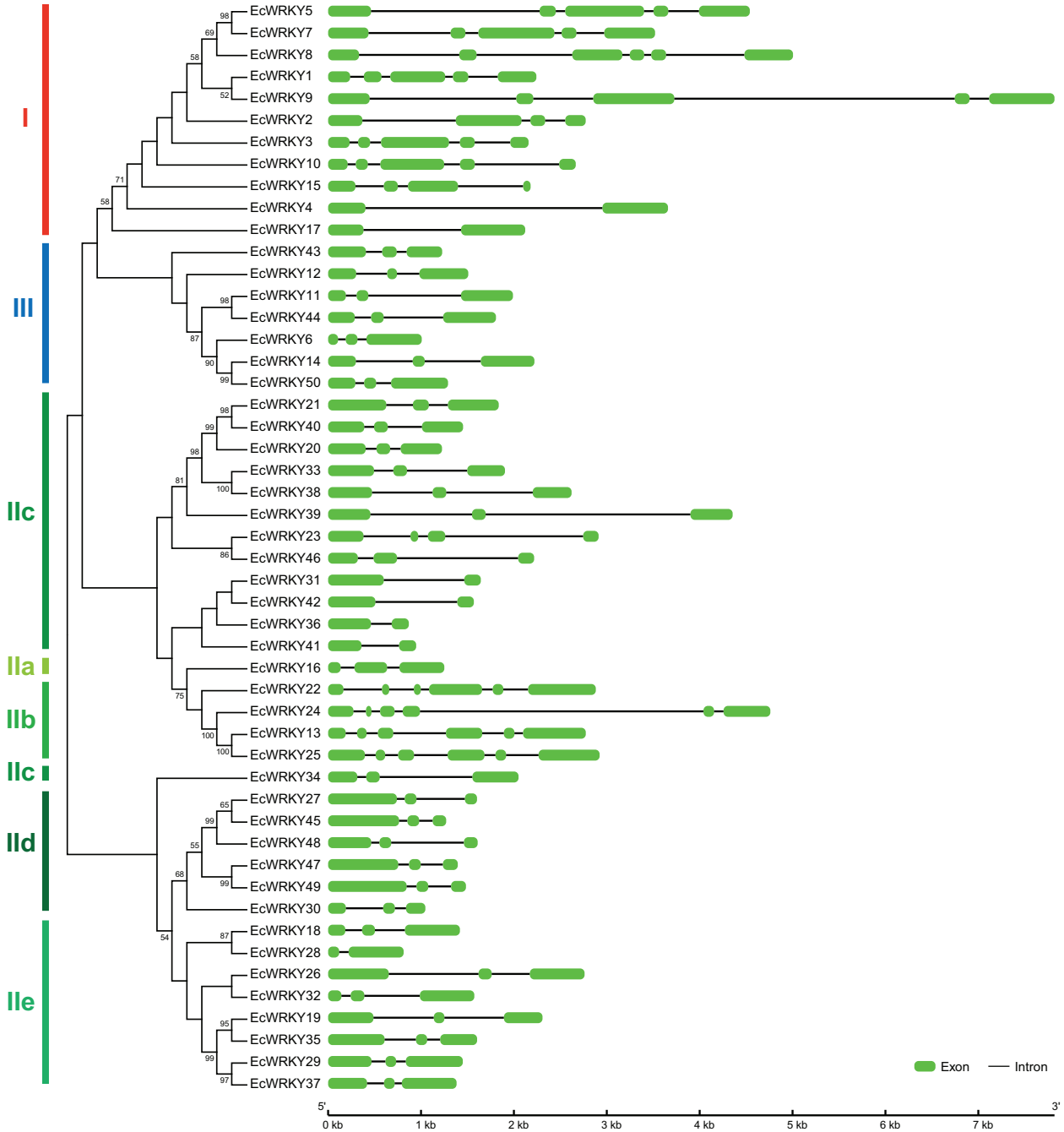
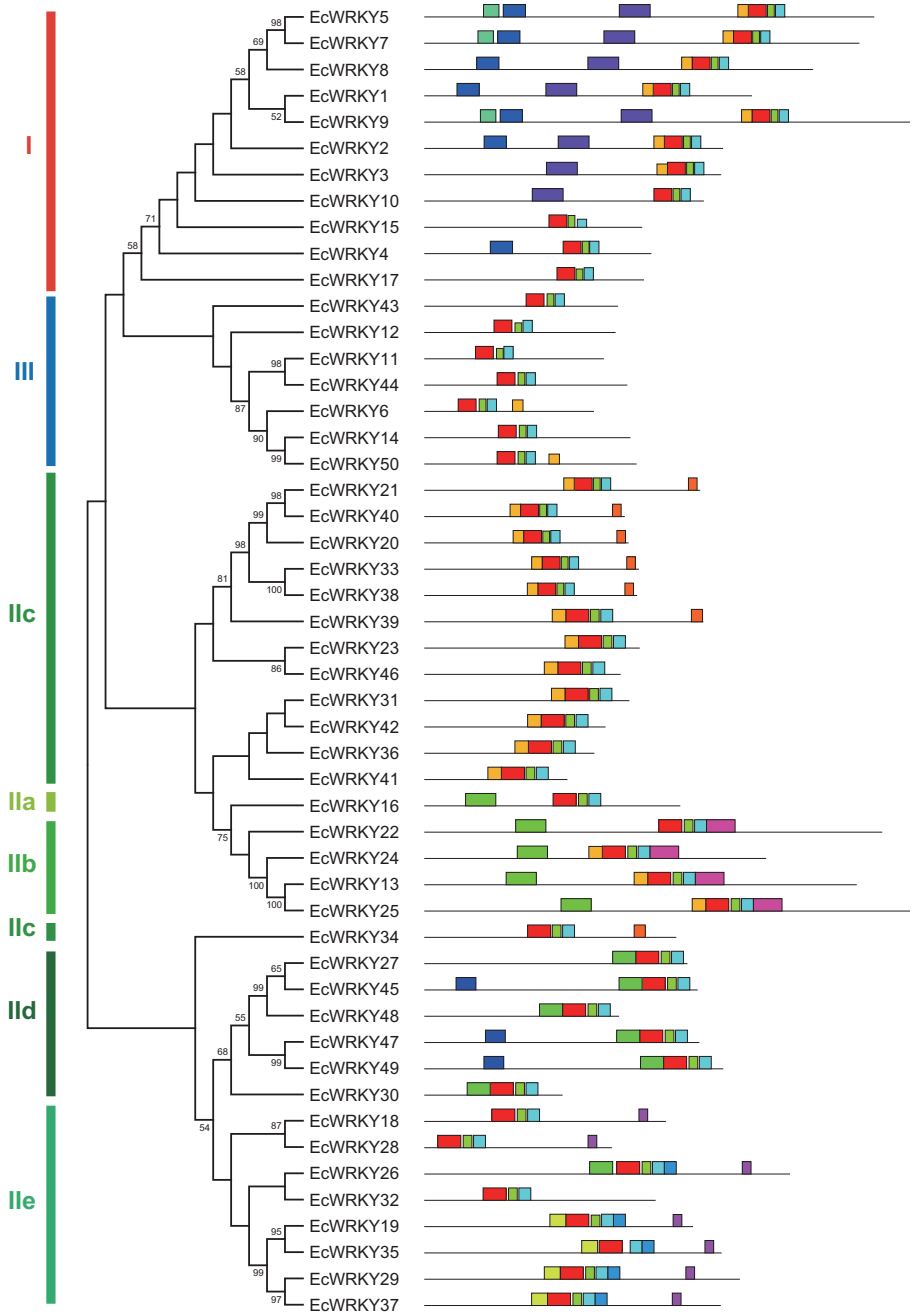


FIGURE 2 | Structure of *EcWRKY* genes. A phylogenetic tree was constructed as shown in **Figure 1**, based on the amino acid sequences of the WRKY domains of 50 EcWRKY proteins. The intron–exon structure of the *EcWRKY* genes was visualized using GSDS. Green and black lines represent the exons and introns, respectively.

The expression of *EcWRKY18*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42* was clearly increased following MeJA treatment. Hierarchical clustering indicated that these *EcWRKY* genes belong to the same clade as the MeJA-responsive BIA biosynthetic enzyme genes and *EcbHLH1-2*, whereas *EcAP2/ERF* genes, which showed earlier induction following MeJA treatment (Yamada et al., 2020), were placed in different clades, as discussed later. The expression profile of four *EcWRKY* genes was rather similar to that of BIA biosynthetic enzyme genes, which were strongly upregulated at 0.5–6 h (log2 FC > 1) and showed the greatest increase in expression after 6 h. In particular, *EcWRKY36* and *EcWRKY42* showed a more than 5-fold increase in expression after 6 h. Among the four EcWRKY TFs, EcWRKY36, EcWRKY41, and EcWRKY42



Motif	Symbol	Motif Consensus
1.		ILDDGYRWRKYGQKPVKGSPYPRSYYRCT
2.		DPSMVITTYEGEHNH
3.		GCPVRKQVERS
4.		DGYNWRKYGQKQVKGSEFPRSYYKCTYPNCQVKKKVERSLDGHITEIIYK
5.		RKIREPRFAFQTRSEVD
6.		SGRCHCSKRRKSRVKRSIKVPAISNKIAD
7.		SPYLTIPPGLSPTSLLDSPVFLSNSQAQPSPTTGTF
8.		PLPPAAMEMASTTSSAASMLLSGSTSSADGLMNNNF
9.		DHGLLQDIVPADMR
10.		RPKRRKNQQKRVVCQVPVDG
11.		ELDZLRAENQKLKGMLNQVTKDYNTLQMHLVEFMQQQQ
12.		AERMAARAGFBAPKLNTARIRSANM
13.		PIPTHRNSLAGSTRH
14.		AVNKFKKVISLLGRTRTGHARFRRA
15.		NDFFMGLEELT

FIGURE 3 | Conserved motifs of EcWRKY proteins. Fifteen motifs were identified using MEME and are indicated by colored rectangles. The height of the rectangles is proportional to the −log (*p*-value), truncated at the height of a motif with a *p*-value of 1e−10.

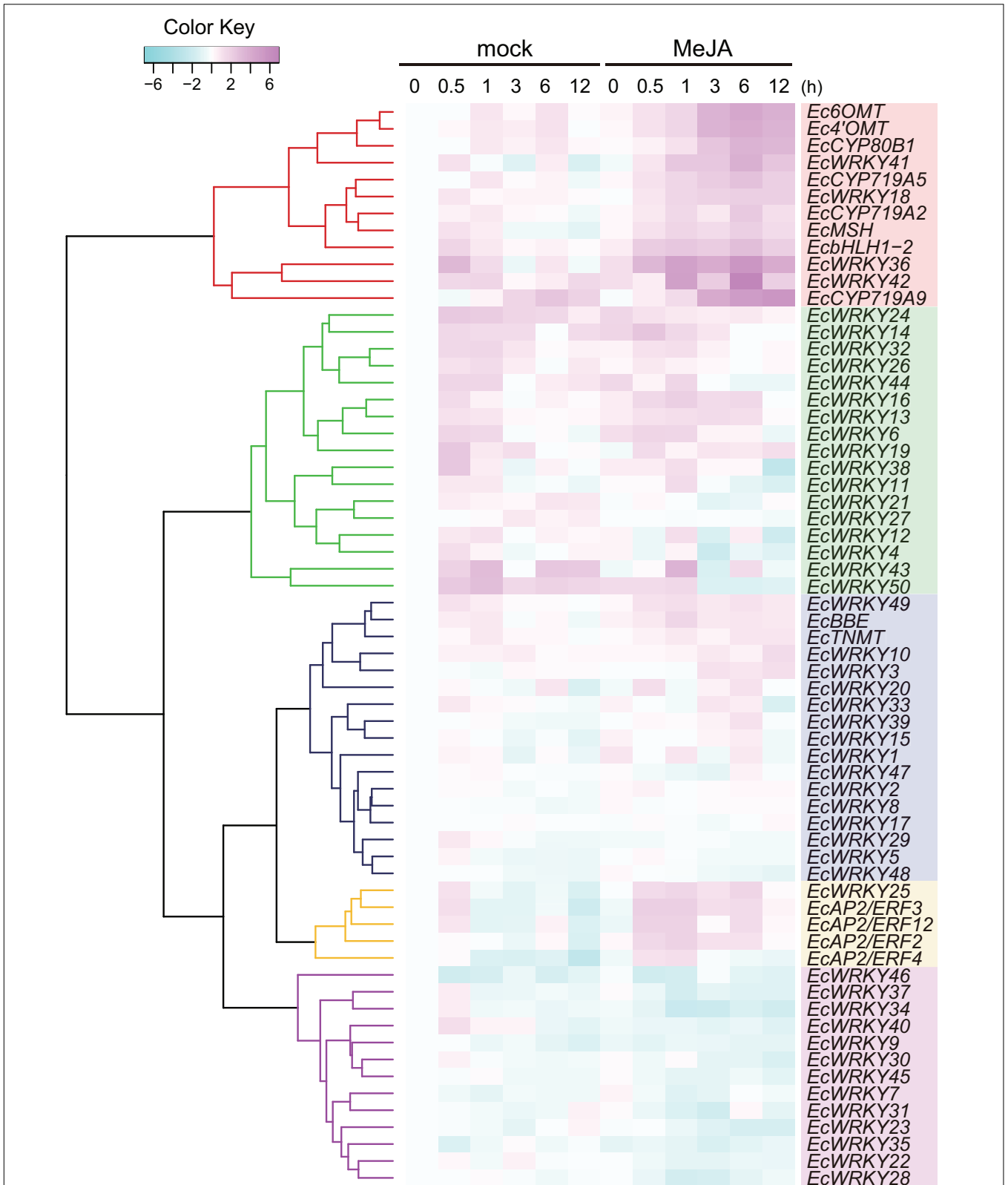


FIGURE 4 | RNA sequencing-based expression profiles of *EcWRKY* genes following methyl jasmonate (MeJA) treatment. Heat maps showing the clustering of *EcWRKY* genes with *EcbHLH1-2*, *Ec6OMT*, *EcCYP80B1*, *Ec4'OMT*, *EcBBE*, *EcCYP719A2*, *EcCYP719A5*, *EcCYP719A9*, *EcTNMT*, *EcMSH*, *EcAP2/ERF2*, *EcAP2/ERF3*, *EcAP2/ERF4*, and *EcAP2/ERF12* were created using log2-based FPKM values in R. Within each row, low and high values are indicated in light blue and pink, respectively. The scale represents the signal intensity of FPKM values.

showed high similarity to CjWRKY1, as mentioned above (**Supplementary Figure 3**).

EcBBE and *EcTNMT* involved in the later stages of BIA biosynthesis were also upregulated by MeJA, although they belonged to a different clade from other biosynthetic enzyme genes. The expression of *EcWRKY3*, *EcWRKY10*, and *EcWRKY49* was weakly induced in response to MeJA, and these genes were placed in clade closely related to *EcBBE* and *EcTNMT* (**Figure 4**). *EcWRKY49* showed the highest expression after 1 h, whereas *EcWRKY3* and *EcWRKY10* showed the highest expression after 12 h, which suggested that EcWRKY49 might act as an early regulator in the JA signaling cascade to control other MeJA-responsive genes including other *EcWRKYs*. Both *EcWRKY3* and *EcWRKY10*, which encode group I proteins, showed relatively similar expression patterns in response to MeJA and their slower response to MeJA indicated that EcWRKY3 and EcWRKY10 might work further downstream of the JA signaling cascade.

Although *EcWRKY13*, *EcWRKY16*, and *EcWRKY25* genes did not show similar expression patterns to BIA biosynthetic enzyme genes in response to MeJA, their expression was upregulated (log2 FC $>$ 1). *EcWRKY13* and *EcWRKY25* encode subgroup IIb proteins, whereas *EcWRKY16* encodes a subgroup IIa protein. Interestingly, the expression pattern of *EcWRKY25* was similar to that of the MeJA-responsive group IX *EcAP2/ERF* genes, which are the possible early regulators of BIA biosynthesis (Yamada et al., 2020).

To verify the expression profiles of *EcWRKY* genes that showed a clear increase in response to MeJA in RNA-Seq analysis, qRT-PCR was performed using cDNA derived from California poppy seedlings treated with MeJA for 0, 0.5, 1, 2, 6, and 24 h, with three biological replicates (**Figure 5**). As previously described (Yamada et al., 2015), two *EcbHLH1* and *EcBBE* genes were markedly upregulated in response to MeJA treatment, which is consistent with the results shown in **Figure 4**. The expression of *EcWRKY18*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42* was highly upregulated in response to MeJA treatment. The expression patterns of subgroup IIc *EcWRKY36*, *EcWRKY41*, and *EcWRKY42* were very similar. In contrast, MeJA did not strongly induce the expression of *EcWRKY3*, *EcWRKY10*, *EcWRKY13*, *EcWRKY16*, *EcWRKY25*, and *EcWRKY49* because of variation in gene expression in each seedling sample. Overall, these results indicate that *EcWRKY18*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42*, which showed a clear response to MeJA, are candidate *WRKY* genes involved in BIA biosynthesis.

Expression Analysis of *EcWRKY* Genes in Different Tissues of California Poppy

California poppy produces several types of BIA, which are accumulated in specific tissues. For instance, sanguinarine and chelerythrine are commonly accumulated in the root, whereas pavine-type BIAs, such as caryachine, californidine, and escholtzine, are only accumulate in aerial parts (**Supplementary Figure 4**). A previous study also revealed that genes involved in sanguinarine biosynthesis were highly expressed in roots (Ikezawa et al., 2007; Yamada et al., 2015). To further investigate the involvement of MeJA-responsive *EcWRKY* genes in the regulation of BIA biosynthesis, we examined the expression profiles of *EcWRKY18*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42* in different tissues, including leaf blades, petioles, roots, flower buds, and flowers (**Figure 6**) and compared them to the profiles of other TF genes involved in BIA biosynthesis, including *EcbHLH* and *EcERF*s. The expression profiles of *EcWRKY18* and *EcWRKY36* were highly similar to those of *Ec6OMT* and *EcBBE*, which encode sanguinarine biosynthetic enzymes; as such, these genes showed the highest expression in roots and relatively high expression in flowers. The expression profile of *EcWRKY42* was also similar to that of *EcWRKY18* and *EcWRKY36*, although it showed quite high expression in flowers as in roots. *EcbHLH1-2*, which is involved in sanguinarine biosynthesis, was highly and exclusively expressed in roots, as reported previously (Yamada et al., 2015). Meanwhile, group IX *EcAP2/ERF* genes were relatively highly expressed in leaves and roots, and these TF genes showed lower expression levels in flowers than *EcWRKY* genes. *EcCYP719A9*, encoding a possible enzyme involved in pavine-type BIA biosynthesis (Ikezawa et al., 2009), was highly expressed in aerial parts, particularly flower buds; however, the expression profile of any *EcWRKY* genes was not similar to that of *EcCYP719A9*. These results indicate that EcWRKY18 and subgroup IIc EcWRKYs, namely EcWRKY36, EcWRKY41, and EcWRKY42, are involved in the regulation of benzophenanthridine-type BIA biosynthesis.

Role of Subgroup IIc EcWRKY Proteins in BIA Biosynthesis

Since EcWRKY36, EcWRKY41, and EcWRKY42 are putative CjWRKY1 homologs in the California poppy, we focused on these proteins and examined their transcriptional activity using a transient LUC reporter assay. While CjWRKY1 showed clear transcriptional activity in *C. japonica* cells (Kato et al., 2007; Yamada et al., 2016), EcWRKY36, EcWRKY41, and EcWRKY42 showed little induction of LUC activity derived from the *Ec6OMT* and *EcCYP719A5* gene promoter:*LUC* constructs (**Supplementary Figure 5**). These results are consistent with our previous findings that the expression of many biosynthetic enzyme genes, including *Ec6OMT* and *EcCYP719A5*, was not significantly upregulated in CjWRKY1-overexpressing California poppy cells. These results also suggest that the regulatory role of WRKY TFs in BIA biosynthesis might be diversified between *C. japonica* and *E. californica*, and EcWRKY proteins serve different functions in the BIA biosynthetic pathway.

Coexpression Analysis of MeJA-Responsive *EcWRKY* Genes With Transporter-Encoding Genes

Our previous work revealed that heterologous expression of CjWRKY1 in California poppy cells increased BIA secretion into the culture medium (Yamada et al., 2017). This result suggests that the WRKY TFs modulate the expression of genes encoding transporter proteins involved in the efflux of alkaloids. To investigate the association between EcWRKY proteins involved in the regulation of genes encoding transporter proteins involved in the efflux of California poppy alkaloids, including BIAs,

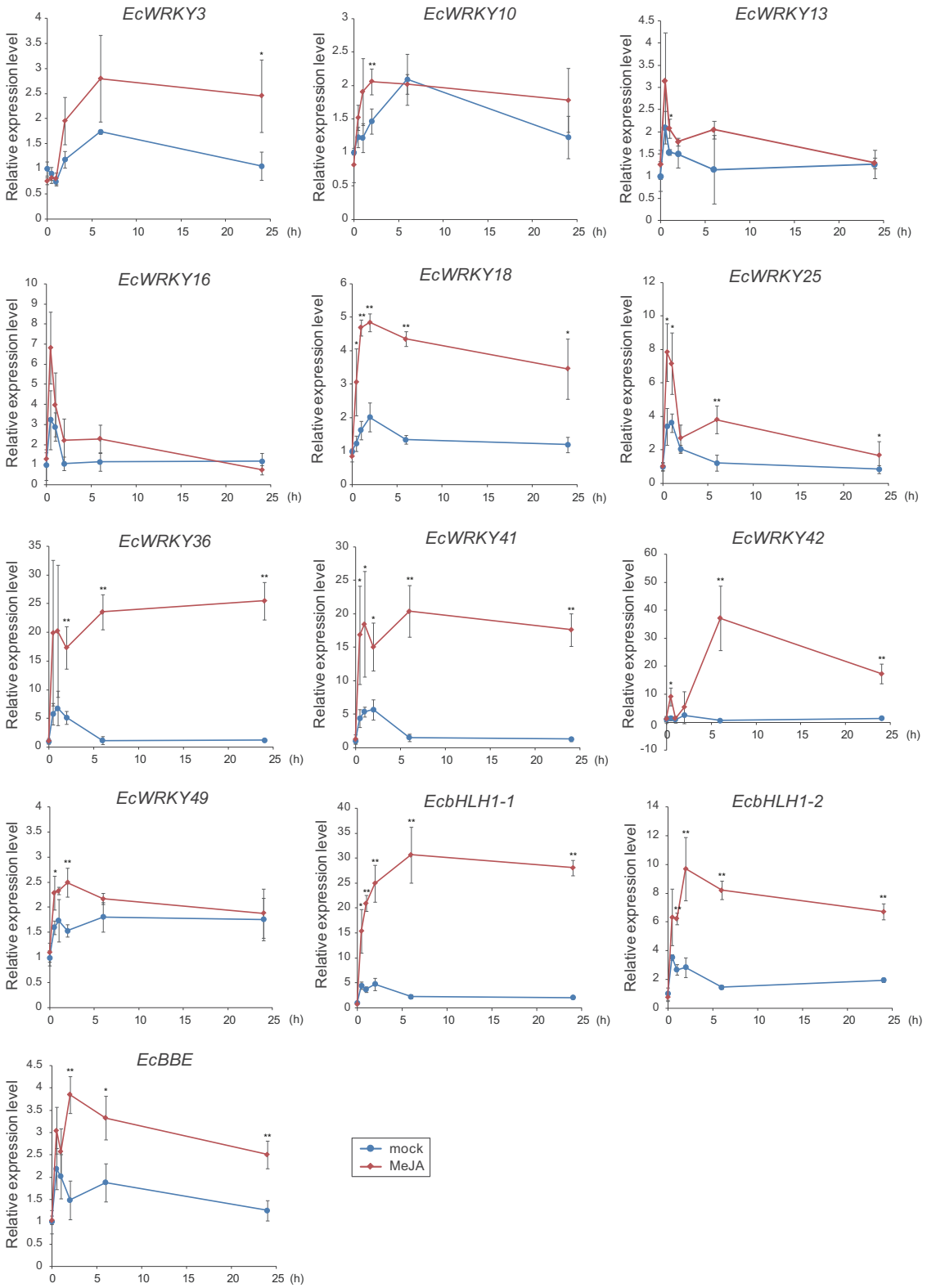


FIGURE 5 | Expression levels of several *EcWRKY* genes in methyl jasmonate (MeJA)-treated seedlings. Expression levels of ten *EcWRKY genes*, *EcbHLH1-1*, *EcbHLH1-2*, and *EcBBE* were determined by qRT-PCR. The relative transcript levels represent the values standardized to those of the mock (0 h) samples set to 1. Error bars indicate the standard deviations calculated from three biological replicates. The asterisks denote significant differences according to Student's *t*-test compared with the mocks: $^{*}P < 0.05$; $^{**}P < 0.01$.

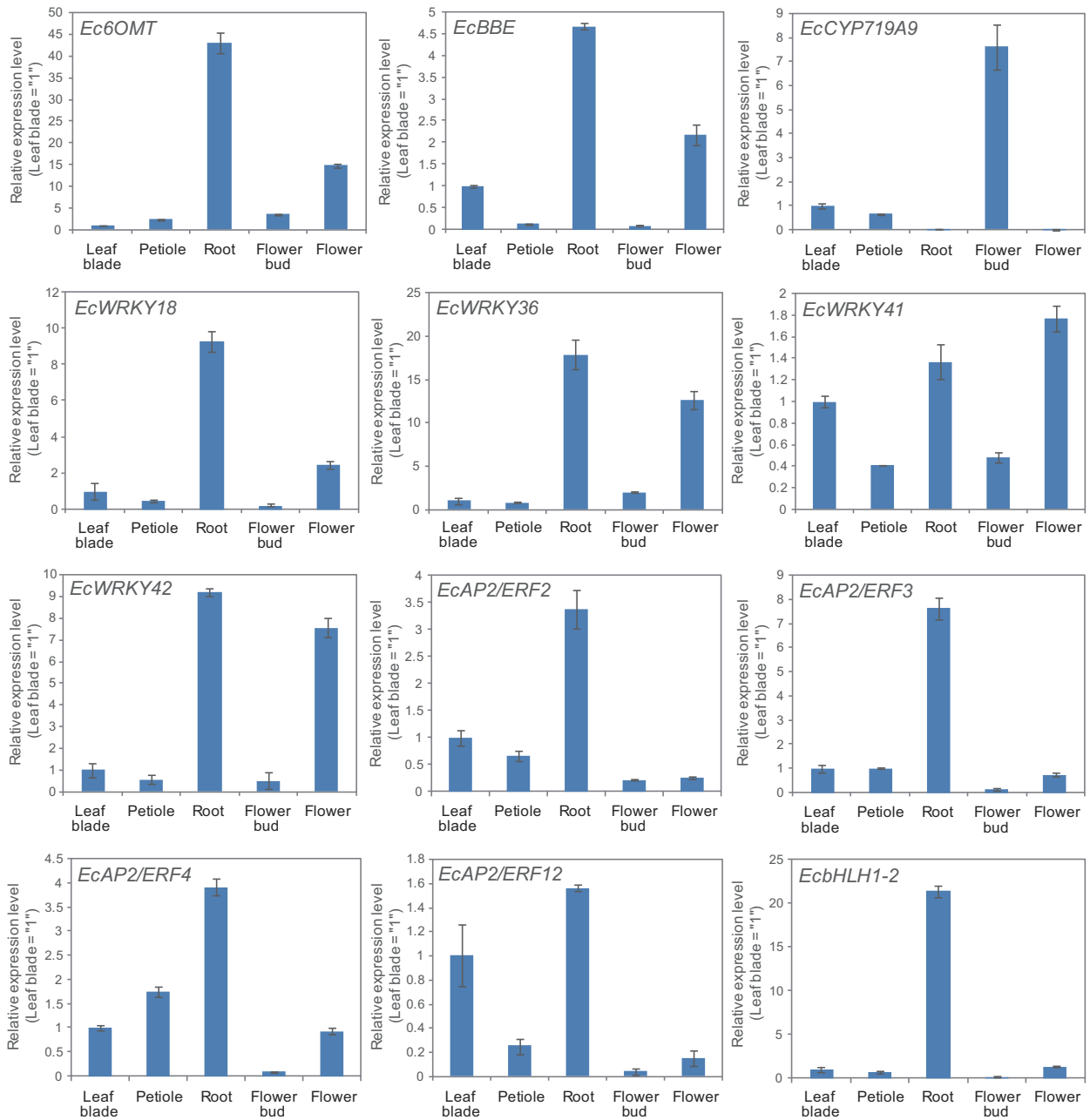


FIGURE 6 | Tissue expression patterns of methyl jasmonate (MeJA)-responsive *EcWRKY* genes. The expression levels of four MeJA-responsive *EcWRKY* genes as well as those of *Ec6OMT*, *EcBBE*, *EcCYP719A9*, *EcbHLH1-2*, *EcAP2/ERF2*, *EcAP2/ERF3*, *EcAP2/ERF4*, and *EcAP2/ERF12* were determined by qRT-PCR using cDNA derived from nine California poppy plants. The relative transcript levels represent the values standardized to those of the leaf blade or petiole samples set to 1. Error bars indicate the standard deviations calculated from three technical replicates.

we explored transporter-encoding genes that showed similar expression patterns to *EcWRKY18*, *EcWRKY36*, and *EcWRKY42* in response to MeJA. We screened 46 transporter-encoding genes that were upregulated (log2 FC > 1) following MeJA treatment for 1–12 h (**Supplementary Table 4**). These candidates included two genes encoding multidrug and toxic compound extrusion (MATE) transporters and three genes encoding B-type ATP-binding cassette (ABC) transporters (**Table 2**), have a possibility to be involved in the translocation of alkaloids, such as berberine in *C. japonica* and nicotine in *Nicotiana tabacum* (Shitan et al., 2003, 2013, 2014; Morita et al., 2009; Shoji et al., 2009; Takanashi et al., 2017). Hierarchical clustering analysis revealed that the expression patterns of Eca_sc001363.1_g1470.1 and Eca_sc100701.1_g2100.1, which are putative B-type ABC transporter genes in response to MeJA, were relatively similar to those of *EcWRKY18* and *EcWRKY36* and *EcWRKY42*,

TABLE 2 | MeJA-responsive genes encoding ABC and MATE transporters.

Gene ID	Annotation	ORF length
Eca_sc001363.1_g1470.1	Nr = XP_010255510.1 PREDICTED: ABC transporter B family member 15-like [*Nelumbo nucifera*] Araport = AT3G28345.1 | ABC transporter family protein | Chr3:10593921-10598775 REVERSE LENGTH = 1240 | 201606	3,858
Eca_sc004559.1_g0090.1	Nr = XP_008796381.1 PREDICTED: protein DETOXIFICATION 33-like [*Phoenix dactylifera*] Araport = AT1G47530.1 | MATE efflux family protein | Chr1:17451724-17454110 FORWARD LENGTH = 484 | 201606	1,452
Eca_sc011255.1_g0480.1	Nr = XP_010271025.1 PREDICTED: ABC transporter B family member 11-like isoform X1 [*Nelumbo nucifera*] Araport = AT1G02520.3 | P-glycoprotein 11 | Chr1:524134-528745 FORWARD LENGTH = 1278 | 201606	3873
Eca_sc100701.1_g2100.1	Nr = XP_002279471.2 PREDICTED: ABC transporter B family member 13-like [*Vitis vinifera*] Araport = AT1G27940.2 | P-glycoprotein 13 | Chr1:9733597-9737211 REVERSE LENGTH = 1031 | 201606	2,685
Eca_sc194586.1_g0360.1	Nr = XP_010260247.1 PREDICTED: MATE efflux family protein LAL5-like [*Nelumbo nucifera*] Araport = AT3G23560.1 | MATE efflux family protein | Chr3:8454361-8456588 REVERSE LENGTH = 477 | 201606	1,440

respectively (**Figure 7**). Therefore, these ABCB transporter genes might be involved in the transport of BIAs and regulated by MeJA-responsive EcWRKY transcription factors.

DISCUSSION

Genes of the WRKY superfamily, which is one of the largest groups of TFs involved in plant development and response to various stresses, have been identified in various plants (Rushton et al., 2010). Recent advances in whole-genome sequencing technologies have enabled us to perform genome-wide analysis of *WRKY* genes in many plant species. To date, 74 *WRKY* genes have been identified in *A. thaliana* (Ulker and Somssich, 2004), 52 in *C. roseus* (Schluttenhofer et al., 2014), 54 in *Ananas comosus* (Xie et al., 2018), 55 in *Cucumis sativus* (Ling et al., 2011), 85 in *Manihot esculenta* (Wei et al., 2016), 70 in *Aquilaria sinensis* (Xu et al., 2020), and 120 in *Gossypium raimondii* (Cai et al., 2014). However, our study is the first report on the genome-wide identification of WRKY TFs from *E. californica*, a BIA-producing plant of the Papaveraceae family. We identified 50 WRKY members in the California poppy draft genome (**Table 1**). The different number of *WRKY* genes among plant species may be implicated in differences in the size of the genome and functional diversification of WRKY family proteins during evolution. Based on phylogenetic analysis (**Figure 1**), the 50 EcWRKY proteins were classified into 11 group I proteins; 32 group II proteins, further divided into 1, 4, 13, 6, and 8 proteins in subgroup IIa, IIb, IIc, IId, and IIe, respectively; and 7 group III proteins. The distribution of each group of WRKY proteins in *E. californica* was quite similar to that in other plant species, although there were lower subgroup IIa and IIb proteins in California poppy than those in other species (**Supplementary Table 3**). Since *E. californica* is a basal eudicot of the Papaveraceae family, this difference likely reflects the evolutionary history of land plant subgroup IIa and IIb genes, which are considered to have evolved from group I genes due to deletion of the domain structure (Wang et al., 2014).

Gene structure and conserved motif analyses indicated that each protein group shared a similar number of introns and similar motifs (**Figures 2**, **3**). All *EcWRKY* genes possessed more than one intron, which is consistent with reports in other plant species (Wei et al., 2016; Xie et al., 2018; Xu et al., 2020). These results suggest that gene duplication and structural diversification of *WRKY* genes may have occurred at the early stages of evolution. Furthermore, the similar motif compositions of each WRKY protein group indicate the potential functional similarity among them, as three subgroup IIc *EcWRKY* genes, namely *EcWRKY36*, *EcWRKY41*, and *EcWRKY42*, showed a marked response to MeJA.

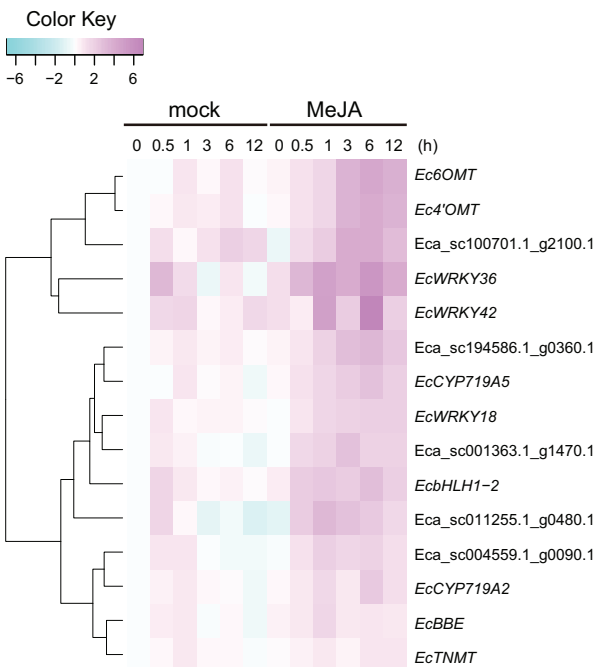


FIGURE 7 | Expression profiles of putative transporter genes and *EcWRKY* genes following methyl jasmonate (MeJA) treatment. Heat maps showing the clustering of three ABCB transporter genes (Eca_sc001363.1_g1470.1, Eca_sc011255.1_g0480.1, and Eca_sc100701.1_g2100.1) and two MATE transporter genes (Eca_sc004559.1_g0090.1 and Eca_sc194586.1_g0360.1) with three *EcWRKY* genes (*EcWRKY18*, *EcWRKY36*, and *EcWRKY42*), *EcbHLH1-2*, *Ec6OMT*, *Ec4'OMT*, *EcBBE*, *EcCYP719A2*, *EcCYP719A5*, and *EcTNMT* were created using log2-based FPKM values. The scale represents the signal intensities of the FPKM values.

Jasmonic acid signaling is a critical axis in defense response, including the biosynthesis of specialized metabolites, which act as chemical defense compounds against herbivores and pathogens. Alkaloid production is modulated by JA signaling, and many JA-responsive TFs, which play vital roles in the regulation of genes involved in the JA signaling cascade, have been identified and characterized (Yamada and Sato, 2013). Hence, JA-responsive *WRKY* genes in *E. californica* may regulate the expression of genes involved in the BIA biosynthetic pathway. The results of RNA-Seq and qRT-PCR revealed that four *EcWRKY* genes, namely *EcWRKY18*, *EcWRKY36*, *EcWRKY41*, and *EcWRKY42*, were upregulated following MeJA treatment (**Figures 4**, **5**). The response pattern of *EcWRKY* genes was relatively similar to that of BIA biosynthetic enzyme genes, including group IX *EcAP2/ERF* and *EcbHLH1-2*, which showed a rapid MeJA response (Yamada et al., 2020). These results indicate that EcWRKY TFs may function downstream of group IX EcAP2/ERF and EcbHLH1-2 TFs in the JA signaling cascade (**Figure 8**). To investigate the detailed transcriptional network of BIA biosynthesis in California poppy, further functional characterization of bHLH, AP2/ERF, and WRKY TFs using stable transformants is warranted. Furthermore, EcWRKY36, EcWRKY41, and EcWRKY42, which are potential CjWRKY1 homologs, showed little transcriptional activity in transient LUC assay using the *Ec6OMT* and *EcCYP719A5* gene promoters (**Supplementary Figure 5**), which is consistent with our previous results of CjWRKY1 overexpression in Californian poppy cells (Yamada et al., 2017). Therefore, the function of WRKY proteins involved in BIA biosynthesis may have diversified during evolution in *E. californica* and *C. japonica*, and EcWRKY proteins may serve additional functions contributing to BIA production. Additionally, post-transcriptional regulation might be involved in BIA biosynthesis such as protein phosphorylation and degradation (Yamada and Sato, 2016).

Group I WRKY proteins are involved in the regulation of BIA biosynthesis. For instance, PsWRKY has been identified as a potent transcriptional activator of BIA biosynthetic genes in *P. somniferum* (Mishra et al., 2013). Moreover,

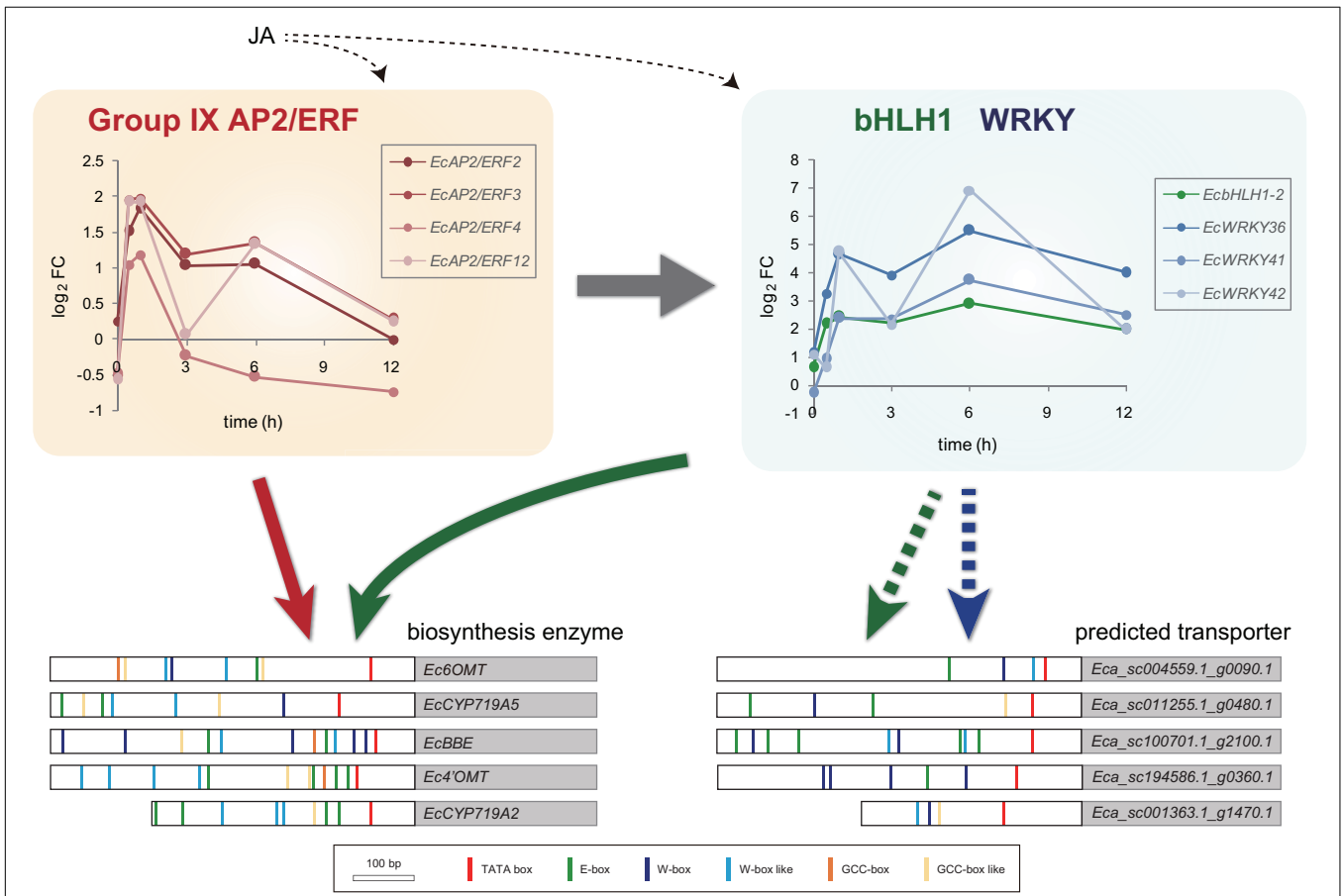


FIGURE 8 | Predicted model of gene expression mechanism in benzylisoquinoline alkaloid biosynthesis via jasmonic acid (JA)-responsive transcription factors. Temporal expression patterns of group IX *EcAP2/ERF*, *EcbHLH1-2*, and *EcWRKY* genes are shown with log2 FC values based on the RNA-Seq data. Large white and gray boxes indicate the partial promoter and coding sequence, respectively. Small boxes with several colors represent the putative *cis*-acting elements. Based on the previous results, EcAP2/ERF and EcbHLH1 show transactivation of the promoters of biosynthetic enzyme genes (indicated by red and green arrows, respectively). Since the promoters of the predicted transporter genes harbor many putative E-box and W-box elements, EcbHLH and EcWRKY transcription factors may regulate them rather than EcAP2/ERF proteins (indicated by green and blue dotted arrows, respectively).

Apuya et al. (2008) reported that AtWRKY1 overexpression in *P. somniferum* and *E. californica* cultured cells enhanced BIA accumulation. In contrast, our RNA-Seq and qRT-PCR data revealed that the expression of group I *EcWRKY* genes, including *EcWRKY1*, the closest homologous gene of *PsWRKY*, was not or weakly induced in response to MeJA (**Figures 4**, **5**). These results also suggest the functional diversification of group I WRKY proteins in the Papaveraceae family during evolution. However, whether *PsWRKY* modulates the expression of enzymes involved in sanguinarine or morphine biosynthetic pathways in *P. somniferum* remains unclear. Therefore, detailed functional characterization of group I WRKY proteins in BIA-producing plants is essential.

Our previous study revealed that CjWRKY1 overexpression in California poppy cultured cells enhanced BIA accumulation in culture medium (Yamada et al., 2017), suggesting that WRKY proteins regulate the expression of genes encoding potential transporters of BIAs in this plant. Several TFs regulating genes encoding transport proteins of specialized metabolites have been identified. For instance, *A. thaliana* MYB TFs regulate the expression of genes involved in the transport of proanthocyanidins (Sharma and Dixon, 2005), and grapevine MYB and WRKY TFs synergistically regulate the expression of genes involved in flavonoid accumulation (Amato et al., 2019). During alkaloid biosynthesis, the expression of genes encoding MATE transporters is regulated by bHLH and AP2/ERF TFs, which also control the expression of biosynthetic enzyme genes (Shoji et al., 2010; Takanashi et al., 2017). The compartmentalization of cytotoxic alkaloids in specific organs or organelles via transporters is important for protection against insects and herbivores, and the regulation of expression of such transporters is important; however, little is known regarding transporters involved in the compartmentalization of BIAs in *E. californica* cells. In this light, we investigated the coexpression patterns of *WRKY* and transporter genes in MeJA-treated *E. californica* seedlings. Three ABCB transporter and two MATE transporter genes that showed a clear MeJA response were coexpressed with *EcWRKY18*, *EcWRKY36*, and *EcWRKY42*. Interestingly, search for putative *cis*-elements using the New PLACE database[10] in the promoter regions of MeJA-responsive transporter genes and biosynthetic enzyme genes (Higo et al., 1999) revealed that there were few GCC-box-like *cis*-elements, which are target sequences of group IX AP2/ERF TFs, in these transporter genes and at least one GCC-box or GCC-box-like nucleotide sequence was present in genes encoding biosynthetic enzymes (**Figure 8**). To reveal the direct interaction of these putative *cis*-elements with AP2/ERF, WRKY, and bHLH TFs, additional analyses are required in future studies. Furthermore, the predicted transporter-encoding genes that were highly upregulated by MeJA included many genes encoding nitrate transporter 1/peptide transporter family (NPF) proteins and purine permeases (**Supplementary Table 4**). CrNPF2.9 involved in the transport of strictosidine from vacuole to cytosol and BIA uptake purine permeases have recently been isolated from *C. roseus* and opium poppy, respectively (Payne et al., 2017; Dastmalchi et al., 2019). Therefore, MeJA-responsive NPF transporters and purine permeases might be involved in the translocation of BIAs in *E. californica*.

In conclusion, our genome-wide analysis and expression profiling of the WRKY family genes in *E. californica* would be useful for understanding the regulatory mechanisms underlying of BIA biosynthesis, accumulation, and translocation. Especially, different EcWRKY proteins might regulate the spatiotemporal expression patterns of genes related to BIA biosynthesis. Further characterization of EcWRKY TFs and transporters is required to elucidate regulatory mechanisms of BIA production and accumulation in California poppy. This information will contribute to the development of metabolic and transport engineering approaches for the efficient production of valuable alkaloids.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ddbj.nig.ac.jp/, BEHA01000001–BEHA01053253.

AUTHOR CONTRIBUTIONS

YY and FS conceived and designed the study and wrote the manuscript. YY and SN analyzed the genomic and transcriptomic data and performed the experiments. NS and FS supervised the project and discussed the results. All authors reviewed the manuscript.

FUNDING

This research was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) [Grant-in-Aid for Scientific Research (S) 26221201 to FS].

ACKNOWLEDGMENTS

We thank Hideki Hirakawa (Kazusa DNA Research Institute, Japan) for depositing the draft genome sequence data of California poppy and establishing the Eschscholzia Genome DataBase. We thank Azusa Hirano and Asuka Tomita (Kobe Pharmaceutical University, Japan) for growing the California poppy plants in flowerpots.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.699326/full#supplementary-material

[10]https://www.dna.affrc.go.jp/PLACE/?action=newplace

REFERENCES

Amato, A., Cavallini, E., Walker, A. R., Pezzotti, M., Bliek, M., Quattrocchio, F., et al. (2019). The MYB5-driven MBW complex recruits a WRKY factor to enhance the expression of targets involved in vacuolar hyper-acidification and trafficking in grapevine. *Plant J.* 99, 1220–1241. doi: 10.1111/tpj.14419

Apuya, N. R., Park, J. H., Zhang, L., Ahyow, M., Davidow, P., Van Fleet, J., et al. (2008). Enhancement of alkaloid production in opium and California poppy by transactivation using heterologous regulatory factors. *Plant Biotechnol. J.* 6, 160–175. doi: 10.1111/j.1467-7652.2007.00302.x

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

Cai, C., Niu, E., Du, H., Zhao, L., Feng, Y., and Guo, W. (2014). Genome-wide analysis of the WRKY transcription factor gene family in *Gossypium raimondii* and the expression of orthologs in cultivated tetraploid cotton. *Crop J.* 2, 87–101. doi: 10.1016/j.cj.2014.03.001

Chen, M., Yan, T., Shen, Q., Lu, X., Pan, Q., Huang, Y., et al. (2017). GLANDULAR TRICHOME-SPECIFIC WRKY 1 promotes artemisinin biosynthesis in *Artemisia annua*. *New Phytol.* 214, 304–316. doi: 10.1111/nph.14373

Dastmalchi, M., Chang, L., Chen, R., Yu, L., Chen, X., Hagel, J. M., et al. (2019). Purine permease-type benzylisoquinoline alkaloid transporters in opium poppy. *Plant Physiol.* 181, 916–933. doi: 10.1104/pp.19.00565

Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/s1360-1385(00)01600-9

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Fabre, N., Claparols, C., Richelme, S., Angelin, M. L., Fourasté, I., and Moulis, C. (2000). Direct characterization of isoquinoline alkaloids in a crude plant extract by ion-pair liquid chromatography-electrospray ionization tandem mass spectrometry: example of *Eschscholtzia californica*. *J. Chromatogr. A.* 904, 35–46. doi: 10.1016/s0021-9673(00)00919-5

Goossens, A., Häkkinen, S. T., Laakso, I., Seppänen-Laakso, T., Biondi, S., De Sutter, V., et al. (2003). A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8595–8600. doi: 10.1073/pnas.1032967100

Gundlach, H., Müller, M. J., Kutchan, T. M., and Zenk, M. H. (1992). Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2389–2393. doi: 10.1073/pnas.89.6.2389

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297

Hori, K., Yamada, Y., Purwanto, R., Minakuchi, Y., Toyoda, A., Hirakawa, H., et al. (2018). Mining of the Uncharacterized Cytochrome P450 genes involved in alkaloid biosynthesis in california poppy using a draft genome sequence. *Plant Cell Physiol.* 59, 222–233. doi: 10.1093/pcp/pcx210

Ikezawa, N., Iwasa, K., and Sato, F. (2007). Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in *Eschscholzia californica*. *FEBS J.* 274, 1019–1035. doi: 10.1111/j.1742-4658.2007.05652.x

Ikezawa, N., Iwasa, K., and Sato, F. (2009). CYP719A subfamily of cytochrome P450 oxygenases and isoquinoline alkaloid biosynthesis in *Eschscholzia californica*. *Plant Cell Rep.* 28, 123–133. doi: 10.1007/s00299-008-0624-8

Kato, N., Dubouzet, E., Kokabu, Y., Yoshida, S., Taniguchi, Y., Dubouzet, J. G., et al. (2007). Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol.* 48, 8–18. doi: 10.1093/pcp/pcl041

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

Ling, J., Jiang, W., Zhang, Y., Yu, H., Mao, Z., Gu, X., et al. (2011). Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genom.* 12:471. doi: 10.1186/1471-2164-12-471

Ma, D., Pu, G., Lei, C., Ma, L., Wang, H., Guo, Y., et al. (2009). Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol.* 50, 2146–2161. doi: 10.1093/pcp/pcp149

Mishra, S., Triptahi, V., Singh, S., Phukan, U. J., Gupta, M. M., Shanker, K., et al. (2013). Wound induced tanscriptional regulation of benzylisoquinoline pathway and characterization of wound inducible PsWRKY transcription factor from *Papaver somniferum*. *PLoS One* 8:e52784. doi: 10.1371/journal.pone.0052784

Morita, M., Shitan, N., Sawada, K., Van Montagu, M. C., Inzé, D., Rischer, H., et al. (2009). Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2447–2452. doi: 10.1073/pnas.0812512106

Payne, R. M., Xu, D., Foureau, E., Teto Carqueijeiro, M. I., Oudin, A., Bernonville, T. D., et al. (2017). An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat. Plants* 3:16208. doi: 10.1038/nplants.2016.208

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006

Schluttenhofer, C., Pattanaik, S., Patra, B., and Yuan, L. (2014). Analyses of *Catharanthus roseus* and *Arabidopsis thaliana* WRKY transcription factors reveal involvement in jasmonate signaling. *BMC Genom.* 15:502. doi: 10.1186/1471-2164-15-502

Sharma, S. B., and Dixon, R. A. (2005). Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsis thaliana*. *Plant J.* 44, 62–75. doi: 10.1111/j.1365-313X.2005.02510.x

Shitan, N., Bazin, I., Dan, K., Obata, K., Kigawa, K., Ueda, K., et al. (2003). Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 751–756. doi: 10.1073/pnas.0134257100

Shitan, N., Dalmas, F., Dan, K., Kato, N., Ueda, K., Sato, F., et al. (2013). Characterization of *Coptis japonica* CjABCB2, an ATP-binding cassette protein involved in alkaloid transport. *Phytochemistry* 91, 109–116. doi: 10.1016/j.phytochem.2012.02.012

Shitan, N., Minami, S., Morita, M., Hayashida, M., Ito, S., Takanashi, K., et al. (2014). Involvement of the leaf-specific multidrug and toxic compound extrusion (MATE) transporter Nt-JAT2 in vacuolar sequestration of nicotine in *Nicotiana tabacum*. *PLoS One* 9:e108789. doi: 10.1371/journal.pone.0108789

Shoji, T., Inai, K., Yazaki, Y., Sato, Y., Takase, H., Shitan, N., et al. (2009). Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. *Plant Physiol.* 149, 708–718. doi: 10.1104/pp.108.132811

Shoji, T., Kajikawa, M., and Hashimoto, T. (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* 22, 3390–3409. doi: 10.1105/tpc.110.078543

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

Takanashi, K., Yamada, Y., Sasaki, T., Yamamoto, Y., Sato, F., and Yazaki, K. (2017). A multidrug and toxic compound extrusion transporter mediates berberine accumulation into vacuoles in *Coptis japonica*. *Phytochemistry* 138, 76–82. doi: 10.1016/j.phytochem.2017.03.003

Ulker, B., and Somssich, I. E. (2004). WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol.* 7, 491–498. doi: 10.1016/j.pbi.2004.07.012

van der Fits, L., and Memelink, J. (2000). ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 289, 295–297. doi: 10.1126/science.289.5477.295

Wang, L., Zhu, W., Fang, L., Sun, X., Su, L., Liang, Z., et al. (2014). Genome-wide identification of WRKY family genes and their response to cold stress in *Vitis vinifera*. *BMC Plant Biol.* 14:103. doi: 10.1186/1471-2229-14-103

Wei, K. F., Chen, J., Chen, Y. F., Wu, L. J., and Xie, D. X. (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res.* 19, 153–164. doi: 10.1093/dnares/dsr048

Wei, Y., Shi, H., Xia, Z., Tie, W., Ding, Z., Yan, Y., et al. (2016). Genome-wide identification and expression analysis of the WRKY Gene family in cassava. *Front. Plant Sci.* 7:25. doi: 10.3389/fpls.2016.00025

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J., Chakrabarty, R., et al. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* 166, 122–134. doi: 10.1016/j.jbiotec.2013.04.004

Xie, T., Chen, C., Li, C., Liu, J., Liu, C., and He, Y. (2018). Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. *BMC Genom.* 19:490. doi: 10.1186/s12864-018-4880-x

Xu, Y. H., Sun, P. W., Tang, X. L., Gao, Z. H., Zhang, Z., and Wei, J. H. (2020). Genome-wide analysis of WRKY transcription factors in *Aquilaria sinensis* (Lour.) Gilg. *Sci. Rep.* 10:3018. doi: 10.1038/s41598-020-59597-w

Xu, Y. H., Wang, J. W., Wang, S., Wang, J. Y., and Chen, X. Y. (2004). Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant Physiol.* 135, 507–515. doi: 10.1104/pp.104.038612

Yamada, Y., Hirakawa, H., Hori, K., Minakuchi, Y., Toyoda, A., Shitan, N., et al. (2021). Comparative analysis using the draft genome sequence of California poppy (*Eschscholzia californica*) for exploring the candidate genes involved in benzylisoquinoline alkaloid biosynthesis. *Biosci. Biotechnol. Biochem.* 85, 851–859. doi: 10.1093/bbb/zbaa091

Yamada, Y., Motomura, Y., and Sato, F. (2015). CjbHLH1 homologs regulate sanguinarine biosynthesis in *Eschscholzia californica* cells. *Plant Cell Physiol.* 56, 1019–1030. doi: 10.1093/pcp/pcv027

Yamada, Y., Nishida, S., Shitan, N., and Sato, F. (2020). Genome-wide identification of AP2/ERF transcription factor-encoding genes in California poppy (*Eschscholzia californica*) and their expression profiles in response to methyl jasmonate. *Sci. Rep.* 10:18066. doi: 10.1038/s41598-020-75069-7

Yamada, Y., and Sato, F. (2013). Transcription factors in alkaloid biosynthesis. *Int. Rev. Cell Mol. Biol.* 305, 339–382. doi: 10.1016/B978-0-12-407695-2.00008-1

Yamada, Y., and Sato, F. (2016). Tyrosine phosphorylation and protein degradation control the transcriptional activity of WRKY involved in benzylisoquinoline alkaloid biosynthesis. *Sci. Rep.* 6:31988. doi: 10.1038/srep31988

Yamada, Y., Shimada, T., Motomura, Y., and Sato, F. (2017). Modulation of benzylisoquinoline alkaloid biosynthesis by heterologous expression of CjWRKY1 in *Eschscholzia californica* cells. *PLoS One* 12:e0186953. doi: 10.1371/journal.pone.0186953

Yamada, Y., Yoshimoto, T., Yoshida, S. T., and Sato, F. (2016). Characterization of the promoter region of biosynthetic enzyme genes involved in berberine biosynthesis in *Coptis japonica*. *Front. Plant Sci.* 7:1352. doi: 10.3389/fpls.2016.01352

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci.* 18, 267–276. doi: 10.1016/j.tplants.2012.09.001

Yamasaki, K., Kigawa, T., Watanabe, S., Inoue, M., Yamasaki, T., Seki, M., et al. (2012). Structural Basis for sequence-specific DNA recognition by an *Arabidopsis* WRKY transcription factor. *J. Biol. Chem.* 287, 7683–7691. doi: 10.1074/jbc.M111.279844

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2021
doi: 10.3389/fpls.2021.691384

Genome-Wide Analysis of *MYB* Gene Family in Chinese Bayberry (*Morella rubra*) and Identification of Members Regulating Flavonoid Biosynthesis

*Yunlin Cao[1,2,3†], Huimin Jia[3†], Mengyun Xing[1,2,3], Rong Jin[4], Donald Grierson[1,2,3,5], Zhongshan Gao[3], Chongde Sun[1,2,3], Kunsong Chen[1,2,3], Changjie Xu[1,2,3] and Xian Li[1,2,3]**

[1] Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology, Zhejiang University, Hangzhou, China, [2] The State Agriculture Ministry Laboratory of Horticultural Plant Growth, Development and Quality Improvement, Zhejiang University, Hangzhou, China, [3] Institute of Fruit Science, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, [4] Agricultural Experiment Station, Zhejiang University, Hangzhou, China, [5] Plant and Crop Sciences Division, School of Biosciences, University of Nottingham, Loughborough, United Kingdom

OPEN ACCESS

Edited by:
Yang Zhang,
Sichuan University, China

Reviewed by:
Shouchuang Wang,
Hainan University, China
Zhichao Xu,
Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Xian Li
xianli@zju.edu.cn

† These authors have contributed equally to this work

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

Received: 06 April 2021
Accepted: 31 May 2021
Published: 24 June 2021

Citation:
Cao Y, Jia H, Xing M, Jin R, Grierson D, Gao Z, Sun C, Chen K, Xu C and Li X (2021) Genome-Wide Analysis of MYB Gene Family in Chinese Bayberry (Morella rubra) and Identification of Members Regulating Flavonoid Biosynthesis. Front. Plant Sci. 12:691384. doi: 10.3389/fpls.2021.691384

Chinese bayberry (*Morella rubra*), the most economically important fruit tree in the Myricaceae family, is a rich source of natural flavonoids. Recently the Chinese bayberry genome has been sequenced, and this provides an opportunity to investigate the organization and evolutionary characteristics of *MrMYB* genes from a whole genome view. In the present study, we performed the genome-wide analysis of *MYB* genes in Chinese bayberry and identified 174 MrMYB transcription factors (TFs), including 122 R2R3-MYBs, 43 1R-MYBs, two 3R-MYBs, one 4R-MYB, and six atypical MYBs. Collinearity analysis indicated that both syntenic and tandem duplications contributed to expansion of the *MrMYB* gene family. Analysis of transcript levels revealed the distinct expression patterns of different *MrMYB* genes, and those which may play important roles in leaf and flower development. Through phylogenetic analysis and correlation analyses, nine MrMYB TFs were selected as candidates regulating flavonoid biosynthesis. By using dual-luciferase assays, MrMYB12 was shown to trans-activate the *MrFLS1* promoter, and MrMYB39 and MrMYB58a trans-activated the *MrLAR1* promoter. In addition, overexpression of 35S:MrMYB12 caused a significant increase in flavonol contents and induced the expression of *NtCHS*, *NtF3H*, and *NtFLS* in transgenic tobacco leaves and flowers and significantly reduced anthocyanin accumulation, resulting in pale-pink or pure white flowers. This indicates that MrMYB12 redirected the flux away from anthocyanin biosynthesis resulting in higher flavonol content. The present study provides valuable information for understanding the classification, gene and motif structure, evolution and predicted functions of the *MrMYB* gene family and identifies MYBs regulating different aspects of flavonoid biosynthesis in Chinese bayberry.

Keywords: Chinese bayberry, MYB transcription factors, transcriptional regulation, anthocyanins, flavonols, proanthocyanidins, flavonoid biosynthesis

INTRODUCTION

Transcription factors (TFs) are important regulators of gene expression and are generally composed of at least a DNA-binding domain, nuclear location signal, transactivation domain, and an oligomerization site. The MYB family is widely present in all eukaryotes and is one of the largest TF families in plants. MYB proteins are characterized by a highly conserved MYB DNA-binding domain (Dubos et al., 2010). This domain usually comprises up to four imperfect repeats of 50–53 amino acids, and each repeat forms a helix-turn-helix (HTH) structure that binds to DNA and intercalates into the major groove of target DNA sequences (Jia et al., 2004). Based on number of adjacent repeats, MYB TFs can be divided into four classes: 1R-MYB (MYB-related and R3-MYB), R2R3-MYB, 3R-MYB (R1R2R3-MYB), and 4R-MYB (Dubos et al., 2010).

MYB TF families have been previously characterized in various plants, from lower plants such as *Physcomitrella patens* (Dubos et al., 2010) to horticultural plants, such as Chinese pear (*Pyrus bretschneideri*) (Cao et al., 2016) and ornamental flower, *Primulina swinglei* (Feng et al., 2020a). R2R3-MYB and 1R-MYB are the main classes of the MYB family identified. A total of 126 R2R3-MYBs and 64 1R-MYBs have been identified from Arabidopsis (*Arabidopsis thaliana*) (Chen et al., 2006; Dubos et al., 2010). MYB TFs from Arabidopsis are involved in the regulation of many plant processes, including cell fate and identity (Jakoby et al., 2008), organ development (Millar and Gubler, 2005; Silva-Navas et al., 2016), plant metabolism in response to abiotic and biotic stresses (Mehrtens et al., 2005; Dubos et al., 2010; Zhang et al., 2015a).

Recent research has paid more attention to MYB TFs regulating flavonoid metabolism, particularly those related to nutritional value or fruit quality traits. Numerous studies on the regulation of flavonoid biosynthesis have focused on anthocyanins accumulation during fruit development, and MYB TFs were identified in various plants, such as VvMYBA1 and VvMYBA2 in grape (*Vitis vinifera*) (Kobayashi et al., 2002), MdMYB1 in apple (*Malus domestica*) (Takos et al., 2006), and PpMYB10.1, PpMYB10.2, and PpMYB9 in peach (*Prunus persica*) (Rahim et al., 2014; Zhou et al., 2016). For flavonol biosynthesis, AtMYB12 was first reported as a flavonol-specific regulator (Mehrtens et al., 2005), followed by the identification of VvMYBF1 in grape (Czemmel et al., 2009, 2017), MdMYB22 in apple (Wang et al., 2017), and PpMYB15 and PpMYBF1 in peach (Cao et al., 2019). In addition, R2R3-MYB TFs regulating proanthocyanidin (PA) biosynthesis were reported in grape (Deluc et al., 2006) and apple (Wang et al., 2017). Therefore, different MYB members may play specific roles in different branches of flavonoids biosynthesis.

Chinese bayberry (*Morella rubra*), a subtropical fruit tree native to China, is a rich source of natural flavonoids such as anthocyanins, PAs, and flavonols (Yang et al., 2011; Zhang et al., 2015b). A series of investigations by our group have shown that flavonoid-rich pulp extracts of the fruit have a variety of bioactivities including anti-cancer (Sun et al., 2012a), anti-diabetes (Sun et al., 2012b; Zhang et al., 2016), and antioxidant (Zhang et al., 2015b) effects, among others. Previous studies have identified an R2R3-MYB protein, MrMYB1, which acts as a positive regulator of anthocyanin biosynthesis (Niu et al., 2010; Liu et al., 2013). However, the *MYB* genes related to flavonol and PA biosynthesis in Chinese bayberry have not yet been identified. Recently, the Chinese bayberry genome has been sequenced (Jia et al., 2019), and this platform provides an opportunity to identify the *MYB* gene family in Chinese bayberry and to characterize MYB proteins regulating flavonoid biosynthesis.

A comprehensive genome-wide identification of the Chinese bayberry *MYB* gene family was performed in the present study. A total of 174 MrMYB proteins (MrMYBs) were identified and subsequently comprehensively analyzed by phylogenetics, gene structure, identification of conserved motifs, collinearity and determination of chromosomal location. Furthermore, RNA-seq was carried out to investigate expression patterns of *MrMYB* genes in different tissues and during fruit development and *MrMYB* genes related to flavonoid biosynthesis were identified. The function of candidate MYBs in flavonol biosynthesis was examined by transactivation and transformation experiments.

MATERIALS AND METHODS

Plant Materials

All plant materials, including fruit, young leaves, and flowers of Chinese bayberry (*M. rubra* cv. Biqi, BQ) were harvested from commercial orchards in Xianju County, Zhejiang Province, China. The fruit were collected at 45 (S1), 75 (S2), 80 (S3), and 85 (S4) days after full bloom (DAFB). Fifteen fruits or approximately 15 g other tissues for each replicate were sampled and frozen in liquid nitrogen immediately after being cut into small pieces, and all samples were stored at −80°C. Three biological replicates were used for all samples.

Identification and Sequence Analysis of the *MrMYB* Gene Family

The Hidden Markov Model (HMM) profile of the MYB DNA-binding domain (PF00249) downloaded from Pfam database[1] was exploited for the identification of *MYB* genes in the Chinese bayberry genome by using the simple HMM search program of TBtools (Chen et al., 2020). The NCBI Conserved Domain Search[2] and SMART program[3] were exploited to test for the presence of the MYB domain. The sequence integrity of MrMYBs were analyzed by performing multiple sequence alignment analysis of all MrMYBs by ClustalW[4] (Chenna et al., 2003). Some MrMYBs containing incomplete MYB domains were found and their coding sequences were individually cloned into pGEM®-T Easy Vectors (Promega, Madison, WI, United States). Primers are listed in **Supplementary Table 1**. After adjusting the multiple sequence alignments manually, we identified the features of R2 and R3 domain repeats by WebLogo[5] (Crooks et al., 2004). The

[1]http://pfam.sanger.ac.uk/

[2]http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

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

[4]http://www.ebi.ac.uk/Tools/msa/clustalw2/

[5]http://weblogo.berkeley.edu/logo.cgi

isoelectric points and protein molecular weights of MrMYBs were obtained through the ExPASy proteomics server[6].

Phylogenetic Analyses and Function Predictions of MrMYBs

The protein sequences of MYB proteins from Chinese bayberry and Arabidopsis were aligned by the ClustalW program and adjusted manually, and the multiple sequence alignments were used for phylogenetic analysis. The phylogenetic tree was constructed by the neighbor-joining method of MEGA 7.0 with 1000 bootstrap replicates (Kumar et al., 2016). For the construction of the phylogenetic trees of R2R3-MYB proteins or other MYB proteins from Chinese bayberry, the same method described above was adopted. The phylogenetic trees of all MrMYBs or 21 selected MrMYBs with 30 functional flavonoid-related MYBs from other plants were constructed by the same method as above. Predictions of the biological functions of some MYB proteins were made, according to the orthology based on the aforementioned phylogenetic tree.

Gene Structure and Conserved Motif Analysis of the *MrMYB* Gene Family

To conduct the classification, GSDS 2.0[7] (Hu et al., 2015) was used to illustrate exon-intron organization of the *MrMYB* gene family. Furthermore, the Simple MEME program of TBtools was used for identification of conserved motifs in the 174 MrMYB protein sequences. The optimized parameters of MEME were employed as follows: Mode, AnyNumberOfOccurPerSeq; number of motifs to find, 10; and the optimum width of each motif, 6–60 residues. The MEME results were also visualized by TBtools software (Chen et al., 2020).

Chromosomal Location and Synteny Analysis of the *MrMYB* Gene Family

MrMYB genes were located on Chinese bayberry chromosomes according to their positions given in annotated documents of the Chinese bayberry genome using the MapChart software (Voorrips, 2002). The whole-genome sequences and annotation documents of Chinese bayberry and five other selected Rosids species were downloaded to our local server. Then the data were applied to analyze synteny relationships between each pair of Chinese bayberry chromosomes and used for interspecies synteny analyses of *MYB* genes between Chinese bayberry and the other five species using the One Step MCScanx program of TBtools (Chen et al., 2020). While tandem duplications were identified according to the custom script TD_identification[8] (Feng et al., 2020b). DnaSP v5.0 software was used to calculate the *Ks* value for tandemly and syntenically duplicated *MrMYB* genes (Librado and Rozas, 2009). The duplication pattern of the *MrMYB* gene family was visualized by the Amaizing Super Circos package of TBtools (Chen et al., 2020). The Dual Synteny Plot package of TBtools was used to exhibit interspecies synteny relationships of orthologous *MYB* genes between Chinese bayberry and the other five Rosid species.

Gene Expression Analysis Using RNA-seq

Total RNA was extracted according to Jia et al. (2019), and its quality was monitored by gel electrophoresis and A_{260}/A_{280}. Libraries for high-throughput Illumina strand-specific RNA-seq were prepared as described previously (Jia et al., 2019). The RNA-Seq data can be found with accession number PRJNA714192. The expression level of *MrMYB* genes was calculated as fragments per kilobase of exon model per million mapped fragments (FPKM). Three biological replicates for various samples were prepared. Transcript profiles for *MrMYB* genes were obtained and displayed in TBtools (Chen et al., 2020).

Flavonoid Analyses by HPLC

Flavonoids were analyzed according to Cao et al. (2019) with some modifications. Sample powder (100 mg) was sonicated in 1 ml extraction solution (50% methanol) for 30 min in the dark. After centrifugation at 13,000 rpm for 15 min, the supernatant was collected for HPLC analysis (e2695 pump, 2998 PDA detector, Waters), coupled to an octadecyl silane (ODS) C18 analytical column (4.6 × 250 mm) operated at 25°C, with an injection volume of 10 μl and flow rate of 1 ml/min. The mobile phase for HPLC consisted of 0.1% (v/v) formic acid in water (eluent A) and acetonitrile: 0.1% formic acid (1:1, v/v) (eluent B). The gradient program was as follows: 0–45 min, 23–50% of B; 45–50 min, 50–100% of B; 50–55 min, 100% of B; 55–56 min, 100–23% of B; 56–60 min, 23% of B. Flavonols, anthocyanins, and PAs were detected at 370, 520, and 280 nm, respectively. Contents of flavonoids were calculated by comparison with commercial standards, including myricetin 3-*O*-rhamnoside, quercetin 3-*O*-rutinoside, quercetin 3-*O*-galactoside, quercetin 3-*O*-glucoside, quercetin 3-*O*-rhamnoside, kaempferol 3-*O*-galactoside, kaempferol 3-*O*-glucoside, cyanidin 3-*O*-glucoside, and epigallocatechin gallate. Kaempferol 3-*O*-rutinoside was quantified as kaempferol 3-*O*-glucoside equivalents, other anthocyanins were quantified as cyanidin 3-*O*-glucoside equivalents, and PAs were quantified as epigallocatechin gallate equivalents. Flavonoids in Chinese bayberry tissues were identified by LC-MS according to Yang et al. (2011) and Zhang et al. (2015b). Total flavonols, anthocyanins and PAs contents were the sum of all detected flavonol glycosides, anthocyanins and PAs respectively.

Dual-Luciferase Assays

Dual-luciferase transactivation activity of TFs on target promoters was performed according to Cao et al. (2019). The full-length coding sequences of MrMYB candidates were individually cloned into pGreenII0029 62_SK vectors. Primers are listed in **Supplementary Table 2**. Promoters of $MrDFR1^{-1557}$, $MrFLS1^{-1705}$, $MrLAR1^{-1534}$, and $MrANR1^{-1512}$ were isolated from 'BQ' genomic DNA and cloned individually into pGreen II0800_LUC vectors. Primers are listed in **Supplementary Table 3**. All constructs were electroporated into *Agrobacterium*

[6]http://www.expasy.org/tools/

[7]http://gsds.gao-lab.org/index.php

[8]https://github.com/scbgfengchao/

tumefaciens GV3101. The bacteria were prepared in infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 150 mM acetosyringone, pH 5.6) when the optical density at 600 nm reached approximately 0.75. The culture mixtures of bacteria containing TFs (1 ml) and promoters (100 μl) were infiltrated into leaves of 4-week-old *Nicotiana benthamiana* plants. The luminescence from Firefly luciferase (LUC) and Renilla luciferase (REN) was detected by Dual-Luciferase Reporter Assay System (Promega, Madison, WI, United States) on the third day after infiltration, and six biological replicates were used. The ratios of LUC and REN were expressed as activation or repression of the promoters by the TFs.

Heterologous Transformation and Overexpression

The full-length coding sequence of *MrMYB12* was cloned into pGreenII0029 62_SK containing the cauliflower mosaic virus 35S promoter and transformed into *A. tumefaciens* GV3101. Tobacco (*Nicotiana tabacum*) transformed plants were regenerated as described by Cao et al. (2019). Kanamycin resistant plants were selected and transplanted to soil. The screening procedure was repeated for T1 generation transgenic lines. Fully extended mature leaves of 3-month-old plants (about 50 days after germination) and full-bloom stage flowers were sampled. Three plants were sampled for each transgenic line. Flavonoids were extracted with 50% methanol, analyzed by HPLC, and identified based on retention times and absorbances according to our previous study for quercetin 3-*O*-rutinoside, kaempferol 3-*O*-rutinoside, and cyanidin 3-*O*-glucoside (Cao et al., 2019).

Real-Time PCR Analysis

Real-time PCR analysis were performed according to Cao et al. (2019). Total RNA of tobacco samples was extracted by TRIzol Reagent kit (Ambion, Unites States). PCRs were performed on a Bio-Rad CFX96 instrument (Bio-Rad), and *NtEF1-α* was used as the internal control for monitoring the abundance of the mRNA. The gene-specific primers proven by melting curves and product resequencing are described in **Supplementary Table 4**. Expression of genes was calculated by $2^{-\Delta t}$.

RESULTS

Identification and Sequence Features of *MYB* Genes in Chinese Bayberry

To identify the *MYB*-encoding genes present in the Chinese bayberry genome, the HMM profile (PF00249) from the Pfam database was used as a query in the HMM search against the genome, and a local BLASTP search was performed by using whole Arabidopsis MYB protein sequences as the query. A total of 276 deduced amino acid sequences that might contain MYB or MYB-like repeats were obtained. All putative *MYB* genes were further examined by the NCBI Conserved Domain Search and SMART program for the presence of the MYB DNA-binding domains. A multiple sequence alignment of all MrMYBs was performed to check the sequence integrity of MrMYBs. Sixteen MrMYBs containing incomplete MYB domains were found, and the sequences of these MrMYBs were corrected through verification of the transcriptome database and cloning and sequence analysis. The updated GenBank numbers of these 16 MrMYBs are provided in **Supplementary Table 1**. As a result, a total of 174 MYBs were identified in the Chinese bayberry genome. The main sequence information of these MYBs is provided in **Supplementary Table 5**. A phylogenetic tree of MrMYBs was constructed by aligning the whole set of predicted MYB protein sequences from Chinese bayberry with 37 Arabidopsis MYB protein sequences. As shown in the phylogenetic tree (**Figure 1**), the MrMYBs were classified into four subfamilies named 1R-MYB (43), R2R3-MYB (122), 3R-MYB (2), and 4R-MYB (1) based on the presence of one, two, three, or four MYB repeats, respectively. Based on the genome data, five MrMYBs contained a complete MYB domain but also contained another incomplete one, and their coding sequences could not be cloned from 'BQ' cDNA library for sequence correction. Therefore, these five MYB members are classified as R2R3-MYBs based on the multiple sequence alignment of all MrMYBs.

The MYB domain is the core motif of MYB TFs and is directly involved in binding to the promoters of their target genes. To investigate conservation at specific positions in the MYB domain, sequence logos were generated by the multiple sequence alignment analysis of 122 R2R3-MYBs from Chinese bayberry. As shown in **Supplementary Figures 1A,B**, the R2 and R3 repeats contain many conserved amino acids, including the characteristic Trp (W) residues, which are recognized landmarks of the MYB domain. Three conserved Trp residues were identified in the R2 repeat. However, only Trp-81 and Trp-100 were conserved in the R3 repeat, and the first Trp at position 62 was substituted with hydrophobic residues, such as Phe (F), Ile (I), or Leu (L), which is a common phenomenon in R2R3-MYB proteins of plants. In addition to the highly conserved Trp residues, Cys-45 and Arg-48 in the R2 repeat, Leu-53 and Pro-55 in the linker region, and Glu-66 and Gly-78 in the R3 repeat were also conserved in the R2R3-MYB proteins.

The Classification, Motif Composition, and Gene Structure of the *MrMYB* Gene Family

To classify the *MrMYB* genes, two neighbor-joining phylogenetic trees were constructed by using the R2R3-MYB protein sequences or other MYB protein sequences from Chinese bayberry. Based on the support of bootstrap value > 50%, R2R3-MYB proteins from Chinese bayberry could be divided into 22 subgroups (designated M1-M22) (**Figure 2A**), and the 1R-MYB and 3R-MYB proteins could be divided into seven subgroups (designated I-VII) (**Supplementary Figure 2A**). Six MrMYBs did not fit into any subgroup, including four R2R3-MYB proteins, one 4R-MYB protein, and one atypical MYB protein.

Subsequently, ten conserved motifs were identified in the MrMYBs through the MEME program (**Supplementary Figure 3**). The MYB DNA-binding domains were represented by motifs 2, 3, 4, 5, 6, 9, and 10 (**Figures 2B** and **Supplementary Figure 2B**). Motif 10 was only present in 1R-MYB TFs, while

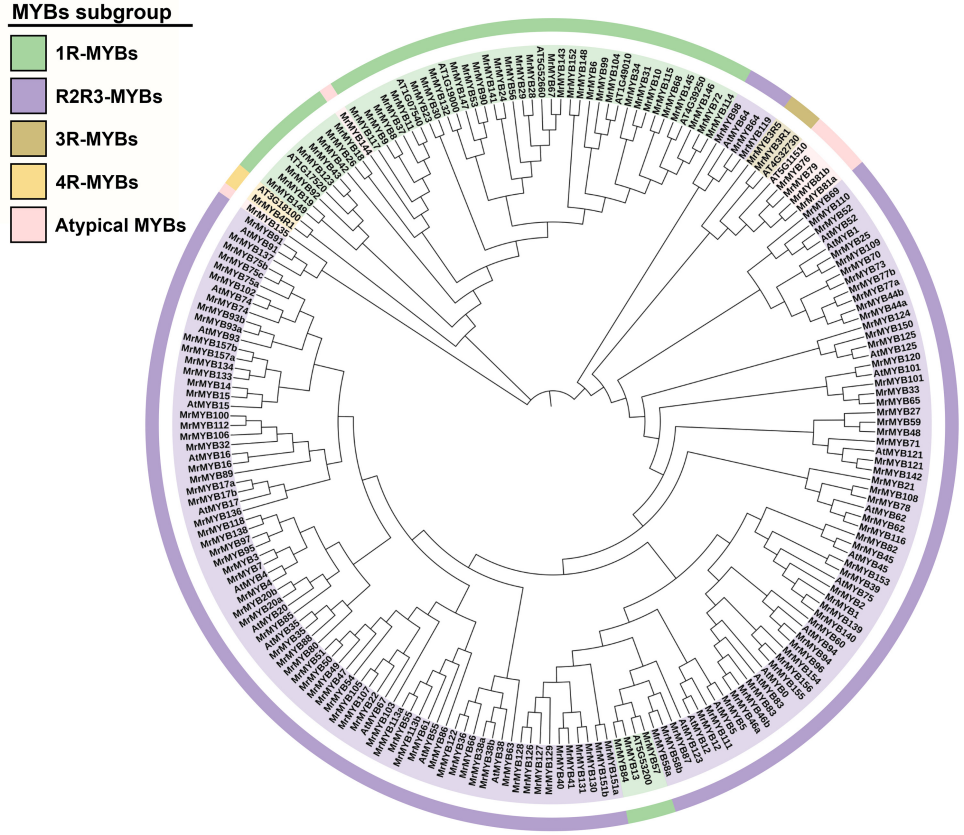


FIGURE 1 | Phylogenetic analysis of MYB proteins from Chinese bayberry and Arabidopsis. A Neighbor-joining phylogenetic tree was constructed by aligning the full-length predicted amino acid sequences of 174 MrMYBs with 37 Arabidopsis MYBs. The classes are shown in different colors.

motifs 1, 5, and 7 only appeared in R2R3-MYB TFs. These results indicated divergence of the MrMYB TFs. Since the analysis of gene structure can help understand the gene function, regulation, and evolution (Feng et al., 2016), the structure of *MrMYB* genes was also examined. As shown in **Figure 2C** and **Supplementary Figure 2C**, the number of exons in *MrMYB* genes ranged from one to 15, with an average of 3.6. Among all *MrMYB* genes, 99 *MrMYB* genes contained three exons and accounted for approximately 57% of *MrMYB* gene family members, whereas only 23% of *MrMYB* genes had more than three exons. Most *R2R3-MYB* genes clustered in related groups with similar exon-intron structures, such as M1, M2, M4, M6, etc. (**Figure 2C**). However, most *1R-MYB*, *3R-MYB,* and atypical *MYB* genes clustered in the same group with different numbers of exons, such as subgroup I-IV, VI, and VII (**Supplementary Figure 2C**).

Chromosomal Location and Synteny Analysis of the *MrMYB* Gene Family

To better understand the genomic distribution of *MrMYB* genes, their positions on each chromosome were marked. This chromosomal location analysis revealed that 158 *MrMYB* genes were unevenly distributed across all eight chromosomes and 16 *MrMYB* genes belonged to unmapped scaffolds (**Figure 3**). Chromosome 6 had the largest number (37) of *MrMYB* genes, followed by 29 on chromosome 3. In contrast, only seven *MrMYB* genes were found on chromosome 8.

Gene duplication has played a very important role in expansion of gene families (Kent et al., 2003), and high segmental and low tandem duplications were found for the *MYB* gene family in some plants (Cannon et al., 2004; Cao et al., 2016; Liu et al., 2020). We identified a similar number of *MrMYB* tandem duplications (15) and *MrMYB* syntenic duplications (16) in the Chinese bayberry genome (**Supplementary Figure 4** and **Supplementary Table 6**), indicating that both tandem and syntenic duplications contribute to the expansion of *MrMYB* genes because of a lack of recent whole-genome duplication in Chinese bayberry (Jia et al., 2019).

To further explore the evolutionary relationships of *MrMYB* genes with other species, we constructed and compared the syntenic maps of Chinese bayberry with five other Rosid species, including walnut (*Juglans regia*) (**Supplementary Figure 5A**), Chinese pear (**Supplementary Figure 5B**), peach (**Supplementary Figure 5C**), *Medicago truncatula* (**Supplementary Figure 5D**), and Arabidopsis (**Supplementary Figure 5E**). A total of 206, 153, 152, 117, and 95 homologous

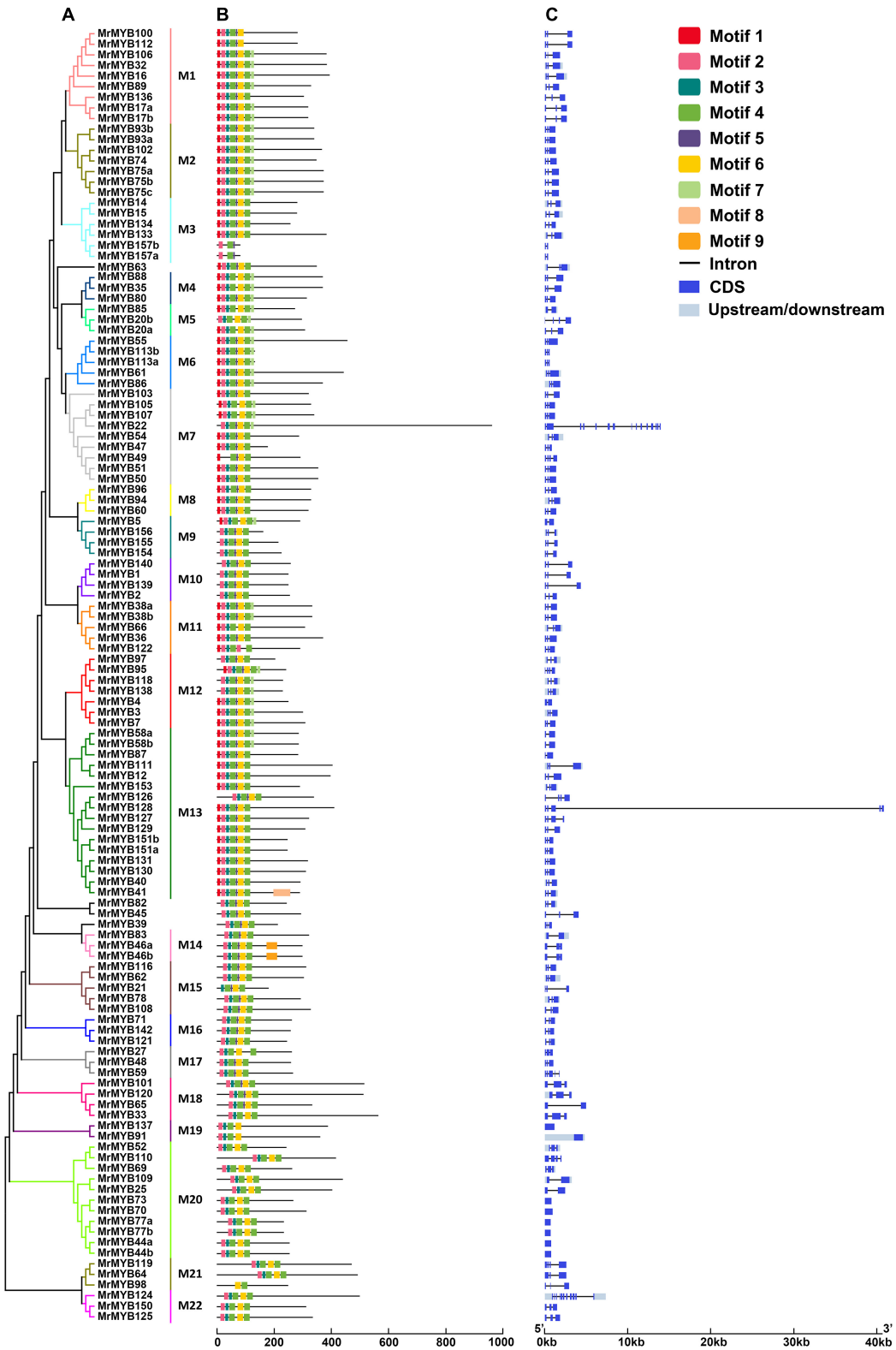


FIGURE 2 | Phylogenetic relationships **(A)**, conserved motifs **(B)**, and gene structure analysis **(C)** of Chinese bayberry *R2R3-MYB* gene family. A Neighbor-joining phylogenetic tree was constructed by aligning the full-length amino acid sequences of 122 R2R3-MYBs in Chinese bayberry. The 22 subgroups are shown with different colors. The blue boxes and black lines in the exon-intron structure diagram represent exons and introns, respectively. The ten conserved motifs are shown in different colors and their specific sequence information is provided in **Supplementary Figure 3**.

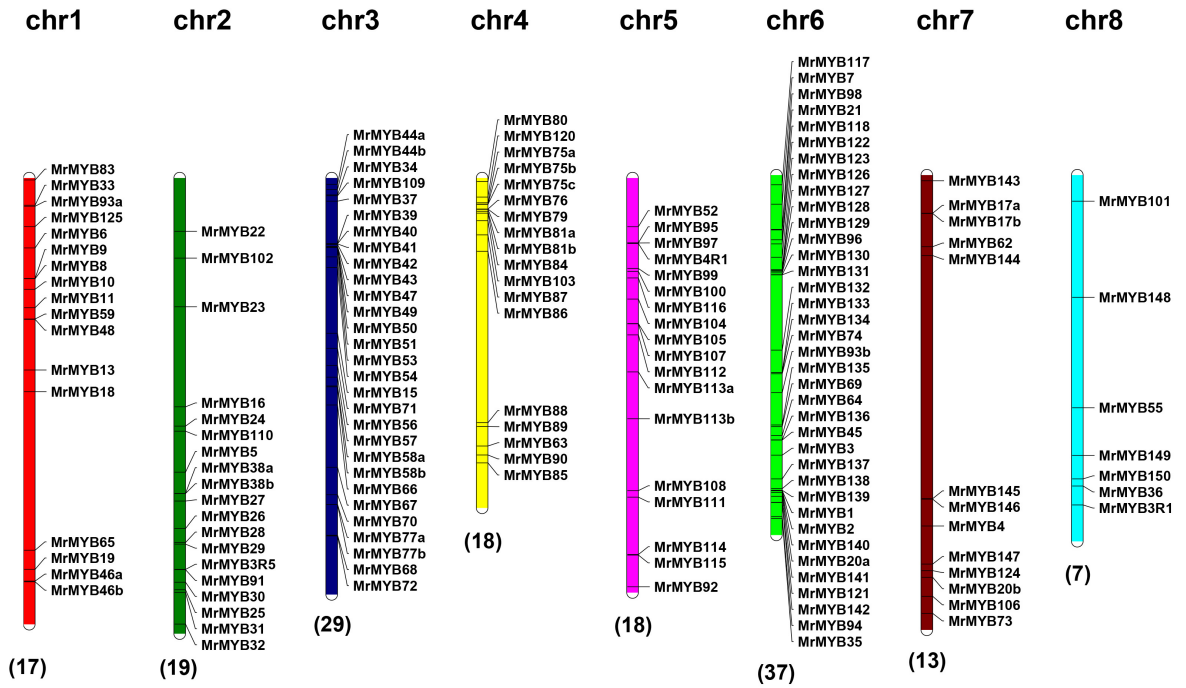


FIGURE 3 | Chromosome distribution of *MrMYB* genes. The chromosome numbers are shown at the top of each chromosome, and the number of *MrMYB* genes in each chromosome is shown at the bottom of each chromosome.

gene pairs were identified between Chinese bayberry and walnut, Chinese pear, peach, *M. truncatula*, and Arabidopsis. These findings likely reflect the closer relationship of Chinese bayberry with walnut, which supported the phylogenetic analysis results of Chinese bayberry and other sequenced species (Jia et al., 2019).

Expression Pattern of *MrMYB* Genes and in Different Tissues

RNA-seq was carried out to examine the expression pattern of 174 *MrMYB* genes in the different tissues, such as leaf, flower, and fruit. The data of transcript levels of *MrMYB* genes is shown in **Supplementary Table 7**. A total of 79, 38, and 25 *MrMYB* genes showed the highest levels of transcripts in the flower, fruit, and leaf, respectively (**Figure 4**). Among these *MrMYB* genes showing fruit-specific expression, transcript levels of 14 *MrMYB* genes were the highest at the S1 fruit stage, followed by nine *MrMYB* genes at S2 fruit stage, nine *MrMYB* genes at S3 fruit stage, and six *MrMYB* genes at S4 fruit stage. We also investigated the roles of *MrMYB* genes in regulating fruit development and ripening by analyzing RNA-seq data in the fruit developmental stages. A total of 132 *MrMYB* genes were expressed in the fruit, and 66 of these had an expression level over 1 (FPKM) at any fruit developmental stage and may be involved in regulating fruit development (**Figure 4**). Of these 132 *MrMYB* genes, transcript levels of 43 *MrMYB* genes were higher at S3 or S4 stage than any other stages, which indicates that these genes may play important roles in regulating aspects of fruit ripening.

Identification of MrMYBs Regulating Flavonoid Biosynthesis in Chinese Bayberry

Flavonoid contents in different tissues and during fruit development of Chinese bayberry were analyzed by HPLC. The results indicated that contents of flavonols and PAs were highest in 'BQ' flowers, reaching 11.78 and 3.54 mg/g fresh weight (FW), respectively (**Figure 5A**). Anthocyanins content significantly increased during fruit development and reached the highest level (1.04 mg/g FW) in the mature fruit. In contrast, the PAs level decreased during fruit developmental and flavonols content showed a reducing trend between the S1 and S3 stage but strongly increased at the S4 stage to 0.20 mg/g FW.

To identify additional *MrMYB* genes participating in the regulation of flavonoid biosynthesis in Chinese bayberry, a phylogenetic tree that included all MrMYBs and 30 functional MYBs regulating flavonoid biosynthesis from other species was generated (**Supplementary Figure 6**). Twenty-one MrMYBs of anthocyanins, flavonols, flavonoids and PAs clades were selected as candidates (**Figure 5B** and **Supplementary Figure 6**). We carried out correlation analyses between expression of *MrMYB* genes in the anthocyanins, flavonols, and PAs and flavonoids clades with contents of anthocyanins, flavonols, and PAs respectively. According to the screening criteria (correlation coefficient $r > 0.6$, $P < 0.05$), two MrMYBs were

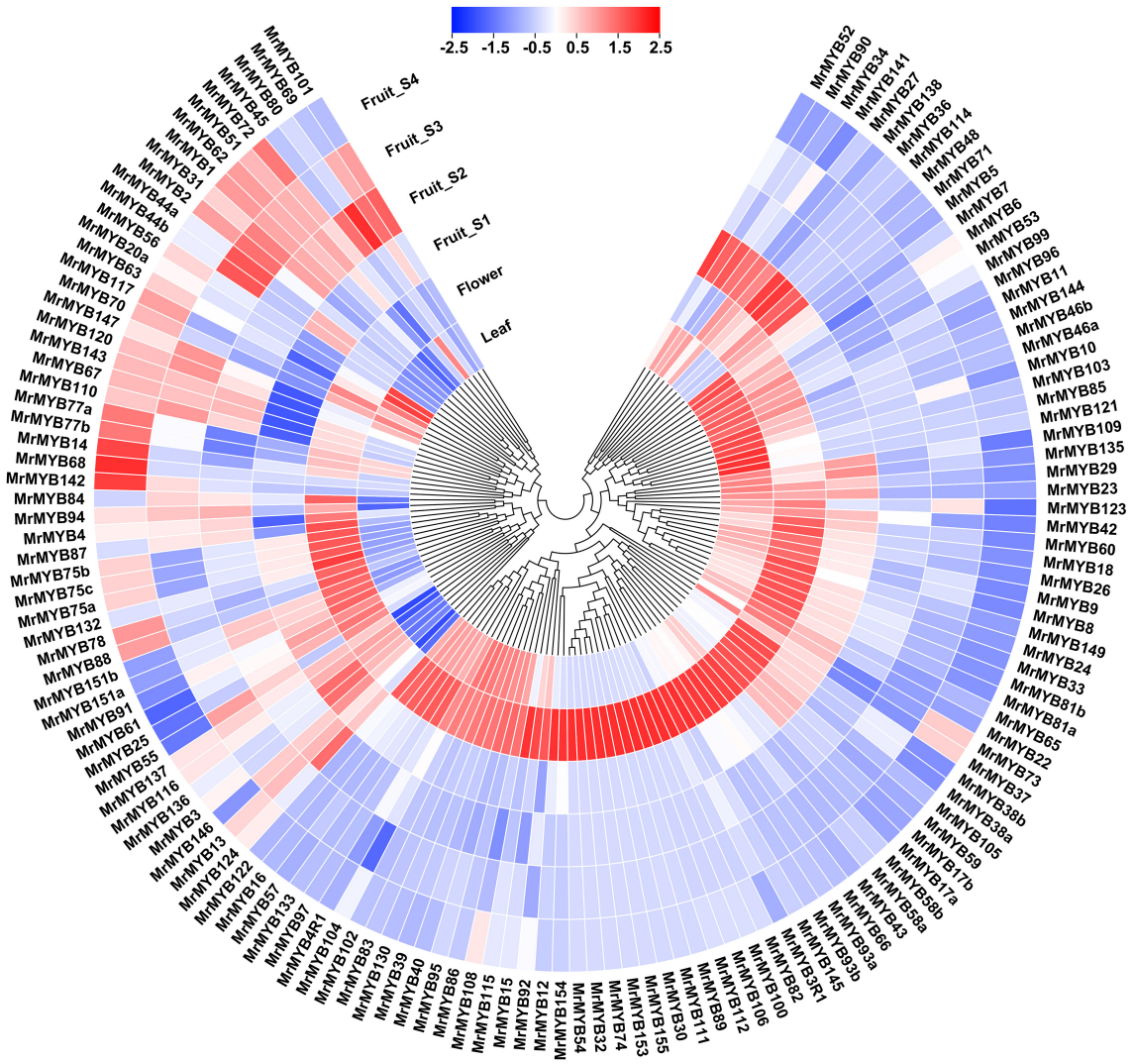


FIGURE 4 | Expression pattern of *MrMYB* genes in different tissues and during fruit development of Chinese bayberry. The expression pattern was generated based on the FPKM after log2 transformation and analyzed by heatmap hierarchical clustering. The color scale (representing −2.5 to 2.5) is shown.

selected as candidates potentially regulating the biosynthesis of anthocyanins, i.e., MrMYB1 (A1), MrMYB2 (A2), two for flavonols, i.e., MrMYB12 (F1), MrMYB111 (F2), five for PAs, i.e., MrMYB40 (P1), MrMYB39 (P2), MrMYB130 (P3), MrMYB58a/b (P4/P5), and one for flavonoids, i.e., MrMYB5 (Fd1) and used for further screening (**Supplementary Table 8**). We noted that the sequences of MrMYB58a (P4) and MrMYB58b (P5) had the same coding sequences, and therefore MrMYB58a (P4) was used for further analysis.

To verify whether the candidate TFs had the ability to regulate flavonoid biosynthesis, dual-luciferase assays were carried out in *N. benthamiana* with potential target genes. It has been well established that DFR, FLS, and LAR and ANR are the key enzymes for anthocyanin, flavonol, and PA biosynthesis, respectively. Therefore, the promoters of these genes were selected as the potential targets of TFs whose transcripts were positively correlated with anthocyanins, flavonols, and PAs accumulation. The results indicated that MrMYB12 (F1) showed a small but significant (1.4-fold) induction of the *MrFLS1* promoter, but MrMYB111 (F2) had no effect (**Figure 6A**). MrMYB1 (A1) could significantly trans-activate the *MrDFR1* promoter (greater than 17-fold induction), compared to the basal activity set as one (**Figure 6B**). This verified the previous research on the function of MrMYB1 (Niu et al., 2010; Liu et al., 2013). However, when MrMYB2 (A2) was tested with the same promoter, the transcriptional activity of the *MrDFR1* promoter only increased 1.3 times. With the *MrLAR1* promoter, only MrMYB39 (P2) (5.4-fold) and MrMYB58a (P4) (2.2-fold) could significantly transactivate the transcriptional activity (**Figure 6C**). However, none of the TFs examined showed significant regulatory effects (neither activation nor repression) on the *MrANR* promoter (**Figure 6D**).

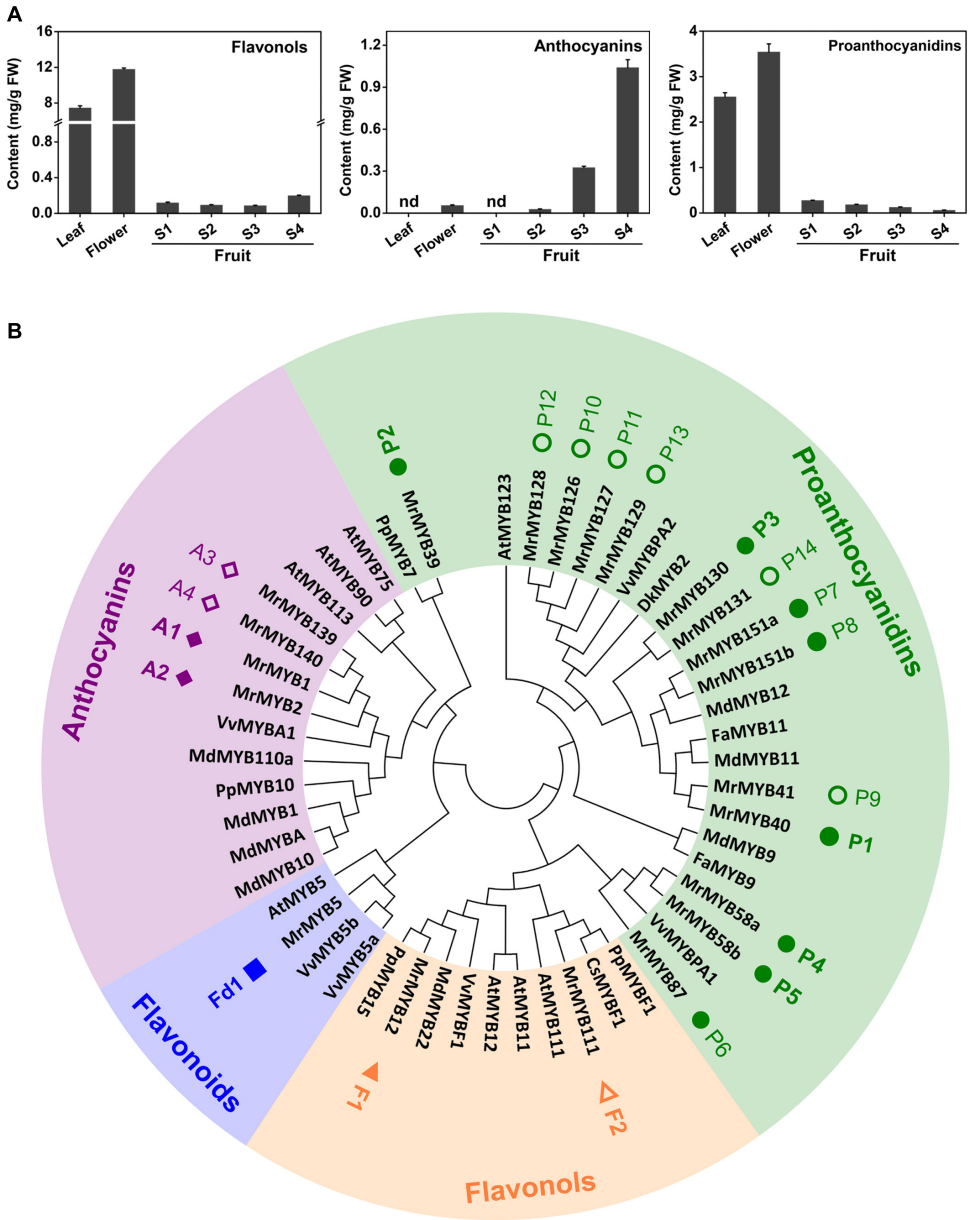


FIGURE 5 | Analysis of Chinese bayberry flavonoid contents **(A)** and phylogeny of MYBs in the anthocyanins, proanthocyanidins, flavonols, and flavonoids clades **(B)**. The square, triangles, rhombuses and circles represent MrMYBs of flavonoids, flavonols, anthocyanins, and proanthocyanidins clades, respectively. The solid symbols indicate the genes expressed in fruit, and the open symbols indicates the genes whose transcripts were not be detected in fruit. Error bars indicate S.E.s from three replicates. FW, fresh weight.

Overexpression of *MrMYB12* (F1) Increased Flavonol Accumulation and Reduced Anthocyanin Biosynthesis in Tobacco

To confirm that MrMYB12 (F1) functioned as a TF positively regulating flavonol biosynthesis, transgenic tobacco plants were generated overexpressing *MrMYB12* (F1) under the control of the CaMV 35S promoter. Two lines of T1 transgenic plants expressing 35S:MrMYB12 (F1) were used for phenotype analysis. Both transgenic lines accumulated significantly higher levels of quercetin and kaempferol in the leaves and flowers than did the WT, while the anthocyanin content in the flowers represented by the cyanidin content was much lower than WT, consistent with their phenotypic pale-pink or pure white colored flowers (**Figures 7A,B**). These results indicated that in the transgenic tobacco flowers MrMYB12 (F1) redirected the flux away from anthocyanin biosynthesis, resulting in higher flavonol content. Real-time quantitative PCR analysis

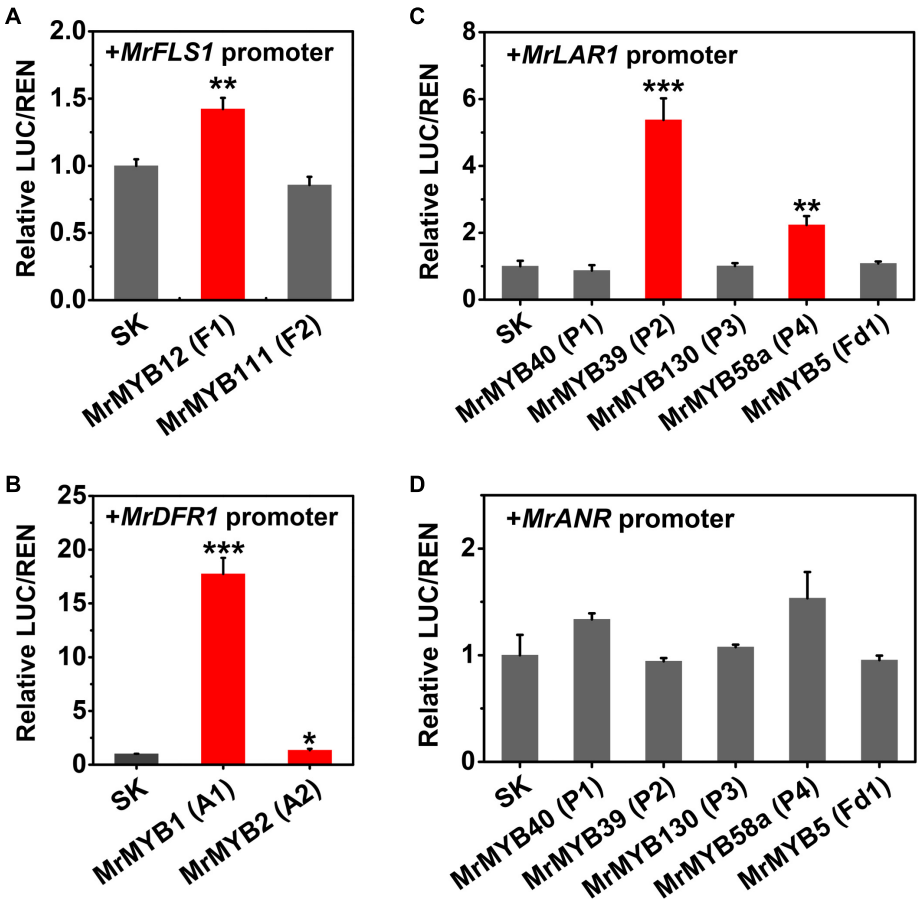


FIGURE 6 | Regulatory effects of flavonoid-related MrMYBs on the transactivation of the promoters of *MrFLS1* **(A)**, *MrDFR1* **(B)**, *MrLAR1* **(C)**, and *MrANR* **(D)**. The ratio of LUC/REN of the empty vector plus promoter was set as 1. SK refers to the empty pGreen II 0029 62-SK vector. Error bars indicate SEs from six replicates (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$).

showed that overexpression of *MrMYB12* (F1) significantly induced accumulation of *NtCHS*, *NtF3H* and *NtFLS* transcripts (**Figure 7C**). These results indicated that MrMYB12 (F1) may be a positive regulator of flavonol biosynthesis in Chinese bayberry.

DISCUSSION

Identification, Sequence Alignment, and Phylogenetic Analyses of *MrMYB* Gene Family

In the present study, 174 *MrMYB* genes were characterized in Chinese bayberry. This is a higher number of *MrMYB* genes than that in Chinese pear (129) (Cao et al., 2016), and grape (170) (Wilkins et al., 2009; Du et al., 2013) but lower than those in citrus (*Citrus sinensis*) (177) (Hou et al., 2014), Arabidopsis (198) (Chen et al., 2006) and soybean (*Glycine max*) (379) (Du et al., 2012, 2013) (**Supplementary Table 9**). This indicates that *MYB*s in different plants have expanded to different degrees. It was found that most *MrMYB* genes were not disrupted by more than two introns, which is consistent with previous studies (Du et al., 2012; Liu et al., 2020). The *R2R3-MYB* gene family from Chinese bayberry was classified into 22 subgroups based on phylogenetic analysis, with mostly similar exon-intron organizations and conserved motif compositions. This result is consistent with the previous reports in Arabidopsis, soybean, Chinese pear, and Japanese plum (*Prunus salicina*) (Chen et al., 2006; Du et al., 2012; Cao et al., 2016; Liu et al., 2020), indicating that a strong correlation exists between the phylogenetic topology and gene structures of *R2R3-MYB* genes. However, in our study, all subgroups of the *1R-MYB* gene family in Chinese bayberry consistently displayed a certain degree of divergent in intron-exon organization.

The *MrMYB* Genes Play Important Roles in Leaf and Flower Development

The combined phylogenetic tree and transcriptomic data analysis provides important information for functional predictions of *MrMYB* genes. Expression analysis by RNA-Seq was conducted in different tissues and during fruit development and ripening in order to investigate the function of *MrMYB*

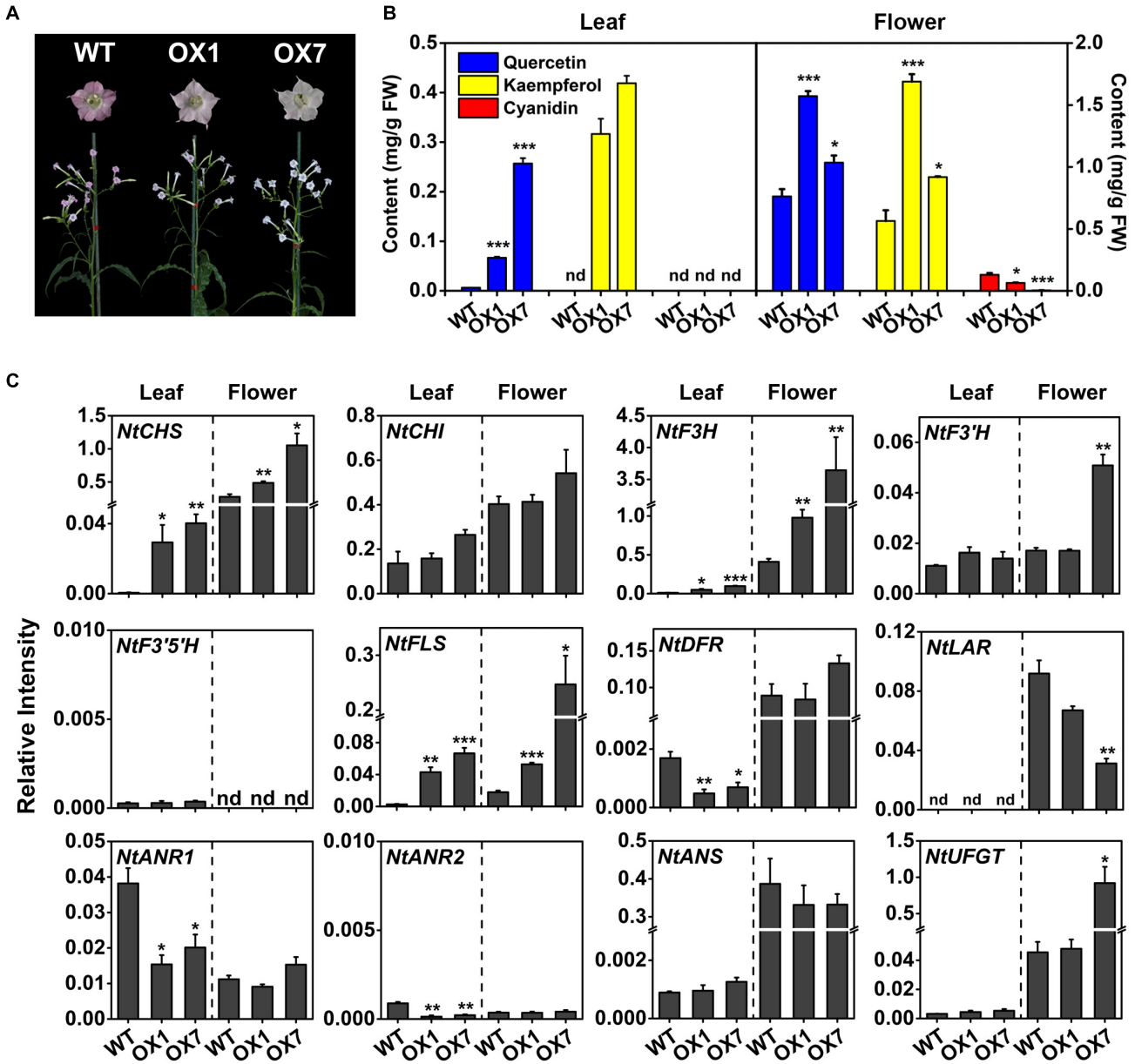


FIGURE 7 | Ectopic expression of *MrMYB12* (F1) genes in tobacco. **(A)** Phenotypes of 35S:MrMYB12 transgenic and wild type (WT) tobacco flowers at the full bloom stage. **(B)** Flavonoid contents of transgenic and WT tobacco flowers and leaves. **(C)** Transcript levels of genes involved in flavonoid biosynthesis in tobacco leaves and flowers overexpressing *MrMYB12* (F1). OX1 and OX7 represent line 1 and line 7 of T1 transgenic plants overexpressing 35S:MrMYB12 (F1), respectively. Student's *t*-test was used for statistical analyses compared with WT (**P* < 0.05, ***P* < 0.01, and ****P* < 0.001). Error bars indicate SEs from three replicates. FW, fresh weight; ND, not detectable.

genes. *MrMYB16* was preferentially expressed in flowers and clustered together with AtMYB16 (**Figure 1**), which contributes to the formation of petal epidermal cells (Baumann et al., 2007), suggesting MrMYB16 may share similar functions in the regulation of petal development. A previous study reported that AtMYB17 may be involved in the regulation of early inflorescence development in Arabidopsis (Zhang et al., 2009), and its homologous genes in Chinese bayberry are *MrMYB17a* and *MrMYB17b*, which are preferentially expressed in the flowers (**Figure 4**) and thus may have a similar function to AtMYB17. *MrMYB20a/b* and *MrMYB85* were preferentially expressed in leaves (**Figure 4**) and had a close relationship with AtMYB20 (**Figure 1**), which participates in regulating lignin and phenylalanine biosynthesis during secondary cell wall formation in Arabidopsis (Geng et al., 2020). This indicates that MrMYB20a/b and MrMYB85 may be involved in the regulation of lignin and phenylalanine biosynthesis in the leaf tissue.

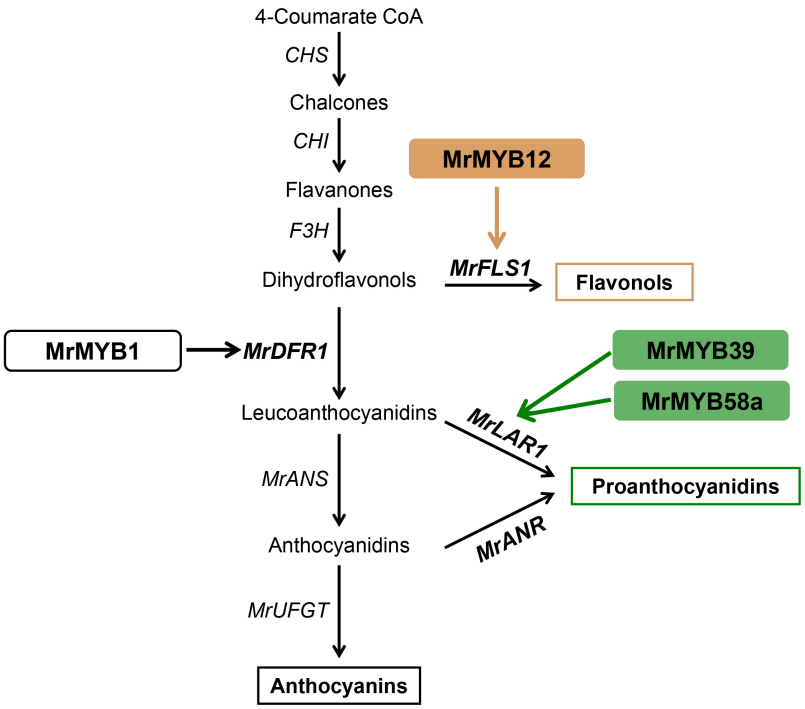


FIGURE 8 | A proposed model for MYB-regulated flavonoid biosynthesis in Chinese bayberry. MrMYB12 activates the expression of *MrFLS1* to regulate flavonol accumulation. The brown and blue arrows indicate the pathways verified in the present work. The black arrows indicate the pathways that have been previously reported in Chinese bayberry. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone-3-*O*-hydroxylase; DFR, dihydroflavonol-4-reductase; FLS, flavonol synthase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; UFGT, UDP-glucose:flavonoid 3-*O*-glucosyltransferase.

MrMYB TFs Are Involved in the Regulation of Anthocyanin and PA Biosynthesis in Fruit

Fresh fruits contain a wide range of health-promoting compounds and their regular consumption is one important way to contribute to a healthy diet. Flavonoids are one of the best-accepted health-promoting compounds in fruits and increasing reports have shown that MYB proteins from fruit species are involved in the transcriptional regulation of flavonoid biosynthesis (Falcone Ferreyra et al., 2012; Liu et al., 2015). However, there has been only limited research about the transcriptional regulation of the flavonoid metabolism in Chinese bayberry. Anthocyanins function as pigments and anthocyanin accumulation is one key determinant of fruit color, an important fruit quality attribute. It was found that four MrMYBs were homologous to and clustered with several functional regulators of anthocyanin biosynthesis, such as VvMYBA1 and VvMYBA2 from grape (Kobayashi et al., 2002), MdMYB1 from apple (Takos et al., 2006). Moreover, expression analysis showed that only *MrMYB1* and *MrMYB2* were expressed in any one tissue and transcript levels of these two genes increased with fruit development and ripening, which is consistent with the anthocyanin accumulation pattern in the fruit of Chinese bayberry (Niu et al., 2010; Liu et al., 2013). Dual-luciferase assays in *N. benthamiana* leaf showed that MrMYB1 could significantly trans-activate the *MrDFR1* promoter, which validates the function of MrMYB1 reported by Niu et al. (2010) and Liu et al. (2013). However, MrMYB2 only induced the transcriptional activity 1.3-fold, indicating that MrMYB1 is the more important regulator of anthocyanin biosynthesis in Chinese bayberry. Further study can use controlled crossing breeding materials (Wang et al., 2020).

PAs are distributed widely in the leaves and fruit of Chinese bayberry and have been associated with health-promoting benefits. A previous study has functionally characterized two key genes of PA biosynthesis, *MrLAR1* and *MrANR*, but the mechanism regulating PA biosynthesis remains unclear. We found (**Figure 5B**) that 14 MrMYBs were clustered with the PA clade of the MYB family, and MrMYB5 (Fd1) in the flavonoid clade was homologous with VvMYB5a and VvMYB5b which are known to be involved in regulating PA biosynthesis (Deluc et al., 2006, 2008). Among five MrMYBs screened by the correlation analyses, only MrMYB39 (P2) and MrMYB58a (P4) significantly activated the promoter of *MrLAR1* but did not activate that of *MrANR* (**Figures 6C,D**). Similar results were found in apple MYB12 and peach MYB7, which regulate the biosynthesis of catechin but not epicatechin (Zhou et al., 2015; Wang et al., 2017). Different results were obtained previously with MdMYB9, VvMYBPA1, and VvMYBPA2, which could regulate the expression of *LAR* and *ANR* to promote the accumulation of catechin and epicatechin (Bogs et al., 2007; Terrier et al., 2009; An et al., 2015), and small differences in the amino acid sequences of these proteins may account for this.

Therefore, the biosynthesis of catechin and epicatechin may be regulated by different MYB TFs. It is clear that MrMYB39 (P2) and MrMYB58a (P4) may function as positive regulators of flavonoid biosynthesis by regulating the transcription of *MrLAR1* (**Figure 8**).

MrMYB12 (F1) Function as A Flavonol-Specific Regulator

Flavonols, a class of colorless flavonoids, are important health-related compounds in the human diet. Previous studies reported that the regulation of flavonol biosynthesis is usually controlled by the SG7 subgroup of the MYB family (Mehrtens et al., 2005; Stracke et al., 2007). Two MrMYBs, MrMYB12 (F1) and MrMYB111 (F2), were clustered in the flavonol clade with the functional flavonol regulators from other plants and expression levels of these two genes were highly correlated with the flavonols content. Dual-luciferase assays *in vivo* indicated that MrMYB12 (F1) trans-activated the *MrFLS1* gene promoter (**Figure 6A**) as does its homologs, AtMYB12, VvMYBF1, MdMYB22, and PpMYB15 (Mehrtens et al., 2005; Czemmel et al., 2009; Wang et al., 2017; Cao et al., 2019). However, MrMYB111, a homolog of PpMYBF1, failed to activate the *MrFLS1* gene promoter. Previously, over-expression of *AtMYB12* resulted in unprecedentedly high levels of kaempferol or quercetin accumulation in both tobacco and tomato, and lower anthocyanin levels (Luo et al., 2008). Accordingly, the content of kaempferol or quercetin was reduced significantly in *Slmyb12* or *Atmyb12* mutants (Mehrtens et al., 2005; Ballester et al., 2010). Consistent with these reports, our data show that over-expression of 35S:MrMYB12 (F1) in tobacco promoted kaempferol or quercetin accumulation and decreased anthocyanin accumulation by upregulating the transcript levels of *NtCHS*, *NtF3H* and *NtFLS*. Therefore, MrMYB12 (F1) may act as a flavonol-specific regulator by redirecting the flux from anthocyanin biosynthesis to flavonol biosynthesis (**Figure 8**).

CONCLUSION

Genome-wide analysis of phylogenetic relationships, gene structures, motif compositions, chromosomal locations, evolutionary relationships, and expression of *MrMYB* genes, was carried out in the present study. A total of 174 MYB family members from Chinese bayberry were identified. Intraspecies synteny analysis indicated that both dispersed syntenic and tandem duplications contributed to expansion of the *MrMYB* gene family. Expression analysis revealed that *MrMYB* genes had tissue-specific expression patterns in leaf, flower and fruit, and some were identified as likely to have important roles in leaf and flower development, consistent with the functional predictions from phylogenetic analysis. Through the combination of phylogenetic analysis and correlation analyses, nine MrMYB TFs were selected as candidates associated with flavonoid biosynthesis. Of these candidates, MrMYB12 trans-activated the *MrFLS1* promoter, and MrMYB39 and MrMYB58a activated the *MrLAR1* promoter. In addition, heterologous overexpression of 35S:MrMYB12 increased flavonol levels and induced the expression of *NtCHS*, *NtF3H*, and *NtFLS* in transgenic tobacco leaves and flowers and significantly reduced anthocyanin accumulation, resulting in pale-pink or pure white flowers. Overall, these results provide information that will facilitate further functional analyses of *MrMYB* genes to elucidate their biological roles. The functional identification of different MYBs regulating flavonoid biosynthesis will help to improve the fruit quality of Chinese bayberry in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: RNA-Seq data can be found with accession number PRJNA714192. The RNA-Seq data is publicly available on National Center for Biotechnology Information.

AUTHOR CONTRIBUTIONS

XL and YC designed the project and drafted the manuscript. YC managed the experiments with help from HJ, MX, and RJ. CX and KC participated in design of the study and provided support for the Morella project. CX, DG, CS, and ZG contributed to the discussion and revision of the manuscript. All authors approved the article.

FUNDING

This work was financially supported by the Programs for the Natural Science Foundation of Zhejiang Province (Z17C150003), the National Natural Science Foundation of China (31872067 and 31972364), the Key R&D Program of Zhejiang Province (2021C0201), and the 111 project (B17039).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.691384/full#supplementary-material

Supplementary Figure 1 | The sequence logos of the R2 **(A)** and R3 **(B)** MYB repeats. These logos were based on the multiple sequence alignment of 122 R2R3-MYBs in Chinese bayberry. The core bits indicate the information content for each position in the sequence. The asterisks indicate the typical conserved Trp residues in the MYB domain. The triangle indicates the residue in the R3 repeat with positions identical to the first conserved Trp residue in the R2 repeat.

Supplementary Figure 2 | Phylogenetic relationships **(A)**, conserved motifs **(B)**, and gene structure analysis **(C)** in Chinese bayberry 1R-, 3R-, and 4R-MYBs. A Neighbor-joining phylogenetic tree was constructed by aligning the full-length amino acid sequences of 1R-, 3R-, and 4R-MYBs. The seven subgroups are shown in different colors. The blue boxes and black lines in the exon-intron structure diagram represent exons and introns, respectively. The ten conserved motifs are exhibited with different colors and their specific sequence information is provided in **Supplementary Figure 3**.

Supplementary Figure 3 | All MEME motif sequence logos in MrMYBs.

Supplementary Figure 4 | Schematic representations for the interchromosomal relationships of *MrMYB* genes. **(A)** Gray lines in the background mean collinear blocks within Chinese bayberry, red lines indicate syntenic *MYB* gene pairs. **(B)** Duplicated gene pairs of *MrMYB* genes are sorted according to their assigned MYB classes.

Supplementary Figure 5 | Syntenic analysis of *MYB* genes between Chinese bayberry and *Juglans regia* **(A)**, *Pyrus bretschneideri* **(B)**, *Prunus persica* **(C)**, *Medicago truncatula* **(D)**, or *Arabidopsis thaliana* **(E)**. Gray lines in the background mean collinear blocks between Chinese bayberry and other plant genomes, and red lines indicate syntenic *MYB* gene pairs.

Supplementary Figure 6 | Phylogenetic analysis of all MrMYBs and functional flavonoid-related MYB proteins from other plants. The clades are shown in different colors.

Supplementary Table 1 | Primers for *MrMYB* gene amplification and T-easy vector constructions.

Supplementary Table 2 | Primers for promoter amplification and pGreenII0029 62_SK vector constructions. Sequences of restriction sites are underlined.

Supplementary Table 3 | Primers for promoter amplification and pGreen II0800_LUC vector constructions. Sequences of restriction sites are underlined.

Supplementary Table 4 | Primers used for quantitative real-time PCR analysis.

Supplementary Table 5 | The isoelectric point, molecular weight, chromosome location, and MYB-domain type of the members of *MrMYB* gene family. The principles of gene naming are as follows: (1) Gene names of the *MrMYB* locus have been reported; (2) *MrMYB* genes were named based on their homologous genes in Arabidopsis; (3) without meeting the first and second principles, *MrMYB* genes were named according to the order of chromosome location.

Supplementary Table 6 | The *Ks* value for tandemly and syntenically duplicated *MrMYB* genes.

Supplementary Table 7 | Transcript levels of *MrMYB* genes in different tissues and during fruit development. L, leaf; Fl, flower; Fr, fruit.

Supplementary Table 8 | Correlation between flavonoid-related *MrMYB* genes with flavonols, anthocyanins or proanthocyanidins contents. Expression patterns of the *MrMYB* genes and flavonoid profiles in different tissues and during fruit development are shown in **Figures 4**, **5**.

Supplementary Table 9 | Numbers of plant MYBs in the four different classes.

REFERENCES

An, X. H., Tian, Y., Chen, K. Q., Liu, X. J., Liu, D. D., Xie, X. B., et al. (2015). MdMYB9 and MdMYB11 are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. *Plant Cell Physiol.* 56, 650–662. doi: 10.1093/pcp/pcu205

Ballester, A. R., Molthoff, J., de Vos, R., Hekkert, B. T. L., Orzaez, D., Fernández-Moreno, J. P., et al. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiol.* 152, 71–84. doi: 10.1104/pp.109.147322

Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., et al. (2007). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* 134, 1691–1701. doi: 10.1242/dev.02836

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

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi: 10.1186/1471-2229-4-10

Cao, Y. L., Xie, L. F., Ma, Y. Y., Ren, C. H., Xing, M. Y., Fu, Z. S., et al. (2019). PpMYB15 and PpMYBF1 transcription factors are involved in regulating flavonol biosynthesis in peach fruit. *J. Agric. Food Chem.* 67, 644–652. doi: 10.1021/acs.jafc.8b04810

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

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., 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. H., Yang, X. Y., He, K., Liu, M. H., Li, J. G., Gao, Z. F., et al. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* 60, 107–124. doi: 10.1007/s11103-005-2910-y

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31, 3497–3500. doi: 10.1093/nar/gkg500

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

Czemmel, S., Höll, J., Loyola, R., Arce-Johnson, P., Alcalde, J. A., Matus, J. T., et al. (2017). Transcriptome-wide identification of novel uv-b- and light modulated flavonol pathway genes controlled by VviMYBF1. *Front. Plant Sci.* 8:1084. doi: 10.3389/fpls.2017.01084

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., et al. (2009). The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* 151, 1513–1530. doi: 10.1104/pp.109.142059

Deluc, L., Barrieu, F., Marchive, C., Lauvergeat, V., Decendit, A., Richard, T., et al. (2006). Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiol.* 140, 499–511. doi: 10.1104/pp.105.067231

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

Du, H., Wang, Y. B., Xie, Y., Liang, Z., Jiang, S. J., Zhang, S. S., et al. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res.* 20, 437–448. doi: 10.1093/dnares/dst021

Du, H., Yang, S. S., Liang, Z., Feng, B. R., Liu, L., Huang, Y. B., et al. (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 12:106. doi: 10.1186/1471-2229-12-106

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222

Feng, C., Ding, D. H., Feng, C., and Kang, M. (2020a). The identification of an R2R3-MYB transcription factor involved in regulating anthocyanin biosynthesis in *Primulina swinglei* flowers. *Gene* 752:144788. doi: 10.1016/j.gene.2020.144788

Feng, C., Wang, J., Wu, L. Q., Kong, H. H., Yang, L. H., Feng, C., et al. (2020b). The genome of a cave plant, Primulina huaijiensis, provides insights into adaptation to limestone karst habitats. *New Phytol.* 227, 1249–1263. doi: 10.1111/nph.16588

Feng, K. W., Liu, F. Y., Zou, J. W., Xing, G. W., Deng, P. C., Song, W. N., et al. (2016). Genome-wide identification, evolution, and co-expression network analysis of mitogen-activated protein kinase kinase kinases in *Brachypodium distachyon*. *Front. Plant Sci.* 7:1400. doi: 10.3389/fpls.2016.01400

Geng, P., Zhang, S., Liu, J. Y., Zhao, C. H., Wu, J., Cao, Y. P., et al. (2020). MYB20, MYB42, MYB43 and MYB85 regulate phenylalanine and lignin biosynthesis during secondary Cell Wall formation. *Plant Physiol.* 182, 1272–1283. doi: 10.1104/pp.19.01070

Hou, X. J., Li, S. B., Liu, S. R., Hu, C. G., and Zhang, J. Z. (2014). Genome-wide classification and evolutionary and expression analyses of citrus MYB

transcription factor families in sweet orange. *PLoS One* 9:e112375. doi: 10.1371/journal.pone.0112375

Hu, B., Jin, J. P., Guo, A. Y., Zhang, H., Luo, J. C., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Jakoby, M. J., Falkenhan, D., Mader, M. T., Brininstool, G., Wischnitzki, E., Platz, N., et al. (2008). Transcriptional profiling of mature Arabidopsis trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiol.* 148, 1583–1602. doi: 10.1104/pp.108.126979

Jia, H. M., Jia, H. J., Cai, Q. L., Wang, Y., Zhao, H. B., Yang, W. F., et al. (2019). The red bayberry genome and genetic basis of sex determination. *Plant Biotechnol. J.* 17, 397–409. doi: 10.1111/pbi.12985

Jia, L., Clegg, M. T., and Jiang, T. (2004). Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.* 134, 575–585. doi: 10.1104/pp.103.027201

Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11484–11489. doi: 10.1073/pnas.1932072100

Kobayashi, S., Ishimaru, M., Hiraoka, K., and Honda, C. (2002). Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta* 215, 924–933. doi: 10.1007/s00425-002-0830-5

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

Librado, P., and Rozas, J. (2009). DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Liu, C. Y., Hao, J. J., Qiu, M. Q., Pan, J. J., and He, Y. H. (2020). Genome-wide identification and expression analysis of the MYB transcription factor in Japanese plum (*Prunus salicina*). *Genomics* 112, 4875–4886. doi: 10.1016/j.ygeno.2020.08.018

Liu, J. Y., Osbourn, A., and Ma, P. D. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012

Liu, X. F., Yin, X. R., Allan, A. C., Lin-Wang, K., Shi, Y. N., Huang, Y. J., et al. (2013). The role of MrbHLH1 and MrMYB1 in regulating anthocyanin biosynthetic genes in tobacco and Chinese bayberry (*Myrica rubra*) during anthocyanin biosynthesis. *Plant Cell Tissue Organ. Cult.* 115, 285–298. doi: 10.1007/s11240-013-0361-8

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., et al. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. *Plant J.* 56, 316–326. doi: 10.1111/j.1365-313X.2008.03597.x

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096. doi: 10.1104/pp.104.058032

Millar, A. A., and Gubler, F. (2005). The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17, 705–721. doi: 10.1105/tpc.104.027920

Niu, S. S., Xu, C. J., Zhang, W. S., Zhang, B., Li, X., Lin-Wang, K., et al. (2010). Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta* 231, 887–899. doi: 10.1007/s00425-009-1095-z

Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 240, 913–929. doi: 10.1007/s00425-014-2078-2

Silva-Navas, J., Moreno-Risueno, M. A., Manzano, C., Téllez-Robledo, B., Navarro-Neila, S., Carrasco, V., et al. (2016). Flavonols mediate root phototropism and growth through regulation of proliferation-to-differentiation transition. *Plant Cell* 28, 1372–1387. doi: 10.1105/tpc.15.00857

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

Sun, C. D., Zhang, B., Zhang, J. K., Xu, C. J., Wu, Y. L., Li, X., et al. (2012b). Cyanidin-3-glucoside-rich extract from Chinese bayberry fruit protects pancreatic β cells and ameliorates hyperglycemia in streptozotocin-induced diabetic mice. *J. Med. Food* 15, 288–298. doi: 10.1089/jmf.2011.1806

Sun, C. D., Zheng, Y. X., Chen, Q. J., Tang, X. L., Jiang, M., Zhang, J. K., et al. (2012a). Purification and anti-tumour activity of cyanidin-3-O-glucoside from Chinese bayberry fruit. *Food Chem.* 131, 1287–1294. doi: 10.1016/j.foodchem.2011.09.121

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

Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verriès, C., Cheynier, V., et al. (2009). Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiol.* 149, 1028–1041. doi: 10.1104/pp.108.131862

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

Wang, N., Xu, H. F., Jiang, S. H., Zhang, Z. Y., Lu, N. L., Qiu, H. R., et al. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. niedzwetzkyana). *Plant J.* 90, 276–292. doi: 10.1111/tpj.13487

Wang, Y., Jia, H. M., Shen, Y. T., Zhao, H. B., Yang, Q. S., Zhu, C. Q., et al. (2020). Construction of an anchoring SSR marker genetic linkage map and detection of a sex-linked region in two dioecious populations of red bayberry. *Hortic. Res.* 7:53. doi: 10.1038/s41438-020-0276-6

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., and Campbell, M. M. (2009). Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* 149, 981–993. doi: 10.1104/pp.108.132795

Yang, H. H., Ge, Y. Q., Sun, Y. J., Liu, D. H., Ye, X. Q., and Wu, D. (2011). Identification and characterisation of low-molecular-weight phenolic compounds in bayberry (*Myrica rubra* Sieb. et Zucc.) leaves by HPLC-DAD and HPLC-UV-ESIMS. *Food Chem.* 128, 1128–1135. doi: 10.1016/j.foodchem.2011.03.118

Zhang, X. N., Huang, H. Z., Zhang, Q. L., Fan, F. J., Xu, C. J., Sun, C. D., et al. (2015b). Phytochemical characterization of chinese bayberry (Myrica rubra Sieb. et Zucc.) of 17 cultivars and their antioxidant properties. *Int. J. Mol. Sci.* 16, 12467–12481. doi: 10.3390/ijms160612467

Zhang, X. N., Lv, Q., Jia, S., Chen, Y. H., Sun, C. D., Li, X., et al. (2016). Effects of flavonoid-rich Chinese bayberry (*Morella rubra* Sieb. et Zucc.) fruit extract on regulating glucose and lipid metabolism in diabetic KK-A(y) mice. *Food Funct.* 7, 3130–3140. doi: 10.1039/c6fo00397d

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., et al. (2015a). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* 6:8635. doi: 10.1038/ncomms9635

Zhang, Y. F., Cao, G. Y., Qu, L. J., and Gu, H. Y. (2009). Characterization of Arabidopsis MYB transcription factor gene AtMYB17 and its possible regulation by LEAFY and AGL15. *J. Genet. Genom.* 36, 99–107. doi: 10.1016/S1673-8527(08)60096-X

Zhou, H., Kui, L. W., Liao, L., Chao, G., Lu, Z. Q., Allan, A. C., et al. (2015). Peach MYB7 activates transcription of the proanthocyanidin pathway gene encoding leucoanthocyanidin reductase, but not anthocyanidin reductase. *Front. Plant Sci.* 6:908. doi: 10.3389/fpls.2015.00908

Zhou, H., Peng, Q., Zhao, J., Owiti, A., Ren, F., Liao, L., et al. (2016). Multiple R2R3-MYB transcription factors involved in the regulation of anthocyanin accumulation in peach flower. *Front Plant Sci.* 7:1557. doi: 10.3389/fpls.2016.01557

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

frontiers in Plant Science

REVIEW
published: 02 July 2021
doi: 10.3389/fpls.2021.682181

Cytochrome P450 Enzymes as Key Drivers of Alkaloid Chemical Diversification in Plants

*Trinh-Don Nguyen and Thu-Thuy T. Dang**

Department of Chemistry, Irving K. Barber Faculty of Science, University of British Columbia, Kelowna, BC, Canada

Plants produce more than 20,000 nitrogen-containing heterocyclic metabolites called alkaloids. These chemicals serve numerous eco-physiological functions in the plants as well as medicines and psychedelic drugs for human for thousands of years, with the anti-cancer agent vinblastine and the painkiller morphine as the best-known examples. Cytochrome P450 monooxygenases (P450s) play a key role in generating the structural variety that underlies this functional diversity of alkaloids. Most alkaloid molecules are heavily oxygenated thanks to P450 enzymes' activities. Moreover, the formation and re-arrangement of alkaloid scaffolds such as ring formation, expansion, and breakage that contribute to their structural diversity and bioactivity are mainly catalyzed by P450s. The fast-expanding genomics and transcriptomics databases of plants have accelerated the investigation of alkaloid metabolism and many players behind the complexity and uniqueness of alkaloid biosynthetic pathways. Here we discuss recent discoveries of P450s involved in the chemical diversification of alkaloids and how these inform our approaches in understanding plant evolution and producing plant-derived drugs.

Keywords: alkaloid, catalysis, scaffold, diversification, oxidation, medicinal plants, P450

OPEN ACCESS

Edited by:
Danièle Werck,
UPR2357 Institut de Biologie Moléculaire des Plantes (IBMP), France

Reviewed by:
Yang Qu,
University of New Brunswick Fredericton, Canada
Amit Rai,
Chiba University, Japan

***Correspondence:**
Thu-Thuy T. Dang
thuy.dang@ubc.ca

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

Received: *17 March 2021*
Accepted: *01 June 2021*
Published: *02 July 2021*

Citation:
Nguyen TD and Dang TTT (2021) Cytochrome P450 Enzymes as Key Drivers of Alkaloid Chemical Diversification in Plants. Front. Plant Sci. 12:682181. doi: 10.3389/fpls.2021.682181

INTRODUCTION

Alkaloids – A Functionally and Structurally Diverse Natural Product Class With Unique Underlying Biosynthesis

Chemical diversity is the key to success for the sessile lifestyle that plants have evolved to adapt. Over hundreds of millions of years, land plants have accumulated a formidable capacity to biosynthesize numerous small molecules, often referred to as natural products or specialized metabolites, that help them thrive in specific environmental niches. Among plant natural products, alkaloids constitute arguably the most intriguing class with thousands-of-years interconnection with human history. Alkaloids have long been used and abused for their potent therapeutic properties and notorious toxic and psychedelic effects with significant geopolitical impacts, as seen in the Anglo-Chinese opium wars of the 19th century and the ongoing war on drugs (Kutchan et al., 2015). Moreover, alkaloid diversity has attracted much attention from chemists, biologists, and pharmacologists alike for its unique structural diversification pathways. Instead of sharing the same biosynthetic routes as observed in terpenoid metabolism, the common nitrogen-containing heterocyclic structure of more than 20,000 known alkaloids can be generated by various Mannich-like condensation of amino acids-derived iminiums (Lichman, 2021). The resulted alkaloid

scaffolds are then decorated and modified extensively to form a wide range of structures, ranging from the poisonous coniine with a simple eight-carbon and one-nitrogen skeleton naturally occurring in hemlock (*Conium maculatum*) to the complex anti-tumor drug vinblastine with a dimeric 45-carbon and four-nitrogen scaffold found in Madagascar periwinkle (*Catharanthus roseus*; Ziegler and Facchini, 2008; O'Connor, 2010).

For example, in monoterpenoid indole alkaloid (MIA) biosynthesis, the amine moiety from tryptamine, a derivative of the amino acid tryptophan, is condensed with the aldehyde moiety from secologanin, a member of the non-canonical monoterpenoid group called iridoids, to yield strictosidine. From this central precursor, different multiple-step pathways are catalyzed by scaffolding and tailoring enzymes such as cytochrome P450 monooxygenases (P450), 2-oxoglutarate-dependent dioxygenases, methyltransferases, dehydrogenases, acetyltransferases, and glycosyltransferases. This leads to more than 2,000 MIA structures mostly found in the dogbane family (Apocynaceae), with vinblastine as the best-known example. Other illustrating examples are found in the biosynthesis of benzylisoquinoline alkaloids (BIAs) which starts with the condensation of the amine moiety of dopamine and the aldehyde group of 4-hydroxyphenylacetaldehyde, both derived from the amino acid tyrosine. The resulted (*S*)-norcoclaurine goes through series of structural changes including oxidation, reduction, methylation, acetylation, and decarboxylation to yield approximately 2,500 BIA structures such as the well-known narcotic painkiller morphine in opium poppy (*Papaver somniferum*; Ziegler and Facchini, 2008; O'Connor, 2010; Dastmalchi et al., 2018; Desgagné-Penix, 2021).

Plant P450s and Chemical Diversity in Plants

Dubbed "nature's most versatile biological catalyst," P450s display incredible adaptability in all domains of life and even in certain viruses (Coon, 2005). Starting as a component of the ancient cell's biochemical response to a world filled with the newly-abundant and poisonous oxygen gas (Wickramasinghe and Villee, 1975), P450s' ability to scissor atmospheric dioxygen at physiological temperatures has turned them into a reservoir of catalysts whose members have been recruited over and again in various metabolic pathways. The structure of a typical P450 consists of a central haem iron tethered by the thiolate group of a cysteine residue. This arrangement allows the formation of the highly reactive Fe^{IV}-oxo species, which abstracts hydrogen from the substrate's chemically inert C–H bond and can kick start a cascade of structural diversification and functionalization with high selectivity, a catalytic feat enviable to chemists (Lewis et al., 2011). Not only from its bond with carbon, hydrogen can also be abstracted by P450 enzymes from bonds with nitrogen, oxygen, and sulfur to allow oxidation and a range of other reactions such as epoxidation, sulfoxidation, dehydrogenation, aryl–aryl coupling and dehalogenation (Coon, 2005; Guengerich and Munro, 2013; Lamb and Waterman, 2013). From these initial chemical changes, carbon skeleton re-arrangements can further expand the chemical space (Tang et al., 2017; Zhang and Li, 2017).

P450 diversity reflects the evolution of metabolism and adaptation in living organisms, especially plants, fungi, and bacteria, whose chemical diversity is extraordinary. In plants, hundreds of thousands of P450s have been identified and grouped in 277 families of sequences sharing 40% or higher identity; of these, more than 16,000 have been named (Nelson, 2018). P450s control many metabolic steps and pathways of plant primary metabolites such as the growth regulators gibberellins, brassinosteroids, and abscisic acid (Helliwell et al., 2001; Turk et al., 2003; Kitahata et al., 2005). P450 enzymes also play a crucial role in plants' eco-physiological adaptation as they catalyze the production of defensive compounds and allelochemicals, among other specialized metabolites. Indeed, the vast majority of plant natural products are oxygenated, and, as most oxidations of chemicals in the living world are catalyzed by P450s, these proteins constitute the largest superfamily of enzymes underlying the diversification of plant natural products (Nelson and Werck-Reichhart, 2011; Hamberger and Bak, 2013; Guengerich, 2018).

Increasing genomic and transcriptomic data in recent years have facilitated the characterization of hundreds of P450s involved in plant specialized metabolism. Here we review the roles of P450s in the structural diversification of plant alkaloids with select recently-elucidated examples being discussed in accordance with the chemical modifications they catalyze (e.g., oxygenation, scaffold re-arrangement, etc.). The metabolism of major alkaloid groups (i.e., BIA, MIA, etc.) which involve these P450s are summarized in the figures with more details available in several excellent reviews published in the past years (Hagel and Facchini, 2013; Larsson and Ronsted, 2013; Thamm et al., 2016; Frick et al., 2017; Dastmalchi et al., 2018; Polturak and Aharoni, 2018; Zenkner et al., 2019; Desgagné-Penix, 2021; Lichman, 2021). Given the unique chemical diversification of alkaloids, insights into the power of P450-based biocatalysts offer essential lessons for exploring unknown pathways as well as generating new-to-nature chemical diversity with tremendous potential applications.

OXYGENATION AS A STARTING POINT FOR CHEMICAL DIVERSIFICATION

The most common reaction catalyzed by P450s is the addition of an oxygen atom into the substrate molecule in the form of a hydroxyl or an epoxide group. This has particularly relevant implications in biotechnologies and pharmaceutical industries as the oxidation of a single C–H bond functionalizes many compounds or makes them more biologically active. For instance, the stereo- and regio-selective oxidations of camptothecin and compactin lead to their more potent hydroxylated forms hydroxycamptothecin (anti-cancer) and pravastatin (lipid-lowering), respectively (Kingsbury et al., 1991; Watanabe et al., 1995), and the underlying oxidases can address industrial-scale drug production issues (Di Nardo and Gilardi, 2020). Furthermore, these simple oxygenations prompt a whole host of additional chemical decorations on the molecules, such as methylation, acetylation, glycosylation, and structural re-arrangements in many pathways.

The recently-elucidated BIA biosynthetic pathways feature several P450s that catalyze such oxygenations. In noscapine biosynthesis in opium poppy (*P. somniferum*), three members of the CYP82 family add single hydroxyl groups to the *N*-methylcanadine scaffold at three different positions with different chemical fates in the end product, noscapine. The first committed step of the pathway was catalyzed by CYP82Y1, hydroxylating (*S*)-*N*-methylcanadine at C1 position (Dang and Facchini, 2014b). The second and third P450s, CYP82X2 and CYP82X1, hydroxylates at C13 and C8 positions, respectively (Dang et al., 2015; **Figure 1**). While the 1-hydroxyl group undergoes a methylation reaction later in the pathway, the 8-hydroxyl group constitutes an unstable structure with the adjacent quaternary ammonium group and is spontaneously converted to an aldehyde group by breaking the C8–N7 bond. This newly-formed C8 aldehyde group then forms a hemiacetal ring with the 13-hydroxyl group. Intriguingly, before forming the hemiacetal structure with the 8-hydroxyl group, the 13-hydroxyl group undergoes acetylation and subsequent deacetylation before and after 8-hydroxylation by CYP82X1. As CYP82X1 and CYP82X2 do not accept each other's substrates, this acetylation seems to protect the oxygenated moiety at C13 and allow both 13- and 8-hydroxylations to occur, albeit in strict order (Dang et al., 2015).

Other members of the CYP82 family have also been found to be responsible for ring hydroxylations of BIAs. In the biosynthesis of the anti-microbial BIA sanguinarine, CYP82N4 catalyzes the hydroxylation at C14 of (*S*)-*cis*-*N*-methylstylopine, breaking the C14–N7 bond to yield protopine. Protopine is in turn hydroxylated by CYP82N3 in opium poppy (Beaudoin and Facchini, 2013) and by CYP82N2 in California poppy (*Eschscholzia californica*; Takemura et al., 2013) to 6-hydroxyprotopine, which is spontaneously converted to dihydrosnaguinarine, illustrating how hydroxylations by P450s can lead to further structural re-arrangement (**Figure 1**).

CYP82S18, a unique P450 involved in MIA metabolism in Indian snakeroot (*Rauwolfia serpentina*), catalyzes not only the ring hydroxylation of vinorine to form vomilenine, but also the non-oxidative isomerization of this product to perakine (**Figure 2**; Dang et al., 2017). Although enzymes are not required for the conversion of vomilenine to perakine, it needs extreme chemical catalysis conditions (Taylor et al., 1962), and biochemical studies showed that plant enzymes facilitate the isomerization (Sun et al., 2008; Dang et al., 2017). It remains unclear how CYP82S18 catalyzes this non-oxidative structural change; however, data suggest that keeping the product vomilenine in the active site after the hydroxylation of vinorine is essential, and a series of re-arrangements *via* ring opening and Michael addition could be facilitated by this active site (Dang et al., 2017). Indian snakeroot's CYP82S18 could be considered as an example of a "moonlighting" P450 that can catalyze different types of structural transformation on the substrate, although it does not use different active sites as seen the "moonlighting" terpene synthase/oxidase CYP170A1 in *Streptomyces coelicolor* (Zhao et al., 2009). More importantly, this unique catalytic capacity of CYP82S18 underlies the divergence of MIA metabolism in Indian snakeroot as vomilenine is the central intermediate leading to a series of MIAs, including the antiarrhythmic drug ajmaline while the bifurcated perakine branch leads to raucaffrinoline. The 21-hydroxyl group of vomilenine resulted from the CYP82S18's hydroxylase activity

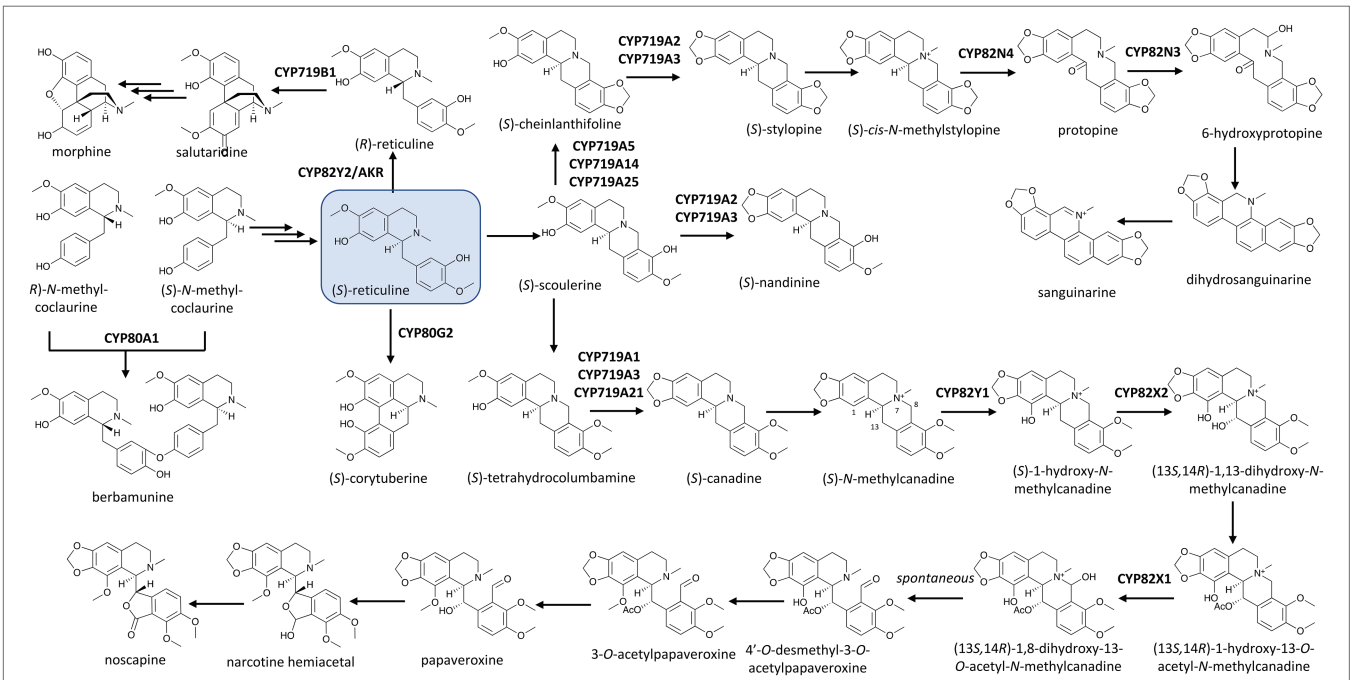


FIGURE 1 | P450s in the biosynthetic network of benzylisoquinoline alkaloids (BIAs) in the Ranunculales with (*S*)-reticuline (boxed) as the central precursor. Multiple and single arrows indicate multiple- and single-step pathways, respectively. Enzymes indicated here are discussed in the text.

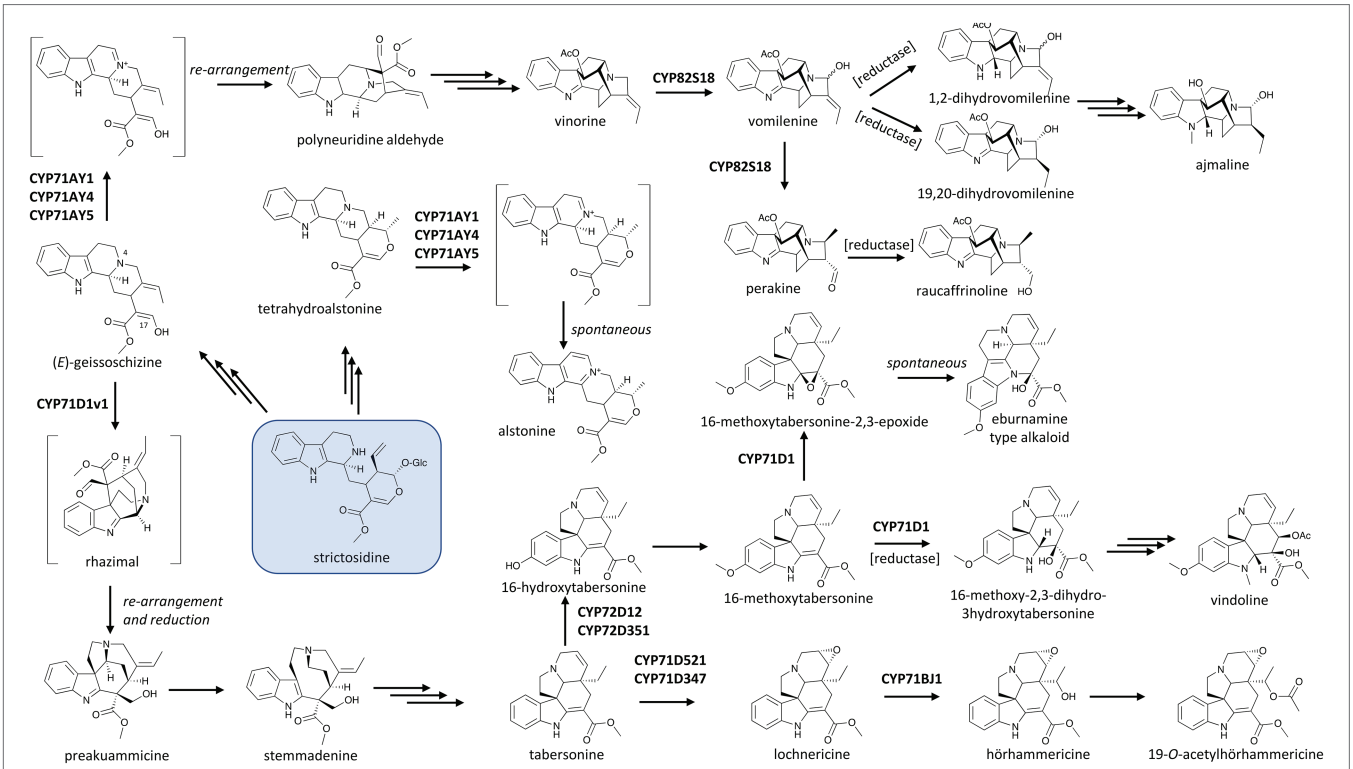


FIGURE 2 | P450s in the biosynthetic network of monoterpenoid indole alkaloids (MIAs) in the Apocynaceae with strictosidine (boxed) as the central precursor. Multiple and single arrows indicate multiple- and single-step pathways, respectively. Enzymes indicated here are discussed in the text. V19H: (+)-vincadifformine 19-hydroxylase, a close homologue of CYP71D1.

also allows subsequent glycosylation in the end products of these divergent pathways (**Figure 2**).

Examples of P450s from other families involved in alkaloid hydroxylation can be widely found in the metabolism of MIAs in the alkaloids-rich dogbane family (Apocynaceae). As early as in the 1990s, the P450-based 16-hydroxylation of tabersonine, a precursor of many MIAs, was identified in Madagascar periwinkle (*C. roseus*; St-Pierre and De Luca, 1995; Schröder et al., 1999). This P450, CYP71D12, yields 16-hydroxytabersonine, the branching precursor leading from tabersonine to vindoline, which together with catharanthine forms the anti-cancer drug vinblastine (**Figure 2**). More recently, a homologue sharing 82% amino acid identity to CYP71D12 and CYP71D351 was found to be another tabersonine 16-hydroxylase. CYP71D351, in contrast to CYP71D12, is expressed in better correlation with other vindoline biosynthetic genes. This suggests that it plays a major role in the biosynthesis of vindoline, which is tightly controlled in an organ-dependent manner and accumulates mostly in leaves of Madagascar periwinkle (Besseau et al., 2013). The 16-hydroxyl group of tabersonine allows a methyl group to be transferred onto the molecule, and both the resulted 16-methoxytabersonine and tabersonine can be further oxidized by another CYP71D subfamily member, CYP71D1, to yield the corresponding 2,3-epoxides (Kellner et al., 2015; Qu et al., 2015; Edge et al., 2018). Intriguingly, yeast feeding assay suggests CYP71D1 converts 16-methoxytabersonine to its 2,3-epoxide, which subsequently undergoes re-arrangement to an eburnamine scaffold similar to that of the anti-hypertension drug vincamine (Kellner et al., 2015; **Figure 2**). Reports by Qu et al. (2015) and Edge et al. (2018) as well as the early work by Wenkert and Wickberg (1965) suggest that such a re-arrangement of the 2,3-epoxide intermediate is induced by the acidification of the yeast culture or extraction process. Furthermore, the concerted activities of CYP71D1 and tabersonine-3-reductase were reported to reduce the C2–C3 double bond and hydroxylate C3 of tabersonine and 16-methoxytabersonine. It is the products of these oxidoreduction catalyzes, not the epoxides, that serve as precursors to vindorosine and vindoline, respectively (Qu et al., 2015; Edge et al., 2018; **Figure 2**). These transformations, underlined by enzymatic activity or otherwise, highlight the frequent occurrence and potential applications of oxidation-induced re-arrangements in MIAs.

The divergence of tabersonine metabolism involves other members of the CYP71 family. CYP71BJ1 was implicated in the biosynthesis of 19-*O*-acetylhörhammericine as it hydroxylates the tabersonine scaffold at C19 and thus allows the acetylation at this position in the final product (**Figure 2**; Giddings et al., 2011). More recently, Carqueijeiro et al. (2018) found two CYP71D homologues, CYP71D347 and CYP71D521, which catalyze the same 6,7-epoxidation of tabersonine to lochnericine. As both of these epoxidases exhibit strict substrate specificity towards tabersonine while the 19-hydroxylase CYP71BJ1 can accept both tabersonine and lochnericine, the 6,7-epoxidation

appears to be the first step in the pathway leading to 19-*O*-acetylhörhammericine from tabersonine (**Figure 2**). Intriguingly, the substrate spectrum of CYP71BJ1 may not extend to other aspidorsperma MIA enantiomers. In addition to tabersonine and catharanthine, the stemmadenine pathway gives rise to (+)-vincadifformine. This compound is hydroxylated at C19 position by (+)-vincadifformine 19-hydroxylase. Although this reaction is almost identical to the 19-hydroxylations of tabersonine and lochnericine (tabersonine-6,7-epoxide) catalyzed by CYP71BJ1, (+)-vincadifformine 19-hydroxylase shares a higher sequence identity (about 80%) to CYP71D1 compared to its identity to CYP71BJ1 (37%; Williams et al., 2019). The hydroxylation of (+)-vincadifformine and subsequent acetylation leading to (+)-echitovenine, parallel with the 19-*O*-acetylhörhammericine route, underscores enzymatic stereo-selectivity as a critical feature in defining similar yet distinct pathways in MIA diversification (**Figure 2**).

A member of the CYP75A subfamily responsible for two hydroxylations of the same pathway was featured in the recent near-complete elucidation of colchicine biosynthesis (Nett et al., 2020). Colchicine from *Colchicum* and *Gloriosa* species has long been used to treat inflammations, including gout and Behçet's disease (Barnes, 2006). It has been hypothesized that the biosynthesis of colchicine involves the condensation of 4-hydroxydihydrocinnamaldehyde and dopamine, derived from L-phenylalanine and L-tyrosine, respectively, to yield the 1-phenethylisoquinoline structure, which is then methylated, hydroxylated, and rearranged in several steps to form the tropolone ring in colchicine. Nett et al. (2020) discovered that CYP75A109 catalyzes not only one but possibly two hydroxylations at two *meta* positions on ring A of the 1-phenethylisoquinoline scaffold (**Figure 3**), and both of the resulting hydroxyl groups are later methylated in the pathway.

OXIDATIVE SCAFFOLD FORMATION

Cyclization reactions that give rise to complex polycyclic scaffolds are hallmarks of alkaloid biosynthetic pathways. While synthetic efforts since the dawn of organic synthesis have demonstrated how different alkaloids can be chemically synthesized from a common scaffold, the enzymes, mostly P450s, that control the regio- and stereo-specific re-arrangement and chemical diversification of the central intermediate found in biological systems have largely remained cryptic until recently. In MIA biosynthesis, a key scaffold forming step is catalyzed by sarpagan bridge enzyme (SBE) and transforms the skeletal scaffold of the central MIA intermediate strictosidine into sarpagan, ajmalan and alstophyllan alkaloid classes (Namjoshi and Cook, 2016), including the class Ia antiarrhythmic agent ajmaline and the anti-cancer compound koumine (Hashimoto et al., 1986; Zhang et al., 2015). More than 20 years after SBE activity was first detected in plants (Schmidt and Stöckigt, 1995), three P450 homologues in *R. serpentina* (CYP71AY4), *Gelsimium sempervirens* (CYP71AY5) and *C. roseus* (CYP71AY1) were found to be responsible for the formation of this scaffold (Dang et al., 2018). The SBE-catalyzed oxidation does not involve oxygenation but the formation of a Schiff base at N4 position of geissoschizine, a strictosidine derivative. This allows a skeletal re-arrangement to form polyneuridine aldehyde, the entry intermediate to the ajmalan-type and alstophyllan-type alkaloids (**Figure 2**). Intriguingly, when tested with a range of other MIAs, these SBEs turnover tetrahydroalstonine, a heteroyimbine alkaloid, to alstonine *via* a similar iminium intermediate. Guided by related yet structurally distinct substrates, the SBEs can catalyze either cyclization to form a sarpagan bridge or aromatization of the alkaloid scaffold and thus redirect and diversify the pathway at critical points (Dang et al., 2018). This illustrates how P450s' catalytic and substrate promiscuity,

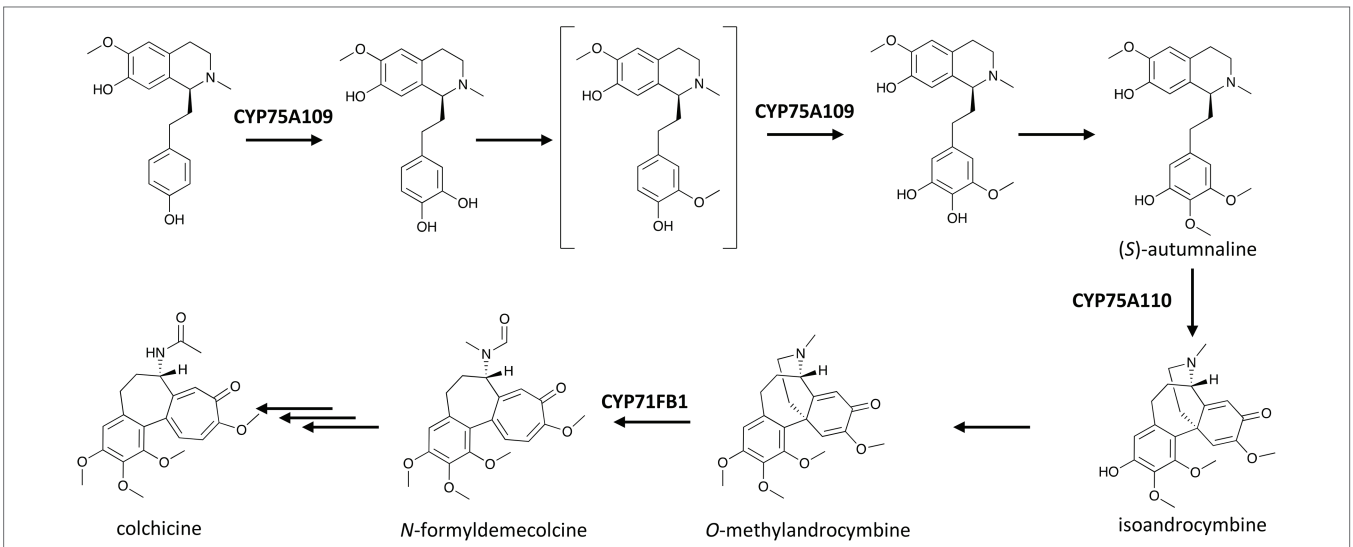


FIGURE 3 | P450s in the recent near elucidation of colchicine biosynthesis in the Colchicaceae. Multiple and single arrows indicate multiple- and single-step pathways, respectively. Enzymes indicated here are discussed in the text.

along with the inherent reactivity of these alkaloid substrates, can create a suite of structurally diverse chemical products in many pathways.

Besides acting alone, P450 enzymes can function in combination with other enzymes to open gateways into rich families of natural products. It is not an exception that concerted action of enzymes from two groups commonly involved in plant natural product metabolism, the P450s and the reductases, produces unexpected re-arrangements leading to different scaffolds. For instance, the CYP71/reductase module was observed in strychnos, sarpagan, ajmalane and β-carboline in MIA biosynthesis. Specifically, geissoschizine synthase (an alcohol dehydrogenase), geissoschizine oxidase (a P450), and two other reductases from *C. roseus*, when assayed simultaneously, catalyzed a series of tandem reactions that lead to the remarkable re-arrangement of the tetrahydro-β-carboline strictosidine substrate into the corynanthean, strychnos, iboga, and aspidosperma scaffolds (Benayad et al., 2016; Tatsis et al., 2017; Qu et al., 2018a). The alcohol group at C17 of geissoschizine is oxidized to an aldehyde by CYP71D1v1, triggering a cascade of skeletal changes and a reduction catalyzed by two conceptive reductases (Tatsis et al., 2017; Qu et al., 2018a,b; **Figure 2**). The resulting preakuammicine is then either spontaneously transformed to akuammicine or reduced to stemmadenine, the precursor to tabersonine and catharanthine (Caputi et al., 2018).

The dual catalytic function of vinorine hydroxylase, CYP82S18, in the ajmaline biosynthetic pathway mentioned earlier is also driven by the presence of different downstream aldo/keto reductases (AKRs). Various combinations of these reductases with CYP82S18 diverge the vomilenine pool to tissue-specific metabolic routes with different end-products, including ajmaline, raucaffrinoline, raucaffricine, rauglucine, and 21-hydroxysarpagane glucoside (**Figure 2**; Dang et al., 2017). Intriguingly, an unusual P450, CYP82Y2, was identified to be a domain of a fused protein that also includes an AKR domain, and this fused AKR/P450 catalyzes the isomerization of (*S*)-reticuline to its (*R*)-epimer, a precursor of morphine biosynthesis in opium poppy (Farrow et al., 2015; Galanie et al., 2015; Winzer et al., 2015; **Figure 1**). What is more unusual is that in contrast with other CYP82 enzymes, which usually catalyze ring hydroxylation of BIAs (Dang and Facchini, 2014b; Dang et al., 2015), CYP82Y2 in reticuline isomerization removes hydrogen from (*S*)-reticuline to yield 1,2-dehydroreticuline. The resulted double bond is in turn reduced by the AKR domain of the AKR/CYP82Y2 fusion to produce (*R*)-reticuline. Similar fusion proteins made up of a CYP82Y2-like portion and an AKR domain were also found in dwarf breadseed poppy (*P. setigerum*) and Persian poppy (*P. bracteatum*), revealing an intriguing evolutionary solution in some poppy species to metabolic flux and/or regulation hurdles of BIA metabolism (Farrow et al., 2015; Winzer et al., 2015).

The abstraction of hydrogen from substrates underlies P450-based catalysis as seen in the oxidation reactions discussed above. In many cases, this dehydrogenation can produce more than one radical, allowing diradical coupling, and consequently, ring formation. This structural transformation can fundamentally alter the core scaffold of the compound. Some of the earliest demonstrated examples of such P450-catalyzed C–C and C–O coupling were observed in BIA biosynthesis (Zenk et al., 1989). CYP80A1 from barberry (*Berberis stolonifera*) was the first P450 identified to catalyze a C–O coupling reaction, condensing two methylcoclaurine molecules with different enantiomeric conformations to yield the (*R*,*S*)-dimer berbamunine (**Figure 1**). Interestingly, although CYP80A1 is regio-specific, it can accept two (*R*)-methylcoclaurine to form the (*R*,*R*)-dimer product guattegaumerine (Kraus and Kutchan, 1995).

The CYP719A subfamily members found in isoquinoline alkaloids-producing plants are responsible for forming the methylenedioxy bridge in these compounds. In meadow rue (*Thalictrum tuberosum*), Rueffer and Zenk (1994) first observed the P450-based conversion of (*S*)-tetrahydrocolumbamine to (*S*)-canadine, also known as (*S*)-tetrahydroberberine, the precursor for many important BIAs such as noscapine, berberine, and sanguinarine. Other (*S*)-canadine synthases were later identified in several species, including CYP719A1 in Japanese goldthread (*Coptis japonica*; Ikezawa et al., 2003) and CYP719A21 in opium poppy (Dang and Facchini, 2014a). The methylenedioxy bridge formation on (*R*,*S*)-cheilanthifoline leading to (*S*)-stylopine is catalyzed by other members of this subfamily, including CYP719A2 in California poppy (Ikezawa et al., 2007) or CYP719A13 in Mexican prickly poppy (*Argemone mexicana*; Díaz Chávez et al., 2011). Ikezawa et al. (2007) also identified CYP719A3, which can accept three substrates (*R*,*S*)-cheilanthifoline, (*S*)-scoulerine, and (*S*)-tetrahydrocolumbamine to yield (*S*)-stylopine, (*S*)-nandinine, and (*S*)-canadine, respectively (**Figure 1**). (*S*)-Scoulerine is also subject to another methoxyphenol cyclization catalyzed by CYP719A5 in California poppy (Ikezawa et al., 2009), CYP719A14 in Mexican prickly poppy (Díaz Chávez et al., 2011), or CYP719A25 in opium poppy (Desgagné-Penix et al., 2010; Dang and Facchini, 2014b) to form (*S*)-cheinlanthifoline. Recently, CYP719A37 in black pepper (*Piper nigrum*) has been found to be responsible for the presence of the methylenedioxy bridge in piperic acid, a precursor of the pungent alkaloid piperine (Schnabel et al., 2021).

More than a decade after the captivating discovery of the P450-based C–O coupling reaction, the C–C coupling activity by a P450, CYP80G2, was identified in the intramolecular phenol coupling of (*S*)-reticuline that produces (*S*)-corytuberine (Ikezawa et al., 2008; **Figure 1**). Not too long after, another P450, CYP719B1, was found to catalyze a similar reaction on (*R*)-reticuline to form salutaridine (Gesell et al., 2009). P450-catalyzed C–C coupling also plays a significant role in the chemical diversification of Amaryllidaceae alkaloids, a group of approximately 600 isoquinoline alkaloids. As the name suggests, these alkaloids are tightly associated with the daffodil family (Amaryllidaceae) and are derived from the phenethylamine norbelladine and its derivative 4'-*O*-methylnorbelladine (Desgagné-Penix, 2021). The intramolecular C–C coupling of 4'-*O*-methylnorbelladine can occurs at different positions and stereochemistry. The *para-para* cyclization yields both (10b*S*,4a*R*)-noroxomaritidine and its enantiomer (10b*R*,4a*S*)-noroxomaritidine, which is the precursor for the biosynthesis of hemanthamine, pancrastatine, montanine and other *para-para* cyclized derivatives. The *para-ortho* cyclization affords *N*-demethylnarwedine (nornarwedine), leading to galanthamine,

chlidanthine, lycoramine and similar compounds. The *ortho-para* coupling forms noroxopluviine, precursor of hippeastrin, lycorine and derivatives. In *Narcissus* sp. *aff. Pseudonarcissus*, CYP96T1 was identified as the enzyme that catalyzes the *para–para* coupling of 4'-*O*-methylnorbelladine to produce two noroxomaritidine enantiomers. This enzyme also displayed some *para–ortho* coupling activity as it yields *N*-demethylnarwedine as a minor product (Kilgore et al., 2016; **Figure 4A**). Despite this structural diversity and a long history of Amaryllidaceae plants being used in traditional

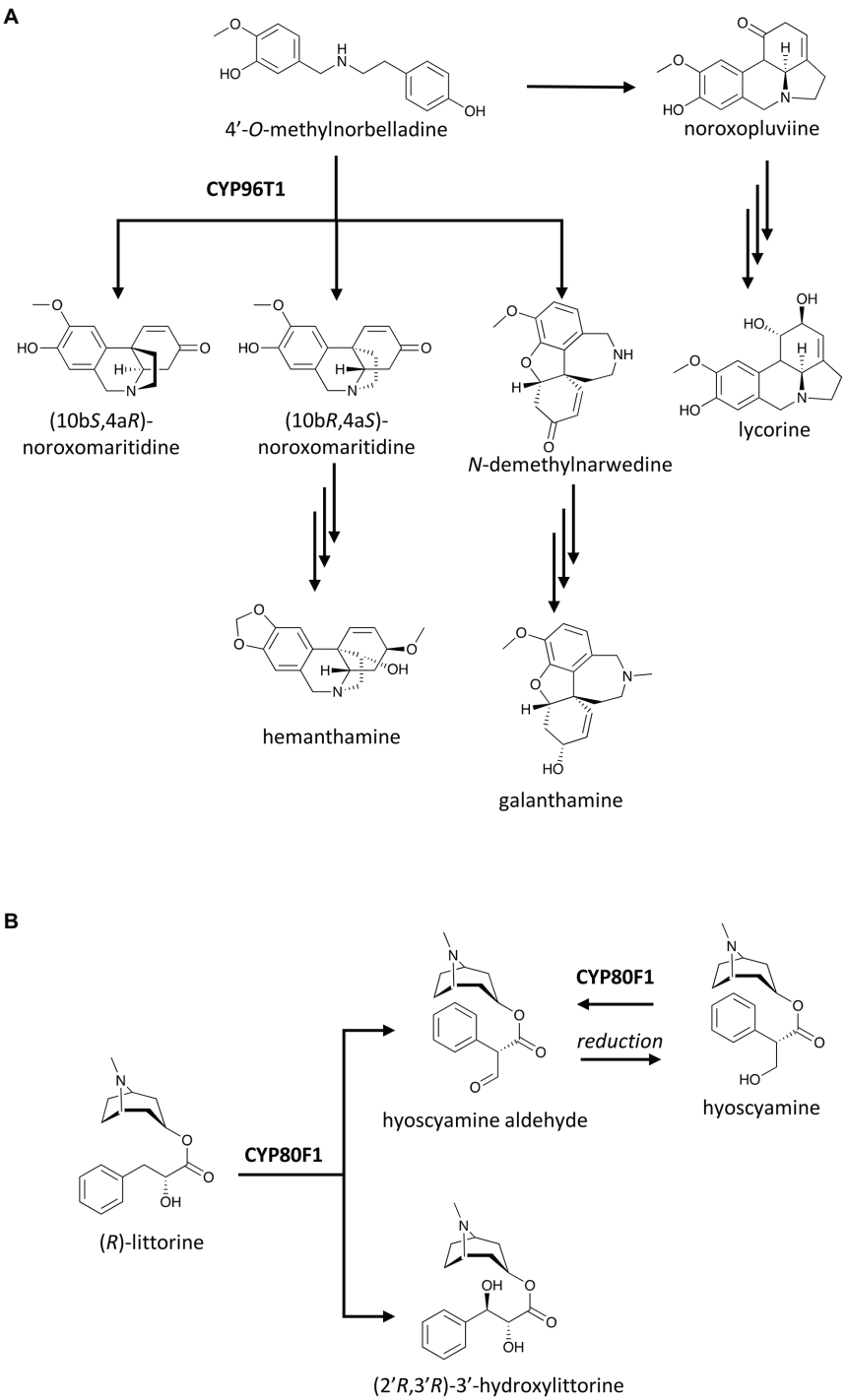


FIGURE 4 | Examples of P450 catalysis beyond oxygenation including: **(A)** C–C couplings of 4'-*O*-methylnorbelladine, a central precursor of Amaryllidaceae alkaloids, leading to various pathways; and **(B)** group migration in tropane alkaloid biosynthesis. Enzymes indicated here are discussed in the text.

medicine, galanthamine has been the only Amaryllidaceae alkaloid commercialized as a drug to treat neurodegenerative disorders. Increasing plant genomics data of Amaryllidaceae plants will undoubtedly reveal more P450s with C–C and C–O coupling activities and help us explore their untapped therapeutic potentials in the near future.

OXIDATIVE RING OPENING

P450s are involved not only in ring formation but also in ring breakage, allowing for the formation of new scaffolds, or activate the chemicals for further condensation in alkaloid metabolism. Ring opening induced by oxygenations such as those catalyzed by CYP82X1 and CYP82N4 in BIA biosynthesis has been described earlier (Dang et al., 2015). Other P450-catalyzed ring openings leading to characteristic scaffolds have also been found. One of the earliest P450s catalyzing such unique reactions is secologanin synthase from *C. roseus*, CYP72A1, which opens the cyclopentane ring of loganin to yield secologanin (**Figure 5**; Irmler et al., 2000).

Secologanin synthase was the first enzyme of the then newly-found CYP72 family (Vetter et al., 1992; Irmler et al., 2000), including many members that catalyze the usual P450-based oxygenation in terpenoid metabolism (Turk et al., 2003; Ikezawa et al., 2011; Fukushima et al., 2013). CYP72A1's unique ring opening activity is shared with close homologues in the biosynthesis of iridoids, a group of non-canonical terpenoids such as secologanin, which is condensed with tryptamine to make the central MIA precursor strictosidine (Mizukami et al., 1979; McCoy et al., 2006). In the Chinese happy tree (*Camptotheca acuminata*), both CYP72A565 and CYP72A610 display secologanin synthase activity and break the cyclopentane ring of loganic acid to produce secologanic acid, the hypothetical precursor of the anti-cancer agent camptothecin. Intriguingly, these two enzymes also hydroxylate 7-deoxyloganic acid to yield loganic acid before proceeding to open its C7–C8 bond (**Figure 5**; Yang et al., 2019).

Rodríguez-López et al. (2021) found two bifunctional CYP72 enzymes in olive (*Olea europaea*) with dehydrogenase activity on the hydroxyl group of 7-*epi*-loganin and C7–C8 bond breaking activity on the resulted ketologanin (**Figure 5**). These two enzymes, named secoxyloganin synthase and oleoside methyl ester synthase based on their products, showed that ring opening activity is not restricted to the CYP72A subfamily as they share less than 50% amino acid identity with the three CYP72A enzymes in *C. roseus* and *C. acuminata* (Irmler et al., 2000; Yang et al., 2019).

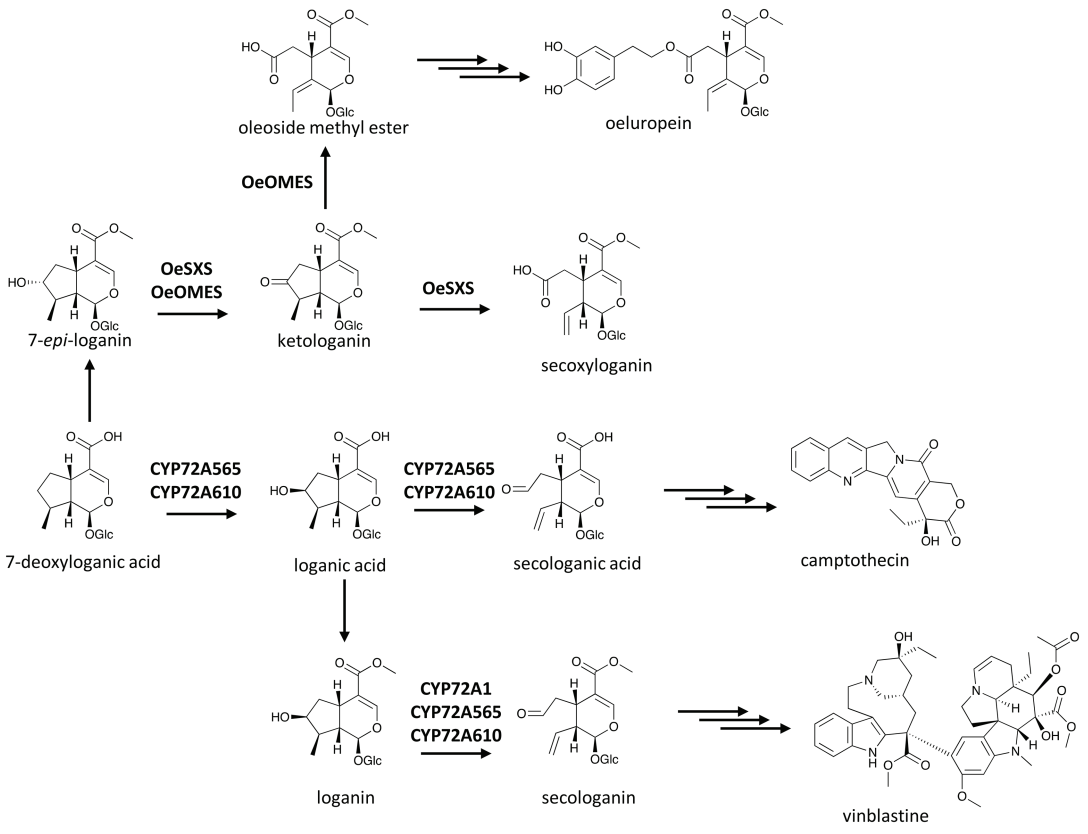


FIGURE 5 | P450s in the terpenoid scaffold formation of MIAs. Multiple and single arrows indicate multiple- and single-step pathways, respectively. Enzymes indicated here are discussed in the text.

OTHER STRUCTURAL RE-ARRANGEMENTS

The skeletal re-arrangement of (*R*)-littorine to (*S*)-hyoscyamine, a tropane alkaloid drug used to manage spasms and symptoms of Parkinson's disease, had intrigued researchers for a long time and was hypothesized to be underlined by a P450 (Robins et al., 1995). Li et al. (2006) demonstrated that this unique migration of the whole 8-methyl-8-azabicyclo[3.2.1]octan-3-yl ester group, or re-arrangement of the 3-phenyllactate to tropate, of (*R*)-littorine is catalyzed by CYP80F1 to produce (*S*)-hyoscyamine aldehyde in black henbane (*Hyocyamus niger*). (*S*)-hyoscyamine is thought to be the precursor of (*S*)-hyoscyamine in the plant, but CYP80F1 can also oxidize the 3'-hydroxyl group of (*S*)-hyoscyamine back to (*S*)-hyoscyamine aldehyde. In addition, CYP80F1 displays remarkable catalytic promiscuity by catalyzing a single oxidation without group migration at C3' position of (*R*)-litorine to yield (2'*R*,3'*R*)-3'-hydroxylittorine (**Figure 4B**; Li et al., 2006).

Colchicine biosynthesis again showcases the fascinating diversity of P450 activities. In addition to the two CYP75A enzymes described earlier, the metabolic pathway of colchicine involves a unique P450 with ring expansion activity. This P450-based ring expansion was demonstrated using NADPH and microsome from the seed of autumn crocus (*Colchicum autumnale*) by Rueffer and Zenk (1998), and the enzyme was identified more than 20 years later by Nett et al. (2020) as CYP71FB1. In this reaction, the 1,4-cyclohexadiene ring of *O*-methylandrocymbine is expanded to yield *N*-formyldemecolcine, the precursor of colchicine. Although the final steps leading to colchicine are yet to be elucidated, the finding of CYP71BF1 activity has established how the characteristic tropolone ring of colchicine structure is built (**Figure 3**).

GENERAL DISCUSSION

P450 enzyme control occurs at many steps in all chemical diversification pathways of alkaloids. The catalytic versatility of P450 enzymes provides not only tremendous chemical diversity and thus adaptability to increase plants' fitness but also blueprints for biocatalyst engineering with applications in medicine, industry, and bioremediation (Bernhardt, 2006; Nelson and Werck-Reichhart, 2011; Sakaki, 2012; Li et al., 2020; Shang and Huang, 2020).

Although enzymes in the same P450 family tend to catalyze similar reactions in alkaloid metabolism, such as single hydroxylation by CYP82 members and methylenedioxy bridge formation by CYP719 enzymes, this is not always the case. Increasingly abundant genomic data will allow for the identification of more P450s and their roles in alkaloid biosynthesis in natural products biochemistry in general (Nelson, 2018). Given the complexity of alkaloid structures and the sheer number of unelucidated pathways, what appear to be "unusual" such as the ring-expanding functions by CYP71BF1 or a P450-dependent oxidoreductase fusion like CYP82Y2 could turn out to occur in other pathways and provide excellent templates for future enzyme engineering to harness these unique chemical prowesses. Likewise, understanding differential substrate specificities of P450s acting in the same pathway and on similar substrates as observed in the noscapine pathway sheds light on the chemical strategies that plants employ and thus inform future pathway engineering approaches. Recent progress in exploring plant P450s have afforded the production in engineered hosts of alkaloids such as dihydrosanguinarine, noscapine (Li et al., 2018), thebaine, hydrocodone (Thodey et al., 2014), strictosidine (Brown et al., 2015), vindoline (Qu et al., 2015), and *N*-formyldemecolcine (Nett et al., 2020).

The broad substrate spectrum of many P450s can complicate biosynthesis studies and metabolic engineering efforts (Hidalgo et al., 2017; De La Peña and Sattely, 2021). Nevertheless, such promiscuity sheds light into the evolution of these enzymes, and indicates their crucial role as part of the catalytic reservoirs whose members can be recruited for emerging pathways and further drive the chemical diversity of plants (Tawfik, 2010; Ikezawa et al., 2011; Weng et al., 2012; Guo et al., 2016; Dang et al., 2017, 2018; Forman et al., 2018; Christ et al., 2019; Nguyen et al., 2019; Lichman et al., 2020; Nguyen and O'Connor, 2020). Even for non-native or new-to-nature substrates including halogenated analogues, P450s display a certain degree of natural tolerance as observed in the multiple-step biotransformation of 7-chlorotryptamine to 12-chloro-19,20-dihydroakuammicine in MIA metabolism of *C. roseus* cultures (Glenn et al., 2011). This feature may, therefore, also provide natural templates for catalytic optimization towards desired and/or novel yet related activities. Despite the general challenge due to the lack of P450 structural information and the requirement of redox partners, P450 engineering will undoubtedly benefit from the cataloguing of new sequences and functions from the ever-expanding plant genome datasets. In addition, modification of the relatively-conserved substrate-recognition sites across P450s may allow product profile customization without experimental protein structural data (Gotoh, 1992; Forman et al., 2018; Shang and Huang, 2020).

There remain other challenges in understanding the catalytic mechanisms of P450 enzymes beyond substrate specificities such as non-oxidative reactions, while the membrane-bound nature of plant P450 enzymes impedes structural studies using crystallography approaches (Shang and Huang, 2020; Zhang et al., 2020). In addition, the interaction and localization of P450 enzymes with respect to other enzymes in the same pathways remain to be explored (Bassard et al., 2017). All of these continue to fascinate scientists for the years to come.

AUTHOR CONTRIBUTIONS

TTTD and TDN wrote the manuscript together. All authors contributed to the article and approved the submitted version.

FUNDING

TTTD received funding from the Canada Natural Science and Engineering Research Council (NSERC RGPIN-2019-05473), the Canada Foundation for Innovation (CFI 38167), UBC Eminence Fund, and the Michael Smith Foundation for Health Research Scholar (SCH-2020-0401).

REFERENCES

Barnes, C. G. (2006). Treatment of Behçet's syndrome. *Rheumatology* 45, 245–247. doi: 10.1093/rheumatology/kei257

Bassard, J.-E., Møller, B. L., and Laursen, T. (2017). Assembly of dynamic P450-mediated metabolons—order versus chaos. *Curr. Mol. Biol. Rep.* 3, 37–51. doi: 10.1007/s40610-017-0053-y

Beaudoin, G. A. W., and Facchini, P. J. (2013). Isolation and characterization of a cDNA encoding (S)-*cis*-*N*-methylstylopine 14-hydroxylase from opium poppy, a key enzyme in sanguinarine biosynthesis. *Biochem. Biophys. Res. Commun.* 431, 597–603. doi: 10.1016/j.bbrc.2012.12.129

Benayad, S., Ahamada, K., Lewin, G., Evanno, L., and Poupon, E. (2016). Preakuammicine: a long-awaited missing link in the biosynthesis of monoterpene indole alkaloids. *Eur. J. Org. Chem.* 2016, 1494–1499. doi: 10.1002/ejoc.201600102

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* 124, 128–145. doi: 10.1016/j.jbiotec.2006.01.026

Besseau, S., Kellner, F., Lanoue, A., Thamm, A. M. K., Salim, V., Schneider, B., et al. (2013). A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant Physiol.* 163, 1792–1803. doi: 10.1104/pp.113.222828

Brown, S., Clastre, M., Courdavault, V., and O'Connor, S. E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3205–3210. doi: 10.1073/pnas.1423555112

Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T. D., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* 360, 1235–1239. doi: 10.1126/science.aat4100

Carqueijeiro, I., Brown, S., Chung, K., Dang, T. T., Walia, M., Besseau, S., et al. (2018). Two tabersonine 6,7-epoxidases initiate lochnericine-derived alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* 177, 1473–1486. doi: 10.1104/pp.18.00549

Christ, B., Xu, C., Xu, M., Li, F. S., Wada, N., Mitchell, A. J., et al. (2019). Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* 10, 3206–3211. doi: 10.1038/s41467-019-11286-7

Coon, M. J. (2005). Cytochrome P450: nature's most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.* 45, 1–25. doi: 10.1146/annurev.pharmtox.45.120403.100030

Dang, T. T. T., Chen, X., and Facchini, P. J. (2015). Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. *Nat. Chem. Biol.* 11, 104–106. doi: 10.1038/nchembio.1717

Dang, T. T. T., and Facchini, P. J. (2014a). Cloning and characterization of canadine synthase involved in noscapine biosynthesis in opium poppy. *FEBS Lett.* 588, 198–204. doi: 10.1016/j.febslet.2013.11.037

Dang, T. T. T., and Facchini, P. J. (2014b). CYP82Y1 is *N*-methylcanadine 1-hydroxylase, a key noscapine biosynthetic enzyme in opium poppy. *J. Biol. Chem.* 289, 2013–2026. doi: 10.1074/jbc.M113.505099

Dang, T. T. T., Franke, J., Carqueijeiro, I. S. T., Langley, C., Courdavault, V., and O'Connor, S. E. (2018). Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat. Chem. Biol.* 14, 760–763. doi: 10.1038/s41589-018-0078-4

Dang, T. T. T., Franke, J., Tatsis, E., and O'Connor, S. E. (2017). Dual catalytic activity of a cytochrome P450 controls bifurcation at a metabolic branch point of alkaloid biosynthesis in *Rauwolfia serpentina*. *Angew. Chem. Int. Ed.* 56, 9440–9444. doi: 10.1002/anie.201705010

Dastmalchi, M., Park, M. R., Morris, J. S., and Facchini, P. (2018). Family portraits: the enzymes behind benzylisoquinoline alkaloid diversity. *Phytochem. Rev.* 17, 249–277. doi: 10.1007/s11101-017-9519-z

De La Peña, R., and Sattely, E. S. (2021). Re-routing plant terpene biosynthesis enables momilactone pathway elucidation. *Nat. Chem. Biol.* 17, 205–212. doi: 10.1038/s41589-020-00669-3

Desgagné-Penix, I. (2021). Biosynthesis of alkaloids in Amaryllidaceae plants: a review. *Phytochem. Rev.* 20, 409–431. doi: 10.1007/s11101-020-09678-5

Desgagné-Penix, I., Khan, M. F., Schriemer, D. C., Cram, D., Nowak, J., and Facchini, P. J. (2010). Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC Plant Biol.* 10:252. doi: 10.1186/1471-2229-10-252

Di Nardo, G., and Gilardi, G. (2020). Natural compounds as pharmaceuticals: the key role of cytochromes P450 reactivity. *Trends Biochem. Sci.* 45, 511–525. doi: 10.1016/j.tibs.2020.03.004

Díaz Chávez, M. L., Rolf, M., Gesell, A., and Kutchan, T. M. (2011). Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy *Argemone mexicana*. *Arch. Biochem. Biophys.* 507, 186–193. doi: 10.1016/j.abb.2010.11.016

Edge, A., Qu, Y., Easson, M. L., Thamm, A. M., Kim, K. H., and De Luca, V. (2018). A tabersonine 3-reductase *Catharanthus roseus* mutant accumulates vindoline pathway intermediates. *Planta* 247, 155–169. doi: 10.1007/s00425-017-2775-8

Farrow, S. C., Hagel, J. M., Beaudoin, G. A. W., Burns, D. C., and Facchini, P. J. (2015). Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* 11, 728–732. doi: 10.1038/nchembio.1879

Forman, V., Bjerg-Jensen, N., Dyekjær, J. D., Møller, B. L., and Pateraki, I. (2018). Engineering of CYP76AH15 can improve activity and specificity towards forskolin biosynthesis in yeast. *Microb. Cell Factories* 17, 181–117. doi: 10.1186/s12934-018-1027-3

Frick, K. M., Kamphuis, L. G., Siddique, K. H. M., Singh, K. B., and Foley, R. C. (2017). Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. *Front. Plant Sci.* 8:87. doi: 10.3389/fpls.2017.00087

Fukushima, E. O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., et al. (2013). Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* 54, 740–749. doi: 10.1093/pcp/pct015

Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., and Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373

Gesell, A., Rolf, M., Ziegler, J., Chávez, M. L. D., Huang, F. C., and Kutchan, T. M. (2009). CYP719B1 is salutaridine synthase, the C–C phenol-coupling enzyme of morphine biosynthesis in opium poppy. *J. Biol. Chem.* 284, 24432–24442. doi: 10.1074/jbc.M109.033373

Giddings, L. A., Liscombe, D. K., Hamilton, J. P., Childs, K. L., DellaPenna, D., Buell, C. R., et al. (2011). A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. *J. Biol. Chem.* 286, 16751–16757. doi: 10.1074/jbc.M111.225383

Glenn, W. S., Nims, E., and O'Connor, S. E. (2011). Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J. Am. Chem. Soc.* 133, 19346–19349. doi: 10.1021/ja2089348

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* 267, 83–90. doi: 10.1016/S0021-9258(18)48462-1

Guengerich, F. P. (2018). Mechanisms of cytochrome P450-catalyzed oxidations. *ACS Catal.* 8, 10964–10976. doi: 10.1021/acscatal.8b03401

Guengerich, F. P., and Munro, A. W. (2013). Unusual cytochrome P450 enzymes and reactions. *J. Biol. Chem.* 288, 17065–17073. doi: 10.1074/jbc.R113.462275

Guo, J., Ma, X., Cai, Y., Ma, Y., Zhan, Z., Zhou, Y. J., et al. (2016). Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. *New Phytol.* 210, 525–534. doi: 10.1111/nph.13790

Hagel, J. M., and Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672. doi: 10.1093/pcp/pct020

Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. Lond.* 368:20120426. doi: 10.1098/rstb.2012.0426

Hashimoto, Y., Hori, R., Okumura, K., and Yasuhara, M. (1986). Pharmacokinetics and antiarrhythmic activity of ajmaline in rats subjected to coronary artery occlusion. *Br. J. Pharmacol.* 88, 71–77. doi: 10.1111/j.1476-5381.1986.tb09472.x

Helliwell, C. A., Chandler, P. M., Poole, A., Dennis, E. S., and Peacock, W. J. (2001). The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2065–2070. doi: 10.1073/pnas.98.4.2065

Hidalgo, D., Martínez-Márquez, A., Moyano, E., Bru-Martínez, R., Corchete, P., and Palazon, J. (2017). Bioconversion of stilbenes in genetically engineered root and cell cultures of tobacco. *Sci. Rep.* 7:45331. doi: 10.1038/srep45331

Ikezawa, N., Göpfert, J. C., Nguyen, D. T., Kim, S. U., O'Maille, P. E., Spring, O., et al. (2011). Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. *J. Biol. Chem.* 286, 21601–21611. doi: 10.1074/jbc.M110.216804

Ikezawa, N., Iwasa, K., and Sato, F. (2007). Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis

in *Eschscholzia californica*. *FEBS J.* 274, 1019–1035. doi: 10.1111/j.1742-4658.2007.05652.x

Ikezawa, N., Iwasa, K., and Sato, F. (2008). Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C–C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* cells. *J. Biol. Chem.* 283, 8810–8821. doi: 10.1074/jbc.M705082200

Ikezawa, N., Iwasa, K., and Sato, F. (2009). CYP719A subfamily of cytochrome P450 oxygenases and isoquinoline alkaloid biosynthesis in *Eschscholzia californica*. *Plant Cell Rep.* 28, 123–133. doi: 10.1007/s00299-008-0624-8

Ikezawa, N., Tanaka, M., Nagayoshi, M., Shinkyo, R., Sakaki, T., Inouye, K., et al. (2003). Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured *Coptis japonica* cells. *J. Biol. Chem.* 278, 38557–38565. doi: 10.1074/jbc.M302470200

Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., et al. (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* 24, 797–804. doi: 10.1046/j.1365-313x.2000.00922.x

Kellner, F., Geu-Flores, F., Sherden, N. H., Brown, S., Foureau, E., Courdavault, V., et al. (2015). Discovery of a P450-catalyzed step in vindoline biosynthesis: a link between the aspidosperma and eburnamine alkaloids. *Chem. Commun.* 51, 7626–7628. doi: 10.1039/C5CC01309G

Kilgore, M. B., Augustin, M. M., May, G. D., Crow, J. A., and Kutchan, T. M. (2016). CYP96T1 of *narcissus* sp. aff. *Pseudonarcissus* catalyzes formation of the *Para-ara*' C–C phenol couple in the Amaryllidaceae alkaloids. *Front. Plant Sci.* 7:225. doi: 10.3389/fpls.2016.00225

Kingsbury, W. D., Boehm, J. C., Jakas, D. R., Holden, K. G., Gallagher, G., Caranfa, M. J., et al. (1991). Synthesis of water-soluble (aminoalkyl)camptothecin analogues: inhibition of topoisomerase I and antitumor activity. *J. Med. Chem.* 34, 98–107. doi: 10.1021/jm00105a017

Kitahata, N., Saito, S., Miyazawa, Y., Umezawa, T., Shimada, Y., Yong, K. M., et al. (2005). Chemical regulation of abscisic acid catabolism in plants by cytochrome P450 inhibitors. *Bioorg. Med. Chem.* 13, 4491–4498. doi: 10.1016/j.bmc.2005.04.036

Kraus, P. F. X., and Kutchan, T. M. (1995). Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C–O phenol-coupling cytochrome P450 from the higher plant *Berberis stolonifera*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 2071–2075. doi: 10.1073/pnas.92.6.2071

Kutchan, T. M., Gershenzon, J., Møller, B. L., and Gang, D. R. (2015). "Natural products," in *Biochemistry and Molecular Biology of Plants*. eds. B. B. Buchanan, W. Gruissem and R. L. Jones (Chichester, UK: John Wiley and Sons), 1132–1206.

Lamb, D. C., and Waterman, M. R. (2013). Unusual properties of the cytochrome P450 superfamily. *Philos. Trans. R. Soc. Lond.* 368:20120434. doi: 10.1098/rstb.2012.0434

Larsson, S., and Ronsted, N. (2013). Reviewing Colchicaceae alkaloids – perspectives of evolution on medicinal chemistry. *Curr. Top. Med. Chem.* 14, 274–289. doi: 10.2174/1568026613666131216110417

Lewis, J. C., Coelho, P. S., and Arnold, F. H. (2011). Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev.* 40, 2003–2021. doi: 10.1039/C0CS00067A

Li, R., Reed, D. W., Liu, E., Nowak, J., Pelcher, L. E., Page, J. E., et al. (2006). Functional genomic analysis of alkaloid biosynthesis in *Hyoscyamus niger* reveals a cytochrome P450 involved in littorine rearrangement. *Chem. Biol.* 13, 513–520. doi: 10.1016/j.chembiol.2006.03.005

Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A., and Smolke, C. D. (2018). Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci.* 115, E3922–E3931. doi: 10.1073/pnas.1721469115

Li, Z., Jiang, Y., Guengerich, X. F. P., Ma, L., Li, S., and Zhang, W. (2020). Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *J. Biol. Chem.* 295, 833–849. doi: 10.1016/S0021-9258(17)49939-X

Lichman, B. R. (2021). The scaffold-forming steps of plant alkaloid biosynthesis. *Nat. Prod. Rep.* 38, 103–129. doi: 10.1039/d0np00031k

Lichman, B. R., Godden, G. T., Hamilton, J. P., Palmer, L., Kamileen, M. O., Zhao, D., et al. (2020). The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* 6:eaba0721. doi: 10.1126/sciadv.aba0721

McCoy, E., Galan, M. C., and O'Connor, S. E. (2006). Substrate specificity of strictosidine synthase. *Bioorg. Med. Chem. Lett.* 16, 2475–2478. doi: 10.1016/j.bmcl.2006.01.098

Mizukami, H., Nordlöv, H., Lee, S. L., and Scott, A. I. (1979). Purification and properties of strictosidine synthetase (an enzyme condensing tryptamine and secologanin) from *Catharanthus roseus* cultured cells. *Biochemistry* 18, 3760–3763. doi: 10.1021/bi00584a018

Namjoshi, O. A., and Cook, J. M. (2016). Sarpagine and related alkaloids. *Alkaloids Chem. Biol.* 76, 63–169. doi: 10.1016/bs.alkal.2015.08.002

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-313X.2011.04529.x

Nelson, D. R. (2018). Cytochrome P450 diversity in the tree of life. *Biochim. Biophys. Acta, Proteins Proteom.* 1866, 141–154. doi: 10.1016/j.bbapap.2017.05.003

Nett, R. S., Lau, W., and Sattely, E. S. (2020). Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* 584, 148–153. doi: 10.1038/s41586-020-2546-8

Nguyen, T. D., Kwon, M., Kim, S. U., Fischer, C., and Ro, D. K. (2019). Catalytic plasticity of germacrene A oxidase underlies sesquiterpene lactone diversification. *Plant Physiol.* 181, 945–960. doi: 10.1104/pp.19.00629

Nguyen, T.-D., and O'Connor, S. E. (2020). The progesterone 5β-reductase/Iridoid synthase family: a catalytic reservoir for specialized metabolism across land plants. *ACS Chem. Biol.* 15, 1780–1787. doi: 10.1021/acschembio.0c00220

O'Connor, S. E. (2010). "Alkaloids," in *Comprehensive Natural Products II*. eds. H.-W. Liu and L. Mander (Amsterdam, Netherlands: Elsevier), 977–1007.

Polturak, G., and Aharoni, A. (2018). "La vie en rose": biosynthesis, sources, and applications of betalain pigments. *Mol. Plant* 11, 7–22. doi: 10.1016/j.molp.2017.10.008

Qu, Y., Easson, M. E. A. M., Simionescu, R., Hajicek, J., Thamm, A. M. K., Salim, V., et al. (2018a). Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19E-geissoschizine. *Proc. Natl. Acad. Sci.* 115, 3180–3185. doi: 10.1073/pnas.1719979115

Qu, Y., Easson, M. L. A. E., Froese, J., Simionescu, R., Hudlicky, T., and DeLuca, V. (2015). Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6224–6229. doi: 10.1073/pnas.1501821112

Qu, Y., Thamm, A. M. K., Czerwinski, M., Masada, S., Kim, K. H., Jones, G., et al. (2018b). Geissoschizine synthase controls flux in the formation of monoterpenoid indole alkaloids in a *Catharanthus roseus* mutant. *Planta* 247, 625–634. doi: 10.1007/s00425-017-2812-7

Robins, R. J., Chesters, N. C. J. E., O'Hagan, D., Parr, A. J., Walton, N. J., and Woolley, J. G. (1995). The biosynthesis of hyoscyamine: the process by which littorine rearranges to hyoscyamine. *J. Chem. Soc. Perkin Trans.* 1, 481–485. doi: 10.1039/p19950000481

Rodríguez-López, C. E., Hong, B., Paetz, C., Nakamura, Y., Koudounas, K., Passeri, V., et al. (2021). Two bi-functional cytochrome P450 CYP72 enzymes from olive (*Olea europaea*) catalyze the oxidative C–C bond cleavage in the biosynthesis of secoxy-iridoids – flavor and quality determinants in olive oil. *New Phytol.* 229, 2288–2301. doi: 10.1111/nph.16975

Rueffer, M., and Zenk, M. H. (1994). Canadine synthase from *Thalictrum tuberosum* cell cultures catalyses the formation of the methylenedioxy bridge in berberine synthesis. *Phytochemistry* 36, 1219–1223. doi: 10.1016/S0031-9422(00)89640-5

Rueffer, M., and Zenk, M. H. (1998). Microsome-mediated transformation of *O*-methylandrocymbine to demecolcine and colchicine. *FEBS Lett.* 438, 111–113. doi: 10.1016/S0014-5793(98)01282-4

Sakaki, T. (2012). Practical application of cytochrome P450. *Biol. Pharm. Bull.* 35, 844–849. doi: 10.1248/bpb.35.844

Schmidt, D., and Stöckigt, J. (1995). Enzymatic formation of the sarpagan-bridge: a key step in the biosynthesis of sarpagan- and ajmaline-type alkaloids. *Planta Med.* 61, 254–258. doi: 10.1055/s-2006-958067

Schnabel, A., Cotinguiba, F., Athmer, B., and Vogt, T. (2021). *Piper nigrum* CYP719A37 catalyzes the decisive methylenedioxy bridge formation in piperine biosynthesis. *Plan. Theory* 10, 1–15. doi: 10.3390/plants10010128

Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V., et al. (1999). Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. *FEBS Lett.* 458, 97–102. doi: 10.1016/S0014-5793(99)01138-2

Shang, Y., and Huang, S. (2020). Engineering plant cytochrome P450s for enhanced synthesis of natural products: past achievements and future perspectives. *Plant Commun.* 1:100012. doi: 10.1016/j.xplc.2019.100012

St-Pierre, B., and De Luca, V. (1995). A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus*. *Plant Physiol.* 109, 131–139. doi: 10.1104/pp.109.1.131

Sun, L., Ruppert, M., Sheludko, Y., Warzecha, H., Zhao, Y., and Stöckigt, J. (2008). Purification, cloning, functional expression and characterization of perakine reductase: the first example from the AKR enzyme family, extending the alkaloidal network of the plant *Rauvolfia*. *Plant Mol. Biol.* 67, 455–467. doi: 10.1007/s11103-008-9331-7

Takemura, T., Ikezawa, N., Iwasa, K., and Sato, F. (2013). Molecular cloning and characterization of a cytochrome P450 in sanguinarine biosynthesis from *Eschscholzia californica* cells. *Phytochemistry* 91, 100–108. doi: 10.1016/j.phytochem.2012.02.013

Tang, M. C., Zou, Y., Watanabe, K., Walsh, C. T., and Tang, Y. (2017). Oxidative cyclization in natural product biosynthesis. *Chem. Rev.* 117, 5226–5333. doi: 10.1021/acs.chemrev.6b00478

Tatsis, E. C., Carqueijeiro, I., Dugé De Bernonville, T., Franke, J., Dang, T. T. T., Oudin, A., et al. (2017). A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate. *Nat. Commun.* 8, 316–319. doi: 10.1038/s41467-017-00154-x

Tawfik, O. K. A. D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79, 471–505. doi: 10.1146/annurev-biochem-030409-143718

Taylor, W. I., Frey, A. J., and Hofmann, A. (1962). Vomilenin und sein Umwandlung in Perakin. *Helv. Chim. Acta* 45, 611–614. doi: 10.1002/hlca.19620450225

Thamm, A. M. K., Qu, Y., and De Luca, V. (2016). Discovery and metabolic engineering of iridoid/secoiridoid and monoterpenoid indole alkaloid biosynthesis. *Phytochem. Rev.* 15, 339–361. doi: 10.1007/s11101-016-9468-y

Thodey, K., Galanie, S., and Smolke, C. D. (2014). A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* 10, 837–844. doi: 10.1038/nchembio.1613

Turk, E. M., Fujioka, S., Seto, H., Shimada, Y., Takatsuto, S., Yoshida, S., et al. (2003). CYP72B1 inactivates brassinosteroid hormones: an intersection between photomorphogenesis and plant steroid signal transduction. *Plant Physiol.* 133, 1643–1653. doi: 10.1104/pp.103.030882

Vetter, H. P., Mangold, U., Schröder, G., Marner, F. J., Werck-Reichhart, D., and Schröder, J. (1992). Molecular analysis and heterologous expression of an inducible cytochrome P-450 protein from periwinkle (*Catharanthus roseus* L). *Plant Physiol.* 100, 998–1007. doi: 10.1104/pp.100.2.998

Watanabe, I., Nara, F., and Serizawa, N. (1995). Cloning, characterization and expression of the gene encoding cytochrome P-450sca-in2 from *Streptomyces carbophilus* involved in production of pravastatin, a specific HMG-CoA reductase inhibitor. *Gene* 163, 81–85. doi: 10.1016/0378-1119(95)00394-L

Weng, J.-K., Li, Y., Mo, H., and Chapple, C. (2012). Assembly of an evolutionarily new pathway for α-pyrone biosynthesis in *Arabidopsis*. *Science* 337, 960–964. doi: 10.1126/science.1221614

Wenkert, E., and Wickberg, B. (1965). General methods of synthesis of indole alkaloids. *J. Am. Chem. Soc.* 87, 1580–1589. doi: 10.1021/ja01085a029

Wickramasinghe, R. H., and Villee, C. A. (1975). Early role during chemical evolution for cytochrome P450 in oxygen detoxification. *Nature* 256, 509–511. doi: 10.1038/256509a0

Williams, D., Qu, Y., Simionescu, R., and De Luca, V. (2019). The assembly of (+)-vincadifformine- and (−)-tabersonine-derived monoterpenoid indole alkaloids in *Catharanthus roseus* involves separate branch pathways. *Plant J.* 99, 626–636. doi: 10.1111/tpj.14346

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., et al. (2015). Morphinan biosynthesis in opium requires a P450-oxidoreductase poppy requires a P450-oxidoreductase fusion protein fusion protein. *Science* 349, 309–312. doi: 10.1126/science.aab1852

Yang, Y., Li, W., Pang, J., Jiang, L., Qu, X., Pu, X., et al. (2019). Bifunctional cytochrome P450 enzymes involved in camptothecin biosynthesis. *ACS Chem. Biol.* 14, 1091–1096. doi: 10.1021/acschembio.8b01124

Zenk, M. H., Gerardy, R., and Stadler, R. (1989). Phenol oxidative coupling of benzylisoquinoline alkaloids is catalysed by regio- and stereo-selective cytochrome P-450 linked plant enzymes: salutaridine and berbamunine. *J. Chem. Soc. Chem. Commun.* 22, 1725–1727. doi: 10.1039/C39890001725%0A

Zenkner, F. F., Margis-Pinheiro, M., and Cagliari, A. (2019). Nicotine biosynthesis in Nicotiana: a metabolic overview. *Tob. Sci.* 56, 1–9. doi: 10.3381/18-063

Zhang, B., Lewis, K. M., Abril, A., Davydov, D. R., Vermerris, W., Sattler, S. E., et al. (2020). Structure and function of the cytochrome p450 monooxygenase cinnamate 4-hydroxylase from *Sorghum bicolor*. *Plant Physiol.* 183, 957–973. doi: 10.1104/pp.20.00406

Zhang, X., Chen, Y., Gao, B., Luo, D., Wen, Y., and Ma, X. (2015). Apoptotic effect of koumine on human breast cancer cells and the mechanism involved. *Cell Biochem. Biophys.* 72, 411–416. doi: 10.1007/s12013-014-0479-2

Zhang, X., and Li, S. (2017). Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* 34, 1061–1089. doi: 10.1039/C7NP00028F

Zhao, B., Lei, L., Vassylyev, D. G., Lin, X., Cane, D. E., Kelly, S. L., et al. (2009). Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *J. Biol. Chem.* 284, 36711–36719. doi: 10.1074/jbc.M109.064683

Ziegler, J., and Facchini, P. J. (2008). Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* 59, 735–769. doi: 10.1146/annurev.arplant.59.032607.092730

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

frontiers in Plant Science

MINI REVIEW
published: 19 July 2021
doi: 10.3389/fpls.2021.699157

Isoprenoid Metabolism and Engineering in Glandular Trichomes of *Lamiaceae*

Soheil S. Mahmoud, Savanna Maddock and Ayelign M. Adal*

Department of Biology, The University of British Columbia, Kelowna, BC, Canada

The isoprenoids play important ecological and physiological roles in plants. They also have a tremendous impact on human lives as food additives, medicines, and industrial raw materials, among others. Though some isoprenoids are highly abundant in nature, plants produce many at extremely low levels. Glandular trichomes (GT), which cover the aerial parts of more than 25% of vascular plants, have been considered as natural biofactories for the mass production of rare industrially important isoprenoids. In several plant genera (e.g., *Lavandula* and *Mentha*), GTs produce and store large quantities of the low molecular weight isoprenoids, in particular mono- and sesquiterpenes, as essential oil constituents. Within each trichome, a group of secretory cells is specialized to strongly and specifically express isoprenoid biosynthetic genes, and to synthesize and deposit copious amounts of terpenoids into the trichome's storage reservoir. Despite the abundance of certain metabolites in essential oils and defensive resins, plants, particularly those lacking glandular trichomes, accumulate small quantities of many of the biologically active and industrially important isoprenoids. Therefore, there is a pressing need for technologies to enable the mass production of such metabolites, and to help meet the ever-increasing demand for plant-based bioproducts, including medicines and renewable materials. Considerable contemporary research has focused on engineering isoprenoid metabolism in GTs, with the goal of utilizing them as natural biofactories for the production of valuable phytochemicals. In this review, we summarize recent advances related to the engineering of isoprenoid biosynthetic pathways in glandular trichomes.

Keywords: glandular trichomes, metabolic engineering, isoprenoids, mint, lavender

OPEN ACCESS

Edited by:
Yang Zhang,
Sichuan University, China

Reviewed by:
Raimund Nagel,
University of Leipzig, Germany
Clarice Noleto-Dias,
Rothamsted Research,
United Kingdom

****Correspondence:***
Soheil S. Mahmoud
soheil.mahmoud@ubc.ca

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

Received: *23 April 2021*
Accepted: *30 June 2021*
Published: *19 July 2021*

Citation:
Mahmoud SS, Maddock S and Adal AM (2021) Isoprenoid Metabolism and Engineering in Glandular Trichomes of Lamiaceae. Front. Plant Sci. 12:699157. doi: 10.3389/fpls.2021.699157

ISOPRENOID DIVERSITY AND BIOSYNTHESIS

The isoprenoids or terpenoids, make up the largest class of plant secondary, or specialized metabolites. They play crucial ecological roles as pollinator attractants and defensive agents, and have important physiological functions as plant hormones and photosynthetic pigments, among others (Gershenzon and Dudareva, 2007; Okada, 2011; Tetali, 2018). Isoprenoids impact human lives through imparting scent, flavor, and health-promoting properties to fruits, vegetables, and medicinal plants (Cordell and Colvard, 2012). Some terpenoids have potent biological activities, and have applications as prescription drugs and over-the-counter medicines. A few

have also been used as sustainable replacements for petroleum-derived chemicals (Cordell and Colvard, 2012; Tetali, 2018).

Isoprenoids are synthesized through the condensation of isopentenyl diphosphate (IPP; C_5) and its isomer dimethylallyl diphosphate (DMAPP; C_5), and are classified by the number of five-carbon units present in the core structure (Bouvier et al., 2005; Defilippi et al., 2009). Major isoprenoid classes include monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpenes (C_{30}), and tetraterpenes (C_{40}), although lower and higher-order isoprenoids (e.g., isoprene and natural rubber, respectively) also exist (**Figure 1**). In general, the biosynthesis of isoprenoids can be classified into the following four states: generation of general precursors IPP and DMAPP, production of specific isoprenyl diphosphates for various isoprenoid classes, the transformation of isoprenyl diphosphates to individual isoprenoids by terpene synthase (TPS) enzymes, and structural modifications catalyzed by other catalysts (**Figure 1**).

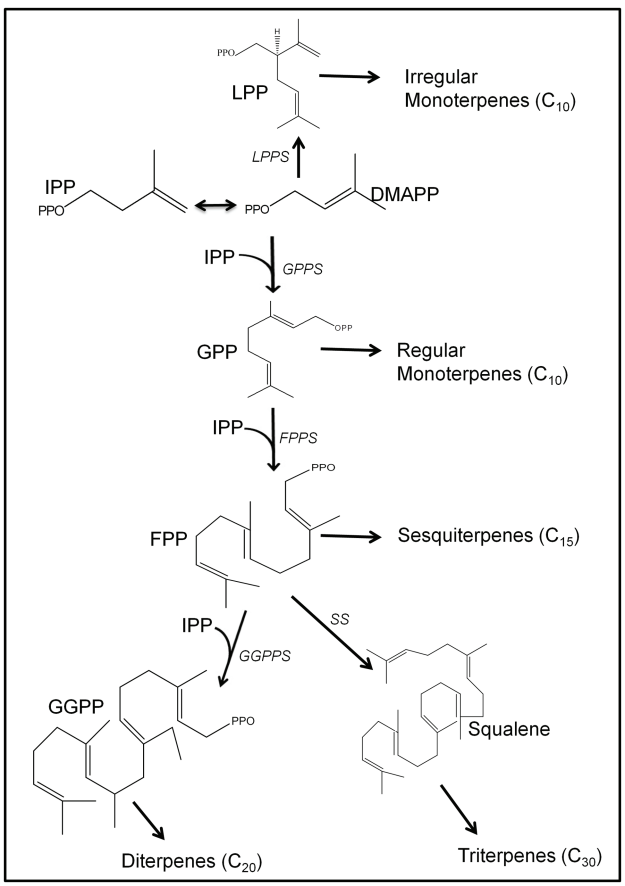


FIGURE 1 | Overview of the biosynthetic pathways for major plant isoprenoids. The biosynthesis of isoprenoids starts with a common pool of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). IPP and DAMPP are initially condensed to form either lavandulyl diphosphate (LPP), the precursor for irregular monoterpenes, or geranyl diphosphate (GPP), the precursor to regular monoterpenes. GPP can also be condensed with one or two IPP units to form FPP for sesquiterpene synthesis, or geranylgeranyl diphosphate (GGPP) for diterpene production. Two FPP units can be condensed to form squalene, the immediate precursor for triterpene metabolism. The isoprenyl diphosphate synthases (IDSs) that catalyzed the precursor(s) into specific prenyl diphosphate include LPP synthase (LPPS), GPP synthase (GPPS), geranylgeranyl diphosphate synthase (GGPPS), FPP synthase (FPPS), and squalene synthase (SS).

In plants, IPP and DMAPP are generated through two distinct compartmentally separated pathways. In plastids, the 1-DXP pathway, aka MEP pathway, produces IPP/DMAPP for monoterpene, diterpene, and tetraterpene biosynthesis (Lois et al., 2000; Turner et al., 2000). In the cytosol, the classical mevalonic acid (MVA) pathway supplies IPP, which is isomerized to DMAPP by an isomerase. The cytosolic IPP/DMAPP pool is primarily used for the production of sesqui- and triterpenes (Laule et al., 2003). All biochemical steps of both pathways have been characterized, and the relevant genes cloned (Tholl and Lee, 2011).

The head-to-tail condensation of IPP and DMAPP yields various linear isoprenoid precursors, including geranyl diphosphate (GPP; C_{10}), neryl diphosphate (NPP; C_{10}), farnesyl diphosphate (FPP; C_{15}), and geranylgeranyl diphosphate (GGPP; C_{20}). The head-to-middle condensation of two DMAPP units also occurs in some plants, and gives rise to lavandulyl diphosphate (LPP; C_{10}) as a linear product. The condensation reactions are catalyzed by a group of enzymes known as isoprenyl diphosphate synthases (IDSs; Ogura and Koyama, 1998; Burke et al., 1999; Schilmiller et al., 2009; Tholl and Lee, 2011; Demissie et al., 2013; Dudareva et al., 2013; Adal and Mahmoud, 2020). IDSs are classificed into "*cis*" or "*trans*" based on their primary structures, and the stereochemistry of the products they generate (Nagel et al., 2019). The *trans*-IDSs catalyze the synthesis of the common terpene precursors, and are distinguished by two conserved aspartate-rich motifs, $DDX_{2-4}D$ and $(N/D)DX_2D$, which serve as substrate and divalent metal ion cofactor binding sites. On the other hand, *cis*-IDSs lack these conserved motifs, and instead share five conserved regions designated as Regions I–V, including catalytically essential aspartate residue in Region IV and the glutamate residue in Region V (Kharel and Koyama, 2003). For example, GPPS, a typical *trans*-IDS, catalyze the head-to-tail condensation of DMAPP with IPP to produce GPP, while LPPS, a *cis*-IDS, generates LPP through the head-to-middle condensation of two DMAPP units in lavenders (Demissie et al., 2013; Adal and Mahmoud, 2020).

Other *trans*-IDSs include FPP, which gives rise to sesquiterpenes and triterpenes, and GGPP, which serves as a precursor for the synthesis of diterpenes and tetraterpenes. The prenyl diphosphates are transformed into various terpenoids by specific TPSs, of which hundreds have been cloned from a wide range of plants.

GLANDULAR TRICHOMES

Glandular trichomes (GTs) can be found on the surfaces of leaves, stems, petals, sepals, petioles, peduncles, and seeds of ca. 30% of vascular plants (Fahn, 2000; Wagner et al., 2004; Glas et al., 2012). Regardless of their location on the plant, GTs are multicellular structures, each consisting of a basal cell, one or more stalk cells, and a group of 4–8 (depending

on species) secretory cells (**Figure 2**; Fahn, 2000; Werker, 2000; Schnittger and Hülskamp, 2002; Huchelmann et al., 2017). A typical secretory cell is non-photosynthetic, and is specialized to produce relatively large amounts of specialized metabolites (Fahn, 2000). In general, two types of glandular trichomes – capitate GT and peltate GT – can often be found in several plant families, including *Lamiaceae* and *Solanaceae*. Both GT types share certain structural features (Fahn, 2000). For example, both include a basal cell, a stalk that can be made up of one to several cells, and a group of secretory cells that are clustered together on the apex of the stalk. However, they produce different classes of metabolites. The capitate GTs mainly produce non-volatile compounds, that are not stored in the trichome but are mostly exuded and accumulate on the surface of the trichome (Glas et al., 2012). Peltate GTs, on the other hand, develop a storage cavity capable of storing relatively large quantities of primarily volatile compounds (Glas et al., 2012; Huchelmann et al., 2017). The latter are responsible for the production and storage of essential oil constituents in members of the *Lamiaceae*, including peppermint, spearmint, lavender, basil, and so forth.

ENGINEERING ISOPRENOID BIOSYNTHESIS IN TRICHOMES

Several platforms, including bacteria (Farrokh et al., 2019), yeast (Rahmat and Kang, 2020), algae (Rico et al., 2017), and plants (Birchfield and McIntosh, 2020), have been successfully used for enhanced (or mass) production of specific isoprenoids of industrial value. Though traditionally microbial systems have been the platform of choice for the production of recombinant natural products, plants are emerging as important alternatives.

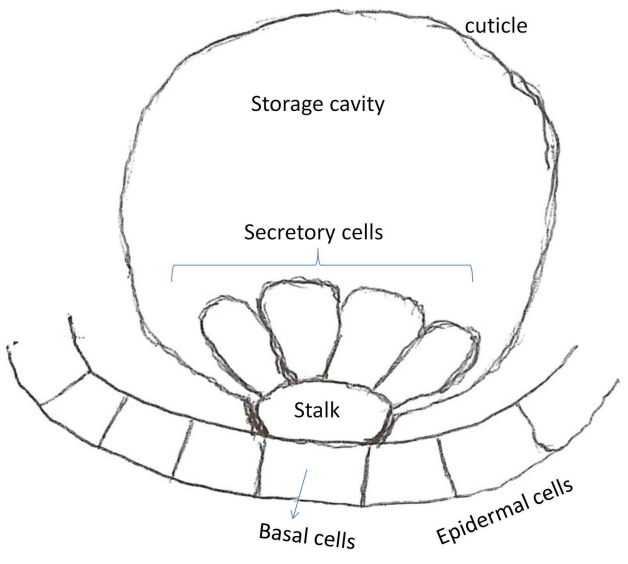


FIGURE 2 | Schematic representation of peltate glandular trichome found in *Lamiaceae* plants. The secretory cells produce and secret essential oil constituents into the trichome's secretory cavity.

This is in part due to the relative ease and cost-effectiveness of mass-producing plants in general, and in part due to the fact that genetic transformation of most plants is now routine. Plants bearing GTs are particularly suited for mass production of secondary metabolites, as GTs have specialized secretory cells, capable of production and secretion of large quantities of phytochemicals (that are often toxic to other cell types) into the GT's storage cavity. In this context, several GT-bearing plants have been investigated as potential biofactories for the mass production of specific metabolites, with most of the studies reported focussing on the plants discussed below.

Mints (*Mentha*)

Metabolic engineering to enhance isoprenoid (essential oil monoterpenes) metabolism has been reported in the two most commercially important mint species, *Mentha x piperita* (peppermint) and *Mentha spicata* (spearmint), which are widely grown for EO production. The EO in these plants is a complex mixture of mainly monoterpenes, many of which are highly abundant while others are present in trace quantities. In both species, GPP is initially converted to (−)-limonene through a reaction catalyzed by the enzyme (−)-limonene synthase (**Figure 3**). In peppermint, hydroxylation at the C-3 position of the limonene ring initiates a cascade of reactions that lead to the production of several monoterpenes, of which (−)-menthol is highly valued for its applications in food and alternative medicine. In spearmint, hydroxylation occurs on the C-6 position of limonene, leading to the production of (−)-*trans*-carveol, which is efficiently oxidized to (−)-carvone as the main EO constituent.

One of the earliest metabolic engineering efforts in mint was reported in 2001, in a study that attempted to increase flux toward monoterpene metabolism by overexpressing the coding sequence for the branch-point enzyme 4S-limonene synthase (*4S-LimS*) in transformed *M. x piperita* plants (Diemer et al., 2001). The expression of *4S-LimS* transgene was driven by the Cauliflower Mosaic Virus (CaMV) 35S promoter. Although monoterpene production was not altered in most plants, a few transformants accumulated higher levels of essential oil constituents, including 56% more pinene, 22–74% higher 1,8-cineole, and 18–40% more pulegone than control plants. A few attempts followed this investigation in Professor Rodney Croteau's laboratory in the early 2000s. In one attempt, the DXP reductoisomerase (DXR) cDNA (**Figure 1**) was overexpressed in peppermint plants to enhance the output of the DXP pathway, and improve the production of precursor (IPP/DMAPP) for monoterpene biosynthesis (Mahmoud and Croteau, 2001). Several transgenic plants produced up to 50% more EO than the wild-type controls. In the same study, the menthofuran synthase gene, responsible for the production of the undesired oil monoterpene constituent (+)-Menthofuran (**Figure 3**), was overexpressed in separate peppermint plants in sense and antisense to evaluate the effects of these overexpressions on menthofuran production (Mahmoud and Croteau, 2001). Most plants expressing the transgene in sense accumulated substantially more menthofuran than wild-type controls. Conversely, plants that overexpressed the gene in

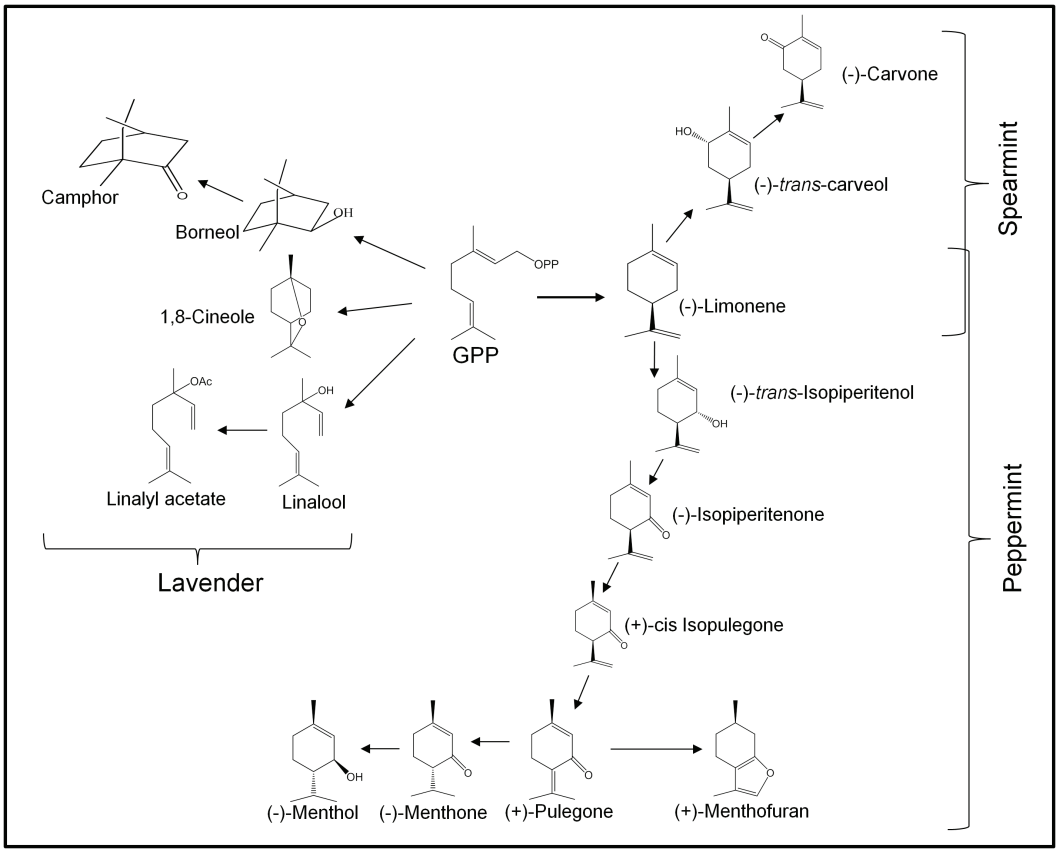


FIGURE 3 | The biosynthetic pathways for monoterpene metabolism in lavender, peppermint, and spearmint. Representative monoterpenes derived from GPP are shown, and the corresponding enzymes involved in the process are previously reported for each species (Croteau et al., 2005; Ringer et al., 2005; Landmann et al., 2007; Demissie et al., 2012; Adal et al., 2019).

antisense orientation, accumulated significantly less menthofuran than control plants. Intriguingly, menthofuran production was almost entirely eliminated in plants in which the menthofuran synthase gene was co-suppressed. Surprisingly, growth and development in co-suppressed plants were compromised as some of the most affected plants appeared bleached, in particular, under stress conditions. Another study that aimed to enhance monoterpene production (oil yield) in peppermint, limonene synthase, and limonene-3-hydroxylase cDNAs were overexpressed in independently (separately) transformed peppermint plants (Mahmoud et al., 2004). Overexpression of either gene failed to improve oil yield significantly. However, limonene levels increased dramatically, surpassing 80% of total oil (compared to ~2% in wild-type plants), in the transformed peppermint lines, in which limonene-3-hydroxylase was co-suppressed. Taken together, studies in peppermint produced evidence that metabolic engineering may be employed to effectively engineer isoprenoid metabolism. However, not all metabolic steps respond to gene overexpression, while gene co-suppression can effectively eliminate unwanted metabolites (e.g., menthofuran) or enhance the accumulation of particular desired end products (e.g., limonene).

Modulating the expression of transcription factors that regulate other EO biosynthetic genes offers another avenue for enhancing isoprenoid production in GTs. A recent study investigated the effects of overexpressing and silencing the spearmint GT-specific transcription factor MsYABBY5 on the production of terpenes in spearmint (Wang et al., 2016). The TF was constitutively overexpressed in stably transformed plants, and silenced using RNA interference (RNAi). Silencing MsYABBY5 increased the levels of terpenes in spearmint from 20 to 77%, while its overexpression led to a 23–52% decrease in the overall levels of terpenes in spearmint plants. This study clearly demonstrated that the MsYABBY5 is a repressor of monoterpene metabolism in spearmint, and that manipulating the expression of transcription factors offers a viable method for enhancing terpene production in plant GTs. In a separate study, the expression of the spearmint GT-specific R2R3-MYB TF was suppressed by RNAi (Reddy et al., 2017). Likewise, suppression of MsR2R3-MYB expression led to a 2.3–4.5-fold increase in total monoterpene abundance in transgenic spearmint. It was further shown that MsR2R3-MYB suppresses the expression of the GPP synthase large subunit gene through binding of its cis-elements, and hence acts as a negative regulator of monoterpene metabolism in spearmint.

A recent study evaluated the potential of spearmint for the production of heterologous monoterpene through metabolic engineering. In this study, a transgenic spearmint line with reduced limonene and carvone was independently transformed with cDNAs of linalool synthase and myrcene synthase from *Picea abies*, and geraniol synthase from *Cananga odorata*, which were also controlled by 35S promoter. Silencing of the limonene synthase gene by RNAi led to a huge reduction in the production of limonene and carvone, and an increase in sesquiterpene, phytosterols, fatty acids, flavonoids, and phenolic metabolites. Surprisingly, overexpression of heterologous TPSs in these lines did not significantly increase the production of the related products, although small quantities of some heterologous terpenes were produced, which were oxygenated by the host plant (Li et al., 2020).

Lavenders (*Lavandula*)

Lavenders (*Lavandula*) are perennial members of the mint family. To date, ca. 34 species have been described, three of which (*L. angustifolia*, *L. latifolia*, and their natural hybrid offspring *L. x intermedia*) are widely grown for EO production. Between 1,200 and 1,500 tons of lavender, EO is produced worldwide per annum for use in foods, cosmetics, and personal care products. As in other members of the Labiatae, lavender bears peltate GT (Huang et al., 2008) that produce and store relatively large quantities of a monoterpene-rich essential oil.

The EO of a typical lavender species consists of over 50 monoterpenes, the most abundant of which include linalool, linalool acetate, borneol, camphor, and 1,8-cineole. Among these, camphor, linalool, and linalool acetate are key determinants of the quality (scent and bioactivity) of lavender EO (Lis-Balchin, 2002; Upson and Andrews, 2004). Lavender EOs with high linalool and linalool acetate, and low borneol, camphor, and 1, 8-cineole content are considered to be of "high quality," and are used in cosmetic products and aromatherapy (Cavanagh and Wilkinson, 2002). These oils are typically obtained from true lavender (*L. angustifolia*) species. Oil marketed to the alternative medicine sector is typically obtained from *L. latifolia* plants, which accumulate high levels of linalool, borneol, camphor, and 1,8-cineole, but no linalool acetate. The EOs obtained from the hybrid *L. x intermedia* plants contain a mixture of monoterpenes present in both parents, and are mainly utilized in personal care and hygiene products, such as soaps, shampoos, mouthwashes, as well as industrial and household cleaners, among others. *Lavandula x intermedia* plants produce up to 10 fold more EO than either parent, and are hence widely grown for EO production (Lis-Balchin, 2002; Upson and Andrews, 2004; Wells et al., 2018).

Metabolic engineering to enhance monoterpene biosynthesis has been reported in two lavender species, *L. latifolia* (spike lavender) and *L. x intermedia* (lavandin). The earliest metabolic engineering study in a lavender species was reported in the early 2000s, when Muñoz-Bertomeu et al. (2006) overexpressed the *Arabidopsis* 1-DXP synthase (DXS) gene in spike lavender. Like the earlier results reported for the overexpression of *DXR* and *DXS* in peppermint (Mahmoud and Croteau, 2001), *DXS* overexpression increased EO (monoterpene) production compared to untransformed controls. In another study, researchers attempted to enhance EO quality (scent) in *L. latifolia* plants by increasing the biosynthesis of *S*-linalool, a sweet-scented monoterpene (Mendoza-Poudereux et al., 2014). In this study, the *Clarkia breweri S*-linalool synthase gene was transformed into *L. latifolia* plants *via Agrobacterium*-mediated transformation. Transgenic lines overexpressing the transgene accumulated substantially (up to 1,000 fold) more *S*-linalool compared to untransformed control plants. Interestingly double transgenic plants, which resulted from a cross between separate transgenic plants overexpressing either DXS or *S*-linalool synthase gene (*LIS*) did not yield the expected results. Both essential oil yield and linalool content in double DXS-LIS transgenic plants were lower than that of their parents. A separate study also aimed at enhancing the fragrance of *L. x intermedia* essential oil by reducing 1,8-Cineole biosynthesis (**Figure 3**) by suppressing the expression of the 1,8-Cineole synthase gene using RNAi (Tsuro et al., 2019). The transcriptional expression of 1,8-Cineole synthase gene was suppressed in the transgenic plants compared to nontransgenic controls. RNAi was effective in reducing production of 1, 8-cineole significantly, and altering the overall scent of the EO of affected plants.

Finally, in a recent study, the *L. x intermedia* BPPS gene (*LiBPPS*) was placed under the control of the CaMV 35S promoter and stably expressed in transgenic *L. latifolia* plants *via Agrobacterium*-transformation in sense and antisense orientations (Adal et al., 2021, Unpublished). As expected, most plants expressing *LiBPPS* in sense produced more borneol and camphor (**Figure 3**), while those expressing *LiBPPS* in antisense accumulated less borneol and camphor than wild-type plants. Notably, the expression of *LiBPPS* in sense severely impeded the growth and development of most transformed plants, many of which (the highest transgene expressers) did not survive past early regeneration stages.

CONCLUSION AND PERSPECTIVES

In conclusion, the results of metabolic engineering efforts in mint and lavender have yielded useful information that paves the way for future investigation. The first lesson learned is that increasing precursor (IPP/DMAPP) through improving the output of the DXP pathway can help boost isoprenoid metabolism substantially in GTs. In this context, overexpression of limiting enzymes of the DXP pathway (DXS and DXR) individually has led to increased monoterpenoid production by up 350% (in spike lavender). The second lesson concerns manipulating the expression of TPS genes. Overexpression of TPSs sometimes leads to increase production of the corresponding terpene product. However, most often overexpression does not yield the desired (improved terpene production) results, clearly demonstrating that factors other than TPS expression are involved in isoprenoid production. Most of such (hypothetical) factors have not been yet defined. However, the use of GT-specific

promoters – rather than constitutive promoters such as the CaMV 35S promoter – may help resolve some of the issues associated with ectopic overexpression of genes. Further, improving other factors potentially limiting isoprenoid biosynthesis, secretion and storage may result in enhanced isoprenoid production. For example, increasing the rate of the transport of the isoprenoids (presumably mediated by lipid transporters) from secretory cells into the storage cavity of the trichome may enhance isoprenoid production. Also, increasing the size and density of glandular trichomes may increase plant capacity to produce and store isoprenoid compounds. These approaches can be examined only when specific genes/proteins that control glandular trichome development and isoprenoid transport are identified.

AUTHOR CONTRIBUTIONS

SSM initiated and wrote the manuscript. SM and AMA contributed to searching and summarizing the articles and reviewing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This review was supported through grants and/or in-kind contributions to SSM by the UBC Okanagan through the Eminence Program, and by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Adal, A. M., and Mahmoud, S. S. (2020). Short—chain isoprenyl diphosphate synthases of lavender (Lavandula). *Plant Mol. Biol.* 102, 517–535. doi: 10.1007/s11103-020-00962-8

Adal, A. M., Sarker, L. S., Malli, R. P. N., Liang, P., and Mahmoud, S. S. (2019). RNA-Seq in the discovery of a sparsely expressed scent-determining monoterpene synthase in lavender (Lavandula). *Planta* 249, 271–290. doi: 10.1007/s00425-018-2935-5

Birchfield, A. S., and McIntosh, C. A. (2020). Metabolic engineering and synthetic biology of plant natural products—A minireview. *Curr. Plant Biol.* 24:100163. doi: 10.1016/j.cpb.2020.100163

Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* 44, 357–429. doi: 10.1016/j.plipres.2005.09.003

Burke, C. C., Wildung, M. R., and Croteau, R. (1999). Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13062–13067. doi: 10.1073/pnas.96.23.13062

Cavanagh, H. M. A., and Wilkinson, J. M. (2002). Biological activities of lavender essential oil. *Phytother. Res.* 16, 301–308. doi: 10.1002/ptr.1103

Cordell, G. A., and Colvard, M. D. (2012). Natural products and traditional medicine: turning on a paradigm. *J. Nat. Prod.* 75, 514–525. doi: 10.1021/np200803m

Croteau, R. B., Davis, E. M., Ringer, K. L., and Wildung, M. R. (2005). (−)-menthol biosynthesis and molecular genetics. *Naturwissenschaften* 92, 562–577. doi: 10.1007/s00114-005-0055-0

Defilippi, B. G., Manríquez, D., Luengwilai, K., and González-Agüero, M. (2009). Chapter 1 aroma volatiles. Biosynthesis and mechanisms of modulation during fruit ripening. *Adv. Bot. Res.* 50, 1–37. doi: 10.1016/S0065-2296(08)00801-X

Demissie, Z. A., Cella, M. A., Sarker, L. S., Thompson, T. J., Rheault, M. R., and Mahmoud, S. S. (2012). Cloning, functional characterization and genomic organization of 1,8-cineole synthases from Lavandula. *Plant Mol. Biol.* 79, 393–411. doi: 10.1007/s11103-012-9920-3

Demissie, Z. A., Erland, L. A. E., Rheault, M. R., and Mahmoud, S. S. (2013). The biosynthetic origin of irregular monoterpenes in Lavandula: isolation and biochemical characterization of a novel cis-prenyl diphosphate synthase gene, lavandulyl diphosphate synthase. *J. Biol. Chem.* 288, 6333–6341. doi: 10.1074/jbc.M112.431171

Diemer, F., Caissard, J. C., Moja, S., Chalchat, J. C., and Jullien, F. (2001). Altered monoterpene composition in transgenic mint following the introduction of 4S-limonene synthase. *Plant Physiol. Biochem.* 39, 603–614. doi: 10.1016/S0981-9428(01)01273-6

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

Fahn, A. (2000). Structure and function of secretory cells. *Adv. Bot. Res.* 31, 37–75. doi: 10.1016/s0065-2296(00)31006-0

Farrokh, P., Sheikhpour, M., Kasaeian, A., Asadi, H., and Bavandi, R. (2019). Cyanobacteria as an eco-friendly resource for biofuel production: a critical review. *Biotechnol. Prog.* 35:e2835. doi: 10.1002/btpr.2835

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

Glas, J., Schimmel, B., Alba, J., Escobar-Bravo, R., Schuurink, R., and Kant, M. (2012). Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *Int. J. Mol. Sci.* 13, 17077–17103. doi: 10.3390/ijms131217077

Huang, S. S., Kirchoff, B. K., and Liao, J. P. (2008). The capitate and peltate glandular trichomes of *Lavandula pinnata L.* (*Lamiaceae*): histochemistry, ultrastructure, and secretion1. *J. Torrey Bot. Soc.* 135, 155–167. doi: 10.3159/07-RA-045.1

Huchelmann, A., Boutry, M., and Hachez, C. (2017). Plant glandular trichomes: natural cell factories of high biotechnological interest. *Plant Physiol.* 175, 6–22. doi: 10.1104/pp.17.00727

Kharel, Y., and Koyama, T. (2003). Molecular analysis of cis-prenyl chain elongating enzymes. *Nat. Prod. Rep.* 20, 111–118. doi: 10.1039/b108934j

Landmann, C., Fink, B., Festner, M., Dregus, M., Engel, K. H., and Schwab, W. (2007). Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). *Arch. Biochem. Biophys.* 465, 417–429. doi: 10.1016/j.abb.2007.06.011

Laule, O., Fürholz, A., Chang, H. S., Zhu, T., Wang, X., Heifetz, P. B., et al. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6866–6871. doi: 10.1073/pnas.1031755100

Li, C., Sarangapani, S., Wang, Q., Nadimuthu, K., and Sarojam, R. (2020). Metabolic engineering of the native monoterpene pathway in spearmint for production of heterologous monoterpenes reveals complex metabolism and pathway interactions. *Int. J. Mol. Sci.* 21:6164. doi: 10.3390/ijms21176164

Lis-Balchin, M. (2002). *Lavender: The genus Lavandula*. New York: Tayler and Francis Inc.

Lois, L. M., Rodríguez-Concepción, M., Gallego, F., Campos, N., and Boronat, A. (2000). Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. *Plant J.* 22, 503–513. doi: 10.1046/j.1365-313X.2000.00764.x

Mahmoud, S. S., and Croteau, R. B. (2001). Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8915–8920. doi: 10.1073/pnas.141237298

Mahmoud, S. S., Williams, M., and Croteau, R. (2004). Cosuppression of limonene-3-hydroxylase in peppermint promotes accumulation of limonene in the essential oil. *Phytochemistry* 65, 547–554. doi: 10.1016/j.phytochem.2004.01.005

Mendoza-Poudereux, I., Muñoz-Bertomeu, J., Navarro, A., Arrillaga, I., and Segura, J. (2014). Enhanced levels of S-linalool by metabolic engineering of the terpenoid pathway in spike lavender leaves. *Metab. Eng.* 23, 136–144. doi: 10.1016/j.ymben.2014.03.003

Muñoz-Bertomeu, J., Arrillaga, I., Ros, R., and Segura, J. (2006). Up-regulation of 1-Deoxy-D-xylulose-5-phosphate synthase enhances production of essential

oils in transgenic spike lavender. *Plant Physiol.* 142, 890–900. doi: 10.1104/pp.106.086355

Nagel, R., Schmidt, A., and Peters, R. J. (2019). Isoprenyl diphosphate synthases: the chain length determining step in terpene biosynthesis. *Planta* 249, 9–20. doi: 10.1007/s00425-018-3052-1

Ogura, K., and Koyama, T. (1998). Enzymatic aspects of isoprenoid chain elongation. *Chem. Rev.* 98, 1263–1276. doi: 10.1021/cr9600464

Okada, K. (2011). The biosynthesis of isoprenoids and the mechanisms regulating it in plants. *Biosci. Biotechnol. Biochem.* 75, 1219–1225. doi: 10.1271/bbb.110228

Rahmat, E., and Kang, Y. (2020). Yeast metabolic engineering for the production of pharmaceutically important secondary metabolites. *Appl. Microbiol. Biotechnol.* 104, 4659–4674. doi: 10.1007/s00253-020-10587-y

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

Rico, M., González, A. G., Santana-Casiano, M., González-Dávila, M., Pérez-Almeida, N., and de Tangil, M. S. (2017). "Production of primary and secondary metabolites using algae" in *Prospects and Challenges in Algal Biotechnology*. eds. B. Tripathi and D. Kumar (Singapore: Springer).

Ringer, K. L., Davis, E. M., and Croteau, R. (2005). Monoterpene metabolism. Cloning, expression, and characterization of (−)-isopiperitenol/(−)-carveol dehydrogenase of peppermint and spearmint. *Plant Physiol.* 137, 863–872. doi: 10.1104/pp.104.053298

Schilmiller, A. L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A. L., Schmidt, A., et al. (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10865–10870. doi: 10.1073/pnas.0904113106\r0904113106

Schnittger, A., and Hülskamp, M. (2002). Trichome morphogenesis: a cell-cycle perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 357, 823–826. doi: 10.1098/rstb.2002.1087

Tetali, S. D. (2018). Terpenes and isoprenoids: a wealth of compounds for global use. *Planta* 249, 1–8. doi: 10.1007/s00425-018-3056-x

Tholl, D., and Lee, S. (2011). Terpene specialized metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* 9:e0143. doi: 10.1199/tab.0143

Tsuro, M., Tomomatsu, K., Inukai, C., Tujii, S., and Asada, S. (2019). RNAi targeting the gene for 1,8-cineole synthase induces recomposition of leaf essential oil in lavandin (*Lavandula* × intermedia Emeric). *Vitr. Cell. Dev. Biol. Plant* 55, 165–171. doi: 10.1007/s11627-018-09949-z

Turner, G. W., Gershenzon, J., and Croteau, R. B. (2000). Development of peltate glandular trichomes of peppermint. *Plant Physiol.* 124, 665–679. doi: 10.1104/pp.124.2.665

Upson, T., and Andrews, S. (2004). *The Genus Lavandula*. Oregon: Timber Press.

Wagner, G. J., Wang, E., and Shepherd, R. W. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* 93, 3–11. doi: 10.1093/aob/mch011

Wang, Q., Reddy, V. A., Panicker, D., Mao, H. Z., Kumar, N., Rajan, C., et al. (2016). Metabolic engineering of terpene biosynthesis in plants using a trichome-specific transcription factor MsYABBY5 from spearmint (*Mentha spicata*). *Plant Biotechnol. J.* 14, 1619–1632. doi: 10.1111/pbi.12525

Wells, R., Truong, F., Adal, A. M., Sarker, L. S., and Mahmoud, S. S. (2018). *Lavandula* essential oils: a current review of applications in medicinal, food, and cosmetic industries of lavender. *Nat. Prod. Commun.* 13, 1403–1417. doi: 10.1177/1934578X1801301038

Werker, E. (2000). Trichome diversity and development department of botany, the hebrew university of Jerusalem. *Adv. Bot. Res.* 31, 1–35. doi: 10.1016/S0065-2296(00)31005-9

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fpls.2021.709197

Integration of miRNAs, Degradome, and Transcriptome Omics Uncovers a Complex Regulatory Network and Provides Insights Into Lipid and Fatty Acid Synthesis During Sesame Seed Development

Yin-Ping Zhang[1†], Yuan-Yuan Zhang[2†], Kiran Thakur[2], Fan Zhang[2], Fei Hu[2], Jian-Guo Zhang[2], Peng-Cheng Wei[3,4] and Zhao-Jun Wei[2*]*

[1] Anhui Academy of Agricultural Sciences, Crop Research Institute, Hefei, China, [2] School of Food and Biological Engineering, Hefei University of Technology, Hefei, China, [3] College of Agronomy, Anhui Agricultural University, Hefei, China, [4] Key Laboratory of Rice Genetic Breeding of Anhui Province, Rice Research Institute, Anhui Academy of Agricultural Sciences, Hefei, China

Sesame (*Sesamum indicum* L.) has always been known as a health-promoting oilseed crop because of its nutrient-rich oil. In recent years, studies have focused on lipid and fatty acid (FA) biosynthesis in various plants by high-throughput sequencing. Here, we integrated transcriptomics, small RNAs, and the degradome to establish a comprehensive reserve intensive on key regulatory micro RNA (miRNA)-targeting circuits to better understand the transcriptional and translational regulation of the oil biosynthesis mechanism in sesame seed development. Deep sequencing was performed to differentially express 220 miRNAs, including 65 novel miRNAs, in different developmental periods of seeds. GO and integrated KEGG analysis revealed 32 pairs of miRNA targets with negatively correlated expression profiles, of which 12 miRNA-target pairs were further confirmed by RT-PCR. In addition, a regulatory co-expression network was constructed based on the differentially expressed gene (DEG) profiles. The *FAD2*, *LOC10515945*, *LOC105161564*, and *LOC105162196* genes were clustered into groups that regulate the accumulation of unsaturated fatty acid (UFA) biosynthesis. The results provide a unique advanced molecular platform for the study of lipid and FA biosynthesis, and this study may serve as a new theoretical reference to obtain increased levels of UFA from higher-quality sesame seed cultivars and other plants.

Keywords: sesame, lipid fatty acid, biosynthesis, network biology, seed development

OPEN ACCESS

Edited by:
Yang Zhang,
Sichuan University, China

Reviewed by:
Mingjie Chen,
Xinyang Normal University, China
Kang Wei,
Tea Research Institute, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Peng-Cheng Wei
weipengcheng@gmail.com
Zhao-Jun Wei
zjwei@hfut.edu.cn

†These authors share first authorship

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

Received: *13 May 2021*
Accepted: *30 June 2021*
Published: *29 July 2021*

Citation:
Zhang Y-P, Zhang Y-Y, Thakur K, Zhang F, Hu F, Zhang J-G, Wei P-C and Wei Z-J (2021) Integration of miRNAs, Degradome, and Transcriptome Omics Uncovers a Complex Regulatory Network and Provides Insights Into Lipid and Fatty Acid Synthesis During Sesame Seed Development. Front. Plant Sci. 12:709197. doi: 10.3389/fpls.2021.709197

INTRODUCTION

Sesame (*Sesamum indicum* L.), as one of the oldest herbaceous diploid oilseed plants (2n = 26) in the Pedaliaceae family with a genome size of approximately 369 Mb, is mainly cultivated in tropical and subtropical regions of Asia, Africa, and South America (Wei X. et al., 2016; Majdalawieh et al., 2020). Among the major cultivated oil crops worldwide, sesame as a commercial material contains

the highest number of fatty acids (FAs) (Wacal et al., 2019; Zhang et al., 2020), and among the FA types, linoleic, oleic, palmitic, and stearic acids are the main components that make up the total FA configuration of sesame oil (Elleuch et al., 2007; Adebowale et al., 2011; Bhunia et al., 2015). In previous studies, unsaturated fatty acids (UFAs) are considered as a quality indicator of edible oils in the core area of research for the development of functional and industrial products (Wang et al., 2016; Ahmed et al., 2020), which can prevent cardiovascular sclerosis and reduce blood pressure and blood lipids (Li et al., 2015). Therefore, increasing the UFAs content is a conventional way to improve the quality of sesame varieties and increase the seed oil content, which are also identified as the most important strategies to meet the current needs of consumers (Wang D.D. et al., 2018).

With the advent of high-throughput sequencing, genome breeding has developed rapidly, such as quantitative genetics, genetic breeding, genetic relatedness, and diversity (Wei X. et al., 2016; Yan et al., 2016; Kumar et al., 2017). In recent years, although the whole transcriptome sequencing of sesame has been described, information on the role of micro RNAs (miRNAs) and target genes, especially miRNA expression profiles involved in lipid and FA biosynthesis of sesame oil, is still poorly understood. miRNAs as endogenous non-coding RNAs (18–26 nt) regulate gene expression through post-transcriptional translational inhibition or direct cleavage of transcripts based on complementarity between miRNA and its target gene (La Sala et al., 2018; Sevgi, 2018). Recent studies have demonstrated the role of miRNAs in almost all life activities, such as biosynthesis, cell reproduction, plant resistance to biotic and abiotic stresses, and metabolic pathways (Han et al., 2016; Zhao et al., 2018, 2020; Zheng et al., 2019). In banana fruits, 128 known miRNAs belonging to 42 miRNA families were determined, and 12 new miRNAs were identified; in olive, 135 conserved miRNAs were identified and 38 putative new miRNAs were discovered; in strawberry, a total of 88 known miRNAs and 103 targets cleaved by 19 known miRNAs families were discovered (Xu et al., 2013; Yanik et al., 2013; Dan et al., 2018). Notably, from 18 sesame libraries, 351 known and 91 novel miRNAs, 116 miRNAs were reported to regulate salt stress response (Zhang et al., 2020). Similarly, major miRNA-target pairs and the complex regulatory network responsible for terpenoid biosynthesis in tea leaves (Zhao et al., 2018) and cadmium phytoremediation in hyperaccumulator *Sedum alfredii* were revealed by integrating small RNAs, degradome, and transcriptomics (Han et al., 2016). Subsequently, the lipid biosynthetic pathway and key periods of FA biogenesis and associated proteins were deciphered in the developing endosperm of *Cocos nucifera* (Reynolds et al., 2019) and tree peony seeds (Yin et al., 2018), respectively. Similarly, regulatory networks of lipid and FA metabolism were revealed in *Carya cathayensis* Sarg (Huang et al., 2016, 2021) and oil palm (Zheng et al., 2019) by transcriptomic and miRNA sequencing, respectively.

From the previous studies, it can be speculated that the integration of transcriptomic and lipidomic studies can provide a comprehensive view of the dynamic processes regulating the seed development and oil biosynthesis in various plants (Dossa et al., 2017; Wang et al., 2019; Chen et al., 2020). However, no such data have been collected for sesame seeds to date. Although there is sparse information on the chemical compositions (lignan, sesamin, and sesamolin), amino acid types and contents, and FA properties of sesame seeds from different regions, so far no transcriptome and degradome analyzes have been performed (Wei L. B. et al., 2016; Gacek et al., 2017; Sui et al., 2018; Ahmed et al., 2020). The specific functional genes and key metabolic pathways that regulate the synthesis of UFA in sesame seeds in particular are still unknown and unexplored. Therefore, a more comprehensive study on the controlling roles of miRNAs and their targets in lipid and FA biosynthesis is needed to further elucidate the underlying mechanism in the differential seed development of sesame.

Herein, the FA compositions in developing sesame seeds were analyzed, and the first-hand data on the combination of transcriptomics, small RNAs, and degradome in sesame seeds harvested in different developmental periods were generated and integrated to develop a comprehensive resource. We anticipate that this study will provide valuable information on the mechanism of lipid and FA biosynthesis in sesame seeds in different developmental periods. Furthermore, this study would promote molecular breeding research to improve the value-adding effects of sesame oil and the resulting functional plant products. Moreover, the collected miRNA expression patterns and extensive miRNA-mRNA regulatory network data can also be used for the upcoming expansion of plant oils with enriched UFA content.

MATERIALS AND METHODS

Chemicals, Plant Materials, and Total RNA Extraction

For this study, sesame seeds named Wan Zhi No.2 (WZ2) were collected from the experimental field of the Crop Research Institute of Anhui Academy of Agricultural Sciences, Hefei, China in 2019. The whole growth period of WZ2 was about 88 days, and its seed yield was recorded at 1,788 kg/ha with a crude fat content of 55% (Wang et al., 2011). It was bred from sexual hybridization and systematic breeding of a local variety "*Xiaozibai*" (female parent) and *Yuzhi No.4* (male parent). Beginning at 7 days after flowering (DAF), three biological replicates of sesame seeds for each sampling development period (7, 14, 21, and 28 DAF in August 2019) were handpicked until fully matured and then stored in liquid nitrogen at −80°C until further use.

For transcriptomic analysis, miRNA sequencing, and degradome analysis, total RNA of each sample was extracted from the collected seeds, including three biological replicates, using Trizol reagent (Invitrogen, Waltham, MA, United States) according to the previous method (Liu et al., 2016).

GC-MS Analysis of FA in Sesame Seed Oil

After drying and crushing, the sesame oil was extracted using the Soxhlet extraction method, and the solvent was evaporated with

a rotary evaporator (Dbrowski et al., 2020). The FA composition of the sesame oil was analyzed by gas chromatography-mass spectrometry (Agilent GC-MS 7890A, Agilent Technologies, Santa Clara, CA, United States) equipped with a DB-5MS column and NIST 08 (Gaithersburg, MD, United States) according to the procedure described (Zhu et al., 2021). Crude fat 50 mg was weighed into a 10-ml centrifuge tube followed by the addition of 2 ml 1% H_2SO_4-CH_3OH reagent and heating for 30 min at 70°C in a water bath. Then, 2 ml of N-hexane was added, and water was filled to the top of the centrifuge tube (10 ml). The whole mixture was shaken well and allowed to stand for 24 h. After the centrifugation, supernatant was analyzed by GC-MS. GC conditions were as follow: DB-5 column (30 mm × 0.32 mm, 0.25 μm), the carrier gas was high purity helium, the injection temperature was 250°C, the column flow rate was 1.25 ml/min, and the micro injector was used to inject 1.0 μl. The temperature rise program was 50°C (2 min), 5°C/min to 270°C. Mass spectrometry conditions were as follow: EI ion source, ion source temperature 220°C, GC-MS interface temperature 250°C, electron energy 70 eV, scan mass range 60–500 m/z.

According to the results of the GC-MS analysis, the percentage of each FA component was calculated by the peak area normalization method (i.e., the percentage of each peak area to the total peak area).

Transcriptome Sequencing and Assembly Analysis

A cDNA library constructed from the pooled RNA from the different samples of sesame seeds was sequenced using the Illumina 6000 sequencing platform according to Illumina paired-end RNA-seq. Prior to assembly, the low-quality reads were omitted, and overall cleaned paired-end reads were generated. The raw sequence data were submitted to the NCBI database, and the reads of the samples were mapped to the UCSC[1] reference genome[2] using the HISAT package. Then, all transcriptomes were generated using StringTie to examine the expression level for FPKM after the final transcriptome was constructed. Here, the differentially expressed genes (DEGs) and mRNAs with p value of less than 0.05 and a $\log_2|$ fold change$|$ of more than 1 were screened with the R package (Lyu et al., 2020).

Small RNA Sequencing and miRNA Identification

In this assay, the 12 small RNA libraries (three biological replicates) of the four different period samples were constructed and processed by screening with a high-throughput sequencing method (Illumina HiSeq 2500 platform) at LC-BIO (Hangzhou, China). Raw RNA reads were generated and estimated using Illumina FastQC to obtain Q30 data. The raw reads were subjected to an in-house program, ACGT101-miR (LC Sciences, Houston, TX, United States) to remove adapter dimers, junk data, low complexity reads, common RNA families (rRNA, tRNA, snRNA, and snoRNA) and repeats, and the valid reads were obtained. Four samples of small RNA sequencing were analyzed based on the previous bioinformatics method (Yuan et al., 2013; Bisgin et al., 2018). Read distribution was calculated for different libraries and normalized for miRNA prediction using global normalization methods (Lei et al., 2021; Zhang et al., 2021).

The differentially expressed miRNAs in different seed development periods were evaluated using the DESeq R package (1.8.3) and found to be significant at $P < 0.05$. Then, to identify the differentially expressed known and novel miRNAs in sesame seeds, unique sequences with length of 18–25 nt were selected to map the specific species precursors in miRBase[3] by BLAST search. The unique sequences that can be further compared with the genome were considered as known miRNAs. The unique sequences that can be compared with the genome but cannot be compared with the pre-miRNAs of the species selected in miRBase were assumed to be novel miRNA candidates.

Degradome Analysis and Target Identification

The RNA libraries containing three biological replicates from four samples of the developmental period of sesame seeds were mixed as one degradome library. Then, the four small RNA samples and one degradome composite sample were analyzed on the sequencing platform (IlluminaHiSeq 2500) at LC-Bio Co., Ltd., (Hangzhou, China), and the degradome libraries were constructed according to the description of the manufacturer. Furthermore, the degradome reads were aligned with the transcriptome data of the sesame samples. Finally, all the identified known and novel miRNA-target genes were mapped to the database (see text footnote 2). Analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment was performed using Agrigo and the Perl script, respectively, and the non-redundant GO enrichment terms were visualized using Revigo. In this study, a miRNA target gene network was constructed using Cytoscape (version 3.4.0).

Analysis of Differentially Expressed Target Genes

To obtain the target genes of lipid and FA synthesis, 12 RNA libraries were constructed from the sesame seed samples collected in four different developmental periods. For each library, all sequences were mapped to the CDS, and the number of each gene was counted using standard parameters. All the clean reads were mapped to the assembled unigenes of *S. indicum* L. for annotation. The RPKM of each gene referred to the number of sequencing fragments mapped to the CDS of the gene and was calculated by one thousand transcripts per million sequencing bases. According to the similarity of the gene expression profiles of samples, the heat map of the clustering expression pattern was constructed using $\log_{10}$ (FPKM + 1) to represent the miRNAs and gene expression. The standard for measuring the DEGs among the different samples was based on *P-value* < 0.05 and fold change greater than 2.

[1] http://genome.ucsc.edu/

[2] https://www.ncbi.nlm.nih.gov/genome/?term=Sesamum+indicum+L

[3] ftp://mirbase.org/pub/mirbase/CURRENT/

Co-expression Analysis

For further exploration of the relationship between each module and FA synthesis, DEGs and expression levels of each module were screened, and the correlation between each module and nine types of FA synthesis was tested. Weighted co-expression network analysis was performed using the three elements (node, edge line, and network) of Cytoscape network graph (LC-Bio Co., Ltd., Hangzhou, China). As a result of the data collection and import with the Cytoscape package, the basic framework of the network graph was created after confirmation. All the other parameters were used with modification and adjustment. The co-expression networks for selected miRNAs and target pairs were visualized using Cytoscape (version 3.4.0).

Analysis of Gene Expression and Validation by qRT-PCR

For data validation, we selected 12 differentially expressed miRNA-target pairs to perform qRT-PCR to verify the high-throughput sequencing results. SYBRPrimeScript^{TM} miRNA RT-PCR Kit (TaKaRa, Dalian, China) was used for reverse transcription reactions for miRNAs, and PrimeScript RT Reagent Kit (TaKaRa, Dalian, China) was used for reverse transcription of cDNAs. U6 and beta-tubulin (TUB) were selected for normalization of miRNAs and target genes, respectively (Zhang et al., 2020). The names of miRNAs and their target genes, sequences, and primers in RT-qPCR test experiment are given in **Supplementary File 1**. The $2^{-\Delta\ \Delta\ Ct}$ method was used to calculate the relative fold change of miRNA target genes, and the PCR reaction of each sample was repeated three times.

RESULTS

Lipid Accumulation in Sesame Seed Development

In this study, we harvested sesame seed samples of the four developmental periods to evaluate their FA composition by gas chromatography-mass spectrometry (GC-MS) (**Figure 1**). The results show five dominant components, namely, oleic acid (C18:1, 43.61% of total FAs at S1), linoleic acid (C18:2, 46.77% of total FAs at S4), palmitic acid (C16:0, 20.96% of total FAs at S3), *cis*-7-hexadecenoic acid (C16:1, 14.72% of total FAs at S2), and stearic acid (C18:0, 19.7% of total FAs at S3) as shown in **Figure 1B**. It is found that all these five FAs had higher percentage and dominated over the other FAs (more than 89% of the total FA) during the four consecutive periods and reached 98.8% of the total FA in the S4 periods. Based on this, the development of sesame seeds can be divided into three distinct periods; starting with a low content of oil and FAs in the initial stage (S1-S2), followed by a rapid accumulation of oil in the second and third stages (S3), and a gradual increase in oil and FAs until full maturation (S4) (**Figures 1B,C**). However, saturated fatty acids (SFAs) showed a slight decrease, and UFAs showed an increase at S4; in particular, oleic acid, linoleic acid, and hexadecenoic acid was the dominant factors to this significant increase (**Figure 1C**). Overall, oleic acid and linoleic acid were maintained at relatively high levels during sesame seed development, and UFAs dominated total FAs with relatively high proportions (**Figure 1C**).

Transcriptome Sequence Analysis in Sesame Seeds Under Different Growing Time Points

RNA pools were obtained from sesame seed samples harvested at four different developmental time points to construct the transcriptome library. After refining the raw sequence data, only 147,561,222 reads remained valid out of 213,772,171 raw reads. **Table 1** shows the annotation of sequence alignments uniq_miRNAs with the NCBI non-redundant protein (Nr) database. In addition, the results of the BLASTX algorithm showed the presence of 6,292 uniq_miRNAs with an average length of 70b (**Table 1**). Twelve RNA sequencing data with three biological replicates from the four different developmental periods were accumulated for the identification of DEGs with the aim of gene expression profile.

Small RNA Sequencing Profile

Small RNA sequencing for 12 libraries was performed to decipher miRNA regulated post-transcriptional changes associated with lipid and FA biosynthesis. **Figure 2A** depicts the size distributions and similar patterns of unique and valid 18- to 25-nt-long sequences obtained after applying the different steps of data processing. Among the 12 libraries, the 24-nt (49.1%) RNAs were found to be most abundant and diverse, followed by 23% small RNAs with 21-nt-long sequences (**Figure 2A** and **Supplementary File 2**).

Based on the abundance in miRNA expression levels, 1,025 miRNAs were obtained and categorized into five different groups after database and sequence reading analysis (**Figure 2C** and **Supplementary File 3**). Among the unique mature miRNAs and pre-miRNAs, 354 unique mature miRNAs were categorized into groups 1, −2, and −3 (gp1, −2, and −3) (**Supplementary File 3**). In addition, the miRNAs from gp4 were not registered in miRBase, and 671 unique mature miRNAs corresponding to 664 pre-miRNAs were considered as new candidate miRNAs (**Supplementary File 3**). Interestingly, the pattern of the proportion of mature miRNAs in each sampling developmental period showed some similarity (**Figure 2A**). Moreover, a total of 220 miRNAs (114 known miRNAs and 106 novel miRNAs) were expressed in all four developmental periods, as shown in **Figure 2B**. In addition, the sum of the first nucleotide bases of miRNA was different between the known and novel miRNAs. Adenosine (A, 49.82%) was reported as the most abundant nucleotide, followed by uracil (U, 20.16%) and cytosine (C, 18.33%) (**Supplementary File 4**).

Differential Expression of miRNAs in Sesame Seeds Under Different Growing Time Points

We used the high-throughput sequencing method to analyze and compare the differential expression of miRNAs (DEmiRNAs) related to lipid and FA biosynthesis in four samples of

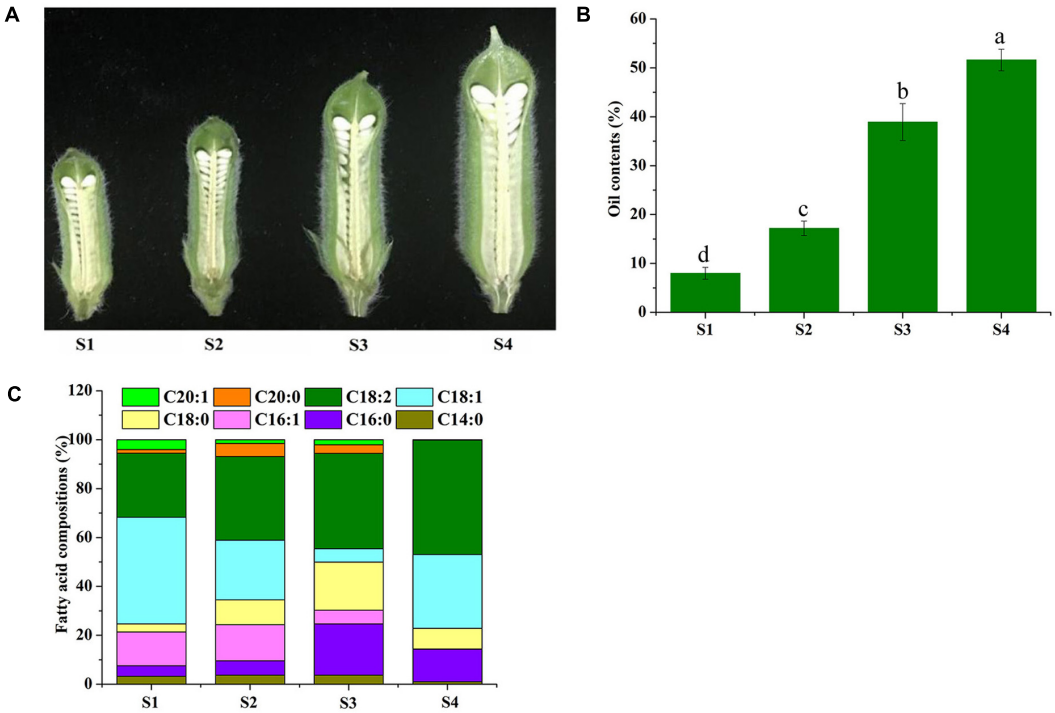


FIGURE 1 | Phenotypic observation and fatty acid determination during sesame seed development. The developmental process of sesame seeds (S1–S4) **(A)**. Collection time was at 7, 14, 21DAF, and until 28 days after flowering (DAF) (the time point of seed maturity). **(B)** The oil content percent of sesame seed at four time points. **(C)** The percent of different fatty acid compositions at four time points of sesame seed.

TABLE 1 | Data analysis of transcriptome sequencing for *Sesamum indicum* L.

Group	S1_Seed	S2_Seed	S3_Seed	S4_Seed
Raw reads	41,739,967	38,964,110	60,033,649	73,034,445
Valid reads	32,450,409	28,931,109	40,219,897	45,959,807
Number of uniq_miRNAs	1,427	1,278	1,833	1,754
Average Uniq_miRNA length (bp)	70	70	69	69
Q20 (%)	98	98	98	98
GC (%)	52	53	52	52

sesame seeds. According to the analysis, among the 220 DEmiRNAs, 126 miRNAs were up-regulated and 74 miRNAs were significantly down-regulated in S3 vs. S2 (**Figure 3**). Moreover, the distribution patterns showed significant up-regulation of 113 miRNAs and down-regulation of 63 in S3 vs. S1 (**Figure 3**). These data suggest that stage 3 may be the critical control point that strongly influences miRNA expression levels during lipid and FA biosynthesis. According to the heat map of potential DEmiRNAs, 220 miRNAs were found to be significantly expressed across the four periods with similar distribution patterns of high or low expression levels verified after clustering analysis (**Supplementary Figure 1**).

Prediction of Known and Novel miRNAs

To further understand the possible function and the regulatory patterns of the identified miRNAs and their target genes responsible for lipid and FA biosynthesis during sesame seed development, degradome sequencing was performed. According to the sequencing results, 33,155,150 (83.46% of the clean reads) refined reads were mapped to the 7,419,022 unigenes (70.84% of the input cDNA sequences) of sesame seeds (**Supplementary File 5**). From the mixed degradome data, 13,801 miRNAs with 1,006 targets were identified, including 65 novel miRNAs with 8,051 transcripts (**Supplementary File 6**). Among the referred 13,801 miRNAs, most miRNAs (804) could cleave six or more different transcript targets, while 202 miRNAs may have cleaved only few transcript targets ($t \leq 5$ in number) (**Supplementary File 6**). Based on the in-depth analysis, the miRNA (aly-MIR408-p3_2ss18CT19TG) was identified with 520 targets, which is the highest number of transcripts cleavages by the same miRNA (**Supplementary File 6**). To summarize, 12 miRNAs were associated with the regulation of lipid and FA biosynthesis during the developmental periods of sesame seed. Among the 12 miRNAs, 11 could cleave only one transcript

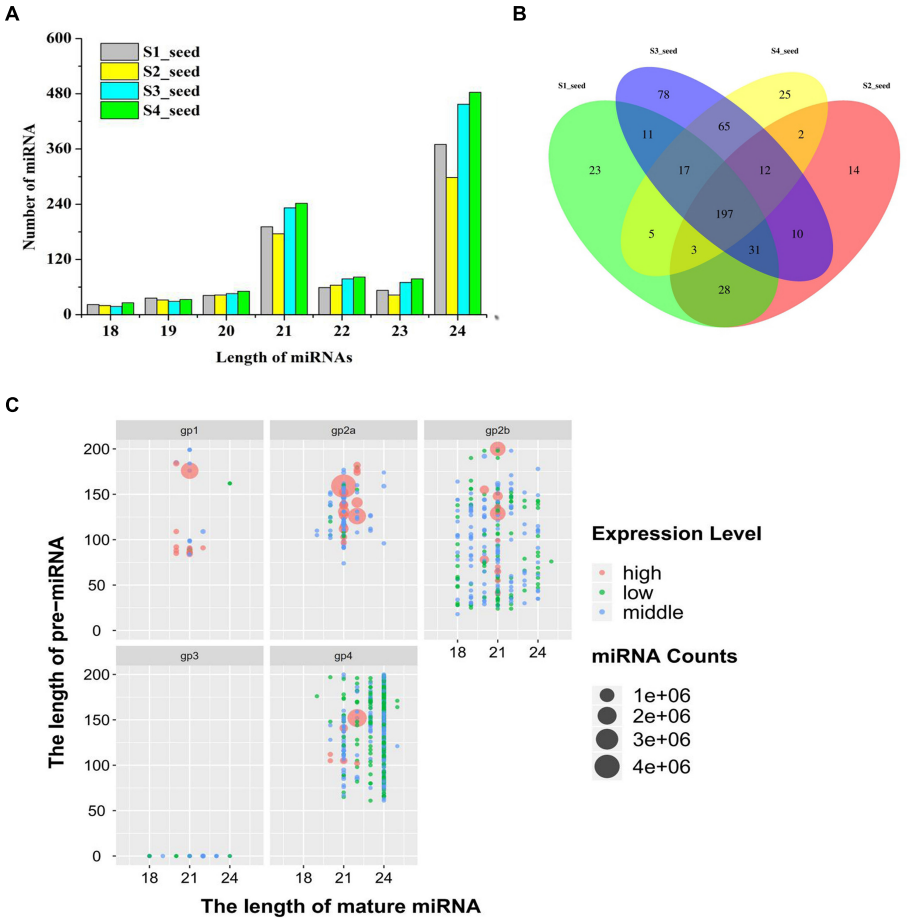


FIGURE 2 | (A) Number of different-length miRNA distribution in different development periods of sesame seed, **(B)** Venn diagrams of differentially expressed miRNAs in different development periods of sesame seed, and **(C)** length of miRNA distribution of five groups for expression level (high-level expression over average; middle-level expression >10 but less than average; low-level expression < 10).

target, while the miRNA (*bol-MIR9410-p3_2ss4TG17TA*) cleaved two transcript targets (*XM_011087111.2* and *XM_011099049.2*), and the miRNA cleavage sites are shown in **Figure 4**.

GO and KEGG Pathway Analyzes of Targets

For the manifestation of the potential biological role of 2,296 miRNAs in sesame seed development, 2,186 target genes were categorized according to GO enrichment (**Supplementary Figure 2**). The enrichment data showed that most of the target genes were associated with biological processes involving "regulation of transcription," "oxidation–reduction process," and "protein phosphorylation" (**Figure 5A**). The GO terms "nucleus," "cytoplasm," and "plasma membrane" in the "cellular component" category accumulated the most frequent targets (**Figure 5A**). The GO terms "protein binding" and "molecular function" in the molecular function category had the highest number of targets compared with the other terms (**Figure 5A**). There, interesting GO terms, such as "embryo development ending in seed dormancy" and "abscisic acid-activated signaling pathway" were revealed (**Figure 5A**). For pathway analysis, KEGG annotations were performed in the top 20 pathways according to the KEGG annotated gene number. Most target genes were found enriched in "ribosome" and "protein processing in endoplasmic reticulum," followed by "glycerolipid metabolism" and "glycerophospholipid metabolism" (**Figure 5B**). In addition, more genes were involved in "biosynthesis of secondary metabolites" consisting of "glycosylphosphatidylinositol (GPI)-anchor biosynthesis" and "biosynthesis of UFAs" (**Figure 5B**). Based on the enrichment data, "fatty acid degradation," "beta-alanine," and "inositol phosphate metabolism" were shown to be the most significantly enriched pathways (FDR 0.05) (**Figure 5B**).

miRNAs and Target Genes in the Context of Lipid and Fatty Acid Biosynthesis

The transcriptomics investigation revealed that a total of 220 DEmiRNAs corresponding to 3,705 target genes were also differentially expressed. After miRNA-target pair analysis, 32 miRNAs with 33 target genes were found to be associated

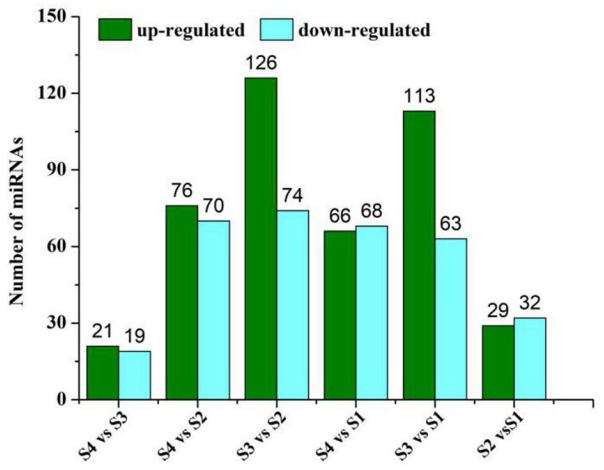


FIGURE 3 | Number of up- and down-regulated miRNAs in different development periods of sesame seed.

with lipid biosynthesis in sesame seeds (**Supplementary File 7**). In this analysis, most DEmiRNAs and their targets showed a similar expression pattern as in high-throughput sequencing (**Figure 6**). According to the experimental qRT-PCR data, eight known miRNAs and their target pairs showed negative expression correlation (**Figure 7**). For example, miRNA (*syl-miR166c-5p-1ss21AG*) showed higher expression in S1 and S2, which decreased in S3 and S4, while target *FAD2* showed lower expression in S1 and S2, which increased in S3 and S4. Moreover, four novel miRNA-target pairs (*PC -3p-15838_568/LOC105162096*, *PC -5p-145800_91/LOC105178588*, *PC -3p-62_86681/LOC105162196*, and *PC -3p-14226_618/LOC105163373*) also showed inverse association at the expression level, which control transcriptional repression by targets and their corresponding miRNAs (**Figure 7**). In addition, we detected three miRNAs-target pairs (*sly-miR166c-5p_1ss21AG/FAD2*, *stu-MIR8005c-p3_1ss13AG/LOC105159459*, and *gma-MIR169o-p3_2ss16AC21CG/LOC105161564*) that exhibited not only a relationship at the high expression level but also an apparent negative relationship. Interestingly, the miRNA *bol-MIR9410-p3_2ss4TG17TA* was found to have two target genes (*LOC105176298* and *LOC105167408*). We hypothesize that the functional analysis and cloning of the three genes *FAD2*, *LOC105159459*, and *LOC105161564* could potentially facilitate the deciphering of the mechanistic details associated with the biosynthesis of UFAs during sesame seed development.

Gene Co-expression Network Associated With Lipid and Fatty Acid Biosynthesis

To elucidate the function of miRNA-target pairs in the context of lipid and FA biosynthesis regulation in different developmental periods, we constructed a coexpression network with 33 genes from nine categories, mainly associated with glycerolipid metabolism, glycerophospholipid metabolism, UFAs, FA biosynthesis, FA elongation, FA degradation, steroid synthesis, arachidonic acid, and ether ester metabolism (**Figure 8**).

After further analysis of the functional assignments, seven and five genes were found to be responsible for the categories of glycerolipid metabolism and glycerophospholipid metabolism, respectively, among the 33 genes (**Supplementary File 7**). According to gene co-expression analysis, 13 genes were associated with FA biosynthesis and metabolism, of which three genes (*FAD2*, *LOC105174578*, and *LOC105169658*) regulated the biosynthesis of UFAs (**Figure 8**). In addition, four miRNAs (*osa-miR167a-5p_2ss2GA21AG*, *osa-miR167a-5p_2ss14AG21A*, *osa-miR167a-5p_2ss2GA21AT*, and *osa-miR167a-5p*) had the same target gene *LOC10514358* involved in FA elongation (**Figure 8**). In the gene co-expression network, we found that miRNA (*bol-MIR9410-p3_2ss4TG17TA*) could simultaneously regulate three target genes (*LOC105176298*, *LOC105167408*, and *LOC105158444*) involved in glycerophospholipid metabolism, FA biosynthesis, and FA degradation, respectively (**Figure 8**). Similarly, some important miRNAs and their target genes involved in glyceride metabolism, ether ester metabolism, arachidonic acid, and steroid synthesis have been linked. Notably, the data suggest that the coexpression network of miRNA and target genes can be used in further studies to identify the function of known and novel genes related to lipid and FA biosynthesis in sesame seed development.

DISCUSSION

To improve the varietal characteristics of sesame, increasing the UFA content (especially oil yield and FA composition) is the first and most important priority for the development of sesame breeding (Wang D.D. et al., 2018; Ahmed et al., 2020). With the possibility of the presence of the sesame genome and the advent of high-throughput sequencing for small RNA analysis in sesame seeds, we hereby elaborated the scope of transcriptomic, miRNA, degradation group analysis data, and gene network data during sesame seed development (Zheng et al., 2019).

In this study, 32 miRNA and target pairs involved in lipid and FA biosynthesis were screened out. Importantly, most of these target pairs were related to glyceride metabolism and FA biosynthesis, and five miRNA target gene pairs related to glycerol phospholipid metabolism were also identified by degradome sequencing data. Interestingly, seven target genes involved in FA degradation were found, but three of them did not have miRNA pairs. We further analyzed and found that the expression levels of four miRNA genes changed significantly with the maturity of sesame, such as, *stu-MIR8005c-p3_1ss13AG/LOC105159459* involved in glyceride metabolism, *sly-miR166c-5p_1ss21AG/FAD2* involved in UFA metabolism, *gma-MIR169o-p3_2ss16AC21CG/LOC105161564* involved in FA biosynthesis, and *PC-3p-62_86681/LOC105162196* involved in steroid synthesis. Thus, these four miRNA-target pairs were identified as potential regulators of the lipid and FA biosynthesis process during sesame seed development. To the knowledge of the authors, this is the first study ever to report novel miRNAs

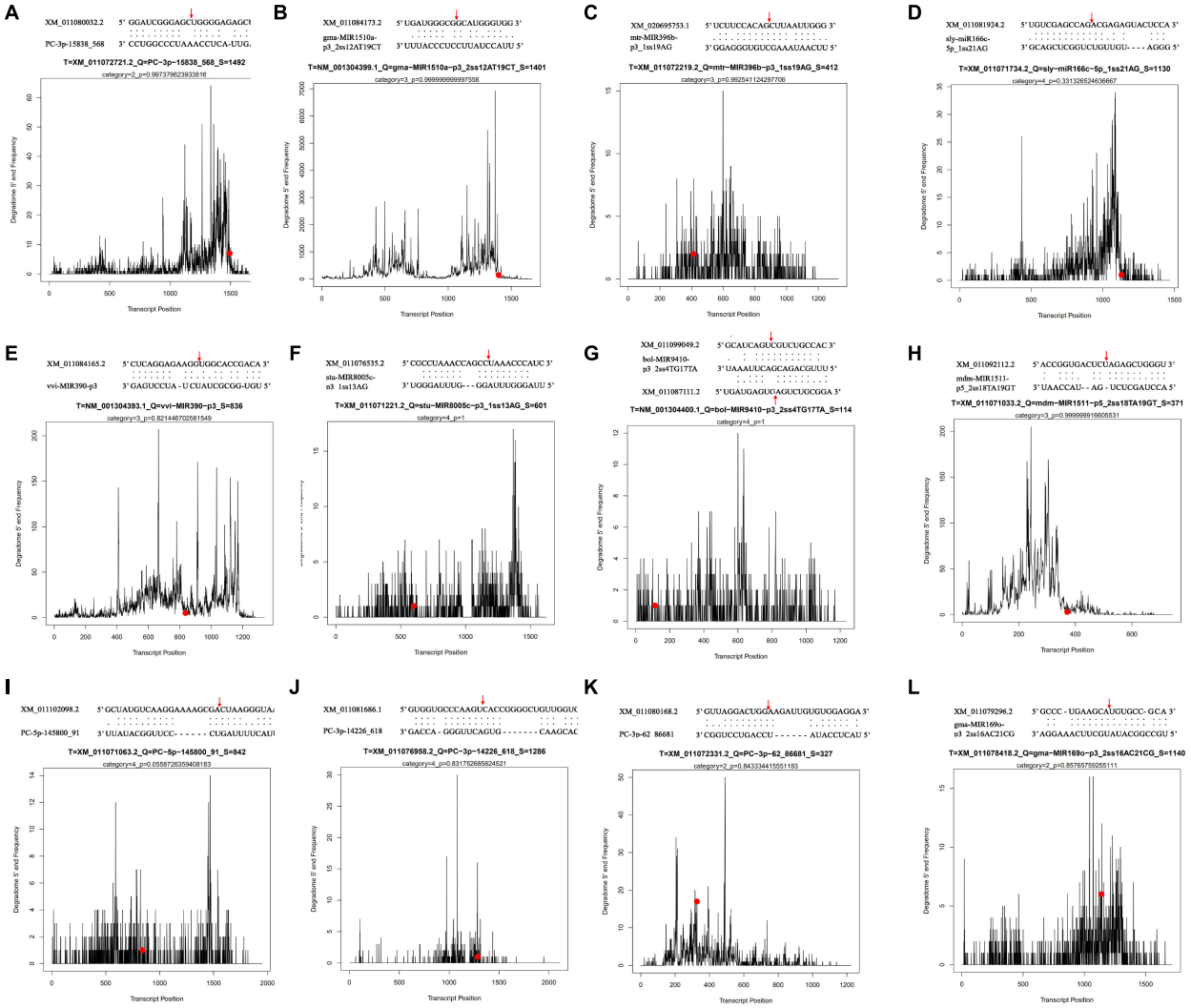


FIGURE 4 | Some differentially expressed miRNA t-plots and cleavage sites identified by degradome sequencing.

targeting miRNA-target pairs that had not been previously discovered, for example, *PC-3p-14226_618*, *PC-5p-228878_56*, *PC-5p-145800_91*, and *PC-3p-59282_203*. The results suggest that these specific miRNA target pairs may be involved in the regulatory network of lipid and FA synthesis, and play a central role in controlling gene expression and metabolic pathways throughout development periods.

Previous studies have shown that oil synthesis regulates *de novo* FA biosynthesis in the plastid and TAG assembly in the endoplasmic reticulum (ER), and that oil body formation (Reynolds et al., 2019) in oil plants is represented (Zheng et al., 2019). This study signified the presence of 106 unigenes involved in oil biosynthesis pathways (**Figure 9**). In general, lipid and FA syntheses are initiated with the supply of acetyl-CoA substrate (Wang et al., 2019), which is first converted to malonyl-CoA (Chen et al., 2020), and carbon flux is initiated *via* FA synthase and most other enzymes associated with lipid and FA biosynthesis (Huang, 1996). In this study, the initiation and acyl chain elongation of *de novo* FA biosynthesis were reported to be possibly regulated by 50 unigenes, and the latter showed apparent down-regulation of at least one isoform in S4 compared with S1, suggesting their functional involvement in the initiation of FA biosynthesis (**Figure 9**). In this study, ACCase catalyzed the first reaction involving four genes (conversion of acetyl-CoA to malonyl-CoA) involved in the biosynthesis of FA (**Figure 9**). Furthermore, FATA is highly specific for 18:1-ACP, and FATB is specific for acyl-A, which are known to regulate chain termination during FA synthesis (Lu et al., 2009). Unigenes *LOC105158149* and *LOC105160370*, which encode FATB, were up-regulated more than 10-fold during the S1 to S4 phase. In addition, four unigenes encoding LACS generating the acyl-CoA pool were detected, all of which were significantly down-regulated (**Figure 9**). The free long-chain FAs (16:0, 18:0, and 20:0) are esterified by LACS and output to

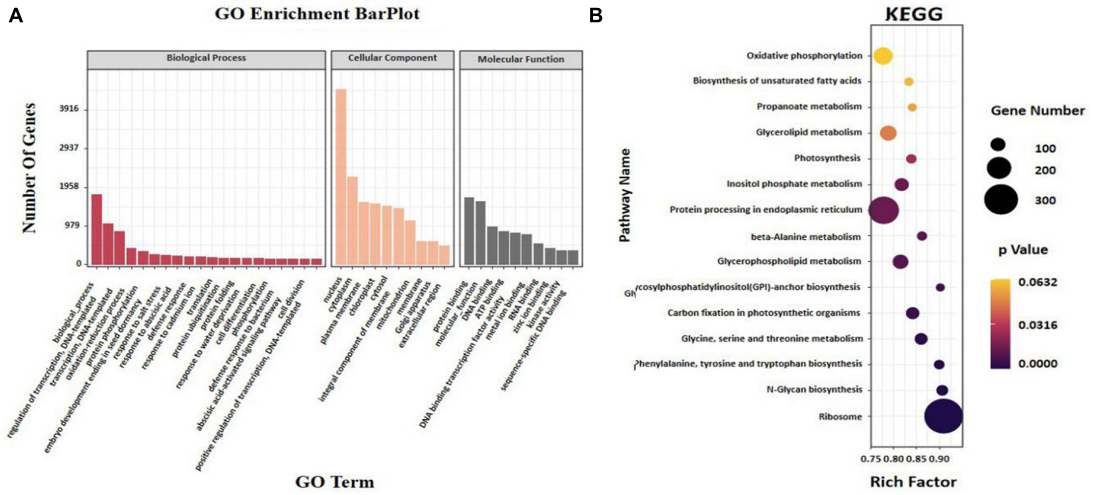


FIGURE 5 | Analysis of **(A)** enriched GO and **(B)** KEGG of miRNAs and target genes in different seed development periods of sesame.

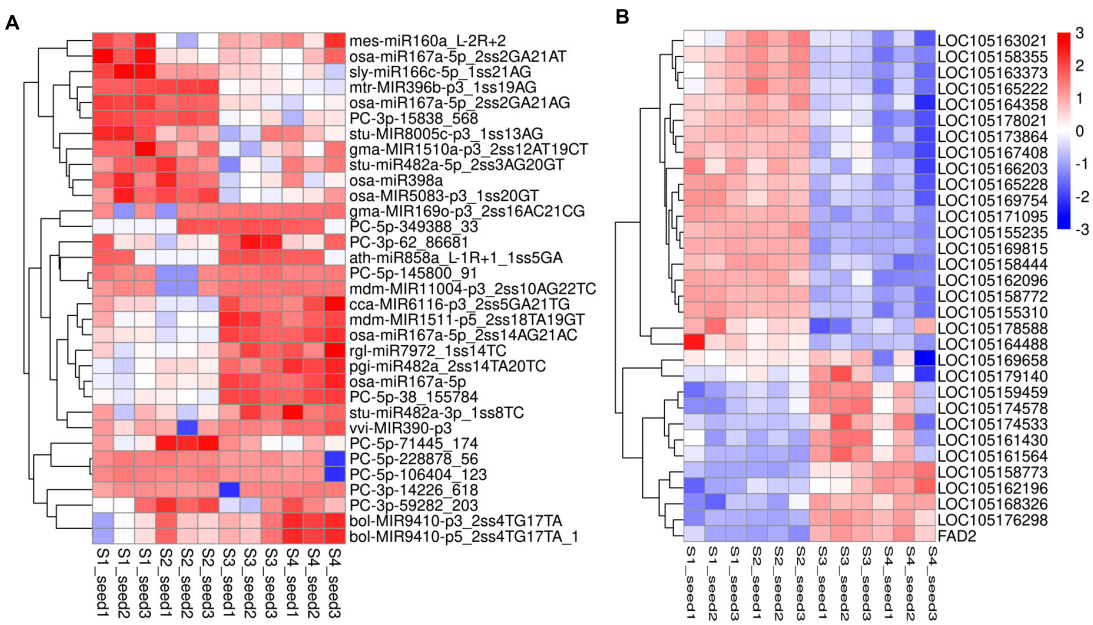


FIGURE 6 | Heat map of DEmiRNAs **(A)** and their target genes **(B)** indicating the expression pattern in different seed development periods of sesame. The corresponding color changes from blue to red with signal intensity ranges from –3.0 to +3.0. The miRNAs expression values were standardized by *Z*-score standardization.

ER for further acyl processing (Bates et al., 2009). Pathways such as glycerolipid metabolism and ether-lipid metabolism are known to be linked by the compound, which converts to glycerone-3P (G3P) under the catalyzation of *NAD* + and *FAD2* (Hu et al., 2009; Dehghan and Yarizade, 2014; Yasemin et al., 2018; Macovei et al., 2021). 49 unigenes related to the formation of TAG were identified, e.g., long-chain acyl-CoA and G3P, which could start the assemble of TAG; diacylglycerol acyltransferase (DGAT), which was the major enzymes that catalyze diacylglycerol (DAG); and phosphatidylcholine (an acyl donor), which could combine with oleosin to form an oil body (Kennedy, 1961; Peng and Weselake, 2011). Most of the unigenes encoding DAG were reported with high transcript levels, suggesting that they are an important signaling pathway during sesame seed development (Wang R.K. et al., 2018; Huai et al., 2020). On the other hand, relatively active unigenes, encoding acyl editing, PE-DAG, and PC-DAG interconversion were likely responsible for the incorporation of UFAs into TAG in sesame seeds (Peng and Weselake, 2011; Wang R.K. et al., 2018; Yang et al., 2018; Tyagi et al., 2019). The data emphasized the

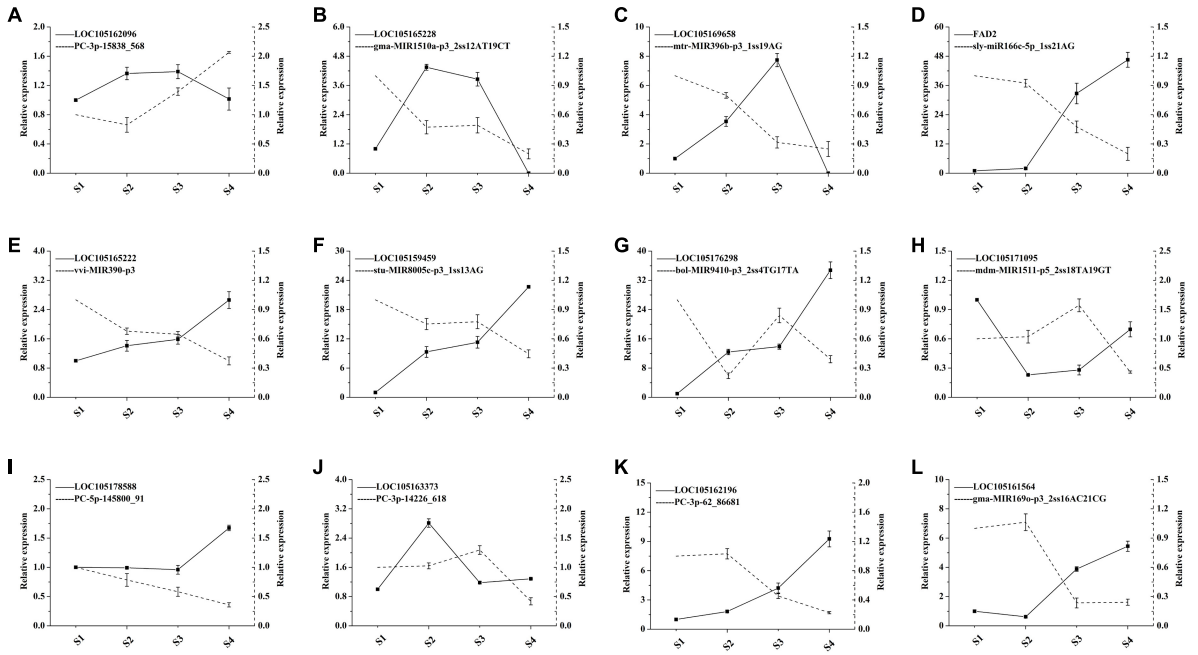


FIGURE 7 | qRT-PCR validation of miRNAs and their target expression correlation in different development periods of sesame seed (n = 3). The dotted and solid lines express miRNAs and target genes, respectively.

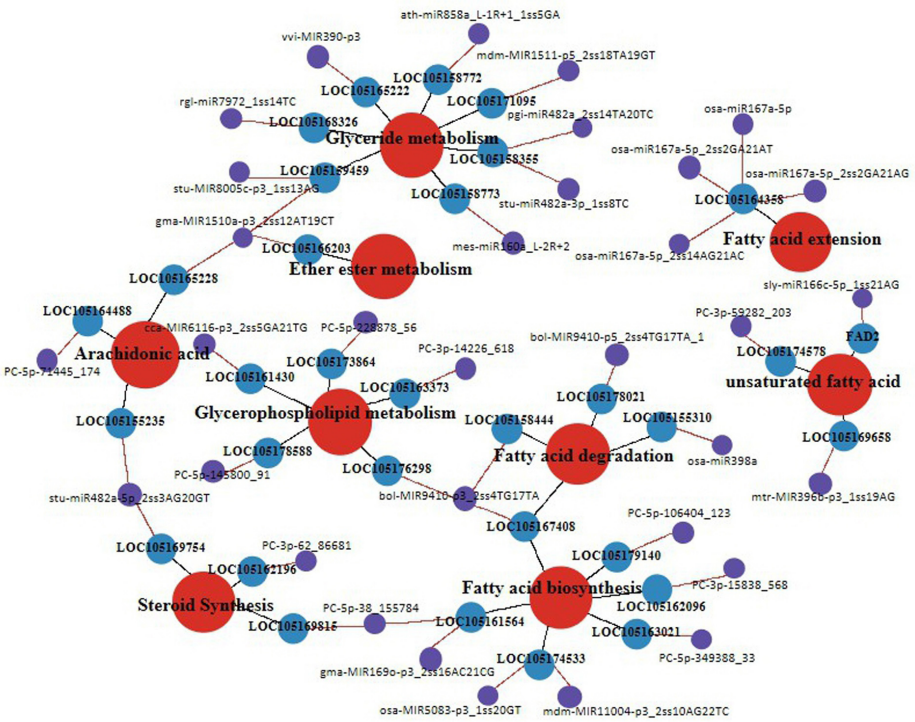


FIGURE 8 | Schemata of lipid synthesis enzymes and genes that related to fatty acid metabolism pathway in the different development periods of sesame seeds. The red color circles show different lipid biosynthesis pathways, blue colors show different expression genes, and purple color shows different expression miRNAs.

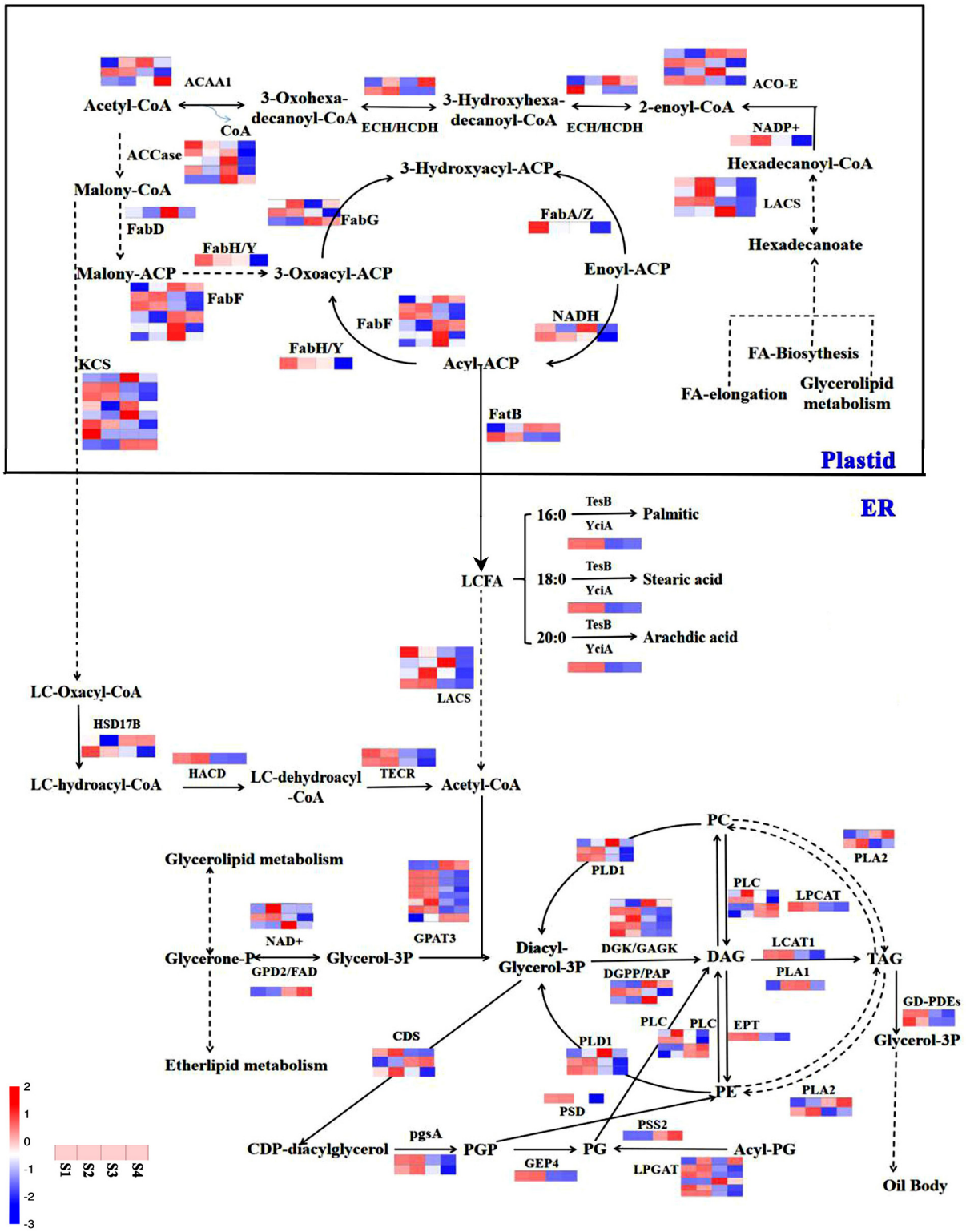


FIGURE 9 | Relation network of miRNAs and target genes with lipid and fatty acid biosynthesis in sesame seeds. The blue color represents the down-regulated gene expression pattern, while the red color represents the up-regulated gene expression pattern.

essential roles of *FAD2*, FabF/KASII, LACS, ACO-E, LPCAT, and some unknown genes during the biosynthesis of FA and other UFAs during sesame seed development (**Figure 9**).

Overall, the regulatory co-expression network construction data showed that the nine kinds of oil synthesis related pathways were the most likely regulatory factors of lipid and FA biosynthesis; and so far, it can be stated that miRNA regulates the complex network of genes and their target genes involved in various physiological and biological functions such as regulation of sesame responses to oil synthesis. These target pairs *gma-MIR1510a-p3_2ss12AT19CT/LOC105159459*, *sly-miR166c-5p_1ss21AG/FAD2*, *gma-MIR169o-p3_2ss16AC21CG/LOC105161564*, and *PC-3p-62_86681/LOC105162196* were found to be potentially relevant to oil synthesis in developing sesame seeds. Nevertheless, future investigation of these four pairs is warranted to reveal their precise molecular functions and regulatory mechanisms. Herein, complementary comparative multi-omics data were generated in four successive developmental periods to unravel the important pathways responsible for the high-oil yield and increased UFA content in sesame seeds. The results collected would also fill the knowledge gap on the mechanism of oil deposition and FA saturation transformation during embryonic development in other oil plant seeds.

CONCLUSION

In conclusion, this study provides a comprehensive framework for a better understanding of the lipid and FA biosynthesis mechanism of sesame under different growth periods. The1,006 target genes of 13,801 miRNAs were extracted from mixed degradome sequencing, which accounted for 65 novel miRNAs targeting 8,051 transcripts, and 3,705 target genes of 220 miRNAs were also differentially manifested. At least 12 potential lipid-related miRNAs and target genes were analyzed by RNA-seq and qRT-PCR. Perfect harmonization of high level of *FAD2* in the mature period promoted the increase of linoleic acid content in sesame. Thirty-three co-expressed genes formed a co-regulatory sub-network, which showed that some important miRNAs and their target genes were involved in the pathways of lipid and FA biosynthesis such as glyceride metabolism, ether ester metabolism, arachidonic acid, and steroid synthesis, and the pathways were also found to be interconnected. In addition, the biosynthetic and regulatory genes identified from the genome-wide co-expression network may expand the understanding of lipid and FA biosynthesis in other plant seeds and increase the functional value of the resulting food products.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA739094.

AUTHOR CONTRIBUTIONS

Y-PZ: conceptualization, methodology, software, and writing–original draft preparation. Y-YZ: data curation and validation. KT: data curation and investigation. FZ and J-GZ: methodology and software. FH: visualization and investigation. P-CW: visualization, supervision, and investigation. Z-JW: funding acquisition, supervision, and writing–reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Major Projects of Science and Technology in Anhui Province (201903a06020021, 201904a06020008, 202004a06020042, and 202004a06020052) and the National Natural Science Foundation of China (31850410476).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.709197/full#supplementary-material

Supplementary Figure 1 | Heatmap of potential DEmiRNA significantly expressed across the four stages with similar distribution patterns. The blue color represents the down-regulated gene expression pattern, while the red color represents the up-regulated gene expression pattern.

Supplementary Figure 2 | GO enrichment of genes significantly expressed across the four stages.

Supplementary File 1 | Sequences and primers of miRNAs and their target genes in RT-qPCR test experiment.

Supplementary File 2 | Length distribution of unique miRNAs.

Supplementary File 3 | miRNAs were categorized into five different groups based on the abundance of miRNAs expression levels.

Supplementary File 4 | Analysis of miRNA first nucleotide bias total.

Supplementary File 5 | Sequencing result of identified miRNAs and their targets genes during sesame seed development.

Supplementary File 6 | Identify of miRNAs and transcript targets from mixed degradome data.

Supplementary File 7 | Co-expression network involved genes of nine categories.

REFERENCES

Adebowale, A. A., Sanni, S. A., and Falore, O. A. (2011). Varietal differencesin the physical properties and proximate. *World J. Agric. Sci.* 7, 42–46.

Ahmed, I. A. M., Aljuhaimi, F., Özcan, M. M., Ghafoor, K., Şimşek, Ş, Babiker, E. E., et al. (2020). Evaluation of chemical properties, amino acid contents and fatty acid compositions of sesame seed provided from different locations. *J. Oleo Sci.* 69, 795–800. doi: 10.5650/jos.ess20041

Bates, P. D., Durrett, T. P., Ohlrogge, J. B., and Pollard, M. (2009). Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. *Plant Physiol.* 150, 55–72. doi: 10.1104/pp.109.137737

Bhunia, R. K., Chakraborty, A., Kaur, R., Gayatri, T., Bhat, K. V., Basu, A., et al. (2015). Analysis of fatty acid and lignan composition of Indian germplasm of sesame to evaluate their nutritional merits. *J. Am. Oil Chem. Soc.* 92, 65–76. doi: 10.1007/s11746-014-2566-3

Bisgin, H., Gong, B. S., Wang, Y. P., and Tong, W. D. (2018). Evaluation of bioinformatics approaches for next-generation sequencing analysis of microRNAs with a toxicogenomics study design. *Front. Genet.* 9:22. doi: 10.3389/fgene.2018.00022

Chen, D. J., Luo, X. G., Yan, L. H., Si, C. L., Wang, N., He, H. P., et al. (2020). Transcriptome analysis of unsaturated fatty acids biosynthesis shows essential genes in sprouting of *Acer truncatum* Bunge seeds. *Food Biosci.* 2020:100739. doi: 10.1016/j.fbio.2020.100739

Dan, M., Huang, M. H., Liao, F., Qin, R. Y., Liang, X. J., Zhang, E. Z., et al. (2018). Identification of ethylene responsive miRNAs and their targets from newly harvested banana fruits using high-throughput sequencing. *J. Agric. Food Chem.* 66, 10628–10639. doi: 10.1021/acs.jafc.8b01844

Dbrowski, G., Czaplicki, S., and Konopka, I. (2020). Composition and quality of poppy (*Papaversomniferum* L.) seed oil depending on the extraction method. *LWT Food Sci. Technol.* 134:110167. doi: 10.1016/j.lwt.2020.110167

Dehghan, N. F., and Yarizade, K. (2014). Bioinformatics study of delta-12 fatty acid desaturase 2 (FAD2) gene in oilseeds. *Mol. Biol. Rep.* 41, 5077–5087. doi: 10.1007/s11033-014-3373-5

Dossa, K., Diouf, D., Wang, L., Wei, X., Zhang, Y., Niang, M., et al. (2017). The emerging oilseed crop sesamum indicum enters the "Omics" Era. *Front. Plant Sci.* 8:1154. doi: 10.3389/fpls.2017.01154

Elleuch, M., Besbes, S., Roiseux, O., Blecker, C., and Attia, H. (2007). Quality characteristics of sesame seeds and byproducts. *Food Chem.* 103, 641–650. doi: 10.1016/j.foodchem.2006.09.008

Gacek, K., Bayer, P. E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D., et al. (2017). Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. *Front. Plant Sci.* 7:2062. doi: 10.3389/fpls.2016.02062

Han, X. J., Yin, H. F., Song, X. X., Zhang, Y. X., Liu, M. Y., Sang, J., et al. (2016). Integration of small RNAs, degradome and transcriptome sequencing in hyperaccumulator Sedum alfredii uncovers a complex regulatory network and provides insights into cadmium phytoremediation. *Plant Biotechnol. J.* 14, 1470–1483. doi: 10.1111/pbi.12512

Hu, Y. P., Wu, G., Cao, Y. L., Wu, Y. H., Xiao, L., Li, X. D., et al. (2009). Breeding response of transcript profiling in developing seeds of *Brassica napus*. *BMC Mol. Biol.* 10:49. doi: 10.1186/1471-2199-10-49

Huai, D. X., Xue, X. M., Li, Y., Wang, P., Li, J. G., Yan, L. Y., et al. (2020). Genome-wide identification of peanut KCS genes reveals that AhKCS1 and AhKCS28 are involved in regulating VLCFA contents in seeds. *Front. Plant Sci.* 11:406. doi: 10.3389/fpls.2020.00406

Huang, A. (1996). Oleosins and oil bodies in seeds and other organs. *Plant Physiol.* 110:1055. doi: 10.1104/pp.110.4.1055

Huang, J. Q., Zhang, T., Zhang, Q. X., Chen, M., Wang, Z. J., Zheng, B. S., et al. (2016). The mechanism of high contents of oil and oleic acid revealed by transcriptomic and lipidomic analysis during embryogenesis in Carya cathayensisSarg. *BMC Genom.* 17:113. doi: 10.1186/s12864-016-2434-7

Huang, R. M., Zhou, Y., Zhang, J. P., Ji, F. Y., Jin, F., Fan, W., et al. (2021). Transcriptome analysis of walnut (*Juglans regia* L.) embryos reveals key developmental stages and genes involved in lipid biosynthesis and polyunsaturated fatty acid metabolism. *Agric. Food Chem.* 69, 377–396. doi: 10.1021/acs.jafc.0c05598

Kennedy, E. P. (1961). Biosynthesis of complex lipids. *Fed Proc.* 20:934.

Kumar, A. P. K., McKeown, P. C., Boualem, A., Ryder, P., Brychkova, G., Bendahmane, A., et al. (2017). Tilling by Sequencing (TbyS) for targeted genome mutagenesis in crops. *Mol. Breed.* 37:14.

La Sala, L., Micheloni, S., De Nigris, V., Prattichizzo, F., and Ceriello, A. (2018). Novel insights into the regulation of miRNA transcriptionalcontrol: implications for T2D and related complications. *Acta Diabetol.* 1:10.

Lei, P., Liu, Z., Hu, Y. B., Kim, H. C., Liu, S., Liu, J. Q., et al. (2021). Transcriptome analysis of salt stress responsiveness in the seedlings of wild and cultivated *ricinus communis* L. *J. Biotechnol.* 327, 106–116. doi: 10.1016/j.jbiotec.2020.12.020

Li, S. S., Wang, L. S., Shu, Q. Y., Wu, J., Chen, L. G., Shao, S., et al. (2015). Fatty acid composition of developing tree peony (Paeonia section Moutan DC.) seeds and transcriptome analysis during seed development. *BMC Genom.* 16:208. doi: 10.1186/s12864-015-1429-0

Liu, H. Y., Tan, M. P., Yu, H. J., Li, L., Zhou, F., Yang, M. M., et al. (2016). Comparative transcriptome profiling of the fertile and sterile flower buds of a dominant genic male sterile line in sesame (*Sesamum indicum* L.). *BMC Plant Biol.* 16:250. doi: 10.1186/s12870-016-0934-x

Lu, C. F., Xin, Z. G., Ren, Z. H., Miquel, M., and Browse, J. (2009). An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18837–18842. doi: 10.1073/pnas.0908848106

Lyu, Y. S., Wei, X. J., Zhong, M., Niu, S., Ahmad, S., Shao, G. N., et al. (2020). Integrated transcriptome, small rna, and degradome analysis to elucidate the regulation of rice seedling mesocotyl development during the passage from darkness to light. *Crop J.* 8, 44–54.

Macovei, A., Rubio-Somoza, I., Paiva, J. A. P., Araújo, S., and Donà, M. (2021). Editorial: MicroRNA signatures in plant genome stability and genotoxic stress. *Front. Plant Sci.* 12:683302. doi: 10.3389/fpls.2021.683302

Majdalawieh, A. F., Dalibalta, S., and Yousef, S. M. (2020). Effects of sesamin on fatty acid and cholesterol metabolism, macrophage cholesterol homeostasis and serum lipid profile: a comprehensive review. *J. Eur J Pharmacol.* 885:173417. doi: 10.1016/j.ejphar.2020.173417

Peng, F. Y., and Weselake, R. J. (2011). Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in *Arabidopsis*. *BMC Genom.* 12:286. doi: 10.1186/1471-2164-12-286

Reynolds, K. B., Cullerne, D. P., Anna, E. T., Rolland, V., Blanchard, C. L., Wood, C. C., et al. (2019). Identification of genes involved in lipid biosynthesis through de novo transcriptome assembly from cocos nucifera developing endosperm. *Plant Cell Physiol.* 60, 945–960. doi: 10.1093/pcp/pcy247

Sevgi, M. (2018). Identification and functional analyses of new sesame miRNAs (*Sesamum indicum* L.) and their targets. *Mol. Biol. Rep.* 45, 2145–2155. doi: 10.1007/s11033-018-4373-7

Sui, N., Wang, Y., Liu, S., Yang, Z., Wang, F., and Wan, S. (2018). Transcriptomic and physiological evidence for the relationship between unsaturated fatty acid and salt stress in peanut. *Front. Plant Sci.* 9:7. doi: 10.3389/fpls.2018.00007

Tyagi, S., Sri, T., Singh, A., Mayee, P., Shivaraj, S. M., Sharma, P., et al. (2019). Suppressor of overexpression of CONSTANS1 influences flowering time, lateral branching, oil quality, and seed yield in *Brassica juncea* cv. *Varuna*. *FunctIntegr. Genomic.* 19, 43–60. doi: 10.1007/s10142-018-0626-8

Wacal, C., Ogata, N., Basalirwa, D., Sasagawa, D., Kato, M., Handa, T., et al. (2019). Nishihara, E. Fatty acid composition of sesame (*Sesamum indicum* L.) seeds in relation to yield and soil chemical properties on continuously monocropped upland fields converted from paddy fields. *Agron* 9:801. doi: 10.3390/agronomy9120801

Wang, D. D., Zhang, L. X., Huang, X. R., Wang, X., Yang, R. N., Mao, J., et al. (2018). Identification of nutritional components in black sesame determined by widely targeted metabolomics and traditional chinese medicines. *Molecules* 23:1180. doi: 10.3390/molecules23051180

Wang, R. K., Liu, P., Fan, J. S., and Li, L. L. (2018). Comparative transcriptome analysis two genotypes of Acer truncatum Bunge seeds reveals candidate genes that influences seed VLCFAs accumulation. *Sci. Rep.* 8:15504.

Wang, L. H., Li, D. H., Zhang, Y. X., Gao, Y., Yu, J. Y., Wei, X., et al. (2016). Tolerant and susceptible sesame genotypes reveal waterlogging stress response patterns. *PLoS One* 11:e0149912. doi: 10.1371/journal.pone.0149912

Wang, Q., Zhao, L., Wang, B. C., Cao, W. X., Xu, G. Z., Chen, P., et al. (2011). Breeding and high yield cultivation techniques of a New Sesame Variety Wanzhi 2 with high quality. *Anhui Agron. Bull.* 17, 62–64.

Wang, X. J., Liang, H. Y., Guo, D. L., Guo, L. L., Duan, X. G., Jia, Q. S., et al. (2019). Integrated analysis of transcriptomic and proteomic data from tree peony (*P. ostii*) seeds reveals key developmental stages and candidate genes related to oil biosynthesis and fatty acid metabolism. *Hortic. Res.* 6:111.

Wei, L. B., Zhang, H. Y., Duan, Y. H., Li, C., Chang, S. X., and Miao, H. M. (2016). Transcriptome comparison of resistant and susceptible sesame (*Sesamum*

indicum L.) varieties inoculated withFusarium oxysporum f. sp. sesami. *Plant Breed.* 135, 627–635. doi: 10.1111/pbr.12393

Wei, X., Zhu, X. D., Yu, J. Y., Wang, L. H., Zhang, Y. X., Li, D. H., et al. (2016). Identification ofsesame genomic variations from genome comparison of landrace and variety. *Front. Plant Sci.* 7:1169. doi: 10.3389/fpls.2016.01169

Xu, X. B., Yin, L. L., Ying, Q. C., Song, H. M., Xue, D. W., Lai, T. F., et al. (2013). High-throughputsequencing and degradome analysis identify miRNAs and theirtargets involved in fruit senescence of *Fragaria ananassa*. *PLoS One* 8:e70959. doi: 10.1371/journal.pone.0070959

Yan, R., Liang, C. Z., Meng, Z. G., Malik, W., Zhu, T., Zong, X. F., et al. (2016). Progress in genomesequencing will accelerate molecular breeding in cotton (*Gossypium* spp.). *3 Biotech.* 6:217.

Yang, T. Q., Yu, Q., Xu, W., Li, D. Z., Chen, F., and Liu, A. Z. (2018). Transcriptome analysis reveals crucial genes involved in the biosynthesis of nervonic acid in woody *Malania oleifera* oilseeds. *BMC Plant Biol.* 18:247. doi: 10.1186/s12870-018-1463-6

Yanik, H., Turktas, M., Dundar, E., Hernandez, P., Dorado, G., and Unver, T. (2013). Genome-wide identification of alternate bearing-associatedmicroRNAs (miRNAs) in olive (*Olea europaea* L). *BMC Plant Biol.* 13:10. doi: 10.1186/1471-2229-13-10

Yasemin, C. A., Mehmet, U. N., Cengiz, B. M., Ulu, F., Can, T. F., and Cetinkaya, R. (2018). Comparative identification and evolutionary relationship of fatty acid desaturase (FAD) genes in some oil crops: the sunflower model for evaluation of gene expression pattern under drought stress. *Biotechnol. Biotec. Eq.* 32, 846–857. doi: 10.1080/13102818.2018.1480421

Yin, D. D., Li, S. S., Shu, Q. Y., Gua, Z. Y., Wu, Q., and Feng, C. Y. (2018). Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. *Gene* 666, 72–82. doi: 10.1016/j.gene.2018.05.011

Yuan, C., Wang, X. L., Geng, R. Q., He, X. L., Qu, L., and Chen, Y. L. (2013). Discovery of cashmere goat (*Capra hircus*) microRNAs in skin and hair follicles by Solexa sequencing. *BMC Genom.* 14:511. doi: 10.1186/1471-2164-14-511

Zhang, Y. J., Gong, H. H., Li, D. H., Zhou, R., Zhao, F. T., Zhang, X. R., et al. (2020). Integrated small RNA and Degradome sequencing provide insights into salt tolerance in sesame (*Sesamum indicum* L.). *BMC Genom.* 21:494. doi: 10.1186/s12864-020-06913-3

Zhang, Y. J., Zhang, Q. Y., Weng, X. C., Du, Y. H., and Zhou, X. (2021). NEase-based amplification for detection of miRNA, multiple miRNAs and circRNA. *Anal. Chim. Acta* 1145, 52–58. doi: 10.1016/j.aca.2020.12.024

Zhao, S. Q., Mi, X. Z., Guo, R., Xia, X. B., Liu, L., An, Y. L., et al. (2020). The biosynthesis of main taste compounds is coordinately regulated by miRNAs and phytohormones in tea plant (*Camellia sinensis*). *J. Agric. Food Chem.* 68, 6221–6236. doi: 10.1021/acs.jafc.0c01833

Zhao, S. Q., Wang, X. W., Yan, X. M., Guo, L. X., Mi, X. Z., Xu, Q. S., et al. (2018). Revealing of the microRNA involved regulatory gene networks on terpenoid biosynthesis in *Camellia sinensis* in different growing time points. *J. Agric. Food Chem.* 66, 12604–12616. doi: 10.1021/acs.jafc.8b05345

Zheng, Y. S., Chen, C. J., Liang, Y. X., Sun, R. H., Gao, L. C., Liu, T., et al. (2019). Genome-wide association analysis of the lipid and fatty acid metabolism regulatory network in the mesocarp of oil palm (*Elaeisguineensis* Jacq.) based on small noncoding RNA sequencing. *Tree Physiol.* 39, 356–371. doi: 10.1093/treephys/tpy091

Zhu, J. F., Zhang, M. M., Guo, J. X., Wu, X. K., and Wang, S. M. (2021). Metabolite profiling of chondrosarcoma cells: a robust gc-ms method for the analysis of endogenous metabolome. *J. Chromatogr. B* 1169:122606. doi: 10.1016/j.jchromb.2021.122606

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

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

frontiers in Plant Science

REVIEW
published: 05 August 2021
doi: 10.3389/fpls.2021.719148

Multiplex Genome Editing in Yeast by CRISPR/Cas9 – A Potent and Agile Tool to Reconstruct Complex Metabolic Pathways

*Joseph Christian Utomo, Connor Lorne Hodgins and Dae-Kyun Ro**

Department of Biological Science, University of Calgary, Calgary, AB, Canada

Numerous important pharmaceuticals and nutraceuticals originate from plant specialized metabolites, most of which are synthesized *via* complex biosynthetic pathways. The elucidation of these pathways is critical for the applicable uses of these compounds. Although the rapid progress of the omics technology has revolutionized the identification of candidate genes involved in these pathways, the functional characterization of these genes remains a major bottleneck. Baker's yeast (*Saccharomyces cerevisiae*) has been used as a microbial platform for characterizing newly discovered metabolic genes in plant specialized metabolism. Using yeast for the investigation of numerous plant enzymes is a streamlined process because of yeast's efficient transformation, limited endogenous specialized metabolism, partially sharing its primary metabolism with plants, and its capability of post-translational modification. Despite these advantages, reconstructing complex plant biosynthetic pathways in yeast can be time intensive. Since its discovery, CRISPR/Cas9 has greatly stimulated metabolic engineering in yeast. Yeast is a popular system for genome editing due to its efficient homology-directed repair mechanism, which allows precise integration of heterologous genes into its genome. One practical use of CRISPR/Cas9 in yeast is multiplex genome editing aimed at reconstructing complex metabolic pathways. This system has the capability of integrating multiple genes of interest in a single transformation, simplifying the reconstruction of complex pathways. As plant specialized metabolites usually have complex multigene biosynthetic pathways, the multiplex CRISPR/Cas9 system in yeast is suited well for functional genomics research in plant specialized metabolism. Here, we review the most advanced methods to achieve efficient multiplex CRISPR/Cas9 editing in yeast. We will also discuss how this powerful tool has been applied to benefit the study of plant specialized metabolism.

Keywords: CRISPR/Cas9, genome editing, multiplex gene integration, *Saccharomyces cerevisiae*, plant specialized metabolites

OPEN ACCESS

Edited by:
Thu Thuy Dang,
University of British Columbia Okanagan, Canada

Reviewed by:
Bjoern Hamberger,
Michigan State University, United States
Isabel Desgagne-Penix,
Université du Québec à Trois-Rivières, Canada

****Correspondence:***
Dae-Kyun Ro
daekyun.ro@ucalgary.ca

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

Received: *01 June 2021*
Accepted: *14 July 2021*
Published: *05 August 2021*

Citation:
Utomo JC, Hodgins CL and Ro D-K (2021) Multiplex Genome Editing in Yeast by CRISPR/Cas9 – A Potent and Agile Tool to Reconstruct Complex Metabolic Pathways. Front. Plant Sci. 12:719148. doi: 10.3389/fpls.2021.719148

INTRODUCTION

Plant specialized metabolites (or secondary metabolites) play important roles in enhancing human health and wellness as sources of pharmaceuticals, nutraceuticals, flavors, and fragrances. However, these specialized metabolites and their biosynthetic enzymes are usually available at miniscule levels in plants, making studies of their biosynthesis difficult. The emergence of

molecular cloning techniques, next-generation sequencing, omics technology, and synthetic biology in recent decades has accelerated the discovery of specialized metabolic pathways in plants (Siddiqui et al., 2012; Pyne et al., 2019). One of the most powerful tools to study these biosynthetic pathways is the heterologous expression of candidate genes in microorganisms. *Escherichia coli* and *Saccharomyces cerevisiae* are the two major workhorses for microbial pathway reconstruction. While the more fully studied *E. coli* system has been useful to study soluble enzymes, expression of membrane-bound enzymes, such as cytochrome P450s in *E. coli*, is difficult (Leonard and Koffas, 2007). As cytochrome P450s have been discovered to be the major enzyme family driving the chemical diversity of specialized metabolites (Chapple, 1998; Bathe and Tissier, 2019), *S. cerevisiae* with a developed endomembrane system has benefited studies of specialized metabolic pathways, including multiple cytochrome P450s. Several key features of yeast exemplify its practicality (Siddiqui et al., 2012; Pyne et al., 2019). Firstly, efficient yeast transformation techniques simplify the day-to-day uses of yeast. Secondly, yeast has limited endogenous specialized metabolism pathways, which minimizes competition with introduced pathways. Thirdly, yeast partially shares primary metabolism pathways with plants, which means heterologous plant specialized metabolism pathways can be easily plugged into the existing pool of yeast primary metabolite precursors, albeit some flux enhancement may be required. Fourthly, it is relatively safe to work with non-pathogenic yeast, known as one of the generally recognized as safe microbes. Finally, yeast can also carry out some post-translational modifications.

Despite all these advantages, reconstructing complex specialized pathways in yeast is still hindered by two major factors. First, there are a limited number of selection markers for yeast transformation. Selection markers can be recycled, but this is time consuming. Secondly, due to plasmid instability and imbalance, it is often difficult to achieve consistent levels of recombinant proteins in individual cells, which can lead to different degrees of toxicity and metabolic burden in each yeast cell (Da Silva and Srikrishnan, 2012). Thus, the integration of heterologous genes into the yeast genome is preferred for stable expression. Gene integration techniques in yeast, such as *in vivo* homologous recombination and pre-CRISPR endonucleases-based systems (e.g., I-SceI, HO endonuclease, ZFNs, and TALENs), have been extensively developed. These techniques have been reviewed in several excellent articles (Da Silva and Srikrishnan, 2012; David and Siewers, 2015; Yang and Blenner, 2020). However, these integration techniques are relatively laborious as they require selection markers, and there is limited availability of efficient integration sites.

In the past decade, the emergence of endonuclease-based techniques, especially clustered regularly interspaced short palindromic repeats (CRISPR) and its associated protein 9 (Cas9), has revolutionized the field of genome editing. This elegant and simple technique has been applied in various organisms, such as human cells, zebrafish, plants, and *S. cerevisiae* (Doudna and Charpentier, 2014). The simplicity, efficiency, and flexibility of CRISPR/Cas9 have allowed for the expansion of its application to include multiple simultaneous genome-editing events, termed "multiplexing" (Mali et al., 2013). There are some excellent reviews on CRISPR/Cas9 in yeast (Jakočiūnas et al., 2016; Stovicek et al., 2017; Deaner and Alper, 2019; Meng et al., 2020), including some articles focusing on multiplex genome editing (Adiego-Pérez et al., 2019; Malcı et al., 2020). Thus, this review will focus on the most up-to-date advances in multiplex genome editing in *S. cerevisiae* with an emphasis on building the complex pathways of plant specialized metabolites. Specifically, this review focuses on the emergence of CRISPR/Cas9, multiplex gene integration in yeast, current developments in multiplex gene integration using other Cas protein (Cas12a), a brief discussion of other applications of multiplex gene editing, and future perspective of using CRISPR/Cas9 multiplex genome editing for studying plant specialized metabolism.

CRISPR/Cas9 DEVELOPMENT

CRISPR-Cas genome editing is derived from the adaptive immune response of archaea and bacteria and consists of CRISPR genomic sequences and *Cas* genes (Jansen et al., 2002). CRISPR genomic sequences were first discovered in the genome of *E. coli* by Ishino et al. (1987), who found repeating palindromic sequences separated by small, evenly sized, and unique spacer sequences (Ishino et al., 1987). In 2002, CRISPR sequences were shown to be transcribed into CRISPR RNA (crRNA) and the *Cas* genes associated with them were predicted to have nuclease and helicase activity (Jansen et al., 2002; Tang et al., 2002). By 2005, the spacer sequences were determined to be viral sequences (Mojica et al., 2005). In 2007, the CRISPR-Cas adaptive immune response was demonstrated to protect *Streptococcus thermophilus* from invading viruses (Barrangou et al., 2007).

Since these initial discoveries, diverse CRISPR-Cas types have been identified but the type II CRISPR system is the most heavily utilized for heterologous gene editing and will be the focus in this review (Makarova et al., 2020). Early studies of the CRISPR immune response of *S. thermophilus* demonstrated that the Cas9 protein uses its catalytic HNH and RuvC-like domains to cleave invading viral DNA (Sapranauskas et al., 2011). At the same time, the requirement of *trans*-activating crRNA (tracrRNA) for the maturation of crRNA in the *Streptococcus pyogenes* CRISPR immune response was demonstrated (Deltcheva et al., 2011). The tracrRNA DNA sequence is located upstream of the CRISPR locus in the bacterial genome. The tracrRNA sequence is complementary to the repeating portion of the crRNA, and when transcribed the tracrRNA and crRNA form a tracrRNA-crRNA duplex. In 2012, the Cas9 protein of *S. pyogenes* (Cas9) was shown to interact with tracrRNA-crRNA duplexes and induce double-stranded breaks (DSBs) in DNA complementary to the spacer sequence, known as the protospacer (Jinek et al., 2012). For recombinant systems, the protospacer sequence represents the target site for gene

editing. Also, the tracrRNA-crRNA duplex can be fused in recombinant systems to a single functional transcript, thereby simplifying the heterologous CRISPR-Cas to two components, the single-guide RNA (gRNA) and Cas9 (**Figure 1**). The gRNA is composed of a 20-nt spacer RNA and scaffold RNA. The scaffold RNA is a fusion of the tracrRNA and the structural repeating portion of the crRNA. A caveat to this system is that a specific sequence called the protospacer adjacent motif (PAM) is required for gRNA recognition of the protospacer (Jinek et al., 2012). The PAM sequence for Cas9 is 5'-NGG-3', while different Cas proteins has different PAM sequences. When binding to a gRNA, Cas9 undergoes a conformational change which forms a channel between the two lobes of the protein (Jinek et al., 2014). This creates room for the binding of the spacer RNA to the target DNA protospacer sequence upstream of the PAM site. After the spacer RNA binds to the target DNA, the HNH domain cleaves the DNA strand complementary to the spacer sequence, and the RuvC-like domain cleaves the opposite strand. The cleavages by HNH and RuvC both occur 3-bp upstream of the PAM creating a blunt DSB (**Figure 1**; Jinek et al., 2012). The first demonstration of CRISPR/Cas9 genome editing in *S. cerevisiae* followed soon (DiCarlo et al., 2013). Due to highly efficient homology-directed repair (HDR) in *S. cerevisiae*, especially when DSBs are induced, CRISPR/Cas9 is a perfect approach for genomic integrations of foreign genes in yeast (Storici et al., 2003; Shao and Zhao, 2009; Gardner and Jaspersen, 2014). The foreign genes can be transformed as donor DNA, which is a linear DNA fragment that consists of the gene cassette with homology arms at its 5' and 3' ends. The homology arms are DNA sequences that are homologous to the 5' and 3' regions of the DSB sites in the yeast genome (**Figure 1**). Therefore, by transforming the Cas9 and gRNA expression cassettes, and donor DNA, heterologous gene integration can be achieved.

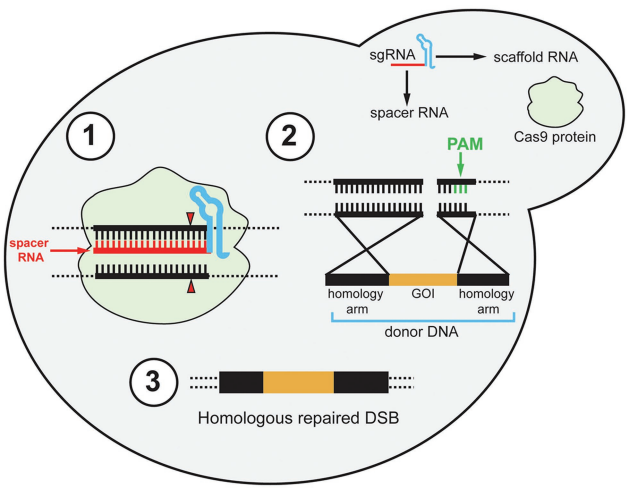


FIGURE 1 | Schematic diagram of the CRISPR/Cas9 mechanism in yeast. The single gRNA (sgRNA) contains two components: scaffold RNA (tracrRNA and structural part of crRNA) and a 20-nt spacer RNA. (1) After the Cas9 protein binds to the sgRNA, Cas9 binds to the target sites in the genomic DNA and undergoes conformational change to cut both strands of the target site 3-bp upstream of the PAM site (red triangle). PAM site is shown in step 2. (2) After a double-stranded break (DSB) is induced, the preferable homology-directed repair (HDR) mechanism repairs the DSB using donor DNA. The donor DNA contains 5' and 3' homology arms and a gene of interest (GOI). The GOI can be non-functional for gene deletion purposes. Note that although the PAM site is emphasized in the second step (green bars), the PAM site is required for Cas9-gRNA complex to recognize the target sites in the first step. (3) After HDR, the DSB is repaired and the GOI is integrated.

CRISPR/Cas9 MULTIPLEX GENE EDITING

As soon as CRISPR/Cas9 genome editing was discovered, the idea of simultaneously editing multiple target sites in the genome (multiplex genome editing) was demonstrated in human and mammalian cells (Cong et al., 2013; Mali et al., 2013). In yeast, the first multiplex genome editing was successfully demonstrated soon after the first application of CRISPR/Cas9 genome editing in yeast (Ryan et al., 2014). Thereafter, various methods have been employed to increase the efficiency of multiplex CRISPR genome editing. Here, these methods will be classified based on their gRNA expression systems. In general, there are three common approaches to express gRNA cassettes for multiplex gene editing: (i) expression of multiple gRNAs in a single gRNA cassette with RNA cleaving mechanisms; (ii) expression of multiple gRNAs in multiple gRNA cassettes; and (iii) editing of multiple, pre-defined sequences in the genome by a single gRNA. The studies reviewed here are summarized in **Table 1**. Unless indicated otherwise, all the plasmids and/or strains are available on request.

Multiple gRNAs in a Single gRNA Cassette With RNA Cleaving Mechanisms

The first strategy of expressing multiple gRNAs for targeting multiple loci is achieved by exploiting RNA cleaving mechanisms from either yeast or other organisms. Using this elegant strategy, multiple gRNAs can be expressed in a single transcript under a single promoter and terminator. Signal sequences are added between each gRNA and can be recognized by RNA cleaving mechanisms thereby producing multiple gRNAs. The multiple gRNAs then bind to multiple target sites simultaneously. Four RNA cleaving mechanisms will be discussed here and included HDV ribozymes, CRISPR direct repeats, Csy4, and tRNA arrays.

Hepatitis Delta Virus Ribozyme

The first demonstration of multiplex genome editing in yeast was demonstrated by Ryan et al. using a plasmid containing a Cas9-expressing cassette and a gRNA cassette containing the self-cleavable hepatitis delta virus (HDV) ribozyme attached to the 5' end of each gRNA (Ryan et al., 2014). The HDV ribozyme cleaves the 5' end of its sequence (**Figure 2A**). The extra nucleotides from the HDV ribozyme attached to the

TABLE 1 | Summary of studies that were reviewed in this manuscript.

Name	[1]Type of editing	Max # edited loci	Maximum # heterologous genes integrated	Total size of genes integrated	Editing efficiency	Genes/pathways integrated	References
CRISPRm (HDV ribozyme)	Deletion (integration)	3 (1)	1	~1.5 kb	20% (85%)	Cellobiose utilization	Ryan et al., 2014
HI-CRISPR (Direct repeats)	Deletion	3	–	100 bp	30–85%	–	Bao et al., 2015
Csy4	Deletion	4	–	120 bp	96%	–	Ferreira et al., 2018
GTR-CRISPR (tRNA)	Deletion	8	–	120 bp	87%	–	Zhang et al., 2019
Mans et al.	Deletion (integration)	6 (2)	6	~15 kb	65% (95%)	Acetyl-CoA biosynthesis	Mans et al., 2015
Modular gRNA	Deletion (integration)	3 (3)	11	~24 kb	64% (4.2%)	Muconic acid biosynthesis	Horwitz et al., 2015
Jakočiūnas et al.	Deletion	4	–	~500 bp	100%	–	Jakočiūnas et al., 2015b
CasEMBLR	Integration	3	3	~18 kb	31%	β-carotene biosynthesis	Jakočiūnas et al., 2015a
CrEdit	Integration	3	3	~18 kb	84%	β-carotene biosynthesis	Ronda et al., 2015
EasyClone-MarkerFree	Integration	3	6	~15 kb	70%	Acetyl-CoA biosynthesis	Jessop-Fabre et al., 2016
Di-CRISPR (delta integration)	Integration	18 (and 10) δ sites	2 (3)	16 kb (24 kb)	95% (85%)	Butadienol biosynthesis and xylose utilization	Shi et al., 2016
CRITGI	Integration	12 Ty sites	1	~5 kb	75%	Pyruvate decarboxylase biosynthesis	Hanasaki and Masumoto, 2019
CMGE	Integration	10 rDNA sites	1	~1.5 kb	46%	GFP expression	Wang et al., 2018
mCAL	Integration	3	3	~4 kb	100%	*HIS3*, *CDC11*, and *SHS1*	Finnigan and Thorner, 2016
Wicket	Integration	3; 6; 9; 12	3	~18 kb	95%; 50%; 50%; 10%	β-carotene biosynthesis	Hou et al., 2018
Landing pads	Integration	4	1	~2 kb	80%	Norcoclaurine biosynthesis	Bourgeois et al., 2018
SGM-CRISPR	Integration	6	6	~15 kb	40%	Kauniolide biosynthesis	Baek et al., 2021
PCR & Go	Integration	5	5	~6 kb	70%	Astaxanthin biosynthesis	Qi et al., 2021
Promiscuous gRNA	Deletion	2	–	120 bp	100%	–	Ferreira et al., 2017

[1]*Type of editing refers to if the study showed gene deletion (donor DNA is non-functional DNA) or gene integration (donor DNA is functional heterologous DNA).*

gRNA do not affect genome-editing efficiency and interestingly the gRNAs modified by the HDV ribozyme significantly improved the efficiency of genome editing. It was shown that the attachment of the HDV ribozyme produces more gRNA transcripts compared to the control system without it (Ryan et al., 2014). This system can knock out a maximum of three loci simultaneously using 120-bp donor DNA with an efficiency around 20 and 80% in diploid and haploid yeast, respectively (Ryan et al., 2014). This approach has been applied to investigate and reconstruct various plant specialized metabolisms, such as tropane alkaloids (Srinivasan and Smolke, 2019, 2020), noscapine (Hafner et al., 2021), and cyanogenic glycosides (Kotopka and Smolke, 2019).

Direct Repeats for crRNA Processing

The homology-integrated CRISPR (HI-CRISPR) system separated the expression of tracrRNA and crRNAs rather than combining them in a single gRNA (Bao et al., 2015). The Cas9, tracrRNA, and crRNA (including the 20-nt spacer sequence) cassettes were expressed in one plasmid (available in Addgene). The crRNA cassette contained multiple crRNAs, with each being flanked with direct repeat sequences which mimicked the natural direct repeats of the *S. pyogenes* CRISPR array. After transcription, these repeats were cleaved by the endogenous yeast RNase III and unknown nucleases, resulting in expression of multiple crRNAs from one cassette (**Figure 2B**). The crRNAs combined with tracrRNAs to form functional gRNAs. This study successfully disrupted up to three different loci in one transformation using 100-bp donor DNA with efficiencies varying from 30 to 85%, depending on the loci targeted.

Heterologous Endoribonuclease (Csy4)

A couple of years before utilizing CRISPR/Cas9 for genome editing, the Doudna lab had characterized an endoribonuclease that cleaves direct repeats from pre-crRNA to produce mature crRNAs in *Pseudomonas aeruginosa* (Haurwitz et al., 2010). This endoribonuclease is called Csy4 and has been tested

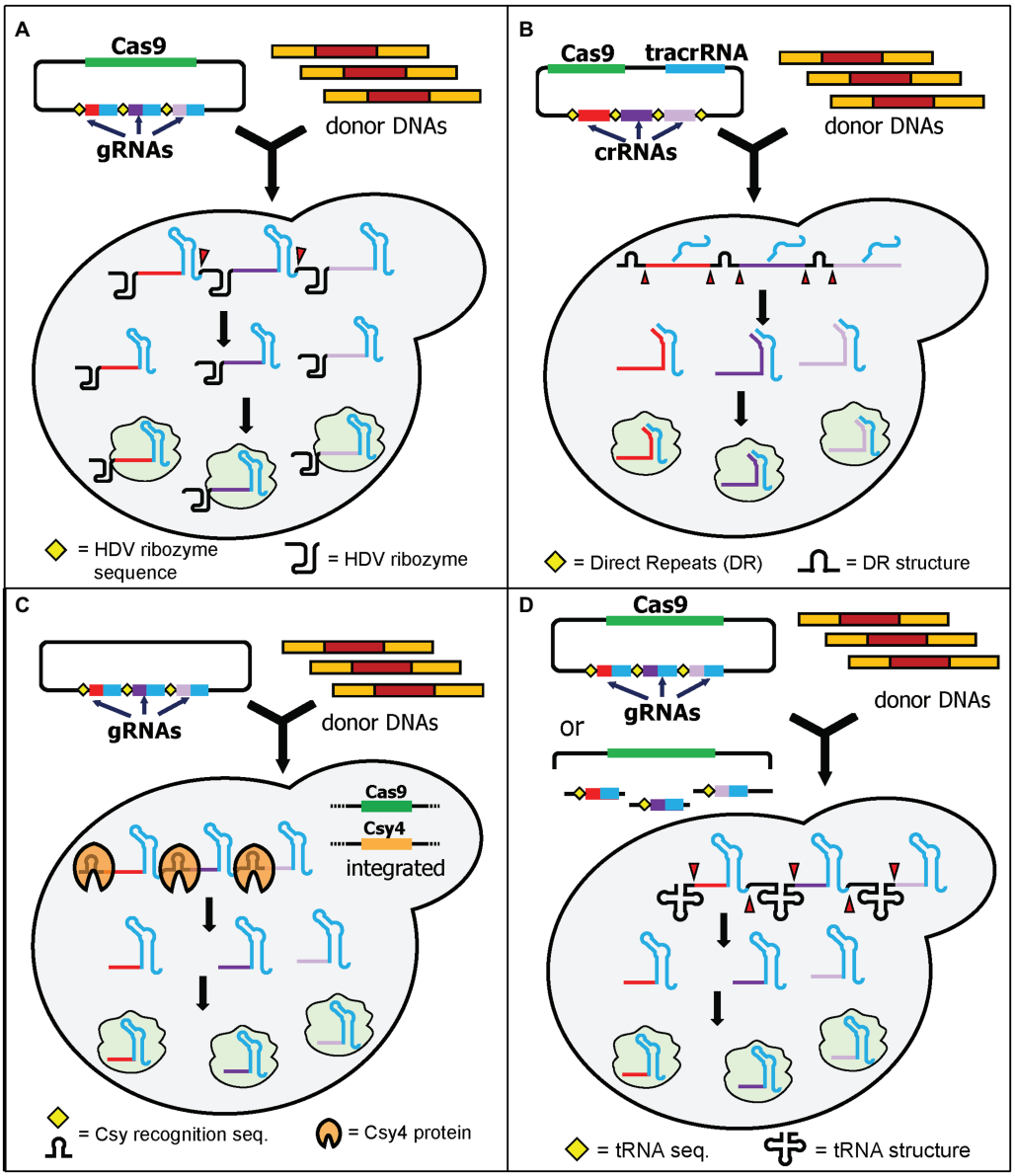


FIGURE 2 | Schematic diagram of multiplex gene editing using multiple gRNAs in a single gRNA cassette with RNA cleaving mechanisms. **(A)** HDV ribozyme self-cleaving mechanism by Ryan et al. After transcription, the HDV ribozyme cuts at its 5' end (red triangles) to produce multiple gRNAs. **(B)** HI-CRISPR by Bao et al. utilized the direct repeats (DR) from *S. pyogenes*. When it is expressed in yeast, the DR will be cleaved (red triangles) by unknown yeast endogenous nucleases and/or RNases, producing multiple gRNAs. Note that the tracrRNA will bind to the structural part of the crRNA before cleaving. **(C)** Csy4 endoribonuclease from *Pseudomonas aeruginosa* can be utilized by providing recognition sites between gRNAs. *Csy4* and *Cas9* were integrated into the yeast genome. Upon transcription, Csy4 recognizes and cleaves the recognition sites. **(D)** Yeast endogenous tRNA can be provided in between the gRNAs. Either fully assembled plasmids or linear plasmids with gRNA fragments can be transformed. The linear plasmid can be assembled *in vivo* using yeast DNA assembly. Upon transcription, the tRNAs are cleaved, resulting in multiple gRNAs (red triangles).

for its functionality to process multiple crRNAs at the same time in various organisms (Qi et al., 2012). The application of Csy4 processing for multiplex genome editing in yeast was first demonstrated by Ferreira et al. (2018). The authors first integrated both Cas9 and Csy4 expression cassettes into the yeast genome. They also built a plasmid containing one gRNA cassette with 28-nt Csy4 recognition sites between each gRNA (**Figure 2C**). Unlike HDV cleavage, but like the HI-CRISPR system, the Csy4 recognition sites are abolished after cleavage. This results in multiple gRNAs without any additional RNA structures. This approach successfully demonstrated quadruple gene disruptions using 120-bp donor DNA with 96% efficiency. The study also showed that utilization of Csy4 recognition sites in the absence of Csy4 still resulted in 50% efficiency for double-gene deletions (Ferreira et al., 2018). This may suggest that Cas9 might

have some flexibility to recognize at least the first two gRNAs in a single long transcript and correctly cleave both genomic DNA targets, albeit with a lower efficiency (Ferreira et al., 2018). The study from Ryan et al. also supports this suggestion as Cas9 can still recognize gRNAs that have additional HDV structures on their 5' ends (Ryan et al., 2014). Alternatively, the endogenous yeast RNase III and nucleases may cleave the Csy4 direct repeats in a similar way to how *S. pyogenes* direct repeats are cleaved in the HI-CRISPR system (Bao et al., 2015). In another independent study, Csy4 processing capability was also demonstrated to express 12 gRNAs simultaneously for CRISPR interference (CRISPRi; McCarty et al., 2019).

Yeast Endogenous tRNA Array

Utilizing the endogenous tRNA-processing mechanism for single transcript expression of multiple gRNAs was first demonstrated in rice (Xie et al., 2015). Zhang et al. showed that the tRNA array can also be successfully used to express multiple gRNAs in yeast (Zhang et al., 2019). The system is called GTR-CRISPR (gRNA-tRNA array for CRISPR/Cas9). GTR-CRISPR used a plasmid that contained a Cas9 expression cassette and a gRNA cassette with $tRNA_{Gly}$ sequences between each gRNA. The tRNA sequences were then cleaved during endogenous yeast tRNA processing, resulting in the release of multiple gRNAs from a single transcript (**Figure 2D**). The authors tested two different arrays to disrupt eight genes at the same time. One array used a single promoter to express all gRNAs, while the other array used two promoters to express eight gRNAs (four gRNAs each). The latter approach resulted in an incredible octuple gene deletion with 87% efficiency, compared to 35.5% efficiency for the former arrangement (Zhang et al., 2019). This tRNA array system has been utilized to integrate the pyruvate dehydrogenase complex (PDH) into the yeast genome (Zhang et al., 2020).

Although these RNA cleavage mechanism approaches are valuable, they have a major limitation. The efficiency of multiplex genome editing using this approach is determined by the lowest expression and cleavage efficiency of its gRNAs. One common issue observed from these studies was that the more downstream a gRNA or crRNA is in the transcript, the less efficient it will be (Ryan et al., 2014; Bao et al., 2015; Ferreira et al., 2018; Zhang et al., 2019). As multiplex gene integration requires multiple factors to be efficient (discussed below), the compounding effects of the inefficiencies of RNA cleavage mechanisms can significantly restrict the application of these approaches for multiplex gene integration. Therefore, the application of these approaches is more widespread for multiple gene disruption (e.g., for disrupting endogenous competing pathways) rather than multiple gene integrations for pathway building.

Expression of Multi-Cassettes gRNAs

The second strategy of expressing multiple gRNAs simultaneously is to express each gRNA with its own cassette. This approach has been simplified by the development of molecular cloning techniques, such as Golden Gate, Gibson Assembly, USER, and *in vivo* DNA assembler (Deaner and Alper, 2019). Readers can refer to a work by Chao et al., who reviewed the current development of molecular cloning and DNA assembly techniques (Chao et al., 2015). These techniques allow the creation of a large plasmid with relatively quick and straightforward steps. The examples that will be discussed below use multiple gRNA cassettes in one plasmid, while Cas9 is expressed separately either by integration into the genome or using a separate plasmid. Thus, these examples will be categorized into two groups: (i) integrated Cas9 expression and (ii) plasmid-based Cas9 expression.

Integrated Cas9 Expression

In the first category, a Cas9 expression cassette is integrated into the yeast genome, while gRNA cassette(s) are expressed in a plasmid (**Figure 3A**). The integration of the Cas9 cassette into the genome has some benefits, such as more stable Cas9 expression, maximizing the availability of selection markers for gRNA-expressing plasmids, and reducing the size of the plasmid, thereby, increasing the transformation efficiency (Stovicek et al., 2017). However, the Cas9 cassette in the genome needs to be removed after genome editing. Mans et al. developed a comprehensive toolbox to simplify CRISPR/Cas9 gene editing in yeast (Mans et al., 2015). This toolbox includes (i) the Yeastriction Web tool,[1] which can help with gRNA design to minimize off-target edits and maximize efficiency; (ii) a set of gRNA expressing plasmids, which includes eight single-gRNA cassette plasmids and eight double-gRNA cassette plasmids with eight different selectable markers (*URA3*, *amdSYM*, *hphNT1*, *kanMX*, *LEU2*, *natNT2*, *HIS3*, and *TRP1*); and (iii) a collection of various haploid, diploid, and auxotrophic CEN.PK strains, into which the Cas9 cassette, under the *TEF1* promoter, has been integrated. Both plasmids and yeast strains were deposited at Euroscarf. The authors showed that using three plasmids, each of which contained two gRNA cassettes, could simultaneously delete six genes with 65% efficiency. Furthermore, they deleted two acetyl-CoA synthetase genes, *ACS1*, and *ACS2*, which are important for cytosolic acetyl-CoA synthesis in yeast. Without these genes, the yeast is not viable unless other sources of acetyl-CoA are supplied. Therefore, they also simultaneously integrated six genes that are part of the *E. faecalis* pyruvate dehydrogenase complex as donor DNAs, to provide the yeast with an alternative acetyl-CoA biosynthesis pathway. All six genes were designed to be assembled *in vivo* into the *ACS2* locus, while the *ACS1* locus was deleted by providing only a 120-bp non-functional donor DNA. Although the efficiency cannot be calculated because failed integrations will not appear as colonies, the study successfully deleted two genes and introduced six heterologous genes (~15 kb) into one locus of the yeast genome in one transformation (Mans et al., 2015).

Similarly, Horwitz et al. also integrated a Cas9 cassette into the yeast genome (Horwitz et al., 2015). They expressed Cas9 under the medium-strength promoter, *FBA1*, as opposed to the high-strength promoter, *TEF1*. More importantly, they used different gRNA plasmid delivery methods. Instead of constructing various

[1]http://yeastriction.tnw.tudelft.nl/

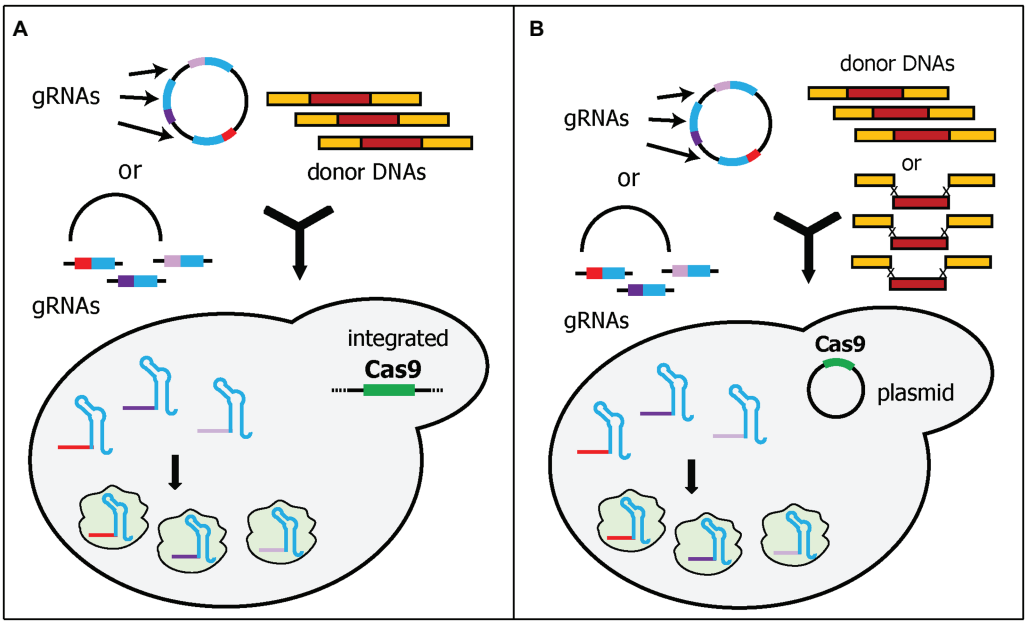


FIGURE 3 | Schematic diagram of multiplex gene editing using multi-cassette gRNA expression. Each gRNA is expressed in one cassette. The gRNA plasmid can be pre-assembled or transformed as linear fragments. The linear fragments can be assembled *in vivo* by yeast endogenous homologous recombination. **(A)** *Cas9* was integrated into the genome to reduce the amount of DNA to be transformed during transformation. **(B)** Cas9 plasmid was pre-transformed into the yeast cells. The studies also demonstrated that donor DNAs can be transformed as separate parts rather than pre-assembled donor DNAs.

plasmids with either one or two gRNAs, Horwitz et al. transformed the yeast cells with one linear plasmid and one, two, or three linear gRNA cassette(s), and then relied upon *in vivo* DNA assembly to circularize the plasmid for gRNA expression (**Figure 3A**). Interestingly, while Mans et al. attempted this approach and had a very low efficiency, Horwitz et al. showed that triple gene deletions can be achieved with 64% efficiency (Horwitz et al., 2015; Mans et al., 2015). The differences between these two studies are the length of the flanking homology sequences for *in vivo* assembly. Horwitz et al. used 500-bp flanking homology sequences (Horwitz et al., 2015), whereas Mans et al. used only 50-bp flanking homology sequences (Mans et al., 2015). Ultimately, Horwitz et al. also demonstrated the capability of this system to integrate 11 genes (~24-kb) in the muconic acid biosynthetic pathway into three loci of the competing pathway. Although the efficiency was low (4.2%), it shows the ability to delete competing pathways and integrate heterologous pathways simultaneously (Walter et al., 2016).

Plasmid-Based Cas9 Expression

Instead of integrating the Cas9 cassette into the yeast genome, the Cas9 cassette and gRNA cassette(s) can be expressed from two different plasmids (**Figure 3B**). This approach has a major benefit over the integration of Cas9 because the removal of a Cas9 expression plasmid is more straightforward, e.g., using counter selection method. However, the transformation efficiency may decrease due to the large size of DNA required to be delivered if all plasmids are transformed at the same time. Therefore, some of the examples below pre-transformed the Cas9-expressing plasmid before the transformation of the gRNA-expressing plasmid. A collaboration between the groups in the Novo Nordisk Foundation for Biosustainability developed four analogous studies using this system with the main differences being the cloning strategies of gRNA plasmids and donor DNAs (Jakočiūnas et al., 2015a,b; Ronda et al., 2015; Jessop-Fabre et al., 2016). In all four studies, the Cas9 expression cassette was expressed in one plasmid and was pre-transformed into a yeast strain. The constructed yeast strain was then transformed with the gRNA cassette(s)-containing plasmid.

Jakočiūnas et al. built the foundation of their system by systematically selecting target sites using the CRISPy Web tool,[2] verifying off-target effects, and utilizing USER cloning to assemble the gRNA cassette(s) plasmid (Jakočiūnas et al., 2015b). Using this approach, they successfully generated a collection of 31 mutant strains with one to five endogenous genes being disrupted. These five genes were chosen as single deletions of these genes resulted in a higher metabolite flux toward the mevalonate (MVA) pathway. The screening of the 31 mutant strains resulted in a yeast strain with a titer of more than 10 μM MVA (Jakočiūnas et al., 2015b). The following study combined this system and yeast *in vivo* DNA assembly to bypass the requirements for donor DNA cloning (Jakočiūnas et al., 2015a). In this approach, called CasEMBLR, up to three donor DNAs, containing five parts each (two flanking homology sequences, promoter, gene, and terminator), were transformed with the gRNA-expressing plasmid into yeast cells. The study demonstrated simultaneous integration of three genes in the

[2]http://staff.biosustain.dtu.dk/laeb/crispy/

β-carotene biosynthetic pathway (~18 kb) into the yeast genome with 31% efficiency (Jakočiūnas et al., 2015b).

The last two studies utilized EasyClone for their target sites. EasyClone consists of a set of standardized plasmids that combines USER cloning and Cre-LoxP-based marker recycling system to enable iterative integration of heterologous genes (Jensen et al., 2014). This method uses previously characterized target sites in the genome with high efficiency, low effects on cell growth, and high expression (Jensen et al., 2014). The following studies used and modified EasyClone target sites and plasmids for CRISPR/Cas9 genome editing (Ronda et al., 2015; Jessop-Fabre et al., 2016). In both studies, the Cas9-expressing plasmid was also pre-transformed into yeast cells. Ronda et al. used three characterized target sites from EasyClone to integrate three genes (~18 kb) for β-carotene biosynthesis into the yeast genome with 85% efficiency (Ronda et al., 2015). In the next study, Jessop-Fabre et al. evaluated 11 previously characterized EasyClone target sites for their efficiencies in CRISPR/Cas9 editing (Jessop-Fabre et al., 2016). All plasmids for EasyClone-based CRISPR/Cas9 editing are available at Addgene. They found that the target sites had a 95–100% targeting efficiency. The practicality of this system for multiplex CRISPR/Cas9 editing was successfully demonstrated by integrating either three (one gene in each locus) or six (two genes in each locus) genes in three different acetyl-CoA synthesis pathways from different species into the yeast genome with 60–70% efficiency. Interestingly, targeting different loci in the same chromosome decreases the editing efficiency. By comparing these three different pathways, 3-hydroxypropionic acid production in yeast was optimized (Jessop-Fabre et al., 2016).

Expression of multiple gRNAs in multiple gRNA cassettes for multiplex genome editing has one major advantage compared to the RNA cleaving approach for producing multiple gRNAs from one cassette. This benefit is that the expression of each gRNA is more comparable, which re-directs the rate-limiting steps toward other factors (e.g., yeast transformation efficiency or overall gRNA expression) instead of the lowest gRNA expression. Consequently, more studies utilize this approach for multiplex heterologous gene integrations (Bond and Tang, 2019; Yee et al., 2019). However, one key limitation persists that is the finite amount of gRNA that can be expressed. This is because the more gRNAs expressed by a plasmid, the larger the plasmid will be, which decreases the transformation efficiency. Additionally, multiple gRNAs might still compete for the yeast endogenous RNA transcription machinery and limit the expression of gRNAs transcripts.

Editing of Multiple Pre-defined Sequences in the Genome by a Single gRNA

So far, the maximum loci that have been demonstrated for simultaneous integration using multiple gRNAs are three (Horwitz et al., 2015). To improve the number of loci for deletions and possibly to increase the number of genes for integrations, other approaches have been developed. In the following approaches, pre-defined target sequences were identified and/or synthetically integrated in the yeast genome. The sequences, which can be endogenous or synthetic, must be found or pre-integrated in multiple sites in the yeast genome. Since multiple loci can be targeted using a single gRNA, it greatly simplifies the integration of the heterologous genes. Several studies have successfully used this approach to integrate multiple genes in multiple loci. These studies can be divided into two categories using either endogenous target sequences or synthetic target sequences.

Yeast Endogenous Target Sequences

Some endogenous sequences can be found in multiple sites in the yeast genome. One such endogenous sequence is the delta (δ) site of Ty (transposons of yeast) elements. Like other retrotransposons, Ty can replicate and insert itself in other sites of the yeast genome (Krastanova et al., 2005). Ty is composed of two genes, which are flanked by identical sequences called long terminal repeats (LTRs). There is at least five known Tys (Ty1 to Ty5), and they are scattered throughout the genome. The δ sites refer to the LTRs of Ty1 and Ty2. There are approximately 40 copies of Ty1 and Ty2 LTRs in haploid yeast (Krastanova et al., 2005). Ty1 and Ty2 LTR sequences have been utilized since the 1990s for the integration of multiple gene copies in yeast (Da Silva and Srikrishnan, 2012; Malcı et al., 2020). However, this approach shows poor efficiencies, especially for integrating larger genes (Shi et al., 2016). Inspired by conventional δ site integration, Shi et al. designed a CRISPR/Cas9 system to target these δ sites, called delta integration-based CRISPR (Di-CRISPR; Shi et al., 2016). It exploits HI-CRISPR (Bao et al., 2015) plasmids to separately express crRNA, tracrRNA, and Cas9, although only one crRNA is expressed in Di-CRISPR. The crRNA was designed to target a characterized δ site sequence (**Figure 4A**). Using this system, the authors demonstrated an astounding 24-kb cassette integration into the δ sites with up to 18 copies of each gene being found in the genome. The 24-kb donor DNA cassette contained seven genes: three xylose utilization genes, three (*R*,*R*)-2,3-butadienol (BDO) biosynthetic genes, and a GFP reporter gene. The resulting strain can utilize xylose as its sole carbon source and produce BDO (Shi et al., 2016). The same group has extended the application of Di-CRISPR for developing automated system to study genotype–phenotype mapping and industrial traits optimization (Si et al., 2017). Recently, another study adopted a similar approach and successfully integrated 25 copies of the BDO biosynthetic pathway and a GFP cassette into δ sites, albeit with a much shorter donor DNA of 4-kb (Huang and Geng, 2020). In another study, Ty elements, instead of δ sites, were targeted for multicopy multiplex genome integration (Hanasaki and Masumoto, 2019). This system, called CRISPR/transposon gene integration, used two plasmids, a Cas9-gRNA expressing plasmid and a donor DNA plasmid. Interestingly, the donor DNA plasmid contains a Ty1 genome sequence, which will be cleaved together with the Ty1 sequences in the genome, by Cas9-gRNA complex. This causes the donor DNA plasmid to be linearized *in vivo* and integrated into the Ty1 sequences. Using this approach, the authors demonstrated the integration of 12 copies of the donor DNA (~5-kb). Additionally, the expression of the genes in the donor DNA can be tuned by exploiting the existence of amino acid markers in the donor DNA plasmid (Hanasaki and Masumoto, 2019).

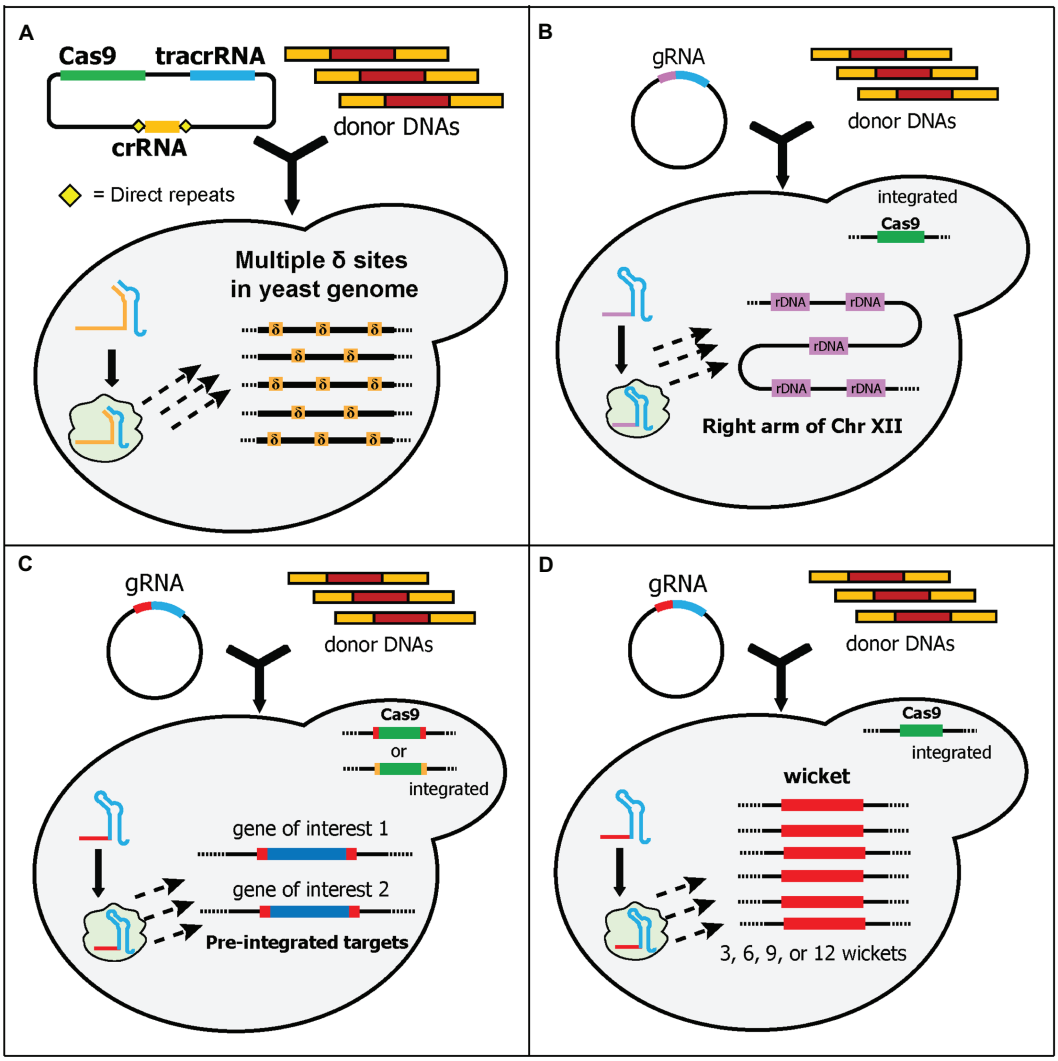


FIGURE 4 | Schematic diagram of multiplex gene-editing pre-defined sequences in the genome using a single gRNA (part 1). **(A)** Shi et al. used δ sites of Ty elements in the yeast genome as the target. More than 30 δ sites (δ, orange boxes) are scattered across all chromosomes in the yeast genome. Therefore, multiple copies of donor DNA can be integrated. Note that they used a separate crRNA and tracrRNA plasmid. **(B)** Wang et al. targeted ribosomal DNA (rDNA, purple boxes) for integration sites. rDNA is located at the right arm of chromosome XII. Around 100–200 copies of rDNA can be found in the yeast genome. *Cas9* was integrated into the genome. **(C)** Finnigan and Thorner generated synthetic target sequences (red or orange boxes) that flanked the genes-of-interest (GOIs) and *Cas9*, which was integrated into the yeast genome. The gRNA cleaves the sequences and causes DSBs at the 5' and 3' ends of GOIs and/or *Cas9*, removing them from the genome. Concurrently, the donor DNAs can be integrated at those sites. Note that although different gRNA sequences can be used, only one gRNA sequence (red box) is shown. **(D)** Hou et al. generated "wicket" sequences (red boxes) which contains a 23-bp synthetic target flanked by 5' and 3' 50-bp synthetic homology arms. Wickets and *Cas9* were integrated into the yeast genome. Yeast strains with 3, 6, 9, or 12 wickets were created. Therefore, multi-copies of donor DNA can be integrated in one transformation.

Another endogenous sequence, ribosomal DNA (rDNA), has also been utilized as a target for multicopy gene integration (Wang et al., 2018). rDNA is located at the *RDN1* locus. The *RDN1* locus is a 1–2 Mb section in chromosome XII of yeast and contains 100–200 copies of a 9.1 kb repeat. Each repeat has regions that encode rRNAs and non-coding regions (Venema and Tollervey, 1999). Wang et al. targeted one of the non-coding regions called the non-transcribed spacers 1 (NTS1) for their system (Wang et al., 2018). In this system, the Cas9 expression cassette was pre-integrated into the genome before the transformation of gRNA plasmid and donor DNA fragment (**Figure 4B**). Although it has been shown that cleaving multiple targets in the same chromosome may disrupt genome stability and reduce the efficiency of CRISPR/Cas9 (Jessop-Fabre et al., 2016), the authors demonstrated the successful integration of up to 10 copies of a GFP donor DNA cassette with 45% efficiency, and the resulting strain maintained stable copy numbers after 55 generations (Wang et al., 2018). In one study, the CRISPR-based sequential integration of four genes into both δ sites and rDNA successfully increased the production of isobutanol in yeast (Park and Hahn, 2019).

Synthetic Target Sequences

Designing an efficient gRNA is one of the most essential factors for successful CRISPR/Cas9 gene editing (DiCarlo et al., 2013). Therefore, the laborious step of testing multiple gRNAs for efficiently integrating heterologous genes in each locus is required. This difficulty is compounded when targeting multiple loci. Moreover, targeting endogenous yeast sequences may have some unintended off-target effects, especially if the targets are not yet tested (Apel et al., 2017). The following studies have developed yeast strains to alleviate these problems by integrating artificial/synthetic gRNA target sequences, which are designed to avoid any potential off-target DSBs, at characterized loci with significant gene expression. The idea of introducing synthetic sequences for a unique target site was first demonstrated by Lee et al. for investigating the effect of a linearized plasmid on integration efficiency (Lee et al., 2015). They termed this synthetic sequence a "landing pad." Although this study did not attempt to develop a multiplex genome-editing system, the term and concept of a "landing pad" have been used in the following five studies.

In the first study, Finnigan and Thorner generated unique 23 bp synthetic sequences, each of which contained a 20-nt target sequence and 3-bp PAM sequence (Finnigan and Thorner, 2016). Each sequence was integrated into two locations that flanked a gene of interest in the genome (**Figure 4C**). Thus, expression of the gRNA and Cas9 cuts both synthetic sequences and can be used to substitute the gene of interest with donor DNA fragments. Additionally, they integrated the Cas9 expression cassette, which is also flanked by the synthetic sequence, into the genome. Depending on the design of the synthetic sequences, the editing could result in different final yeast strains (**Figure 4C**). If the synthetic sequences for the Cas9 cassette and the genes of interest are identical, the final strain will not only contain the substituted genes but also remove the Cas9 cassette in the genome. Otherwise, the final strain could contain the substituted genes but still retain the Cas9 cassette in the genome. Using this system, the authors demonstrated the successful substitution of three genes, with and without the removal of the Cas9 cassette (Finnigan and Thorner, 2016).

The next study from Hou et al. expanded the landing pad concept by synthesizing and integrating an artificial sequence containing a 50-bp 5' homology arm, 20-nt target sequence, 3-bp PAM sequence, and 50-bp 3' homology arm, which they called a "wicket" (Hou et al., 2018). Wicket is a wooden structure in cricket, composed of three stumps (i.e., signifying a left homology arm, central gRNA target sequences, and a right homology arm). Therefore, only one gRNA cassette and one set of universal homology sequence for all donor DNAs are required for editing multiple loci. They also pre-integrated the Cas9 cassette into the genome and constructed yeast strains containing 3, 6, 9, or 12 wickets in various intergenic regions (**Figure 4D**). To evaluate this system, they used three genes in the β-carotene biosynthetic pathway as donor DNAs. In the process called "pre-assembled integration," they used one donor DNA containing the 5' and 3' homology arms with all three genes cassettes in the middle. Using this process, the efficiency was between 50 and 100% for strains with 3, 6, and 9 wickets, while the efficiency for 12 wickets was very low. Alternatively, in the process called "cocktail integration," they used three donor DNAs containing the same 5' and 3' homology arms with only one gene cassette in the middle of each donor DNA. Cocktail integration was intended to control the copy numbers of each donor DNA cassette in the genome. Using this approach, a variety of strains with different amounts of β-carotene were generated, although the efficiency was much lower compared to the pre-assembled donor DNA. Interestingly, they found some tandem duplication events between the donor DNA cassettes in both processes, which caused some strains with three wickets to have up to 20 copies of the donor DNA cassette (Hou et al., 2018).

The study by Bourgeois et al. attempted to build a yeast strain that can be used to compare and optimize the copy number of heterologous gene integration. To achieve this, they meticulously evaluated 10 synthetic landing pads (with 280-bp 5' and 3' homology arms and a 23-nt synthetic target sequence) and 16 genomic loci for their integration efficiency and gene expression levels (Bourgeois et al., 2018). After the rigorous characterization, they ranked and picked four synthetic landing pads and 10 genomic loci to be used. The four landing pads were integrated into one, two, three, and four genomic loci, respectively (**Figure 5A**). As a result, this yeast strain can be utilized for integrating fixed copies of genes in one transformation. They successfully demonstrated the utility of these strains by comparing 10 different norcoclaurine synthases (NCSs), each of which had one, two, three, or four copies, resulting in 40 strains being generated and evaluated in a relatively short time. Based on these strains, they unambiguously determined the best NCS and the optimum copy number to produce the highest titer of (S)-norcoclaurine (Bourgeois et al., 2018).

Recently, Baek et al. also characterized synthetic sequences (consisting of a 20-nt target sequence and 3-bp PAM sequence) and genomic loci, albeit on a smaller scale (two synthetic sequences and 12 intergenic loci) than Bourgeois et al. (Bourgeois et al., 2018; Baek et al., 2021). The authors integrated the most efficient synthetic sequence into one to six intergenic loci with the highest integration efficiency, resulting in six different strains with one to six copies of the synthetic sequence in the genome (**Figure 5B**). To simplify the cloning process for the donor DNA, six corresponding plasmids were generated. Using the strain containing six synthetic targets, this platform successfully integrated six genes for kauniolide biosynthesis (~15 kb) into six different loci with 40% efficiency (Baek et al., 2021). So far, this is the highest number of unique gene cassettes to be integrated into unique loci in yeast (**Table 1**). Although Shi et al. managed to remarkably edit 18 loci of δ sites, each locus contained the same cassette (Shi et al., 2016). Similarly, Horwitz et al. astoundingly integrated 11 gene cassettes, but these cassettes were integrated into three different loci (Horwitz et al., 2015).

In another study, Qi et al. utilized the highly efficient genomic loci from Apel et al. and built a yeast strain containing eight well-defined cassettes in eight different loci (Apel et al., 2017; Qi et al., 2021). Each cassette contained a unique set of promoters and terminators with a synthetic linker in the middle, resulting in a strain with eight promoter-linker-terminator cassettes in the genome (**Figure 5C**). Since each linker had different sequences, unique gRNA cassettes were required to target each site, much

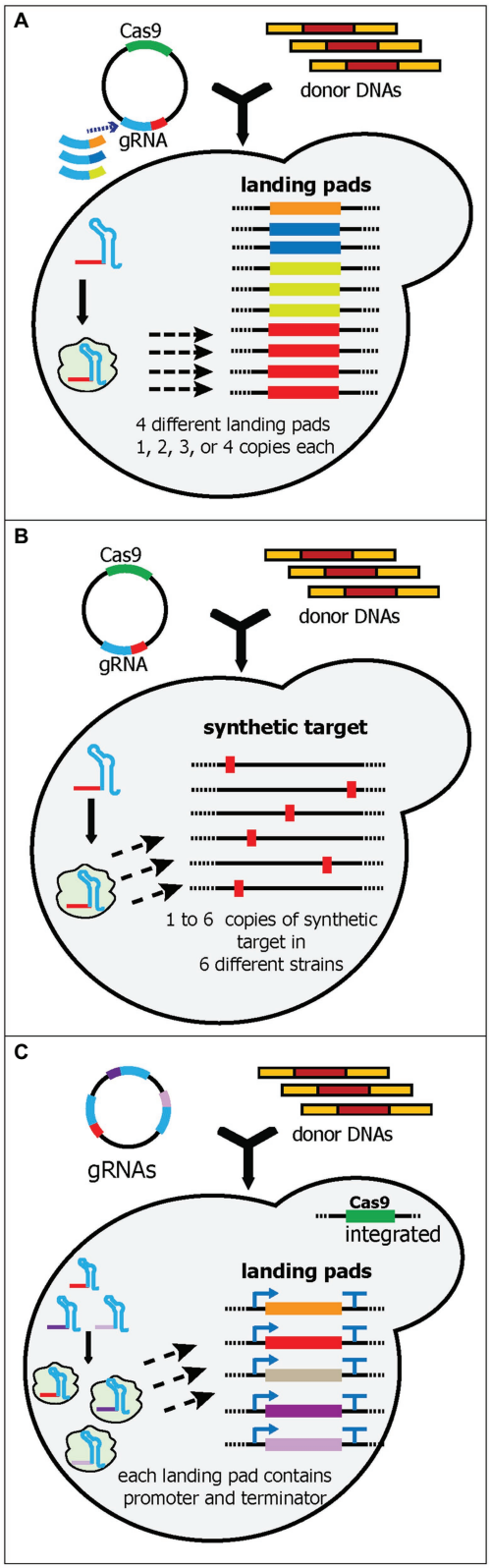


FIGURE 5 | Schematic diagram of multiplex gene-editing pre-defined sequences in the genome using a single gRNA (part 2). **(A)** Bourgeois et al. *(continued)*

FIGURE 5 | tested and selected various landing pads sequences. Like wickets, the landing pads contain a 23-bp synthetic target flanked by 280-bp 5' and 3' synthetic homology arms. Four unique landing pads (orange, dark blue, light green, and red boxes) were integrated into 10 characterized loci in the yeast genome. Each landing pad had a different copy number in the genome; i.e., landing pad 1 has one, and landing pad 2 has two. Note that four different plasmids were created, each of which targeted a different landing pad. Only one landing pad gRNA (red) was shown in the yeast cell. **(B)** Baek et al. generated and integrated a 23-bp synthetic sequence (red boxes) into one to six characterized loci in the yeast genome. Six yeast strains were constructed, each of which had from one to six synthetic sequence(s). Therefore, up to six genes could be integrated simultaneously. **(C)** Qi et al. created a system called PCR & Go. Up to eight 23-bp synthetic sequences flanked with unique sets of promoters and terminators were integrated into the yeast genome. The gRNA plasmid contains multiple gRNA cassettes instead of a single gRNA. However, donor DNA preparation can be simplified as unique sets of promoters and terminators had already been integrated into the genome.

like using multiple gRNA cassettes to target multiple sites. However, because the Cas9 cassette, promoters, and terminators were pre-integrated into the genome, the transformation only required plasmid(s) with gRNA cassettes and donor DNA genes (without promoters and terminators). This platform successfully integrated five genes in the astaxanthin biosynthetic pathway into five different loci with 69% efficiency (Qi et al., 2021).

The utilization of pre-defined endogenous or synthetic target sequences has dramatically increased the maximum number of genes that can be integrated into unique loci. This is largely because the total amount of DNA transformed into yeast can be reduced due to either shorter gRNA (typically only one is required) plasmids (Finnigan and Thorner, 2016; Shi et al., 2016; Bourgeois et al., 2018; Baek et al., 2021) or shorter donor DNAs (Qi et al., 2021). Nevertheless, this approach has two major limitations. First, the targets are pre-determined and therefore cannot be used to disrupt endogenous competing pathways. Second, constructing the base strain requires additional effort for the characterization and integration of the pre-defined synthetic sequences.

Another study that does not completely fit into these three multiplex categories but contains interesting ideas was carried out by Ferreira et al. They attempted to use off-target effects to knock-out multiple genes simultaneously (Ferreira et al., 2017). The authors built a bioinformatics tool to predict the promiscuity of gRNA sequences and used promiscuous gRNA to target multiple genes at once. Using one promiscuous gRNA, the double knockout of *FAA1* and *FAA4* was achieved with 100% efficiency (Ferreira et al., 2017). Despite this ingenious approach, there is a major drawback in the limited availability of promiscuous sequences. As one may expect, most of the promiscuous gRNA sequences are within the same gene families, transposons, or paralogs.

MULTIPLEX GENOME EDITING USING Cas12a

Cas9 is the first and most popular Cas protein for genome-editing purposes in yeast (Jinek et al., 2012; Malcı et al., 2020). However, another Cas protein, called Cas12a (previously Cpf1),

has started to gain attentions for genome editing, including multiplex gene editing in yeast. There are three important differences between Cas9 and Cas12a: (i) Cas12a recognizes a 5'-(T)TTV-3' PAM sequence, and it cleaves the sequences downstream of the PAM instead of upstream like Cas9; (ii) Cas12a only needs crRNA to function as an endonuclease, instead of the crRNA and tracrRNA used by Cas9; (iii) Cas12a has an inherent capability to process pre-crRNA, whereas Cas9 requires host RNase activity for pre-crRNA processing. As shown by Bao et al. for multiplex CRISPR/Cas9 (see above), pre-crRNA processing can be useful for expressing multiple gRNAs in one cassette (Bao et al., 2015). Therefore, Cas12a has huge potential to be used for an efficient multiplex genome-editing system. For a more comprehensive comparison between Cas9 and Cas12a, the readers can refer to a recent review (Paul and Montoya, 2020).

The first utilization of Cas12a for multiplex genome editing was demonstrated in mammalian cells and the mouse brain (Zetsche et al., 2017). In yeast, at least two studies have demonstrated the utilization of Cas12a for multiplex genome integration. The first study used one plasmid containing a Cas12a cassette and crRNA array. They successfully integrated three genes (~9 kb) for the β-carotene biosynthesis pathway into three different loci with 91% efficiency (Verwaal et al., 2018). Similarly, the second study also used a one plasmid system and resulted in the integration of four genes (~13 kb) for β-carotene production into three different loci (two genes in the same locus) with up to 32% efficiency. They also integrated four genes (~7.5 kb) for the biosynthesis of the sesquiterpene, patchoulol, into three different loci (*FPPS* and *PTS* were linked) with 30% efficiency (Li et al., 2018). Although studies for multiplex genome editing using Cas12a are still rare, increasing understanding of Cas12a mechanism will accelerate the applications of this system for multiplex genome editing in yeast.

OTHER APPLICATIONS OF MULTIPLEX CRISPR/Cas9

Other than gene integration, multiplex CRISPR/Cas9 can also be used to optimize yeast strains *via* metabolic engineering strategies, such as gene disruption for eliminating competing pathways, gene downregulation for diminishing competing but important pathways, and gene upregulation for boosting endogenous yeast pathways (Siddiqui et al., 2012; Sander and Joung, 2014; Pyne et al., 2019). As mentioned above, multiple gene disruptions or gene deletions are usually implemented using RNA-cleaving mechanisms (HDV, tRNA, Csy4, and HI-CRISPR) due to their simplicity and high efficiencies in disrupting multiple genes (Ryan et al., 2014; Bao et al., 2015; Ferreira et al., 2018; Zhang et al., 2019). The systems for endogenous gene downregulation and upregulation are less developed, but some exciting progress has been demonstrated. CRISPR/Cas9-based gene downregulation (CRISPR interference - CRISPRi) usually exploits deactivated Cas9 (dCas9), which was generated by mutating the nucleases domains of Cas9 (Qi et al., 2013). dCas9 does not create DSBs in the target sites, but still tightly binds to the target sites, this causes repression of gene expression downstream of the target sites. This repression, which efficiency can be increased by fusion of dCas9 with different repressive chromatin modifier domains, is caused by the dCas9 sterically hindering RNA polymerase binding (Gilbert et al., 2013; Qi et al., 2013). Multiplex CRISPRi in yeast has been successfully demonstrated by simultaneously repressing seven yeast genes to increase β-amyrin production (Ni et al., 2019). The septuple site targeting efficiency was 40%, with the repression of each gene being between 60 and 80% (Ni et al., 2019). CRISPR/Cas9-based gene upregulation (CRISPR activation – CRISPRa) has also been developed by fusion of the dCas9 protein with strong transcriptional activator domains, such as VP64, p65AD, Rta, or a combination of them (Farzadfard et al., 2013). Two studies have demonstrated at least simultaneous double-gene activation using either VP64 (Zalatan et al., 2015) or V64-p65AD-Rta (Deaner et al., 2017) as the activator domains. Moreover, both studies did not only show the double activations but also show interference of other genes, which allows simultaneous activation and repression in one transformation (Zalatan et al., 2015; Deaner et al., 2017). Another study combined CRISPRa, CRISPRi, and gene deletion using optimized versions of dCas12a, dCas9, and Cas9, respectively, which they called CRISPR-AID (Lian et al., 2017). CRISPR-AID integrated the three nucleases into the genome and exploited the Csy4 system to process the gRNAs. Using this approach, upregulation of *HMG1*, downregulation of *ERG9*, and deletion of *ROX1* were achieved, and the β-carotene titer increased by 3-fold (Lian et al., 2017). More complete reviews of CRISPRi and CRISPRa are available elsewhere (Jensen, 2018).

PERSPECTIVES

The advancement of CRISPR/Cas9 from its discovery as a bacterial immune system to multiplex genome-editing applications has been revolutionary. Despite these remarkable innovations, CRISPR/Cas9 multiplex gene integration has several challenges that need to be addressed to improve the capacity and efficiency of this technique. First, the gRNA design and target loci selection steps need to be minimized. Designing gRNAs is a crucial step for successful CRISPR/Cas9 gene editing, including multiplex gene integration (DiCarlo et al., 2013; Adiego-Pérez et al., 2019). Although the development of various assisting tools for gRNA designs has been greatly improved for various Cas proteins (Concordet and Haeussler, 2018; Labuhn et al., 2018; Labun et al., 2019; Liao et al., 2019), the *in vivo* efficiency of these gRNAs still has to be scrutinized, and this is even more significant for multiplex gene integration (Bourgeois et al., 2018). Moreover, the selection of target loci also plays a significant role in the integration efficiency. For example, Baek et al. found that certain target sites in gene-sparse loci were highly inefficient, which may be due to limited chromatin accessibility (Baek et al., 2021). Similar observations were also demonstrated elsewhere (Mans et al., 2015; Bourgeois et al., 2018; Verkuijl and Rots, 2019).

The characterization of efficient gRNAs and target loci, as well as the installation of synthetic landing pads, would greatly increase the capability and efficiency of multiplex gene integration (Apel et al., 2017; Bourgeois et al., 2018; Baek et al., 2021). Second, HDR as the DSB repair mechanism in yeast needs to be optimized. Baker's yeast is known to prefer HDR over non-homologous end joining (NHEJ) for repairing DSBs. This is demonstrated by expressing Cas9 and gRNA without donor DNA in a cell which can cause toxicity in yeast. However, some studies of multiplex gene integration showed that some yeast colonies can still survive without donor DNAs, indicating that NHEJ operates to repair the DSBs (Baek et al., 2021). This can cause an increase in false positive colonies and reduce the efficiency of multiplex genes integration. Deletions of some genes involved in the NHEJ pathway, such as *POL4*, *DNL4*, and *Ku70*, were shown to reduce false positive rates and increase the success of HDR (Lemos et al., 2018; Yan and Finnigan, 2018). Therefore, the deletion of the NHEJ genes may increase the capability and efficiency of multiplex genome integration by reducing the number of background colonies, generated by chromosome repairs by NHEJ. Third, yeast transformation efficiency needs to be improved. CRISPR/Cas9 multiplex gene integration in yeast requires the introduction of a large amount of foreign DNA (donor DNA cassettes, gRNA plasmids, and Cas9 plasmids) to yeast cells. The more targets to be edited, the more donor DNA and/or gRNA cassettes will be required. Transformation techniques, such as electroporation or the addition of amino acids, can be incorporated to improve the transformation efficiency (Benatuil et al., 2010; Yu et al., 2019). Finally, the development of CRISPR/Cas9 multiplex genome integration in yeast can be improved by combining multiple methods. For example, the combination of the pre-installed target sites (Bourgeois et al., 2018; Baek et al., 2021) with RNA cleaving mechanisms (Bao et al., 2015; Shi et al., 2016; Ferreira et al., 2018) or with a sequential integration approach (Li et al., 2020) can minimize the required time and maximize the number of genes for multiplex integration. The currently developed methods (**Table 1**) have been remarkable and have endless potential. However, the success of addressing these limitations and creatively combining several methods will expand the scope of multiplex genome editing. In turn, this will accelerate the study and optimization of complex specialized metabolic pathways in yeast.

AUTHOR CONTRIBUTIONS

JCU and D-KR conceived the work. JCU and CH wrote the manuscript. JCU prepared the figures. D-KR revised the manuscript. All the authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the Technology Development Program (grant number, 20014582) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and by the Natural Sciences and Engineering Research Council of Canada (NSERC) to D-KR.

ACKNOWLEDGMENTS

We thank Susan Roth and Rahul Kumar for proofreading this article.

REFERENCES

Adiego-Pérez, B., Randazzo, P., Daran, J. M., Verwaal, R., Roubos, J. A., Daran-Lapujade, P., et al. (2019). Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.* 366:fnz086. doi: 10.1093/femsle/fnz086

Apel, A. R., d'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., et al. (2017). A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 45, 496–508. doi: 10.1093/nar/gkw1023

Baek, S., Utomo, J. C., Lee, J. Y., Dalal, K., Yoon, Y. J., and Ro, D.- K. (2021). The yeast platform engineered for synthetic gRNA-landing pads enables multiple gene integrations by a single gRNA/Cas9 system. *Metab. Eng.* 64, 111–121. doi: 10.1016/j.ymben.2021.01.011

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., et al. (2015). Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4, 585–594. doi: 10.1021/sb500255k

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140

Bathe, U., and Tissier, A. (2019). Cytochrome P450 enzymes: a driving force of plant diterpene diversity. *Phytochemistry* 161, 149–162. doi: 10.1016/j.phytochem.2018.12.003

Benatuil, L., Perez, J. M., Belk, J., and Hsieh, C.- M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* 23, 155–159. doi: 10.1093/protein/gzq002

Bond, C. M., and Tang, Y. (2019). Engineering *Saccharomyces cerevisiae* for production of simvastatin. *Metab. Eng.* 51, 1–8. doi: 10.1016/j.ymben.2018.09.005

Bourgeois, L., Pyne, M. E., and Martin, V. J. J. (2018). A highly characterized synthetic landing pad system for precise multicopy gene integration in yeast. *ACS Synth. Biol.* 7, 2675–2685. doi: 10.1021/acssynbio.8b00339

Chao, R., Yuan, Y., and Zhao, H. (2015). Recent advances in DNA assembly technologies. *FEMS Yeast Res.* 15, 1–9. doi: 10.1111/1567-1364.12171

Chapple, C. (1998). Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 311–343. doi: 10.1146/annurev.arplant.49.1.311

Concordet, J.- P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. doi: 10.1093/nar/gky354

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Da Silva, N. A., and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12, 197–214. doi: 10.1111/j.1567-1364.2011.00769.x

David, F., and Siewers, V. (2015). Advances in yeast genome engineering. *FEMS Yeast Res.* 15, 1–14. doi: 10.1111/1567-1364.12200

Deaner, M., and Alper, H. S. (2019). Enhanced scale and scope of genome engineering and regulation using CRISPR/Cas in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 19:foz076. doi: 10.1093/femsyr/foz076

Deaner, M., Mejia, J., and Alper, H. S. (2017). Enabling graded and large-scale multiplex of desired genes using a dual-mode dCas9 activator in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 6, 1931–1943. doi: 10.1021/acssynbio.7b00163

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607. doi: 10.1038/nature09886

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., and Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336–4343. doi: 10.1093/nar/gkt135

Doudna, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

Farzadfard, F., Perli, S. D., and Lu, T. K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* 2, 604–613. doi: 10.1021/sb400081r

Ferreira, R., Gatto, F., and Nielsen, J. (2017). Exploiting off-targeting in guide-RNAs for CRISPR systems for simultaneous editing of multiple genes. *FEBS Lett.* 591, 3288–3295. doi: 10.1002/1873-3468.12835

Ferreira, R., Skrekas, C., Nielsen, J., and David, F. (2018). Multiplexed CRISPR/Cas9 genome editing and gene regulation using Csy4 in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 7, 10–15. doi: 10.1021/acssynbio.7b00259

Finnigan, G. C., and Thorner, J. (2016). mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3* 6, 2147–2156. doi: 10.1534/g3.116.029801

Gardner, J. M., and Jaspersen, S. L. (2014). "Manipulating the yeast genome: deletion, mutation, and tagging by PCR," in *Yeast Genetics: Methods and Protocols*. eds. J. S. Smith and D. J. Burke (New York, NY: Springer), 45–78.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044

Hafner, J., Payne, J., Mohammadi, P. H., Hatzimanikatis, V., and Smolke, C. (2021). A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives. *Nat. Commun.* 12:1760. doi: 10.1038/s41467-021-22022-5

Hanasaki, M., and Masumoto, H. (2019). CRISPR/transposon gene integration (CRITGI) can manage gene expression in a retrotransposon-dependent manner. *Sci. Rep.* 9:15300. doi: 10.1038/s41598-019-51891-6

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355–1358. doi: 10.1126/science.1192272

Horwitz, A. A., Walter, J. M., Schubert, M. G., Kung, S. H., Hawkins, K., Platt, D. M., et al. (2015). Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts *via* CRISPR-Cas. *Cell Syst.* 1, 88–96. doi: 10.1016/j.cels.2015.02.001

Hou, S., Qin, Q., and Dai, J. (2018). Wicket: a versatile tool for the integration and optimization of exogenous pathways in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 7, 782–788. doi: 10.1021/acssynbio.7b00391

Huang, S., and Geng, A. (2020). High-copy genome integration of 2,3-butanediol biosynthesis pathway in *Saccharomyces cerevisiae via in vivo* DNA assembly and replicative CRISPR-Cas9 mediated delta integration. *J. Biotechnol.* 310, 13–20. doi: 10.1016/j.jbiotec.2020.01.014

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433. doi: 10.1128/jb.169.12.5429-5433.1987

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S. J., Kristensen, M., Pedersen, L. E., et al. (2015b). Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 213–222. doi: 10.1016/j.ymben.2015.01.008

Jakočiūnas, T., Jensen, M. K., and Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.* 34, 44–59. doi: 10.1016/j.ymben.2015.12.003

Jakočiūnas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., et al. (2015a). CasEMBLR: Cas9-facilitated multiloci genomic integration of *in vivo* assembled DNA parts in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4, 1226–1234. doi: 10.1021/acssynbio.5b00007

Jansen, R., Embden, J. D. A., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x

Jensen, M. K. (2018). Design principles for nuclease-deficient CRISPR-based transcriptional regulators. *FEMS Yeast Res.* 18:foy039. doi: 10.1093/femsyr/foy039

Jensen, N. B., Strucko, T., Kildegaard, K. R., David, F., Maury, J., Mortensen, U. H., et al. (2014). EasyClone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 14, 238–248. doi: 10.1111/1567-1364.12118

Jessop-Fabre, M. M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M. K., Keasling, J. D., et al. (2016). EasyClone-markerfree: a vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae via* CRISPR-Cas9. *Biotechnol. J.* 11, 1110–1117. doi: 10.1002/biot.201600147

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997. doi: 10.1126/science.1247997

Kotopka, B. J., and Smolke, C. D. (2019). Production of the cyanogenic glycoside dhurrin in yeast. *Metab. Eng. Commun.* 9:e00092. doi: 10.1016/j.mec.2019.e00092

Krastanova, O., Hadzhitodorov, M., and Pesheva, M. (2005). Ty elements of the yeast *Saccharomyces cerevisiae*. *Biotechnol. Biotechnol. Equip.* 19, 19–26. doi: 10.1080/13102818.2005.10817272

Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., et al. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res.* 46, 1375–1385. doi: 10.1093/nar/gkx1268

Labun, K., Montague, T. G., Krause, M., Torres, C. Y. N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47, W171–W174. doi: 10.1093/nar/gkz365

Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4, 975–986. doi: 10.1021/sb500366v

Lemos, B. R., Kaplan, A. C., Bae, J. E., Ferrazzoli, A. E., Kuo, J., Anand, R. P., et al. (2018). CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2040–E2047. doi: 10.1073/pnas.1716855115

Leonard, E., and Koffas, M. A. G. (2007). Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Appl. Environ. Microbiol.* 73, 7246–7251. doi: 10.1128/AEM.01411-07

Li, Z.- H., Meng, H., Ma, B., Tao, X., Liu, M., Wang, F.- Q., et al. (2020). Immediate, multiplexed and sequential genome engineering facilitated by CRISPR/Cas9 in *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 47, 83–96. doi: 10.1007/s10295-019-02251-w

Li, Z.- H., Wang, F.- Q., and Wei, D.- Z. (2018). Self-cloning CRISPR/Cpf1 facilitated genome editing in *Saccharomyces cerevisiae*. *Bioresour. Bioprocess.* 5:36. doi: 10.1186/s40643-018-0222-8

Lian, J., HamediRad, M., Hu, S., and Zhao, H. (2017). Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat. Commun.* 8:1688. doi: 10.1038/s41467-017-01695-x

Liao, C., Ttofali, F., Slotkowski, R. A., Denny, S. R., Cecil, T. D., Leenay, R. T., et al. (2019). Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat. Commun.* 10:2948. doi: 10.1038/s41467-019-10747-3

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x

Malcı, K., Walls, L. E., and Rios-Solis, L. (2020). Multiplex genome engineering methods for yeast cell factory development. *Front. Bioeng. Biotechnol.* 8:589468. doi: 10.3389/fbioe.2020.589468

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering *via* Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033

Mans, R., van Rossum, H. M., Wijsman, M., Backx, A., Kuijpers, N. G. A., van den Broek, M., et al. (2015). CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15:fov004. doi: 10.1093/femsyr/fov004

McCarty, N. S., Shaw, W. M., Ellis, T., and Ledesma-Amaro, R. (2019). Rapid assembly of gRNA arrays *via* modular cloning in yeast. *ACS Synth. Biol.* 8, 906–910. doi: 10.1021/acssynbio.9b00041

Meng, J., Qiu, Y., and Shi, S. (2020). CRISPR/Cas9 systems for the development of *Saccharomyces cerevisiae* cell factories. *Front. Bioeng. Biotechnol.* 8:594347. doi: 10.3389/fbioe.2020.594347

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182. doi: 10.1007/s00239-004-0046-3

Ni, J., Zhang, G., Qin, L., Li, J., and Li, C. (2019). Simultaneously down-regulation of multiplex branch pathways using CRISPRi and fermentation optimization for enhancing β-amyrin production in *Saccharomyces cerevisiae*. *Synth. Syst. Biotechnol.* 4, 79–85. doi: 10.1016/j.synbio.2019.02.002

Park, S.- H., and Hahn, J.- S. (2019). Development of an efficient cytosolic isobutanol production pathway in *Saccharomyces cerevisiae* by optimizing copy numbers and expression of the pathway genes based on the toxic effect of α-acetolactate. *Sci. Rep.* 9:3996. doi: 10.1038/s41598-019-40631-5

Paul, B., and Montoya, G. (2020). CRISPR-Cas12a: functional overview and applications. *Biom. J.* 43, 8–17. doi: 10.1016/j.bj.2019.10.005

Pyne, M. E., Narcross, L., and Martin, V. J. J. (2019). Engineering plant secondary metabolism in microbial systems. *Plant Physiol.* 179, 844–861. doi: 10.1104/pp.18.01291

Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A., and Arkin, A. P. (2012). RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* 30, 1002–1006. doi: 10.1038/nbt.2355

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Qi, M., Zhang, B., Jiang, L., Xu, S., Dong, C., Du, Y.- L., et al. (2021). PCR & go: a pre-installed expression chassis for facile integration of multi-gene biosynthetic pathways. *Front. Bioeng. Biotechnol.* 8:613771. doi: 10.3389/fbioe.2020.613771

Ronda, C., Maury, J., Jakočiūnas, T., Baallal, J. S. A., Germann, S. M., Harrison, S. J., et al. (2015). CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb. Cell Factories* 14:97. doi: 10.1186/s12934-015-0288-3

Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., et al. (2014). Selection of chromosomal DNA libraries using a multiplex CRISPR system. *elife* 3:e03703. doi: 10.7554/eLife.03703

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi: 10.1038/nbt.2842

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282. doi: 10.1093/nar/gkr606

Shao, Z., and Zhao, H. (2009). DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37:e16. doi: 10.1093/nar/gkn991

Shi, S., Liang, Y., Zhang, M. M., Ang, E. L., and Zhao, H. (2016). A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 33, 19–27. doi: 10.1016/j.ymben.2015.10.011

Si, T., Chao, R., Min, Y., Wu, Y., Ren, W., and Zhao, H. (2017). Automated multiplex genome-scale engineering in yeast. *Nat. Commun.* 8:15187. doi: 10.1038/ncomms15187

Siddiqui, M. S., Thodey, K., Trenchard, I., and Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* 12, 144–170. doi: 10.1111/j.1567-1364.2011.00774.x

Srinivasan, P., and Smolke, C. D. (2019). Engineering a microbial biosynthesis platform for *de novo* production of tropane alkaloids. *Nat. Commun.* 10:3634. doi: 10.1038/s41467-019-11588-w

Srinivasan, P., and Smolke, C. D. (2020). Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* 585, 614–619. doi: 10.1038/s41586-020-2650-9

Storici, F., Durham, C. L., Gordenin, D. A., and Resnick, M. A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14994–14999. doi: 10.1073/pnas.2036296100

Stovicek, V., Holkenbrink, C., and Borodina, I. (2017). CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Res.* 17:fox030. doi: 10.1093/femsyr/fox030

Tang, T.- H., Bachellerie, J.- P., Rozhdestvensky, T., Bortolin, M.- L., Huber, H., Drungowski, M., et al. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7536–7541. doi: 10.1073/pnas.112047299

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* 33, 261–311. doi: 10.1146/annurev.genet.33.1.261

Verkuijl, S. A. N., and Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Curr. Opin. Biotechnol.* 55, 68–73. doi: 10.1016/j.copbio.2018.07.005

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., and Roubos, J. A. (2018). CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae*. *Yeast* 35, 201–211. doi: 10.1002/yea.3278

Walter, J. M., Chandran, S. S., and Horwitz, A. A. (2016). CRISPR-Cas-assisted multiplexing (CAM): simple same-day multi-locus engineering in yeast. *J. Cell. Physiol.* 231, 2563–2569. doi: 10.1002/jcp.25375

Wang, L., Deng, A., Zhang, Y., Liu, S., Liang, Y., Bai, H., et al. (2018). Efficient CRISPR–Cas9 mediated multiplex genome editing in yeasts. *Biotechnol. Biofuels* 11:277. doi: 10.1186/s13068-018-1271-0

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3570–3575. doi: 10.1073/pnas.1420294112

Yan, Y., and Finnigan, G. C. (2018). Development of a multi-locus CRISPR gene drive system in budding yeast. *Sci. Rep.* 8:17277. doi: 10.1038/s41598-018-34909-3

Yang, Z., and Blenner, M. (2020). Genome editing systems across yeast species. *Curr. Opin. Biotechnol.* 66, 255–266. doi: 10.1016/j.copbio.2020.08.011

Yee, D. A., DeNicola, A. B., Billingsley, J. M., Creso, J. G., Subrahmanyam, V., and Tang, Y. (2019). Engineered mitochondrial production of monoterpenes in *Saccharomyces cerevisiae*. *Metab. Eng.* 55, 76–84. doi: 10.1016/j.ymben.2019.06.004

Yu, S.- C., Kuemmel, F., Skoufou-Papoutsaki, M.- N., and Spanu, P. D. (2019). Yeast transformation efficiency is enhanced by TORC1- and eisosome-dependent signaling. *Microbiol. Open* 8:e00730. doi: 10.1002/mbo3.730

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La, R. M., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350. doi: 10.1016/j.cell.2014.11.052

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., et al. (2017). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35, 31–34. doi: 10.1038/nbt.3737

Zhang, Y., Su, M., Qin, N., Nielsen, J., and Liu, Z. (2020). Expressing a cytosolic pyruvate dehydrogenase complex to increase free fatty acid production in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 19:226. doi: 10.1186/s12934-020-01493-z

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J., et al. (2019). A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nat. Commun.* 10:1053. doi: 10.1038/s41467-019-09005-3

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

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

frontiers in Plant Science

REVIEW
published: 13 August 2021
doi: 10.3389/fpls.2021.697318

Plant Metabolic Gene Clusters: Evolution, Organization, and Their Applications in Synthetic Biology

Revuru Bharadwaj[1], *Sarma R. Kumar*[1], *Ashutosh Sharma*[2]* *and Ramalingam Sathishkumar*[1]*

[1] Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore, India, [2] Tecnologico de Monterrey, Centre of Bioengineering, Querétaro, Mexico

Plants are a remarkable source of high-value specialized metabolites having significant physiological and ecological functions. Genes responsible for synthesizing specialized metabolites are often clustered together for a coordinated expression, which is commonly observed in bacteria and filamentous fungi. Similar to prokaryotic gene clustering, plants do have gene clusters encoding enzymes involved in the biosynthesis of specialized metabolites. More than 20 gene clusters involved in the biosynthesis of diverse metabolites have been identified across the plant kingdom. Recent studies demonstrate that gene clusters are evolved through gene duplications and neofunctionalization of primary metabolic pathway genes. Often, these clusters are tightly regulated at nucleosome level. The prevalence of gene clusters related to specialized metabolites offers an attractive possibility of an untapped source of highly useful biomolecules. Accordingly, the identification and functional characterization of novel biosynthetic pathways in plants need to be worked out. In this review, we summarize insights into the evolution of gene clusters and discuss the organization and importance of specific gene clusters in the biosynthesis of specialized metabolites. Regulatory mechanisms which operate in some of the important gene clusters have also been briefly described. Finally, we highlight the importance of gene clusters to develop future metabolic engineering or synthetic biology strategies for the heterologous production of novel metabolites.

Keywords: plant gene clusters, specialized metabolites, defensive functions, gene duplications, metabolic engineering, synthetic biology

OPEN ACCESS

Edited by:
Jakob Franke,
Leibniz University Hannover, Germany

Reviewed by:
Raimund Nagel,
University of Leipzig, Germany
Mariam Gaid,
Independent Researcher,
Braunschweig, Germany

***Correspondence:**
Ramalingam Sathishkumar
rsathish@buc.edu.in
Ashutosh Sharma
asharma@tec.mx

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

Received: *19 April 2021*
Accepted: *05 July 2021*
Published: *13 August 2021*

Citation:
Bharadwaj R, Kumar SR, Sharma A and Sathishkumar R (2021) Plant Metabolic Gene Clusters: Evolution, Organization, and Their Applications in Synthetic Biology. Front. Plant Sci. 12:697318. doi: 10.3389/fpls.2021.697318

INTRODUCTION

Plants produce an array of specialized metabolites to evade biotic and abiotic stressors. Therefore, the production of specialized metabolites is influenced by various environmental cues. These metabolites have been extensively employed in preparing herbal formulations for human health care. For instance, specialized metabolites, such as vincristine, vinblastine, paclitaxel, and curcumin, are recognized as effective inhibitors of cell proliferation and being used in cancer therapeutics (Seca and Pinto, 2018). The significance of plant secondary metabolites in human medicine led researchers to explore the plant kingdom for understanding the biosynthetic machinery of novel metabolites. Plant specialized metabolites are classified according to their chemical backbone and functional groups. Biosynthetic pathways of several specialized metabolites have been elucidated by characterizing the pathway genes, regulators, and gene products involved

in their biosynthesis. Transcriptomic, functional genomics studies combined with metabolomic approaches revealed insights into the operational features of novel metabolite pathways in different medicinal plants (Verma et al., 2014; Meena et al., 2017; Anand et al., 2019; Nagegowda and Gupta, 2020). It is quite challenging to understand the evolutionary aspects of the plant metabolic diversity at the molecular level as several metabolite-encoding gene cascades might be present in the plant genome as well, which are yet to be deciphered (Nützmann et al., 2016).

The clustering of nonhomologous genes of catabolic enzymes and the genes involved in the biosynthesis of specialized metabolites is common in prokaryotes with Lac-operon being the best example (Jacob et al., 1960). Further, in *Streptomyces* sp., genes encoding enzymes involved in the biosynthesis of antibiotics such as granaticin, actinorhodin, are reported to be clustered (Caballero et al., 1991; Ichinose et al., 1998). A few classes of filamentous fungi are known to possess clusters of both primary and secondary metabolic pathway genes and are coordinately expressed (Nützmann et al., 2018). However, the functional genes in animals and plants are scattered throughout the genome, except in a few cases of gene complexes, such as homeobox (*Hox*) and major histocompatibility complexes (*MHCs*), which exist as clusters in animals and are expressed in a synchronized manner (Horton et al., 2004; Holland, 2013; Nützmann et al., 2018). In plants, until the discovery of a gene cluster in *Zea mays* (maize) involved in the biosynthesis of hydroxamic acid 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), it was assumed that secondary metabolite-producing genes occur randomly in the plant genome (Frey et al., 1997). The genes encoding the enzymes for DIMBOA biosynthesis have been reported to be clustered together on the chromosome 4 in maize, and this cluster is found to be widely distributed among the monocots (Frey et al., 2009). In general, a biosynthetic gene cluster is defined as the occurrence of two or more non-homologous genes, in the vicinity on a particular chromosome, which are involved in a common biosynthetic pathway to produce a specialized metabolite or group of similar metabolites (Medema et al., 2015). In particular, plant gene cluster size ranges from 35 kb to several hundred kb, and gene clusters comprises primarily the genes responsible for determining class of metabolites and secondarily one or more genes whose role is to modulate metabolite scaffold to create metabolic diversity (Schneider et al., 2017; Nützmann et al., 2020).

Fungal and plant gene clusters generally share several similarities in the cluster architecture and evolutionary aspects except in the case of the concept of horizontal gene transfer (HGT) of clusters in fungi, which is absent in plants (Slot and Rokas, 2010; Nützmann et al., 2018). Filamentous fungi possess both primary and secondary metabolite producing gene clusters, however specialized metabolite producing gene clusters have been predominantly characterized in plants (Nützmann et al., 2018; Rokas et al., 2018). While fungal gene clusters are equipped with pathway specific transcriptional regulators and transporters, this salient feature is not common in plant gene clusters (Rokas et al., 2020). However, a few plant gene clusters have been reported to possess transporters and coordinately activated regulatory genes (Darbani et al., 2016; Hen-Avivi et al., 2016; Shen et al., 2021). In addition, multigene clusters for a single biosynthetic pathway, the co-occurrence of two clusters of genes on a single chromosome, and intertwined clusters for producing different metabolites are common in filamentous fungi (Yu et al., 2000; Bradshaw et al., 2013; Wiemann et al., 2013). These distinct features of gene cluster organization is not common in plants, except for a few cases such as the co-occurrence of steroidal glycoalkaloid- (SGA) and acyl sugar producing gene clusters on a single chromosome in the members of the *Solanaceae* and the multi-functionality of CYP76M8 in the phytoalexin production in rice (Fan et al., 2020; Kitaoka et al., 2021).

In the past two decades, more than 20 specialized metabolite-producing gene clusters involved in the biosynthesis of different classes of compounds have been identified in various plants (Nützmann et al., 2016). Recent discoveries of high-value metabolite-producing noscapine and thebaine gene clusters in *Papaver somniferum* (poppy) contributed significantly to the understanding of cluster organization, and these findings could aid in developing metabolic engineering/synthetic biology strategies for overproducing these compounds in heterologous production platforms (Guo et al., 2018). Plant metabolic gene clusters typically contain primarily the genes of committed or rate limiting enzymes of various pathways and secondarily the other enzymes required for the modification of backbone to form end-products. It has been shown that, in most cases genes are recruited from the primary metabolism through gene duplications followed by neofunctionalization (Qi et al., 2006). Most of the gene clusters are inferred to have evolved as adaptive strategies to defend against pathogen attack by producing defense metabolites, and a few other clusters are found to have a role in plant development (Qi et al., 2006; Field and Osbourn, 2008; Krokida et al., 2013; **Table 1**). It could be possible that plant systems favor the clustering of genes for controlled and regulated accumulation of metabolites, thus avoiding the formation of toxic intermediates and co-inheritance to progeny (Nützmann et al., 2016). Moreover, the metabolites produced by the gene clusters have a significant agronomic and human therapeutic importance, and in-depth studies are needed to develop the strategies for improving the bioproduction of target metabolites in heterologous hosts. Furthermore, plant genome mining through the genomic resources followed by functional genomic approaches together with metabolomics could reveal the existence of novel gene clusters involved in the production of high-value metabolites.

Herein, we discuss the evolutionary aspects of plant gene clusters along with their molecular features related to the organization of the clusters and regulatory mechanisms governing the organization. In addition, we elaborate the physiological role of cluster-derived metabolites in plant defense and other metabolic functions. Finally, we propose guiding principles toward the development of novel strategies related to metabolic engineering and synthetic biology by utilizing the repository of studies on plant gene clusters.

TABLE 1 | Details of the gene clusters present in the different plant species and diversity of metabolites produced along with their physiological function.

S.no	Species	Metabolite	Compound class	Chromosome (core genes)	Function	Tissue of expression	References
1	*Arabidopsis thaliana*	Thalianol	Triterpenes	5	Unknown physiological function, unregulated expression of cluster genes leads to dwarfing of plant. Modulate root microbiome content	Roots	Field and Osbourn, 2008; Field et al., 2011; Chen et al., 2019a
2	*A. thaliana*	Marneral	Triterpenes	5	Unknown physiological function, unregulated expression of cluster genes leads to dwarfing of plant	Roots	Field and Osbourn, 2008; Field et al., 2011
3	*Avena strigosa*	Avenacins	Triterpenes	1	Defense against pathogens	Roots	Li et al., 2021a
4	*Cucumis sativus*	Cucurbitacins	Triterpenes	6	Insect deterrent properties and possess medicinal value	Leaves and fruits	Shang et al., 2014
5	*Ricinus communis and Jatropha curcas*	Casbenes	Diterpenes	1	Possess medicinal value and used in treating cancers and HIV infection	Constitutive expression in leaves, roots and stems	King et al., 2014, 2016
6	*Zea mays*	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA)	Hydroxamic acids	4	Defense related activities	Mainly expressed during seedling stages and in roots.	Frey et al., 1997
7	*Oryza sativa*	Momilactones	Diterpenes	4	Insect deterring properties and anti-fungal properties	Induced expression during pathogen attack	Shimura et al., 2007; Wang et al., 2011
8	*O. sativa*	Phytocassanes Oryzalides	Diterpenes	2	Defense related activities	Induced expression during pathogen attack	Swaminathan et al., 2009; Wu et al., 2011
9	*Lotus japonicus*	Linamarin Lotuastralin	Cyanogenic glucosides	3	Herbivore deterrent activities	Above ground plant parts	Takos et al., 2011
10	*Sorghum bicolor*	Dhurrin	Cyanogenic glucosides	1	Herbivore deterrent activities	Above ground plant parts	Takos et al., 2011
11	*L. japonicus*	20-hydroxy-betulinic acid	Triterpene	3	Possible role in plant development and nodule formation	Elevated expression in roots and nodules	Krokida et al., 2013
12	*Solanum lycopersicum*	α-tomatine	Steroidal glycoalkaloid	7	Anti-pathogenic activity and toxic to humans	Elevated expression during pathogen attack	Itkin et al., 2013
13	*Solanum tuberosum*	α-solanine	Steroidal alkaloid glycoalkaloid	7	Anti-pathogenic activity and toxic to humans	Elevated during stress conditions	Itkin et al., 2013
14	*S. lycopersicum*	Mono & di terpenes	β-phellandrene Lycosantalonol	8	Involved in attracting pollinators and possible role in herbivore and fungal defenses	Localized expression in young leaves and trichomes	Matsuba et al., 2013, 2015; Zhou and Pichersky, 2020
15	*Papaver somniferum*	Noscapine	Benzylisoquinoline alkaloid	11	High value secondary metabolites	Stem and capsule	Guo et al., 2018
16	*P. somniferum*	Thebaine	Benzylisoquinoline alkaloid	11	High value secondary metabolites	Stem and capsule	Guo et al., 2018
17	*S. lycopersicum*	Falcarindiol	Fatty acid cluster	12	Elicited by pathogen attack	Above ground parts	Jeon et al., 2020
18	*Capsicum annuum*	Capsidiol	Sesquiterpene	2, 12	Pathogen induced production	Above ground parts	Lee et al., 2017

EVOLUTIONARY DYNAMICS OF PLANT GENE CLUSTERS

Genes that exist in close proximity on chromosomes, are often co-expressed (Elizondo et al., 2009). In prokaryotes, a set of non-homologous genes form clusters, which are generally referred as operons, and the genes in the operons are transcribed together to form a polycistronic messenger RNA (mRNA) to encode the proteins involved in a specific metabolic function (Jacob et al., 1960). The "Selfish operon model" describes that cluster arrangement of genes can improve HGT to other species, thereby increasing the chances of cluster survival (Ballouz et al., 2010). In fungi, the genes encoding enzymes involved in formation of β-lactam antibiotics and the biosynthesis of nitrate, proline, and galactose (GAL) occur as clusters with a similar pattern of co-expression (Nützmann et al., 2018). β-lactam antibiotic clusters might be of bacterial origin and could have been transferred to fungi through HGT (Liras and Martín, 2006; Slot, 2017). Similarly, GAL clusters originated differently in three species of yeast, whereas in *Saccharomyces cerevesiae* and *Candida sp.* GAL clusters are known to be originated independently through gene relocation. Comparably, in the case of *Schizosaccharomyces,* GAL cluster is found to have been acquired from *Candida* species through HGT (Slot and Rokas, 2010).

It is inferred that plant gene clusters are not evolved through HGT. Further, it has been shown that the cluster development in plants could have occurred through gene duplications, relocalization, neofunctionalization, or an independent evolution of genes toward the acquisition of specialized metabolism (Nützmann et al., 2016). To understand the evolution of plant gene clusters, Liu et al. (2020a) explored the genomes of the members of *Brassicaceae* for genome neighborhood (GNs) regions spanning around oxidosqualene cyclases (*OSCs* are key enzymes involved in sterol biosynthesis in plants) and identified that clade II *OSC*s were surrounded by cytochrome P450s (CYP450s) and acyltransferase genes possibly indicating the cluster organization. These GNs were found to be in highly dynamic chromosomal regions and lacked synteny toward the regions originated from whole genome duplication (WGD) in *Brassicaceae*, depicting an independent mode of evolution. Interestingly, functional characterization of GNs together with *OSC* in different species revealed that these putative clusters are equipped with a similar set of genes, even though they exhibited a diverse spatial and temporal expression (Liu et al., 2020a). Thalianol-producing gene cluster is known to be identical in *Arabidopsis thaliana* and *Arabidopsis lyrata*. Genome analysis of various *Arabidopsis* species revealed the evolution of thalianol cluster occurred before the divergence of *A. thaliana* and *A. lyrata*. In both the species, the cluster is organized with four core genes, whereas in *A. thaliana* three additional genes (*THAA2, THAR1 and THAR2*) are required for producing thalianin, which might have occurred through a chromosomal inversion event, but in *A. lyrata* five genes [four core genes; *THAS*, thalianol hydroxylase *(THAH)*, *THAO*, *THAA1* and one linked gene; *THAA2*] are responsible for epithalianin production (Liu et al., 2020b). In addition, similar clusters from *Capsella rubella* and *Brassica rapa* are known to produce tirucallol derivatives (produced in buds) and euphol, respectively (Liu et al., 2020a). Boutanaev and Osbourn (2018) reported that transposable elements, such as miniature inverted-repeat transposable elements (MITEs), are present within the gene clusters of eudicots and they are known to play a predominant role in cluster formation by chromosomal rearrangements such as deletions, translocations, and inversions. Accordingly, triterpene clusters of *A. thaliana* are rich in transposable elements, which might have contributed to cluster formation through the segmental duplication of a committed step followed by an independent recruitment of tailoring enzymes (Field et al., 2011). In addition, transposable element-mediated genetic recombination and duplications led to the formation of sesterterpene-producing gene clusters in *A. thaliana* (Chen et al., 2019a). In addition, *Brassicaceae* members also possess pairs of terpene synthase and prenyl transferase on the genome that produces different sesterterpenes (Huang et al., 2017).

Local gene duplication could also contribute to the formation of gene clusters which are explicitly lineage specific (Schläpfer et al., 2017). For example, SGA gene clusters emerged in *Solanum lycopersicum* (tomato) and *Solanum tuberosum* (potato) through a duplication event from the common ancestor. Similar orthologous clusters have been identified in eggplant and pepper. In pepper, few genes for SGA biosynthesis underwent deletions during the course of evolution, and as a result pepper plant produces steroidal saponins instead of SGAs (Itkin et al., 2013; Barchi et al., 2019; **Table 1**). In addition, terpene synthase gene cluster in *Solanum* species might have evolved through several segment duplications of terpene synthase and *cis*-prenyltransferase genes, though few genes in the cluster are found to be non-functional (Matsuba et al., 2013). In *Solanaceae*, three genes *AsAT1*, *AACS1*, and *AECH1* are characterized to be (occuring in multi-chromosome synteny regions) responsible for the production of medium-chain acyl sugars. These genes might have evolved through the insertion and segmental duplication events that occurred in a common ancestor before divergence (Fan et al., 2020).

In monocots, avenacin cluster and metabolite biosynthesis is highly specific to *Avena strigosa* (diploid oat), and this cluster has been reported to have evolved independently (Qi et al., 2004). Genome analysis of *A. strigosa* revealed the presence of avenacin cluster in sub-telomeric regions, with the presence of the early pathway genes near to the telomere, and the existence of terminal pathway genes away from the telomere. These typical positioning co-linearity of genes in the cluster is considered to avoid the deletions of terminal pathway genes, thus preventing the accumulation of toxic intermediates (Li et al., 2021a). On the other hand, maize DIMBOA gene cluster is also located at the tip (telomeric region) of the chromosome for facilitating an adaptive evolution and a coordinated regulation (Dutartre et al., 2012). Hydroxamic acids of maize (DIMBOA) are also produced in rye and wheat, and DIMBOA-producing genes are present on two different chromosomes in respective species without disrupting metabolite biosynthesis (Frey et al., 1997, 2009). Barnyard grass, a noxious weed of rice fields, has acquired orthologous gene clusters involved in the biosynthesis of momilactone and DIMBOA, respectively. In addition, barnyard grass is known

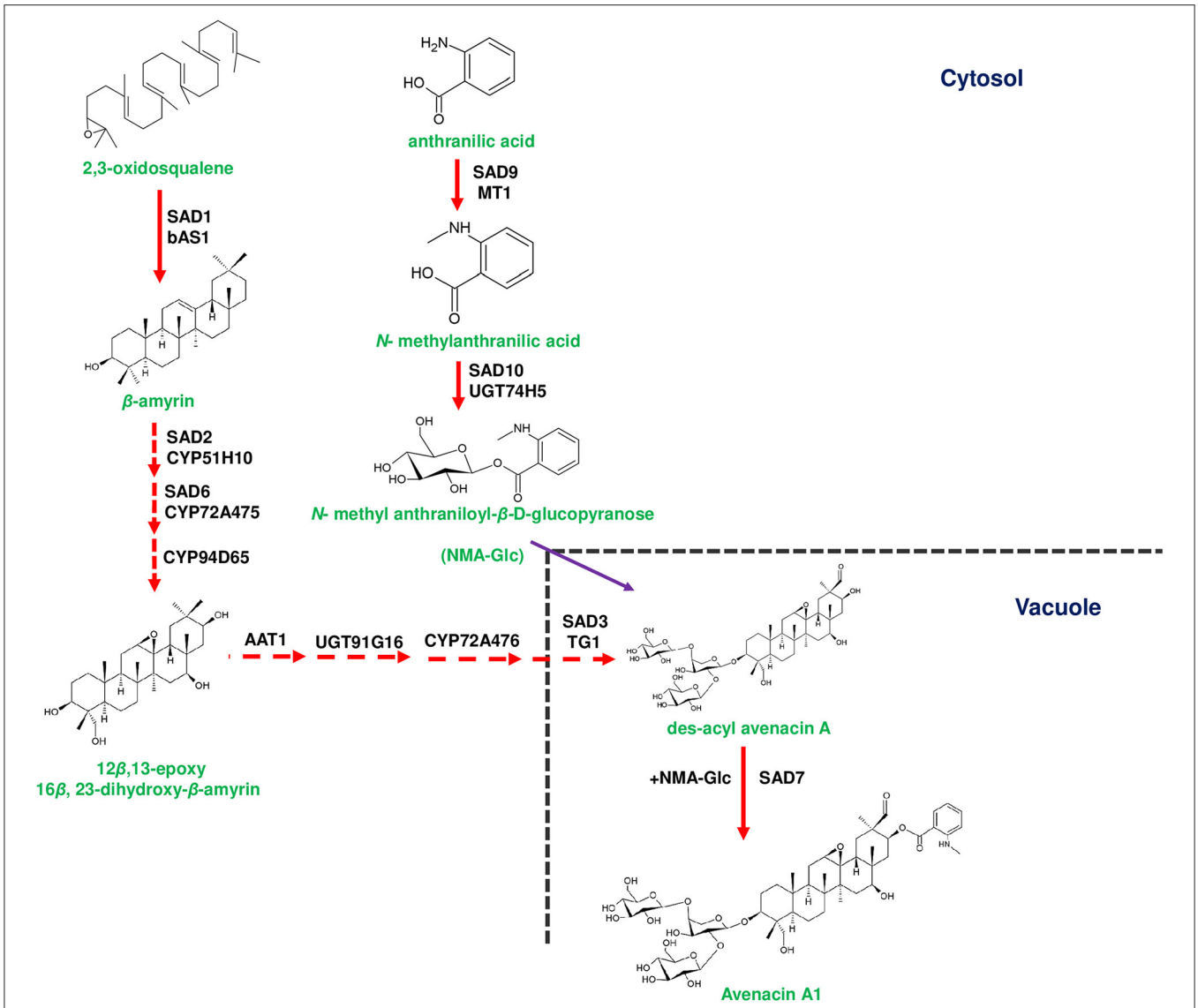


FIGURE 1 | Biosynthetic pathway of avenacin in the roots of *Avena strigosa*. Avenacin biosynthesis starts with the conversion of 2,3-oxidosqaulene to β-amyrin by Sad1 (Saponin deficient 1). Conversion of β-amyrin to 12β,13-epoxy 16β, 23-dihydroxy-β-amyrin that occurs in the cytosol is mediated by Sad2, Sad6 and CYP94D65, respectively. 12β,13-epoxy 16β, 23-dihydroxy-β-amyrin is subsequently conjugated with one arabinose and two glucose moieties for the formation of des-acyl avenacin. Two glycosylation steps occur in the cytosol catalyzed by *A. strigosa* arabinosyltransferase (*As*AAT1) and *As*UGT91G16 (UDP-glucosyl transferase). The final glucose moiety is added by a unique vacuolar glycosyltransferase *A. strigosa* transglucosidase 1 (*As*TG1). Des-acyl avenacin is acylated by Sad7 aided by acyl donors such as *N*-methyl anthranilate glucopyranose (NMA-Glc) to form avenacin A1, in the vacuole. NMA-Glc is formed from anthranilic acid through a two-step reaction catalyzed by Sad9 and Sad10, respectively. Genes of the plant metabolic gene cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represents multi-step pathway.

to possess a quercetin producing gene cluster, which is highly upregulated during the interaction of the weed with the host (Sultana et al., 2019). Gene clusters of rice (phytocassanes and momilactones) might have evolved as an adaptive strategy to counter pathogenic invasions (Swaminathan et al., 2009). Miyamoto et al. (2016) suggested that the evolution of these two gene clusters could have occurred in the common ancestor of *Oryza* species before domestication. In addition, momilactone cluster might have evolved through an assembly of individual genes into a physical proximity through duplications, whereas the phytocassane cluster might have already existed in the common ancestor of *Oryza* species but got lost in different lineages during the course of evolution (Miyamoto et al., 2016). Fascinating occurrence of the momilactone biosynthetic cluster in *Calohypnum plumiforme,* a bryophyte species has opened up the speculations of convergent evolution of gene clusters for species survival in challenging environments (Mao et al., 2020; Zhang and Peters, 2020). Casbene diterpenoid producing gene

cluster in rice is specific to *Oryza* genus and have evolved through gene duplications (Zhan et al., 2020). Furthermore, an intact casbene production is observed in japonica cultivars (Medema et al., 2015) compared to indica cultivars of rice. This has been attributed to the natural selection of cluster in japonica varieties during the process of domestication to impart innate resistance against blight disease (Zhan et al., 2020). In addition, recent documentation of hydroxycinnamoyl tyramine producing gene cluster being specific to rice lineage and a high level of induced expression of the clustered genes during pathogen attack sheds light on the adaptive evolution of plant gene clusters in response to environmental cues (Shen et al., 2021).

In the case of cyanogenic glucoside clusters, parallel and independent evolution might have occurred in higher plants possessing the same scaffold of genes. *Lotus japonicus* and *Sorghum bicolor* are known to possess similar classes of genes in the cluster responsible for the production of plant-specific glucosides (Takos et al., 2011; **Table 1**). Intriguingly, an orthologous gene cluster of cyanogenic glucosides was also observed in white clover, a distant relative of *L. japonicus* (both belong to the subfamily *Papilionoideae*). They are considered to share a common ancestry, even though other members of *Papilionoideae* lost this cluster during the course of evolution as an adaptive strategy toward specific environmental niche (Olsen and Small, 2018). An *in silico* analysis using Plantismash revealed the presence of 10 orthologous gene clusters in the genomes of *Amaranthus cruentus* and *Amaranthus hypochondriacus* related to secondary metabolism. However, their *in planta* functional role is not conclusive (Ma et al., 2021). Comparative genome analysis of four genera of *Amaranthaceae* (*Amaranthus*, *Beta*, *Chenopidum,* and *Spinacia*) provided conclusive evidence about the co-occurrence of betalain pigment-producing genes on a specific chromosome, but in spinach (*Spinacia oleracea*) and quiona (*Chenopodium quinoa*) additional copies of the genes were found to exist. This could be due to tandem gene duplication mechanisms (Ma et al., 2021).

Available reports on gene cluster evolution indicate that plants developed gene clusters to reprogram their functional attributes toward an adaptation to different ecological niches by recruiting genes through the duplication of primary metabolism genes and acquiring new functions to them. For instance, *Sad1* (saponin deficient), *Sad2* genes in oat avenacin cluster are recruited from sterol metabolism, and novel functions were acquired subsequently (Qi et al., 2004, 2006; **Figure 1**; **Table 2**). Sonawane et al. (2016) also reported the duplication events followed by the neofunctionalization of primary metabolite biosynthesis led to the evolution of cholesterol biosynthetic genes in *Solanum* species. Moreover, the formation of these gene clusters is facilitated by various chromosomal recombination events, the presence of transposable elements, and sub-telomeric positions of clusters, and these events further support the notion that gene clusters are dynamic and evolving rapidly. Further, these gene clusters are co-inherited to progeny as an environmental adaptation for functions such as development and defense responses. In addition, negative selection pressure against the accumulation of toxic intermediates could lead to the formation of gene clusters in plants (Li et al., 2021a). Based on the observations related to the dynamic evolution of gene clusters, it is conclusive that plant genome is highly flexible and capable of associating non-homologous genes into a single coordinated cluster. Nevertheless, the lack of genome information of several plant species is a major constraint to reach a plausible conclusion on the evolution of gene clusters. It will be interesting to study the precise molecular mechanisms of evolutionary pressure that prompted plant genome plasticity. Genome mining of several plants through bioinformatics tools followed by analyzing the emergence of gene clusters in closely related and as well as distantly related species through phylogenomics, and comparative genomics approaches could help in understanding the evolutionary dynamics. Finally functional characterization of these clusters help to identify their *in planta* role as part of a major adaptive evolution.

ORGANIZATION OF PLANT GENE CLUSTERS

General Rules of Plant Gene Clusters

Plant gene clusters have been characterized in both monocots (phytoalexin clusters in rice, DIMBOA cluster in maize, and avenacin cluster in oat) and dicots (thalianol, marneral clusters of *Arabidopsis*, noscapine, and thebaine clusters of poppy; **Table 1**). The common principle of gene clusters is the occurrence of set of non-homologous genes producing a specific metabolite in physical proximity on a chromosome (Medema et al., 2015; **Table 1**). Recent chromosomal analysis by Nützmann et al. (2020) in *Arabidopsis* revealed that the active cluster regions occur in special local hotspot regions away from the heterochromatin region and nuclear periphery. Furthermore, a similar kind of organization has been reported in rice, tomato, and maize (Nützmann et al., 2020). Genes of the first committed pathway step followed by downstream tailoring enzymes are the typical components of plant gene clusters. The number of downstream tailoring enzymes in a cluster ranges from 3 to 12 depending on the complexity of different metabolic pathways (**Table 2**). A signature enzyme catalyzes the first step of the pathway, which outlines the class of metabolite to be produced, and this signature enzyme is assigned to draw the primary metabolite flux toward a more specialized metabolism (Nützmann et al., 2016; **Figures 1–6**). However, cyanogenic glucoside, SGA pathways of *L. japonicus,* and several other members of the *Solanaceae* deviate from the above rule by catalyzing the first step of the pathway through CYP450s (Takos et al., 2011; Itkin et al., 2013; **Figure 3**). The arrangement of genes in plant gene clusters differs significantly, for instance, noscapine and falcarindiol clusters possess non-functional genes whose role in the metabolism is yet to be deciphered while thebaine cluster is tightly packed and contains the genes responsible for the biosynthesis of thebaine from *R*-reticuline (Guo et al., 2018; Jeon et al., 2020; Li et al., 2020; **Table 1**; **Figure 5**). In addition, genome analysis of poppy revealed that noscapine and thebaine clusters occur separately on the same chromosome while *R*-reticuline-synthesizing genes were found to be loosely clustered (Guo et al., 2018; Li et al., 2020). A similar pattern of compact clustering is

TABLE 2 | Different classes of tailoring enzymes occurring in the plant gene clusters, including CYP450s, acyl transferases, glycosyl transferases, and alcohol dehydrogenases.

S. no	Cluster	CYP450s	Acyl transferases	Glycosyl transferases	Alcohol dehydrogenases	Other classes of enzymes	References
1	Thalianol	THAH-CYP708A2 THAD-CYP705A5, THAR1, THAR2	BAHD acyltransferases THAA1, THAA2, THAA3				Field and Osbourn, 2008; Liu et al., 2020a,b; Bai et al., 2021
2	Marneral	MRO-CYP71A16 CYP705A12					Field and Osbourn, 2008; Field et al., 2011
3	Avenacin	Sad2-CYP51H10 Sad6-CYP72A475, CYP72A476 CYP94D65	Sad7 acyl transferase (Serine carboxypeptidase like protein)	Sad3–glucosyl hydrolase glycosyl transferases-Sad10, *As*AAT1 and UGT91G16		Sad9 Methyl transferase	Qi et al., 2006 and references related to avenacin biosynthesis mentioned in main text
4	Cucurbitacins	CYP81Q58, CYP89A140 CYP87D19 (within cluster) CYP87D20 (Chromosome1) CYP712D8 CYP88L2 CYP88L3 (Chromosome 3)	Acyl transferase (Csa6G088700)				Shang et al., 2014; Zhou et al., 2016
5	Casbenes	CYP76A14 CYP76A15 CYP76A16 CYP76A19 (*R. communis*) CYP71D495 CYP726A35 (*Jatropha carcus*)					King et al., 2014, 2016
6	Momilactones	CYP99A2 CYP99A3 CYP701A8 CYP76M14			*Os*MAS Short chain alcohol dehydrogenase		Shimura et al., 2007; Wang et al., 2011; De La Peña and Sattely, 2021
7	*Oryzalides* *Ph*ytocassanes	CYP71Z6 CYP71Z7 CYP76M7 CYP76M8					Swaminathan et al., 2009; Wu et al., 2011
8	Cyanogenic glucosides (*L.japonicus*)	CYP736A2 CYP79D3/D4		UGT85K2 (UDP-glucosyltransferase)			Takos et al., 2011
9	Cyanogenic glucosides (*S.bicolor*)	CYP71E1 CYP79A1		UGT85B1 (UDP-glucosyltransferase)			Takos et al., 2011
10	20-hydroxy-betulinic acid	CYP71D353					Krokida et al., 2013
11	Steroidal glucoside alkaloid cluster	GAME6 GAME7 GAME4 GAME8 (CYP72)		GAME1, GAME17, GAME18, GAME2 (UDP-glycosyl transferases)	GAME25 (Short chain dehydrogenase)	GAME11 (Putative dioxygenase) GAME12 (Transaminase)	Itkin et al., 2013; Sonawane et al., 2018

(Continued)

TABLE 2 | Continued

S. no	Cluster	CYP450s	Acyl transferases	Glycosyl transferases	Alcohol dehydrogenases	Other classes of enzymes	Reference
12	Falcarindiol cluster					Acetylenase Fatty acid desaturase Decarbonylase	Jeon et al., 2020
13	Capsidiol					5-epi-aristolochene synthase (EAS) 5-epi-Aristolochene 3-hydroxylase, (EAH)	Lee et al., 2017

also observed in oat avenacin cluster (Qi et al., 2004; Li et al., 2021a). Recently, elucidation of complete cluster organization of avenacins revealed that all genes are organized in a co-linear manner in relation to the biosynthetic steps of the pathway, and ten genes are arranged on the end of long arm of the chromosome 1 while the other two glycosyl transferases (GTs) are present in proximal scaffold (Li et al., 2021a).

Few metabolite pathway genes are known to exist adjacently in a cluster, and the three genes, *viz*., *Sad7*, *Sad9,* and *Sad10* are found to form an acylation module, which is responsible for acylating avenacins (Mugford et al., 2013; Nützmann et al., 2016). In contrast to the gene arrangement of avenacin cluster, *Glycoalkaloid Metabolism 7* (*GAME7*) and *GAME8* genes of SGA biosynthesis are found to be located away from the cluster in both tomato and potato (Itkin et al., 2013). *Berberine bridge enzyme* (*BBE*) and *tetrahydroprotoberberine N-methyltransferase* (*TNMT*) genes of noscapine biosynthesis are known to lack a clustered arrangement due to their additional functions in sanguinarine biosynthesis (Guo et al., 2018). In addition, two CYP450s (*THAR1* and *THAR2*) involved in the production of thalianin (the final product derived from thalianol) are located away from the thalianol cluster (Liu et al., 2020a,b). In few cases, an identical set of homologous genes are known to produce slightly different compounds. For instance, gene clusters of tomato and potato produce two different SGAs, viz., α-tomatine and α-solanine, respectively (Itkin et al., 2013; **Figure 3B**). Conversely, the orthologous genes of *L. japonicus* and *S. bicolor* encode a similar class of enzymes *albeit* producing different cyanogenic glucosides (Takos et al., 2011; **Figure 2B**). In addition, a few plants are known to possess multifunctional gene clusters equipped with the genes responsible for producing different compounds, for example, rice phytoalexin gene cluster contain the genes involved in the bioproduction of both phytocassanes and oryzalides (Swaminathan et al., 2009; **Figure 4**). Similarly, tomato terpene gene cluster is known to contain the genes producing mono- and di-terpenes such as lycosantalonol and β-phellandrene (Matsuba et al., 2015). In yet another situation, a few genes of monoterpene indole alkaloid (MIA) pathway are arranged together in different scaffolds in *Catharanthus roseus*. Interestingly, similar small clusters have also been observed in *Gelsemium sempervirens,* which produces the similar MIAs depicting the conserved nature of genes in different plant families (Kellner et al., 2015; Franke et al., 2019). A gene cluster producing medium-chain acyl sugars in the *Solanacaeae* occurs adjacent to the SGA gene cluster on chromosome 7. However, the evolutionary aspects regarding the co-localization of both clusters are yet to be deciphered (Fan et al., 2020). Despite the occurrence of several complicated organizational patterns of gene clusters in different plant species, the genes responsible for the bioproduction of specialized metabolites are found to be expressed in a coordinated manner independent of their occurrence within or outside the cluster. Dhurrin producing gene cluster of *S. bicolor* is found to be stringently regulated and forms metabolon for channeling intermediates, but the enzymes of avenacin and DIMBOA biosynthesis occur in different cellular compartments highlighting the complexity of cluster expression, channeling of the metabolic intermediates (Frey et al., 2009; Takos et al., 2011; Li et al., 2021a). Along with the genes responsible for metabolite bioproduction, a few clusters are also equipped with transporters and other regulatory elements. For example, the dhurrin cluster contains a co-expressed multi-antimicrobial extrusion protein (MATE) transporter gene (multidrug and toxic compound extrusion), which could bind and transport dhurrin (Darbani et al., 2016). Additionally, a cofactor synthase *OsPDX3* of rice, which is involved in the biosynthesis of hydroxycinnamoyl tyramine, occurs within the cluster (Shen et al., 2021). Hen-Avivi et al. (2016) reported the presence of a regulatory gene within the vicinity of the β-diketone cluster in wheat. Considering all the above mentioned examples, at this point of time, it is indeed difficult to draw firm conclusions related to the general organization and behavior of plant gene clusters. An in-depth analysis could reveal not only their functional role but also indicate novel evolutionary strategies of these gene clusters in conferring adaptability to the ever-changing environmental conditions.

Significance of First Pathway Step in Metabolite Production

In most of the gene clusters identified so far, the first step of the pathway is catalyzed by a specific class of enzymes (e.g., terpene cyclases and terpene synthases) that divert the flux of primary metabolites to synthesize the cluster-specific specialized metabolites. For instance, OSC of thalianol (*THAS; thalianol synthase*) and marneral (*MRN1: marneral synthase*) clusters initiate metabolite bioproduction in *Arabidopsis* by converting 2,3-oxidosqualene, a branch point intermediate to a different set

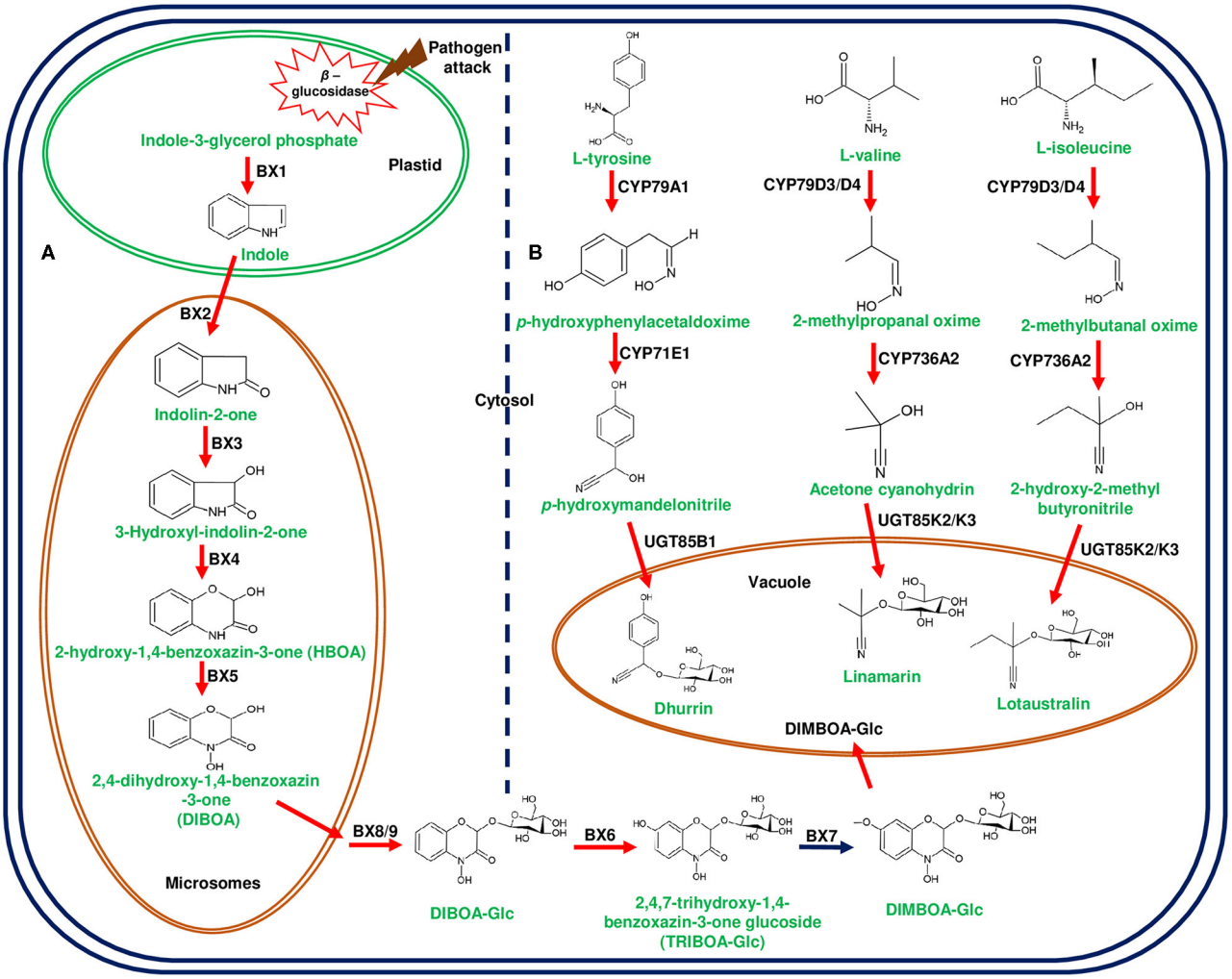


FIGURE 2 | Biosynthetic pathways of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) in *Zea mays* and cyanogenic glucosides in *Lotus japonicus* and *Sorghum bicolor*. **(A)** represents the biosynthesis of DIMBOA from indole-3-glycerol phosphate mediated by DIMBOA gene cluster in maize. The initial step of indole formation from indole-3-glycerol-phosphate occurs in the plastid catalyzed by benzoxazinless1 (Bx1). The conversion of indole to 2,4-dihydroxy-1, 4-benzoxazin-3-one (DIBOA) occurs in the microsomes through the catalytic reactions of cytochrome P450 enzymes (CYP450s);Bx2-Bx5. DIBOA is transported to cytosol and converted to DIMBOA-glucoside (DIMBOA-Glc) by sequential enzymatic reactions of Bx8/9, Bx6, and Bx7. The glycosylated DIMBOA is transported to the vacuole for sequestration to avoid auto-toxicity. **(B)** represents cyanogenic glucosides (linamarin, lotaustralin in *L. japonicus*, and dhurrin in *S. bicolor*) produced from different amino acids. L-tyrosine, L-valine, and L-isoleucine are converted into aglycones through the catalytic action of CYP450s belonging to CYP71 family. The final step of glucose addition is catalyzed by UGTs and sequestered in vacuole. Herbivore attack and cell disruption induces the plastidial β-glucosidase activity and releases the glucose moiety from DIMBOA-Glc and cyanogenic glucosides. Genes of the plant metabolic cluster encoding enzymatic reactions are indicated in red arrows.

of metabolites. *THAS* and *MRN1* belong to the clade II OSC of *Arabidopsis,* and other enzymes belonging to this clade are also found to be flanked and are co-expressed (Field and Osbourn, 2008; Field et al., 2011; Liu et al., 2020a,b). Interestingly, oat-specific *A. strigosa β-amyrin synthase* (*AsbAS1*) gene is similar to cycloartenol synthase (has a role in sterol precursor production) and mediate the conversion of 2,3-oxidosqualene into β-amyrin (Haralampidis et al., 2001; Qi et al., 2004; **Figure 1**). OSC encoded by *bitterness (Bi)* gene of the cucurbitacin cluster catalyzes the first step to produce cucurbitadienol, and *Bi* is known to be highly conserved across cucurbits (Shang et al., 2014; **Figure 3A**). Additionally, *AMY2* (catalyzes the first step of 20-hydroxy-betulinic acid synthesis) in *L. japonicus,* encodes an unusual OSC to produce dihydro-lupeol along with a relatively lower amounts of β-amyrin (Krokida et al., 2013; **Figure 6A**).

Diterpene synthases of gene cluster of *Ricinus communis* are involved in the diversion of geranylgeranyl pyrophosphate (GGPP) pool from the primary metabolism to the synthesis of casbene oxidases (King et al., 2014). Similarly, phytoalexins produced by diterpene synthases of *O. sativa* (rice) share an analogy with the enzymes of gibberellin biosynthesis (Wilderman et al., 2004; Zhang and Peters, 2020). Interestingly,

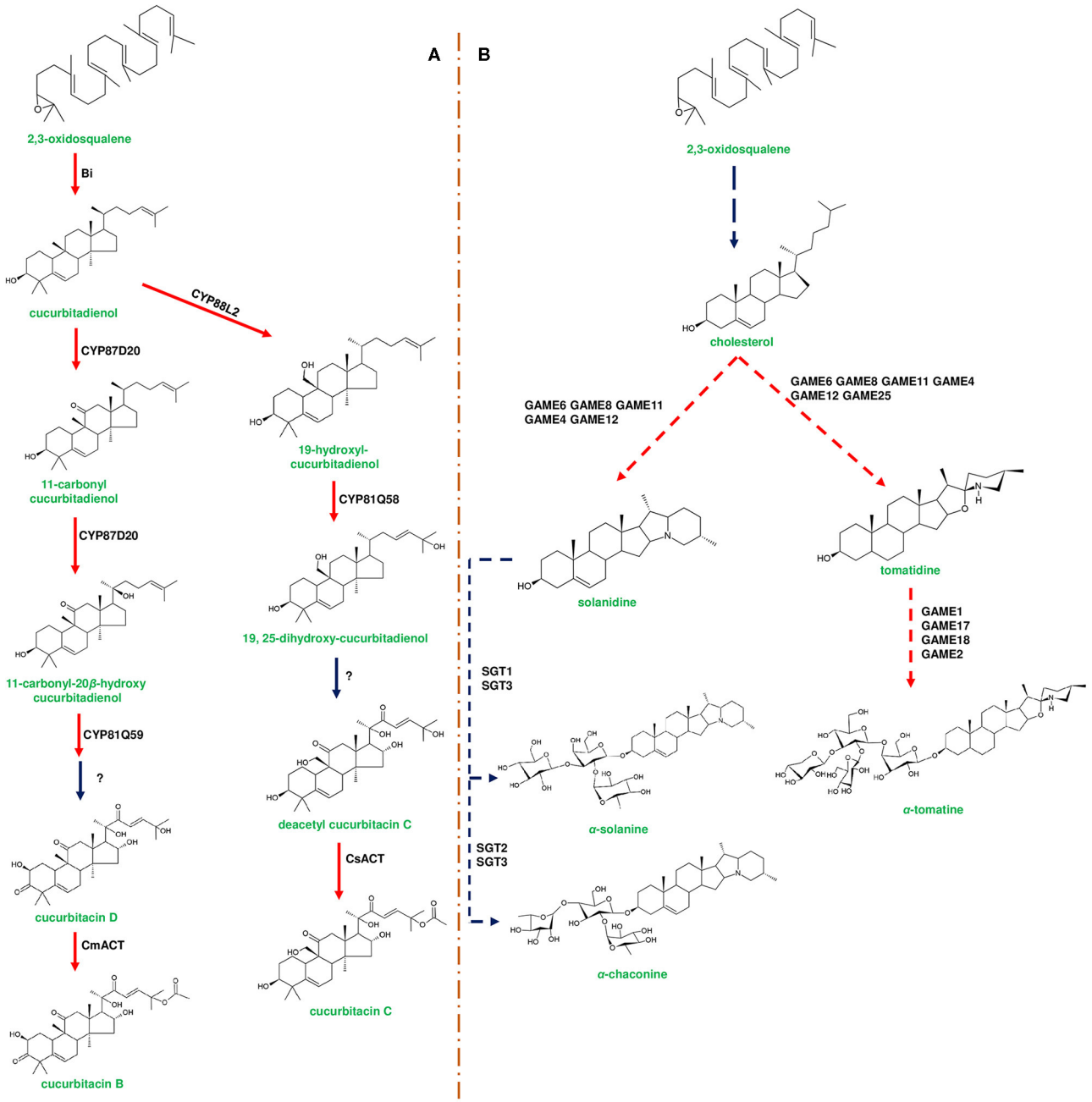


FIGURE 3 | Representative biosynthetic pathway of cucurbitacins in *Cucurbitaceae* and steroidal glycoalkaloids (SGAs) in *Solanaceae*. **(A)** represents the biosynthesis of cucurbitacins through enzymes encoded by the cucurbitacin gene cluster. The first step in cucurbitacin biosynthesis occurs by the enzymatic conversion of 2,3-oxidosqualene to cucurbitadienol mediated by enzyme Bitterness (Bi) belonging to oxidosqualene cyclase family. Cucurbitadienol is converted to cucurbitacin C (CuC) in *Cucumis sativus* mediated by CYP450s and acyltransferase [*C. sativus* acyltransferase (*Cs*ACT)]. In *Cucumis melo*, cucurbitadienol is converted to cucurbitacin B (CuB) by CYP450s and *C. melo* acyl transferase (*Cm*ACT). Modification of the backbone in different *Cucumis* species is attributed to the activity of CYP450s. **(B)** represents the biosynthesis of SGAs from cholesterol using the enzymes encoded by the *SGA* gene cluster. α-solanine and α-chaconine are the signature metabolites of potato (*Solanum tuberosum*). In potato, cholesterol is converted to solanidine by catalytic reactions of GAME8 (glycoalkaloid metabolism), GAME6, GAME11, and GAME12. Further, solanidine is converted to α-solanine and α-chaconine by Sterol alkaloid glycosyl transferases (SGT1, SGT2, SGT3). α-tomatine is produced in *Solanum lycopersicum* (tomato) through a similar set of genes present in the *SGA* gene cluster. Tomatidine is formed by the modification of cholesterol mediated by GAME6, GAME8, GAME11, GAME4, GAME12, and GAME25, respectively. Tomatidine is conjugated with four sugar moieties [one moiety each of galactose and xylose and two moieties of glucose] by UDP-glycosyl transferases, GAME1, GAME17, GAME18, and GAME2, respectively. Genes of the plant metabolic cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represent a multi-step pathway.

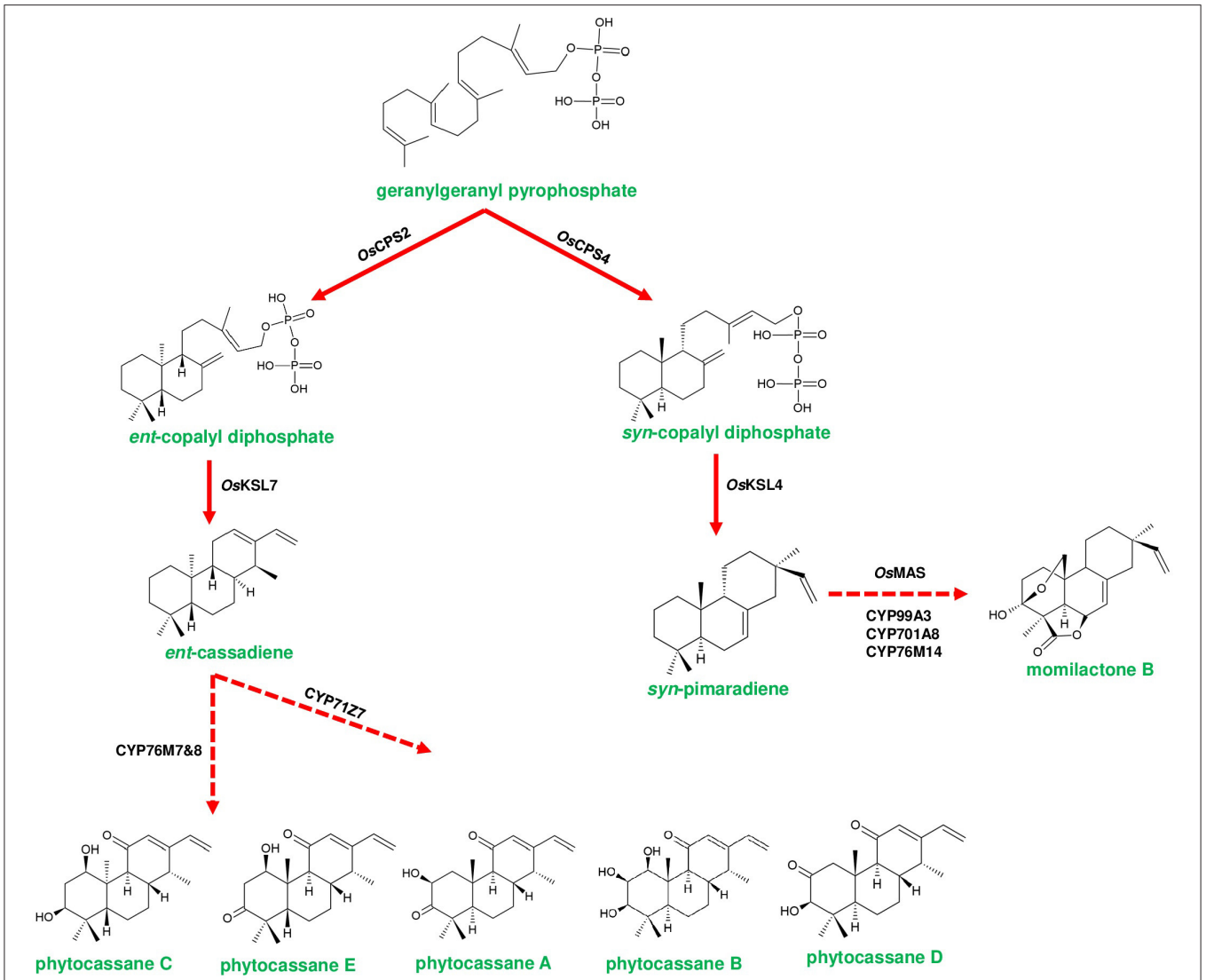


FIGURE 4 | Representative biosynthetic pathway of phytoalexins in *Oryza sativa* (rice). Phytocassanes bioproduction is initiated by the conversion of geranylgeranyl pyrophosphate (GGPP) to *ent*-copalyl diphosphate (*ent*-CDP) by OsCPS2 (copalyl diphosphate synthase 2). *ent*-CDP is converted to *ent*-cassadiene by OsKSL7 (kaurene synthase like 7). *ent*-cassadiene is converted to different derivatives of phytocassanes by CYP450s. Momilactone biosynthesis is initiated by the conversion of GGPP to *syn*-copalyl diphosphate (*syn*-CDP) through OsCPS4. *syn*-CDP is converted to *syn*-pimaradiene by OsKSL4. *syn*-pimaradiene is converted to momilactone through the reactions mediated by CYP99A2/A3, CYP701A8, CYP76M14 and short-chain alcohol dehydrogenase (momilactone synthase; OsMAS). Dotted arrows represent a multi-step pathway.

a tryptophan synthase homolog [*benzoxazinless1 (Bx1)*] of maize diverts the flux of indole-3-glycerol phosphate from tryptophan synthesis to indole for the biosynthesis of benzoxazinoids, and *Bx1* is considered to have evolved independently in monocots through duplication and neofunctionalization (Frey et al., 1997; **Figure 2A**). In addition, the identification of orthologous *Bx1* gene in dicots implicates the convergent evolution of benzoxazinoid production across the plant kingdom (Schullehner et al., 2008; Dick et al., 2012). Tomato falcarindiol cluster (a modified fatty acid) involved in the biosynthesis of specialized metabolite was shown to be initiated by acetynalase enzyme catalyzing the conversion of the primary metabolite linoleic acid to crepenyic acid (Jeon et al., 2020; **Figure 6B**). There are a few other gene clusters in which the first step of the metabolic pathway is not catalyzed by any signature enzyme. An example of this kind of gene cluster function is found in *L. japonicus*, where cyanogenic glucoside bioproduction is initiated by CYP450 (CYP79D3/D4), which catalyzes the conversion of amino acids (L-valine, and L-isoleucine) into respective oximes (2-methylpropanal oxime, 2-methylbutanal oxime). These oximes are subsequently converted to linamarin and lotaustralin (Takos et al., 2011; **Figure 2B**). CYP79 class is found to be highly conserved among higher plants and is involved in the conversion of amino acids to oximes

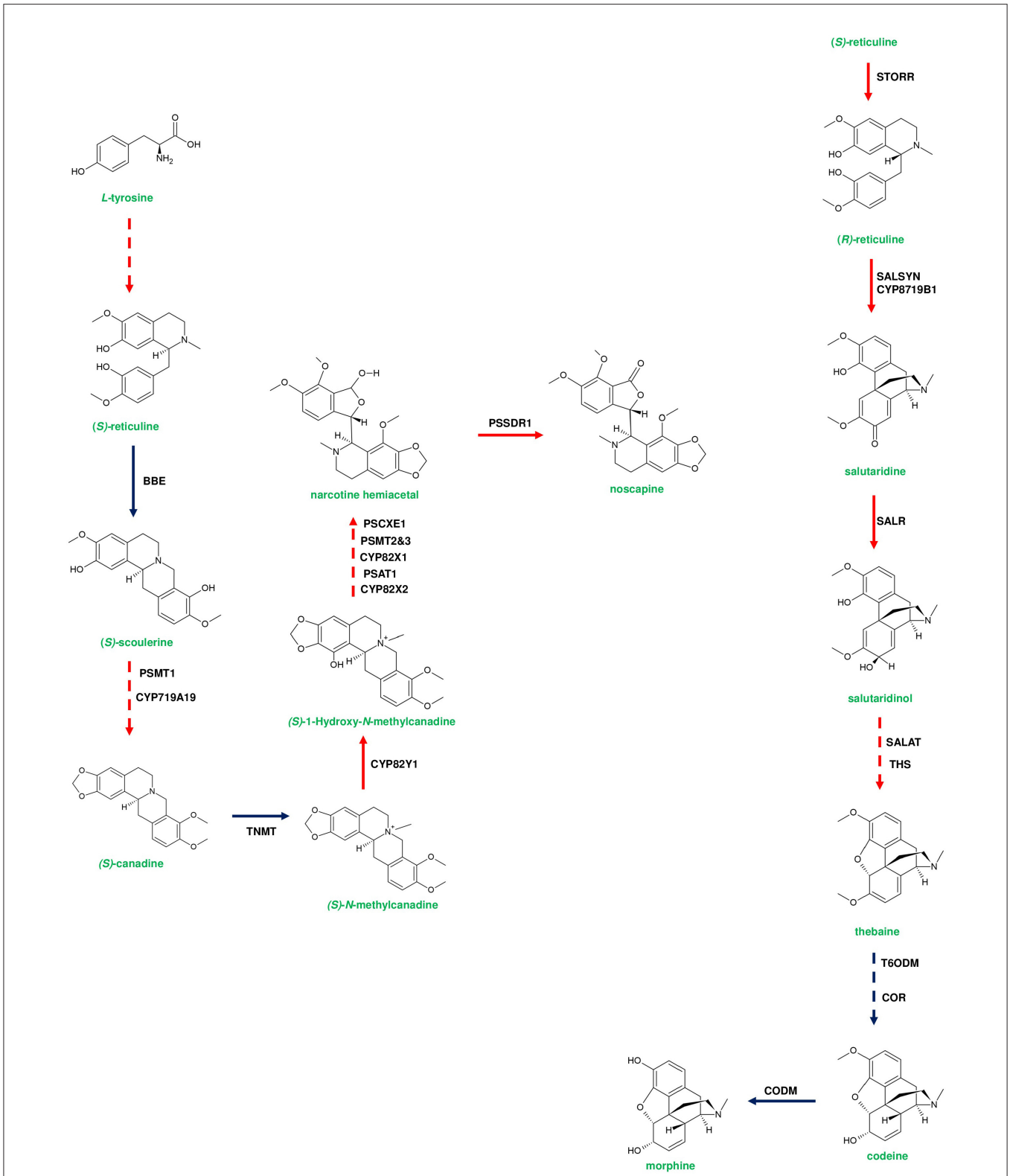


FIGURE 5 | Representative pathway showing the biosynthetic events of benzylisoquinoline alkaloids (BIAs) in *Papaver somniferum*. Genes responsible for *(S)*-reticuline biosynthesis from L-tyrosine are loosely clustered in the poppy genome. *(S)*-reticuline is converted to noscapine by a set of enzymes; berberine bridge enzyme

(Continued)

FIGURE 5 | (BBE), *P. somniferum* methyltransferase 1 (PSMT1), Tetrahydroprotoberberine N-methyltransferase (TNMT), *P. somniferum* acyltransferase 1 (PSAT1), PSMT2&3, *P. somniferum* carboxylesterase 3 (PSCXE3), and *P. somniferum* shortchain dehydrogenase 1 (PSSDR1). BBE and TNMT genes are not present in the noscapine gene cluster due to their additional function in sanguinarine biosynthesis. Thebaine-producing genes are clustered and occur adjacent to the noscapine gene cluster. Thebaine biosynthesis is initiated from *(R)*-reticuline. *(S)*-reticuline is converted to *(R)*-reticuline by *(S)*- to *(R)*-reticuline (STORR) enzyme. The formation of thebaine occurs through the activity of salutaridine synthase (SALSYN), salutaridine reductase (SALR), salutaridinol-7-*O*-acetyl transferase (SALAT), and thebaine synthase (THS). Genes involved in morphine biosynthesis are not clustered. Thebaine is converted to codeine by two subsequent steps catalyzed by thebaine 6-*O*-demethylase (T6ODM), and codeinone reductase (COR). Finally, codeine is converted to morphine by Codeine3-*O*-demethylase (CODM). Genes of plant metabolic gene cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represent a multi-step pathway.

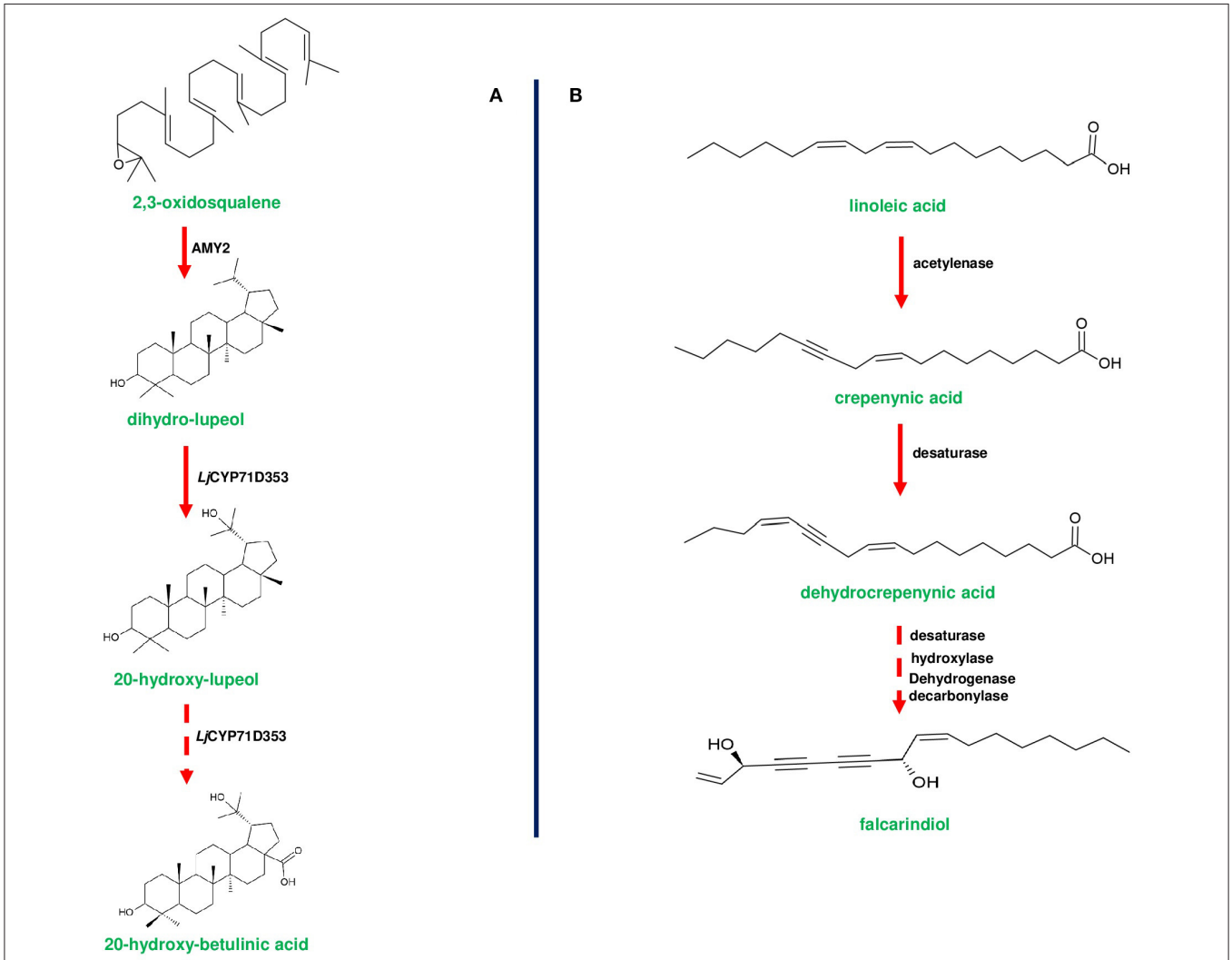


FIGURE 6 | Biosynthetic pathways of 20-hydroxy-betulinic acid in *Lotus japoinucs* and anti-pathogenic fatty acid falcarindiol in tomato. **(A)** shows the biosynthesis of 20-hydroxybetulinic acid from 2,3-oxdiosqaulene by the enzymes encoded by a triterpene producing gene cluster in *Lotus japonicus.* AMY2 (*β*-amyrin synthase) converts 2,3-oxidosqualene into a unique metabolite dihydro-lupeol. Subsequently dihydro-lupeol is converted to 20-hydroxy-betulinic acid by two consecutive reactions catalyzed by *L. japonicus* CYP71D353. **(B)** represents the bioproduction of falcarindiol from the primary metabolite linoleic acid by the enzymes of anti-pathogenic fatty acid gene cluster in *S. lycopersicum* (tomato). Linoleic acid is converted to crepenyic acid by acetylenase enzyme encoded by *Solyc12g100240* gene. Crepenyic acid is converted into dehydrocrepenyic acid by a desaturase (Solyc12g100250). Subsequently, dehydrocrepenyic acid is converted to falcarindiol by a set of desaturase, hydroxylase, and decarboxylase. Dotted arrows represent a multi-step pathway.

for producing defense compounds such as glucosinolates and camalexin (Halkier et al., 2002; Takos et al., 2011). Similarly, *GAME7* encoding CYP72 class CYP450 modifies cholesterol to 22-hydroxycholestrol for SGA production in *Solanum* species (**Figure 3B**). Interestingly, *GAME7* was reported to be 7,880 kb away from the SGA cluster on chromosome 7 (Itkin et al., 2013; Cárdenas et al., 2016; Nützmann et al., 2016). Yet another interesting inference has been made in *Solanum* species where the biosynthetic machinery of cholesterol, the precursor for SGAs is evolved from phytosterol pathway through gene duplications

and also *GAME12* of SGA pathway is likely to be evolved from gamma amino butyric acid transaminases (*GABA2*) to impart nitrogen into the steroidal backbone (Sonawane et al., 2016, 2020). In poppy, a loosely arranged gene cluster (seven genes, spanning a distance of 5 Mb across the chromosome) diverts the flux of tyrosine toward *(S)* - reticuline biosynthesis, which is the precursor for the biosynthesis of benzylisoquinoline alkaloids (BIA). On the other hand, *P. somniferum methyltransferase 1 (PSMT1)* and *(S)- to (R)-reticuline* (*STORR*) mediate the committed step in noscapine and thebaine biosynthesis, and these two genes are found to be arranged in proximity within their respective clusters (Winzer et al., 2015; Guo et al., 2018; Li et al., 2020; **Figure 5**).

Tailoring Enzymes of Plant Gene Clusters: Importance of Cytochrome P450s

Plant gene clusters possess various tailoring enzymes for modifying the backbone of signature metabolite produced through the first catalytic reaction. These enzymes include CYP450s, acyl transferases, UDP-glycosyl transferases (UGTs), short-chain alcohol dehydrogenases, transaminases, and decarboxylases (**Table 2**). CYP450s represent the largest enzyme family playing a significant role in the structural diversification of terpene scaffolds (Ghosh, 2017). Certain classes of terpene synthases and CYP450s (TPS/CYP450) form non-random pairing and are distributed among different gene clusters in both eudicots and monocots (Boutanaev et al., 2015). Boutanaev et al. (2015) reported that TPS/CYP450s pairs might have an independent evolutionary origin among different families of angiosperms. In *Arabidopsis*, triterpene cyclases (TTC) and CYP71 clan (CYP705 family) exist as pairs in both thalianol and marneral clusters, and it has been observed that TTC/CYP71 pairs are duplicated from a single founder cluster and then both clusters independently recruited another CYP71 family genes, which include *THAH* of thailanol, marneral oxidase (*MRO*) of marneral clusters to complete the cluster organization (Field et al., 2011; **Table 2**). Two CYP450s of thalianol cluster are involved in the conversion of thalianol to (-)-16-keto-3β,7β,15-thaliantriol, and this metabolic intermediate is converted to thalianin upon subsequent reactions catalyzed by four gene products (two genes encoding BAHD acyl transferases: *THAA1*, *THAA2*, and other two oxido-reductases: *THAR1*, *THAR2*), a key metabolite involved in modulating microbiome of the *Arabidopsis* roots (Huang et al., 2019). Further two oxido-reductases, *THAR1*, *THAR2* are unlinked to the thalianol cluster and can also act on other triterpenes of *Arabidopsis* such as arabidiol and tirucalladienol resulting in their respective oxidative products (Huang et al., 2019; Liu et al., 2020a,b).

CYP71D353 belonging to the triterpene cluster of *L. japonicus* catalyzes the final steps of converting dihydro lupeol to 20-hydroxy-betulinic acid, and this enzyme is found to be phylogenetically similar to MRO (CYP71A16; marneral oxidase) of marneral cluster belonging to CYP71 clan CYP450s (Krokida et al., 2013; **Figure 6A**). Additionally, it is reported that CYP71 clan enzymes can oxidize various triterpenes, such as, lupeol, β-amyrin, and α-amyrin, leading to more specialized metabolites (Yasumoto et al., 2016). TTC/CYP81Q58 pair along with CYP88L2 and *Cs sativus* acyl transferase (*cs*ACT) accounts for the bioproduction of cucurbitadienol and its further conversion to cucurbitacin C (CuC) in *Cucumis sativus* (Shang et al., 2014). Interestingly, the other two *Cucumis* species (musk melon and water melon) do not produce CuC due to non-functional CYP88L2, instead they produce cucurbitacin B (CuB) and cucurbitacin E (CuE) through the catalytic activity of CYP87D20 and CYP81Q59, respectively (Zhou et al., 2016; **Figure 3A**). Moreover, CYP450s and tissue specific regulators involved in the biosynthesis of cucurbitacins were differentially expressed in both wild and domesticated species of *Cucumis,* indicating the operation of environmental selection force on cluster organization (Zhou et al., 2016).

Rice phytoalexin gene cluster (phytocassanes and oryzalides) has a combination of diterpene synthase genes with CYP450s of CPY71 clan (TPS/CYP71). GGPP is converted into *ent*-copalyl diphosphate (*ent*-CDP) by CPS2, then two kauerene synthase like genes (*KSL6* and *KSL7*) produce *ent*-isokaurene and *ent*-cassadiene from *ent*-CDP, which are the immediate precursors for oryzalides and phytocassanes, respectively (Wu et al., 2011; **Figure 4**). CYP71Z6 mediates primarily the biosynthesis of oryzalides, and on the other hand, CYP76M7 & CYP76M8 and CYP71Z7 catalyze subsequent steps of phytocassane biosynthesis (Swaminathan et al., 2009; Wu et al., 2011; **Figure 4**). Similarly, rice momilactone cluster contains CDP synthase (*OsCPS4*), kaurene synthase like (*KSL4*), CYP99 family (*CYP99A2/CYP99A3*) genes, and dehydrogenase (*Os*MAS), which are collectively involved in the production of momilactone (Shimura et al., 2007; Wang et al., 2011; **Figure 4**). The elucidation of the momilactone pathway by De La Peña and Sattely (2021) led to the identification of two other CYP450s (CYP701A8 and CYP76M14) that are responsible for the final conversion of momilactone A to momilactone B (De La Peña and Sattely, 2021). Furthermore, Kitaoka et al. (2021) proposed a novel biosynthetic route of momilactone biosynthesis through deciphering the role of CYP76M8, which converts *syn*-pimaradien-19-al to 6β-hydroxy-*syn*-pimaradien-19-al, a key intermediate in momilactone production. Afterward, this metabolite gets converted to momilactone through subsequent steps catalyzed by *Os*MAS1 or *Os*MAS2 and CYP701A8, respectively (Kitaoka et al., 2021). CYP76M8 is located in the phytocassane cluster, which is a close relative of CYP76M7. However, the functioning of CYP76M8 together with momilactone cluster genes for metabolite biosynthesis indicates the interdependent evolution of gene clusters as a selective advantage (Kitaoka et al., 2021). The promiscuous activity of CYP76M8 is further supported by its ability to hydroxylate several diterpene metabolites of rice (Wang et al., 2012). Moreover, the bryophyte *C. plumiforme* possess four genes *CpDTC1*/*HpDTC1*, *CpMAS*, *CpCYP970A14,* and *CpCYP964A1* related to momilactone production, and the complete biosynthetic pathway is yet to be elucidated (Mao et al., 2020; Zhang and Peters, 2020).

CYP726A14 and CYP726A15 of casbene cluster in *R. communis* perform the conserved step reaction by oxidizing casbene and neocembrene at the fifth position, and the oxidation of the fifth carbon atom is conserved among the casbene-derived metabolites distributed across *Euphorbiaceae* (King et al., 2014). In *Jatropha curcas,* CYP726A35, orthologous to the CYP726A18 oxidizes the fifth position of casbene metabolites (King et al., 2016). Few other CYP450s (CYP726A19, CYP71D495) in the clusters of *R. communis* and *J. curcas* catalyzes the oxidation of casbene and its derivatives for further metabolic diversification (King et al., 2014, 2016). Recently, Zhan et al. (2020) identified a similar 5,10-diketo-casbene producing gene cluster (*OsTPS28*, *OsCYP71Z2*, *and OsCYP71Z21*) on the seventh chromosome of rice while *Os*TPS28 converts GGPP to casbene and the other two CYP450s converts casbene to 5,10-diketo-casbene (Zhan et al., 2020).

Oat avenacin cluster possesses an unique combination of TTC and CYP51 (*AsCyp51H10*) encoding enzymes of the first two steps of avenacin biosynthesis (**Figure 1**). CYP51 family has been shown to play a specific role in sterol biosynthesis, but it has been demonstrated that *AsCyp51H10* was recruited from ancient CYP51 family and acquired new functions in avenacin metabolism (Haralampidis et al., 2001; Qi et al., 2006; **Figure 1**). On the other hand, CYP51 gene of sterol biosynthesis in *Solanum* species performs additional function in cholesterol biosynthesis, which is a precursor of SGAs (Sonawane et al., 2016). Moreover, oat avenacin cluster has been proposed to have evolved independently and orthologous genes and metabolite production were not found in other monocots (Qi et al., 2004). Meanwhile, similar pairing of TTC/CYP51 is observed in a cereal member, *viz*., *Brachypodium distachyon*, its functional similarity with the avenacin cluster is still found to be inconclusive (Boutanaev et al., 2015). Along with *CYP51H10*, three more CYP450s have been shown to be associated with avenacin biosynthesis, out of which two CYP450s belong to the CYP72 clan. CYP72A475 and CYP72A476 participate in C-21-β-hydroxylation and C-30 aldehyde group addition to β-amyrin scaffold, respectively (Leveau et al., 2019; Li et al., 2021a). CYP94D65 is responsible for adding a hydroxyl group at C-23 position of β-amyrin scaffold (Li et al., 2021a). In a comparable way, CYP71C family genes (*Bx2* to *Bx5*) of maize involved in DIMBOA biosynthesis are of monophyletic origin and homologous genes were not observed in other monocots other than *Bx2* (Frey et al., 2009; **Figure 2A**). These set of genes might have evolved before divergence of *Triticeae* and *Panicoideae* (Frey et al., 2009).

The non-random pairing of TPS and CYP450s were also present in the genomes of the members of *Lamiaceae* (Lichman et al., 2020). For instance, the genome assembly of *Tectona grandis* revealed the presence of 41 terpene synthase genes in 14 tandem clusters, and 20 TPS together with 31 CYP450s were physically clustered in the genome. *TPS-c*/*CYP76AH31* combination of *T. grandis* is involved in diterpene biosynthesis displaying a high percentage of homology toward *Salvia miltiorrhiza Sm*CPS1/CYP76AH12 responsible for the production of tanshinones (Zhao et al., 2019). In the case of *S. miltiorrhiza SmCPS1* and *SmCPS2,* genes were flanked with CYP76AH subfamily genes involved in tanshinone biosynthesis, whereas these clusters might have formed from the ancestral duplication event of *CPS*/*CYP76AH* pair (Xu et al., 2016). Similarly, in *Salvia splendens* genome, eight clusters of TPS/CYP450 combinations were found, even though the functional role of these combinations are yet to be deciphered (Dong et al., 2018). The occurrence of gene pairs of terpenoid biosynthesis is not restricted to TPS and CYP450 combinations alone in *Brassicaceae.* The distinct pairs of prenyl transferase and terpene synthases were spread across the genome of several members, which are responsible for the biosynthesis of sesterterpenoids and these pairs might have evolved from a common ancestral pair *via* duplication and functional divergence (Huang et al., 2017). It has been reported that the construction of Lavender (*Lavandula angustifolia*) genome assembly led to the identification of TPS-TPS, TPS-BAHD acyltransferases, and TPS-CYP450 gene pairs, which are found to be induced upon stress conditions (Li et al., 2021b). It has been inferred that tandem duplications might have been the driving force for the emergence of these gene combinations (Li et al., 2021b). In addition, it has been reported that the coupling of terpene synthases and CYP450s P450s in gene clusters are conserved in eudicots, but not in monocots. The occurrence of transposable elements (in eudicots) and the sub-telomeric position of clusters (in monocots) indicate the possibilities of the occurrence of recombination events leading to a novel cluster formation resulting in metabolic diversity (Qi et al., 2006; Field et al., 2011; Boutanaev et al., 2015).

Acyl Transferases and Glycosyl Transferases: Avoiding the Accumulation of Toxic Intermediates

Plants have evolved in such a way to physically link genes as clusters to avoid the accumulation of toxic intermediates (Nützmann et al., 2016). The addition of sugar moieties and acyl groups to metabolic pathway intermediates, respectively, through glycosyl transferases and acyl transferases results in reducing the cytotoxicity (Itkin et al., 2011). It is known that the biological activity of triterpenes requires acylation at C-21 position (Podolak et al., 2010). Accordingly, des-acyl avenacins are acylated at C-21 position by serine carboxypeptidase like (SCPL) acyltransferase (Sad7), which incorporates *N*-methyl anthraniloyl group to produce avenacin A1 (Mugford et al., 2009; **Table 2**; **Figure 1**). In general, SCPL acyltransferases utilize β-acetyl glucose esters as acyl donors and, in oat Sad7 utilizes *N*-methyl anthraniloyl-*O*-glucopyranose (NMA-Glc) as the acyl donor (Mugford and Osbourn, 2010; Ciarkowska et al., 2019). NMA-Glc is synthesized from anthranillic acid through two sequential steps catalyzed by methyltransferase (Sad9) and UDP-glucosyltransferase (Sad10), respectively (Owatworakit et al., 2012; Mugford et al., 2013; **Figure 1**). It is interesting to note that *Sad7*, *Sad9,* and *Sad10* occur adjacently in the avenacin gene cluster and regarded as acylation module (Mugford et al., 2013). In addition, Sad7 can acylate des-acyl avenacins with benzoyl group donated by benzoyl-β-glucopyranose to form avenacin A2 (Mugford et al., 2009; Owatworakit et al., 2012). Hence,

the scaffold diversity of avenacins is attributed to the SCPL acyltransferase (Sad7) and availability of acyl donors (**Figure 1**).

Besides acylation, triterpene backbone of avenacin is glycosylated with branched trisaccharide moiety (one arabinose and two glucose molecules), which is essential for its antimicrobial activity (Louveau et al., 2018). In the first instance, the addition of arabinose to β-amyrin scaffold is mediated by an arabinosyltransferase (*As*AAT1) and the mutants of *asaat1* exhibited susceptibility to pathogenic *Gaeumannomyces* sp. (Louveau et al., 2018). Subsequently, two glucose molecules are attached to L-arabinose through glucosyltransferase and glycosyl hydrolase encoded by *AsUGT91G16* and *A. strigosa* transglucosidase 1 (*As*TG1), respectively (**Figure 1**). *As*UGT91G16 adds 1,2 linked glucose to arabinose, and *As*TG1 performs the final addition of 1,4 linked glucose (Orme et al., 2019). Remarkably, *AsTG1* (synonymous to *Sad3*; a core gene of avenacin cluster) encodes a unique vacuolar glycosyl hydrolase 1 (GH1), and the final addition of glucose moiety is known to occur in the vacuole, although an activity of glycosyl transferases is predominant in the cytosol as well (Hansen et al., 2013; Orme et al., 2019; **Figure 1**). Furthermore, *sad3* mutants accumulated non-glycosylated avenacins, which resulted in stunted root growth and deformed root hairs in oat, implicating the importance of glycosylation for normal avenacin bioactivity (Mylona et al., 2008). While the core cluster of avenacin exists in a sub-telomeric region of chromosome 1, *AsUGT91G16* and *AsTG1* genes exist in a location away from the core cluster in a proximal scaffold to avoid telomere gene deletions, which in turn help to avoid the accumulation of toxic intermediates (Li et al., 2021a). In *Arabidopsis*, three BAHD acyltransferases, *THAA1*, *THAA2*, and *THAA3*, are involved in triterpene metabolism, and acetylates signature metabolites, such as thalianol, arabidiol, and deletions of the *THAA2* locus, could lead to the development of short roots as compared to the wild type although an internal molecular cue is yet to be deciphered (Huang et al., 2019; Liu et al., 2020a,b; Bai et al., 2021).

Plants producing benzoxazinoids (DIMBOA) glycosylate them to avoid auto toxicity and store these biomolecules in the vacuole, while during defense response β-glucosidase cleaves sugar molecule and activates the function (Von Rad et al., 2001; Dick et al., 2012; **Figure 2A**). *Solanum* species also glycosylate SGAs to avoid auto toxic effects, and four UGTs (*GAME1*, *GAME17*, *GAME18*, and *GAME2*) of α-tomatine gene cluster adds tetrasaccharide moieties (one galactose, one xylose, and two glucose molecules) to tomatidine for the formation of α-tomatine (Itkin et al., 2013; **Figure 3B**). Similarly, α-solanine and α-chaconine are glycosylated with glucose:galactose:rhamnose and rhamnose:galactose:rhamnose, respectively (Ohyama et al., 2013). Alterations in GAME1 activity have been shown to result in the accumulation of high levels of tomatidine and as a consequence plants exhibited defective fruit development and increased susceptibility to pathogen attack. It has been reported that the accumulation of α-tomatine and its subsequent conversion to esculeoside A is important for the normal development of tomato fruits (Itkin et al., 2011). Additionally, *GAME25* encodes a short-chain dehydrogenase, which catalyzes the reduction of a double bond at the C-5,C-6 position and is considered as a key step in determining the diversity of SGAs in *Solanum* species. In addition, this modification has been shown to result in the reduction of the toxic effects of SGAs (Sonawane et al., 2018).

In a similar way, cyanogenic glucosides are glycosylated at the final step and transported to the vacuole (Gleadow and Møller, 2014). Cyanogenic glucoside producing the gene clusters of *L. japonicus* and *S. bicolor* are equipped with UGTs so as to catalyze the final glycosylation step for producing linamarin, lotaustralin, and dhurrin (Takos et al., 2011; **Figure 2B**). Independent recruitment of genes and their coordinated expression for metabolic channeling are the characteristic features of plant gene clusters to avoid the accumulation of toxic metabolic intermediates. Extensive mining of plant genomes can reveal the novel gene clusters, and it would be interesting to study their architecture related to the arrangement of the constituent genes of the cluster and their functional attributes in metabolite production.

REGULATION OF PLANT GENE CLUSTERS

Plant Gene Clusters Are Stress-Induced and Exhibit Spatiotemporal Expression

In general, plant gene clusters are spatiotemporally regulated (Qi et al., 2006; Shimura et al., 2007; **Table 1**). The expression of *AsbAs1* of avenacin cluster and other genes are found to be localized specifically in root tips, depicting that the biosynthesis and storage of metabolites are tissue specific (Haralampidis et al., 2001; Qi et al., 2004, 2006). In maize, the biosynthesis of DIMBOA is reported to be developmentally regulated with a localized expression restricted to leaves and roots (Frey et al., 1997). The expression of thalianol and marneral gene clusters of *Arabidopsis* are localized in root epidermis. However, the physiological function of the clusters is yet to be deciphered. The overexpression of these clusters resulted in dwarfing and negatively affected the overall plant development in *Arabidopsis* (Field and Osbourn, 2008; Field et al., 2011). In addition, triterpenes produced by *Arabidopsis* roots are reported to modulate the microbial diversity in the rhizosphere by promoting the enrichment of species-specific bacteria on root surfaces (Huang et al., 2019) though the role of microbiota in plant development is not conclusive. Mutant *Arabidopsis* lines of triterpene biosynthetic genes significantly reduced microbial operational taxonomical units (OTUs) compared to wild type (Huang et al., 2019). In continuation with this, Chen et al. (2019a) identified a group of sesterterpene cluster-derived metabolites in *A. thaliana,* which were shown to modulate the density of root microbiota. This observation was further confirmed by analyzing OTUs in wild-type and mutant lines (Chen et al., 2019a).

Diterpene clusters in rice, such as phytocassanes, momilactones, and oryzalides, are activated by chitin oligosaccharide-based elicitation and UV irradiations, and their biosynthesis is localized in leaves and roots (Shimura et al., 2007; Swaminathan et al., 2009). Pathogen-induced capsidiol gene cluster has been identified in plants such as capsicum and tobacco, and the cluster is highly conserved in both the species

(Lee et al., 2017; Chen et al., 2019b). CYP450s catalyzing the biosynthesis of linamarin and lotaustralin was found to be highly expressed in younger apical leaves and the expression gradually diminished with plant age (Takos et al., 2011). In a similar way, the expression of lycosantalonol-producing genes leaf petiole, and β-phellandrene gene expression is confined to leaves (Matsuba et al., 2015; Zhou and Pichersky, 2020). Interestingly, 20-hydroxy-betulinic acid-producing gene cluster in *L. japonicus* is highly expressed during nodule formation, thus portraying its role in plant development (Krokida et al., 2013). Therefore, it is evident that gene clusters evolved as adaptive strategies to produce defense-specific metabolites in a rapid and coordinated fashion to evade both biotic and abiotic stress conditions.

Role of Transcription Factors and Chromatin Remodeling

Three basic helix–loop–helix (bHLH) transcription factors (TFs) related to cucurbitacin biosynthesis have been identified in cucurbits, *viz*., *bitter fruit* (*Bt*), *bitter leaf* (*Bl*), and *bitter root* (*Br*), and the expression of these TFs is specific to fruits, leaves, and roots, respectively. These regulators can strongly bind to the promoter regions of nine cucurbitacin biosynthesis genes and influence the biosynthesis of CuC, CuB, and CuE (Shang et al., 2014; Zhou et al., 2016). In addition, it has been shown that domesticated cucurbit fruits (cucumber, melon, and watermelon) lost their bitterness due to a mutation in the *Bt* gene affecting its binding properties to promoter regions, although the wild species retained Bt activity (Zhou et al., 2016; Chomicki et al., 2020). In tomato and potato, *GAME9* TF [belonging to a class of Ethylene Responsive Factor (ERF)] is found to regulate SGA bioproduction. While the overexpression of *GAME9* leads to increased levels of α-tomatine, α-chaconine, the knockout of *GAME9* activity reversed the expression of SGA biosynthetic genes (Cárdenas et al., 2016). Recently, Yu et al. (2020) identified two allelic variants ($GAME9^{135A}$ and $GAME9^{135V}$) of GAME9 in wild and domesticated *Solanum* species. In the wild species, $GAME9^{135A}$ exhibited a strong binding affinity to *GAME7* and *GAME17* gene promoters (Yu et al., 2020). However, $GAME9^{135V}$ did not display strong interactions with the promoter regions and is coordinated with another TF MYC2 to regulate SGA biosynthesis (Yu et al., 2020). In addition, JRE4 TF of tomato increased SGA accumulation by binding to the promoter region of *GAME4* gene. In addition, overexpression and knockout experiments established the role of JRE4 in SGA biosynthesis (Thagun et al., 2016; Nakayasu et al., 2017).

In rice, elicitor-induced bZIP TF *OsTGAP1* has been identified along with its possible role in momilactone biosynthesis, and it has been shown that the overexpression of *OsTGAP1* significantly enhanced metabolite accumulation (Okada et al., 2009). Intriguingly, *Os*TGAP1 does not directly regulate the transcription of the momilactone cluster, instead it binds to the intergenic regions adjacent to the momilactone cluster (Miyamoto et al., 2014). However, *Os*TGAP1 binds strongly to the promoter region of *OsDXS* gene and upregulates its transcription, probably to increase the precursor pool (Miyamoto et al., 2014). Interestingly, the overexpression of TGA factor, *Os*bZIP79 suppresses momilactone biosynthesis, and *Os*bZIP79 and *Os*TGAP1 interact with each other to form a heterodimer, which might have a role in phytoalexin biosynthesis in rice (Miyamoto et al., 2015). In addition, TFs regulating the synthesis of nicotine and terpene indole alkaloids are clustered in the genome for facilitating a coordinated activity in *Nicotiana tabacum* and *C. roseus,* respectively (Shoji et al., 2010; Singh et al., 2020).

In *Arabidopsis*, metabolic clusters possess histone 3 lysine trimethylation (H3K27me3) and histone 2 variant H2A.Z chromatin signatures regulated by sick with RSC/Rat1 (SWR) complex-mediated chromatin remodeling (Yu et al., 2016). Actin-related protein (Arp6), a subunit of SWR complex, incorporates a histone 2 variant (H2A.Z) into nucleosomes of cluster genes and facilitate metabolite bioproduction. The downregulation of the cluster gene expression in the mutants of *arp6* and *h2a.z* further confirmed their regulatory role in the cluster (Nützmann and Osbourn, 2015). Reimegård et al. (2017) identified the similar modifications performed by histone methylases and histone deacetylases to facilitate the expression of development-specific gene clusters of *Arabidopsis*. Zhan et al. (2020) reported a histone demethylase JMJ705 activated upon methyl jasmonate treatment, which acted antagonistically on H3K27me3 of a chromatin region and upregulated the expression of rice casbene gene cluster. Recently, Nützmann et al. (2020) reported that metabolic gene clusters occurred in the local interactive domains of plant genome, and these domains surrounding the clusters facilitated the tightly coordinated expression of genes in the cluster. Further, the localization of the silenced clusters occurred in the periphery of the chromosomes while the expressing clusters are positioned in a location away from the periphery (Nützmann et al., 2020). In addition, chromatin-level remodeling is evident in the regulation of specialized metabolite gene clusters in filamentous fungi (Bok et al., 2009). Wegel et al. (2009) demonstrated that chromatin decondensation at *Sad1*, *Sad2* genes occurs only in root tip epidermis to initiate avenacin synthesis. Similar decondensation process was minimal in other tissues, although internal molecular cascades of cell-specific chromatin decondensation is yet to be understood.

PHYSIOLOGICAL ACTIVITIES OF PLANT GENE CLUSTER-DERIVED METABOLITES

Metabolites produced by gene clusters exhibit a specific role in defense responses and few other metabolites are involved in regulating the cell physiology during plant development (Krokida et al., 2013; Louveau et al., 2018; **Table 1**). Avenacins synthesized by oat exhibit antifungal properties and are known to inhibit fungal infection, by acting against *Blumeria graminis*, *Bipolaris oryzae*, and *Magnaporthe oryzae* (Inagaki et al., 2013). Non-glycosylated avenacins (*Sad3* mutants) severely affected the formation of root epidermis, implicating the biosynthesis and accumulation of specific intermediates of specialized metabolites is cytotoxic to the plant (Mylona

et al., 2008). On the other hand, rice momilactones are induced by UV irradiation and are active against a wide range of fungal pathogens such as *Magnaporthe grisea*, *Botrytis cinerea*, *Fusarium solani*, *Colletrotrichum gloeosporioides*, and *Phytophthora infestans* (Zhao et al., 2018). Similarly, DIMBOA and its derivatives are found to be effective against *Ralstonia solanacearum* and *Rhizoctonia solani,* which are the causative agents of bacterial wilt and sheath blight disease, respectively (Song et al., 2011; Guo et al., 2016). Besides, cyanogenic glucosides, such as, linamarin, lotaustralin, and dhurrin, have been shown to exhibit anti-herbivore and anti-insect properties (Gleadow and Møller, 2014). The removal of the saccharide moiety by glucosidase from cyanogenic glucosides and DIMBOA is essential for its activity. Plants sequester specific glucosidases in plastids to avoid auto toxicity. When cells get ruptured due to mechanical injury, these enzymes get released and cleave glucose moiety to induce defense responses (Gleadow and Møller, 2014). Few insects such as burnet moth can ingest cyanogenic glucosides and use it for its own defense, and also this moth has been shown to possess cyanogenic glucoside-producing genes and capable of synthesizing them on its own (Jensen et al., 2011). Cyanogenic glucoside gene machinery might have co-evolved with insects to develop an immunity against plant defense.

Metabolites, such as α-tomatine and tomatidine, display neurotoxic effects in humans by producing truncated proteins and deactivating proteasomes (da Silva et al., 2017). Unripe fruits are known to produce relatively high amounts of α-tomatine and during the process of ripening it gets gradually converted into esculeoside A, while the disruption of α-tomatine biosynthesis leads to the development of deformed fruits (Itkin et al., 2011). SGAs also possess anticancer, anti-inflammatory, antioxidant, and cardiovascular curative properties (Friedman, 2013; Al Sinani and Eltayeb, 2017). α-tomatine suppressed the metastasis of human lung cells by downregulating focal adhesion kinase (FAK), phosphatidylinositol 3-kinase (PI3K), and nuclear factor kappa B (NF-κB), which are potentially involved in cancer cell migration (Shieh et al., 2011). Interestingly, *Fusarium oxysporum*-derived tomatinases detoxify the effects of α-tomatine and suppresses host defense mechanisms (Ito et al., 2004). It can be postulated that pathogenic fungi might produce these enzymes as adaptive mechanisms to counteract host immune responses.

Casbene-derived metabolites, such as prostratin and ingenol-3,20-dibenzoate (IDB), are known to activate protein kinase C- (PKC-) mediated signaling, which is involved in repolarizing cardiac muscle cells, and could be used for treating cardiovascular disorders (Jiang et al., 2018). Prostratin and Ingenol derivatives are known to be potential latency reversal agents (LRA) and are found to clear the latently infected cells of HIV (Sloane et al., 2020). In addition, casbene-derived ingenol mebutate upsurges the neutrophil-mediated tumor cell degradation of subcutaneous melanoma (Braun et al., 2018). Noscapine, a notable BIA, is analgesic and used in the preparation of cough syrups due to its antitussive properties (Winzer et al., 2015). Thebaine is widely used in industries to develop semisynthetic drugs, such as oxycodone, oxymorphone, etorphine, nalbuphine, and naloxone, which are used as analgesics and also in the treatement of opoid poisoning (Hagel and Facchini, 2013; Singh et al., 2019). Metabolic engineering approaches for the heterologous expression of these clusters could be developed for the heterologous production of these metabolites for agriculture and pharmaceutical applications.

PLANT GENE CLUSTERS FOR METABOLIC ENGINEERING

The biosynthesis of specialized metabolites in plants underpins many traits of ecological, pharmaceutical, and agronomic importance. However, plant-derived products have not gained importance till the recent times due to the domination by synthetic chemical analogs. Recent era has evidenced the emergence of plant-based natural compounds as the potential alternatives against synthetic counterparts in both agricultural and pharmaceutical sectors. Furthermore, advances in the omics approach led to the identification of plant metabolic gene clusters involved in the biosynthesis of terpenoids, alkaloids, benzoxazinoids, and cyanogenic glycosides, that have several applications. Novel findings from several studies shed light on developing sustainable metabolic engineering strategies for the overproduction of some of these specialized metabolites to meet market demands. In this regard, *in silico* approaches, including cluster predicting tools, such as PhytoClust, plantiSMASH, ClusterFinder, and ClustScan to name a few, have accelerated the discovery of diverse metabolic pathway-related gene clusters in plants (Chavali and Rhee, 2018). Technological advances in genome sequencing and availability of high-throughput functional genomics tools resulted in a shift from a single-step characterization to the validation of entire metabolic pathways. Increased knowledge of metabolic gene clusters accelerated genome engineering strategies for the biosynthesis of alkaloids and terpenoids in heterologous host systems. Nevertheless, low precursor availability, the accumulation of unwanted intermediates, and hindrance due to the lack of information regarding genes/regulatory steps are some of the bottlenecks to overproduce these compounds through synthetic biology approaches.

Primarily, the biosynthetic pathways of plant specialized metabolism are complex with several gene cascades, enzymatic reactions, and compartmentalized intermediates and or end products. Hence, a thorough understanding of the pathway dynamics is necessary to develop a strategy for heterologous metabolite production. For instance, the introduction of 12 artemisinin biosynthetic genes into tobacco chloroplasts severely affected the growth of the plant (Saxena et al., 2014). However, dividing the pathway and expressing the genes into two compartments (plastids and cytosol) successfully yielded a reasonable titer of artemisinin in tobacco that could be further used for commercial application (Malhotra et al., 2016). De La Peña and Sattely (2021) increased the biosynthesis of momilactone by expressing plastid localized geranylgeranyl diphosphate synthases (GGPPS) and CPS in cytosol through

plastid tag truncation (De La Peña and Sattely, 2021). Another strategy is to modulate the precursor flux by engineering additional copies of rate limiting step genes or by silencing the branch point genes so that the flux can be diverted to the synthesis of the desired metabolite. This approach was attempted in *Artemisia annua* by downregulating the expression of *β-caryophyllene synthase* gene to increase the bioproduction of artemisinin (Lv et al., 2016). A similar strategy of pathway shunting can be achieved by gene editing tools, such as CRISPR-Cas9, which is currently being used widely in metabolic engineering (Sabzehzari et al., 2020). Furthermore, the identification of TFs that regulate multiple steps in a biosynthetic pathway is yet another approach for overproducing specific metabolites. GAME9 TF of the SGA cluster regulates both the MEP pathway and SGA pathway (Cárdenas et al., 2016). Finally, improving the cell numbers in the plant, such as increasing the trichome on the leaves, can also improve the metabolite bioproduction specific to a particular tissue (Fu et al., 2018).

The use of robust microbial hosts, such as *Escherichia coli* and yeast, is another promising approach for overproducing target metabolites by synthetic biology (Pyne et al., 2019). Celedon et al. (2016) engineered a multistep pathway in yeast cells for producing sandalwood oil similar to the one extracted from the heartwood of *Santalum album*. In addition, opoids were successfully produced in yeast through the introduction of morphine biosynthesis genes (Galanie et al., 2015). Nevertheless, some of the challenges which the researchers should anticipate and foresee while using a microbial host include: (a) ways to avoid the accumulation of toxic metabolic intermediates; (b) a fine control of gene expression; and (c) precursor limitation. Taken together, the combination of *in silico* analysis, the availability of robust functional genomic tools, and the knowledge about metabolic gene clusters could shed light into new directions in synthetic biology research at an accelerated level. The present scenario in this area of research holds a great promise to translate the basic knowledge of plant metabolism into tangible benefits for agricultural and pharmaceutical applications.

AUTHOR CONTRIBUTIONS

RB, SRK, and AS wrote the manuscript. RS revised it. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Dr. V. S. Pragadheesh, Scientist, CSIR-Central Institute of Medicinal and Aromatic Plants, Bengaluru, India for the help rendered with chemical structures.

REFERENCES

Al Sinani, S. S. S., and Eltayeb, C. A. (2017). The steroidal glycoalkaloids solamargine and solasonine in solanum plants. *S. Afr. J. Bot.* 112, 253–269. doi: 10.1016/j.sajb.2017.06.002

Anand, A., Jayaramaiah, R. H., Beedkar, S. D., Dholakia, B. B., Lavhale, S. G., Punekar, S. A., et al. (2019). Terpene profiling, transcriptome analysis and characterization of cis-β-terpineol synthase from ocimum. *Physiol. Mol. Biol. Plants.* 25, 47–57. doi: 10.1007/s12298-018-0612-6

Bai, Y., Fernández-Calvo, P., Ritter, A., Huang, A. C., Morales-Herrera, S., Bicalho, K. U., et al. (2021). Modulation of *Arabidopsis* root growth by specialized triterpenes. *New Phytol.* 230, 228–243. doi: 10.1111/nph.17144

Ballouz, S., Francis, A. R., Lan, R., and Tanaka, M. M. (2010). Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* 6:e1000672. doi: 10.1371/journal.pcbi.1000672

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.* 9:11769. doi: 10.1038/s41598-019-47985-w

Bok, J. W., Chiang, Y. M., Szewczyk, E., Reyes-Dominguez, Y., Davidson, A. D., Sanchez, J. F., et al. (2009). Chromatin-level regulation of biosynthetic gene clusters. *Nat. Chem. Biol.* 5, 462–464. doi: 10.1038/nchembio.177

Boutanaev, A. M., Moses, T., Zi, J., Nelson, D. R., Mugford, S. T., Peters, R. J., et al. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E81–E88. doi: 10.1073/pnas.1419547112

Boutanaev, A. M., and Osbourn, A. E. (2018). Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 115, E6650–E6658. doi: 10.1073/pnas.1721318115

Bradshaw, R. E., Slot, J. C., Moore, G. G., Chettri, P., de Wit, P. J. G. M., Ehrlich, K. C., et al. (2013). Fragmentation of an aflatoxin-like gene cluster in a forest pathogen. *New Phytol.* 198, 525–535. doi: 10.1111/nph.12161

Braun, S. A., Baran, J., Schrumpf, H., Buhren, B. A., Bölke, E., Homey, B., et al. (2018). Ingenol mebutate induces a tumor cell-directed inflammatory response and antimicrobial peptides thereby promoting rapid tumor destruction and wound healing. *Eur. J. Med. Res.* 23:45. doi: 10.1186/s40001-018-0343-8

Caballero, J. L., Martinez, E., Malpartida, F., and Hopwood, D. A. (1991). Organisation and functions of the actVA region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol. Genet. Genom.* 230, 401–412. doi: 10.1007/BF00280297

Cárdenas, P. D., Sonawane, P. D., Pollier, J., Vanden Bossche, R., Dewangan, V., Weithorn, E., et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* 7:10654. doi: 10.1038/ncomms10654

Celedon, J. M., Chiang, A., Yuen, M. M., Diaz-Chavez, M. L., Madilao, L. L., Finnegan, P. M., et al. (2016). Heartwood-specific transcriptome and metabolite signatures of tropical sandalwood (*Santalum album*) reveal the final step of (Z)-santalol fragrance biosynthesis. *Plant J.* 86, 289–299. doi: 10.1111/tpj.13162

Chavali, A. K., and Rhee, S. Y. (2018). Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief Bioinform.* 19, 1022–1034. doi: 10.1093/bib/bbx020

Chen, Q., Jiang, T., Liu, Y. X., Liu, H., Zhao, T., Liu, Z., et al. (2019a). Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. *Sci. China Life Sci.* 62, 947–958. doi: 10.1007/s11427-019-9521-2

Chen, X., Liu, F., Liu, L., Fang, D., Wang, W., Zhang, X., et al. (2019b). Characterization and evolution of gene clusters for terpenoid phytoalexin biosynthesis in tobacco. *Planta* 250, 1687–1702. doi: 10.1007/s00425-019-03255-7

Chomicki, G., Schaefer, H., and Renner, S. S. (2020). Origin and domestication of *Cucurbitaceae* crops: insights from phylogenies, genomics and archaeology. *New Phytol.* 226, 1240–1255. doi: 10.1111/nph.16015

Ciarkowska, A., Ostrowski, M., Starzyńska, E., and Jakubowska, A. (2019). Plant SCPL acyltransferases: multiplicity of enzymes with various functions in secondary metabolism. *Phytochem. Rev.* 18, 303–316. doi: 10.1007/s11101-018-9593-x

da Silva, D. C., Andrade, P. B., Valentão, P., and Pereria,. D. M. (2017). Neurotoxicity of the steroidal alkaloids tomatine and tomatidine is RIP1

kinase- and caspase-independent and involves the eIF2α branch of the endoplasmic reticulum. *J. Steroid Biochem. Mol. Biol.* 171, 178–186. doi: 10.1016/j.jsbmb.2017.03.009

Darbani, B., Motawia, M. S., Olsen, C. E., Nour-Eldin, H. H., Møller, B. L., and Rook, F. (2016). The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in Sorghum bicolor contains its co-expressed vacuolar MATE transporter. *Sci. Rep*. 6:37079. doi: 10.1038/srep37079

De La Peña, R., and Sattely, E. S. (2021). Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat. Chem. Biol*. 17, 205–212. doi: 10.1038/s41589-020-00669-3

Dick, R., Rattei, T., Haslbeck, M., Schwab, W., Gierl, A., and Frey, M. (2012). Comparative analysis of benzoxazinoid biosynthesis in monocots and dicots: independent recruitment of stabilization and activation functions. *Plant Cell.* 24, 915–928. doi: 10.1105/tpc.112.096461

Dong, A. X., Xin, H. B., Li, Z. J., Liu, H., Sun, Y. Q., Nie, S., et al. (2018). High-quality assembly of the reference genome for scarlet sage, salvia splendens, an economically important ornamental plant. *Gigascience* 7:giy068. doi: 10.1093/gigascience/giy068

Dutartre, L., Hilliou, F., and Feyereisen, R. (2012). Phylogenomics of the benzoxazinoid biosynthetic pathway of *Poaceae*: gene duplications and origin of the Bx cluster. *BMC Evol. Biol*. 12:64. doi: 10.1186/1471-2148-12-64

Elizondo, L. I., Jafar-Nejad, P., Clewing, J. M., and Boerkoel, C. F. (2009). Gene clusters, molecular evolution and disease: a speculation. *Curr. Genomics* 10, 64–75. doi: 10.2174/138920209787581271

Fan, P., Wang, P., Lou, Y. R., Leong, B. J., Moore, B. M., and Schenck, C. A. (2020). Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *Elife* 9:e56717. doi: 10.7554/eLife.56717

Field, B., Fiston-Lavier, A. S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U.S.A*. 108, 16116–16121. doi: 10.1073/pnas.1109273108

Field, B., and Osbourn, A. E. (2008). Metabolic diversification–independent assembly of operon-like gene clusters in different plants. *Science* 320, 543–547. doi: 10.1126/science.1154990

Franke, J., Kim, J., Hamilton, J. P., Zhao, D., Pham, G. M., Wiegert-Rininger, K., et al. (2019). Gene discovery in gelsemium highlights conserved gene clusters in monoterpene indole alkaloid biosynthesis. *Chembiochem* 20, 83–87. doi: 10.1002/cbic.201800592

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A., et al. (1997). Analysis of a chemical plant defense mechanism in grasses. *Science* 277, 696–699 doi: 10.1126/science.277.5326.696

Frey, M., Schullehner, K., Dick, R., Fiesselmann, A., and Gierl, A. (2009). Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70, 1645–1651. doi: 10.1016/j.phytochem.2009.05.012

Friedman, M. (2013). Anticarcinogenic, cardioprotective, and other health benefits of tomato compounds lycopene, α-tomatine, and tomatidine in pure form and in fresh and processed tomatoes. *J. Agric. Food Chem.* 61, 9534–9550. doi: 10.1021/jf402654e

Fu, R., Martin, C., and Zhang, Y. (2018). Next-Generation plant metabolic engineering, inspired by an ancient chinese irrigation system. *Mol. Plant* 11, 47–57. doi: 10.1016/j.molp.2017.09.002

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., and Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373

Ghosh, S. (2017). Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.* 8:1886. doi: 10.3389/fpls.2017.01886

Gleadow, R. M., and Møller, B. L. (2014). Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* 65, 155–185. doi: 10.1146/annurev-arplant-050213-040027

Guo, B., Zhang, Y., Li, S., Lai, T., Yang, L., Chen, J., et al. (2016). Extract from maize (Zea mays L.): antibacterial activity of DIMBOA and its derivatives against *Ralstonia solanacearum*. *Molecules* 21:1397. doi: 10.3390/molecules21101397

Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., He, Z., et al. (2018). The opium poppy genome and morphinan production. *Science* 362, 343–347. doi: 10.1126/science.aat4096

Hagel, J. M., and Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672. doi: 10.1093/pcp/pct020

Halkier, B. A., Hansen, C. H., Mikkelsen, M. D., Mikkelsen, M. D., Naur, P., and Wittstock, U. (2002). "The role of cytochromes P450 in biosynthesis and evolution of glucosinolates," in *Recent Advances in Phytochemistry, Vol. 36,* ed J. T. Romeo, and R. A. Dixon (Elsevier), 223–248. doi: 10.1016/S0079-9920(02)80029-6

Hansen, S. F., Harholt, J., Oikawa, A., and Scheller, H. V. (2013). Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Front. Plant Sci.* 3:59. doi: 10.3389/fpls.2012.00059

Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., et al. (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc. Natl. Acad. Sci. U.S.A*. 98, 13431–13436. doi: 10.1073/pnas.231324698

Hen-Avivi, S., Savin, O., Racovita, R. C., Lee, W. S., Adamski, N. M., Malitsky, S., et al. (2016). A metabolic gene cluster in the wheat W1 and the barley Cer-cqu loci determines β-diketone biosynthesis and glaucousness. *Plant Cell* 28,1440–1460. doi: 10.1105/tpc.16.00197

Holland, P. W. (2013). Evolution of homeobox genes. *WIREs Dev. Biol.* 2, 31–45. doi: 10.1002/wdev.78

Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., et al. (2004). Gene map of the extended human MHC. *Nat. Rev. Genet.* 5, 889–899. doi: 10.1038/nrg1489

Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J., Qu, B., et al. (2019). A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* 364:eaau6389. doi: 10.1126/science.aau6389

Huang, A. C., Kautsar, S. A., Hong, Y. J., Medema, M. H., Bond, A. D., Tantillo, D. J., et al. (2017). Unearthing a sesterterpene biosynthetic repertoire in the *Brassicaceae* through genome mining reveals convergent evolution. *Proc. Natl. Acad. Sci. U.S.A*. 114, E6005–E6014. doi: 10.1073/pnas.1705567114

Ichinose, K., Bedford, D. J., Tornus, D., Bechthold, A., Bibb, M. J., Revill, W. P., et al. (1998). The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: sequence analysis and expression in a heterologous host. *Chem. Biol.* 5, 647–659. doi: 10.1016/S1074-5521(98)90292-7

Inagaki, Y., Noutoshi, Y., Fujita, K., Imaoka, A., Arase, S., Toyoda, K., et al. (2013). Infection-inhibition activity of avenacin saponins against the fungal pathogens *Blumeria graminis* f. sp. hordei, *Bipolaris oryzae, and Magnaporthe oryzae. J. Gen. Plant Pathol.* 79, 69–73. doi: 10.1007/s10327-012-0412-8

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A. J., Shinde, B., Cardenas, P. D., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 341, 175–179. doi: 10.1126/science.1240230

Itkin, M., Rogachev, I., Alkan, N., Rosenberg, T., Malitsky, S., Masini, L., et al. (2011). Glycoalkaloid Metabolism1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* 23, 4507–4525. doi: 10.1105/tpc.111.088732

Ito, S., Eto, T., Tanaka, S., Yamauchi, N., Takahara, H., and Ikeda, T. (2004). Tomatidine and lycotetraose, hydrolysis products of alpha-tomatine by *Fusarium oxysporum* tomatinase, suppress induced defense responses in tomato cells. *FEBS Lett.* 571, 31–34. doi: 10.1016/j.febslet.2004.06.053

Jacob, F., Perrin, D., Sánchez, C., Monod, J., and Edelstein, S. (1960). The operon: a group of genes with expression coordinated by an operator. *C. R. Acad. Sci. Paris* 250, 1727–1729.

Jensen, N. B., Zagrobelny, M., Hjern,ø, K., Olsen, C. E., Houghton-Larsen, J., Borch, J., et al. (2011). Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. *Nat. Commun.* 3:820. doi: 10.1038/ncomms1865

Jeon, J. E., Kim, J. G., Fischer, C. R., Mehta, N., Dufour-Schroif, C., Wemmer, K., et al. (2020). A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 180, 176–187.e19. doi: 10.1016/j.cell.2019.11.037

Jiang, Q., Li, K., Lu, W. J., Li, S., Chen, X., Liu, X. J., et al. (2018). Identification of small-molecule ion channel modulators in *C. elegans* channelopathy models. *Nat Commun.* 9:3941. doi: 10.1038/s41467-018-06514-5

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., et al. (2015). Genome-guided investigation of plant natural product biosynthesis. *Plant J.* 82, 680–692. doi: 10.1111/tpj.12827

King, A. J., Brown, G. D., Gilday, A. D., Forestier, E., Larson, T. R., and Graham, I. A. (2016). A cytochrome P450-mediated intramolecular carbon-carbon ring closure in the biosynthesis of multidrug-resistance-reversing lathyrane diterpenoids. *Chembiochem* 17, 1593–1597. doi: 10.1002/cbic.201600316

King, A. J., Brown, G. D., Gilday, A. D., Larson, T. R., and Graham, I. A. (2014). Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 26, 3286–3298. doi: 10.1105/tpc.114.129668

Kitaoka, N., Zhang, J., Oyagbenro, R. K., Brown, B., Wu, Y., Yang, B., et al. (2021). Interdependent evolution of biosynthetic gene clusters for momilactone production in rice. *Plant Cell* 33, 290–305. doi: 10.1093/plcell/koaa023

Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L. M., et al. (2013). A metabolic gene cluster in Lotus japonicus discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol.* 200, 675–690. doi: 10.1111/nph.12414

Lee, H. A., Kim, S., Kim, S., and Choi, D. (2017). Expansion of sesquiterpene biosynthetic gene clusters in pepper confers nonhost resistance to the irish potato famine pathogen. *New Phytol.* 215, 1132–1143. doi: 10.1111/nph.14637

Leveau, A., Reed, J., Qiao, X., Stephenson, M. J., Mugford, S. T., Melton, R. E., et al. (2019). Towards take-all control: a C-21β oxidase required for acylation of triterpene defence compounds in oat. *New Phytol.* 221, 1544–1555. doi: 10.1111/nph.15456

Li, J., Wang, Y., Dong, Y., Zhang, W., Wang, D., Bai, H., et al. (2021b). Correction: The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Hortic Res.* 8:90. doi: 10.1038/s41438-021-00536-9

Li, Q., Ramasamy, S., Singh, P., Hagel, J. M., Dunemann, S. M., Chen, X., et al. (2020). Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat. Commun.* 11:2899. doi: 10.1038/s41467-020-16467-3

Li, Y., Leveau, A., Zhao, Q., Feng, Q., Lu, H., Miao, J., et al. (2021a). Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* 12:2563. doi: 10.1038/s41467-021-22920-8

Lichman, B. R., Godden, G. T., and Buell, C. R. (2020). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of *Lamiaceae*. *Curr. Opin. Plant Biol.* 55, 74–83. doi: 10.1016/j.pbi.2020.03.005

Liras, P., and Martín, J. F. (2006). Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate? *Int. Microbiol.* 9, 9–19.

Liu, Z., Cheema, J., Vigouroux, M., Hill, L., Reed, J., Paajanen, P., et al. (2020b). Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat. Commun.* 11:5354. doi: 10.1038/s41467-020-19153-6

Liu, Z., Suarez Duran, H. G., Harnvanichvech, Y., Stephenson, M. J., Schranz, M. E., Nelson, D., et al. (2020a). Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the *Brassicaceae*. *New Phytol.* 227, 1109–1123. doi: 10.1111/nph.16338

Louveau, T., Orme, A., Pfalzgraf, H., Stephenson, M. J., Melton, R., Saalbach, G., et al. (2018). Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family1 provides insights into disease resistance and sugar donor specificity. *Plant Cell* 30, 3038–3057. doi: 10.1105/tpc.18.00641

Lv, Z., Zhang, F., Pan, Q., Fu, X., Jiang, W., Shen, Q., et al. (2016). Branch pathway blocking in artemisia annua is a useful method for obtaining high yield artemisinin. *Plant Cell Physiol.* 57, 588–602. doi: 10.1093/pcp/pcw014

Ma, X., Vaistij, F. E., Li, Y., Jansen van Rensburg, W. S., Harvey, S., and Bairu, M. W. (2021). A chromosome-level *Amaranthus cruentus* genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *Plant J.* doi: 10.1111/tpj.15298

Malhotra, K., Subramaniyan, M., Rawat, K., Kalamuddin, M., Qureshi, M. I., Malhotra, P., et al. (2016). Compartmentalized metabolic engineering for artemisinin biosynthesis and effective malaria treatment by oral delivery of plant cells. *Mol. Plant.* 9, 1464–1477. doi: 10.1016/j.molp.2016.09.013

Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y., et al. (2020). Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc. Natl. Acad. Sci. U.S.A.* 117, 12472–12480. doi: 10.1073/pnas.1914373117

Matsuba, Y., Nguyen, T. T., Wiegert, K., Falara, V., Gonzales-Vigil, E., Leong, B., et al. (2013). Evolution of a complex locus for terpene biosynthesis in solanum. *Plant Cell.* 25, 2022–2036. doi: 10.1105/tpc.113.111013

Matsuba, Y., Zi, J., Jones, A. D., Peters, R. J., and Pichersky, E. (2015). Biosynthesis of the diterpenoid lycosantalonol via nerylneryl diphosphate in *Solanum lycopersicum*. *PLoS ONE* 10:e0119302. doi: 10.1371/journal.pone.0119302

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631. doi: 10.1038/nchembio.1890

Meena, S., Rajeev Kumar, S., Dwivedi, V., Singh, A. K., Chanotiya, C. S., Akthar, M. Q., et al. (2017). Transcriptomic insight into terpenoid and carbazole alkaloid biosynthesis, and functional characterization of two terpene synthases in curry tree (*Murraya koenigii*). *Sci. Rep.* 7:44126. doi: 10.1038/srep44126

Miyamoto, K., Fujita, M., Shenton, M. R., Akashi, S., Sugawara, C., Sakai, A., et al. (2016). Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. *Plant J.* 87, 293–304. doi: 10.1111/tpj.13200

Miyamoto, K., Matsumoto, T., Okada, A., Komiyama, K., Chujo, T., Yoshikawa, H., et al. (2014). Identification of target genes of the bZIP transcription factor OsTGAP1, whose overexpression causes elicitor-induced hyperaccumulation of diterpenoid phytoalexins in rice cells. *PLoS ONE* 9:e105823. doi: 10.1371/journal.pone.0105823

Miyamoto, K., Nishizawa, Y., Minami, E., Nojiri, H., Yamane, H., and Okada, K. (2015). Overexpression of the bZIP transcription factor OsbZIP79 suppresses the production of diterpenoid phytoalexin in rice cells. *J. Plant Physiol.* 173, 19–27. doi: 10.1016/j.jplph.2014.09.001

Mugford, S. T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., et al. (2013). Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat [published correction appears. *Plant Cell.* 25, 1078–1092. doi: 10.1105/tpc.113.110551

Mugford, S. T., and Osbourn, A. (2010). Evolution of serine carboxypeptidase-like acyltransferases in the monocots. *Plant Signal Behav.* 5, 193–195. doi: 10.4161/psb.5.2.11093

Mugford, S. T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R. K., et al. (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell.* 21, 2473–2484 doi: 10.1105/tpc.109.065870

Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., et al. (2008). Sad3 and sad4 are required for saponin biosynthesis and root development in oat. *Plant Cell* 20, 201–212. doi: 10.1105/tpc.107.056531

Nagegowda, D. A., and Gupta, P. (2020). Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Sci.* 294:110457. doi: 10.1016/j.plantsci.2020.110457

Nakayasu, M., Umemoto, N., Ohyama, K., Ujimoto, Y., Lee, H. J., and Watanabe, B. (2017). A dioxygenase catalyzes steroid 16α-hydroxylation in steroidal glycoalkaloid biosynthesis. *Plant Physiol.* 175, 120–133. doi: 10.1104/pp.17.00501

Nützmann, H. W., Doerr, D., Ramírez-Colmenero, A., Sotelo-Fonseca, J. E., Wegel, E., Di Stefano, M., et al. (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. U.S.A.* 117, 13800–13809. doi: 10.1073/pnas.1920474117

Nützmann, H. W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters - from genetics to genomics. *New Phytol.* 211, 771–789. doi: 10.1111/nph.13981

Nützmann, H. W., and Osbourn, A. (2015). Regulation of metabolic gene clusters in *Arabidopsis thaliana*. *New Phytol.* 205, 503–510. doi: 10.1111/nph.13189

Nützmann, H. W., Scazzocchio, C., and Osbourn, A. (2018). Metabolic gene clusters in eukaryotes. *Annu. Rev. Genet.* 52, 159–183. doi: 10.1146/annurev-genet-120417-031237

Ohyama, K., Okawa, A., Moriuchi, Y., and Fujimoto, Y. (2013). Biosynthesis of steroidal alkaloids in Solanaceae plants: involvement of an aldehyde intermediate during C-26 amination. *Phytochemistry* 89, 26–31. doi: 10.1016/j.phytochem.2013.01.010

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., et al. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J. Biol. Chem.* 284, 26510–26518. doi: 10.1074/jbc.M109.036871

Olsen, K. M., and Small, L. L. (2018). Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytol.* 219, 757–766. doi: 10.1111/nph.15184

Orme, A., Louveau, T., Stephenson, M. J., Appelhagen, I., Melton, R., Cheema, J., et al. (2019). A noncanonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat. *Proc. Natl. Acad. Sci. U.S.A.* 116, 27105–27114. doi: 10.1073/pnas.1914652116

Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M., Hughes, R. K., et al. (2012). Glycosyltransferases from oat (Avena) implicated in the acylation of avenacins. *J. Biol. Chem.* 288, 3696–3704. doi: 10.1074/jbc.M112.426155

Podolak, I., Galanty, A., and Sobolewska, D. (2010). Saponins as cytotoxic agents: a review. *Phytochem. Rev.* 9, 425–474. doi: 10.1007/s11101-010-9183-z

Pyne, M. E., Narcross, L., and Martin, V. J. J. (2019). Engineering plant secondary metabolism in microbial systems. *Plant Physiol.* 179, 844–861. doi: 10.1104/pp.18.01291

Qi, X., Bakht, S., Leggett, M., Maxwell, C., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8233–8238. doi: 10.1073/pnas.0401301101

Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., et al. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18848–18853 doi: 10.1073/pnas.0607849103

Reimegård, J., Kundu, S., Pendle, A., Irish, V. F., Shaw, P., Nakayama, N., et al. (2017). Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in *Arabidopsis thaliana*. *Nucleic Acids Res.* 45, 3253–3265. doi: 10.1093/nar/gkx087

Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A., and Oberlies, N. H. (2020). Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* 37, 868–878. doi: 10.1039/C9NP00045C

Rokas, A., Wisecaver, J. H., and Lind, A. L. (2018). The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* 16, 731–744. doi: 10.1038/s41579-018-0075-3

Sabzehzari, M., Zeinali, M., and Naghavi, M. R. (2020). CRISPR-based metabolic editing: Next-generation metabolic engineering in plants. *Gene* 759:144993. doi: 10.1016/j.gene.2020.144993

Saxena, B., Subramaniyan, M., Malhotra, K., Bhavesh, N. S., Potlakayala, S. D., and Kumar, S. (2014). Metabolic engineering of chloroplasts for artemisinic acid biosynthesis and impact on plant growth. *J. Biosci.* 39, 33–41. doi: 10.1007/s12038-013-9402-z

Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., et al. (2017). Genome-Wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* 173, 2041–2059. doi: 10.1104/pp.16.01942

Schneider, L. M., Adamski, N. M., Christensen, C. E., Stuart, D. B., Vautrin, S., Hansson, M., et al. (2017). The Cer-cqu gene cluster determines three key players in a β-diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J. Exp. Bot.* 68, 2715–2730. doi: 10.1093/jxb/erw105

Schullehner, K., Dick, R., Vitzthum, F., Schwab, W., Brandt, W., Frey, M., et al. (2008). Benzoxazinoid biosynthesis in dicot plants. *Phytochemistry* 69, 2668–2677. doi: 10.1016/j.phytochem.2008.08.023

Seca, A. M. L., and Pinto,. D. C. G. A. (2018). Plant secondary metabolites as anticancer agents: successes in clinical trials and therapeutic application. *Int. J. Mol. Sci.* 19:263. doi: 10.3390/ijms19010263

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Shen, S., Peng, M., Fang, H., Wang, Z., Zhou, S., Jing, X., et al. (2021). An Oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci. Bull.* doi: 10.1016/j.scib.2021.03.015

Shieh, J. M., Cheng, T. H., Shi, M. D., Wu, P. F., Chen, Y., Ko, S. C., et al. (2011). α-Tomatine suppresses invasion and migration of human non-small cell lung cancer NCI-H460 cells through inactivating FAK/PI3K/Akt signaling pathway and reducing binding activity of NF-κB. *Cell Biochem. Biophys.* 60, 297–310. doi: 10.1007/s12013-011-9152-1

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K. W., Toyomasu, T., et al. (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *J. Biol. Chem.* 282, 34013–34018. doi: 10.1074/jbc.M703344200

Shoji, T., Kajikawa, M., and Hashimoto, T. (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell.* 22, 3390–3409. doi: 10.1105/tpc.110.078543

Singh, A., Menéndez-Perdomo, I. M., and Facchini, P. J. (2019). Benzylisoquinoline alkaloid biosynthesis in opium poppy: an update. *Phytochem. Rev.* 18, 1457–1482. doi: 10.1007/s11101-019-09644-w

Singh, S. K., Patra, B., Paul, P., Liu, Y., Pattanaik, S., and Yuan, L. (2020). Revisiting the ORCA gene cluster that regulates terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Sci.* 293:110408. doi: 10.1016/j.plantsci.2020.110408

Sloane, J. L., Benner, N. L., Keenan, K. N., Zang, X., Soliman, M. S. A., Wu, X., et al. (2020). Prodrugs of PKC modulators show enhanced HIV latency reversal and an expanded therapeutic window. *Proc. Natl. Acad. Sci. U.S.A.* 117, 10688–10698. doi: 10.1073/pnas.1919408117

Slot, J. C. (2017). Fungal gene cluster diversity and evolution. *Adv. Genet.* 100, 141–178. doi: 10.1016/bs.adgen.2017.09.005

Slot, J. C., and Rokas, A. (2010). Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10136–10141. doi: 10.1073/pnas.0914418107

Sonawane, P. D., Heinig, U., Panda, S., Gilboa, N. S., Yona, M., Kumar, S. P., et al. (2018). Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus solanum. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5419–E5428. doi: 10.1073/pnas.1804835115

Sonawane, P. D., Jozwiak, A., Panda, S., and Aharoni, A. (2020). 'Hijacking' core metabolism: a new panache for the evolution of steroidal glycoalkaloids structural diversity. *Curr. Opin. Plant Biol.* 55, 118–128. doi: 10.1016/j.pbi.2020.03.008

Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., et al. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* 22:17101. doi: 10.1038/nplants.2017.101

Song, Y. Y., Cao, M., Xie, L. J., Liang, X. T., Zeng, R. S., Su, Y. J., et al. (2011). Induction of DIMBOA accumulation and systemic defense responses as a mechanism of enhanced resistance of mycorrhizal corn (Zea mays L.) to sheath blight. *Mycorrhiza* 21, 721–731. doi: 10.1007/s00572-011-0380-4

Sultana, M. H., Liu, F., Alamin, M., Mao, L., Jia, L., Chen, H., et al. (2019). ene modules co-regulated with biosynthetic gene clusters for allelopathy between rice and barnyardgrass. *Int. J. Mol. Sci.* 20:3846. doi: 10.3390/ijms20163846

Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B., and Peters, R. J. (2009). CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21, 3315–3325. doi: 10.1105/tpc.108.063677

Takos, A. M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., and Motawia, M. S. (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *Plant J.* 68, 273–286. doi: 10.1111/j.1365-313X.2011.04685.x

Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohyama, K., Mori, T., et al. (2016). Jasmonate-Responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.* 57, 961–975. doi: 10.1093/pcp/pcw067

Verma, M., Ghangal, R., Sharma, R., Sinha, A. K., and Jain, M. (2014). Transcriptome analysis of *Catharanthus roseus* for gene discovery and expression profiling. *PLoS ONE* 9:e103583. doi: 10.1371/journal.pone.0103583

Von Rad, U., Hüttl, R., Lottspeich, F., Gierl, A., and Frey, M. (2001). Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J.* 28, 633–642. doi: 10.1046/j.1365-313x.2001.01161.x

Wang, Q., Hillwig, M. L., Okada, K., Yamazaki, K., Wu, Y., Swaminathan, S., et al. (2012). Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* 287, 6159–6168. doi: 10.1074/jbc.M111.305599

Wang, Q., Hillwig, M. L., and Peters, R. J. (2011). CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J.* 65, 87–95. doi: 10.1111/j.1365-313X.2010.04408.x

Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell type-specific chromatin decondensation of a metabolic gene cluster in oat. *Plant Cell.* 21, 3926–3936. doi: 10.1105/tpc.109.072124

Wiemann, P., Guo, C. J., Palmer, J. M., Sekonyela, R., Wang, C. C., and Keller, N. P. (2013). Prototype of an intertwined secondary-metabolite supercluster. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17065–17070. doi: 10.1073/pnas.1313258110

Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., and Peters, R. J. (2004). Identification of syn-pimara-7,15-diene synthase reveals functional clustering

of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol.* 135, 2098–2105. doi: 10.1104/pp.104.045971

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., et al. (2015). Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* 349, 309–312. doi: 10.1126/science.aab1852

Wu, Y., Hillwig, M. L., Wang, Q., and Peters, R. J. (2011). Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* 585, 3446–3451. doi: 10.1016/j.febslet.2011.09.038

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the Genome Sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9, 949–952. doi: 10.1016/j.molp.2016.03.010

Yasumoto, S., Fukushima, E. O., Seki, H., and Muranaka, T. (2016). Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *FEBS Lett.* 590, 533–540. doi: 10.1002/1873-3468.12074

Yu, G., Li, C., Zhang, L., Zhu, G., Munir, S., Shi, C., et al. (2020). An allelic variant of GAME9 determines its binding capacity with the GAME17 promoter in the regulation of steroidal glycoalkaloid biosynthesis in tomato. *J. Exp. Bot.* 71, 2527–2536. doi: 10.1093/jxb/eraa014

Yu, J., Chang, P., Bhatnagar, D., and Cleveland, T. E. (2000). Cloning of a sugar utilization gene cluster in *Aspergillus parasiticus*. *Biochim. Biophys. Acta*. 1493, 211–214. doi: 10.1016/S0167-4781(00)00148-2

Yu, N., Nützmann, H. W., MacDonald, J. T., Moore, B., Field, B., Berriri, S., et al. (2016). Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* 44, 2255–2265. doi: 10.1093/nar/gkw100

Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X., et al. (2020). Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants*. 6, 1447–1454. doi: 10.1038/s41477-020-00816-7

Zhang, J., and Peters, R. J. (2020). Why are momilactones always associated with biosynthetic gene clusters in plants? *Proc. Natl. Acad. Sci. U.S.A.* 117, 13867–13869. doi: 10.1073/pnas.2007934117

Zhao, D., Hamilton, J. P., Bhat, W. W., Johnson, S. R., Godden, G. T., Kinser, T. J., et al. (2019). A chromosomal-scale genome assembly of Tectona grandis reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *Gigascience* 8:giz005. doi: 10.1093/gigascience/giz005

Zhao, M., Cheng, J., Guo, B., Duan, J., and Che, C. T. (2018). Momilactone and related diterpenoids as potential agricultural chemicals. *J. Agric. Food Chem.* 66, 7859–7872. doi: 10.1021/acs.jafc.8b02602

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

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in *Cucurbitaceae*. *Nat Plants* 2:16183. doi: 10.1038/nplants.2016.183

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fpls.2021.671487

Silencing of Oleuropein β-Glucosidase Abolishes the Biosynthetic Capacity of Secoiridoids in Olives

Konstantinos Koudounas[1†‡§], Margarita Thomopoulou[1†‡], Aimilia Rigakou[2†], Elisavet Angeli[1†], Eleni Melliou[2†], Prokopios Magiatis[2†] and Polydefkis Hatzopoulos[1†]*

[1] Laboratory of Molecular Biology, Department of Biotechnology, Agricultural University of Athens, Athens, Greece,
[2] Laboratory of Pharmacognosy and Natural Products Chemistry, Department of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece

Specialized metabolism is an evolutionary answer that fortifies plants against a wide spectrum of (a) biotic challenges. A plethora of diversified compounds can be found in the plant kingdom and often constitute the basis of human pharmacopeia. Olive trees (*Olea europaea*) produce an unusual type of secoiridoids known as oleosides with promising pharmaceutical activities. Here, we transiently silenced oleuropein β-glucosidase (OeGLU), an enzyme engaged in the biosynthetic pathway of secoiridoids in the olive trees. Reduction of *OeGLU* transcripts resulted in the absence of both upstream and downstream secoiridoids *in planta,* revealing a regulatory loop mechanism that bypasses the flux of precursor compounds toward the branch of secoiridoid biosynthesis. Our findings highlight that *OeGLU* could serve as a molecular target to regulate the bioactive secoiridoids in olive oils.

Keywords: *Olea europaea*, olive, secoiridoids, oleuropein, β-glucosidase, virus-induced gene silencing, tobacco rattle virus, one-dimensional quantitative nuclear magnetic resonance (1D-qNMR)

OPEN ACCESS

Edited by:
Thu Thuy Dang,
University of British Columbia Okanagan, Canada

Reviewed by:
Laura Perez Fons,
Royal Holloway, University of London, United Kingdom
Vonny Salim,
Louisiana State University Shreveport, United States
Yang Qu,
University of New Brunswick, Fredericton, Canada

****Correspondence:***
Polydefkis Hatzopoulos
phat@aua.gr

†ORCID:
Konstantinos Koudounas
orcid.org/0000-0002-6000-6565
Margarita Thomopoulou
orcid.org/0000-0002-4999-7257
Elisavet Angeli
orcid.org/0000-0003-4104-909X
Aimilia Rigakou
orcid.org/0000-0002-1034-7484
Eleni Melliou
orcid.org/0000-0003-1419-1752
Prokopios Magiatis
orcid.org/0000-0002-0399-5344
Polydefkis Hatzopoulos
orcid.org/0000-0002-0074-2552

‡ These authors have contributed equally to this work

§Present address:
Konstantinos Koudounas, EA2106 Biomolécules et Biotechnologies Végétales, Université de Tours, Tours, France

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

Received: *23 February 2021*
Accepted: *11 August 2021*
Published: *03 September 2021*

INTRODUCTION

Olive (*Olea europaea* L., Oleaceae) is an emblematic crop of the Mediterranean Basin and over the last decades, the cultivation has expanded to the Americas, Asia, and Oceania mostly due to the high nutritional value of olive oil. An essential trait that evolution has gifted to olives is the production of a certain type of secondary metabolites, known as oleosides, with an extremely narrow taxonomic distribution. Oleosides are terpene-derived secoiridoids, conjugates of glycosylated elenolic acid with a characteristic exocyclic 8,9-olefinic bond (Soler-Rivas et al., 2000), and are only present in the Oleaceae family and the genus *Caiophora* (Loasaceae) (Obied et al., 2008). Oleoside derivatives synergistically contribute to the beneficial aspects of olive oil in human health (Tripoli et al., 2005; Papanikolaou et al., 2019) and determine the flavor and quality of olive oil (Vitaglione et al., 2015). Recently, the European Medicinal Agency published a risk-benefit report on olives, and the European Food Safety Authority has already approved a health claim related to polyphenols in olive oil (EFSA, 2012; HMPC, 2017).

The dominant secoiridoid in olives accumulating up to 14% in young drupes is oleuropein (Amiot et al., 1986), an ester of elenolic acid with 2′-(3′,4′-dihydroxyphenyl)ethanol (hydroxytyrosol) that exhibits antioxidant, anti-inflammatory, antiproliferative, antimicrobial, and

antiviral activities, hence being a metabolite of high interest for humans (Bulotta et al., 2013). The first identified enzyme that is engaged in the metabolism of oleuropein is the oleuropein-specific β-glucosidase (OeGLU; E.C. 3.2.1.206) (Koudounas et al., 2015, 2017). OeGLU is a homomultimeric enzyme, member of the Glycoside Hydrolase 1 family (GH1), and localized in the nucleus. Oleuropein is localized in the vacuoles or cytosol (Konno et al., 1999), thus physically separated from the OeGLU enzyme, and upon cell disruption, the dual-partner defensive system comes in contact. Deglycosylation of oleuropein by OeGLU produces an unstable aglycone form that is rapidly converted to a highly reactive molecule with a glutaraldehyde-like structure. This compound covalently binds to amino acids and exhibits strong protein denaturing/cross-linking activities, providing a mighty arsenal against herbivores (Konno et al., 1999).

In addition to a pivotal role in the plant chemical defense of Oleaceae species, the OeGLU-mediated enzymatic detonation of oleuropein is also crucial during olive oil extraction. Crushing of olive drupes followed by malaxation (coalescence of oil droplets through the mixing of olive paste) causes cell disruption. Oleuropein is exposed to OeGLU, and the aglycone products are massively produced, thus shaping the organoleptic properties of olive oil (Obied et al., 2008; Romero-Segura et al., 2012; Hachicha Hbaieb et al., 2015; Vitaglione et al., 2015).

Apart from OeGLU, only three enzymatic hubs are known to be engaged in the biosynthesis and biotransformations of secoiridoids in olive. The first one is an iridoid synthase (OeISY) which converts 8-oxogeranial into the iridoid scaffold in a two-step reduction-cyclization sequence (Alagna et al., 2016). Recently, two bi-functional cytochrome P450s (oleoside methyl ester synthase, OeOMES and secoxyloganin synthase, OeSXS) which convert 7-*epi*-loganin into the secoiridoid scaffold in two sequential oxidation steps were characterized (Rodríguez-López et al., 2021). Additionally, two methylesterase enzymes (elenolic acid methylesterase 1, OeEAME1 and elenolic acid methylesterase 2, OeEAME2) that generate oleacein and oleocanthal after the concerted activity of β-glucosidase on oleuropein and ligstroside, respectively, have been identified (Volk et al., 2019). Finally, a *bona fide* geraniol synthase (OeGES1) has been reported to generate geraniol – monoterpene alcohol that is the precursor of iridoids (Vezzaro et al., 2012). Despite the recent advances toward the elucidation of this pathway, the proposed biosynthetic pathway of oleuropein in Oleaceae involves at least 18 enzymatic steps, therefore, remains largely uncharacterized (**Figure 1**).

The olive tree is a non-model plant, member of a plant family that comprises perennial woody species (Green, 2004) recalcitrant to genetic transformation. Therefore, studies on the Oleaceae-specific biosynthetic pathway of secoiridoids and characterization of the aforementioned enzymatic hubs have been limited to heterologous *in vivo* systems (i.e., *Nicotiana benthamiana*, *Saccharomyces cerevisiae*, and *Escherichia coli*) and *in vitro* enzymatic assays, thus raising the question of the actual contribution of these enzymes in olives. Plant genomes are known to typically encode more than 30 members of the GH1 family (Ketudat Cairns and Esen, 2010) therefore olive possess other β-glucosidases that could deglycosylate oleuropein. However, we have previously determined that *OeGLU* is a single copy gene in the olive genome (Koudounas et al., 2015), and defensive GH1 β-glucosidases are known to be highly diversified and exhibit exceptional specificity against their respective substrates (Verdoucq et al., 2004; Xia et al., 2012).

In this study, we recruited the recently described *Agrobacterium*-mediated virus-induced gene silencing (VIGS) methodology (Koudounas et al., 2020) to transiently silence *OeGLU* in olive seedlings and validate whether OeGLU is the major oleuropein β-glucosidase *in planta*. Besides, downregulation of the OeGLU enzyme, transient silencing of OeGLU unexpectedly affected the biosynthesis of upstream and downstream secoiridoids suggesting the existence of a feedback regulatory loop in this pathway.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Ripe fruits were harvested from *Olea europaea* L. cv. "Koroneiki" grown in a natural environment at the Agricultural University of Athens. The black mesocarps were removed, and the woody endocarps were subjected to 10% sodium hydroxide for 10 min and then thoroughly washed to remove any fleshy remnants. The endocarps were dried naturally and stored until use. Seeds were extracted after gently cracking the woody endocarp with a bench vice. After surface sterilization followed by stratification for 7 days at 4°C, seeds were transferred to soil for germination and grown at 22°C in a Fitotron growth chamber (Weiss Gallenkamp, Loughborough, United Kingdom) with a 16/8 h light/darkness cycle and 100 μmol m^{-2} s^{-1} light intensity.

Plasmid Construction

The pTRV1 (stock no.: CD3-1039) and pTRV2-MCS (stock no.: CD3-1040) plasmids encoding the bipartite RNA genome of tobacco rattle virus (TRV) were obtained from the Arabidopsis Biological Resource Center (ABRC[1]). RNA was extracted from olive young leaves using a phenol/chloroform procedure, and fragments of *OeGLU* (211 bp) and *OeChlH* (297 bp) were amplified by real-time polymerase chain reaction (RT-PCR) using the OeGLUi-F and OeGLUi-R, or OeChlH-F and OeChlH-R primers, respectively (**Supplementary Table 1**). To produce the pTRV2-OeGLU and pTRV2-OeChlH constructs, the PCR products were blunt-end ligated into the *Sma*I site of pUC19, and after verification by sequencing, the PCR products were directly sub-cloned into pTRV2-MCS (empty vector, EV) in antisense orientation relative to the TRV coat protein, utilizing the *Kpn*I and *Xba*I restriction sites of pUC19.

Agrobacterium-Mediated VIGS

The pTRV1 and pTRV2 constructs with either fragment of the targeted genes for silencing or an EV were electroporated in the *Agrobacterium tumefaciens* strain C58C1 Rif^R (GV3101) containing the T-DNA-deficient Ti plasmid pMP90. Positive transformants were harvested, and cells harboring the pTRV1

[1]www.arabidopsis.org

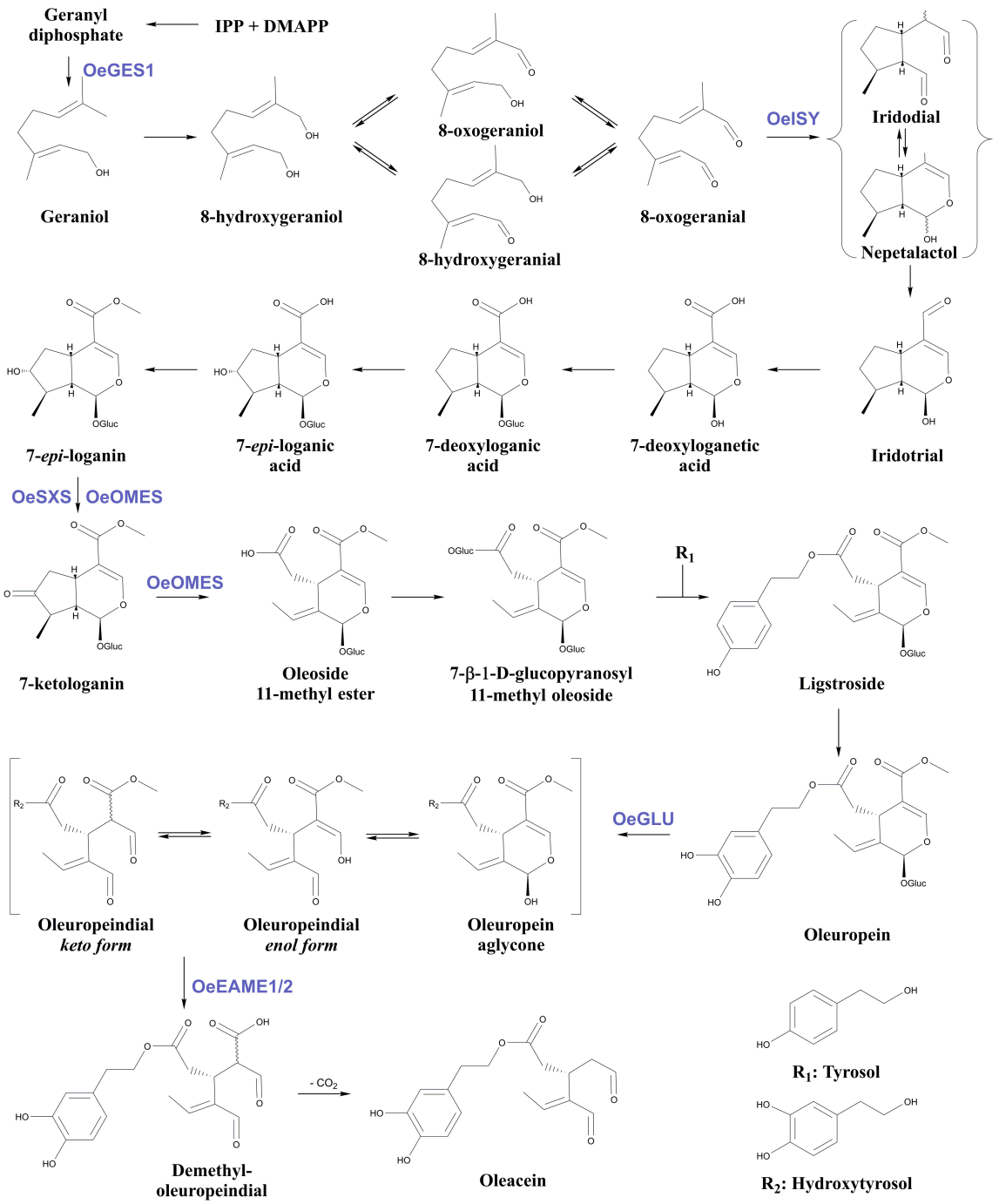


FIGURE 1 | Proposed biosynthetic pathway of oleuropein and oleacein in Oleaceae. Characterized enzymes of the pathway as referred in the literature (Obied et al., 2008; Vezzaro et al., 2012; Koudounas et al., 2015; Alagna et al., 2016; Volk et al., 2019; Rodríguez-López et al., 2021). IPP, isopentenyl diphosphate; DMAPP, Dimethylallyl diphosphate; Glu, glucose; OeGES1, geraniol synthase; OeISY, iridoid synthase; OeOMES, oleoside methyl ester synthase; OeSXS, secoxyloganin synthase; OeGLU, oleuropein β-glucosidase; OeEAME1/2, elenolic acid methylesterase 1 and 2.

plasmid were mixed 1:1 with cells harboring the pTRV2 constructs in inoculation medium (10 mM 2-N-morpholino-ethanesulfonic acid pH 5.6, 10 mM $MgCl_2$, and 150 μM acetosyringone). Olive plantlets having at least one pair of fully expanded leaves were Agroinoculated by gently pricking the leaves at the abaxial side as described (Koudounas et al., 2020). This procedure was repeated every 2 weeks until the observation of a phenotype. Leaves from the first two leaf pairs that emerged after Agroinoculation were harvested and used for molecular and biochemical studies. The presence of TRV in

non-Agroinoculated (i.e., newly emerged) leaves was validated by detecting the transcripts of TRV coat protein (**Supplementary Table 1**; Senthil-Kumar and Mysore, 2014).

Gene Expression Analysis

Reverse transcription was performed with SuperScript II RT (Invitrogen, Life Technologies, Carlsbad, CA, United States) using 200 ng of DNA-free RNA and the oligo(dT)$_{17}$ primer. Quantitative gene expression analysis was performed in a PikoReal 96 Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, United States) using the SYBR Select Master Mix (Applied Biosystems, Life Technologies, Carlsbad, CA, United States) and calculated by the ΔΔCt method. *OeGLU*, *OeGES1*, *OeISY*, *OeOMES*, *OeSXS*, and *OeEAME1/2* transcripts were amplified with the OeGLUq-F and OeGLUq-R, OeGES1q-F and OeGES1q-R, OeISYq-F and OeISYq-R, OeOMESq-F and OeOMESq-R, OeSXSq-F and OeSXSq-R, OeEAME1q-F and OeEAME1q-R, or OeEAME2q-F and OeEAME2q-R primers, respectively (**Supplementary Table 1**). Normalization of gene expression data was performed by using the *OeActin* housekeeping gene as a reference with primers OeActin_q-F and OeActin_q-R (**Supplementary Table 1**). Standard curves for both the target and the reference genes were generated to determine the amplification efficiency of each gene.

Enzymatic Assays

Harvested leaves were frozen in liquid nitrogen and ground into powder using a mortar and pestle. Crude proteins were extracted in ice-cold extraction buffer (100 mM Tris–HCl, pH 8.8, 100 mM EDTA, 1 mM PMSF, 100 mM KCL, 10 mM Na_2SO_3, and 100 mM glycine), samples were centrifuged, and the soluble fractions were quantified by the Bradford assay. The enzymatic reactions with oleuropein were performed by incubating 10 μg of the soluble protein extracts in 100-μl hydrolysis buffer (5 mM oleuropein, 150 mM sodium acetate, pH 5.5, and 0.05% bovine serum albumin [BSA]) at 37°C for 6 min. The deglycosylation degree of oleuropein was measured by high-performance liquid chromatography (HPLC) as described previously (Koudounas et al., 2015). The enzymatic reactions with *p*-nitrophenyl β-D-glucopyranoside (pNPGlu) were performed by incubating 10 μg of the soluble protein extracts in 500 μl hydrolysis buffer (10 mM pNPGlu, 150 mM sodium acetate, pH 5.5, and 0.05% BSA) at 37°C for 30 min, and reactions were stopped by mixing with an equal volume of 0.2 M sodium carbonate. The deglycosylated *p*-nitrophenyl was determined by measuring the absorbance at 405 nm. Relative activity was calculated by arbitrarily setting the oleuropein or pNPGlu β-glucosidase activity of soluble extracts from plants Agroinoculated with pTRV2-EV to 100%.

Chemicals and Standards

All solvents were of analytical grade and purchased from Merck. Syringaldehyde (98% purity) used as internal standard (IS) was purchased from Sigma-Aldrich (Steinheim, Germany). IS solution was prepared in acetonitrile at a concentration of 0.5 mg/mL and kept in a refrigerator. Prior to use, the IS solution was left to come to room temperature. Oleuropein was purchased from Extrasynthese (Genay, France), stock solution (15 mM) was prepared in ddH_2O, and aliquots were stored in −20°C until usage.

Extraction and NMR Spectra Analysis

Olive leaves were dried at room temperature in a dark place for a week. Then, the leaves were pulverized at room temperature, and 20 ml of MeOH were added to 100 mg of powdered dried olive leaves. The mixture was placed for 45 min in an ultrasonic bath followed by centrifugation at 4,000 rpm for 3 min. A portion of the methanolic extract was collected (10 mL), and 0.5 ml syringaldehyde in acetonitrile solution was added. The mixture was evaporated to dryness under vacuum in a rotary evaporator. The residue of the above procedure was dissolved in MeOD (600 μl), transferred to a 5 mm NMR tube, and the ^{1}H NMR was recorded at 400 MHz. Oleuropein was identified and quantitated by integrating the peak of proton at 5.9 ppm, and based on the equation y = 0.512x + 0.0904, the results were expressed per 100 mg olive leaf sample (Mousouri et al., 2014). NMR spectra were recorded on a DRX 400 MHz and analyzed with MestreNova. A total of 32 scans were collected into 32K data points over a spectral width of 0–16 ppm with a relaxation delay of 1 s and an acquisition time of 1.7 s. Prior to Fourier transformation (FT), an exponential weighting factor corresponding to a line broadening of 0.3 Hz was applied. The spectra were phase corrected, and accurate integration was performed manually for the peaks of interest.

Statistical Analysis

Three technical replicates from three biological replicates per treatment were used in real-time quantitative polymer chain reaction (RT-qPCR) analyses and *in vitro* enzymatic assays. Three or six biological replicates of plants Agroinoculated with pTRV2-EV or pTRV2-OeGLU, respectively, were used for NMR spectra analysis. Data were analyzed using the GraphPad Prism program (GraphPad Software). The statistical significance of differences between control and *OeGLU*-silenced samples was tested by unpaired Student's *t*-test.

Accession Numbers

Sequence data discussed in this article are available in GenBank under accession numbers: AY083162 (OeGLU), GABQ01080755 (OeChlH), GABQ01079399 (OeActin), AF406990 (pTRV1), AF406991 (pTRV2-MCS), JN408072 (OeGES1), KT954038 (OeISY), MT909123 (OeOMES), MT909125 (OeSXS), MK234850 (OeEAME1), and MK160486 (OeEAME2).

RESULTS

TRV-Mediated Silencing of *OeGLU*

To silence *OeGLU*, we first performed bioinformatic analysis to identify specific regions that would not trigger any off-target silencing (**Supplementary Figure 1**). cDNA fragments of either *OeGLU* or the H subunit of Mg-protoporphyrin chelatase (*OeChlH*) were cloned in plasmids encoding the bipartite RNA genome of TRV. ChlH is involved in the biosynthesis of chlorophyll and serves as a positive control

of the successful VIGS process (Senthil-Kumar and Mysore, 2014). Since the fragment used to trigger silencing of *OeChlH* had significant homology (**Supplementary Figure 2A**) with the corresponding cDNA from *Nicotiana benthamiana* (*NbChlH*), the efficiency of the construct to trigger silencing of *ChlH* was firstly validated in Agroinfiltrated tobacco plants. As expected, a yellowish leaf phenotype due to chlorophyll reduction was observed (**Supplementary Figure 2B**) indicating that *NbChlH* was successfully silenced. We next validated the efficiency of the TRV-based VIGS constructs by Agroinoculating olive seedlings. An intense yellowing phenotype indicative of successful silencing of the targeted gene in olive plants Agroinoculated with the *pTRV2-OeChlH* construct confirmed the successful application of the methodology (**Supplementary Figure 3**). Olive plants Agroinoculated with either the pTRV2-EV or the pTRV2-OeGLU constructs had no detectable phenotype (**Supplementary Figure 3**). Detection of the TRV coat protein transcripts in non-Agroinoculated leaves confirmed that TRV successfully infected the olive plants and moved systemically toward newly emerged leaves (**Supplementary Figure 4**).

To analyze the degree of silencing of *OeGLU*, we performed RT-qPCR analysis in olive young leaves, which clearly demonstrated that *OeGLU* was successfully silenced and the expression level was reduced by 80.82% compared to control conditions (**Figure 2A**). This result was further validated by semi-quantitative RT-PCR (**Figure 2B**). Additionally, we performed RT-qPCR analysis to monitor the expression levels of other characterized enzymes from the secoiridoid pathway in these plants. Silencing of *OeGLU* did not affect the transcript levels of either upstream (i.e., OeGES1, OeISY, and OeOMES) or downstream (i.e., OeEAME2) enzymes (**Supplementary Figure 5**).

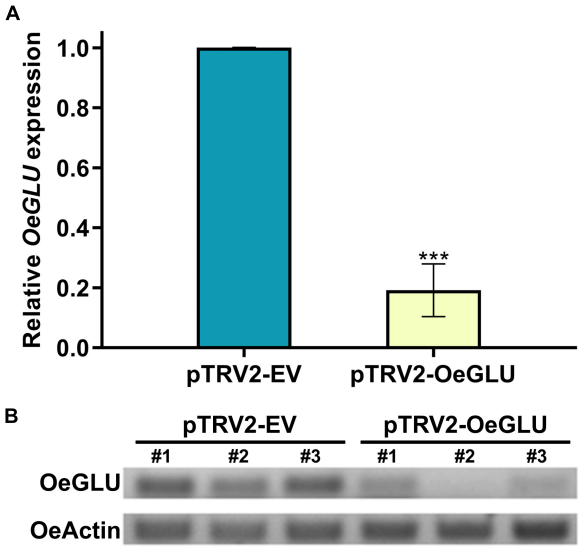


FIGURE 2 | VIGS-mediated silencing of *OeGLU* in olive plants. Real Time quantitative PCR (RT-PCR) **(A)** and semi-quantitative RT-PCR **(B)** analysis of *OeGLU* expression level of plants Agroinoculated with either pTRV2-EV or pTRV2-OeGLU constructs. OeActin was used as an internal control to normalize the samples. Three technical replicates of three biological replicates per treatment **(A)** or three biological replicates per treatment **(B)** were analyzed. Mean ± SE and asterisks denote statistical significance (***$P \leq 0.001$; Student's *t*-test). Numbers indicate biological replicates. EV, empty vector; OeGLU, oleuropein β-glucosidase.

Silencing of *OeGLU* Drastically Affects the Degree of Oleuropein Deglycosylation *in vitro*

We next investigated whether this biochemical phenotype would have an impact on the deglycosylation of oleuropein. HPLC analysis of the *in vitro* enzymatic assays using protein extracts from plants that *OeGLU* was silenced with oleuropein demonstrated that the relative *OeGLU* enzymatic activity was drastically reduced (**Figure 3A**). The degree of deglycosylation of oleuropein was reduced by 65.96% compared to control conditions after 6 min of incubation. As expected, this result coincides with the RT-qPCR (**Figure 2A**) and is in accordance with the abundance of *OeGLU* transcripts in the RNAi-silenced plants. The progress of the enzymatic deglycosylation of oleuropein revealed that the relative activity of OeGLU during the first 3 min of incubation was null, at 6 min was 30.09%, and even after 15 min of incubation, it remained at 33.93% compared to control (**Figure 3B**).

This result can be attributed to the fact that the VIGS-based approach of silencing genes varies and rarely reaches 100% effectiveness, as would be in a knockout mutant. Therefore, a number of remaining "escaping" transcripts were translated to the reduced quantity of OeGLU, nevertheless adequate to catalyze the enzymatic reaction to less extend. In addition, other non-specific β-glucosidases present in the olive tree that exhibit a lower affinity for oleuropein could potentially deglycosylate this substrate but within a worth noting delayed time lapse.

The specificity of silencing of *OeGLU* was validated by screening for the relative enzymatic activity of pNPGlu deglycosylation – a general synthetic substrate often used with β-glucosidases (**Figure 3C**). Since the relative non-specific β-glucosidase activity of the crude extracts remained at similar levels, the silencing of *OeGLU* was highly selective.

Silencing of *OeGLU* Abolishes the Biosynthesis of Both Upstream and Downstream Secoiridoids *in planta*

We next questioned whether the reduction of OeGLU would have an impact *in planta*. Instead of substrate accumulation, the content of the oleuropein secoiridoid in *OeGLU* silenced plants was lower compared to control non-silenced plants (**Figure 4** and **Supplementary Table 2**). In certain plants, oleuropein was even non-detectable as observed in the NMR spectra of the leaf extracts (**Figures 4**, **5A,B**; **Supplementary Figure 6**). In our experimental setup, the lowest concentration of oleuropein that can be quantified is 70 μg of oleuropein per 100 mg of dry weight (Mousouri et al., 2014), therefore in these plants the amount of oleuropein was reduced by at least 1,000 times compared to control plants. Additionally, the content of

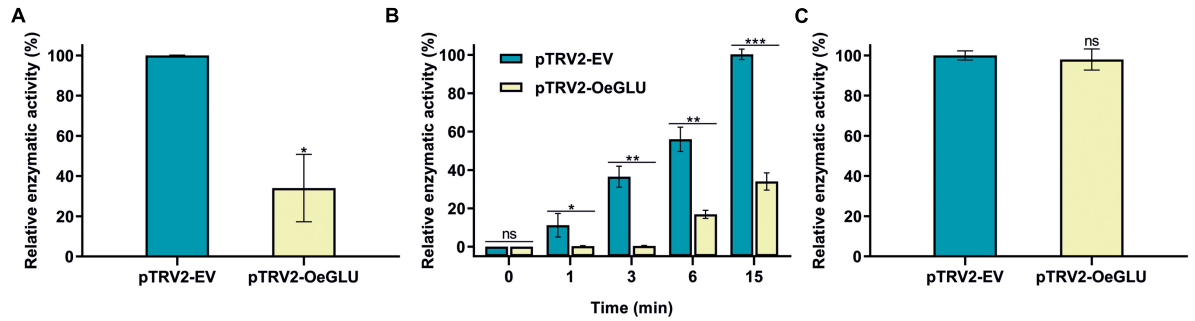


FIGURE 3 | *In vitro* enzymatic assays of crude leaf protein extracts of olives Agroinoculated with either pTRV2-EV or pTRV2-OeGLU constructs. **(A)** Relative enzymatic activity against oleuropein at 6 min. **(B)** Relative enzymatic activity against oleuropein at regular time intervals. **(C)** Relative enzymatic activity against pNPGlu. Three technical replicates of three biological replicates per treatment were analyzed. Mean ± SE and asterisks denote statistical significance (***$P \leq 0.001$, **$P \leq 0.01$, *$P \leq 0.05$, ns: not significant; Student's *t*-test). EV, empty vector; OeGLU, oleuropein β-glucosidase; pNPGlu, *p*-nitrophenyl β-D-glucopyranoside.

oleuropein aglycones and any related downstream secoiridoid were almost non-detectable. This result is in agreement with the low amount of the precursor oleuropein since the concerted activity of OeGLU on oleuropein followed by methylesterase activity results in a pool of aglycone derivatives (**Supplementary Figure 7**). It is worth noting that the existence of the oleuropein aglycones *in planta* (**Figure 5A**) has not been highlighted in the literature potentially pointing out a continuous, nevertheless at a low level, catabolic pathway of oleuropein. How is oleuropein transported from the vacuoles to the nucleus to be deglycosylated by OeGLU and especially how the cells cope with the highly reactive aglycones remain to be addressed.

In contrast, oleanolic acid or maslinic acid (triterpenic acids with characteristic peaks observed between 0.8 and 1.2 ppm), secondary compounds not related to the biosynthetic pathway of secoiridoids, were unaffected by the silencing of *OeGLU* (**Figure 6A**). The methyl group of oleuropein observed at 1.7 ppm and the characteristic H-1 at 5.90 ppm presented significant quantitative differences among the silenced and the control samples confirming that the effect *in planta* was apparently secoiridoid specific. No other peaks of related secoiridoids of any precursor molecule could be observed in the plants that *OeGLU* was silenced (**Figures 5A,B**, **6B**). Control Agroinoculated plants contained high amount of oleuropein, and a detectable amount of decarboxymethylated oleuropein aglycone products, which are observed in the aldehyde region (**Figures 4**, **5A,B**).

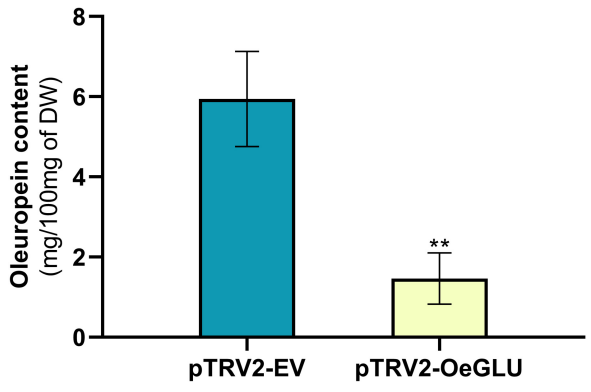


FIGURE 4 | Oleuropein content (mg) per 100 mg (dry weight) of olive leaves. Olive plantlets Agroinoculated either with pTRV2-EV ($n = 3$) or with pTRV2-OeGLU ($n = 6$) constructs were screened (**Supplementary Table 2**). Mean ± SE and asterisks denote statistical significance (**$P \leq 0.01$; Student's *t*-test). EV, empty vector; OeGLU, oleuropein β-glucosidase.

DISCUSSION

Plant chemical defense is often formed as a dual-partner system composed of glycosylated secondary compounds and dedicated detonating β-glucosidases ensuring that a reactive defensive aglycone is released only after deglycosylation (Morant et al., 2008; Piasecka et al., 2015). In Oleaceae species, the oleuropein/OeGLU system serves as a mighty chemical arsenal against herbivores (Konno et al., 1999) and this enzymatic reaction also determines the quality and flavor of olive oil (Romero-Segura et al., 2012; Hachicha Hbaieb et al., 2015; Vitaglione et al., 2015). Previous *in vitro* determination of enzymatic properties of OeGLU (Koudounas et al., 2015, 2017) urged us to functionally characterize this enzyme in the olive tree.

Molecular analyses of the Agroinoculated olive plants confirmed the successful VIGS-mediated silencing of *OeGLU*, and transcript abundance was reduced by more than 80%. VIGS processes, as all the RNA interference (RNAi) approaches, are known to result in a broad range of silencing (Meins et al., 2005), and comparable transcript reduction was observed in the woody tree *Jatropha curcas* using TRV-based constructs (Ye et al., 2009). *In vitro* enzymatic assays with oleuropein revealed that the relative deglycosylation activity of leaf protein extracts from *OeGLU*-silenced plants was drastically reduced compared to control plants and was arrested to the ratio of almost one-third independently of the incubation time of the enzymatic reactions. In contrast, enzymatic assays with pNPGlu confirmed that the relative non-specific β-glucosidase activity of the crude extracts remained at similar levels. In agreement with previously determined biochemical characteristics and kinetics of heterologously expressed OeGLU (Koudounas et al., 2017), these

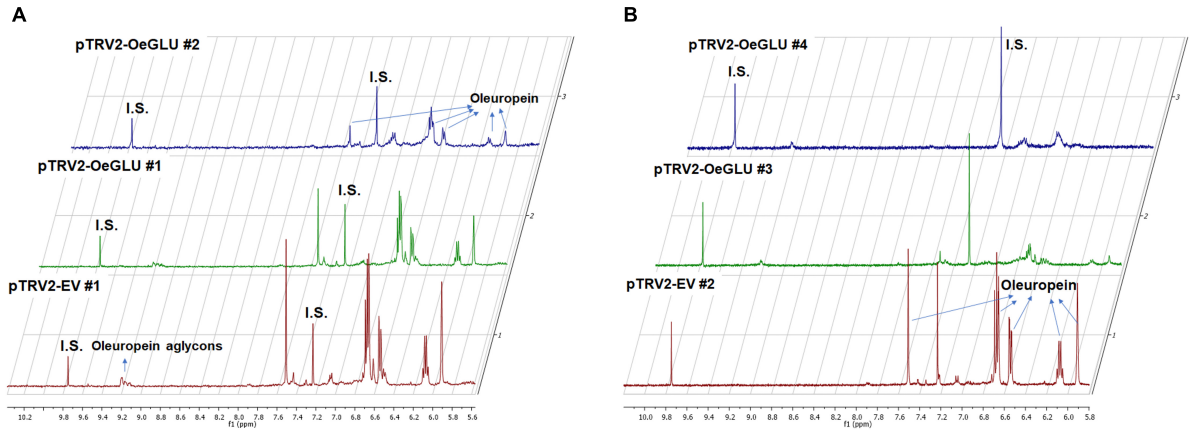


FIGURE 5 | Representative 1D-qNMR spectra of the *in planta* oleuropein content. Comparison between the methanolic extracts of olive plants Agroinoculated with pTRV2-EV or pTRV2-OeGLU constructs **(A,B)**. Numbers indicate biological replicates. EV, empty vector; IS, internal standard; OeGLU, oleuropein β-glucosidase; 1D qNMR, one-dimensional quantitative nuclear magnetic resonance.

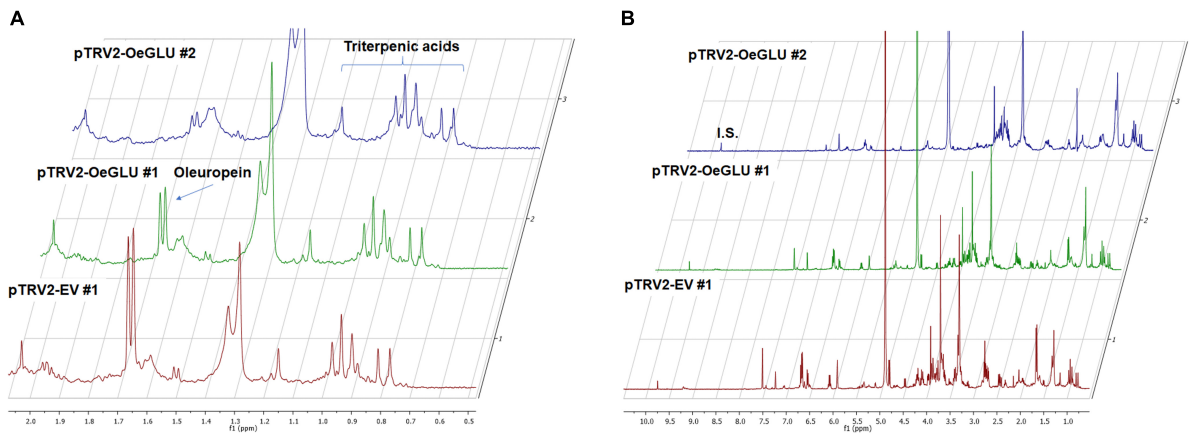


FIGURE 6 | Representative 1D-qNMR spectra. Comparison between the methanolic extracts of plants Agroinoculated with pTRV2-EV and pTRV2-OeGLU constructs for their triterpenic acids content **(A)** or their full spectrum **(B)**. Numbers indicate biological replicates. EV, empty vector; IS, internal standard; OeGLU, oleuropein β-glucosidase; 1D-qNMR, one-dimensional quantitative nuclear magnetic resonance.

results further confirm that OeGLU is the major oleuropein-specific β-glucosidase in olives since the deglycosylation activity of oleuropein was directly related to the abundance of *OeGLU* transcripts *in planta*. VIGS-mediated silencing is a powerful tool to study the biosynthetic pathway of secoiridoids in the olive tree, and this approach is expected to complement the functional characterization of other enzymatic hubs.

The *in planta* amount of oleuropein was found significantly lower in *OeGLU*-silenced olive plants compared to controls. The content of upstream and downstream secoiridoids was also reduced or even non-detectable revealing that reduction of *OeGLU* transcripts affected unexpectedly this pathway. The characteristic peak at 5.95 ppm that corresponds to H-1 of oleoside 11-methyl ester (Park et al., 1999; Mousouri et al., 2014), the first secoiridoid of the biosynthetic pathway of oleuropein (**Figure 1**), was non-detectable. Additionally, no differences between control and *OeGLU*-silenced olive plants were observed in the characteristic peaks of precursor iridoids (**Figure 1**), such as 7-deoxyloganic acid (5.20, 7.41 ppm) (Teng et al., 2005), 7-*epi*-loganic acid (5.38, 7.37 ppm) (Damtoft et al., 1993), 7-*epi*-loganin (5.33, 7.41 ppm) (Itoh et al., 2005), and 7-ketologanin (5.51, 7.39 ppm) (Kamoldinov et al., 2011); therefore, none of these compounds was accumulated in the *OeGLU*-silenced plants.

In contrast, the amount of characteristic triterpenoids was unaffected by *OeGLU* silencing. Although a possible effect in other classes of terpenoids cannot be excluded, the effect was apparently secoiridoid specific. These results strongly suggest that a regulatory loop mechanism controls the flux toward the branch of secoiridoids in olives.

It is worth noting that 7-β-1-D-glucopyranosyl-11-methyl oleoside (**Figure 1**) has been proposed to be an intermediate in the biosynthetic pathway of olive secoiridoids (Obied et al., 2008) only as the precursor of ligstroside but it is unknown whether it is deglycosylated back to oleoside 11-methyl ester

(Damtoft et al., 1993). This compound is found in trace amounts in olive leaves (De Nino et al., 1999) and has not been tested as a potential substrate of OeGLU since it is not commercially available. Even if OeGLU could potentially deglycosylate this diglucoside, the absence of the characteristic peak of its direct precursor – oleoside 11-methyl ester (**Figure 1**) – is an unexpected result.

In our experimental setup, olive leaves were dried at room temperature for a week prior to extraction and NMR spectra analysis. Although all secoiridoids and secoiridoid glucosides are not volatile compounds, this could potentially have an impact on volatile compounds found at the very early steps of the biosynthetic pathway before iridodial (**Figure 1**), like for example in the case of 8-hydroxygeraniol. On the other hand, no differences could be observed in the early iridoids among the samples, therefore most likely the biosynthesis of iridoids was supplied with comparable quantities of precursor molecules.

Among the characterized dual-partner defense systems, the absence of one partner typically does not affect the other. For example, cyanogenic glucosides are biosynthesized in barley (*Hordeum vulgare*) leaves despite the lack of co-localized cyanide releasing β-glucosidase and transient overexpression of a cyanogenic β-glucosidase from sorghum (*Sorghum bicolor*) reconstitutes cyanogenesis in barley leaves (Nielsen et al., 2006). In white clover (*Trifolium repens*) either both cyanogenic glucosides and the respective β-glucosidase or one or even none of the two partners are present (Olsen et al., 2007, 2008). In the glucosinolate-myrosinase defense system of Brassicaceae, knockout mutations of two functional myrosinases (TGG1 and TGG2) in *Arabidopsis thaliana* resulted in slightly higher glucosinolate content in certain developmental stages (Barth and Jander, 2006).

One example of unexpected phenotype was recently reported in the medicinal plant *Catharanthus roseus* after unbalancing the metabolic flux of strictosidine biosynthesis – a monoterpene indole alkaloid closely related to oleuropein with similar protein-crosslinking activity after deglycosylation (Guirimand et al., 2010; Sudžuković et al., 2016). VIGS-mediated silencing of either strictosidine β-glucosidase (Carqueijeiro et al., 2021) or the vacuolar exporter of strictosidine (Payne et al., 2017) resulted in necrotic symptoms. Possibly olives have evolved a regulatory framework within the secoiridoid pathway to avoid any substantial accumulation of cytotoxic intermediates that eventually could result in cell death.

A growing body of evidence highlights that the secondary metabolism may regulate plant defense responses which suggests the existence of synergies among distinct biosynthetic pathways (Erb and Kliebenstein, 2020). Mutants of enzymes involved in the biosynthesis of glucosinolates in *Arabidopsis* may affect the pathogen-triggered callose regulation (Clay et al., 2009), the biosynthesis of phenylpropanoids (Kim et al., 2020), hormonal signaling (Burow et al., 2015), or biosynthesis of other tryptophan-derived compounds (Frerigmann et al., 2016). Similar regulatory cross-talks exist in the biosynthesis of benzoxazinoids (Ahmad et al., 2011; Meihls et al., 2013; Li et al., 2018). Even though, feedback or feedforward regulation among specialized metabolism and distinct biosynthetic pathways was observed, in most cases, the underlying molecular etiology remains elusive.

Reports about regulatory loop mechanisms that self-govern a specialized biosynthetic pathway in plants are scarce. Identified examples include epistasis in carotenoid biosynthesis (Kachanovsky et al., 2012) and feedback regulation through intermediates in phenylpropanoids (Blount et al., 2000). To the best of our knowledge, the only known example partially resembling the phenotype observed in *OeGLU*-silenced olive plants has been reported in opium poppy (*Papaver somniferum*). Silencing of codeinone reductase (COR) resulted in accumulation of (*S*)-reticuline, an intermediate compound upstream of seven enzymatic steps proposing a feedback regulatory loop mechanism (Allen et al., 2004) and preventing intermediates of benzylisoquinoline synthesis to enter the morphine-specific branch. In agreement with the absence of any transcriptional effect in other enzymes engaged in the biosynthetic pathway of secoiridoids after silencing *OeGLU* in olives, silencing of *COR* did not affect the abundance of transcripts encoding for enzymes engaged upstream and downstream of (*S*)-reticuline (Allen et al., 2004). Although oleosides- and morphinan-type alkaloids are structurally and biosynthetically different, a fundamental regulatory mechanism could govern both secondary metabolic pathways. The subcellular compartmentalization of the secoiridoid pathway in Oleaceae engaging at least two physically separated compartments at the last enzymatic steps (i.e., vacuole and nucleus) (Koudounas et al., 2017) points out retrograde signaling and a sophisticated cross-talk within the cell.

Biosynthesis of specialized metabolites has energetic and metabolic costs for the plants (Gershenzon, 1994; Neilson et al., 2013; Cipollini et al., 2018). Possibly the absence of OeGLU *in planta* is perceived as a failure in the defensive pathway of secoiridoids and olives completely bypass the flux toward oleuropein to recoup the costs and avoid investing resources in the absence of defensive benefits. The molecular mechanism(s) by which olive cells sense the loss of a key enzymatic hub and bypass the flux toward the biosynthetic pathway of secoiridoids remain to be addressed. Although no difference was observed at the transcript levels of the selected genes engaged in this pathway, possibly the expression levels of other regulatory genes, such as transcription factors or specialized transporters, may have been altered after the silencing of *OeGLU*, and a comparative transcriptomic approach is expected to shed light on the etiology of the observed phenotype. Nonetheless, this study highlights an unexpected direct link between OeGLU and total content of secoiridoids, thus this enzyme could serve as a molecular target of high biotechnological interest in order to produce tailor-made olive oils with adjustable secoiridoid content having an optimized balance of flavor (Vitaglione et al., 2015) and high nutritional value with beneficial aspects in human health.

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

KK and PH designed the research. KK, MT, and EA prepared the constructs, performed VIGS experiments, analyzed gene expressions, and performed *in vitro* enzymatic assays. EM and PM performed method development for qualitative and quantitative analyses of olive leaf ingredients by NMR. AR analyzed application of NMR methodology and performed data collection. KK, EM, PM, and PH wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

PH acknowledges funding from the General Secretariat of Research and Technology, Greece, the National Emblematic Action "Olive Routes," and the European Regional Development Fund of the European Union and Greek National Funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE ("ELIADA," project code: T2EΔK-01315). MT acknowledges funding from Greece and the European Union (European Social Fund, ESF) through the Operational Program "Human Resources Development, Education and Lifelong Learning" in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY). AR acknowledges scholarship from World Olive Center for Health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.671487/full#supplementary-material

REFERENCES

Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y., Martin, J., Smart, L., et al. (2011). Benzoxazinoid Metabolites Regulate Innate Immunity against Aphids and Fungi in Maize. *Plant Physiol.* 157, 317–327. doi: 10.1104/pp.111.180224

Alagna, F., Geu-Flores, F., Kries, H., Panara, F., Baldoni, L., O'Connor, S. E., et al. (2016). Identification and characterization of the iridoid synthase involved in oleuropein biosynthesis in olive (*Olea europaea*) fruits. *J. Biol. Chem.* 291, 5542–5554. doi: 10.1074/jbc.M115.701276

Allen, R. S., Millgate, A. G., Chitty, J. A., Thisleton, J., Miller, J. A. C., Fist, A. J., et al. (2004). RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. *Nat. Biotechnol.* 22, 1559–1566. doi: 10.1038/nbt1033

Amiot, M. J., Fleuriet, A., and Macheix, J. J. (1986). Importance and evolution of phenolic compounds in olive during growth and maturation. *J. Agric. Food Chem.* 34, 823–826. doi: 10.1021/jf00071a014

Barth, C., and Jander, G. (2006). Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* 46, 549–562. doi: 10.1111/j.1365-313X.2006.02716.x

Blount, J. W., Korth, K. L., Masoud, S. A., Rasmussen, S., Lamb, C., and Dixon, R. A. (2000). Altering Expression of Cinnamic Acid 4-Hydroxylase in Transgenic Plants Provides Evidence for a Feedback Loop at the Entry Point into the Phenylpropanoid Pathway. *Plant Physiol.* 122, 107–116. doi: 10.1104/pp.122.1.107

Bulotta, S., Oliverio, M., Russo, D., and Procopio, A. (2013). "Biological activity of oleuropein and its derivatives," in *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*, eds K. G. Ramawat and J.-M. Mérillon (Berlin: Springer), 3605–3638.

Burow, M., Atwell, S., Francisco, M., Kerwin, Rachel, E., Halkier, et al. (2015). The Glucosinolate Biosynthetic Gene AOP2 Mediates Feed-back Regulation of Jasmonic Acid Signaling in Arabidopsis. *Mol. Plant* 8, 1201–1212. doi: 10.1016/j.molp.2015.03.001

Carqueijeiro, I., Koudounas, K., Dugé de Bernonville, T., Sepúlveda, L. J., Mosquera, A., Bomzan, D. P., et al. (2021). Alternative splicing creates a pseudo-strictosidine β-D-glucosidase modulating alkaloid synthesis in Catharanthus roseus. *Plant Physiol.* 185, 836–856. doi: 10.1093/plphys/kiaa075

Cipollini, D., Walters, D., and Voelckel, C. (2018). "Costs of Resistance in Plants: from Theory to Evidence," in *Annual Plant Reviews Online*, ed. J. A. Roberts (Hoboken: Wiley), 263–307.

Clay, N. K., Adio, A. M., Denoux, C., Jander, G., and Ausubel, F. M. (2009). Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. *Science* 323, 95–101. doi: 10.1126/science.1164627

Damtoft, S., Franzyk, H., and Jensen, S. R. (1993). Biosynthesis of secoiridoid glucosides in Oleaceae. *Phytochemistry* 34, 1291–1299. doi: 10.1016/0031-9422(91)80018-V

De Nino, A., Mazzotti, F., Morrone, S. P., Perri, E., Raffaelli, A., and Sindona, G. (1999). Characterization of cassanese olive cultivar through the identification of new trace components by ionspray tandem mass spectrometry. *J. Mass Spectrometry* 34, 10–16. doi: 10.1002/(SICI)1096-9888(199901)34:1<10::AID-JMS744<3.0.CO;2-X

EFSA. (2012). Scientific Opinion on the substantiation of a health claim related to polyphenols in olive and maintenance of normal blood HDL cholesterol concentrations (ID 1639, further assessment) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. *EFSA J.* 10:2848. doi: 10.2903/j.efsa.2012.2848

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

Frerigmann, H., Piślewska-Bednarek, M., Sánchez-Vallet, A., Molina, A., Glawischnig, E., Gigolashvili, T., et al. (2016). Regulation of Pathogen-Triggered Tryptophan Metabolism in Arabidopsis thaliana by MYB Transcription Factors and Indole Glucosinolate Conversion Products. *Mol. Plant* 9, 682–695. doi: 10.1016/j.molp.2016.01.006

Gershenzon, J. (1994). Metabolic costs of terpenoid accumulation in higher plants. *J. Chem. Ecol.* 20, 1281–1328. doi: 10.1007/BF02059810

Green, P. S. (2004). "Oleaceae," in *Flowering Plants · Dicotyledons: Lamiales (except Acanthaceae including Avicenniaceae)*, ed. J. W. Kadereit (Berlin: Springer), 296–306.

Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., et al. (2010). Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.* 10:182. doi: 10.1186/1471-2229-10-182

Hachicha Hbaieb, R., Kotti, F., García-Rodríguez, R., Gargouri, M., Sanz, C., and Pérez, A. G. (2015). Monitoring endogenous enzymes during olive fruit ripening and storage: correlation with virgin olive oil phenolic profiles. *Food Chem.* 174, 240–247. doi: 10.1016/j.foodchem.2014.11.033

HMPC. (2017). *Final Assessment Report on Olea europaea L., Folium - First Version*. Amsterdam: European Medicines Agency.

Itoh, A., Kumashiro, T., Yamaguchi, M., Nagakura, N., Mizushina, Y., Nishi, T., et al. (2005). Indole Alkaloids and Other Constituents of *Rauwolfia serpentina*. *J. Nat. Products* 68, 848–852. doi: 10.1021/np058007n

Kachanovsky, D. E., Filler, S., Isaacson, T., and Hirschberg, J. (2012). Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *Proc. Nat. Acad. Sci. U. S. A.* 109, 19021–19026. doi: 10.1073/pnas.1214808109

Kamoldinov, K. S., Eshbakova, K. A., Bobakulov, K. M., and Abdullaev, N. D. (2011). Components of Fraxinus raibocarpa. *Chem. Nat. Compounds* 47, 448–449. doi: 10.1007/s10600-011-9958-5

Ketudat Cairns, J. R., and Esen, A. (2010). β-Glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2

Kim, J. I., Zhang, X., Pascuzzi, P. E., Liu, C.-J., and Chapple, C. (2020). Glucosinolate and phenylpropanoid biosynthesis are linked by proteasome-dependent degradation of PAL. *New Phytologist* 225, 154–168. doi: 10.1111/nph.16108

Konno, K., Hirayama, C., Yasui, H., and Nakamura, M. (1999). Enzymatic activation of oleuropein: a protein crosslinker used as a chemical defense in the privet tree. *Proc. Nat. Acad. Sci. U. S. A.* 96, 9159–9164. doi: 10.1073/pnas.96.16.9159

Koudounas, K., Banilas, G., Michaelidis, C., Demoliou, C., Rigas, S., and Hatzopoulos, P. (2015). A defence-related *Olea europaea* β-glucosidase hydrolyses and activates oleuropein into a potent protein cross-linking agent. *J. Exp. Bot.* 66, 2093–2106. doi: 10.1093/jxb/erv002

Koudounas, K., Thomopoulou, M., Angeli, E., Tsitsekian, D., Rigas, S., and Hatzopoulos, P. (2020). "Virus-Induced Gene Silencing in Olive Tree (Oleaceae)," in *Virus-Induced Gene Silencing in Plants: Methods and Protocols*, eds V. Courdavault and S. Besseau (New York: Springer), 165–182.

Koudounas, K., Thomopoulou, M., Michaelidis, C., Zevgiti, E., Papakostas, G., Tserou, P., et al. (2017). The C-Domain of Oleuropein β-Glucosidase Assists in Protein Folding and Sequesters the Enzyme in Nucleus. *Plant Physiol.* 174, 1371–1383. doi: 10.1104/pp.17.00512

Li, B., Förster, C., Robert, C. A. M., Züst, T., Hu, L., Machado, R. A. R., et al. (2018). Convergent evolution of a metabolic switch between aphid and caterpillar resistance in cereals. *Sci. Adv.* 4:eaat6797. doi: 10.1126/sciadv.aat6797

Meihls, L. N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M. M., et al. (2013). Natural Variation in Maize Aphid Resistance Is Associated with 2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-One Glucoside Methyltransferase Activity. *Plant Cell* 25, 2341–2355. doi: 10.1105/tpc.113.112409

Meins, F., Si-Ammour, A., and Blevins, T. (2005). RNA silencing systems and their relevance to plant development. *Ann. Rev. Cell Dev. Biol.* 21, 297–318. doi: 10.1146/annurev.cellbio.21.122303.114706

Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., et al. (2008). β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69, 1795–1813. doi: 10.1016/j.phytochem.2008.03.006

Mousouri, E., Melliou, E., and Magiatis, P. (2014). Isolation of Megaritolactones and Other Bioactive Metabolites from 'Megaritiki' Table Olives and Debittering Water. *J. Agric. Food Chem.* 62, 660–667. doi: 10.1021/jf404685h

Neilson, E. H., Goodger, J. Q. D., Woodrow, I. E., and Møller, B. L. (2013). Plant chemical defense: at what cost? *Trends Plant Sci.* 18, 250–258. doi: 10.1016/j.tplants.2013.01.001

Nielsen, K., Hrmova, M., Nielsen, J., Forslund, K., Ebert, S., Olsen, C., et al. (2006). Reconstitution of cyanogenesis in barley (*Hordeum vulgare* L.) and its implications for resistance against the barley powdery mildew fungus. *Planta* 223, 1010–1023. doi: 10.1007/s00425-005-0158-z

Obied, H. K., Prenzler, P. D., Ryan, D., Servili, M., Taticchi, A., Esposto, S., et al. (2008). Biosynthesis and biotransformations of phenol-conjugated oleosidic secoiridoids from *Olea europaea* L. *Nat. Prod. Rep.* 25, 1167–1179. doi: 10.1039/b719736e

Olsen, K. M., Hsu, S.-C., and Small, L. L. (2008). Evidence on the Molecular Basis of the Ac/ac Adaptive Cyanogenesis Polymorphism in White Clover (*Trifolium repens* L.). *Genetics* 179, 517–526. doi: 10.1534/genetics.107.080366

Olsen, K. M., Sutherland, B. L., and Small, L. L. (2007). Molecular evolution of the Li/li chemical defence polymorphism in white clover (*Trifolium repens* L.). *Mol. Ecol.* 16, 4180–4193. doi: 10.1111/j.1365-294X.2007.03506.x

Papanikolaou, C., Melliou, E., and Magiatis, P. (2019). "Olive Oil Phenols," in *Functional Foods*, ed. V. Lagouri. (London: IntechOpen), 29–46.

Park, H. J., Lee, M. S., Lee, K. T., Sohn, I. C., Han, Y. N., and Miyamoto, K. I. (1999). Studies on constituents with cytotoxic activity from the stem bark of Syringa velutina. *Chem. Pharmaceutical Bull.* 47, 1029–1031. doi: 10.1248/cpb.47.1029

Payne, R. M. E., Xu, D., Foureau, E., Teto Carqueijeiro, M. I. S., Oudin, A., Bernonville, T. D. D., et al. (2017). An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat. Plants* 3:16208. doi: 10.1038/nplants.2016.208

Piasecka, A., Jedrzejczak-Rey, N., and Bednarek, P. (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol.* 206, 948–964. doi: 10.1111/nph.13325

Rodríguez-López, C. E., Hong, B., Paetz, C., Nakamura, Y., Koudounas, K., Passeri, V., et al. (2021). Two bi-functional cytochrome P450 CYP72 enzymes from olive (*Olea europaea*) catalyze the oxidative C-C bond cleavage in the biosynthesis of secoxy-iridoids – flavor and quality determinants in olive oil. *New Phytol.* 229, 2288–2301. doi: 10.1111/nph.16975

Romero-Segura, C., García-Rodríguez, R., Sánchez-Ortiz, A., Sanz, C., and Pérez, A. G. (2012). The role of olive β-glucosidase in shaping the phenolic profile of virgin olive oil. *Food Res. Int.* 45, 191–196. doi: 10.1016/j.foodres.2011.10.024

Senthil-Kumar, M., and Mysore, K. S. (2014). Tobacco rattle virus–based virus-induced gene silencing in *Nicotiana benthamiana*. *Nat. Protoc.* 9, 1549–1562. doi: 10.1038/nprot.2014.092

Soler-Rivas, C., Espín, J. C., and Wichers, H. J. (2000). Oleuropein and related compounds. *J. Sci. Food Agric.* 80, 1013–1023. doi: 10.1002/(SICI)1097-0010(20000515)80:7<1013::AID-JSFA571<3.0.CO;2-C

Sudžuković, N., Schinnerl, J., and Brecker, L. (2016). Phytochemical meanings of tetrahydro-β-carboline moiety in strictosidine derivatives. *Bioorg. Med. Chem.* 24, 588–595. doi: 10.1016/j.bmc.2015.12.028

Teng, R. W., Wang, D. Z., Wu, Y. S., Lu, Y., Zheng, Q. T., and Yang, C. R. (2005). NMR assignments and single-crystal X-ray diffraction analysis of deoxyloganic acid. *Magn. Reson. Chem.* 43, 92–96. doi: 10.1002/mrc.1502

Tripoli, E., Giammanco, M., Tabacchi, G., Di Majo, D., Giammanco, S., and La Guardia, M. (2005). The phenolic compounds of olive oil: structure, biological activity and beneficial effects on human health. *Nutr. Res. Rev.* 18, 98–112. doi: 10.1079/NRR200495

Verdoucq, L., Moriniere, J., Bevan, D. R., Esen, A., Vasella, A., Henrissat, B., et al. (2004). Structural determinants of substrate specificity in family 1 β-glucosidases: novel insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with strict specificity, in complex with its natural substrate. *J. Biol. Chem.* 279, 31796–31803. doi: 10.1074/jbc.M402918200

Vezzaro, A., Krause, S. T., Nonis, A., Ramina, A., Degenhardt, J., and Ruperti, B. (2012). Isolation and characterization of terpene synthases potentially involved in flavor development of ripening olive (*Olea europaea*) fruits. *J. Plant Physiol.* 169, 908–914. doi: 10.1016/j.jplph.2012.01.021

Vitaglione, P., Savarese, M., Paduano, A., Scalfi, L., Fogliano, V., and Sacchi, R. (2015). Healthy Virgin Olive Oil: a Matter of Bitterness. *Crit. Rev. Food Sci. Nutr.* 55, 1808–1818. doi: 10.1080/10408398.2012.708685

Volk, J., Sarafeddinov, A., Unver, T., Marx, S., Tretzel, J., Zotzel, J., et al. (2019). Two novel methylesterases from *Olea europaea* contribute to the catabolism of oleoside-type secoiridoid esters. *Planta* 250, 2083–2097. doi: 10.1007/s00425-019-03286-0

Xia, L., Ruppert, M., Wang, M., Panjikar, S., Lin, H., Rajendran, C., et al. (2012). Structures of alkaloid biosynthetic glucosidases decode substrate specificity. *ACS Chem. Biol.* 7, 226–234. doi: 10.1021/cb200267w

Ye, J., Qu, J., Bui, H. T. N., and Chua, N. H. (2009). Rapid analysis of *Jatropha curcas* gene functions by virus-induced gene silencing. *Plant Biotechnol. J.* 7, 964–976. doi: 10.1111/j.1467-7652.2009.00457.x

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

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

Citation: Koudounas K, Thomopoulou M, Rigakou A, Angeli E, Melliou E, Magiatis P and Hatzopoulos P (2021) Silencing of Oleuropein β-Glucosidase Abolishes the Biosynthetic Capacity of Secoiridoids in Olives. Front. Plant Sci. 12:671487. doi: 10.3389/fpls.2021.671487

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us:frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership