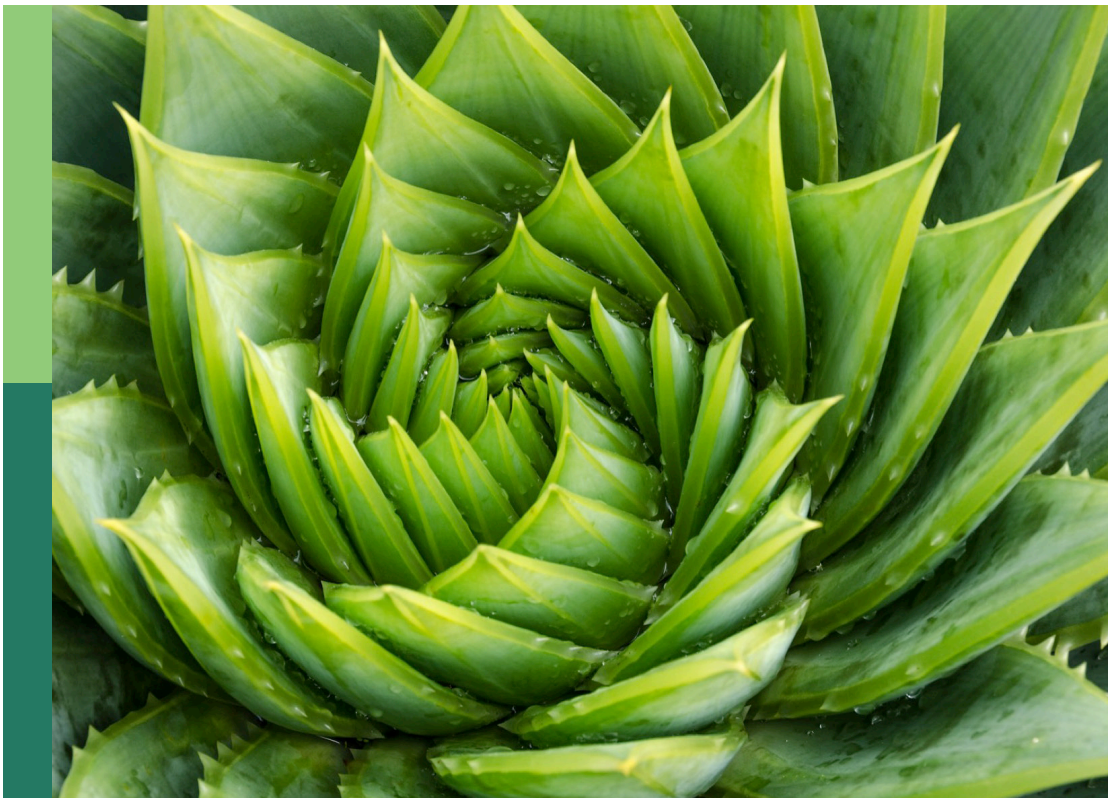
frontiers | Research Topics

Insights in plant metabolism and chemodiversity: 2021

Edited by
Laigeng Li

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83251-328-6
DOI 10.3389/978-2-83251-328-6

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

Insights in plant metabolism and chemodiversity: 2021

Topic editor
Laigeng Li — Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences (CAS), China

Citation
Li, L., ed. (2023). *Insights in plant metabolism and chemodiversity: 2021*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-328-6

Table of contents

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fpls.2022.827453

Using CRISPR-Cas9 Technology to Eliminate Xyloglucan in Tobacco Cell Walls and Change the Uptake and Translocation of Inorganic Arsenic

Meng Wang[1†], Xinxin Song[1†], Shuaiqiang Guo[1], Peiyao Li[1], Zongchang Xu[2], Hua Xu[3], Anming Ding[2], Rana Imtiaz Ahmed[2], Gongke Zhou[4,5], Malcom O'Neill[6], Dahai Yang[7] and Yingzhen Kong[1]**

[1]College of Agronomy, Qingdao Agricultural University, Qingdao, China, [2]Key Laboratory of Tobacco Gene Resources, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, China, [3]Key Laboratory of Biofuels, Qingdao Engineering Research Center of Biomass Resources and Environment, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, [4]College of Resources and Environment, Qingdao Agricultural University, Qingdao, China, [5]Academy of Dongying Efficient Agricultural Technology and Industry on Saline and Alkaline Land in Collaboration With Qingdao Agricultural University, Dongying, China, [6]Complex Carbohydrate Research Center, University of Georgia, Athens, GA, United States, [7]China Tobacco Breeding and Biotechnology Research Center, Yunnan Academy of Tobacco Agricultural Sciences, Kunming, China

Xyloglucan is a quantitatively major polysaccharide in the primary cell walls of flowering plants and has been reported to affect plants' ability to tolerate toxic elements. However, it is not known if altering the amounts of xyloglucan in the wall influences the uptake and translocation of inorganic arsenic (As). Here, we identified two *Nicotiana tabacum* genes that encode xyloglucan-specific xylosyltransferases (XXT), which we named NtXXT1 and NtXXT2. We used CRISPR-Cas9 technology to generate *ntxxt1*, *ntxxt2*, and *ntxxt1/2* mutant tobacco plants to determine if preventing xyloglucan synthesis affects plant growth and their ability to accumulate As. We show that NtXXT1 and NtXXT2 are required for xyloglucan biosynthesis because no discernible amounts of xyloglucan were present in the cell walls of the *ntxxt1/2* double mutant. The tobacco double mutant (*ntxxt1/2*) and the corresponding Arabidopsis mutant (*atxxt1/2*) do not have severe growth defects but do have a short root hair phenotype and a slow growth rate. This phenotype is rescued by overexpressing *NtXXT1* or *NtXXT2* in *atxxt1/2*. Growing *ntxxt* mutants in the presence of AsIII or AsV showed that the absence of cell wall xyloglucan affects the accumulation and translocation of As. Most notably, root retention of As increased substantially and the amounts of As translocated to the shoots decreased in *ntxxt1/2*. Our results suggest that xyloglucan-deficient plants provide a strategy for the phytoremediation of As contaminated soils.

Keywords: cell wall, xyloglucan xylosyltransferases, CRISPR-Cas9, arsenic, uptake

OPEN ACCESS

Edited by:
Laigeng Li,
Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences (CAS), China

Reviewed by:
Markus Pauly,
Heinrich Heine University of Düsseldorf, Germany
Baocai Zhang,
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Dahai Yang
bioresearch2013@126.com
Yingzhen Kong
kongyingzhen@qau.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: *02 December 2021*
Accepted: *21 January 2022*
Published: *16 February 2022*

Citation:
Wang M, Song X, Guo S, Li P, Xu Z, Xu H, Ding A, Ahmed RI, Zhou G, O'Neill M, Yang D and Kong Y (2022) Using CRISPR-Cas9 Technology to Eliminate Xyloglucan in Tobacco Cell Walls and Change the Uptake and Translocation of Inorganic Arsenic. Front. Plant Sci. 13:827453. doi: 10.3389/fpls.2022.827453

INTRODUCTION

The polysaccharide-rich cell wall is plants' first line of defense against toxic elements present in soils and water as it may prevent them from entering the cytoplasm (Parrotta et al., 2015). For example, flax hypocotyls adapt to the "b" class soft metal cadmium (Cd^{2+}) by changing the methyl-esterification pattern of homogalacturonan (Douchiche et al., 2010). Increases in pectin and hemicellulose in rice leaves have been reported to be responsible for increased Cd^{2+} accumulation in root cell walls and a decrease in soluble Cd^{2+} (Xiong et al., 2009). However, a decrease of pectin and hemicelluloses resulting from phosphorous deficiency has also been reported to enhance Cd^{2+} exclusion in Arabidopsis root walls (Zhu et al., 2012a). Cd^{2+} tolerance in tobacco has been increased by overexpressing a *Populus euphratica* gene encoding a xyloglucan (XyG) endotransglucosylase/hydrolase (XTH) to decrease the amount of wall XyG (Han et al., 2014). A decrease of XyG (*xth31* mutant; Han et al., 2014) or the absence of XyG (*xxt1/2* double mutant; Zhu et al., 2012b) reduces the accumulation of the "a" class hard metal aluminum (Al) in Arabidopsis root cell walls. The Al-binding capacity is determined in part by the extent of XyG *O*-acetylation, since more Al accumulates in the walls of Arabidopsis XyG *O*-acetyltransferase mutants than their wild-type counterpart (Zhu et al., 2014).

The primary cell wall is a versatile and dynamic structure (Fry, 1990). It provides mechanical strength yet is capable of expanding to allow cell growth and also has roles in water and mineral uptake, pathogen resistance, as well as developmental and physiological processes. Cell walls have become a target for bioengineering to improve the value of biomass for renewable energy production (Loque et al., 2015) or to enhance a plants' ability to bind toxic element present in contaminated soils (Zhu et al., 2012a,b, 2014; Han et al., 2014).

XyG is present in the cell walls of all land plants and is one of the most abundant non-cellulosic polysaccharides in dicot primary walls (Pauly and Keegstra, 2016). Xyloglucans have a 1,4-linked β-D-glucan backbone with α-D-xylopyranosyl (Xyl*p*) residues attached at *O*-6. The xylosyl residue is often substituted with D- and L-galactosyl, L-fucosyl, D-galacturonosyl, L-arabinopyranosyl, and/or L-arabinofuranosyl moieties (Pauly and Keegstra, 2016). To date, 24 different side chains, which are described using a one letter code have been identified (Pauly and Keegstra, 2016). The type and order of XyG sidechains depend on the plant species, the tissue and cell type, and the developmental state of the cell (Pauly and Keegstra, 2016). XyGs are classified either as "XXXG type" or "XXGG type." Tobacco leaf XyG is composed predominantly of XXGG and XSGG subunits. S represents an α-L-Ara*f*-(1→2)-α-D Xyl*p*-[1→side chain linked to *O*-6 of a 4-linked β-D-Glc*p* residue (Hoffman et al., 2015)].

The biosynthesis of XyG has been studied in detail (Pauly and Keegstra, 2016). Xyloglucan-specific xylosyltransferases (XXTs) catalyze the addition of Xyl residues to *O*-6 of the glucan backbone (Cavalier and Keegstra, 2006; Zabotina, 2012; Culbertson et al., 2016). In Arabidopsis, the XXT genes belong to the GT34 family which containing AtXXT1 to AtXXT5 and two mannan:galactosyltransferases (MGs) AtGT6 and AtGT7 (Vuttipongchaikij et al., 2012). In Arabidopsis, XXT1 and XXT2 are responsible for the most, if not all, of this xylosylation (Zabotina et al., 2012), since no XyG is discernible in the walls of the *xxt1/2* double mutant (Cavalier et al., 2008). This mutant grows and develops normally, which has led plant scientists to question the notion that a cellulose-XyG network is the major load-bearing structure in the walls of growing plant cells (Schultink et al., 2014). Nevertheless, the *xxt1/2* double mutant does have some visible phenotypes including short root hairs with bulging bases (Cavalier et al., 2008), shorter and wider hypocotyls, and bent stems (Xiao et al., 2016). Such changes are consistent with the notion that some of the cells of the *xxt1/2* mutant have walls with altered mechanical and chemical properties.

Arsenic (As) is a naturally occurring toxic metalloid element ranked as a top 20 priority hazardous substance by the Agency for Toxic Substances and Disease Registry (Abbas et al., 2018). Most environmental As occurs as oxyanions in either of two oxidation states: +3 (AsIII, arsenites) or +5 (AsV, arsenates and organoarsenic compounds). Low amounts of AsIII or AsV may cause substantial morphological, physiological, and biochemical changes once they enter plant cells (18). These include stunted growth, reduced photosynthetic efficiency, and decreased biomass accumulation. Arsenic also causes increased generation of reactive oxygen species (ROS), which interferes with numerous metabolic pathways (Abbas et al., 2018). The As absorbed by plants may accumulate to toxic levels in vegetables, grains, and fruits (Chen et al., 2017a), which can lead to As poisoning in humans *via* the food chain.

It is not known if the uptake or accumulation of As by plants is affected by cell wall XyG or whether XyG can lessen the deleterious effect of this toxic element. Common tobacco (*Nicotiana tabacum*) is an important agricultural crop and a model plant widely used for studying fundamental biological processes (Sierro et al., 2014). The genome of *Nicotiana tabacum* was sequenced in 2014 (Sierro et al., 2014). In this study, we identified 11 *XXT* orthologs in the tobacco genome. Two of these genes, *NtXXT1* and *NtXXT2*, are required for xyloglucan biosynthesis. CRISPR-Cas9 technology was used to generate single (*ntxxt1* and *ntxxt2*) and double (*ntxxt1/2*) mutants. The double mutant lacked XyG in its cell walls. We show that eliminating XyG increases the amounts of As bound to the root walls, which results in less As being translocated to the tobacco shoots.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

N. tabacum var. K326 was used for genetic transformations. The seeds of K326 and the *xxt* mutants were surface sterilized with aq. 10% sodium hypochlorite (NaClO) and then with aq. 75% alcohol. The seeds were germinated on half-strength Murashige and Skoog (MS) medium. Two-week-old seedlings were transferred to potting soil and grown at 25±1°C with a long-day photoperiod (16-h light/8-h dark) and 60% humidity in a greenhouse.

NtXXT Identification and Phylogenetic Analysis

XXT orthologs in *N. tabacum* were identified using the protein sequences of AtXXT1 to AtXXT5 and AtGT6 to AtGT7 downloaded from TAIR[1] and used as queries to BLAST the International Tomato Genome Database.[2] *AtXXTs* and *NtXXTs* protein sequences were aligned with ClustalW2[3] with a gap extension penalty of 0.1. Redundant sequences of NtXXTs were excluded. Phylogenetic analysis with a Poisson model method was performed using MEGA6.0 with a neighbor-joining method. The robustness of the tree topology was assessed using 1,000 bootstrap replicates.

Tissue Expression Analysis of *NtXXTs*

The spatiotemporal expression patterns of *NtXXTs* were obtained using semi-quantitative PCR. Selected tissues (young leaves, mature leaves, and senescent leaves, stem, roots, vein, and flowers) were collected during the tobacco growing season. Three biological independent replicates of each tissue were used. The tissues were frozen in liquid nitrogen and kept at −80°C. RNA extraction, cDNA synthesis, and semi-quantitative PCR reactions were performed as described (Wang et al., 2018). Tobacco *NtACTIN1* (XM_019370655.1) was used as the reference gene (Wang et al., 2018).

Single Guide RNA Design and Construction of CRISPR-Cas9 Binary Vectors

Single guide RNA (sgRNA) sequences of *Ntab-K326_AWOJ-SS412* (*NtXXT1*) and *Ntab-K326_AWOJ-SS848* (*NtXXT2*) were designed with the web-based tool CRISPR MultiTargeter.[4] Three different sgRNA sequences were designed for each gene to improve the success rate of editing. Each sgRNA sequence contained 20 nucleotides followed by the NGG trinucleotide protospacer adjacent motif (PAM) at the 3′-end of the target region. A pair of complementary sgRNA DNA oligonucleotides were synthesized in Qingke sequencing company and annealed to generate dimers. The dimers were inserted into a modified CRISPR-Cas9 pORE vector plasmid driven by the Arabidopsis U6-26 promoter (Gao et al., 2015). PCR and Sanger sequencing were used to ensure that no polymorphisms existed between the sgRNAs and the corresponding target sequences.

Tobacco Leaf Cell Transformation by *Agrobacterium* Infiltration

Tobacco leaf disc were used for transformation. *Agrobacterium tumefaciens* strain LBA4404 containing the CRISPR-Cas9 plasmid constructs was grown at 28°C in Luria–Bertani medium containing rifampicin and kanamycin antibiotics to an OD_{600} of 0.8. The bacteria were collected by centrifugation and the pellet suspended in MS liquid medium containing 30 g/L sucrose and 20 mg/L acetosyringone (MS_0 medium) and adjusted to an OD_{600} of 0.8. Tobacco leaf discs were then immersed for 8 to 10 min in the MS_0 medium containing the bacteria. The leaf discs were transferred to antibiotic-free MS medium and kept for 3 days in the absence of light. The transfected discs were then transferred to differentiation medium (MS containing 2 mg/L 6-benzylaminopurine, 0.1 mg/L 1-naphthaleneacetic acid [NAA], 50 mg/L kanamycin and hygromycin, and 500 mg/L cefotaxime) until buds formed. The regenerated plantlets were transferred on root-inducing medium (MS + 0.1 mg/L NAA + hygromycin + 200 mg/L cefotaxime) to produce the roots. Lastly, the drug-resistant seedlings were planting in the soil and used for further analysis.

Monosaccharide Compositions of the Cell Walls

The leaves of 6-week-old tobacco plants (K326, *ntxxxt1*, *ntxxt2*, and *ntxxt1/2*) were collected, frozen in liquid nitrogen, and ground to a powder. The materials were then prepared as their alcohol-insoluble residues (AIR) as described (Pena et al., 2008). In brief, the powder was sequentially extracted for 30 min each with aqueous 70 and 80% ethanol and then with absolute ethanol. The residue was suspended in acetone, filtered through Whatmans filter paper, and allowed to dry in a fume hood. The AIR was de-starched with α-amylase and amyloglucosidase (Sigma–Aldrich, St. Louis, MO, United States) at 37°C overnight.

Three different AIR (2 mg) from each plants was hydrolyzed for 2 h at 120°C with 2 M trifluoroacetic acid (TFA). The hydrolysates were then reacted for 30 min at 70°C with 1-phenyl-3-methyl-5-pyrazolone (PMP). The mixture was extracted three times with chloroform and the PMP-monosaccharides were analyzed by high-performance liquid chromatography (HPLC; Wang et al., 2020). Three biological replicates were used for each sample.

Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

The de-starched AIRs prepared from K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* leaves were treated with 50 mM ammonium oxalate to solubilize pectin and then with 4 M KOH to solubilize XyG. The XyG subunit composition (Kong et al., 2015) was determined by treating the 4 M KOH-soluble material in 50 mM ammonium formate, pH 5, for 24 h at room temperature with two units XEG (Kong et al., 2015). Ethanol was then added to 70% (v/v). The suspension was centrifuged ($2000 \times g$) and the soluble fraction transferred to a clean tube and concentrated to dryness under a flow of warm air. To ensure the ammonium formate was completely removed the residue was dissolved in water and freeze-dried three times. Solutions of the XEG-treated material in water (~1 mg/ml) were analyzed using a Bruker Microflex LT matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) in the positive ion mode and Bruker workstation (Bruker, Billerica, MA, United States) as described (Kong et al., 2015).

[1] http://www.arabidopsis.org
[2] https://solgenomics.net/organism/Solanum_lycopersicum/genome
[3] http://www.ebi.ac.uk/Tools/msa/clustalw2/
[4] http://www.multicrispr.net/

Arsenic Absorption by Tobacco Plants

Six-week-old tobacco were watering the HoaglandArnon nutrient solution (HNS) containing 20 μM arsenite (AsIII, $NaAsO_2$) or 20 μM arsenate (AsV, $Na_2HAsO_4{\cdot}7H_2O$) and cultivated for 3 days. The roots, shoots, and leaves were separately collected and kept at −80°C. For determining the As concentration in roots and shoots, the frozen tissues (~10 g) were kept for 30 min at 120°C, and then at 75°C until they reached a constant weight. For determining the As concentration in roots cell walls and leaves cell walls, the frozen tissues were grounded into powders and extracted in subcellular extraction buffer (250 mmol/L sucrose, 50 mmol/L Tris-HCl (pH7.4), and 1 mmol/L erythritol dithiocarcinol). After centrifugation at 3,000 rpm for 15 min, the cell wall components were precipitated and then dried at 75°C. Approximately 0.5 g of dry materials (three replicates performed in each plants) were digested in concentrated nitric acid (4 ml) for overnight. H_2O_2 (2 ml) was added, and the mixture was kept for 4 h at 120–130°C. After being cooled, the solution were then heated at 180–230°C to heat acid. The remaining solution was used to determine the As concentration by inductively coupled plasma mass spectrometry (ICP-MS; United States, PerkinElmer; Chen et al., 2016). This experiment was repeated three times.

RESULTS

Expression Profiles of Xyloglucan Xylosyltransferase Orthologs in Tobacco

A total of 11 putative *XXTs* genes were identified in the *N. tabacum* L. genome. Our phylogenetic analysis classified the *NtXXTs* into two sub-clusters. Seven *NtXXTs* (*NtXXT1* to *NtXXT7*) were clustered together with all five *AtXXTs*. Surprisingly, the remaining four *NtXXTs* (*NtXXT8* to *NtXXT11*) were clustered into a second group with MGs AtGT6 and AtGT7 (**Figure 1A**). Thus, these four genes are putative mannan:galactosyltransferases (MGs), and we renamed them as *NtGT6-like*, *NtGt7-like*, *NtGT6*, and *NtGT7*, respectively. *Ntab-K326_AWOJ-SS412* and *Ntab-K326_AWOJ-SS848* were grouped into the same branch as *AtXXT1* and *AtXXT2*, so we named them *NtXXT1* and *NtXXT2*, respectively, (**Figure 1A**).

We used semi-quantitative PCR to explore the expression levels of the *NtXXTs* in young leaves, mature leaves, senescent leaves, stem, roots, vein, and flowers. The expression patterns of *NtXXTs* varied somewhat. *NtXXT1* and *NtGT7-like* were expressed at a relatively high level in all tissues, whereas *NtXXT3*, *NtXXT5*, and *NtGT6-like* showed relatively low expression level in all tissues. The expression of *NtXXT4*, *NtGT6*, and *NtGT7* was barely detectable in any of the tissues (**Figure 1B**). NtXXT2 showed weak expression in leaves, stem, roots, and flowers (**Figure 1B**).

CRISPR-Cas9 Induced Mutation of Tobacco XXTs

Based on our phylogenetic and expression pattern analyses, we hypothesized that NtXXT1 and NtXXT2 are required for xyloglucan biosynthesis. To test this, we generated *ntxxt1* and *ntxxt2* knockout mutants by transforming K326 plants with pORE vectors that contained the Cas9 gene and sequences for single-stranded guide RNAs (sgRNA) that target *NtXXT1* and *NtXXT2* individually.

Three sgRNAs were designed for each gene (**Figure 2**). A total of 43 and 60 transgenic plants for targets 2 and 3, respectively, were obtained for *NtXXT1*. No transgenic plants were obtained for target 1. PCR sequencing revealed that four individuals had base variations at target 2 and two at target 3, with editing efficiencies of 9 and 3%, respectively. These transgenic plants showed deletion of one to three bases in the target region of *NtXXT1* (**Figure 2B**).

A total of 49, 19, and 82 transgenic plants were obtained for targets 1, 2, and 3, respectively, of *NtXXT2*. However, no plants with base changes at the target 1 site were identified. One target 2 transgenic plant and three target 3 transgenic plants were identified with editing efficiencies of 5 and 4%, respectively. Deletions of one to five bases and base insertions were identified at the target 2 and 3 regions of *NtXXT2* (**Figure 2E**). All the indel edit types resulted in termination codon occurrences in subsequent sequences both in *NtXXT1* and *NtXXT2* (**Figures 2B,E**). After self-fertilization, the F2 generation plants were used to screen for homozygous plants that lack Cas9. Only the "AC" (Adenine deoxyribonucleotide and Cytosine deoxyribonucleotide) and "CT" (Cytosine deoxyribonucleotide and Thymine deoxyribonucleotide) deletion homozygote plants of *NtXXT1* and *NtXXT2* (**Figures 2C,F**) were obtained and were referred to as knockout lines *ntxxt1* and *ntxxt2*, respectively. The *ntxxt1/2* double mutant was obtained by crossing the homozygous *ntxxt1* and *ntxxt2* mutants. The homozygous *ntxxt1/2* double mutant was identified by gold-standard Sanger sequencing.

The *ntxxt1/2* Mutant Has Growth Slower Defects and Has a Severe Root Hair Phenotype

The Arabidopsis *atxxt1/2* double mutant does not have a severe visible phenotype, even though no xyloglucan is synthesized (Cavalier et al., 2008; Xiao et al., 2016). Our *ntxxt1*, *ntxxt2*, and *ntxxt1/2* plants grew somewhat more slowly (**Figures 3A–C,E,F**) but only the *ntxxt2* and *ntxxt1/2* were appreciably smaller than wild-type at the adult stage (**Figure 3C**). Reduced growth, as indicated by leaf area, was more discernible in the *ntxxt1/2* double mutant than in either single mutant (**Figure 3E**). The growth rate indicated by plant height was also examined in the first 8 weeks (**Figure 3F**). The growth was reduced in *ntxxt1*, *ntxx2*, and *ntxxt1/2* mutants compared to K326. Moreover, the *ntxxt1/2* double mutant leaves were smaller and rounder than the leaves of K326 (**Figure 3B**). Similar visible differences in growth were also apparent in 4-week-old and adult plants (**Figure 3B**).

We also examined the root hair phenotype of the mutants. The root hairs of the *ntxxt1* and *ntxxt2* mutants are indistinguishable from those of K326 (**Figure 3D**). By contrast, the number of root hairs formed is substantially reduced in the double mutant and they are much shorter compared to K326 root hairs. The severe root hair phenotypes *ntxxt1/2* and *atxxt1/2* are similar to *atxxt1/2* with one exception:

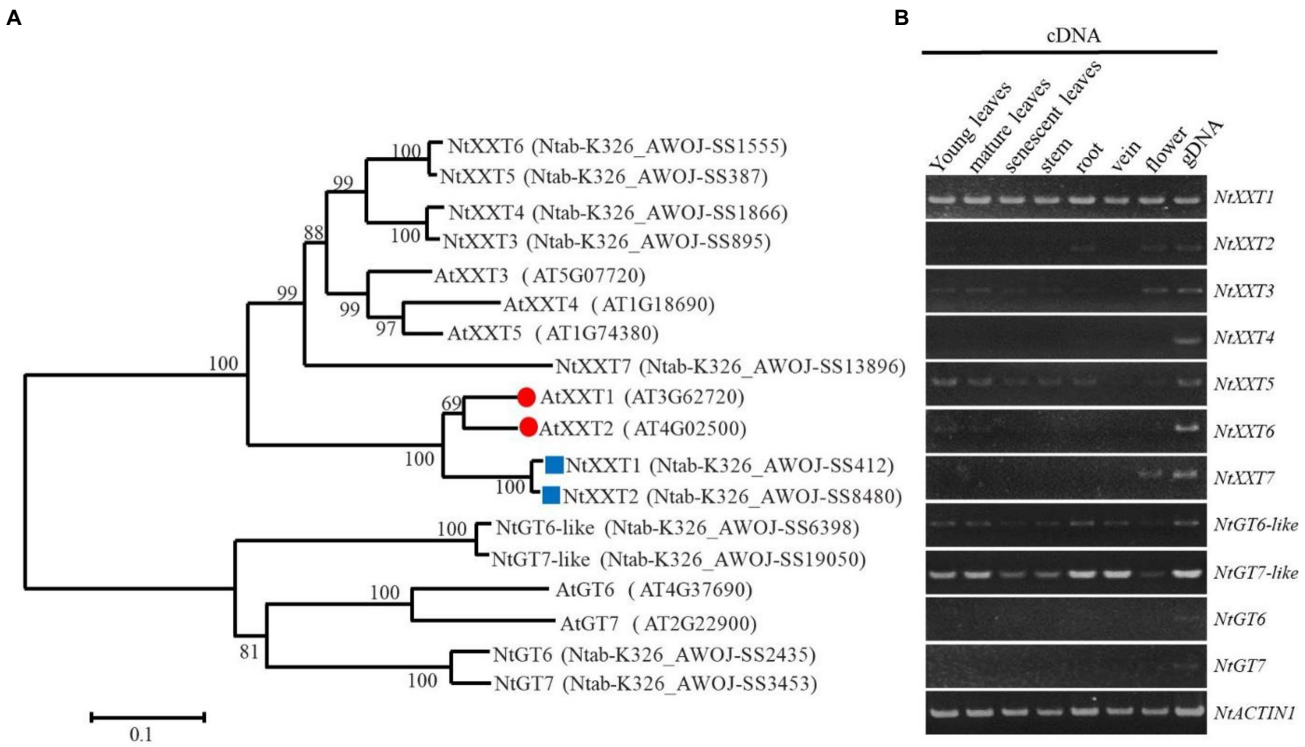


FIGURE 1 | Phylogenetic analysis of NtXXT and AtXXT protein sequences and tissue-specific expression of the corresponding *NtXXT* genes. **(A)** Phylogenetic analysis of the AtXXT and NtXXT, NtGT protein sequences. The red circles indicated the genes that been well studied in *Arabidopsis*. The blue boxes indicated the NtXXTs that are tightly closed to the well-studied AtXXTs. **(B)** The expression pattern of all the *NtXXTs* in different stages of leaves, stems, roots, vein, and flowers.

there is no bulging at the *ntxxt1/2* root hair base. Taken together these data show that eliminating both NtXXT1 and NtXXT2 has a discernible effect of the growth and development of tobacco plants.

The Cell Walls of *ntxxt1/2* Lack Xyloglucan

The phenotypes of the *ntxxt1/2* mutant indicated that these two tobacco XXTs have a function similar to AtXXT1 and AtXXT2. To determine if xyloglucan is absent in *ntxxt1/2*, we treated the 4 M KOH-soluble material from the leaves of 4-week-old wild-type and mutant plants with XyG-specific endoglucanase (XEG) and analyzed the products formed with MALDI-TOF-MS (Kong et al., 2015). Our results show that there is no discernible difference in the relative abundance of XyG oligosaccharides in wild-type and *ntxxt1* or *ntxxt2* plants (**Figure 4**). However, no signals for XyG oligosaccharides were detected in *ntxxt1/2* mutant (**Figure 4**). We also found that the xylose content of the *ntxxt1/2* double mutant AIR was ~50% lower than that in the wild-type (**Figure 5**). The xylose content of the *ntxxt1* and *ntxxt2* single mutants decreased 6.7 and 11.2%, respectively (**Figure 5**). The xylose content of the AIR material was mainly from xyloglucan and xylan. Thus, the remaining 50% xylose content in *ntxxt1/2* double mutants is maybe from xylan. These results suggest that no detectable amounts of XyG is synthesized by the *ntxxt1/2* double mutant.

Overexpression of *NtXXT1* or *NtXXT2* Rescues the Root Hair and Slow Growth Phenotype of the *atxxt1/2* Mutant

Our results have demonstrated that tobacco plants lacking both NtXXT1 and NtXXT2 have growth and root hair phenotypes and wall chemotypes that are similar to the Arabidopsis *xxt1/2* mutant plant. Thus, tobacco and Arabidopsis XXT1 and XXT2 likely have similar biochemical functions. To test this, we separately transformed the *atxxt1/2* double mutant with the coding sequences of NtXXT1 and NtXXT2 driven by a *35S* promoter. We obtained 10 complemented transgenic lines for *35S-NtXXT1* and 15 for *35S-NtXXT2*. The short root hair defect was rescued in all these transgenic lines (**Figure 6A**). And the slow growth phenotype of 3-week-old plants was also restored to WT (**Figure 6B**) in all these transgenic lines.

XyG Deficiency in the Cell Wall Affects as Accumulation and Translocation

It is not known if plants' ability to accumulate As is altered by modifying its cell walls. Thus, we exposed wild-type and mutant tobacco plants to As to determine if the presence or absence of *NtXXT1* and *NtXXT2* affects As accumulation or translocation. A low As concentration (20 μM) was used as this does not cause severe growth inhibition in K326. We first determined the As concentration of root and leaf cell walls, since XyG is the predominant hemicellulose in tobacco primary cell walls and its absence may

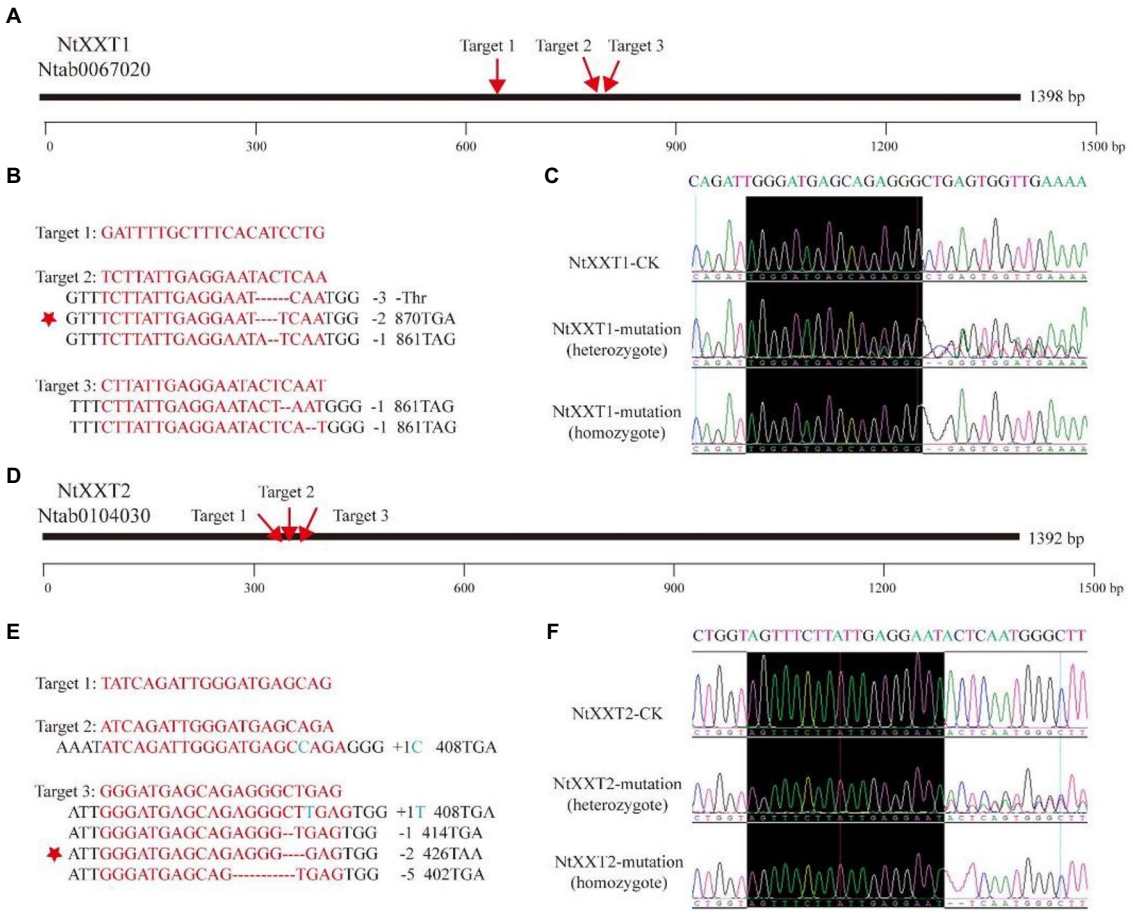


FIGURE 2 | Regions of *NtXXT1* and *NtXXT2* targeted by CRISPR-Cas9. **(A)** The gRNA targets in the *NtXXT1*. The red arrows indicate the positions of the three target sites. **(B)** Editing profiles at *NtXXT1*. The target sequence is shown in red letters, for which red dashes indicated deletions. Different editing types were detected which can result in amino acid deletion and termination codon. The red stars indicated the deletion lines which were obtained homozygous lines and used for further analysis. **(C)** The knockout of *NtXXT1* was verified by gold-standard Sanger sequencing in heterozygous and homozygous lines which comes from the red star labeling in **(B)**. The black box indicated the target sequence upstream of the editing location. **(D)** The gRNA targets in the *NtXXT2*. The red arrows indicate the positions of the three target sites. **(E)** Editing profiles at *NtXXT2*. The target sequence is shown in red letters, for which red dashes indicated deletions, and the blue letters indicated insertion. Different editing types were detected which can result in termination codon. The red stars indicated the deletion lines which were obtained homozygous lines and used for further analysis. **(F)** The knockout of *NtXXT2* was verified by gold-standard Sanger sequencing in heterozygous and homozygous lines. The black box indicated the target sequence upstream of the editing location.

affect their mechanical and biochemical properties. The *ntxxt1/2* mutant grown with AsIII or AsV has increased amounts of As in its root cell walls and decreased amounts of As in its leaf cell walls compared to K326 (**Figures 7A,B**). The As contents of *ntxxt1* and *ntxxt2* root walls were not significantly different from K326, irrespective of whether they were grown with AsIII or AsV. Growing *ntxxt1/2* in the presence of AsIII and AsV also resulted in a significant reduction of leaf wall As. Leaf wall As was significantly lower than K326 for all three mutants grown in the presence of AsV.

The roots of *ntxxt1*, *ntxxt2*, and *ntxxt1/2* mutants grown with 20 μM AsIII, had 55, 118, and 154% more As, respectively, than K326 roots. Shoot As decreased by 7% in *ntxxt2* and by 49% in *ntxxt1/2* (**Figure 7C**). The calculated translocation factors (As concentration ratio in shoots to roots) for *ntxxt1*, *ntxxt2*, and *ntxxt1/2* were reduced by 21, 19, and 64%, respectively (**Figure 7D**). Similarly, As concentration increased in mutants grown with 20 μM AsV. Root As increased by 34, 107, and 140% in *ntxxt1*, *ntxxt2*, and *ntxxt1/2*, respectively. Shoot As decreased by 57, 53, and 75% in *ntxxt1*, *ntxxt2*, and *ntxxt1/2*, respectively (**Figure 7E**). The calculated translocation factors for *ntxxt1*, *ntxxt2*, and *ntxxt1/2* were reduced by 38, 62, and 71%, respectively (**Figure 7F**). Taken together these results suggest that As is retained more effectively in the roots of *ntxxt1/2* mutant and that As translocation from root to shoots is attenuated in the double mutants.

DISCUSSION

In this study, we used CRISPR-Cas9 technology to generate a tobacco mutant (*ntxxt1/2*) with cell walls that lack XyG.

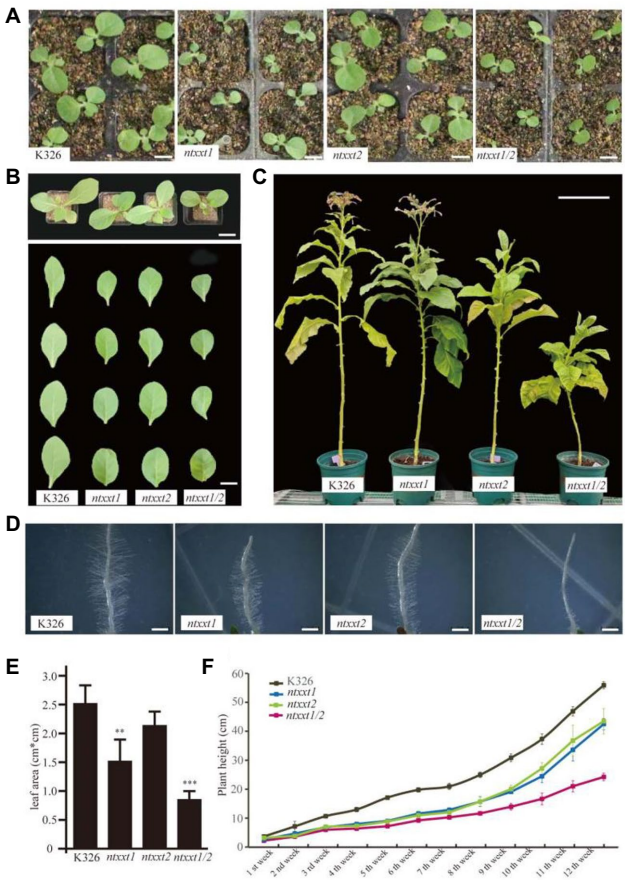


FIGURE 3 | The morphology phenotype of *ntxx1/2* double mutants. **(A)** Two-week-old seedlings of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* mutants. Bar = 1.5 cm. **(B)** The morphological phenotypes and leaf shape of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* 4-week-old seedlings. Bar = 3.5 cm. **(C)** The morphological phenotypes of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* adult plants. Bar = 0.25 m. **(D)** Root hair phenotypes of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* seedlings. Bar = 0.5 mm. **(E)** The statistical data leaf area of the 2-week-old K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* seedlings. ** Statistically significant differences from K326 using Student's *t*-test ($p < 0.01$), *** statistically significant differences from K326 using Student's *t*-test ($p < 0.001$). **(F)** The growth rates indicated by plant height of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* mutants. These experiments were performed in triplicate. Error bars, mean ± SE.

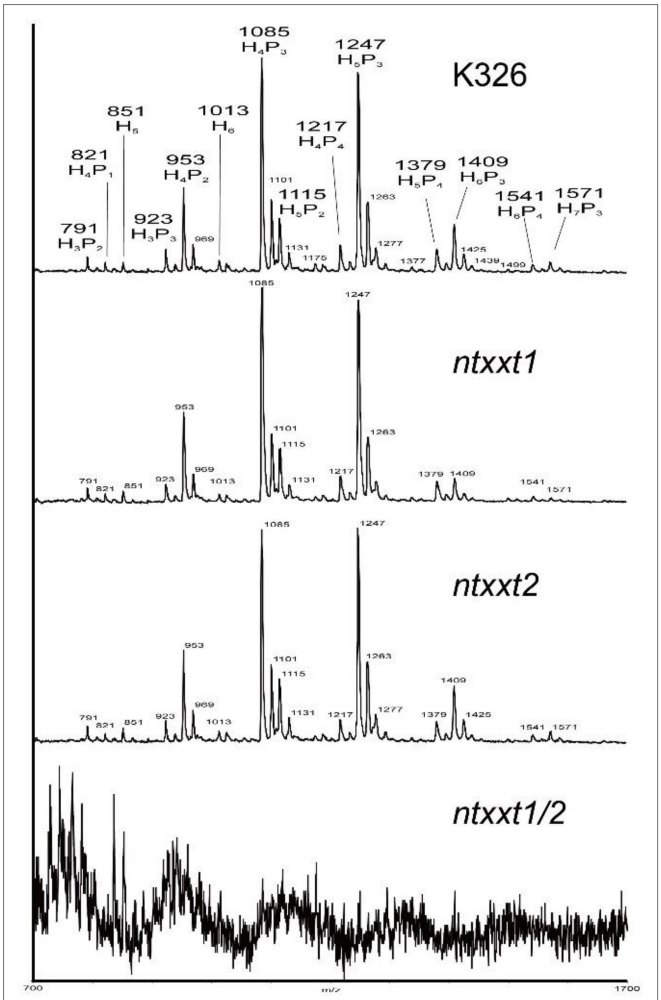


FIGURE 4 | MALDI-TOF spectra of the XyG oligosaccharides generated from the 4 M KOH-soluble of AIR obtained from K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* mutants plants.

This modification resulted in root cell walls that accumulate more As than wild-type plants and thereby decreases the amounts of As that is translocated to the shoots (**Figure 7**). The relatively big increase of As in *ntxxt2* may be the result of the more decrease of xylose concentration. Thus, targeted engineering of cell walls provides an opportunity to generate plants capable of remediating soils contaminated with As.

Our data show that root cell walls lacking XyG bind more As than walls containing XyG. However, the nature of the interactions remains to be determined. In aqueous solutions, AsIII and AsV exist as protonated oxyanions with degrees of protonation that are pH-dependent (Ferguson and Gavis, 1972). AsIII has two pK_as (pH 9.2 and 12.1) whereas AsV has three (pH 2.2, 6.9, and 11.5). Thus, in the pH range of 4 to 8 typically encountered in the plant apoplast (Martiniere et al., 2013), AsIII will exist predominantly as the neutral arsenite H_3AsO_3, whereas AsV will exist as the negatively charged arsenates $H_2AsO_4^-$ and $HAsO_4^{2-}$ (Wang and Mulligan, 2006). Arsenite has a high affinity for sulfhydryl groups (Shen et al., 2013), whereas arsenite and arsenate may form weak complexes with carboxylic acids, phenols, amines, and alcohols (Fakour and Lin, 2014).

There are only a limited number of studies that provide insight into the interactions between arsenites, arsenates, and cell wall polysaccharides. For example, As-pectin interactions have been proposed to occur *via* hydrogen-bonded bridging between the protons of the As species and the ionized carbonyl and carboxyl groups of the pectin (Fox et al., 2012). The As-pectin interaction has been reported to increase if the As exists as an iron oxide–arsenic complex (Wang and Mulligan, 2006; Fakour and Lin, 2014). Much more information on the interactions between As and polyelectrolytes has been obtained from studies of humic and fulvic acids (Wang and Mulligan,

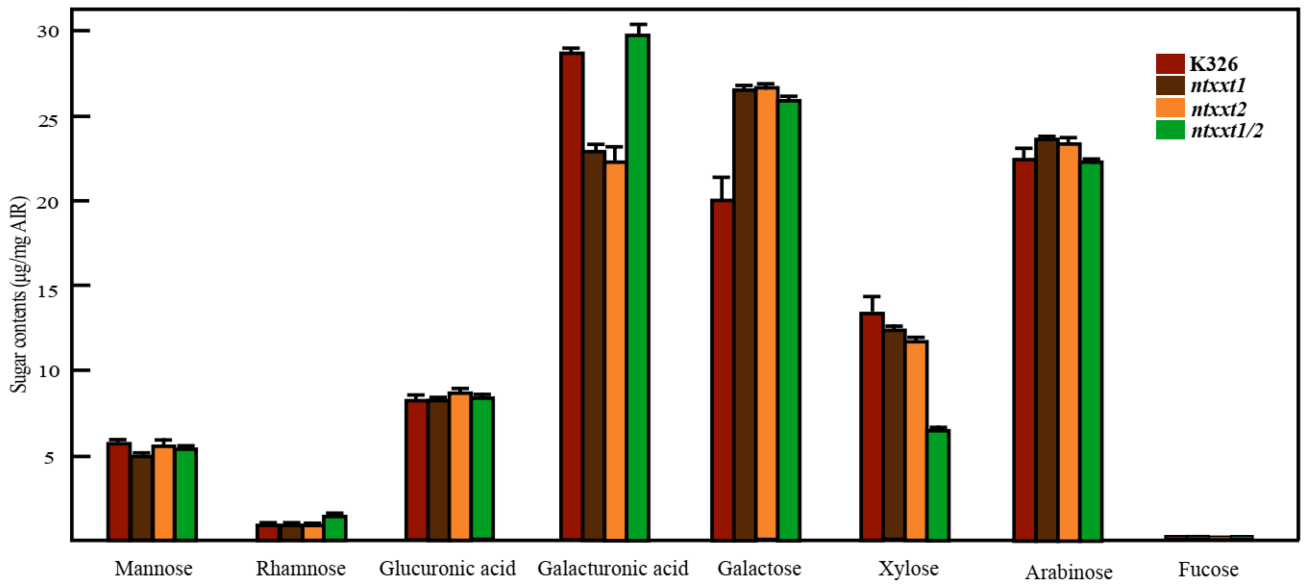


FIGURE 5 | The glycosyl residue compositions of the AIR from K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* leaves. These experiments were performed in triplicate. Error bars, mean ± SE.

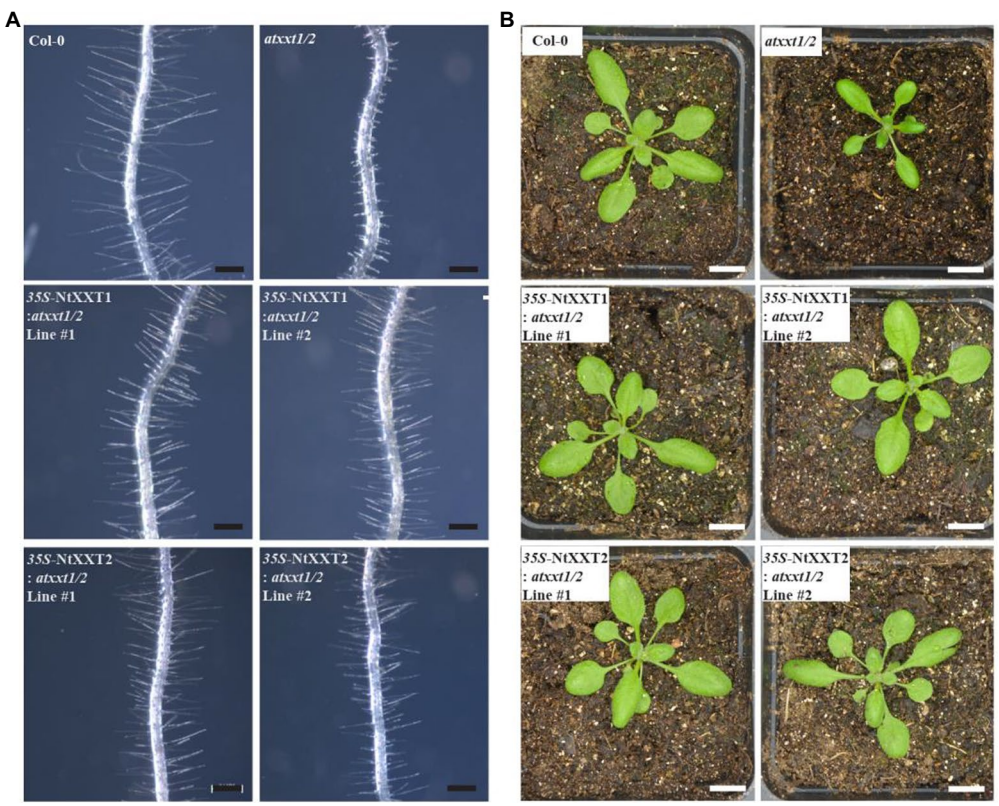


FIGURE 6 | The root hair and plants phenotypes of complemented *atxxt1/2* lines. **(A)** The root hair phenotypes of 7 day-old Arabidopsis seedlings were observed. The root hairs of the *atxxt1/2* double mutant are far shorter than Col-0. Overexpressing *35S*-NtXXT1 or *35S*-NtXXT2 in *atxxt1/2* plants rescues the root hair defect of the *atxxt1/2* double mutant. Bar = 100 μm. **(B)** The 3-week-old plants phenotype was observed. The *atxxt1/2* double mutant plants were smaller than Col-0, and overexpressing *35S*-NtXXT1 or *35S*-NtXXT2 in *atxxt1/2* plants were recovered. Bar = 10 mm.

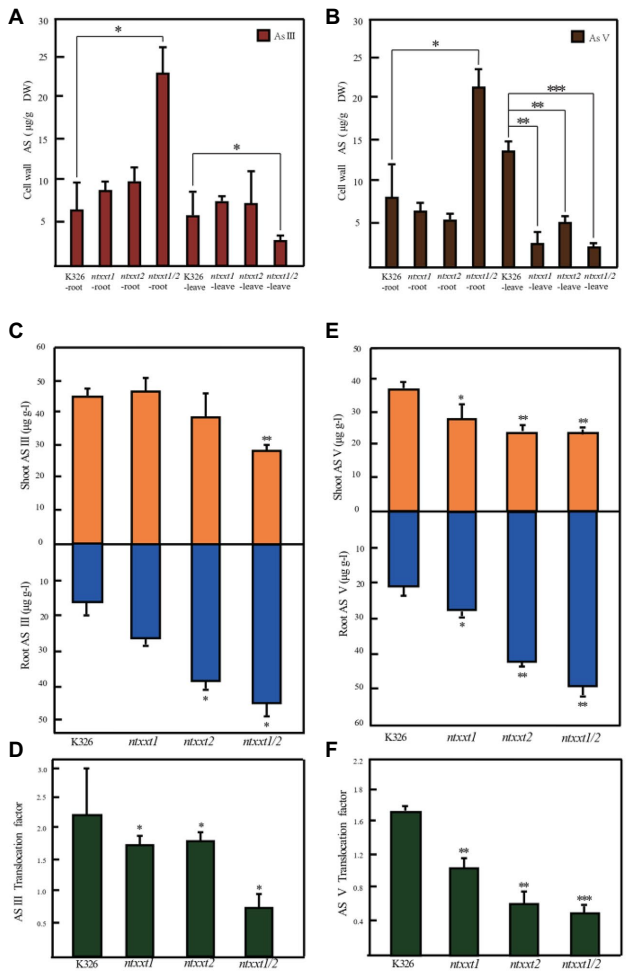


FIGURE 7 | The As accumulation and translocation in *ntxxt1/2* double mutants grown with AsIII or AsV. **(A)** The amounts of As in the leaf and roots cell walls of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* grown with AsIII. **(B)** The amounts of As in the leaf and root cell walls of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* grown with AsV. **(C,E)** The amounts of As in shoots and roots of K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* grown in the presence of AsIII or AsV. **(D,F)** Translocation factor (As concentration ratio in shoots to roots) for K326, *ntxxt1*, *ntxxt2*, and *ntxxt1/2* grown in the presence of AsIII or AsV. Experiments were performed in three biological replicates, and each biological replicates were performed in three technological replicates. Error bars, mean ± SE. * Statistically significant differences from K326 using Student's *t*-test ($p < 0.05$), ** statistically significant differences from K326 using Student's *t*-test ($p < 0.01$), *** statistically significant differences from K326 using Student's *t*-test ($p < 0.001$).

2006; Fakour and Lin, 2014). These interactions may in part involve As anions forming metal bridges with the cations that are themselves bound to the polyelectrolyte. Thus, it is conceivable that As interacts with the metal cations bound to pectin. The galacturonic acid of cell wall pectin is typically associated with divalent cations, most notably calcium (Pérez et al., 2000; Braccini and Perez, 2001). Pectin has been reported to be a major contributor to As retention in rice cell walls *via* an interaction that is enhanced by silicates (Cui et al., 2020). Arsenic binding sites may also involve metal ions that form coordination complexes with the hydroxyl groups of mono- and polysaccharides (Angyal, 1989; Joly et al., 2020).

Xyloglucan is a quantitatively major polysaccharide in the primary cell wall of flowering plants. It has been proposed to interact with cellulose to form the major load-bearing network of these walls (Zhu et al., 2014). However, this notion has been challenged by the demonstration that an Arabidopsis mutant (*atxxt1/2*) unable to synthesize XyG does not have severe growth and developmental defects (Cavalier et al., 2008). Our study has shown that Tobacco plants lacking XyG also have no severe growth phenotypes. Thus, laboratory grown plants are able to survive in the absence of XyG. It is possible that in the absence of XyG, pectin takes on a greater share of the biomechanical load of the walls (Rui and Dinneny, 2020). Moreover, in XyG-deficient plants, the relative abundance or accessibility of wall pectin may increase and account for the increased amounts of bound As.

The absence of XyG is known to affect the organization of cellulose microfibrils and lead to a decrease in cellulose biosynthesis (Xiao et al., 2016). The cellulose microfibrils are largely parallel to one another, and the spacing between them is increased in the etiolated hypocotyls of the *atxxt1/2* mutant. Furthermore, the orientation of the cellulose microfibrils changes. In 3-day-old seedlings, the fibrils were predominantly transverse relative to the cell axis but longitudinal in 6-day-old *atxxt1/2* seedlings. Different orientations of cellulose may affect wall mechanics and properties. Thus, in the absence of XyG, cellulose organization may be altered and lead to changes in pectin organization within the wall. This pectin and other wall components may then become more accessible for binding with As.

We have provided evidence that no XyG is present in the walls of the *ntxxt1/2* double mutant. We expected that four tobacco orthologs of XXT1 and XXT2 would exist since *N. tabcum* is tetraploid. However, only two orthologs were identified and a tobacco mutant lacking functional *NtXXT1* and *NtXXT2* produced no discernible amounts of XyG. Thus, NtXXT1 and NtXXT2 must be the main xylosyltransferases involved in *N. tabcum* XyG biosynthesis. Both of these genes are expressed, albeit at different levels, in leaves, stems, veins and flowers. Nevertheless, *NtXXT1* expression is much higher than *NtXXT2* in these tissues (**Figure 1B**), which suggests that NtXXT1 is the predominant XyG-specific xylosyltransferase in most tissues. It is notable that reduced flowering and reduced plant height were more pronounced in the *ntxxt2* mutant than in *ntxxt1* (**Figure 3C**), which indicates that NtXXT2 may function predominantly in mature plants and at flowering time.

Elevated levels of As in the environment are harmful to virtually all organisms, limit agricultural productivity, and have severe effects on human health. Arsenic poisoning in humans often occurs *via* the food chain as plants grown in As-rich environments may accumulate toxic amounts of this element. Economically viable methods are required to remediate soils containing hazardous levels of As. Previous studies have shown that AsIII and AsV are transported into plant cells by aquaporin and phosphate transporters

(Chen et al., 2017b, 2019, 2021). These transporters are required for normal plant growth and development. Thus, it is unlikely that they can be engineered to limit As uptake without affecting plant productivity. The cell wall is the first barrier to prevent As entry into plant cells. Our study and a previous report (Cavalier et al., 2008) have shown that this wall is malleable and that targeted elimination of XyG from the wall does not cause severe growth defects. We have shown that CRISPR-Cas9 technology can be used to generate tobacco plants that lack XyG in their cell walls. This modification allows the plant to accumulate more As in its root cell walls. Field studies are now required to demonstrate that such plants can be used to phytoremediate As contaminated soils.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MW, XS, SG, PL, ZX, YK, and DY conceived the study. MW, XS, SG, PL, ZX, AD, HX, and RA performed the experiments. MW analyzed the data. XS prepared the figures. MW, YK, MO'N, and GZ drafted the manuscript. All authors discussed the results, edited the manuscript, and approved the final manuscript.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (32070330, 31670302, and 31470291), Yunnan Academy of Tobacco Agricultural Sciences (Nos. 2018530000241002 and 2019530000241003), Project of Shandong Natural Science Foundation (ZR2021QC138), the First-Class Grassland Science Discipline Program of Shandong Province, the Taishan Scholar Program of Shandong (to GZ), Grant DE-SC0008472 from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the United States Department of Energy, and Foundation of Shandong Province Modern Agricultural Technology System (SDAIT-25-05).

REFERENCES

Abbas, G., Murtaza, B., Bibi, I., Shahid, M., Niazi, N. K., Khan, M. I., et al. (2018). Arsenic uptake, toxicity, detoxification, and speciation in plants: physiological, biochemical, and molecular aspects. *Int. J. Environ. Res. Public Health* 15:59. doi: 10.3390/ijerph15010059

Angyal, S. J. (1989). Complexes of metal cations with carbohydrates in solution. *Adv. Carbohydr. Chem. Biochem.* 47, 1–4. doi: 10.1016/S0065-2318(08)60411-4

Braccini, I., and Perez, S. (2001). Molecular basis of $Ca2^{+}$−induced gelation in alginates and pectins: the egg-box model revisited. *Biomacromolecules* 2, 1089–1096. doi: 10.1021/bm010008g

Cavalier, D. M., and Keegstra, K. (2006). Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. *J. Biol. Chem.* 281, 34197–34207. doi: 10.1074/jbc.M606379200

Cavalier, D. M., Lerouxel, O., Neumetzler, L., Yamauchi, K., Reinecke, A., Freshour, G., et al. (2008). Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. *Plant Cell* 20, 1519–1537. doi: 10.1105/tpc.108.059873

Chen, Y., Fu, J. W., Han, Y. H., Rathinasabapathi, B., and Ma, L. Q. (2016). High as exposure induced substantial arsenite efflux in as-hyperaccumulator pteris vittata. *Chemosphere* 144, 2189–2194. doi: 10.1016/j.chemosphere.2015.11.001

Chen, Y., Han, Y. H., Yue, C., Zhu, Y. G., and Ma, L. Q. (2017a). Arsenic transport in rice and biological solutions to reduce arsenic risk from rice. *Front. Plant Sci.* 8:268. doi: 10.3389/fpls.2017.00268

Chen, Y., Hua, C. Y., Chen, J. X., Rathinasabapathi, B., Cao, Y., and Ma, L. Q. (2019). Expressing arsenite antiporter PvACR3;1 in rice (*Oryza sativa* L.) decreases inorganic arsenic content in rice grains. *Environ. Sci. Technol.* 53, 10062–10069. doi: 10.1021/acs.est.9b02418

Chen, Y., Hua, C. Y., Jia, M. R., Fu, J. W., Liu, X., Han, Y. H., et al. (2017b). Heterologous expression of pteris vittata arsenite antiporter PvACR3;1 reduces arsenic accumulation in plant shoots. *Environ. Sci. Technol.* 51, 10387–10395. doi: 10.1021/acs.est.7b03369

Chen, Y., Wang, H. Y., and Chen, Y. F. (2021). The transcription factor MYB40 is a central regulator in arsenic resistance in *Arabidopsis*. *Plant Commun.* 2:100234. doi: 10.1016/j.xplc.2021.100234

Cui, J., Li, Y., Jin, Q., and Li, F. (2020). Silica nanoparticles inhibit arsenic uptake into rice suspension cells via improving pectin synthesis and the mechanical force of the cell wall. *Environ. Sci. Nano* 7, 162–171. doi: 10.1039/C9EN01035A

Culbertson, A. T., Chou, Y. H., Smith, A. L., Young, Z. T., Tietze, A. A., Cottaz, S., et al. (2016). Enzymatic activity of xyloglucan xylosyltransferase 5. *Plant Physiol.* 171, 1893–1904. doi: 10.1104/pp.16.00361

Douchiche, O., Driouich, A., and Morvan, C. (2010). Spatial regulation of cell-wall structure in response to heavy metal stress: cadmium-induced alteration of the methyl-esterification pattern of homogalacturonans. *Ann. Bot.* 105, 481–491. doi: 10.1093/aob/mcp306

Fakour, H., and Lin, T. F. (2014). Experimental determination and modeling of arsenic complexation with humic and fulvic acids. *J. Hazard. Mater.* 279, 569–578. doi: 10.1016/j.jhazmat.2014.07.039

Ferguson, J. F., and Gavis, J. A. (1972). Review of the arsenic cycle in natural waters. *Water Res.* 6, 1259–1274. doi: 10.1016/0043-1354(72)90052-8

Fox, D. I., Pichler, T., Yeh, D. H., and Alcantar, N. A. (2012). Removing heavy metals in water: the interaction of cactus mucilage and arsenate (As (V)). *Environ. Sci. Technol.* 46, 4553–4559. doi: 10.1021/es2021999

Fry, S. C. (1990). "Roles of the primary cell wall in morphogenesis," in *Progress in plant cellular and molecular biology*. eds. H. J. J. Nijkamp and L. H. W. van der Plas, J. van Aartrijk (Kluwer Academic Publishers: Dordrecht, Netherlands), 9, 504–513.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., et al. (2015). Crispr/cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Mol. Biol.* 87, 99–110. doi: 10.1007/s11103-014-0263-0

Han, Y., Sa, G., Sun, J., Shen, Z. D., Zhao, R., Ding, M. Q., et al. (2014). Overexpression of populus euphratica xyloglucan endotransglucosylase/hydrolase gene confers enhanced cadmium tolerance by the restriction of root cadmium uptake in transgenic tobacco. *Environ. Exp. Bot.* 100, 74–83. doi: 10.1016/j.envexpbot.2013.12.021

Hoffman, M., Jia, Z., Peña, M. J., Cash, M., Harper, A., Blackburn, A. R., et al. (2015). Structural analysis of xyloglucans in the primary cell walls of plants in the subclass *Asteridae*. *Carbohydr. Res.* 340, 1826–1840. doi: 10.1016/j.carres.2005.04.016

Joly, N., Ghemati, D., Aliouche, D., and Martin, P. (2020). Interaction of metal ions with mono- and polysaccharides for wastewater treatment: a review. *Nat. Prod. Chem. Res.* 8:373. doi: 10.35248/2329-6836.20.8.373

Kong, Y., Pena, M. J., Renna, L., Avci, U., Pattathil, S., and Tuomivaara, S. T. (2015). Galactose-depleted xyloglucan is dysfunctional and leads to

dwarfism in *Arabidopsis*. *Plant Physiol.* 167, 1296–1306. doi: 10.1104/pp.114.255943

Loque, D., Scheller, H. V., and Pauly, M. (2015). Engineering of plant cell walls for enhanced biofuel production. *Curr. Opin. Plant Biol.* 25, 151–161. doi: 10.1016/j.pbi.2015.05.018

Martiniere, A., Bassil, E., Jublanc, E., Alcon, C., Reguera, M., Sentenac, H., et al. (2013). In vivo intracellular ph measurements in tobacco and *Arabidopsis* reveal an unexpected ph gradient in the endomembrane system. *Plant Cell* 25, 4028–4043. doi: 10.1105/tpc.113.116897

Parrotta, L., Guerriero, G., Sergeant, K., Cai, G., and Hausman, J. F. (2015). Target or barrier? The cell wall of early- and later-diverging plants vs cadmium toxicity: differences in the response mechanisms. *Front. Plant Sci.* 6:133. doi: 10.3389/fpls.2015.00133

Pauly, M., and Keegstra, K. (2016). Biosynthesis of the plant cell wall matrix polysaccharide xyloglucan. *Annu. Rev. Plant Biol.* 67, 235–259. doi: 10.1146/annurev-arplant-043015-112222

Pena, M. J., Darvill, A. G., Eberhard, S., York, W. S., and O'Neill, M. A. (2008). Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* 18, 891–904. doi: 10.1093/glycob/cwn078

Pérez, S., Mazeau, K., and du Penhoat, C. H. (2000). The three-dimensional structures of the pectic polysaccharides. *Plant Physiol. Biochem.* 38, 37–55. doi: 10.1016/S0981-9428(00)00169-8

Rui, Y., and Dinneny, J. R. (2020). A wall with integrity: surveillance and maintenance of the plant cell wall under stress. *New Phytol.* 225, 1428–1439. doi: 10.1111/nph.16166

Schultink, A., Liu, L., Zhu, L., and Pauly, M. (2014). Structural diversity and function of xyloglucan sidechain substituents. *Plan. Theory* 3, 526–542. doi: 10.3390/plants3040526

Shen, S., Li, X. F., Cullen, W. R., Weinfeld, M., and Le, X. C. (2013). Arsenic binding to proteins. *Chem. Rev.* 113, 7769–7792. doi: 10.1021/cr300015c

Sierro, N., Battey, J. N., Ouadi, S., Bakaher, N., Bovet, L., and Willig, A. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 5:3833. doi: 10.1038/ncomms4833

Vuttipongchaikij, S., Brocklehurst, D., Steele-King, C., Ashford, D. A., Gomez, L. D., and McQueen-Mason, S. J. (2012). Arabidopsis GT34 family contains five xyloglucan α-1,6-xylosyltransferases. *New Phytol.* 195, 585–595. doi: 10.1111/j.1469-8137.2012.04196.x

Wang, S., and Mulligan, C. N. (2006). Effect of natural organic matter on arsenic release from soils and sediments into groundwater. *Environ. Geochem. Health* 28, 197–214. doi: 10.1007/s10653-005-9032-y

Wang, M., Xu, Z., Ding, A., and Kong, Y. (2018). Genome-wide identification and expression profiling analysis of the xyloglucan endotransglucosylase/hydrolase gene family in tobacco (*Nicotiana tabacum* L.). *Genes* 9:273. doi: 10.3390/genes9060273

Wang, M., Xu, Z., Guo, S., Zhou, G., O'Neill, M., and Kong, Y. (2020). Identification of two functional xyloglucan galactosyltransferase homologs BrMUR3 and BoMUR3 in brassicaceous vegetables. *PeerJ* 8:e9095. doi: 10.7717/peerj.9095

Xiao, C., Zhang, T., Zheng, Y., Cosgrove, D. J., and Anderson, C. T. (2016). Xyloglucan deficiency disrupts microtubule stability and cellulose biosynthesis in *Arabidopsis*, altering cell growth and morphogenesis. *Plant Physiol.* 170, 234–249. doi: 10.1104/pp.15.01395

Xiong, J., An, L., Lu, H., and Zhu, C. (2009). Exogenous nitric oxide enhances cadmium tolerance of rice by increasing pectin and hemicellulose contents in root cell wall. *Planta* 230, 755–765. doi: 10.1007/s00425-009-0984-5

Zabotina, O. A. (2012). Xyloglucan and its biosynthesis. *Front. Plant Sci.* 3:134. doi: 10.3389/fpls.2012.00134

Zabotina, O. A., Avci, U., Cavalier, D., Pattathil, S., Chou, Y. H., Eberhard, S., et al. (2012). Mutations in multiple XXT genes of *Arabidopsis* reveal the complexity of xyloglucan biosynthesis. *Plant Physiol.* 159, 1367–1384. doi: 10.1104/pp.112.198119

Zhu, X. F., Lei, G. J., Jiang, T., Liu, Y., Li, G. X., and Zheng, S. J. (2012a). Cell wall polysaccharides are involved in P-deficiency-induced cd exclusion in *Arabidopsis thaliana*. *Planta* 236, 989–997. doi: 10.1007/s00425-012-1652-8

Zhu, X. F., Shi, Y. Z., Lei, G. J., Fry, S. C., Zhang, B. C., Zhou, Y. H., et al. (2012b). XTH31, encoding an in vitro XEH/XET-active enzyme, regulates aluminum sensitivity by modulating in vivo XET action, cell wall xyloglucan content, and aluminum binding capacity in Arabidopsis. *Plant Cell* 24, 4731–4747. doi: 10.1105/tpc.112.106039

Zhu, X. F., Sun, Y., Zhang, B. C., Mansoori, N., Wan, J. X., and Liu, Y. (2014). TRICHOME BIREFRINGENCE-LIKE27 affects aluminum sensitivity by modulating the O-acetylation of xyloglucan and aluminum-binding capacity in *Arabidopsis*. *Plant Physiol.* 166, 181–189. doi: 10.1104/pp.114.243808

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fpls.2022.850711

RNAi of Sterol Δ24-Isomerase Implicated Its Involvement in Physalin Biosynthesis in *Physalis angulata* L.

*Jiao Yang[1], Jingyi Tian[2], Yuhui Yang[2], Yaru Zhu[3,4], Changfu Li[2] and Yansheng Zhang[2]**

[1]School of Environmental and Chemical Engineering, Shanghai University, Shanghai, China, [2]School of Life Science, Shanghai University, Shanghai, China, [3]Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, [4]College of Life Science, University of Chinese Academy of Science, Beijing, China

Physalis angulata is a renowned traditional Chinese medicine for the treatment of various conditions. Physalin is the major type of bioactive constituents conferring medicinal properties of *P. angulata*. Despite the medicinal importance, the pathways leading to physalin are largely unknown. In this study, we employed a transcriptomic approach to identify a *Pa24ISO* gene from *P. angulata*. Through heterologous expression in yeast, Pa24ISO was revealed to catalyze an isomerization reaction in converting 24-methylenecholesterol to 24-methyldesmosterol. Real-time PCR analysis showed that the abundance of *Pa24ISO* transcripts correlated with the accumulation pattern of physalin B in different tissues of *P. angulata*. A direct role of Pa24ISO in channeling of 24-methylenecholesterol for physalin B biosynthesis was illustrated by suppressing the gene in *P. angulata via* the VIGS approach. Down-regulation of *Pa24ISO* led to reduced levels of 24-methyldesmosterol and physalin B, accompanied with an increase of campesterol content in *P. angulata*. The results supported that 24ISO is involved in physalin biosynthesis in plants.

Keywords: ***Physalis angulata*, physalin, 24-methyldesmosterol, 24ISO, campesterol**

OPEN ACCESS

Edited by:
Sheng-Xiong Huang, Kunming Institute of Botany (CAS), China

Reviewed by:
Edmundo Lozoya-Gloria, Unidad Irapuato (CINVESTAV), Mexico
Zhihua Liao, Southwest University, China

****Correspondence:***
Yansheng Zhang
zhangys1@shu.edu.cn

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

Received: *08 January 2022*
Accepted: *10 February 2022*
Published: *04 March 2022*

Citation:
Yang J, Tian J, Yang Y, Zhu Y, Li C and Zhang Y (2022) RNAi of Sterol Δ24-Isomerase Implicated Its Involvement in Physalin Biosynthesis in Physalis angulata L. Front. Plant Sci. 13:850711. doi: 10.3389/fpls.2022.850711

INTRODUCTION

Physalis angulata L, popularly known as Kuzhi in Chinese, has been prescribed for centuries for treatments of a variety of diseases, such as fever, malaria, liver disorders, and diabetes (Abe et al., 2006; Damu et al., 2007; Reyes-Reyes et al., 2013). Currently, crude extracts of *P. angulata* are reported to show antinociceptive (Choi and Hwang, 2003), anti-breast cancer (Hsieh et al., 2006), and anti-inflammatory activities (Bastos et al., 2008). The medicinal effects of *P. angulata* are mainly due to the presence of polyoxygenated steroids with an ergosterol backbone possessing a C-22, 26-lactone (Damu et al., 2007). Concerning modifications of their side chains, especially the cleavage of C13-C14 bond and the formation of a 16,24-carboncyclic bond, these steroid lactones can be classified as physalin and withanolide (**Figure 1A**; Huang et al., 2020; Xia et al., 2021). Of the *P. angulata* physalin, physalin B has received the most interest due to its diversely pharmacological activities.

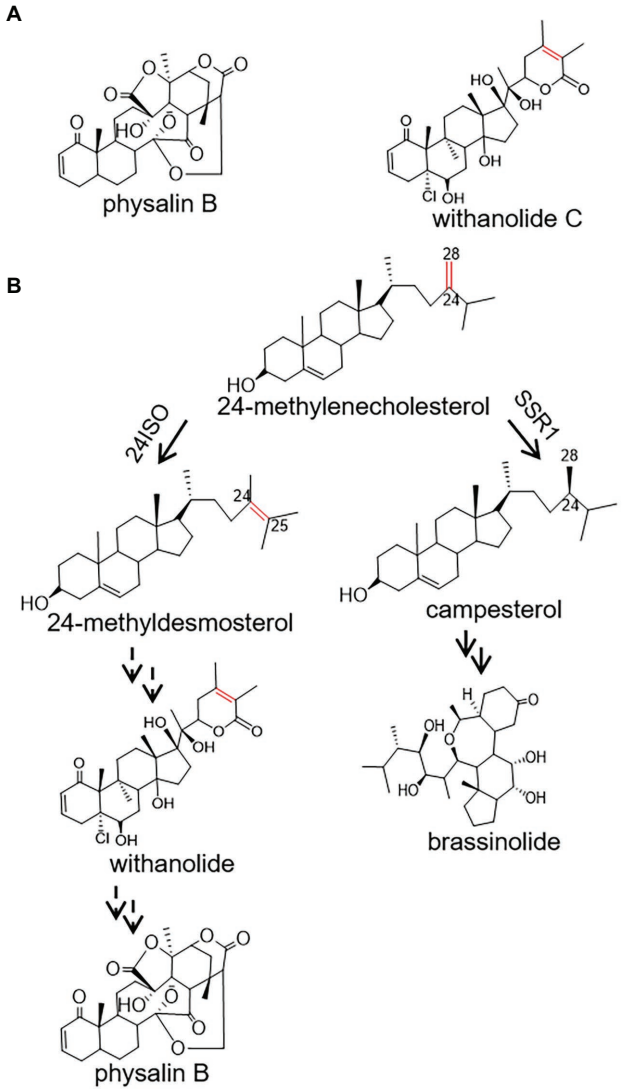


FIGURE 1 | A proposed pathway for the biosynthesis of physalin B in *Physalis angulata* starting from 24-methylenecholesterol. **(A)** The structures of withanolide C and physalin B; **(B)** Proposed biosynthetic routes to physalin B and brassinolide. Biosyntheses of physalin B and brassinolide diverge at the level of 24-methylenecholesterol, which is channeled to the production of physalin B through withanolide *via* 24-methyldesmosterol under the activity of 24ISO, or of brassinolide *via* campesterol, catalyzed by SSR1.

For example, it has been reported that physalin B showed significant activities in inhibition of several cancers, including breast (Wang et al., 2018) and prostate (Han et al., 2011) tumors.

Despite the pharmacological importance of physalin, the pathways leading to their biosyntheses are largely unknown. Physalin or withanolide are derived from a 24-alkyl sterol, and accordingly 24-methylenecholesterol is believed to be an intermediate in their biosynthetic steps (Pal et al., 2019). The 24-methylenecholesterol represents a branching point leading to biosynthesis of brassinolide *via* campesterol (i.e., 24-methyl cholesterol), catalyzed by sterol side chain reductase 1 (SSR1), or withanolide *via* 24-methyldesmosterol that is catalyzed by a sterol Δ^{24}-isomerase (24ISO; **Figure 1B**; Knoch et al., 2018). The C24-C25 double bond in the intermediate 24-methyldesmosterol (**Figure 1B**) is retained in accumulated withanolide (see the structure of a representative withanolide, withanolide C, in **Figure 1A**). On the contrary, the C24-C25 double bond is not present in physalin-type compounds (see the structure of a representative physalin, physalin B, in **Figure 1A**). Probably because of this, Wang et al. have proposed an intermediacy role of campesterol (see its structure in **Figure 1B**) in physalin biosynthesis (Zhan et al., 2018). However, when taking analogous structures of physalin and withanolide (**Figure 1A**) into consideration, this study contemplated that physalin may be derived from withanolide, and this C24-C25 double bond may be rearranged later in the conversion stage from withanolide to physalin. If this hypothesis is followed, the enzyme 24ISO (**Figure 1B**) must involve in physalin biosynthesis.

To investigate whether 24ISO is involved in physalin biosynthesis *in vivo*, here, we reported the cDNA cloning and functional analysis of a *24ISO* gene, designated as *Pa24ISO*, from *P. angulata*. The biochemical activity of Pa24ISO was confirmed by heterologous expression in yeast. To further study its function *in vivo*, we are the first to develop a virus-induced gene silencing (VIGS) system for *P. angulata*. A direct role of Pa24ISO in physalin biosynthesis was discovered by silencing it in *P. angulata* through the VIGS approach. The *Pa24ISO* expression pattern also matched well with the accumulation of physalin B across different tissues.

MATERIALS AND METHODS

Plant Materials

Seeds of *P. angulata* were harvested from the Langxi county, Anhui province of China in August 2018. The identity of the plant material was confirmed by Dr. Xiaodong Li at the Wuhan Botanical Garden, Chinese Academy of Science. *P. angulata* plants were grown in a growth chamber at 22°C under long-day (16 h of light/8 h of darkness) conditions. Transgenic plants were grown at an open field, located at the campus of Shanghai University, China.

Identification of *Pa24ISO* and Phylogenetic Analysis

Using *Ws24ISO* (GenBank accession: AXG64144.1) as a query, a BLASTP analysis of our recently constructed *P. angulata* calyx-derived transcriptome (unpublished data) revealed the candidate *Pa24ISO*. Deduced amino acid sequences of Pa24ISO were aligned with the previously reported sequences of SSR1 and SSR2 (see their accession numbers in **Supplementary Table 2**) using ClustalW software embedded in MEGA7. A maximum likelihood tree was inferred in MEGA7 using bootstrap method with1000 replications. *Arabidopsis* DWF1 was used to root the tree.

Establishment of a VIGS System for *Physalis angulata* and Generation of the *Pa24ISO*-Silenced Plants

The *PDS* (*phytoene desaturase*) gene, associated with the photobleaching phenotype, was used as a reporter gene to test the TRV-based VIGS system for *P. angulata*, The cDNA sequence of *PaPDS*, as shown in **Supplementary Figure 1**, was obtained from the *P. angulata* calyx transcriptome, and a 382 bp-fragment of *PaPDS* was amplified from the *P. angulata* calyx cDNA. For silencing the *Pa24ISO* gene expression, a 297 bp-region, which is specific to *Pa24ISO*, but is not conserved in the putative *PaSSR1* and *PaSSR2*, was selected (see their sequences in **Supplementary Figures 2–4**). All the primers used in this study are shown in **Supplementary Table 1** in the Supporting Information. The TRV system constitutes the pTRV1 and pTRV2 vectors, which represent two RNA strands of the TRV virus system. The amplified DNA fragments of targeted genes were cloned into the pTRV2 vector *via* restriction cloning sites, yielding the pTRV2-PaPDS and pTRV2-Pa24ISO plasmids.

For down-regulating the expression of *PaPDS* gene, the pTRV2 (empty vector), pTRV1, and pTRV2-PaPDS were separately transformed into the *Agrobacterium* GV3101. The transgenic *Agrobacterium* strains were inoculated into LB medium supplemented with kanamycin at 50 mg/ml and rifampicin at 50 mg/ml at 28°C, and used for infection when OD_{600} value of the *Agrobacterium* culture reaching 1.5. For infiltration of the *P. angulata* leaf, the TRV1-contained *Agrobacterium* culture was mixed at a 1:1 ratio with the cultures containing pTRV2-PaPDS, as well as the empty vector pTRV2 as a control. To improve the infection efficiency, 0.01% Silwet L-77 was added into the *Agrobacterium* mixer. Four to six-leaf-stage plants were selected for the infection. The infected plants were kept in darkness for 48 h, and then grown at an open field at the campus of the Shanghai University, China. The phenotype was then evaluated after 6–8 weeks. The same procedure was applied to generate the *Pa24ISO*-silenced plants.

Quantitative Real-Time PCR Analysis

The Real-time PCR (qRT-PCR) was performed on Bio-Rad CFX96™ Real-Time PCR instrument (Bio-Rad, Inc., United States) using TransStart Green qPCR SuperMix (Transgen). Primers used in this article are listed in **Supplementary Table 1**. The PCR program consisted of an initial step of 94°C for 30 s; 40 cycles of 94°C for 5 s and 60°C for 30 s; and then a dissociation stage of 95°C for 10 s, 65°C for 5 s and 95°C for 5 s. Each gene was assessed at least three biological replicates. The relative expression levels of the genes were calculated by the $2^{-\Delta\Delta Ct}$ method.

Extraction of Physalin and UPLC–MS Analysis

Every 1 g of powdered *P. angulata* samples was extracted with 1.7 ml of 90% methanol by sonication for 30 min, and this procedure was repeated three times. To normalize variations between the different extractions, isofraxidin at a final concentration of 1.5 μg/ml was added as an internal standard. The solvent extracts were collected by centrifugation at 12,000 *g* for 20 min, and passed through a membrane filter (0.22 mm) prior to the Ultra Performance Liquid Chromatography–Tandem Mass Spectrometry (UPLC–MS) analysis.

One microliter of the extracts was injected for the UPLC-MS analysis. LC-MS analysis was performed using a Q-Exactive Focus mass spectrometer, coupled with a VanquishTM UPLC system (Thermo Fisher Scientific, Bremen, Germany) and a HESI source (Thermo Fisher Scientific). The column (100 mm × 2.1 mm, 1.8 μm) was used to separate the sample, the column temperature was 40°C, and the flow rate was 0.35 ml/min. The mobile phases contain 0.1% formic acid (solvent A) and acetonitrile (solvent B), and the solvent gradient is set as follows: 15% solution B (0–1.0 min), 20–25% solution B (2.0–13.0 min), 25–40% solution B (13.0–25.0 min), 40–90% solution B (25–26 min), 90% solution B (26.0–29.0 min), and 15% solution B (29.0–33.0 min). The MS detection was performed with a full MS mode. The parameters of the mass spectrometers were as follows: spray voltage, 3.2 kV; source capillary temperature, 320°C; sheath gas flow rate (nitrogen), 25 ml/min; Aux gas flow rate (nitrogen), 8 ml/min; Aux gas heater temperature, 30°C, Scan range 150.0–700.0 m/z.

Heterologous Expression of *Pa24ISO* in Yeast

To optimize the expression of *Pa24ISO* in yeast, the full-length sequences of *Pa24ISO* were manually synthesized following the *S. cerevisiae* codon usage preference (see the optimized *Pa24ISO* cDNA sequence in **Supplementary Figure 5**), and inserted into a yeast expression vector pESC-HIS3 at *BamHI*/*SalI* sites, yielding the construct pESC-HIS3-Pa24ISO. The plasmid pESC-HIS3-Pa24ISO was transferred into a 24-methylenecholesterol-producing yeast strain YS11, which was recently developed by our group (Yang et al., 2021). As a negative control, the empty vector pESC-HIS3 was also transformed. The transgenic yeast colonies were selected on selection medium omitting histidine. Expression of the transformed gene was induced by 2% galactose.

To extract sterols from the transformed yeast cultures, yeast cells were pelleted from 2 ml of the yeast culture, broken with acid-washed glass beads, and extracted with 500 μl of 20% potassium hydrate in methanol (w/v) at 90°C for 2 h. When the saponified sample was cooled down, it was extracted with 2 ml of hexane for 30 min, and then was repeatedly extracted two times with another 1 ml hexane. The pooled hexane extracts were then dried in a vacuum freeze drier, and derivatized in 50 μl of N,O-bis(trimethylsilyl)-trifluoroacetamide (BSTFA) at 60°C for at least 1 h for sterol analysis by GC–MS.

Phytosterol Extraction and Analysis by GC–MS

Phytosterol was extracted as described previously (Knoch et al., 2018) with small modifications. One gram of powered sample was extracted with 3 ml chloroform/methanol (2:1, v/v) at 75°C for 1.5 h. After removing the solvent, the dried residue was saponified in 1 ml 10% (w/v) KOH in MeOH for 1 h at 80°C. When the samples were cooled to room temperature, 500 μl of H_2O and 500 μl of hexane were added, and extracted with shaking for 30 min. After centrifugation, the upper clear hexane extract was collected. The lower aqueous phase was then extracted

twice with an additional 500 µl of hexane. The pooled hexane extracts were evaporated to dryness, and treated with 50 µl of the silylating reagent BSTFA [bis(trimethylsilyl) trifluoroacetamide]. The trimethylsilyl esters were dissolved in 80 µl of hexane, and a 1-µl sample was used for GC–MS analysis. For quantifying the sterol content, 40 µg of cholesterol (YuanYe, Shanghai, China) was added to each sample as an internal standard. GC–MS analysis was performed using a Shimadzu 2010 Plus GC machine coupled to a QP2020 Ultra mass spectrometer (Shimadzu Corporation, Canby, OR, United States). Samples were separated using a RTX-5 MS column (30 m × 0.25 mm × 0.25 mm) with helium being used as a carrier gas at a flow rate of 1.0 ml/min. The injection temperature was 280°C. The GC oven temperature program was as follows: 90°C for 1 min, then increased by 30°C/min to 300°C and hold at 300°C for 15 min. The mass of sterol compounds was detected in a select ion-scan mode with electronic ionization (70 eV), and monitored at multiple ions of *m/z* 343, 382, and 472 for campesterol, *m/z* 341, 365, and 470 for 24-methylenecholesterol, *m/z* 365, 386, and 470 for 24-methyldesmosterol, and *m/z* 329, 368, and 458 for the internal standard cholesterol.

RESULTS

Identification of the *24ISO* Gene From *Physalis angulata*

Firstly we have shown that physalin B accumulates to higher levels in leaf, compared to other tissues of *P. angulata* (**Figure 2A**). This result is different from a previous report (Huang et al., 2014), in which calyx was identified to be the physalin-accumulating site with the highest amounts. This discrepancy is probably due to the difference of the *P. angulata* cultivars utilized between this and that studies. The identity of physalin B was determined by comparing its MS/MS fragmented pattern (**Supplementary Figure 7**) with that published in a previous literature (Huang, 2014). A *Ws24ISO* gene was previously cloned from *Withania somnifera* (Knoch et al., 2018). In this work, using Ws24ISO as a query, we performed a BLASTP analysis against the transcriptome derived from *P. angulata* calyx that was recently constructed by our group. One unigene (ID number: PB.4414.9), hereafter refer to as Pa24ISO, was found to show the highest amino acid identity (96.0%) with Ws24ISO. A phylogenetic analysis shows that Pa24ISO closely clustered with Ws24ISO in the tree (**Supplementary Figure 7**), indicating that it is a potential 24ISO candidate. The expression of *Pa24ISO* was profiled in different organs with real-time RT-PCR. *Pa24ISO* transcripts were detected in the highest levels in the leaf of *P. angulata*, but in much less levels in other tissues (**Figure 2B**). The expression pattern of *Pa24ISO* matched well with the accumulation profile of physalin B across these tissues (**Figure 2A**).

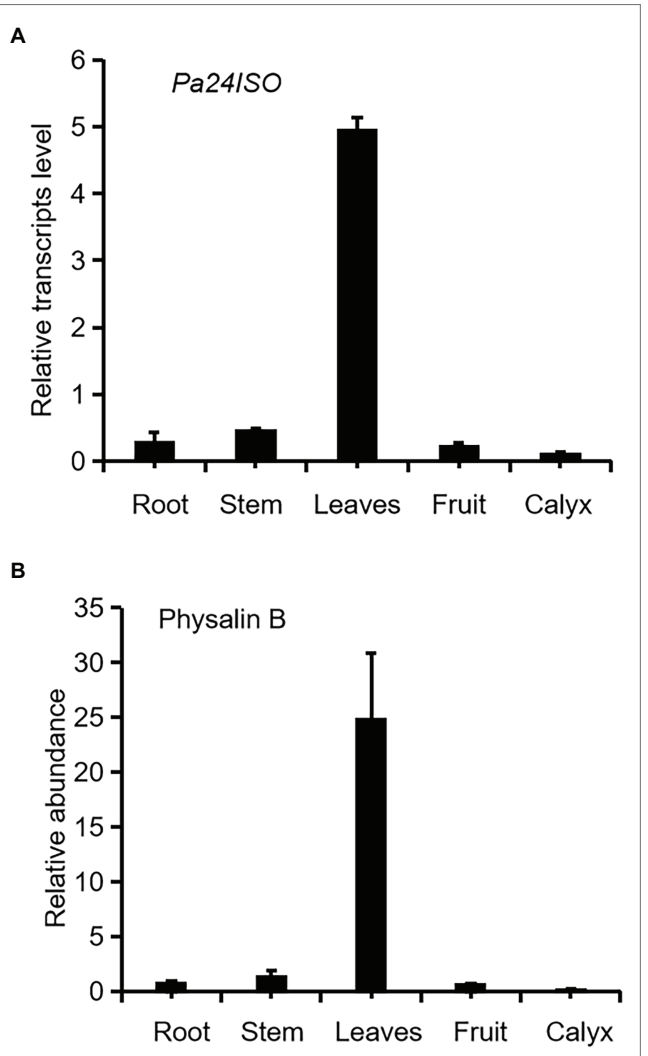


FIGURE 2 | Transcript abundance of *Pa24ISO* **(A)** correlates with the accumulation pattern of physalin B **(B)** in different tissues of *P. angulata*. The tissues, including root, stem, leaf (the fifth leaf materials from the top), fruit, and calyx, were collected from four and half-month old *P. angulata* plants for the analyses. At this stage, the calyx was in the size of about 0.7 cm of diameter. Means and standard errors are shown ($n = 3$).

Pa24ISO Converts 24-Methylene Cholesterol to 24-Methyldesmosterol

To test the enzymatic function of Pa24ISO, the *Pa24ISO* gene was codon-optimized by manually chemical synthesis for preferential expression in *Saccharomyces cerevisiae*, and inserted into a yeast expression vector pESC-HIS3. As a positive control, the known gene *Ws24ISO* was also synthesized, and expressed in the same way as *Pa24ISO*. *Pa24ISO* or *Ws24ISO* was expressed in a 24-methylenecholesterol-producing yeast strain YS11, which was previously made by Yang et al. (2021). The yeast extracts were subsequently prepared, and analyzed by GC–MS for the presence of 24-methyldesmosterol. The negative control was prepared by transferring the empty vector pESC-HIS3 into the YS11 strain.

Cells expressing *Pa24ISO* showed a clear GC–MS product peak, which was missing from the empty vector control. This new GC–MS peak showed the same retention times (**Figure 3A**) and the same mass spectra with the 24-methyldemosterol product synthesized by Ws24ISO (**Figure 3B**). Thus, it is

confirmed that the protein encoded by *Pa24ISO* is able to convert 24-methylenecholesterol to 24-methyldesmosterol (see the proposed reaction in **Figure 3C**).

Development of a VIGS System for *Physalis angulata*

To test functions of genes *in vivo* in *P. angulata*, it is necessary to establish a virus-induced gene silencing (VIGS) system for *P. angulata*, as there have been no reports of VIGS in this plant. For this purpose, the tobacco rattle virus (TRV; Burch-Smith et al., 2006) vector based VIGS system was employed, using *phytoene desaturase* (*PDS*; Cunningham and Gantt, 1998) as a reporter gene. The sequence of *P. angulata PDS* (*PaPDS*) was retrieved from the calyx transcriptome, which is available in our laboratory, and a 382 bp-fragment of *PaPDS* was selected as a target sequence. Six-to-eight week-old seedlings were infected with the *Agrobacterium* slurry containing the pTRV2-PaPDS vector or the empty pTRV2 vector as a control. At around 4 weeks post infiltration, a greenish–whitish variegated phenotype was observed in the upper leaves. With the growing of the infected seedlings, some of the newly emerging leaves exhibited a completely white phenotype (**Figure 4A**). Interestingly, at around 10–12 weeks post infiltration, a complete white phenomenon was also observed in the newly developed flowers and calyx of some of the infected plants (**Figure 4A**). We harvested the seeds that were generated from the plants showing completely white flowers or calyx, and grew them in soils to see whether the photobleaching phenomenon can be inherited in their offspring. However, they grew as wild type plants, and no photobleaching phenomenon was observed in their newly grown leaves.

To check whether the white phenotype correlated with the decreased *PaPDS* transcripts by VIGS, the *PaPDS* expression was monitored by real-time PCR. As shown in **Figure 4B**, the transcript levels of *PaPDS* were largely decreased in the photobleached leaf, flower and calyx, respectively, by the *PaPDS*-VIGS, compared with their corresponding empty vector controls. This result suggests that the TRV-based VIGS system can be used to suppress expression of genes of interest in *P. angulata*.

The silencing of *Pa24ISO* Led to Reduced Levels of Physalin B in *Physalis angulata*

After successfully establishing the VIGS system for *P. angulata*, we embarked experiments to silence the expression of *Pa24ISO in vivo*. Seven-week old plants were separately infiltrated with each of the *Agrobacterium* strains, which carried pTRV2-Pa24ISO,

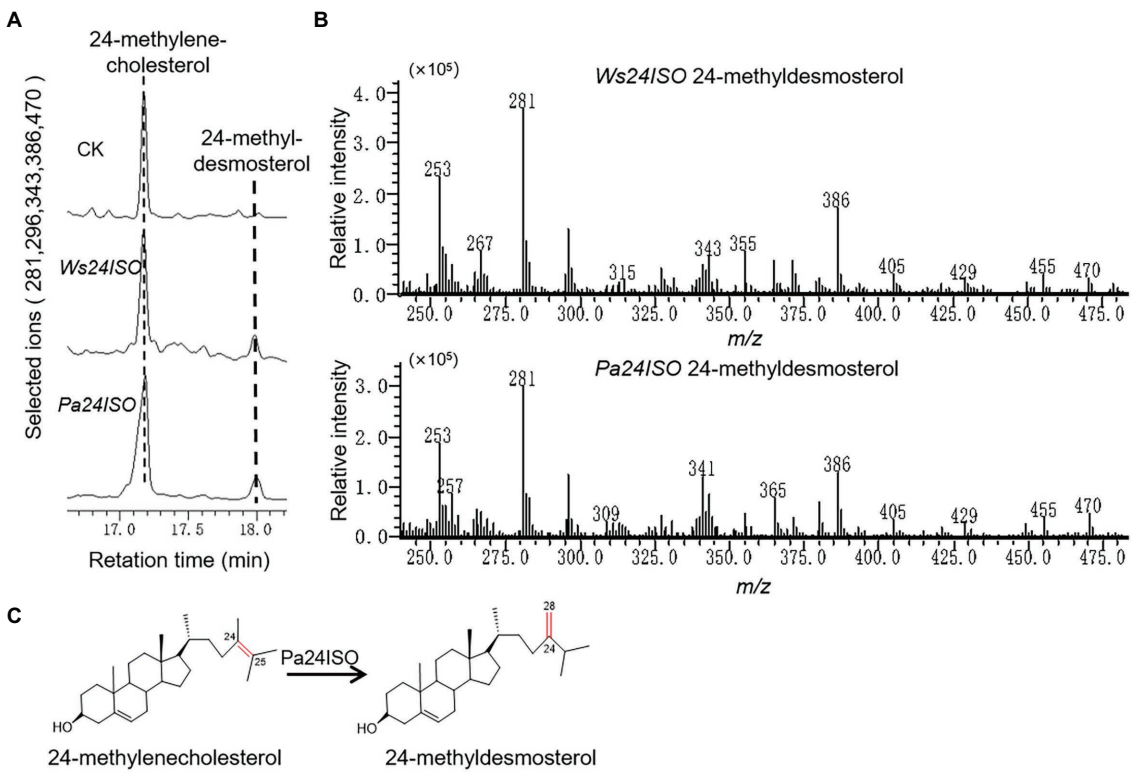


FIGURE 3 | Enzymatic characterization of Pa24ISO in a 24-methylenecholesterol-producing yeast. **(A)** GC–MS chromatograms of yeast extracts from the yeast strain YS11 expressing *Pa24ISO*, *Ws24ISO* or empty vector. The 24-methyldesmosterol product was detected in the cells expressing *Pa24ISO* or *Ws24ISO* but not in the cells harboring the empty vector. **(B)** Mass spectra of the 24-methyldesmosterol product formed by Pa24ISO or Ws24ISO. **(C)** The reaction scheme showing the activity of Pa24ISO in converting 24-methylenecholesterol to 24-methyldesmosterol.

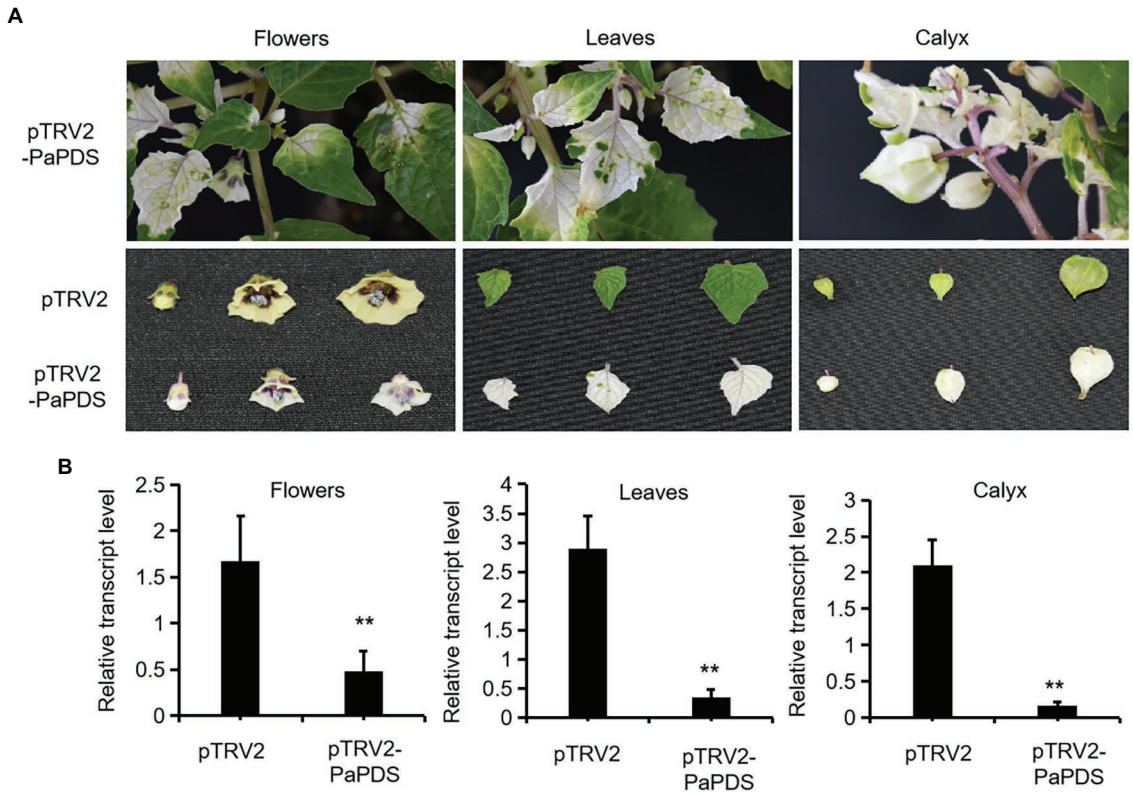


FIGURE 4 | Silencing of *PaPDS* in *P. angulata* using the TRV2-based VIGS. **(A)** Representative photobleaching phenotype of flower, leaf and calyx tissues from the *PaPDS*-silenced plants. The upper row shows the intact *PaPDS*-silenced plants, and the below row shows the detached various tissues from the *PaPDS*-silenced plants, compared to those from the empty vector-transferred control. **(B)** qRT-PCR analysis of *PaPDS* in various tissues of *P. angulata* inoculated with pTRV2-PaPDS and the empty vector pTRV2. The error bars represent ± SE of three independent experiments (**$p \leq 0.01$ by student t test).

pTRV2-PaPDS, or pTRV2 empty vector, respectively. The pTRV2-PaPDS construct was utilized, as an indicator group, to monitor the silencing efficiency under different time lengths post the inoculation. Based on the visible phenomenon by pTRV2-PaPDS, the highest silencing efficiency was achieved at 8 weeks post the infection. Thus, after 8 weeks, the newly developed leaves from the infected plants transformed with the pTRV2-Pa24ISO, as well as the empty vector pTRV2, were harvested. To evaluate the silencing efficiency of *Pa24ISO*, quantification of the transcript level of *Pa24ISO* was performed with real-time PCR. In comparison with the control plants, which were infiltrated with the empty vector, the transcript level of *Pa24ISO* was down-regulated by 46% in the *Pa24ISO*-silenced plants (**Figure 5A**). It should be noted that the silencing of *Pa24ISO* also led to a decrease in the expression of a putative *PaSSR1*, however, it was not statistically significant (**Supplementary Figure 8**). Consistent with the reduced levels of *Pa24ISO* transcript, the *Pa24ISO*-silenced plants showed a sharp decline (54.3%) in the content of 24-methyldesmosterol, which was accompanied with increased accumulation of 24-methylenecholesterol and campesterol (**Figure 5B** and **Supplementary Figure 8**), when compared to the empty vector control samples. This result demonstrates that Pa24ISO is indeed the branch enzyme diverting the common substrate 24-methylenecholesterol toward biosynthesis of 24-methyldesmosterol-derived sterols, in competition with the campesterol-mediated carbon flux, as depicted in **Figure 1B**.

To investigate the effect of *Pa24ISO* silencing on biosynthesis of physalin, we quantified the content of physalin B in the newly developed leaves of the infected plants, through LC–MS/MS anlaysis. Physalin B appeared to be the major physalin compound in the *P. angulata* cultivar used in this study. The *Pa24ISO*-silenced plants displayed a 47% reduction in physalin B (**Figure 5B**), as compared with control leaves, indicating that *Pa24ISO* play a pivotal role in biosynthesis of physalin in *P. angulata*.

DISCUSSION

Physalin are the characteristic constituents identified in *P. angulata*, and they exhibit a broad spectrum of medicinal properties (Damu et al., 2007). The pathways leading to physalin are largely unknown. Since physalin is a class of 24-alkyl sterols, thus it is believed that they are undoubtedly synthesized from 24-methylenecholesterol. As depicted in **Figure 1B**, 24-methylenecholesterol is the branching point, which can be either directed into 24-methyldemosterol-mediated

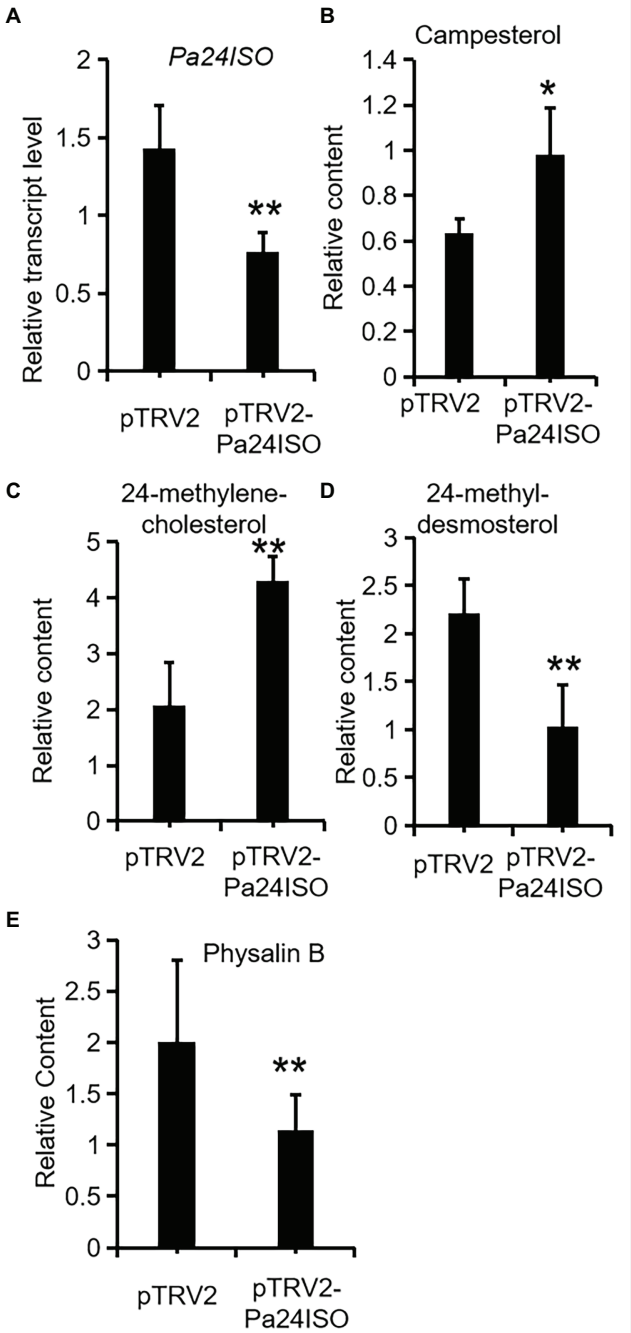


FIGURE 5 | Downregulation of *Pa24ISO* resulted in reduced physalin B levels in *P. angulata*. The *Pa24ISO*-silenced plants (pTRV2-Pa24ISO) were compared to the empty vector-transferred plants (pTRV2), in analysis of expression of *Pa24ISO* **(A)**, and the relative abundances of campesterol **(B)**, 24-methylenecholesterol **(C)**, 24-methyldesmosterol **(D)**, and physalin B **(E)**. The error bars represent ± SE of three or six independent experiments (**$p \leq 0.01$, *$p \leq 0.05$; student *t* test).

flux for withanolide *via* 24ISO (Knoch et al., 2018), or into campesterol-mediated flux for regular phytosterols *via* SSR1. Apparently, 24ISO competes with SSR1 at the level of 24-methylenecholesterol in the steps leading to withanolide production (Knoch et al., 2018). To date, there are no reports of physalin biosynthetic pathways in literatures. However, two possible pathways below were previously predicted by other groups: (1) physalins are biosynthesized from withanolides and (2) physalins are formed through the campesterol-pathway. The route 1 was recently proposed primarily based on a chemical viewpoint (Araki et al., 2021), in which withanolides can be transformed into compounds with a physalin-skeleton, through multiple oxidations. The route 2 was suggested, because that the withanolide content in *Withania somnifera* L was increased by feeding campesterol (Singh et al., 2014). For either route, there is no direct experimental evidence provided so far. For this reason, this project was designed with an aim to provide insight of which route that really occurs in nature for physalin production, through the silencing of *Pa24ISO* gene in *P. angulata*. This study is the first to isolate the *Pa24ISO* gene from *P. angulata*. Upon expression of *Pa24ISO* in a 24-methylenecholesterol-producing yeast strain YS11 (Yang et al., 2021), 24-methylenecholesterol was isomerized to 24-methyldesmosterol (**Figure 3**), implying that it is a functional 24ISO in *P. angulata*. The silencing of *Pa24ISO* was significant, with its transcript level being almost halved, as compared to the empty vector-transferred plants (**Figure 5A**). Physalin B was selected for the analysis, as it is present in high amounts in the *P. angulata* cultivar of this study, and could be reliably identified by comparing MS/MS fragmentation pattern with literature data (Huang, 2014). Compared to the empty vector controls, the *Pa24ISO* silencing led to significantly reduced level of physalin B (**Figure 5**). We also observed a 63% reduction for one unknown compound in the *Pa24ISO*-silenced plants (data not shown). Based on its mass spectra (**Supplementary Figure 6**), the molecular mass of this unknown compound matched that of withanolide C. If it was the withanolide C, the data that we have acquired is consistent with the previous finding, which clearly concluded that 24ISO is the first committed enzyme in the steps beyond 24-methylenecholesterol toward withanolide biosynthesis (Knoch et al., 2018). Therefore, it is highly likely that physalin is biogenetically synthesized from withanolide, and 24ISO is involved in physalin biosynthesis. An involvement of Pa24ISO in physalin biosynthesis could also be supported by a positive correlation of its transcripts with the accumulation profile of physalin B across different tissues of *P. angulata* (**Figure 2**). Taken together, these data corroborate that 24ISO plays a pivotal role in regulating physalin biosynthesis *in vivo*.

Upon the *Pa24ISO* silencing, physalin B and the intermediate 24-methyldesmosterol displayed a concurrently decreased accumulation pattern, whereas the change trend of campesterol was the reverse (**Figure 5**). This data further favors the proposal that physalin is biosynthesized through the 24ISO-catalyzed 24-methyldesmosterol pathway, while not through the SSR1-catalyzed campesterol (**Figure 1B**). The data of this study seem to be contradictory with a previous report (Singh et al., 2014), in which increased withanolide contents were found in *W. somnifera* plants fed with campesterol. However, campesterol is generally a precursor of

C28-brassinosteroids (BRs; Bajguz et al., 2020), and feeding of campesterol would modulate levels of BRs, which, as signaling compounds (Kim and Russinova, 2020), may in turn regulate expression of withanolide biosynthesis genes. One might argue that the intermediate 24-methyldesmosterol is also the intermediate in the SSR1-catalyzed two-step conversions of 24-methylene cholesterol to campesterol (**Figure 1B**; Klahre et al., 1998; Takahashi et al., 2006), and physalin biosynthesis enzymes may also have access to the 24-methyldesmosterol produced by SSR1. However, although both 24ISO and SSR1 may co-localize to the same subcellular site (i.e., endoplasmic reticulum; Silvestro et al., 2013; Knoch et al., 2018), the 24-methyldesmosterol formed by SSR1 seems to be quickly subjected to the next reduction step instead of being released by SSR1. For example, expression of *Physalis alkekengi* SSR1 in a 24-methylenecholesterol-producing yeast strain led to the production of only campesterol with no trace of 24-methyldesmosterol (Knoch et al., 2018).

In short, this study provides experimental evidence to support that 24ISO is involved in physalin biosynthesis. Elucidation of the physalin biosynthetic pathway needs further isolation of the genes in the later steps beyond *24ISO*, such as the genes encoding cytochrome P450s, which probably catalyze multiple oxidations of the 24ISO product 24-methyldesmosterol to form the physalin.

CONCLUSION

This study reports the isolation and functional analysis of a 24ISO-encoding gene (*Pa24ISO*) from *P. angulata*. Heterologous expression of *Pa24ISO* in yeast suggested that Pa24ISO is able to convert 24-methylenecholesterol to 24-methyldesmosterol. Silencing of *Pa24ISO* in *P. angulata* by the VIGS approach revealed a direct role of Pa24ISO in biosynthesis of physalin, such as physalin B at this case.

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

YZ received the funding to support this project. YZ and JY contributed to the writing. JY performed the experiments. CL provided the calyx transcriptome database and conducted the LC–MS/MS analysis. JT, YY, and YZ provided assistances in gene cloning, construct preparation, and chemical analysis. All authors contributed to the article and approved the submitted version.

FUNDING

The work was jointly supported by a grant from the National Key R&D Program of China (2018YFC1706200) and the start-up funding grant to YZ from Shanghai University (N.13-G210-20-327).

ACKNOWLEDGMENTS

We thank Xiaodong Li, who is from the Wuhan Botanical Garden, Chinese Academy of Sciences, for the plant identification.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.850711/full#supplementary-material

REFERENCES

Abe, F., Nagafuji, S., Okawa, M., and Kinjo, J. (2006). Trypanocidal constituents in plants 6. 1 minor withanolides from the aerial parts of *Physalis angulata*. *Chem. Pharm. Bull.* 54, 1226–1228. doi: 10.1248/cpb.54.1226

Araki, Y., Nakazaki, A., and Nishikawa, T. (2021). Biomimetic synthesis of the CDE ring moiety of Physalins, complex 13,14-Secosteroids. *Org. Lett.* 23, 989–994. doi: 10.1021/acs.orglett.0c04198

Bajguz, A., Chmur, M., and Gruszka, D. (2020). Comprehensive overview of the Brassinosteroid biosynthesis pathways: substrates, products, inhibitors, and connections. *Front. Plant Sci.* 11:1034. doi: 10.3389/fpls.2020.01034

Bastos, G. N., Silveira, A. J., Salgado, C. G., Picanco-Diniz, D. L., and do Nascimento, J.L., (2008). *Physalis angulata* extract exerts anti-inflammatory effects in rats by inhibiting different pathways. *J. Ethnopharmacol.* 118, 246–251. doi: 10.1016/j.jep.2008.04.005

Burch-Smith, T. M., Schiff, M., Liu, Y., and Dinesh-Kumar, S. P. (2006). Efficient virus-induced gene silencing in *Arabidopsis*. *Plant Physiol.* 142, 21–27. doi: 10.1104/pp.106.084624

Choi, E. M., and Hwang, J. K. (2003). Investigations of anti-inflammatory and antinociceptive activities of *Piper cubeba*, *Physalis angulata* and *Rosa hybrida*. *J. Ethnopharmacol.* 89, 171–175. doi: 10.1016/s0378-8741(03)00280-0

Cunningham, F. X., and Gantt, E. (1998). Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 557–583. doi: 10.1146/annurev.arplant.49.1.557

Damu, A. G., Kuo, P. C., Su, C. R., Kuo, T. H., Chen, T. H., Bastow, K. F., et al. (2007). Isolation, structures, and structure - cytotoxic activity relationships of withanolides and physalins from *Physalis angulata*. *J. Nat. Prod.* 70, 1146–1152. doi: 10.1021/np0701374

Han, H., Qiu, L., Wang, X., Qiu, F., Wong, Y., and Yao, X. (2011). Physalins A and B inhibit androgen-independent prostate cancer cell growth through activation of cell apoptosis and downregulation of androgen receptor expression. *Biol. Pharm. Bull.* 34, 1584–1588. doi: 10.1248/bpb.34.1584

Hsieh, W. T., Huang, K. Y., Lin, H. Y., and Chung, J. G. (2006). Physalis angulata induced G2/M phase arrest in human breast cancer cells. *Food Chem. Toxicol.* 44, 974–983. doi: 10.1016/j.fct.2005.11.013

Huang, C. (2014). The use of UPLC-MS/MS in rapid characterization of physalins in Ph.alkekengi L.var.franchetii and in quality evalutation of the herb[D]. Master thesis, Huazhong University of Science and Technology. doi: 10.7666/d.D613074

Huang, M., He, J. X., Hu, H. X., Zhang, K., Wang, X. N., Zhao, B. B., et al. (2020). Withanolides from the genus Physalis: a review on their phytochemical and pharmacological aspects. *J. Pharm. Pharmacol.* 72, 649–669. doi: 10.1111/jphp.13209

Huang, C., Xu, Q., Chen, C., Song, C., Xu, Y., Xiang, Y., et al. (2014). The rapid discovery and identification of physalins in the calyx of *Physalis alkekengi* L.var.franchetii (mast.) Makino using ultra-high performance liquid chromatography-quadrupole time of flight tandem mass spectrometry together with a novel three-step data mining strategy. *J. Chromatogr. A* 1361, 139–152. doi: 10.1016/j.chroma.2014.08.004

Kim, E. J., and Russinova, E. (2020). Brassinosteroid signalling. *Curr. Biol.* 30, R294–R298. doi: 10.1016/j.cub.2020.02.011

Klahre, U., Noguchi, T., Fujioka, S., Takatsuto, S., Yokota, T., Nomura, T., et al. (1998). The Arabidopsis DIMINUTO/DWARF1 gene encodes a protein involved in steroid synthesis. *Plant Cell* 10, 1677–1690. doi: 10.1105/tpc.10.10.1677

Knoch, E., Sugawara, S., Mori, T., Poulsen, C., Fukushima, A., Harholt, J., et al. (2018). Third DWF1 paralog in *Solanaceae*, sterol Delta(24)-isomerase, branches withanolide biosynthesis from the general phytosterol pathway. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8096–E8103. doi: 10.1073/pnas.1807482115

Pal, S., Rastogi, S., Nagegowda, D. A., Gupta, M. M., Shasany, A. K., and Chanotiya, C. S. (2019). RNAi of sterol methyl Transferase1 reveals its direct role in diverting intermediates towards Withanolide/Phytosterol biosynthesis in *Withania somnifera*. *Plant Cell Physiol.* 60, 672–686. doi: 10.1093/pcp/pcy237

Reyes-Reyes, E. M., Jin, Z., Vaisberg, A. J., Hammond, G. B., and Bates, P. J. (2013). Physangulidine A, a withanolide from *Physalis angulata*, perturbs the cell cycle and induces cell death by apoptosis in prostate cancer cells. *J. Nat. Prod.* 76, 2–7. doi: 10.1021/np300457g

Silvestro, D., Andersen, T. G., Schaller, H., and Jensen, P. E. (2013). Plant sterol metabolism. Delta(7)-sterol-C5-desaturase (STE1/DWARF7), Delta(5,7)-sterol-Delta(7)-reductase (DWARF5) and Delta(24)-sterol-Delta(24)-reductase (DIMINUTO/DWARF1) show multiple subcellular localizations in *Arabidopsis thaliana* (Heynh) L. *PLoS One* 8:e56429. doi: 10.1371/journal.pone.0056429

Singh, S., Pal, S., Shanker, K., Chanotiya, C. S., Gupta, M. M., Dwivedi, U. N., et al. (2014). Sterol partitioning by HMGR and DXR for routing intermediates toward withanolide biosynthesis. *Physiol. Plant.* 152, 617–633. doi: 10.1111/ppl.12213

Takahashi, K., Nasu, K., Mashino, T., Morisaki, M., Hara, N., and Fujimoto, Y. (2006). Metabolic conversion of 24-methyl-Delta25-cholesterol to 24-methylcholesterol in higher plants. *Bioorg. Med. Chem.* 14, 732–738. doi: 10.1016/j.bmc.2005.08.061

Wang, A., Wang, S., Zhou, F., Li, P., Wang, Y., Gan, L., et al. (2018). Physalin B induces cell cycle arrest and triggers apoptosis in breast cancer cells through modulating p53-dependent apoptotic pathway. *Biomed. Pharmacother.* 101, 334–341. doi: 10.1016/j.biopha.2018.02.094

Xia, G. Y., Cao, S. J., Chen, L. X., and Qiu, F. (2021). Natural withanolides, an update. *Nat. Prod. Rep.* doi: 10.1039/d1np00055a, [Epub ahead of print]

Yang, J., Li, C., and Zhang, Y. (2021). Engineering of Saccharomyces cerevisiae for 24-methylene-cholesterol production. *Biomol. Ther.* 11:1710. doi: 10.3390/biom11111710

Zhan, X., Liao, X., Luo, X., Zhu, Y., Feng, S., Yu, C., et al. (2018). Comparative Metabolomic and proteomic analyses reveal the regulation mechanism underlying MeJA-induced bioactive compound accumulation in Cutleaf Groundcherry (*Physalis angulata* L.) hairy roots. *J. Agric. Food Chem.* 66, 6336–6347. doi: 10.1021/acs.jafc.8b02502

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fpls.2022.861036

Differential Regulation of an OsIspH1, the Functional 4-Hydroxy-3-Methylbut-2-Enyl Diphosphate Reductase, for Photosynthetic Pigment Biosynthesis in Rice Leaves and Seeds

Yeo Jin Lee[1], Jae Kwang Kim[2], Seung-A Baek[2], Ji-Su Yu[1], Min Kyoung You[1] and Sun-Hwa Ha[1]**

[1] Department of Genetics and Biotechnology, Graduate School of Biotechnology, College of Life Sciences, Kyung Hee University, Yongin, South Korea, [2] Division of Life Sciences, Bio-Resource and Environmental Center, Incheon National University, Incheon, South Korea

OPEN ACCESS

Edited by:
Guodong Wang, Institute of Genetics and Developmental Biology (CAS), China

Reviewed by:
De-Yu Xie, North Carolina State University, United States
Raimund Nagel, Leipzig University, Germany

***Correspondence:**
Min Kyoung You
minkyou@khu.ac.kr
Sun-Hwa Ha
sunhwa@khu.ac.kr

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

Received: *24 January 2022*
Accepted: *10 March 2022*
Published: *13 April 2022*

Citation:
Lee YJ, Kim JK, Baek S-A, Yu J-S, You MK and Ha S-H (2022) Differential Regulation of an OsIspH1, the Functional 4-Hydroxy-3-Methylbut-2-Enyl Diphosphate Reductase, for Photosynthetic Pigment Biosynthesis in Rice Leaves and Seeds. Front. Plant Sci. 13:861036. doi: 10.3389/fpls.2022.861036

The methylerythritol 4-phosphate (MEP) pathway is responsible for providing common precursors for the biosynthesis of diverse plastidial terpenoids, including chlorophylls, carotenoids, and phytohormones, in plants. In rice (*Oryza sativa*), the last-step genes encoding 4-hydroxy-3-methylbut-2-enyl diphosphate reductase [HDR/isoprenoid synthesis H (IspH)] have been annotated in two genes (*OsIspH1* and *OsIspH2*) in the rice genome. The spatial transcript levels indicated that *OsIspH1* is highly expressed in all tissues at different developmental stages, whereas *OsIspH2* is barely expressed due to an early stop in exon 1 caused by splicing error. OsIspH1 localized into plastids and *osisph1*, a T-DNA inserted knockout mutant, showed an albino phenotype, indicating that *OsIspH1* is the only functional gene. To elucidate the role of *OsIspH1* in the MEP pathway, we created two single (H145P and K407R) and double (H145P/K407R) mutations and performed complementation tests in two *hdr* mutants, including *Escherichia coli* DLYT1 strains and *osisph1* rice plants. The results showed that every single mutation retained HDR function, but a double mutation lost it, proposing that the complementary relations of two residues might be important for enzyme activity but not each residue. When overexpressed in rice plants, the double-mutated gene, $OsIspH1^{MUT}$, reduced chlorophyll and carotenoid biosynthesis in the leaves and seeds. It confirmed the crucial role of *OsIspH1* in plastidic terpenoid biosynthesis, revealing organ-specific differential regulation of *OsIspH1* in rice plants.

Keywords: carotenoids and chlorophylls, duplicated pseudogene, functional complementation, methylerythritol 4-phosphate pathway, OsIspH1, OsIspH2, rice

INTRODUCTION

Plants synthesize an enormous variety of terpenoids that serve as growth regulators (cytokines, gibberellins, abscisic acids, strigolactones, and brassinosteroids), photosynthetic and respiratory components (chlorophylls, carotenoids, prenylquinones, and ubiquinone), membranous sterols, and other secondary metabolites (Nagegowda and Gupta, 2020). All terpenoids are derived by

the condensation of two universal precursors, namely, isopentenyl diphosphate (IPP, C5) and dimethylallyl diphosphate (DMAPP, C5). While most organisms retain one pathway for the biosynthesis of these isoprene building units, plants have two independent pathways, namely, the cytosolic mevalonate (MVA) and plastidic methylerythritol 4-phosphate (MEP) pathways (Hemmerlin et al., 2012). The latter has been evolutionally acquired due to endosymbiosis with prokaryotes and consists of seven enzymatic steps that start with the condensation of pyruvate (C3) and glyceraldehyde 3-phosphate (C3) derived from glycolysis into 1-deoxy-D-xylulose 5-phosphate (DXP, C5) by the action of DXP synthase (DXS) and end with the production of IPP and DMAPP from 4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP, C5) by the catalyst of HMBPP reductase (HDR), also known as isoprenoid synthesis H (IspH) or lysis-tolerant B (LytB; Vranova et al., 2013).

The MEP genes have been reported mainly in photosynthetic organisms, such as cyanobacteria, algae, and plants (Hemmerlin et al., 2012). However, Apicomplexa (malaria causative *Plasmodium* spp.), a non-photosynthetic protist, possesses MEP genes, although its vestigial chloroplasts and apicoplasts result from convergent reduction (Mathur et al., 2021). Bacterial MEP genes have also been discovered in *Eubacteria*, such as *Escherichia coli* and *Streptomyces* species, *Chlamydia* (obligate intracellular bacteria causing a sexually transmitted infection), and *Mycobacterium tuberculosis* (tuberculosis causative bacteria) (Rohmer et al., 1993; Grieshaber et al., 2004; Eoh et al., 2009). Therefore, MEP genes, including *IspH*, have been identified as promising targets for developing algicides, herbicides, anti-malaria drugs, and bactericides to combat infectious diseases, such as *Chlamydia* and tuberculosis (Singh et al., 2007; Masini and Hirsch, 2014; Wang and Dowd, 2018).

At present, several *IspH* genes have been characterized in diverse species, including cyanobacteria (*Synechocystis* strain PCC 6803, Cunningham et al., 2000), *E. coli* (Altincicek et al., 2002), *Aquifex* (*Aquifex aeolicus*, Wang et al., 2010), green microalgae (*Botryococcus braunii*, Uchida et al., 2018), ginkgo (*Ginkgo biloba*, Kim et al., 2008), red pine (*Pinus densiflora*, Kim et al., 2009), tobacco (*Nicotiana benthamiana*, Page et al., 2004), tomato (*Lycopersicon esculentum*, Botella-Pavia et al., 2004), *Arabidopsis* (*Arabidopsis thaliana*, Hsieh and Goodman, 2005), corn (*Zea mays*, Lu et al., 2012), danshen (*Salvia miltiorrhiza*, Hao et al., 2013), melon (*Cucumis melo*, Saladie et al., 2014), sweet wormwood (*Artemisia annua*, Ma et al., 2017), and rice (*Oryza sativa*, Liu et al., 2020). Among them, the protein structures were first examined in bacterial systems, such as *A. aeolicus* and *E. coli*, to discover *IspH* inhibitors that have potential for novel cancer immunotherapy because the substrate of *IspH*, HMBPP, has been known as a potent phosphoantigen that activates γδ T cells to kill tumor cells (Hintz et al., 2001; Grawert et al., 2004; Rao and Oldfield, 2016). IspH is identified as a 4Fe-4S cluster-containing protein with a trefoil-like structure, and its key residues are elucidated: the Fe-S cluster acts as an electron transfer cofactor in the active site, with direct involvement of the fourth Fe in substrate binding and catalysis, and it requires coordination with three Cys-residues for ligand binding (Wang et al., 2010; Hsieh et al., 2014; Rao and Oldfield, 2016).

In planta IspH functions have been reported for diverse terpenoid production in several plants. The overexpression of the *L. esculentum* HDR gene (*LeHDR*) increased carotenoids in tomato fruits during ripening and *Arabidopsis* seedlings during de-etiolation, and co-expression with taxadiene synthase gene increased taxadiene levels, indicating synergistic effects (Botella-Pavia et al., 2004). The overexpression of the *S. miltiorrhiza* HDR gene (*SmHDR1*) enhanced tanshinone production in cultured hairy roots of *S. miltiorrhiza* Bge. F. alba (Hao et al., 2013). The overexpression of the *A. annua* HDR gene (*AaHDR1*) increased the contents of artemisinin, arteannuin B, other sesquiterpenes, and monoterpenes, while its antisense-induced suppression exhibited opposite effects (Ma et al., 2017). Recently, the overexpression of the *G. biloba* HDR2 gene (*GbHDR2/GbIspH2*) in *Nicotiana tabacum* cv. Xanthi resulted in an increased photosynthetic rate through the upregulation of biosynthetic genes for photosynthetic pigments, including chlorophylls and carotenoids, as well as the elevated contents of duvatrienediol, the major diterpene of the *N. tabacum* leaf surface (Kim et al., 2021). However, the loss-of-function studies on *IspH* genes have been reported in several plants. Tobacco plants infected with the tobacco rattle virus posttranscriptionally silence the expression of several MEP genes, including *IspG*, *IspH*, and *IPI*, resulting in an albino phenotype (Page et al., 2004). T-DNA knockout mutant of *Arabidopsis*, *atisph*, exhibited an albino phenotype, whereas transgenic-induced gene silencing of *AtIspH* exhibited variegated albino phenotypes (Hsieh and Goodman, 2005). In the case of Z. may IspH (ZmIspH), an ethyl methanesulfonate (EMS)-induced recessive mutant and foxtail mosaic virus-induced gene silencing both showed albino phenotypes, suggesting that it is fatal to chlorophyll biosynthesis (Lu et al., 2012; Liu et al., 2016).

Interestingly, in the rice genome, the existence of two IspH alleles is predicted unlike in corn that has one *IspH* gene. Although a recent study using the CRISPR-Cas9 system presented an albino phenotype in seedlings caused by defective chloroplast biogenesis of a rice *lethal albinic seedling 1* (*las1*) mutant, corresponding to the *osisph1* (Liu et al., 2020), the rice *IspH* function has not been fully understood. In this study, we analyzed the molecular characteristics of *OsIspH1* and *OsIspH2*, performed the complementation assays of *OsIspH1* in two *hdr* mutant backgrounds, namely, *E. coli* DLYT1 strains and rice T-DNA insertional *osisph1* mutant plants, identified key residues crucial to HDR activity, and investigated the *IspH* effects on the biosynthesis of terpenoids, including chlorophylls and carotenoids, in the source and sink organs of rice plants.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Two commercial varieties of Korean rice, *O. sativa* L. cv. Dongjin and Ilmi, were used to analyze endogenous gene expression, amplify the genes, isolate protoplasts, and perform a stable transformation. The Crop Biotech Institute in Kyung Hee University (Yongin, South Korea) generously provided a T-DNA-insertional rice mutant. Mature seeds were dehulled, surface-sterilized by 70% ethanol for 2 min and 2% sodium

hypochlorite for 40 min, washed five times with distilled water, germinated for a week on Murashige and Skoog agar medium (Duchefa, Haarlem, Netherlands) in a plant growth chamber, transplanted into soil, and cultivated in a greenhouse under the conditions of 14-h light/10-h dark cycle at 28°C or in the paddy field until maturity during the summer. Mature seeds were harvested 60 days after flowering, and their endosperm color was visually compared after dehusking (TR-200 Electromotion rice husker, Kett, Tokyo, Japan) and polishing (Pearlest Polisher, Kett).

RNA and DNA Analysis

Rice samples of diverse organs obtained at different developmental stages and 2-week-old seedlings treated with six phytohormones, including 100-μM abscisic acid (ABA), 100-μM gibberellic acid (GA3), 100-μM auxin (IAA), 100-μM kinetin (KT), 100-μM methyl jasmonic acid (MeJA), and 200-μM salicylic acid (SA), were used for total RNA extraction using mainly RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and PureLink® Plant RNA Reagent (Invitrogen, Waltham, MA, United States) for seed case. The real-time quantitative reverse transcriptase PCR (qRT-PCR) was performed using iQ™ SYBR® Green Supermix (Bio-Rad) and CFX Connect™ Real-Time System (Bio-Rad), with PCR conditions and data calculated relative to the rice ubiquitin 5 gene (Os01g22490) to normalize RNA amounts, as previously described (You et al., 2020), and each qRT-PCR was performed in three technical repeats. Semiquantitative RT-PCR was performed under the following conditions: a cycle at 95°C for 3 min, 30 cycles at 95°C for 15 s, 55°C for 15 s, 72°C for 1 min/kb, followed by 3 min at 72°C with KOD FX DNA polymerase (Toyobo, Osaka, Japan).

We extracted the rice genomic DNAs using a Genomic DNA Preparation Kit (QIAGEN). The genotyping PCRs were performed using an EmeraldAmp® PCR Master Mix (TaKaRa Bio, Shiga, Japan). To examine the transgene homozygosity, TaqMan PCR was performed to detect *Bar* as a transgene with an internal reference in the rice genome, i.e., the α-tubulin gene (Os11g14220), as previously described (Ha et al., 2019). **Supplementary Table 10** contains all sequence information of primers.

Cloning and Subcellular Localization Using Rice Protoplasts

The coding region of *OsIspH1* (Os03g0731900) was amplified using gene-specific primer pair from the total RNA of 10-day-old seedlings of rice (*O. sativa* L. cv. Ilmi) and cloned into the *pDONR221* vector using the Gateway® BP® Clonase II Enzyme Mix (Invitrogen), yielding *pDONR221-OsIspH1*. To determine the subcellular localization, *OsIspH1* was fused to the N-terminus of superfolder green fluorescent protein (sGFP) and cloned into the *pB2GW7* binary vector using Gateway® LR Clonase® II Enzyme Mix (Invitrogen). Protoplast isolation, DNA transfection, and confocal microscopy analysis were conducted, as previously described (You et al., 2016). The fluorescence was detected and imaged using a confocal laser scanning microscope (Carl Zeiss LSM 880, Jena, Germany).

Site-Directed Mutagenesis and Complementation Assay in *E. coli* DLYT1 Strain

Using the *pDONR221-OsIspH1* as a template, two non-synonymous substitution mutations (H145P and K407R) in OsIspH1 were created to have one or both of each through site-directed mutagenesis with sequence-substituted primers (**Supplementary Table 10**). Original and mutated genes were cloned into *pMW118* (Nippon Gene, Tokyo, Japan) and transformed into the *E. coli* DLYT1 strain, which was disrupted in *LytB/EcIspH* encoding *E. coli* HDR gene (Kim et al., 2008). An *E. coli* DLTY1 strain transformed with *pTTQ18-EcIspH/LytB* was used as a positive control for complementation analysis. The complementation tests were performed on the Luria-Bertani (LB) medium containing 50 μg/ml kanamycin, 100 μg/ml spectinomycin, and 0.01% (w/v) (±)-mevalonolactone (MVA) (Sigma-Aldrich, St. Louis, MO, United States).

Transformation and Cross-Fertilization of Rice Plants

After amplification of a double mutated gene using the *pMW118-OsIspH1*$^{H145P/K407R}$ as a template, a constitutive expression cassette was prepared by consecutive cloning into the *pDONR221* vector and the *p600-PGD1* vector containing a rice phosphogluconate dehydrogenase 1 promoter using the Gateway procedures (Park et al., 2010), resulting in *pIspH1*^{MUT}. It was further introduced into *pGlb:stPAC* vector, in which endosperm specifically produces β-carotene (Jeong et al., 2017), resulting in *pIspH1*^{MUT}*-stPAC*. Two final binary vectors were transformed *via* triparental mating of *Agrobacterium* LBA4404 harboring *pSB1* plasmids and cocultivated with embryogenic calli differentiated from matured rice seeds (*O. sativa* L. cv. Ilmi), as previously described (You et al., 2020). The putative transgenic plants were generated on a selection medium containing phosphinothricin (4 mg/L) and cefotaxime (500 mg/L) using shooting and rooting procedures, according to a previously described method (Ko et al., 2018).

In planta complementation assay was performed using conventional interbreeding during the flowering season at the T3 generation between heterozygous plants of *osisph1* as the female parent and homozygous *pIspH1*^{MUT}-overexpressed plants as the male parent. Filial (F) 1 progenies were self-pollinated in the paddy field for three more generations to obtain homozygous seeds for transgene traits of both parents.

Analyses of Chlorophyll and Carotenoid Metabolites

Rice leaves and dehusked seeds were ground to a fine powder using a pestle in liquid nitrogen. Total chlorophylls were extracted from rice leaf powders with 100% methanol at 70°C for 30 min using ThermoMixer Comfort (Eppendorf AG, Hamburg, Germany) at 500 rpm and centrifuged for 10 min at 4°C and 3,000 rpm. A spectrophotometer was used to measure the absorbance of the supernatant at 666 and 653 nm (Mecasys Co., Daejeon, South Korea). The chlorophyll content was calculated using the formula stated by Wellburn (1994).

Carotenoids were extracted from 100 mg powders of each rice leaf and seed, following the previous report with minor modifications (Kim et al., 2010). Notably, 0.1% ascorbic acid-containing ethanol was added to 100 mg powder to be incubated at 85°C for 5 min, after vortex-mixing for 20 s. After 10 min of saponification with potassium hydroxide (80%, w/v) in an 85°C water bath, samples were immediately placed on ice, and we added cold deionized water with β-apo-8′-carotenal (25 μg/ml) as an internal standard. Hexane-extracted carotenoid layers were separated two times by centrifugation at 1,200 × *g*. Aliquots were dried under a nitrogen stream and redissolved in 50:50 (v/v) dichloromethane/methanol solution. High-performance liquid chromatography (HPLC) was performed on a C30 YMC column (4.6 mm × 250 mm, 3 μm; Waters Corporation, Milford, MA, United States) using gradient elution at 1 ml/min with solvent A [methanol/water (92.8, v/v) with 10 mM ammonium acetate] and B (100% methyl *tert*-butyl ether) under the following conditions: 0 min, 83% A/17% B; 23 min, 70% A/30% B; 29 min, 59% A/41% B; 35 min, 30% A/70% B; 40 min, 30% A/70% B; 44 min, 83% A/17% B; 55 min, 83% A/17% B. Chromatograms were generated at 450 nm. The calibration curves were set using 5 concentrations of carotenoid standards, except for lycopene, from 0.3 to 5.0 μg/ml based on the peak area ratios obtained with the internal standard. Linearity was tested by least-squares regression analysis of the corrected peak area ratios against increasing weight ratios. The calibration curves of lutein, zeaxanthin, α-carotene, and β-carotene were linear (r^2 = 0.9988–0.9998). Lycopene undergoes degradation *via* isomerization and oxidation. Thus, the quantitative calculation of lycopene was based on the peak area of 5.0 μg/ml of lycopene standard (Kim et al., 2017).

RESULTS

Phylogenetic Tree Analysis of IspH Family

To investigate the evolutionary relationship of the rice IspH family, the predicted amino acid sequences of 79 IspHs were first collected from GreenPhylDB[1], confirmed whether they contained a full-length open reading frame (ORF) in the National Center for Biotechnology Information Web BLAST[2], and finally selected for the phylogenetic tree analysis using the maximum-likelihood method (**Figure 1**). As a result, the tree classified them into six clades, namely, bacteria, mosses, ferns, gymnosperms, monocotyledons, and dicotyledons, as they showed a sequential relationship based on general evolutionary processes. All bacteria have a single *IspH* gene, whereas plants possess one to three copies. Considering the closer relationship between cyanobacteria and plants, it was proposed that plants obtain *IspH* genes from cyanobacteria *via* endosymbiosis, which might have evolved independently with non-photosynthetic bacteria from the common prokaryotic ancestor, as previously reported (Hsieh and Hsieh, 2015). Particularly, family genes belonging to the same species were mostly grouped into the same clade or closely related, indicating that they might be multiplicated during evolutionary processes.

In the *O. sativa* genome, two *IspH* genes with 10 introns, *OsIspH1* (Os03t0731900-01) and *OsIspH2* (Os03t0732000-00), designated in this study are annotated in Rice Annotation Project Database (RAP-DB). They have 68% amino acid sequence homology and are adjacent genes about 2.8 Mbp apart in the same chromosome 3. In monocotyledon plants, *IspH* genes are analyzed to exist as two copies in date palm (*Phoenix dactylifera*), banana (*Musa acuminata*), and a domesticated-rice (*O. sativa*; AA genotype) genomes and as just a single copy in other wild rice species (*Oryza brachyantha*, FF genotype; *Oryza meyeriana*, GG genotype), sorghum, corn, barley, wheat, and so on (**Figure 1**). As shown in the phylogenetic tree, *IspH* genes exist mostly as two or three copies in moss, fern, and gymnosperm genomes, one or two copies in monocotyledon genomes of angiosperm plants, and mostly a single copy in bacteria, dicotyledon genomes of angiosperm plants, indicating that the copy number of *IspH* genes has evolved to increase from a single copy in bacteria to two or three copies in gymnosperms but has evolved to decrease back to a single copy in angiosperms, during the genomic evolution. Of two rice *IspH* genes, an *OsIspH1* was evolutionarily grouped with a single-copy *IspH* gene of other monocotyledon plants in a phylogenetic tree, but *OsIspH2* was distant from them (**Figure 1**), supposing that OsIspH1 might have a conservative IspH function in angiosperms, rather than OsIspH2 having it.

Molecular Identification and Characterization of Rice *IspH* Genes

To define the structures of two rice *IspH* genes, we extracted genomic DNA sequence information from the RAP-DB[3] and constructed their genome structures, which are equally composed of ten exons and nine introns (**Figure 2A**). The 1st and 3rd introns of *OsIspH2* have unusually longer sequences, 1,061 bp and 13.5 kbp, respectively, compared with 652 and 409 bp of *OsIspH1* when predicted only with the genomic DNA sequence because the cDNA of *OsIspH2* has never been reported. It meant that the functionality of *OsIspH2* was largely unknown.

To compare the expression patterns of two rice *IspH* genes, we examined them in various tissues at different developmental stages with the gene-specific primers located in a 3′-untranslated region (**Figure 2B** and **Supplementary Table 3**). The spatial expression pattern revealed that *OsIspH1* transcripts were 20–40 times higher in the leaves than roots at all developmental stages, indicating a preference for leaves. They were also expressed in florets and seeds. *OsIspH2* was not expressed in any leaf tissue and florets but was found in low levels in roots and seeds, indicating that *OsIspH1* might be mainly functional in all rice tissues. Furthermore, we examined the expression patterns of two rice *IspH* genes against the treatments of six phytohormones, namely, ABA, CK, GA, IAA, MeJA, and SA (**Figure 2C** and **Supplementary Table 4**). Transcript levels of *OsIspH1* were slightly suppressed by ABA but were induced by other five

[1]https://www.greenphyl.org/cgi-bin/index.cgi

[2]https://blast.ncbi.nlm.nih.gov/Blast.cgi

[3]https://rapdb.dna.affrc.go.jp/

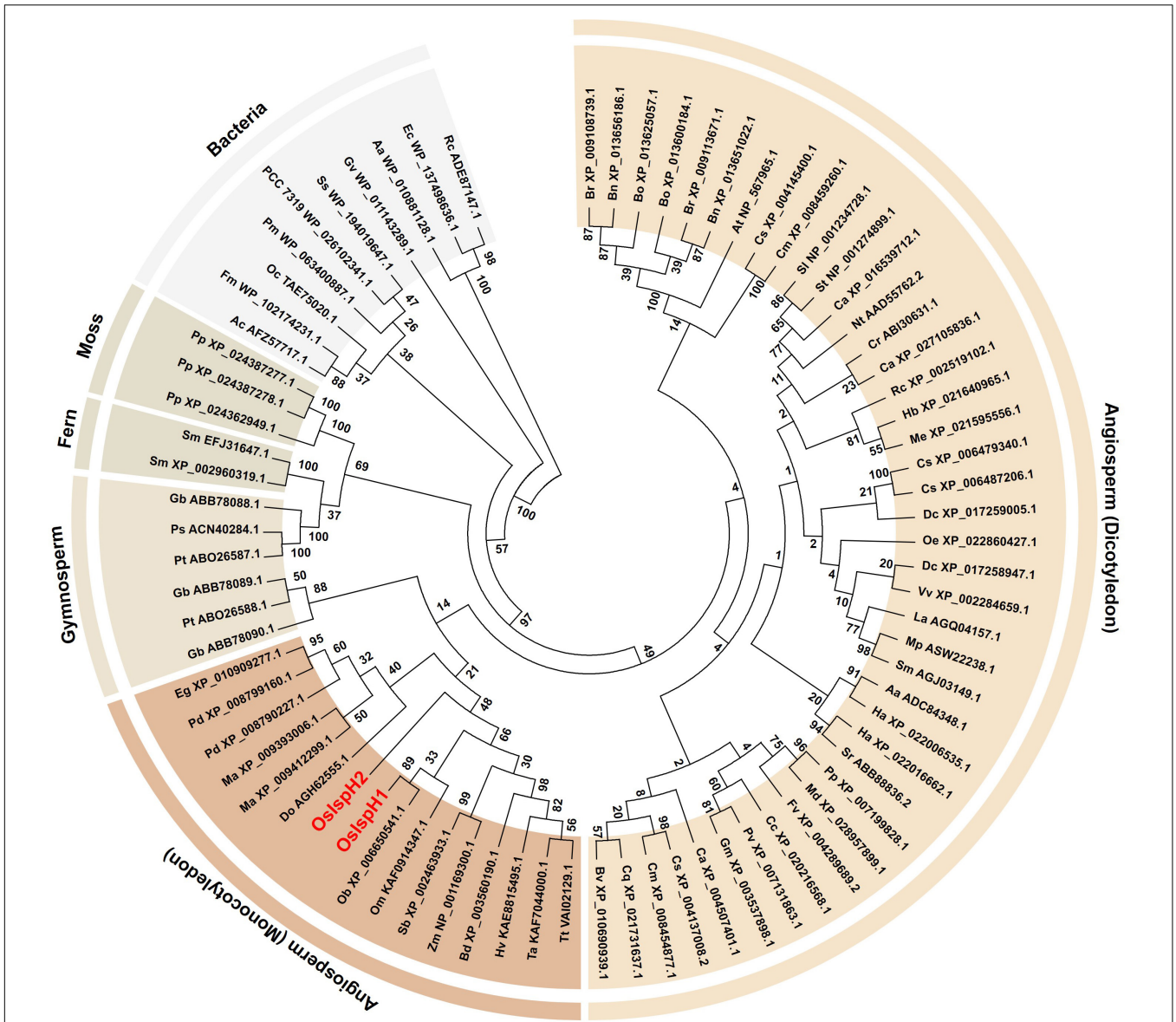


FIGURE 1 | Evolutionary relationships among isoprenoid synthesis H (IspH) families. The phylogenetic tree was inferred using the maximum-likelihood method based on the JTT matrix-based model. The bootstrap consensus tree was inferred from 100 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The IspH family is named by a combination of a species abbreviation and an accession number, with rice IspHs marked in red letters. Evolutionary analyses were conducted on 79 amino acid sequences using MEGA7, and **Supplementary Table 1** contains all information of the IspH family used in this study.

hormones in early time posttreatment, proposing different responses. Those of *OsIspH2* were barely detectable before and after treatment.

To simultaneously identify the cDNAs of two rice *IspH* genes, total RNAs were prepared from roots at the seedling stage using the spatial expression pattern in **Figure 2B**, and RT-PCRs were performed using diverse primer combinations (**Figure 2D**). *OsIspH1* exhibited a single 1,410-bp-long band (including 1,380 bp encoding 459 amino acids) as predicted in amplification, including the full-length ORFs; however, *OsIspH2* unexpectedly showed a longer single 2,705-bp band than the predictable 1,644 bp (including 1,449 bp encoding 482 amino acids). Through further RT-PCRs of *OsIspH2* with different combinations of exon-specific primers and sequencing of corresponding amplicons, we confirmed the only splicing error in the 1st intron, generating the longer 1st exon, shorter 1st intron, and quite longer 2nd exon than expected size, as indicated by the dotted line in **Figure 2A**. Interestingly, the extended 1st exon region contained the early stop codon, indicating that *OsIspH2* could be a non-functional gene by causing a disabled small ORF of 324 bp encoding 107 amino acids.

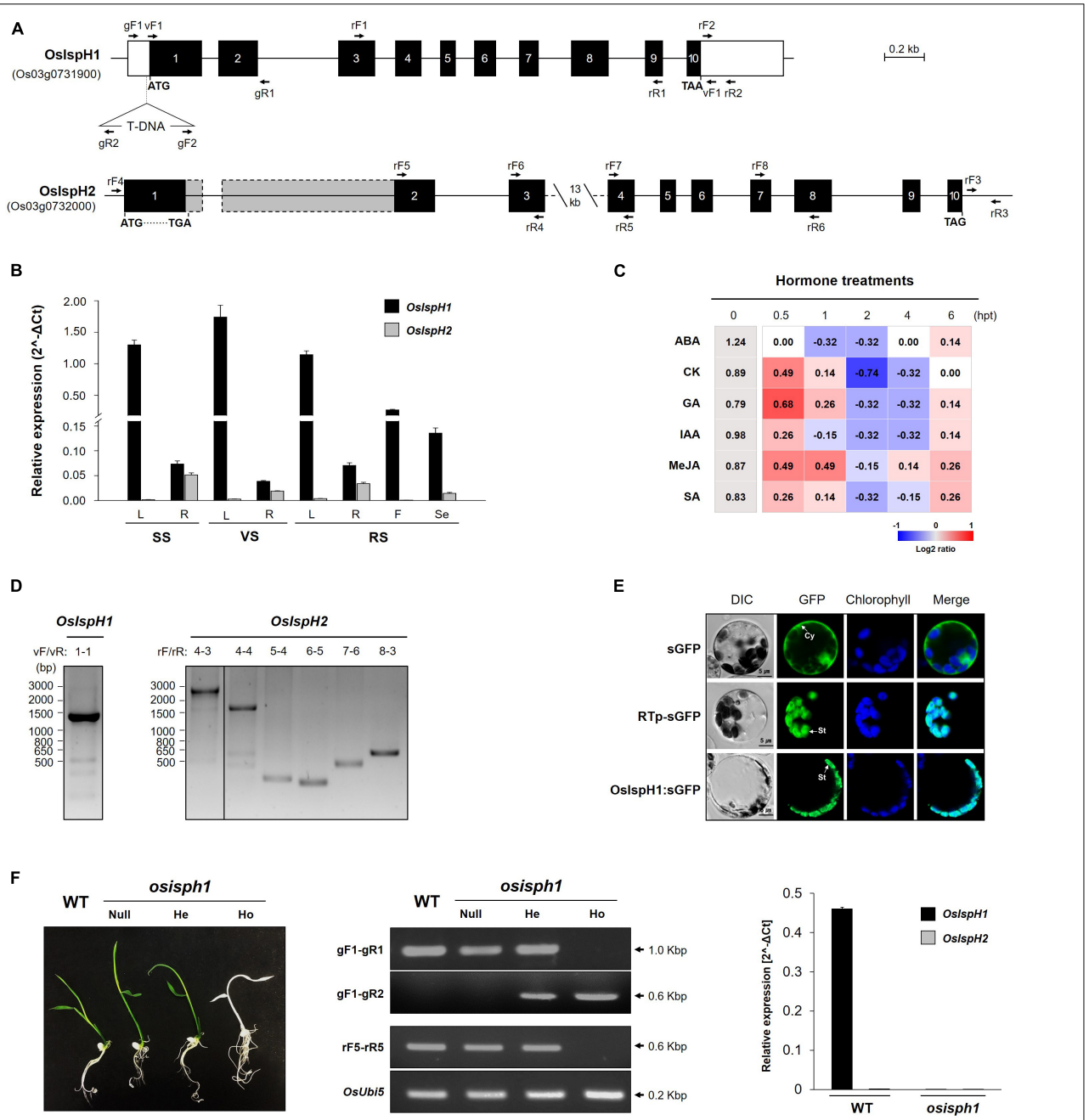


FIGURE 2 | Molecular characterization of rice *lspH* genes. **(A)** Genomic structures of the *OslspH1* and *OslspH2* genes and verification of T-DNA position in knockout mutant *osisph1*. Solid boxes represent exon (black) and untranslated regions (white); the actual sizes of exons 1 and 2 of *OslspH2* are represented by gray boxes with dotted line, and solid lines indicate introns. Triangle indicates the position of T-DNA insertion. **(B)** Spatial and temporal expression of *OslspH1* and *OslspH2* in various tissues, including leaves (L), roots (R) of the seedling stage (SS), vegetative stage (VS), and reproductive stage (RS) tissues, and florets (F) of RS and seeds (Se) harvested at 60 days after flowering. **(C)** Expression profiles of endogenous *OslspH1* in response to different hormone treatments in 10-day-old leaves. The heatmap graph is presented by the relative log2 fold change values of each time point, which were obtained by normalizing the ΔCt values of each 0 h time point (gray-colored boxes) as individual control, and the ΔCt value of each 0 h time point is presented in column 0 hpt. **(D)** Reverse transcriptase PCRs (RT-PCRs) verify the predicted transcript sizes of *OslspH1* (left panel) and *OslspH2* (right panel) using rice young seedling roots. **(E)** Subcellular localization of the OslspH1:superfolder green fluorescent protein (sGFP) fusion protein with sGFP and RTp-sGFP, which are used as cytosol (Cy)- and stroma (St)-localized markers, respectively. A confocal microscope was used to detect the sGFP fluorescence and chlorophyll autofluorescence. **(F)** Characterization of the loss-of-function mutant for *OslspH1* using 5-day-old seedling phenotypes (left panel), genotyping and semiquantitative reverse-transcriptase PCRs (middle panel), and real-time quantitative reverse transcriptase PCR (qRT-PCR, right panel). WT, *Oryza sativa* Japonica cv. Dongjin as a wild-type control; null, nullizygous plants; He, heterozygous plants; Ho, homozygous plants. The used primer locations are indicated with arrows in **(A)** and **Supplementary Table 10** shows the sequence information. All qRT-PCRs were triplicated and calculated using the ΔCt equation against the *OsUbi5* gene as an internal control. The results are expressed as the mean ± standard error (SE), and **Supplementary Tables 3, 4** show the original values.

Accordingly, *OsIspH1* turned out to be the only functional HDR gene in rice. Its localization was examined using a GFP-fused protein transformation in rice protoplasts prepared from green seedlings and confirmed in stroma inside chloroplasts, with a similar pattern to stroma marker, rice ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit transit peptide, but different from cytosol marker (**Figure 2E**).

To evaluate the role of *OsIspH1* in rice plants, we obtained a rice mutant with a T-DNA insertion in the region of 5′-untranslated region of *OsIspH1* (PFG_1B-11039)[4], as depicted in **Figure 2A**. We confirmed that the *osisph1* knockout mutant exhibits albinism in homozygous plants based on genotyping and qRT-PCR (**Figure 2F**), and the albino phenotypes were observed at a 3:1 segregation ratio to identify a single-copy insertion of T-DNA, supposing that the albino phenotype might be derived from the knockout of *OsIspH1* gene functions. Similarly, another knockout mutant of an *OsIspH1* gene showed also the lethal albinism and the defectiveness of chloroplast biogenesis, which was constructed by a CRISPR-Cas9 gene-editing system (Liu et al., 2020), supposing that the albinism phenotype of *osisph1* might also be mediated by the defective chloroplast biogenesis. Taken together, *OsIspH1* is suggested to be crucial in photosynthetic pigments, including each or both chlorophylls and carotenoids, and that *OsIspH2* could not compensate for the loss of function of *OsIspH1*.

Elucidation of Key Residues of OsIspH1 by Functional Complementation in *E. coli*

To confirm the conservation in key residues of *OsIspH1*, we aligned amino acid sequences among 12 monocot plants, including rice (**Figure 3**). Previously known key residues, an N-terminal conserved domain (NCD), four Cys-residues consisting of Fe-S clusters at active sites, and eight substrate-binding sites are highly conserved among these IspHs, indicating that *OsIspH1* could function as an HDR. Interestingly, a potential key residue, such as protein kinase C (PKC) phosphorylation site in the C-terminal region, was newly predicted to be conserved among IspHs in the corresponding position of *OsIspH1*, Ser-Tyr-Lys$^{405-407}$, through a Motif Scan analysis[5], and Lys^{407} was also predicted as one of the ubiquitination sites through computational prediction of ubiquitination site[6].

Referring to the well-studied key residue of His^{152} in *Arabidopsis* (Hsieh et al., 2014), we mutated OsIspH1 to have non-synonymous substitutions in the corresponding His^{145} residue and newly predicted Lys^{407} residue into Pro^{145} and Arg^{407}, respectively. Two single-mutated *OsIspH1*^{H145P} and *OsIspH1*^{K407R} genes and a double-mutated gene of *OsIspH1*$^{H145P/K407R}$ were transformed into an *E. coli IspH1/LytB*-lesion mutant strain, a DLYT1 (Kim et al., 2008). As a result, *OsIspH1*, *OsIspH1*^{H145P}, and *OsIspH1*^{K407R} successfully complemented the lethal phenotype of DLYT1 when *EcIspH1/LytB* was transformed as a positive control (**Figure 4A**). It confirmed that individual single mutation does not affect HDR activity. However, a double-substitution mutant, *OsIspH1*$^{H145P/K407R}$ (also called *OsIspH1*^{MUT} hereafter), barely restored DLYT1 growth, indicating the complementary relationship of two residues for HDR enzyme activity. Additionally, to inquire whether substitution mutations of amino acids cause structural change in *OsIspH1*, we simulated the three-dimensional modeling based on PDB ID: 3DNF from *A. aeolicus* IspH using SWISS-MODEL[7], as depicted in **Figure 4B**. According to a previous report, the first mutation from a positively charged His to a non-polar Pro is crucial for the "induced fit" effect to bind the substrate HMBPP (Hsieh et al., 2014). In this study, the predicted modeling showed the possibility that 1st key His^{145} residue might be faced to the Lys^{407} residue at nearly position (**Figure 4B**), supporting the possibility that the complementary relationship of two residues might be important for HDR enzyme activity in rice plants, although the mechanism is largely unknown.

In planta Effects of OsIspH1 on the Biosynthesis of Chlorophylls and Carotenoids

Based on the albino phenotype of the loss-of-function mutants, *osisph1*, as depicted in **Figure 2F**, we constitutively overexpressed an *OsIspH1*^{MUT} to elucidate how an *OsIspH1* affects chlorophyll and carotenoid biosynthesis in rice leaves and seeds. To learn more about its effects on carotenoid metabolism in seeds involved in the Golden Rice trait, the *OsIspH1*^{MUT} cassette was further combined with a *stPAC* cassette, which produces β-carotene in rice endosperm (Jeong et al., 2017), as a single T-DNA vector (**Figure 5A**). Using three-independent homozygous T3 plants, which were selected by considering the lower copy number of one and two (**Figure 5B**), the expression patterns of a double-mutated transgene and an endogenous gene encoding rice HDR/IspH1 were examined in leaves (**Figure 5C**) and seeds (**Figure 5D**) of T4 rice plants. The overexpression of transgene, *OsIspH1*^{MUT}, in both organs and seed-specific expression of *stPAC* was confirmed relative to non-transgenic (NT) rice plants. Interestingly, the ectopic expression of an *OsIspH1*^{MUT} distinctly affected the expression patterns of endogenous *OsIspH1* in both organs. In leaves of the IspH1^{MUT} transgenic plants, the expression of endogenous *OsIspH1* was significantly suppressed by an average of about 50% ($P < 0.001$ in IspH1^{MUT} and $P < 0.05$ in IspH1^{MUT}-stPAC), and the combined transcript levels of *OsIspH1*^{MUT} and *OsIspH1* were analyzed to be also significantly lower than the normal transcript level of endogenous *OsIspH1* ($P < 0.01$ in IspH1^{MUT} and $P < 0.05$ in IspH1^{MUT}-stPAC), even under the ectopic expression of *OsIspH1*^{MUT} transgene (**Figure 5C** and **Supplementary Table 5**). In contrast to the suppression patterns in leaves, the combined transcript levels of *OsIspH1*^{MUT} and *OsIspH1* were analyzed to be at least 4 times higher than the transcript level of endogenous *OsIspH1* ($P < 0.001$ in IspH1^{MUT} and $P < 0.05$ in IspH1^{MUT}-stPAC), and the transcript levels of endogenous *OsIspH1* were not suppressed in seeds of all transgenic plant (**Figure 5D** and **Supplementary Table 6**),

[4]http://signal.salk.edu/cgi-bin/RiceGE

[5]http://myhits.isb-sib.ch/cgi-bin/motif_scan

[6]http://bdmpub.biocuckoo.org/prediction.php

[7]https://swissmodel.expasy.org/

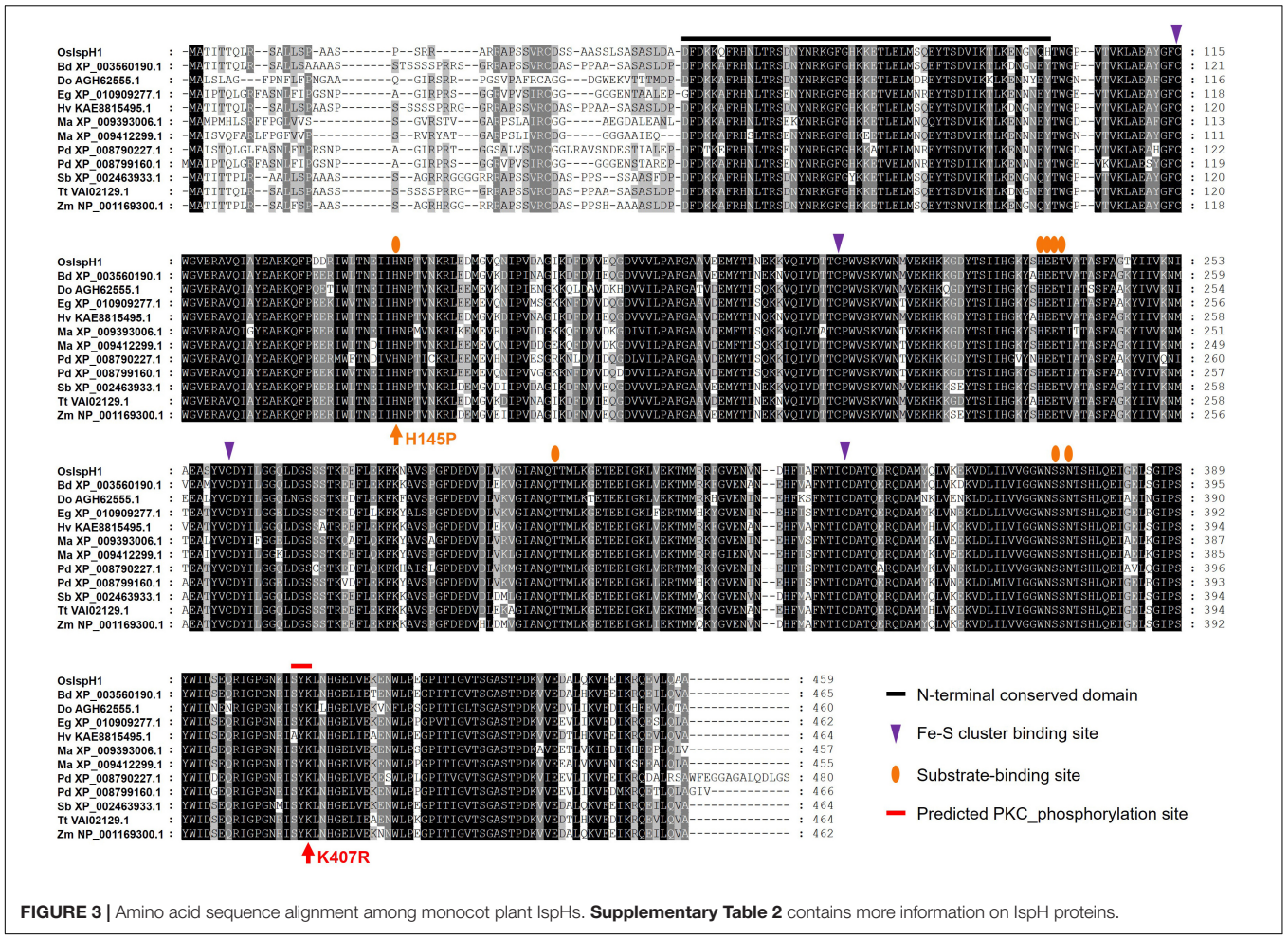


FIGURE 3 | Amino acid sequence alignment among monocot plant IspHs. **Supplementary Table 2** contains more information on IspH proteins.

indicating that *OsIspH1*^{MUT} was expressed four times more than the intrinsic *OsIspH1* gene. Finally, the distinct expression patterns between leaves and seeds were statistically significantly conservative in six independent transgenic plants of IspH1^{MUT} and IspH1^{MUT}-stPAC, and taken collectively, it was supposed that the expression of *OsIspH1* might be differentially regulated between leaves and seeds.

In the next step, we analyzed two terpene-derived photosynthetic pigments, including chlorophylls and carotenoids, using HPLC. In leaves, although there were differences between the independent transgenic plants, the total contents of chlorophylls (an average of 84.6% level, **Supplementary Table 7**) and carotenoids (an average of 84.3% level, **Supplementary Table 8**) were decreased in two transgenic plants with the significant differences compared to NT plants (**Figure 6A**). In addition to the functional complementation of *OsIspH1*^{MUT} as HDR/IspH in *E. coli* (**Figure 4A**), *in planta* functional complementation was examined through cross-interbreeding between two transgenic rice plants, namely, *osisph1* and IspH1^{MUT}. In rice plants, the albinism of *osisph1* was not restored by *OsIspH1*^{MUT} overexpression at all, of which homozygosity was ascertained by genotyping (**Figure 6B**), indicating that OsIspH1^{MUT} was a protein defective in HDR activity in rice plants. Collectively, it was supposed that the overexpression of OsIspH1^{MUT} might cause the relative decrease in active HDR proteins through the competition of OsIspH1^{MUT} with the endogenous OsIspHs and consequently decrease chlorophyll and carotenoid contents in the leaves. In the seeds, the carotenoid content of IspH1^{MUT} (**Figure 7A**) and IspH1^{MUT}-stPAC (**Figure 7B**) was statistically significantly reduced on the average of 61.4 and 54.2%, respectively, compared to NT, indicating a more severe pattern in seeds than leaves (**Figure 7** and **Supplementary Table 9**). The results were considered to be closely related to the higher expression level of OsIspH1^{MUT} than the endogenous *OsIspH1* gene in the seeds.

Taken together, OsIspH1 is supposed to have crucial roles such as the rate-limiting regulation on carotenoid biosynthesis in seeds and photosynthetic pigments in leaves.

DISCUSSION

OsIspH1 Is the Unique Functional HDR in Rice Plants

The MEP pathway consists of seven enzymatic steps that start with the catalysis of DXS and end with that of IspH. While the

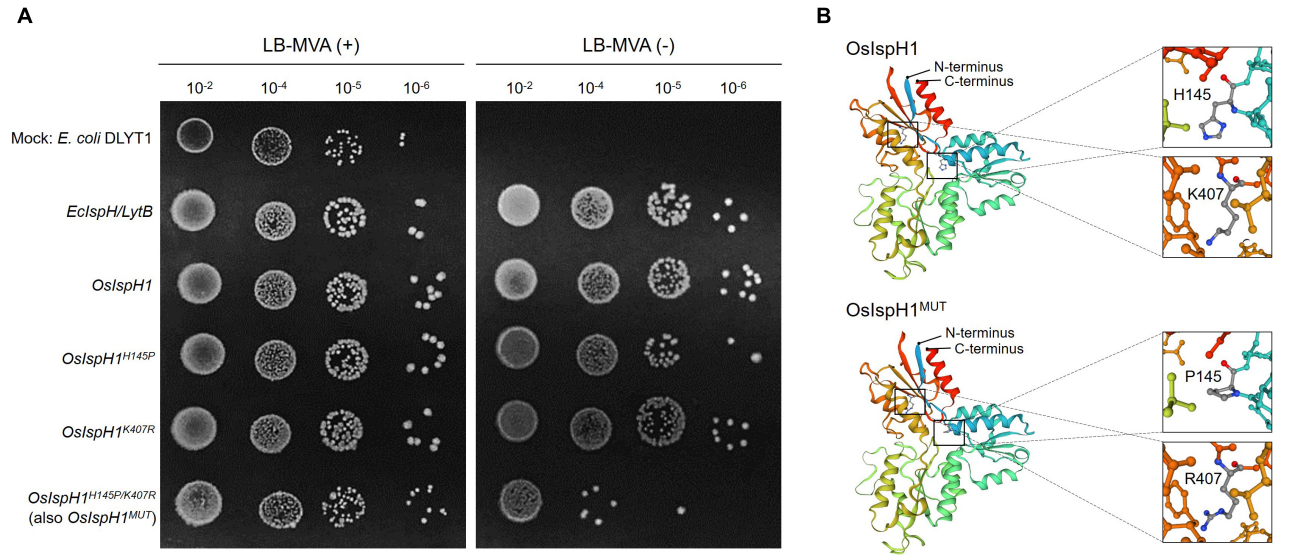


FIGURE 4 | The 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) enzyme assay in an *Escherichia coli* system and simulated protein structures of IspHs. **(A)** Functional complementation in *E. coli* DLYT1, an *EcIspH* lesion strain. An *E. coli IspH* gene, *EcIspH/LytB*, was used as a positive control. Mevalonate (MVA) (+) and (-) mean 0.01% and none of mevalonolactone, respectively. **(B)** Simulation of the 3D modeling to predict protein structure as rainbow ribbon drawing based on PDB ID: 3DNF from *Aquifex aeolicus* IspH built using SWISS-MODEL. Enlarged views showed the region with H145 and K407 residues as a ball and stick model with carbon (gray), nitrogen (blue), and oxygen (red) atoms.

five intermediate enzymes are encoded by a single gene, DXS and IspH have been reported as one or multiple genes in diverse plants, proposing the potent roles as flux-controlling steps for terpenoid production. Plant DXSs have been classified into three clades, with different influences on the downstream pathways: DXS1s for chlorophyll biosynthesis with a housekeeping role, DXS2s for the production of specialized terpene metabolites, including artemisinins, ginkgolides, resins, and carotenoids, in different plants, and DXS3 for considerably understudied but still tentative roles in a few plants, as recently summarized (Cordoba et al., 2011; Zhang et al., 2018; de Luna-Valdez et al., 2021; Tian et al., 2021). Meanwhile, plant IspH has been reported as a single-gene family in most plant species but as a multiple-gene family of 2–3 copies in some species. In *G. biloba* and *Pinus taeda* of gymnosperm plants, GbIspH1 and PtIspH1 are for housekeeping roles, and GbIspH2 and PtIspH2 are for individually specialized ginkgolide and resin roles (Kim et al., 2008, 2009). In angiosperm plants, two *C. melo* genes showed different spatial expression: *CmIspH1* for the most abundant in all organs and *CmIspH2* in specialized contexts, such as de-etiolating seedlings and orange-fleshed fruit during ripening (Saladie et al., 2014). Thereby, it could be proposed that *IspH1*s play housekeeping roles, but *IspH2s* can contribute to plant species-specific terpene production. In the case of *O. sativa*, two *IspH* genes are predicted in the genome and have been previously presumed the different function: *OsIspH1* as a predominantly light-induced gene but *OsIspH2* as an upregulated gene in the *dxr* mutants without light response, proposing a compensating route awaiting validation (Jung et al., 2008). However, Liu et al. (2020) presented that the genome-edited *OsIspH1* defective mutant showed the albino lethal phenotype, and similarly in this study, the T-DNA mutant of *OsIspH1*, *osisph1*, showed the albino lethal phenotype (**Figure 6**), supposing that *OsIspH2* might not have any compensate function for it. Interestingly in this study, the sequence analysis of *OsIspH2* mRNA presented that the extended 1st exon region had an early stop codon to consequentially generate the truncated-short polypeptide fragments, which was spliced at the quite different region from the predicted-splicing position provided by the RAP-DB (**Figures 2A,D**). Collectively, an *OsIspH2* gene is supposed to be non-functional, indicating that an *OsIspH1* is a unique functional HDR gene in rice.

The Complementary Relationship of His^{145} and Lys^{407} Residues Is Crucial for the HDR Activity of OsIspH1 in Both Bacteria and Rice Plants

The X-ray crystal structure of IspH protein has been first reported in the hyperthermophilic eubacterium *A. aeolicus* (Rekittke et al., 2008), its critical residues have been elucidated mainly in bacterial systems, such as *A. aeolicus*, *E. coli*, and cyanobacteria (Rekittke et al., 2008; Grawert et al., 2009; Hsieh and Hsieh, 2015). Meanwhile, the plant-type IspH structure has also been reported in *Arabidopsis* and the same key residues to bacterial ones by complementation assay in *E. coli isph* mutant (Hsieh et al., 2014). By Hsieh et al. (2014), two bacterial critical residues showed the difference in the corresponding ones of *Arabidopsis*, His^{152} and His^{241}, not individually but cooperatively critical for HDR activity with more significance of His^{152} than His^{241}. In this study, we found another conserved motif, referred to as a predicted PKC phosphorylation site in **Figure 3**, and to determine whether it could be a novel

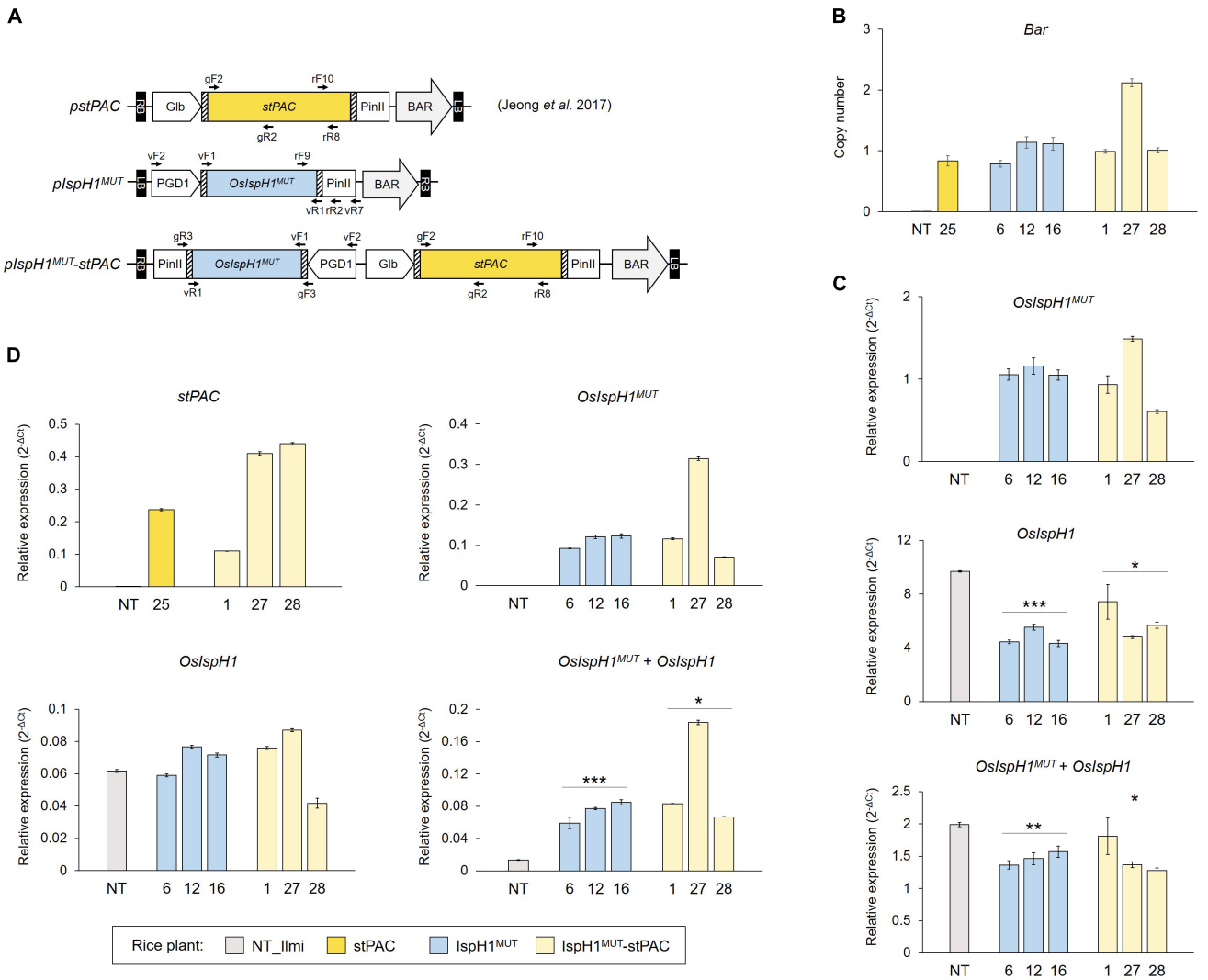


FIGURE 5 | Generation of transgenic rice plants for *in planta* characterization. **(A)** Schematic representation of binary vectors for rice transformation with the indicated positions of primer sets listed in **Supplementary Table 10**. **(B)** TaqMan PCR to verify the transgene copy numbers using bialaphos-resistant (Bar) gene cassette. The previously reported homozygous stPAC 25 line was used as a reference for the single T-DNA insertion (Jeong et al., 2017). The qRT-PCR to examine transcript levels of transgenes, *OsIspH1*^{MUT} or *stPAC*, endogenous *OsIspH1*, and combined genes, *OsIspH1*^{MUT} and *OsIspH1*, in transgenic rice leaves **(C)** and seeds **(D)**, respectively. The rice ubiquitin gene (Os01g22490) was used as an internal reference for quantitative normalization, and the values were represented as the mean ± SE of three independent measurements. Statistical analyses were performed using a one-tailed Student's *t*-test, and significant differences were indicated by a *p*-value (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

crucial residue of OsIspH1, we mutated each or together at both positions of His^{145} and Lys^{407} to generate *OsIspH1*^{H145P} and *OsIspH1*^{K407R}, and *OsIspH1*$^{H145P/K407R}$, and the former is the corresponding site to His^{152} of AtIspH, and the latter is a newly predicted site for PKC phosphorylation or ubiquitination in this study. The complementary experiments in the *isph* mutants of bacteria and rice plants showed that the simultaneous mutation of His^{145} and Lys^{407}, an OsIspH1$^{H145P/K407R}$, largely reduced the HDR activity of OsIspH1 not only in the bacteria expression system (**Figure 4A**) but also in the rice plant system (**Figure 6B**), but every single mutation did not have any effects on the HDR activity of it in bacteria system. These results supposed that the complementary relationship of two residues, namely, His^{145} and Lys^{407}, might be crucial to the HDR activity of OsIspH1.

OsIspH1 Plays an Important Role in the Biosynthesis of Chlorophylls and Carotenoids, Through the Differential Expression Patterns Between Leaves and Seeds

The regulation of plant genes involved in MEP and MVA pathways occurs at multiple levels of transcription,

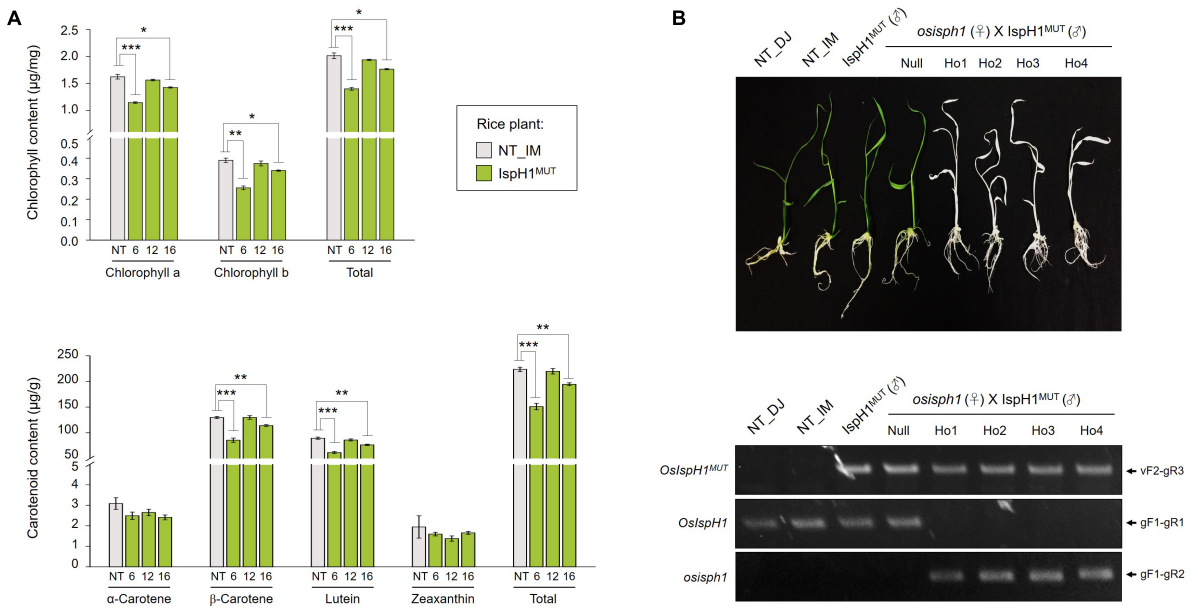


FIGURE 6 | Effects of the *OslspH1*^{MUT} on rice leaves. **(A)** Contents and composition of chlorophylls and carotenoids in the leaves at T4 plants. Non-transgenic (NT) rice plants (*O. sativa* L. cv. Ilmi) and three independent IspH1^{MUT} transgenic plants were used for metabolite analysis. All data are represented as the mean ± SE of three independent measurements. Statistical analyses were performed using a one-tailed Student's *t*-test, and significant differences were indicated by a *p*-value (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). **(B)** Functional complementation in *osisph1* rice plants. Phenotypes of 10-day-old seedlings by interbreeding between heterozygous *osisph1* and IspH1^{MUT} transgenic plants at F4 generation and genomic PCR to confirm genotypes using primer sets listed in **Supplementary Table 10**.

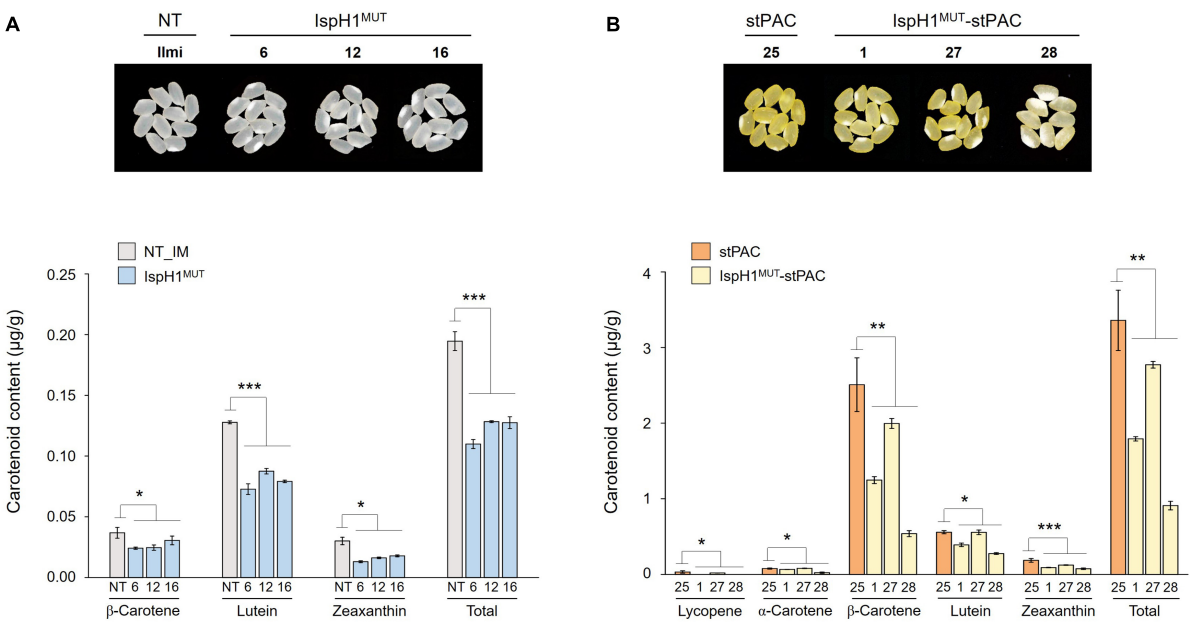


FIGURE 7 | Color phenotypes and carotenoid levels in the rice seeds of IspH1^{MUT} **(A)** and IspH1^{MUT}-stPAC **(B)** transgenic lines. Colors of mature seed harvested 60 days after flowering were compared, dried for 2 weeks at room temperature, threshed, dehusked, and polished in the homozygous T4 generation. Carotenoid contents were analyzed using high-performance liquid chromatography (HPLC), and all data are represented as the mean ± SE of three independent measurements. Statistical analyses were performed using a one-tailed, and significant differences were indicated by a *p*-value (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

posttranscription, translation, and posttranslation (Cordoba et al., 2009; Hemmerlin, 2013; Vranova et al., 2013). Particularly, plant IspHs contribute to the MEP pathway primarily through transcriptional regulation with tissue specificity and different responses against environmental conditions, such as light, circadian, mechanical wounding, and fungal elicitors in diverse plants (Kim et al., 2008, 2009; Cordoba et al., 2009; Ma et al., 2017). To scrutinize the intrinsic rice IspH regulation, we constitutively overexpressed the malfunctional OsIspH1^{MUT} (**Figure 5**) in rice plants, which was identified in two *hdr* mutants, namely, *E. coli* DLYT1 strains (**Figure 4**) and rice *osisph1* plants (**Figure 6B**). Interestingly, the effects of *OsIspH1*^{MUT}-ectopic expression were observed to be quite differential between rice leaves and seeds. In leaves, not only the transcript level of an endogenous *OsIspH1* was downregulated, but also the combined transcript level of endogenous *OsIspH1* and exogenous *OsIspH1*^{MUT} remained lower than that of NT (**Figure 5C**). In contrast, the expression of an endogenous *OsIspH1* was not downregulated in the seeds and remained higher than that of NT (**Figure 5D**), suggesting that the expression of an intrinsic *OsIspH1* gene was differentially regulated in leaves and seeds. Similarly, organ-specific differential roles of OsDXS2 and OsDXR on carotenoid accumulation between leaves and seeds have also been reported (You et al., 2020). Collectively, these results suggest that the DXS, DXR, and IspH steps belonging to the MEP pathway might be regulated to differentially affected carotenoid accumulation between leaves and seeds.

Despite such differential regulation, the ectopic expression of *OsIspH1*^{MUT} simultaneously caused the decrease of chlorophyll and carotenoid content in leaves and seeds (**Figures 6**, **7**), and the reduction patterns were proportional to the overexpression patterns of *OsIspH1*^{MUT}, indicating the relative decrease in active HDR enzyme function by competition between endogenous OsIspH1 and OsIspH1^{MUT}. In other words, the intrinsic OsIspH1 is supposed to play an important role in the biosynthesis of chlorophyll and carotenoid levels in rice leaves and seeds and to be a promising target to biofortify the functional terpene metabolites in rice plants.

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

S-HH coordinated the project, supervised the manuscript, and was responsible for all contacts and correspondence. MY contributed to the interpretation of data for the work, and revised it critically for important intellectual content. YL performed all the experiments *in planta* and wrote the draft of the manuscript. JK and S-AB analyzed the contents of chlorophylls and carotenoids. J-SY contributed to the experiments in *E. coli.* All authors have read and approved the final manuscript.

FUNDING

This study was supported by two grants from the BioGreen21 Agri-Tech Innovation Program (PJ01567101 to S-HH), funded by the Rural Development Administration in South Korea, and National Research Foundation of Korea (NRF), funded by the Korean government (MSIT) (2021R1A2C2012227 to S-HH).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.861036/full#supplementary-material

REFERENCES

Altincicek, B., Duin, E. C., Reichenberg, A., Hedderich, R., Kollas, A. K., Hintz, M., et al. (2002). LytB protein catalyzes the terminal step of the 2-C-methyl-D-erythritol-4-phosphate pathway of isoprenoid biosynthesis. *FEBS Lett.* 532, 437–440. doi: 10.1016/s0014-5793(02)03726-2

Botella-Pavia, P., Besumbes, O., Phillips, M. A., Carretero-Paulet, L., Boronat, A., and Rodriguez-Concepcion, M. (2004). Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. *Plant J.* 40, 188–199. doi: 10.1111/j.1365-313X.2004.02198.x

Cordoba, E., Porta, H., Arroyo, A., Roman, C. S., Medina, L., Rodriguez-Concepcion, M., et al. (2011). Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *J. Exp. Bot.* 62, 2023–2038. doi: 10.1093/jxb/erq393

Cordoba, E., Salmi, M., and Leon, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60, 2933–2943. doi: 10.1093/jxb/erp190

Cunningham, F. X. Jr., Lafond, T. P., and Gantt, E. (2000). Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. *J. Bacteriol.* 182, 5841–5848. doi: 10.1128/JB.182.20.5841-5848.2000

de Luna-Valdez, L., Chenge-Espinosa, M., Hernandez-Munoz, A., Cordoba, E., Lopez-Leal, G., Castillo-Ramirez, S., et al. (2021). Reassessing the evolution of the 1-deoxy-D-xylulose 5-phosphate synthase family suggests a possible novel function for the DXS class 3 proteins. *Plant Sci.* 310:110960. doi: 10.1016/j.plantsci.2021.110960

Eoh, H., Brennan, P. J., and Crick, D. C. (2009). The *Mycobacterium tuberculosis* MEP (2C-methyl-d-erythritol 4-phosphate) pathway as a new drug target. *Tuberculosis (Edinb)* 89, 1–11. doi: 10.1016/j.tube.2008.07.004

Grawert, T., Kaiser, J., Zepeck, F., Laupitz, R., Hecht, S., Amslinger, S., et al. (2004). IspH protein of *Escherichia coli*: studies on iron-sulfur cluster implementation and catalysis. *J. Am. Chem. Soc.* 126, 12847–12855. doi: 10.1021/ja0471727

Grawert, T., Rohdich, F., Span, I., Bacher, A., Eisenreich, W., Eppinger, J., et al. (2009). Structure of active IspH enzyme from *Escherichia coli* provides mechanistic insights into substrate reduction. *Angew. Chem. Int. Ed Engl.* 48, 5756–5759. doi: 10.1002/anie.200900548

Grieshaber, N. A., Fischer, E. R., Mead, D. J., Dooley, C. A., and Hackstadt, T. (2004). Chlamydial histone-DNA interactions are disrupted by a metabolite in the methylerythritol phosphate pathway of isoprenoid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7451–7456. doi: 10.1073/pnas.0400754101

Ha, S. H., Kim, J. K., Jeong, Y. S., You, M. K., Lim, S. H., and Kim, J. K. (2019). Stepwise pathway engineering to the biosynthesis of zeaxanthin, astaxanthin

and capsanthin in rice endosperm. *Metab. Eng.* 52, 178–189. doi: 10.1016/j.ymben.2018.11.012

Hao, G., Shi, R., Tao, R., Fang, Q., Jiang, X., Ji, H., et al. (2013). Cloning, molecular characterization and functional analysis of 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase (HDR) gene for diterpenoid tanshinone biosynthesis in *Salvia miltiorrhiza* Bge. f. *alba*. *Plant Physiol. Biochem.* 70, 21–32. doi: 10.1016/j.plaphy.2013.05.010

Hemmerlin, A. (2013). Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Sci.* 203-204, 41–54. doi: 10.1016/j.plantsci.2012.12.008

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

Hintz, M., Reichenberg, A., Altincicek, B., Bahr, U., Gschwind, R. M., Kollas, A. K., et al. (2001). Identification of (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate as a major activator for human gammadelta T cells in *Escherichia coli*. *FEBS Lett.* 509, 317–322. doi: 10.1016/s0014-5793(01)03191-x

Hsieh, M. H., and Goodman, H. M. (2005). The *Arabidopsis* IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. *Plant Physiol.* 138, 641–653. doi: 10.1104/pp.104.058735

Hsieh, W. Y., and Hsieh, M. H. (2015). The amino-terminal conserved domain of 4-hydroxy-3-methylbut-2-enyl diphosphate reductase is critical for its function in oxygen-evolving photosynthetic organisms. *Plant Signal. Behav.* 10:e988072. doi: 10.4161/15592324.2014.988072

Hsieh, W. Y., Sung, T. Y., Wang, H. T., and Hsieh, M. H. (2014). Functional evidence for the critical amino-terminal conserved domain and key amino acids of *Arabidopsis* 4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE REDUCTASE. *Plant Physiol.* 166, 57–69. doi: 10.1104/pp.114.243642

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

Jung, K. H., Lee, J., Dardick, C., Seo, Y. S., Cao, P., Canlas, P., et al. (2008). Identification and functional analysis of light-responsive unique genes and gene family members in rice. *PLoS Genet.* 4:e1000164. doi: 10.1371/journal.pgen.1000164

Kim, J. K., Lee, S. Y., Chu, S. M., Lim, S. H., Suh, S. C., Lee, Y. T., et al. (2010). Variation and correlation analysis of flavonoids and carotenoids in Korean pigmented rice (*Oryza sativa* L.) cultivars. *J. Agric. Food Chem.* 58, 12804–12809. doi: 10.1021/jf103277g

Kim, S. M., Kuzuyama, T., Kobayashi, A., Sando, T., Chang, Y. J., and Kim, S. U. (2008). 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms *Ginkgo biloba* and *Pinus taeda*. *Planta* 227, 287–298. doi: 10.1007/s00425-007-0616-x

Kim, T. J., Choi, J., Kim, K. W., Ahn, S. K., Ha, S. H., Choi, Y., et al. (2017). Metabolite profiling of peppers of various colors reveals relationships between tocopherol, carotenoid, and phytosterol content. *J. Food Sci.* 82, 2885–2893. doi: 10.1111/1750-3841.13968

Kim, Y. B., Kim, S. M., Kang, M. K., Kuzuyama, T., Lee, J. K., Park, S. C., et al. (2009). Regulation of resin acid synthesis in *Pinus densiflora* by differential transcription of genes encoding multiple 1-deoxy-D-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase genes. *Tree Physiol.* 29, 737–749. doi: 10.1093/treephys/tpp002

Kim, Y. B., Kim, S. M., Sathasivam, R., Kim, Y. K., Park, S. U., and Kim, S. U. (2021). Overexpression of *Ginkgo biloba* Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase 2 gene (GbHDR2) in *Nicotiana tabacum* cv. Xanthi. *3 Biotech* 11:337. doi: 10.1007/s13205-021-02887-5

Ko, M. R., Song, M. H., Kim, J. K., Baek, S. A., You, M. K., Lim, S. H., et al. (2018). RNAi-mediated suppression of three carotenoid-cleavage dioxygenase genes, OsCCD1, 4a, and 4b, increases carotenoid content in rice. *J. Exp. Bot.* 69, 5105–5116. doi: 10.1093/jxb/ery300

Liu, N., Xie, K., Jia, Q., Zhao, J., Chen, T., Li, H., et al. (2016). Foxtail mosaic virus-induced gene silencing in monocot plants. *Plant Physiol.* 171, 1801–1807. doi: 10.1104/pp.16.00010

Liu, X., Cao, P. H., Huang, Q. Q., Yang, Y. R., and Tao, D. D. (2020). Disruption of a rice chloroplast-targeted gene OsHMBPP causes a seedling-lethal albino phenotype. *Rice (N. Y.)* 13:51. doi: 10.1186/s12284-020-00408-1

Lu, X. M., Hu, X. J., Zhao, Y. Z., Song, W. B., Zhang, M., Chen, Z. L., et al. (2012). Map-based cloning of zb7 encoding an IPP and DMAPP synthase in the MEP pathway of maize. *Mol. Plant* 5, 1100–1112. doi: 10.1093/mp/sss038

Ma, D., Li, G., Zhu, Y., and Xie, D. Y. (2017). Overexpression and suppression of *Artemisia annua* 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 1 gene (AaHDR1) differentially regulate Artemisinin and terpenoid biosynthesis. *Front. Plant Sci.* 8:77. doi: 10.3389/fpls.2017.00077

Masini, T., and Hirsch, A. K. (2014). Development of inhibitors of the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway enzymes as potential anti-infective agents. *J. Med. Chem.* 57, 9740–9763. doi: 10.1021/jm501 0978

Mathur, V., Kwong, W. K., Husnik, F., Irwin, N. A. T., Kristmundsson, A., Gestal, C., et al. (2021). Phylogenomics identifies a new major subgroup of apicomplexans, *Marosporida* class nov., with extreme apicoplast genome reduction. *Genome Biol. Evol.* 13:evaa244. doi: 10.1093/gbe/evaa244

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

Page, J. E., Hause, G., Raschke, M., Gao, W., Schmidt, J., Zenk, M. H., et al. (2004). Functional analysis of the final steps of the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway to isoprenoids in plants using virus-induced gene silencing. *Plant Physiol.* 134, 1401–1413. doi: 10.1104/pp.103.038133

Park, S. H., Yi, N., Kim, Y. S., Jeong, M. H., Bang, S. W., Choi, Y. D., et al. (2010). Analysis of five novel putative constitutive gene promoters in transgenic rice plants. *J. Exp. Bot.* 61, 2459–2467. doi: 10.1093/jxb/erq076

Rao, G., and Oldfield, E. (2016). Structure and function of four classes of the 4Fe-4S protein, IspH. *Biochemistry* 55, 4119–4129. doi: 10.1021/acs.biochem.6b00474

Rekittke, I., Wiesner, J., Rohrich, R., Demmer, U., Warkentin, E., Xu, W., et al. (2008). Structure of (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase, the terminal enzyme of the non-mevalonate pathway. *J. Am. Chem. Soc.* 130, 17206–17207. doi: 10.1021/ja806668q

Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* 295(Pt. 2), 517–524. doi: 10.1042/bj2950517

Saladie, M., Wright, L. P., Garcia-Mas, J., Rodriguez-Concepcion, M., and Phillips, M. A. (2014). The 2-C-methylerythritol 4-phosphate pathway in melon is regulated by specialized isoforms for the first and last steps. *J. Exp. Bot.* 65, 5077–5092. doi: 10.1093/jxb/eru275

Singh, N., Cheve, G., Avery, M. A., and McCurdy, C. R. (2007). Targeting the methyl erythritol phosphate (MEP) pathway for novel antimalarial, antibacterial and herbicidal drug discovery: inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) enzyme. *Curr. Pharm. Des.* 13, 1161–1177. doi: 10.2174/138161207780618939

Tian, S., Wang, D., Yang, L., Zhang, Z., and Liu, Y. (2021). A systematic review of 1-Deoxy-D-xylulose-5-phosphate synthase in terpenoid biosynthesis in plants. *Plant Growth Regul.* 96, 221–235. doi: 10.1007/s10725-021-00784-8

Uchida, H., Sumimoto, K., Oki, T., Nishii, I., Mizohata, E., Matsunaga, S., et al. (2018). Isolation and characterization of 4-hydroxy-3-methylbut-2-enyl diphosphate reductase gene from *Botryococcus braunii*, race B. *J. Plant Res.* 131, 839–848. doi: 10.1007/s10265-018-1039-4

Vranova, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116

Wang, W., Wang, K., Liu, Y. L., No, J. H., Li, J., Nilges, M. J., et al. (2010). Bioorganometallic mechanism of action, and inhibition, of IspH. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4522–4527. doi: 10.1073/pnas.0911087107

Wang, X., and Dowd, C. S. (2018). The methylerythritol phosphate pathway: promising drug targets in the fight against *Tuberculosis*. *ACS Infect. Dis.* 4, 278–290. doi: 10.1021/acsinfecdis.7b00176

Wellburn, A. R. (1994). The spectral determination of chlorophylls a and b, as well as total carotenoids, using various sovents with spectrophotometers of different resolution. *J. Plant Physiol.* 144, 307–313. doi: 10.1016/s0176-1617(11)81192-2

You, M. K., Kim, J. H., Lee, Y. J., Jeong, Y. S., and Ha, S. H. (2016). Plastoglobule-targeting competence of a putative transit peptide sequence from rice phytoene synthase 2 in plastids. *Int. J. Mol. Sci.* 18:18. doi: 10.3390/ijms18010018

You, M. K., Lee, Y. J., Kim, J. K., Baek, S. A., Jeon, Y. A., Lim, S. H., et al. (2020). The organ-specific differential roles of rice DXS and DXR, the first two enzymes of the MEP pathway, in carotenoid metabolism in *Oryza sativa* leaves and seeds. *BMC Plant Biol.* 20:167. doi: 10.1186/s12870-020-02357-9

Zhang, F., Liu, W., Xia, J., Zeng, J., Xiang, L., Zhu, S., et al. (2018). Molecular characterization of the 1-deoxy-D-xylulose 5-phosphate synthase gene family in *Artemisia annua*. *Front. Plant Sci.* 9:952. doi: 10.3389/fpls.2018.00952

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fpls.2022.854842

Untargeted Metabolomics Sheds Light on the Diversity of Major Classes of Secondary Metabolites in the Malpighiaceae Botanical Family

Helena Mannochio-Russo[1,2], Rafael F. de Almeida[3,4], Wilhan D. G. Nunes[5], Paula C. P. Bueno[6,7], Andrés M. Caraballo-Rodríguez[2], Anelize Bauermeister[2], Pieter C. Dorrestein[2] and Vanderlan S. Bolzani[1]**

[1]NuBBE, Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, Brazil, [2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, San Diego, CA, United States, [3]Royal Botanical Gardens Kew, Science, Ecosystem Stewardship, Diversity and Livelihoods, Richmond, United Kingdom, [4]Department of Biological Sciences, Lamol Lab, Feira de Santana State University (UEFS), Feira de Santana, Brazil, [5]Federal Institute of Education, Science and Technology of Rondônia (IFRO), Ji-Paraná, Brazil, [6]Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, [7]Institute of Chemistry, Federal University of Alfenas (UNIFAL), Alfenas, Brazil

Natural products produced by plants are one of the most investigated natural sources, which substantially contributed to the development of the natural products field. Even though these compounds are widely explored, the literature still lacks comprehensive investigations aiming to explore the evolution of secondary metabolites produced by plants, especially if classical methodologies are employed. The development of sensitive hyphenated techniques and computational tools for data processing has enabled the study of large datasets, being valuable assets for chemosystematic studies. Here, we describe a strategy for chemotaxonomic investigations using the Malpighiaceae botanical family as a model. Our workflow was based on MS/MS untargeted metabolomics, spectral searches, and recently described *in silico* classification tools, which were mapped into the latest molecular phylogeny accepted for this family. The metabolomic analysis revealed that different ionization modes and extraction protocols significantly impacted the chemical profiles, influencing the chemotaxonomic results. Spectral searches within public databases revealed several clades or genera-specific molecular families, being potential chemical markers for these taxa, while the *in silico* classification tools were able to expand the Malpighiaceae chemical space. The classes putatively annotated were used for ancestral character reconstructions, which recovered several classes of metabolites as homoplasies (i.e., non-exclusive) or synapomorphies (i.e., exclusive) for all sampled clades and genera. Our workflow combines several approaches to perform a comprehensive evolutionary chemical study. We expect it to be used on further chemotaxonomic investigations to expand chemical knowledge and reveal biological insights for compounds classes in different biological groups.

Keywords: chemotaxonomy, mass spectrometry, metabolite annotation, metabolomics, evolution, ancestral character reconstruction, systematics, malpighiales

OPEN ACCESS

Edited by:
Jens Rohloff, Norwegian University of Science and Technology, Norway

Reviewed by:
José Juan Ordaz-Ortiz, Centro de Investigaciones y Estudios Avanzados, Instituto Politécnico Nacional de México (CINVESTAV), Mexico
Evangelos Tatsis, Center for Excellence in Molecular Plant Sciences (CAS), China

****Correspondence:***
Vanderlan S. Bolzani
vanderlan.bolzani@unesp.br
Helena Mannochio-Russo
helena.russo@unesp.br

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

Received: *14 January 2022*
Accepted: *23 March 2022*
Published: *14 April 2022*

Citation:
Mannochio-Russo H, de Almeida RF, Nunes WDG, Bueno PCP, Caraballo-Rodríguez AM, Bauermeister A, Dorrestein PC and Bolzani VS (2022) Untargeted Metabolomics Sheds Light on the Diversity of Major Classes of Secondary Metabolites in the Malpighiaceae Botanical Family. Front. Plant Sci. 13:854842. doi: 10.3389/fpls.2022.854842

INTRODUCTION

Plant metabolites have been widely explored since the 1930s, mainly aiming at an in-depth study of species with ethnopharmacological value, bioactive extracts and compounds, and new chemical structures (Newman and Cragg, 2020; Atanasov et al., 2021). Despite its invaluable importance, the establishment of phylogenetic diversification and distribution patterns of plant secondary metabolites is still in its early steps, and several plant families have not been deeply explored to date in this context.

The development of highly sensitive detection techniques, such as mass spectrometry (MS), allowed the investigation of plant extracts as a whole, providing a comprehensive overview of the metabolites biosynthesized by plant species (De Vos et al., 2007; Gemperline et al., 2016). In this context, chemosystematics studies employing MS, multivariate analyses, and bioinformatic tools for exploring large plant datasets have been successfully performed, raising valuable insights regarding the biosynthetic pathways involved in different phylogenetic groups (Gallon et al., 2018; Martucci et al., 2018; Ernst et al., 2019; Kang et al., 2019). It is now possible to support DNA-based phylogenetic studies at a molecular level and assist in elaborating evolutionary hypotheses based on natural products and metabolomics analyses (Schmitt and Barker, 2009). The rapid development of natural product bioinformatics tools and databases (Li and Gaquerel, 2021; Medema, 2021; Bauermeister et al., 2022) can be of great value to assist and accelerate more comprehensive chemosystematics studies aiming at taxa-specific metabolic pathways.

Even though evolutionary studies at a metabolite level are of great interest, certain factors can significantly impact the detection of metabolites and, consequently, the conclusions drawn. For instance, the extraction solvent used and the ionization mode in MS can prioritize one class of compounds over another (Floros et al., 2016; Creydt and Fischer, 2017), leading to biased results. Even though several studies compare different extraction protocols and ionization modes in metabolomics analyses, an investigation of these factors in chemosystematics studies is still poorly explored. Therefore, selecting a plant family with a broad diversity of classes of compunds identified, in addition to an extensive record of DNA-based phylogenetic studies, would be ideal for evaluating these variables, enabling a deeper chemotaxonomic investigation.

The Malpighiaceae plant family is an excellent example of both the high diversity of secondary metabolites produced by plant species, and a flowering plant family with all its genera sampled in DNA-based phylogenetic studies (Davis and Anderson, 2010; Mannochio-Russo et al., 2020). Malpighiaceae is one of the 36 families of flowering plants placed in Malpighiales by several phylogenetic studies based on chloroplast genes (Angiosperm Phylogeny Group, 1998, 2003, 2009; Angiosperm Phylogeny Group et al., 2016), being also one of the most important and diverse orders of angiosperms in tropical forests (Xi et al., 2012; Cai et al., 2019). This family currently comprises 74 genera and *ca.* 1,300 species, mostly confined to the American tropics, with Brazil being its most representative country. Few genera and species reach the tropics of Africa, Asia, and Oceania (Davis and Anderson, 2010; de Almeida et al., 2021; de Almeida and van den Berg, 2021). Some Amazonian species of Malpighiaceae are traditionally known for their psychedelic or aphrodisiac properties, with several studies focusing on the chemical characterization of these species (Samoylenko et al., 2010; Queiroz et al., 2014). On the other hand, several extra-Amazonian Malpighiaceae species are long reported as toxic to cattle (i.e., *Amorimia*, and *Niedenzuella* spp.; Riet-Correa et al., 2012; Lee et al., 2014), causing significant economic losses in the Brazilian growing cattle industry. Additionally, *Malpighia* and *Byrsonima* spp. also stand out for the nutritional value of their fruits (Belwal et al., 2018; Neri-Numa et al., 2018).

In the past two decades, traditional intrafamilial classifications of several angiosperm families (i.e., subfamily and tribe ranks), solely based on macromorphology, were proven to be non-monophyletic (i.e., artificial, and not reflecting common ancestry; Angiosperm Phylogeny Group, 1998, 2003, 2009; Angiosperm Phylogeny Group et al., 2016). During this time, Malpighiaceae has gone through unprecedented changes in its traditional classification due to several DNA-based phylogenetic studies (Cameron et al., 2001; Davis et al., 2001; Davis and Anderson, 2010; de Almeida et al., 2017, 2018). Key morphological characters of its traditional classification system (i.e., fruit types) were recovered as highly homoplastic (i.e., non-exclusive; Cameron et al., 2001; Davis et al., 2001). The inevitable recognition of unforeseen relationships within Malpighiaceae brought to light deep taxonomic problems regarding the monophyly of several genera [e.g., *Banisteriopsis* C.R.Rob., *Mascagnia* (Bertero ex DC.) Bertero, *Stigmaphyllon* A.Juss., and *Tetrapterys* Cav.], tribes (e.g., just Gaudichaudieae Horan. was recovered as monophyletic), and all its subfamilies (e.g., Byrsonimoideae W.R.Anderson and Malpighioideae A.Juss.; Cameron et al., 2001; Davis et al., 2001; Davis and Anderson, 2010; de Almeida et al., 2017). Since then, different authors have gradually proposed new genera and combinations to accommodate these newly identified lineages (Anderson, 2006, 2011; Anderson and Davis, 2006; Davis et al., 2020; de Almeida and van den Berg, 2021). Even though some morphological characters were used to reconstruct the last generic phylogeny of Malpighiaceae, no morphological characters were ever recovered, circumscribed, or discussed for its major phylogenetic clades (de Almeida and van den Berg, 2021). As a result, its traditional classification was rejected, and 10 informal clades, without any morphological circumscription, were recognized in the most recent generic phylogeny for Malpighiaceae (Davis and Anderson, 2010). More recently, phylogenomic studies were performed with six Malpighiaceae species and strongly corroborated previous phylogenetic studies within this family (Menezes et al., 2018; Ramachandran et al., 2018; Cai et al., 2019; Jo et al., 2019; Gong et al., 2020). In this context, a deeper chemical investigation of the family and mapping its chemical traits in a phylogenetic tree that reflects the evolutionary relationships among organisms can be of great value in chemotaxonomic investigations (Schmitt and Barker, 2009). It can help predict metabolically interesting groups of organisms

to assist future studies of this taxon and give more support to the evolutionary hypotheses.

In this study, we present a new approach for chemosystematics studies, by combining natural products research with phylogenetic methods (Schmitt and Barker, 2009). We performed metabolomics analyses in combination with recently described *in silico* fragmentation predictors, chemical hierarchy analysis, and ancestral character reconstructions to map the presence/absence of the annotated metabolites in the most recent generic DNA-based phylogenetic tree of Malpighiaceae (**Figure 1**). For this, we evaluated a unique collection of Malpighiaceae samples, comprising 39 genera (out of 74) and 137 species from each of the major phylogenetic groups currently accepted for the family (collection distribution is shown in **Figure 2**). This study comprised representative samples from all the currently accepted phylogenetic clades of the family, which enabled us to obtain a comprehensive overview of the metabolites produced by this family (percentage of genera covered by each clade: 67, 50, 67, 100, 33, 60, 20, 59, 31, and 63% for clades A–J, respectively[1]). We evaluated (i) the impact of different extraction protocols and ionization modes in MS for chemotaxonomic investigations; (ii) the metabolites annotated based on spectral matches and *in silico* tools; and (iii) how the chemical diversity in Malpighiaceae evolved over the geological time in this family. With this enhanced approach, we were able to provide insights regarding the complementarity of the different ionization modes, provide the first chemical information of several Malpighiaceae genera, and draw conclusions regarding the evolution of the classes of secondary metabolites annotated in the Malpighiaceae plant family.

[1]http://powo.science.kew.org/

MATERIALS AND METHODS

General Information

The ethyl alcohol (proof, for molecular biology) used for the extraction procedure was acquired from Sigma-Aldrich (St. Louis, United States). The ethyl acetate (HPLC grade) used for the extraction procedure was obtained by J.T. Baker (J.T. Baker-Avantor, Radnor, United States). Acetonitrile and water, both LC–MS grade, were obtained from Fisher Scientific (Fair Lawn, NJ, United States).

Collection of Plant Material

Most plant samples were collected by R.F. Almeida on field expeditions throughout Brazil from 2013 to 2017, or were retrieved from discarded fragmented samples used for DNA extraction in molecular studies (de Almeida et al., 2017, 2018; de Almeida and van den Berg, 2020, 2021). For information regarding all sampled specimens, see **Supplementary Table 1**. After each collection, the plant materials were dried in a desiccator containing silica at room temperature. Samples were then frozen in liquid nitrogen and grounded in a ball mill.

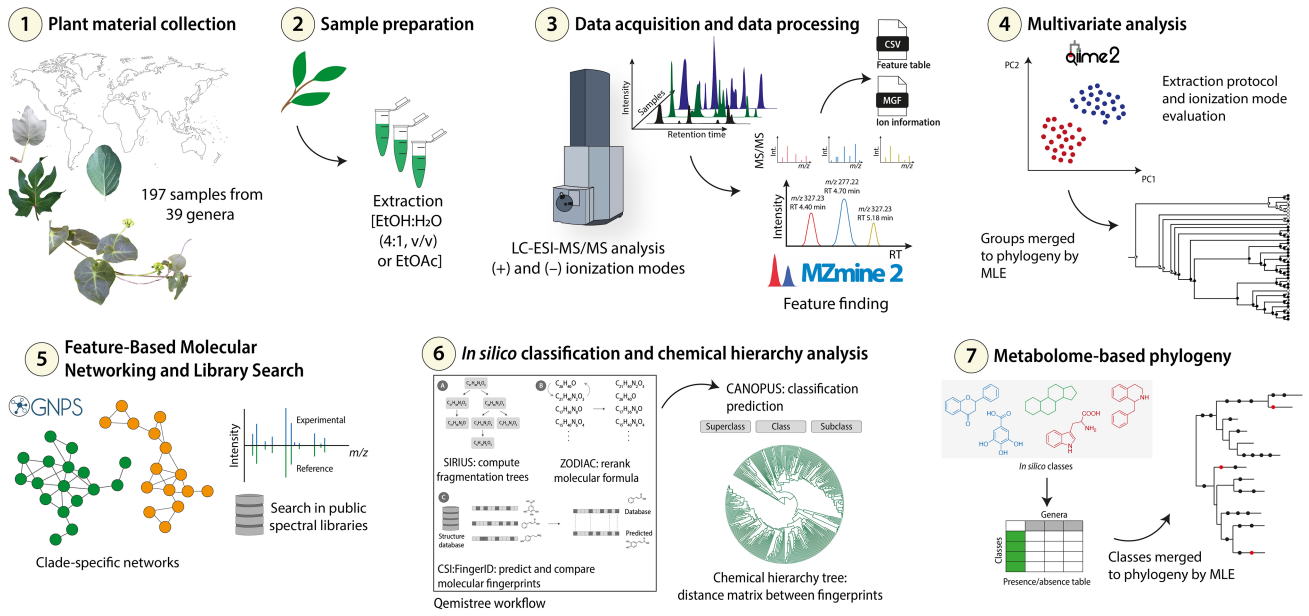


FIGURE 1 | Experimental workflow followed for the metabolomics and chemosystematics analyses of Malpighiaceae samples. (1) The samples were initially collected, (2) the extracts were prepared with different solvents [EtOH:H_2O (4:1, v/v) or EtOAc], and then (3) subjected to LC-ESI-MS/MS analysis in positive and negative ionization modes in an untargeted method. (4) The data acquired were processed for feature finding, and the exported data were used for multivariate analysis. The clustering groups observed were merged to the phylogeny using the Maximum Likelihood Estimation (MLE) for preliminary chemotaxonomic investigations. (5) The data were also used for Feature-Based Molecular Networking and library searches workflows to observe clade-specific molecular families. (6) A chemical hierarchy analysis and *in silico* classifications were obtained and finally (7) merged to the currently accepted Malpighiaceae phylogeny to determine the ubiquitous and the taxa-specific *in silico* classes.

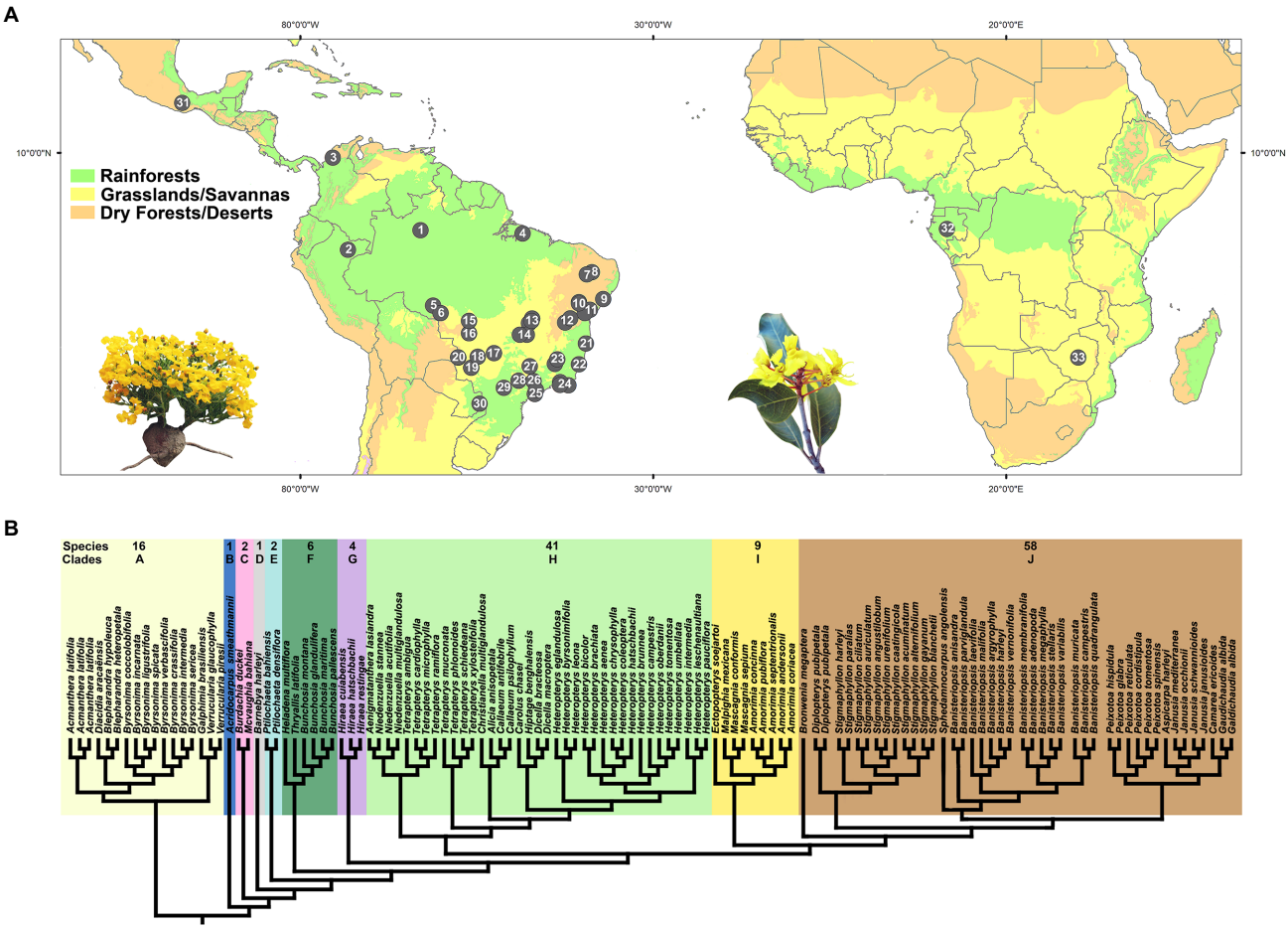


FIGURE 2 | (A) Distribution map showing the collection sites of all samples within the American and African continents. A complete record of all collection sites (numbers on black circles) is listed in **Supplementary Table 1**. Photograph on the left represents a New World tropic species of Malpighiaceae (*Camarea ericoides* by R.F. Almeida). Photograph on the right represents an Old World tropic species of Malpighiaceae (*Acridocarpus excelsus* by T. Randrianarivony). **(B)** Ten major phylogenetic clades currently accepted in Malpighiaceae, based on plastid and nuclear genes, according to Davis and Anderson (2010). Major clades are shaded in different colors. Species = number of species sampled by each clade in our study. Clade A, Byrsonimoid clade; B, Acridocarpoid clade; C, Mcvaughioid clade; D, Barnebyoid clade; E, Ptilochaetoid clade; F, Bunchosioid clade; G, Hiraeoid clade; H, Tetrapteroid clade; I, Malpighioid clade; and J, Stigmaphylloid clade.

The samples were stored in a freezer at −20°C until the preparation of the extracts.

The authorization for conducting this study was conceded by the National System for Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN), n° A6FDC2E.

Extraction Procedure

The plant sample materials were weighed and extracted with EtOH:H_2O (4:1, v/v) or EtOAc (100%) in a proportion of 20 mg of plant material to 1 mL of extraction solvent. The samples were homogenized in a Qiagen TissueLyzer II (Qiagen, Hilden, Germany) for 5 min at 25 MHz and extracted for additional 30 min at room temperature. The samples were centrifuged (5,000 *g*) for 15 min, and 300 µL of the supernatants were transferred to a 96-deep-well plate. The solvent was dried in a Labconco CentriVap (United States), and the plates were sealed and stored at −80°C prior to analyses.

UHPLC-MS/MS Analysis

The extracts were initially resuspended in 200 µL MeOH:H_2O (4:1) containing sulfachloropyridazine (2 µM) as internal standard [to monitor sample injection during the Ultra High Performance Liquid Chromatography (UHPLC)-tandem Mass Spectrometry (MS/MS) data acquisition], and sonicated for 15 min. The plates were centrifuged for 10 min at 1,300 *g*, and the supernatants were then transferred to a new 96-well plate for metabolomics analyses.

The analyses were carried out with a Thermo Scientific UltiMate 3000 UHPLC system coupled to a Maxis Impact QTOF mass spectrometer (Bruker Daltonics, Germany), controlled by the Otof Control and Hystar software packages, and equipped with ESI source. The extracts were analyzed using a Kinetex 1.7 µm C18 reversed-phase UHPLC column (50 × 2.1 mm; Phenomenex,

Torrance, CA, United States), at 40°C, and an injection volume of 5 μL. The pump system consisted of water (A) and acetonitrile (B), both acidified with formic acid (0.1%, v/v), and the flow rate was set at 0.5 ml/min. The metabolites separation was achieved with 5% solvent B for 1 min, followed by a linear gradient from 5 to 100% in 5 min. The column was washed at 100% solvent B for 2 min, then returned to the initial 5% in 1 min, and the equilibration of the column was achieved for 1 min at 5% solvent B. The mass spectra were acquired in both positive and negative ionization modes, separately, in a mass range of 50–2,000 Da in data-dependent acquisition (DDA) mode. The parameters used for data acquisition were set as follows: nitrogen used as nebulizer gas with pressure at 2 bar, a capillary voltage of 4,200 V, ion source temperature of 200°C, dry gas flow at 9 L/min, and spectra rate acquisition of three spectra/s. The five most intense selected ions per spectrum were fragmented (MS/MS) using ramped collision-induced dissociation energy, ranging from 22 to 50 eV. MS/MS active exclusion was set after five spectra and released after 30 s.

The UHPLC–MS/MS data were deposited in the MassIVE Public GNPS dataset[2] (MSV000085119) and are publicly available.

MS/MS Data Pre-Processing

The raw data files (.d) were converted to .mzXML format using DataAnalysis software (Bruker) after lock mass correction using hexakis(1H,1H,2H-difluoroethoxy) phosphazene (Synquest Laboratories, Alachua, FL, United States), with *m/z* 622.029509. The quality of the analyses was evaluated considering the retention time and the *m/z* of a standard solution containing a mixture of six standards, which was analyzed after the completion of each row in a 96-well plate.

The .mzXML files were processed in MZmine2 (version 2.37.corr17.7_kai_merge2) for positive and negative ionization modes, separately. The parameters used for feature finding were as follows: mass detection (centroid, 1.0E3 and 1.0E1 for MS1 and MS2, respectively); chromatogram builder (minimum time span of 0.01 min, minimum height of 3.0E3, and *m/z* tolerance of 20 ppm); chromatogram deconvolution (baseline cut-off algorithm: minimum peak height: 1.0E3, peak duration range: 0.01–3 min; and baseline level: 1.0E3) with median *m/z* center calculation, *m/z* range for MS2 scan pairing of 0.02 Da and retention time (RT) range for MS2 scan pairing of 0.1 min; isotope peaks grouper (*m/z* tolerance set at 20 ppm, RT tolerance of 0.1 min, maximum charge of 3, and representative isotope set to most intense), join alignment (*m/z* tolerance of 20 ppm, weight for *m/z* and RT of 75 and 25, respectively, and RT tolerance of 0.1 min). A filter was applied in order to keep only the features with MS/MS spectra. This feature list was exported as a feature quantification table (.csv), as a MS2 spectral summary (.mgf), and with the SIRIUS export module (.mgf) for downstream analyses.

Feature-Based Molecular Networking

To investigate the metabolic profile of the dataset, the processed LC–MS/MS data (.mgf and .csv) were used to create a Feature-Based Molecular Network (FBMN) (Nothias et al., 2020) on the GNPS platform (Wang et al., 2016) with input files containing only the features detected in the hydroethanolic extracts. The data were filtered by removing all MS/MS fragment ions within +/− 17 Da of the precursor ion in order to remove possible residual precursor ions, which can sometimes be observed in MS/MS spectra acquired in QToF equipment. Additionally, MS/MS spectra were window filtered to select only the top six fragment ions in the +/− 50 Da window throughout the spectrum. Both the precursor ion and the MS/MS fragment ion tolerance were set to 0.02 Da. A molecular network was created, in which the edges were filtered to have a cosine score above 0.7 and at least four matched peaks. Similarly, the parameters for the library search (for comparison between the experimental and library spectra) were set to have a score above 0.7 and at least four matched peaks to assist in the metabolites annotation—level three according to the metabolomics standards initiative (MSI; Sumner et al., 2007). The FBMN jobs on GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2c5f11403ac847a298e4d7866a491143 (positive mode) and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=501c16500476451f978311057266fbdf (negative mode).

The molecular network visualization was performed in Cytoscape (version 3.7.2, Cytoscape Consortium, San Diego, CA, United States; Shannon et al., 2003), in which the nodes correspond to ion features, while the edges between the nodes represent the MS/MS cosine scores calculated. Subnetworks in which the nodes were found in significant abundances in blanks were excluded from the Cytoscape visualization to avoid misinterpretations due to contaminants in the analyses. Sample type information was added to color the nodes as pie charts representing the relative abundance of the features across the samples (colors based on the phylogenetic clades A–J). Node size was scaled relative to the sum of the peak areas obtained in the samples in which the feature was detected. Compounds with the same MS/MS spectra, but with different retention times, were represented as separate nodes, indicating isomers.

Chemical Hierarchy Analysis

A chemical hierarchy analysis (Qemistree; Tripathi et al., 2021) was performed with the metabolites detected in the hydroethanolic extraction protocol. For this, we used the q2-qemistree qiime2 plugin,[3] in which the feature quantification table (.csv) and the file obtained from the SIRIUS (Dührkop et al., 2019) export module (.mgf) from MZmine were used as input. Briefly, the Qemistree workflow consists of applying SIRIUS (version 4.8.2) to the .mgf file (containing ion information), generating predicted molecular formulas for each feature. The predicted molecular formulas were reranked using ZODIAC (Ludwig et al., 2020), and the predicted molecular fingerprints were subsequently generated using fragmentation trees *via* CSI:FingerID (Dührkop et al., 2015). The chemical taxonomy of the predicted metabolite structures was obtained

[2]http://massive.ucsd.edu

[3]https://github.com/biocore/q2-qemistree

by CANOPUS (superclass, class, and subclass; Dührkop et al., 2021). The Euclidean pairwise distances between the molecular fingerprints were calculated, and the fingerprint vectors were hierarchically clustered to generate a tree representing the structural chemical relationships of this dataset. The tree was then pruned in order to keep only the fingerprints classified up to a superclass level. The trees were visualized interactively in EMPRESS (Cantrell et al., 2021) for data exploration, in which clade information was added as relative abundance stacked barcharts to each feature. The dendrogram obtained can be interactively visualized with the .qzv files found in the Github repository.[4]

Statistical Analysis

The feature table exported from MZmine was used to perform unsupervised analysis using Qiime 2 (version 2020.2; Bolyen et al., 2019) bioinformatics pipeline within a Jupyter notebook. The metabolomic profiles were compared using the Bray–Curtis distance metric for comparing different extraction protocols, and using Canberra metric to investigate the four subsets individually (two extraction solvents, and two ionization modes). The Principal Coordinates Analysis (PCoA) plots showing the top three principal coordinates were visualized using EMPeror (Vázquez-Baeza et al., 2013). Permutational multivariate ANOVA (PERMANOVA; Anderson, 2001) was also performed in Qiime 2 on metabolite distance matrices to test for clustering significance (with 999 permutations), and the *F* statistic was reported as a measure of effect size.

The *in silico* classes retrieved from the Qemistree workflow (Tripathi et al., 2021) and CANOPUS (Dührkop et al., 2021) were tested for differential enrichment in the most sampled clades in study (clades A, F–J). Initially, the relative abundances were summed for the class group, the data were normalized by arc-sine square root transformation, and the effect of the clades was tested using a simple ANOVA. The adjusted values of *p* from ANOVA were obtained from the Benjamini-Hochberg method. The pairwise differences between the clades were tested through a *post-hoc* Tukey test, and the magnitude of differential enrichment was calculated through the $\log_2$ fold change in mean relative abundance between clades. The subclasses statistically enriched for specific clades were then selected to build a heatmap.

Phylogenetic Analyses

In total, 39 genera (out of 74) and 139 species (out of 1,300) of Malpighiaceae were sampled, representing all of its 10 major clades recognized by recent molecular phylogenetic studies (**Supplementary Table 1**). Since our sampling focused on the diversity of genera and clades within the Malpighiaceae family, genera not sampled on the chemical analyses were not included in the molecular phylogenetic analysis. Sequences for the genes *matK*, *ndhF*, PHYC, and *rbcL* were retrieved from GenBank,[5] edited using Geneious (Kearse et al., 2012), and aligned using Muscle (Edgar, 2004), with subsequent adjustments in the preliminary matrices to ensure that the nitrogenous bases were correctly aligned. The complete data matrices are available at TreeBase (https://www.treebase.org/treebase-web/search/studySearch.html, accession number S11008).

All trees were rooted in the Byrsonimoid clade (clade A), which is considered the sister-group of the other Malpighiaceae clades according to Davis and Anderson (2010). Combined analysis of plastid + nuclear regions was performed using Bayesian inference and Maximum Likelihood criteria to reconstruct our phylogenetic hypotheses. Both model-based methods were conducted with a mixed substitution model (GTR + G + I) and unlinked parameters, using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and raxmlGUI2 (Edler et al., 2021). For the Bayesian inference, the Markov Chain Monte Carlo (MCMC) was run using two simultaneous independent runs with four chains each (one cold and three heated), saving one tree every 1,000 generations, for a total of 10 million generations. We excluded 20% of retained trees as "burn in", and checked for a stationary phase of likelihood, checking for ESS values higher than 200 for all parameters on Tracer 1.6 (Rambaut et al., 2014). The posterior probabilities (PP) of clades were based on the majority rule consensus, using the stored trees, and calculated with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003).

Ancestral State Reconstruction

Chemical profiles were selected based on the clustering trends of the PCoA analysis for EtOH80 and EtOAc100% extracts, in both positive and negative ionization modes (four subsets). The subsets in which samples clustered into two groups (A and B) represented different chemical profiles. Character coding followed the recommendations of (Sereno, 2007) for morphological phylogenies. Primary homology hypotheses (de Pinna, 1991) were proposed for different chemical profiles, with a total of two subgroups of classes scored (i.e., presence and absence). All characters were optimized on the concatenated tree from the Bayesian inference using the maximum likelihood function with Mesquite 2.73 (Maddison and Maddison, 2006) using the mk1 algorithm.

After annotating the main classes of compounds present in the hydroethanolic extracts, we transformed the obtained quantitative matrices for the positive and negative ionization modes into qualitative matrices considering only the presence and absence of all annotated classes. Since we do not have all the analyzed species of this study sampled in the latest generic molecular phylogeny of Malpighiaceae, we summarized our results at the generic level using an arithmetic mean equation. After summarizing both matrices (retrieved from the positive and negative ionization modes), we compared all classes of secondary metabolites identified in each of them, merging their binary coding into a single column. By doing so, we obtained a total of 78 classes of secondary metabolites that were optimized into a summarized version of the phylogenetic tree presented by us in the previous step, including only the genera sampled in this study. Ancestral character reconstructions were performed using Maximum

[4]https://github.com/helenamrusso/Malpighiaceae_supplementary
[5]https://www.ncbi.nlm.nih.gov/genbank/

Likelihood approaches into the software Mesquite 2.73 (Maddison and Maddison, 2006) and visualized in the software Winclada (Nixon, 1999) using the fast optimization method, which favors homoplasies in the analyses.

RESULTS AND DISCUSSION

Evaluation of Extraction Solvent and Ionization Mode in Malpighiaceae Chemical Diversity

The extracts (EtOH:H_2O, 4:1 v/v; EtOAc) obtained from the 197 Malpighiaceae samples were analyzed by LC–MS/MS in both positive and negative ionization modes, and the LC–MS data were processed in MZmine2. The feature finding step resulted in a total of 24,440 and 20,150 features detected in positive and negative ionization modes, respectively. Usually, most of the dereplication tools based on LC–MS/MS consider the fragmentation of $[M+H]^+$ or $[M-H]^-$ adducts. Generally, molecules with pH greater than 7 (basic compounds) could be easily ionized in the positive mode making adducts with proton(s). The formation of deprotonated molecules is usually limited to compounds able to form acidic protons in the negative ionization mode. Although it is of high importance for the characterization of secondary metabolites, the signal intensity for data acquired in the negative mode is usually lower compared to the positive ionization mode. In addition, the use of positively charged ion fragmentation is more relevant due to the larger spectral library availability (Steckel and Schlosser, 2019; Wolfender et al., 2019).

The Venn diagrams obtained in positive and negative ionization modes for the different extraction solvents (**Figure 3A**) showed that, in the positive ionization mode, 74% of the features were shared between the two extraction protocols, while in the negative ionization mode, this number dropped to 59%, with the number of metabolites exclusively detected in the hydroethanolic extraction rising from 20 to 36%. Only about 5% of the features were detected exclusively with ethyl acetate as extraction solvent,

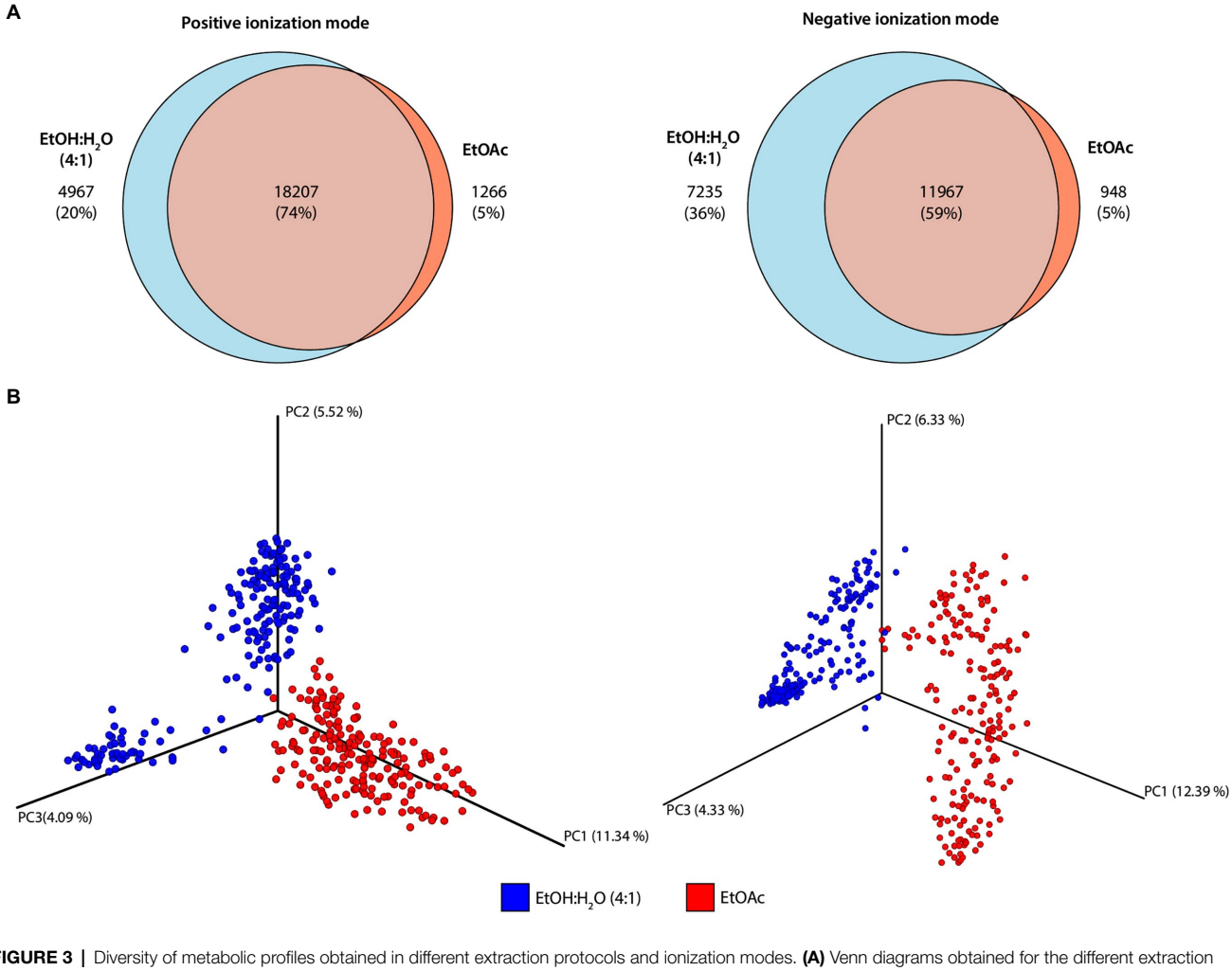


FIGURE 3 | Diversity of metabolic profiles obtained in different extraction protocols and ionization modes. **(A)** Venn diagrams obtained for the different extraction protocols in positive and negative ionization modes. **(B)** Three-dimensional Principal Coordinates Analysis (PCoA) plots of the samples analyzed in different ionization modes (positive: left; negative: right) determined by Bray–Curtis distance metric. The percentage of variance explained by the principal coordinates is presented on each axis.

regardless of the ionization mode used. These results are in accordance with previous reports since even though there is a common core metabolome, solvent-specific metabolites are likely to be observed (Crüsemann et al., 2017).

Usually, distinct classes of metabolites are extracted with solvents of different polarities (usually hydroalcoholic mixtures, methanol, ethyl acetate, or methylene chloride), which will allow the enrichment of specific classes of metabolites (such as flavonoids, coumarins, glucosides, alkaloids, diterpenes, and saponins, among others), depending on the solvent polarity (Wolfender et al., 2019; Pilon et al., 2020). In this way, different extraction solvents can be used to obtain a broader chemical diversity. Our results show that the diversity of solvent-specific metabolites will also vary depending on the ionization mode employed.

To observe the chemical space provided by the metabolomic profiles obtained by the different extraction protocols, PCoA of the positive and negative subsets were created (**Figure 3B**). In summary, differently than Principal Component Analysis (PCA) which measures correlations among the samples, PCoA analysis is used to calculate distances among them, and the way these distances are calculated can result in different clustering trends in the plots. When the Euclidean distance is used in PCoA analysis, the result will be the same as if PCA was employed (Mohammadi and Prasanna, 2003; Bauermeister et al., 2022). To evaluate the impact of different extraction protocols, PCoA plots obtained by Bray–Curtis distance metric showed that the two extraction solvents resulted in very different metabolomic profiles on both ionization modes (for positive ionization mode: PERMANOVA $F = 40.39$, $p = 0.001$; for negative ionization mode: PERMANOVA $F = 38.37$, $p = 0.001$). Even though most of the features detected are shared between the two extraction protocols, the compounds' relative abundance may significantly vary resulting in different metabolomic profiles.

In order to determine whether there would be clustering trends for each of the extracts obtained and in both ionization modes, PCoAs for these four subsets were also obtained. For the hydroethanolic extract, two main groups were formed in both ionization modes (**Supplementary Figure 1A**; for positive ionization mode: PERMANOVA $F = 28.59$, $p = 0.001$; for negative ionization mode: PERMANOVA $F = 43.92$, $p = 0.001$). For the ethyl acetate extract, a separation into two groups was obtained in the positive ionization mode (PERMANOVA $F = 18.71$, $p = 0.001$; **Supplementary Figure 2A**), while no group separation was observed in the negative ionization mode in any of the distance matrices tested (**Supplementary Figure 1B**), reinforcing that the extraction solvent and ionization mode influence data acquisition.

The clustering trends observed for the hydroethanolic extracts were optimized using the maximum likelihood criteria into the most recent molecular phylogenetic tree of Malpighiaceae (obtained from Davis and Anderson, 2010), resulting in the cladograms shown in **Supplementary Figure 1B**. These trees are graphical representations of how these different chemical profiles obtained from hydroethanolic extracts originated in specific clades (i.e., groups) through time (i.e., geological time). For the metabolites detected in the positive ionization mode, we can infer that the production of metabolites from group B originated in at least three separate geological times (firstly in part of clade A, secondly in clade C, and thirdly in the most recent common ancestor of clade E shared with all remaining clades of Malpighiaceae; i.e., clades E–J). In other words, the samples from group B are, in general, evolutionary more recent than the ones from group A. Some specific genera (such as *Amorimia*, *Alicia*, and *Banisteriopsis*), which are more recent in terms of their origins in past geological times, also cluster together with group A. A possible explanation for this tendency is that these genera are mostly found in dry environments, just like some of the early-diverging genera of Malpighiaceae (Davis et al., 2014; de Almeida et al., 2021).

On the other hand, the metabolomic profile obtained in the negative ionization mode for the hydroethanolic extracts showed a different tendency. The different chemical profiles obtained from group B in the negative ionization mode were only recovered as homoplastic to clade J (i.e., *Bronwenia*, *Diplopterys*, *Stigmaphyllon*, *Sphedamnocarpus*, and *Peixotoa*), besides also independently originated in some genera from clades A, E, F, and H.

We also evaluated the metabolomic profile obtained for the ethyl acetate extracts in both ionization modes (**Supplementary Figures 2A**,**C**). Contrary to what we previously observed, the clustering trends observed in the positive ionization mode did not show any clear correlation with the phylogeny of the Malpighiaceae family (**Supplementary Figure 1B**), while the negative ionization mode did not show any clustering trends in the metrics tested. These differences in both extraction protocols could be explained due to the different polarity ranges covered by these two solvents. From these results, we observed that different extraction protocols and ionization modes in MS can significantly impact the results from multivariate analysis and chemosystematics investigations.

Molecular Families and Metabolite Annotation

Since the chemical profiles obtained for the hydroethanolic extracts showed a promising correlation with the evolution of Malpighiaceae, we made an in-depth investigation of the data obtained in both positive and negative ionization modes. The MS/MS library search performed in GNPS resulted in 1,070 spectral matches for the positive and 1,025 matches for the negative ionization modes, resulting in 4.6 and 5.3% of the detected chemical spaces, respectively. These matches were manually evaluated and compared with the literature, resulting in level 2 or 3 annotations according to the MSI (Sumner et al., 2007). In addition, these compounds were also searched in the *Dictionary of Natural Products*, LOTUS (Rutz et al., 2021), and Scifinder databases to inspect for previous reports in Malpighiaceae.

In addition to the library searches, the MS/MS data were visualized by molecular networking analysis. The molecular families constructed by such analysis represent the similarity of fragmentation patterns obtained by tandem mass spectrometry (MS/MS) analysis. In other words, structurally similar compounds will present similar chemical stability and functional groups, leading to similar fragmentation patterns (Yang et al., 2013). These molecular families consist of nodes (representing MS/MS spectra) and of edges connecting these nodes (representing

the cosine similarity between two nodes, which measure the relatedness in MS/MS spectra; Aron et al., 2020). In this way, depending on the cosine score set in the analysis, the connections between the nodes can be more, or less strict. All the molecular families discussed in this section can be found in the **Supplementary Material**.

The library matches retrieved from the analyses obtained in the positive ionization mode showed the presence of a high diversity of classes of compounds, including *C*-glycosylated and *O*-glycosylated flavonoids, lipids, alkaloids, quinic acid derivatives, amides, triterpenes, iridoids, and lignans (**Figure 4**). Some of these chemical classes were widely detected in all phylogenetic

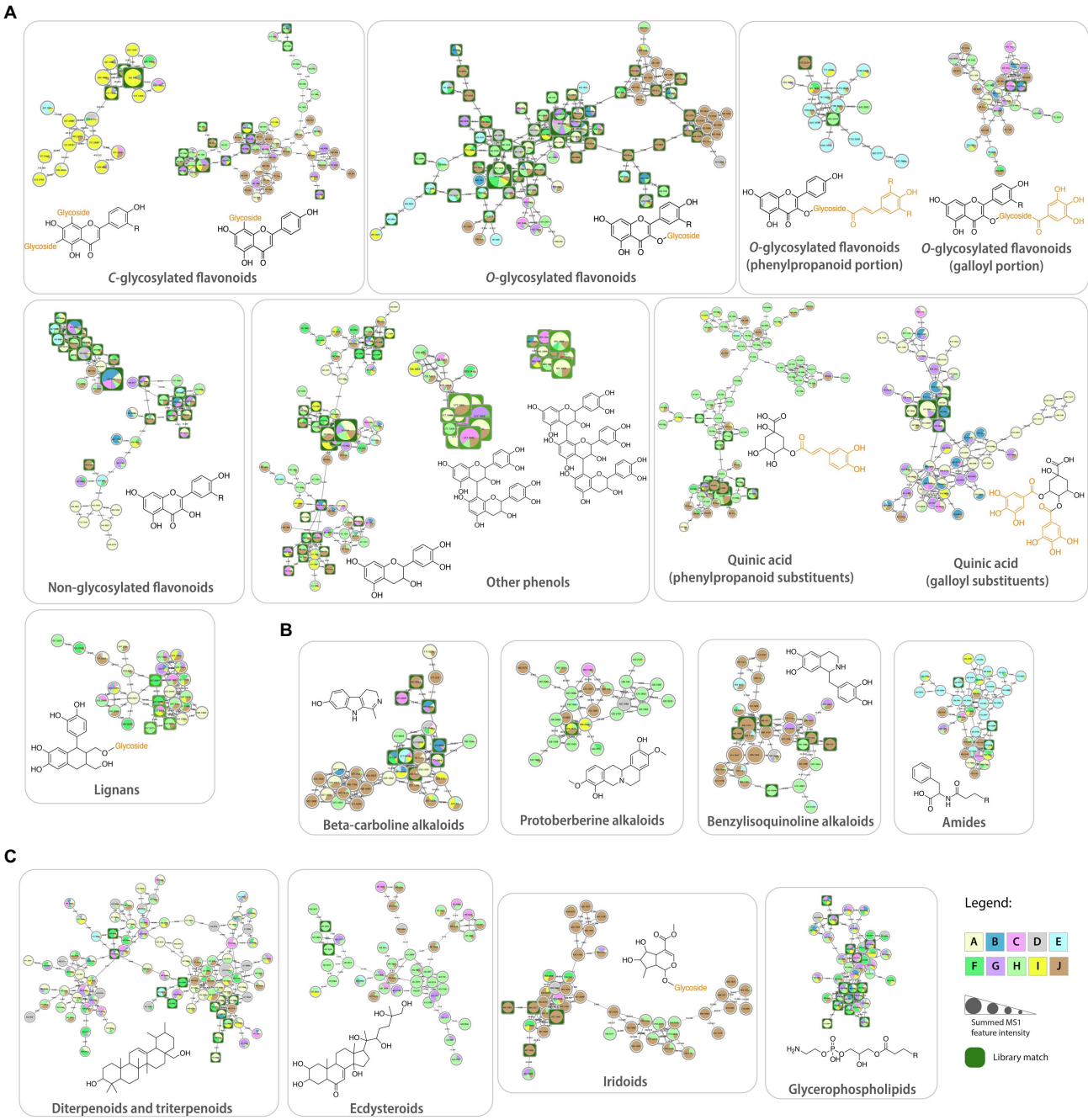


FIGURE 4 | Molecular families obtained from the Feature-Based Molecular Networking workflow and annotated based on spectral matches within the GNPS platform: **(A)** phenolic compounds, **(B)** alkaloids, and **(C)** lipids and terpenoids. Each node represents a tandem mass spectrometry spectra (MS/MS), while the edges that connect them represent the MS/MS fragmentation similarity (cosine >0.7). Pie charts indicate the relative abundance of ion features in each Malpighiaceae phylogenetic clade (A–J). Node sizes are relative to the summed peak areas of the precursor ion in MS1 scans. These are level 2 or 3 annotations according to the 2007 metabolomics standards initiative (Sumner et al., 2007).

clades, while others were more specific to particular clades or even genera.

Several molecular families related to flavonoid compounds were obtained, highlighting the difference in the fragmentation pattern in MS/MS spectra among these subclasses with different substituents. For instance, the usual MS/MS spectra obtained for *O*-glycosylated flavonoids in electrospray ionization consist of the neutral losses of the glycosidic substituents, such as hexosides (162 Da), deoxyhexosides (146 Da), and pentosides (132 Da), eventually reaching the aglycone. On the other hand, *C*-glycosylated flavonoids present a very distinct fragmentation pathway, with many more fragments observed between the precursor ion and the aglycone. The loss of the glycosidic portion is usually reached by the consecutive losses of water molecules and 120 Da, characteristic of *C*-glycosylated flavonoids (Mannochio-Russo et al., 2020). Therefore, the distinction of these two molecular families in the molecular networking analysis is expected. In addition, compounds containing two *C*-glycosidic bounds, even more fragments are expected, and we can observe two molecular families for *C*-glycosylated flavonoids.

Flavonoids with two *C*-glycosylated portions were mainly detected in samples from clade I, especially from the genera *Amorimia* and *Mascagnia* (**Supplementary Figure 3**). Library matches from this molecular family included apigenin-di-*C*-hexoside-pentoside and luteolin-di-*C*-hexoside, being in accordance with the previous reports for the *Amorimia* genus (Mannochio-Russo et al., 2020). On the other hand, spectral matches to flavonoids containing only one *C*-glycosylated portion (**Supplementary Figure 4**) were less clade-specific, being detected in samples from clades A, G, H, I, and J.

In addition to the *C*-glycosylated flavonoids, three main molecular families were observed for *O*-glycosylated ones: (1) flavonoids containing only glycosides, (2) glycosides in addition to phenylpropanoids, and (3) glycosides in addition to galloyl portions. These are also explained by the difference in general MS/MS spectra of each of these groups, in which the flavonoids containing the galloyl portion present a characteristic fragment at *m/z* 153, and the ones containing phenylpropanoids will show the fragment ions regarding this substituent (*m/z* 147 for coumaric acid, for instance). Finally, *O*-glycosylated flavonoids containing only glycosides will only present neutral losses regarding each glycosidic portion.

Several matches to *O*-glycosylated flavonoids were observed in all Malpighiaceae clades (**Supplementary Figure 5**), which is expected since this class of compounds is considered ubiquitous to plant species (Buer et al., 2010). This network is primarily composed of quercetin and kaempferol derivatives bound to glycosidic portions. A small cluster in this molecular family corresponds to compounds mostly present in clade J, specifically from *Stigmaphyllon* species. These compounds were annotated as flavonoids with glucuronide, acetylated, and malonylated hexoside substituents, which have not been reported in the literature for the genus *Stigmaphyllon* to date.

O-glycosylated flavonoids containing sugars and phenylpropanoid portions as substituents were also observed (**Supplementary Figure 6**), mainly present in the *Ptilochaeta* genus (clade E), which is in accordance with previous reports (Mannochio-Russo et al., 2020). In addition, sugars bound to galloyl portions (**Supplementary Figure 7**) were also detected as flavonoid substituents and widely found in all Malpighiaceae clades.

A network containing mainly non-glycosylated flavonoids (**Supplementary Figure 8**), widely distributed in all clades, was also observed, in which only the fragments relative to aglycones were detected in the MS/MS spectra. These compounds can represent, indeed, aglycones, or even in-source fragments of the *O*-glycosylated flavonoids. Additionally, a molecular family with spectral matches to catechin and afzelechin and their derivatives (**Supplementary Figure 9**) was also formed, containing mainly methoxylated portions and galloyl and sugar substituents.

Quinic acid derivatives were detected in higher amounts in clades H and J for phenylpropanoid substituents, and in clades A, B, and G for gallic acid substituents (**Supplementary Figures 10**, **11**). Usually, the MS/MS patterns will contain an ion relative to the quinic acid moiety (usually with additional neutral loss of water), in addition to the fragment at *m/z* 153, relative to the galloyl portion, while other characteristic fragments will be observed for the phenylpropanoid derivatives (Mannochio-Russo et al., 2020). Previous reports described the presence of galloylquinic acids in *Byrsonima* (clade A) species (Fraige et al., 2018; Mannochio-Russo et al., 2020), while quinic acids containing phenylpropanoids substituents have been described for the genus *Heteropterys* (clade H; Huerta-Reyes et al., 2013; Paula-Freire et al., 2013).

Many lipid-like molecules presented library matches, mainly corresponding to the glycerophospholipids (**Supplementary Figure 12**), fatty acids, and fatty esters (**Supplementary Figure 13**) classes, besides jasmonic acid derivatives (**Supplementary Figure 14**), which were largely distributed in all Malpighiaceae clades. Glycerophospholipids also present key fragments in MS/MS analysis, such as the cleavage of the choline group, which can generate a fragment of the choline group itself (if positively charged) and the fragment of the long-chain portion (Ivanova et al., 2007). Lipids represent an important class of compounds, widely found in plants, with key roles in multiple signaling processes (Mamode Cassim et al., 2019). Along with sugars, such compounds are among the main constituents of the oil glands present in Malpighiaceae leaves and flowers (Possobom et al., 2010). Lipids are mainly produced by oil glands that play an important ecological role in Malpighiaceae, a botanical family which is mainly pollinated by oil collecting bees, and the oldest family characterized by oil-bee pollination (Anderson, 1990; Renner and Schaefer, 2010; Davis et al., 2014).

β-carboline alkaloids and other tryptophan derivatives were detected in samples from many clades (**Supplementary Figure 15**), including *Tetrapterys* and *Banisteriopsis* species, corroborating previous reports (Samoylenko et al., 2010; Queiroz et al., 2014). In general, β-carboline alkaloids present MS/MS spectra with fragments relative to the loss of a hydroxyl group (if present) and to the formation of a four-membered ring (*m/z* 184 and *m/z* 160, respectively, harmalol as an example). A small cluster into this molecular family showed the presence of compounds

without library matches with higher *m/z* in samples from clade A and J, mainly in *Janusia*, *Banisteriopsis*, and *Byrsonima* species, which can indicate the presence of glycosylated portions due to the mass differences observed between the nodes.

Networks containing library matches corresponding to isoquinoline, protoberberine, and benzylisoquinoline alkaloids were also observed (**Supplementary Figures 16–18**, respectively). Isoquinoline alkaloids were mainly observed in samples from clades G, H, and J. Berberine alkaloids usually show a retro Diels-Alder (RDA) reaction and a B-ring cleavage, forming fragments *m/z* 151 and *m/z* 178, respectively, in the case of scoulerine (Qing et al., 2020). On the other hand, benzylisoquinoline alkaloids will usually present some key fragments, such as the initial loss of the nitrogen atom as ammonia or as methylamine (in the case of methylated nitrogens), in addition to an "even electron"-type McLafferty rearrangement with a reversed charge distribution, and a fragment relative to the benzyl moiety. Therefore, these key fragmentation pathways will generate ions at *m/z* 299, *m/z* 192, and *m/z* 137, taking reticuline as an example (Schmidt et al., 2005). The molecular family representing protoberberine alkaloids showed nodes corresponding to compounds mostly detected in clades H and J, particularly in *Stigmaphyllon* and *Alicia* species. Similarly, benzylisoquinoline alkaloids were also mainly detected in samples from clades H and J, especially in *Stigmaphyllon* species. These classes of compounds have not been described to date for Malpighiaceae species and represent important traits for chemosystematics studies in Malpighiaceae due to their specificity to the above-mentioned clades, especially for protoberberine and benzylisoquinoline alkaloids.

Spectral matches corresponding to amides and polyamines were also observed in molecular families (as shown in **Supplementary Figures 19**, **20**, respectively). These compounds can generate ions relative to the *N*-cleavage of the amide bond (Barrère et al., 2014). Amides were mainly detected in the *Ptilochaeta* genus (clade E) as fatty amides and small dipeptides, while the polyamines were primarily observed in samples from clades E, H, I, and J, being this the first report of these classes of compounds for Malpighiaceae.

Several molecular families related to terpenoids were also observed, which are characterized for presenting several fragments in MS/MS spectra depending on the general skeleton of the molecule, such as RDA, McLafferty rearrangement, water losses, among others (Demarque et al., 2016). One of the largest networks obtained (**Supplementary Figure 21**) showed library matches to triterpenoids and their precursors. These compounds were also largely distributed among all Malpighiaceae clades, in accordance with previous reports. In fact, triterpenoids are the most described class of compounds from Malpighiaceae species to date, with numerous reports in the literature for *Acridocarpus*, *Byrsonima*, and *Galphimia* genera (Cao et al., 2004; Cardoso Taketa et al., 2004; Aguiar et al., 2005).

It is important to mention at this point that the molecular networking approach highlights the chemical similarity of compounds based on their fragmentation patterns, which may not necessarily reflect their biosynthetic origin. For instance, some diterpenoids are grouped in the same molecular family as the triterpenoids, even though their biosynthesis differs from the plastidial MEP and the cytosolic mevalonate pathway. The connection between them can be observed since these two classes of compounds present similar MS/MS fragmentation patterns. A possible way to separate these compounds into two distinct molecular families would be to significantly increase the threshold for the cosine score similarity (set to 0.7 in this work). However, in the molecular networking analysis, the cosine value is set for the entire dataset, and thus, we used an intermediary value that was adequate for most of the families discussed in this work. Since different chemical classes present different key MS/MS fragments, it is very likely that while separating triterpenoids from diterpenoids, other molecular families can be fragmented into smaller networks, even though they can present significant correlations. In addition, it is important to emphasize that even though we can increase this cosine threshold, it is very likely that the diterpenoids and triterpenoids will not be perfectly separated since these two classes share several key MS/MS fragments.

A separate molecular family relative to ecdysteroids, another class of terpenoids, was also observed (**Supplementary Figure 22**), and mainly detected in samples from clade H, particularly in *Niedenzuella* species. These compounds are mainly characterized by the presence of consecutive neutral losses of water molecules, in addition to the hydrocarbon side chain cleavage (Lavrynenko et al., 2013; Mannochio-Russo et al., 2020). The presence of ecdysteroids in *Niedenzuella multiglandulosa* was recently described in the literature (Mannochio Russo et al., 2020; Mannochio-Russo et al., 2020), being the only reports of this class of compounds in the Malpighiaceae family. Our results showed a wide diversity of ecdysteroids in all six *Niedenzuella* species sampled, being possible chemical markers for the *Niedenzuella* genus. This represents about 40% of all the *Niedenzuella* species reported to date, and future studies with other *Niedenzuella* species should be performed to corroborate this hypothesis. Moreover, these steroids were also detected in *Hiraea*, *Tetrapterys* (including in the recently segregated *Glicophyllum* species, evolutionary close to *Niedenzuella*), and *Peixotoa* genera (clades G, H, and J, respectively).

Iridoids, another relevant class of terpenoids, were also annotated based on library matches, which are shown in **Supplementary Figure 23**. These compounds present key fragments relative to the neutral losses of the glycosidic portions, in addition to water losses and other possible substituents (Wu et al., 2010). The correspondent networks revealed that these iridoids are mainly present in samples from clade J, especially in *Stigmaphyllon* species, being in accordance with previous reports (Sainty et al., 1981; Davioud et al., 1985). To date, iridoids have only been reported for *Stigmaphyllon* genus in Malpighiaceae, and in this study, we observe that these compounds are also present in samples from other clades, in particular, *Heteropterys oberdanii* species (clade H). Furthermore, a molecular family with a spectral match to a secoiridoid (**Supplementary Figure 24**) was also observed, with part of the compounds widely distributed among all the clades, and part of them mainly detected in *Stigmaphyllon* species. Finally, molecular families relative to

neolignans and furofuranoid lignans (**Supplementary Figure 25**) were observed and widely distributed in Malpighiaceae clades, which have not been previously reported in this family to date.

The library matches obtained for the negative ionization mode and the molecular networks formed were also inspected, and these analyses showed mainly the same classes of compounds described for the positive ionization mode. Two molecular families stood out for presenting library matches to classes of compounds that were not observed in the positive ionization mode: proanthocyanidins and lignans (**Supplementary Figures 26**, **27**). Proanthocyanidins dimers and trimers were observed in all Malpighiaceae clades, with characteristic fragmentation patterns (such as RDA), corroborating previous reports (Fraige et al., 2018; Mannochio-Russo et al., 2020). Lignans have already been reported for *Tetrapterys mucronata* (Queiroz et al., 2014), and here we observe that, in addition to *T. mucronata* species (clade H), these compounds are also present in other clades.

It is important to emphasize that the compounds discussed above represent only a part of the compounds detected in this study. In fact, about 87% of the molecular families (composed of two nodes or more) obtained in both ionization modes did not show any spectral match, including networks composed mainly or exclusively by nodes representing specific clades/genera. These numbers point to the possibility of undescribed natural products. On the other hand, even considering only the spectral matches, it was possible to obtain important information regarding the classes of compounds produced by these plant species. Another crucial point to be considered is that the sampling in this study comprises a larger number of samples from clades A, H, I, and J, and that these clades presented most of the spectral matches observed. It is possible that if a greater number of samples from other clades are included, more spectral matches can be retrieved to give more insights on other clades as well.

In silico Metabolite Annotation and Chemical Hierarchy Analysis

In order to amplify the chemical space from the Malpighiaceae dataset and have additional information about the classes of compounds detected, we used the Qemistree workflow combined with the CANOPUS classification tool for systematic compound class annotation (Dührkop et al., 2021; Tripathi et al., 2021). These *in silico* classifications consist of level 3 annotations according to the MSI (Sumner et al., 2007). In this way, it was possible to construct a chemical tree based on molecular fingerprints from MS/MS spectra and *in silico* classification tools. A total of 7,489 and 3,773 fingerprints were generated and classified at a superclass level for the positive and negative ionization modes, respectively. In this way, two chemical hierarchy trees were obtained, as shown in **Figure 5A**.

From the results obtained, it is evident that the ionization mode employed greatly influences the classifications obtained. At a superclass level, the ones most retrieved in the positive ionization modes were the "organic acids and derivatives," followed by "benzenoids" and "organoheterocyclic compounds." In contrast, most of the superclasses retrieved in the negative ionization mode consisted of "lipids and lipid-like molecules," "organic acids and derivatives," and "benzenoids" (**Supplementary Figure 28A**). At a CANOPUS class level, the "carboxylic acids and derivatives" was the main recovered class in both ionization modes. Ion features classified as "benzene and substituted derivatives," and "azoles" were also observed several times for the positive ionization mode, while for the negative ionization mode, "organooxygen compounds" and "fatty acyls" were the second and third most abundant classifications.

These results confirm some of the conclusions drawn from the molecular networks, such as the presence of lipids and lipid-like molecules in several Malpighiaceae clades. The molecular networks and library searches in spectral libraries, combined with the *in silico* approaches based on structural databases allowed us to expand the Malpighiaceae chemical space. Several hypotheses raised from the molecular networks were corroborated with the *in silico* classifications, giving higher confidence in these results.

The classifications obtained in the different ionization modes are also shown distributed in the chromatographic run (**Figure 5B**). In addition to being possible to observe differences in classifications between ionization modes, the retention time ranges also vary for specific superclasses. For instance, in the negative ionization mode, organic oxygen compounds, and phenylpropanoids and polyketides elute from near the dead volume to approximately 4 min, while the ion features classified as lipids and lipid-like molecules present higher retention times.

The putative chemical classes retrieved as highly correlated with the most sampled clades (clades A, F, G, H, I, and J; ANOVA $p < 0.05$) were selected to build a heatmap (**Supplementary Figure 29**). The normalized distribution pattern of the different classes within the sampled genera showed that specific classes are significantly enriched in determined genera/clades, which corroborates our observations retrieved from the molecular networks. In addition, once again, the results obtained for the positive and negative ionization modes differ.

These results show that both ionization modes result in complementary chemical classifications, which is crucial for comprehensive chemotaxonomic investigations. In fact, the *in silico* tools used in this study relies on public spectral and structural databases, which are known to be more populated with data on the positive ionization mode (Tripathi et al., 2021). In this way, the results obtained for the negative ionization mode are less extensive compared to the positive ionization mode.

Ancestral Character Reconstructions for the Classes of Secondary Metabolites Annotated in Malpighiaceae

Ancestral character reconstruction analyzes have been increasingly encouraged in natural products studies since the early 2010s (Schmitt and Barker, 2009). It has been used in the chemistry of natural products to investigate chemical evolutionary relationships comprising different organisms, such as plants, fungi, and animals (Lumbsch et al., 2006; Bondoc et al., 2013; Allevato et al., 2019; Coley et al., 2019; Chen et al., 2020; Beaulieu et al., 2021). Phylogenetic methods have proven to be a promising approach to explore the evolution

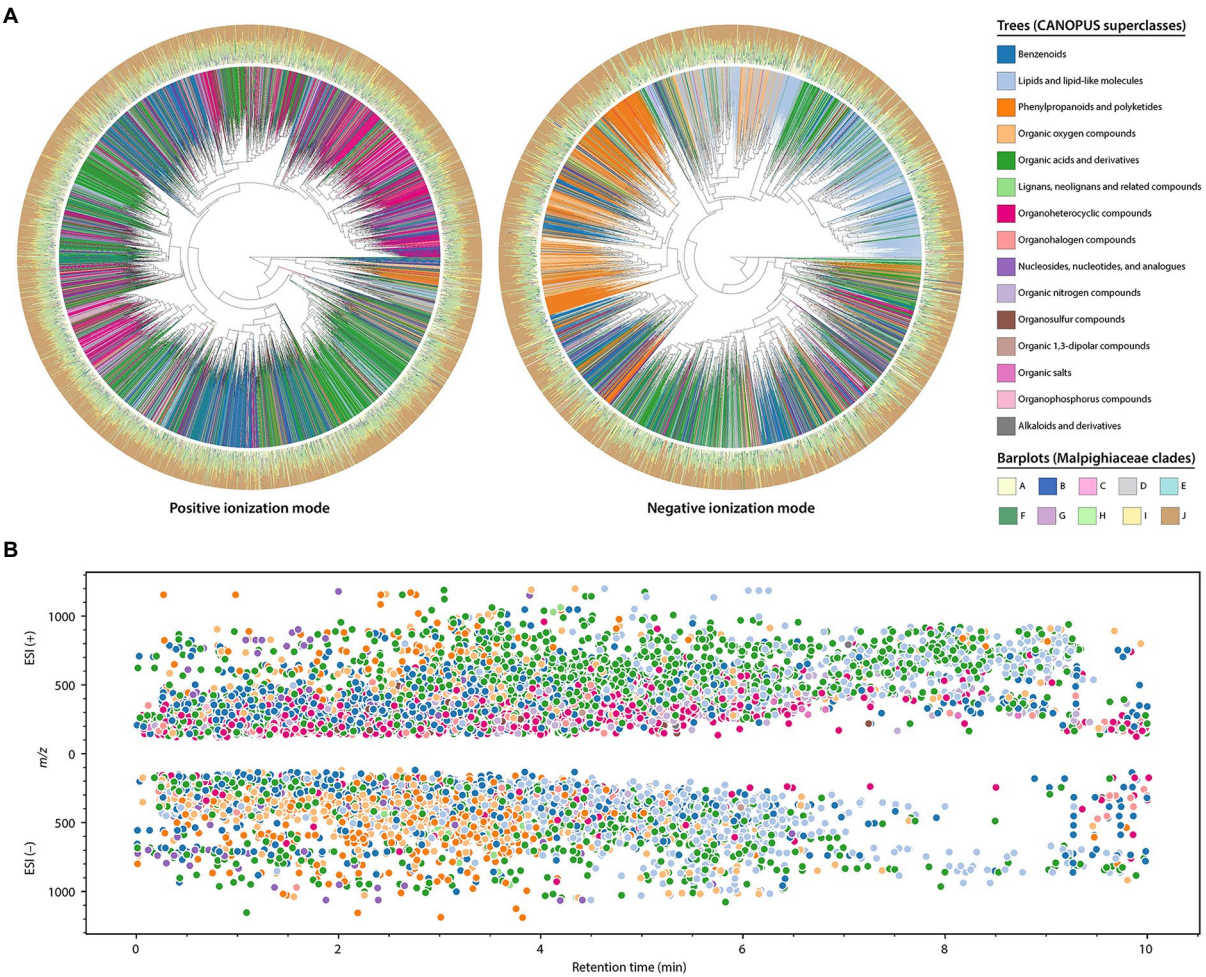


FIGURE 5 | *In silico* annotations obtained for the Malpighiaceae dataset from the Qemistree workflow combined with the CANOPUS classification tool. These are level 3 annotations according to the 2007 metabolomics standards initiative (Sumner et al., 2007). **(A)** Chemical hierarchies of the predicted molecular fingerprints from the Malpighiaceae plant samples analyzed in positive (left) and negative (right) ionization modes. The trees are pruned to keep fingerprints which were classified up to a superclass level in CANOPUS. The branch colors indicate the superclasses, while the barplots of the outer ring indicate the relative abundance of a molecular fingerprint in each Malpighiaceae clade. **(B)** The ion features classified *in silico* are mapped based on the CANOPUS superclass (same colormap described in **A**). The *x* and *y* axes indicate the retention time and *m/z* value, respectively.

of chemical compounds in a specific genus or family of plants and other living organisms. This analysis consists of optimizing binary character states (presence/absence) into a DNA-based molecular phylogenetic tree by statistically testing using Maximum Likelihood Estimation, and recovering which character states (such as chemical classes) characterizes a given clade or taxonomic group within the molecular phylogenetic tree (i.e., genera or major clades recovered by DNA-based phylogenetic studies in Malpighiaceae). In this way, we can determine the statistical probability of the ancestor of a particular genus or clade to exclusively (i.e., synapomorphy) or non-exclusively (i.e., homoplasy) show the presence or absence of a specific chemical class. These exclusive or non-exclusive chemical classes can be used in future studies to circumscribe the analyzed taxa into a new classification, which additionally considers chemical information.

In the present study, the *in silico* classifications obtained were used to map the evolutionary history of the classes retrieved in our analyses using the maximum likelihood criteria in the Malpighiaceae DNA-based molecular phylogeny (**Figure 6**). The classifications obtained in the positive and negative ionization modes were combined to provide a general overview of the total number of chemical classes obtained for the family. Classes retrieved as homoplasies or synapomorphies for the major Malpighiaceae clades (A–J) are described in **Table 1**, while the ones retrieved for all Malpighiaceae clades and genera are extensively presented in **Supplementary Table 2**. From the 113 *in silico* classes retrieved, 35 were present in all genera, such as fatty acyls, flavonoids, glycerolipids, phenols, and purine nucleosides, compounds widely distributed in plants, with a variety of ecological roles in these organisms. Nonetheless, future additional studies sampling the remaining botanical

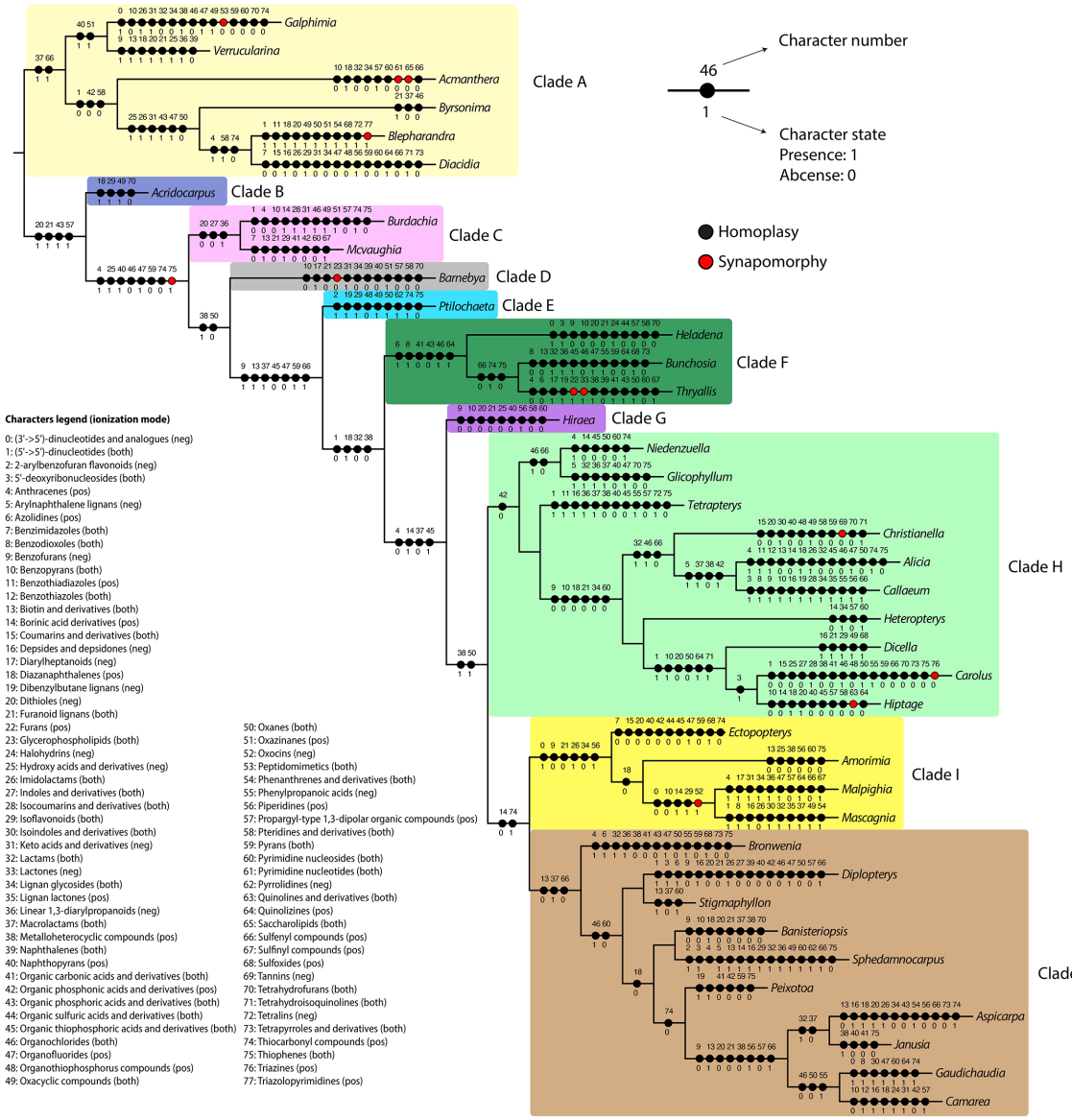


FIGURE 6 | Summary of the maximum likelihood ancestral state reconstruction for the *in silico* classifications obtained at a class level. Each chemical class was treated as a character (0–77), and character states were binary-coded for each genus (1: present; 0: absent). Black and red circles represent homoplasies and synapomorphies, respectively. Clades highlighted represent the Malpighiaceae major clades recognized by recent molecular phylogenetic studies according to Davis and Anderson (2010).

families included in the order Malpighiales are needed to properly evaluate the relevance of these 35 classes of metabolites for the Malpighiaceae family as a whole.

Regarding the 10 major clades of Malpighiaceae, all of them were recovered with at least one homoplastic or synapomorphic class of metabolites supporting them (**Table 1**; **Supplementary Table 2**). The 10 major clades in Malpighiaceae were characterized by the presence and absence of 22 and 23 classes of metabolites (i.e., homoplasies), respectively, with the absence of glycerophospholipids being recovered as a synapomorphy of Clade D. It is worth mentioning that even though most classes of secondary metabolites circumscribing all the 10 major clades of Malpighiaceae were recovered as homoplasies, their presence or absence was recorded exclusively for each clade, with reversions (i.e., parallelisms) only being recorded within a few distantly related subclades or genera. The results retrieved from this analysis corroborate some of the conclusions obtained from the molecular networks and from the *in silico* classifications. For instance, glycerophospholipids, fatty acids, fatty esters, furofuranoid lignans, and prenol lipids classes were widely distributed in our molecular networking analyses for all phylogenetic clades, and these classes were also extensively recovered in the ancestral character reconstructions (i.e., appeared in different geological

TABLE 1 | Characters retrieved from the ancestral characters reconstruction (clades) based on the classifications obtained *in silico* for Malpighiaceae samples.

	Classes present[†]	Classes absent
Clade A (Byrsonimoid clade)	Macrolactams (both); Sulfenyl compounds (pos)	–
Clade B (Acridocarpoid clade)	Diazanaphthalenes (pos); Isoflavonoids (both); Oxacyclic compounds (both)	Tetrahydrofurans (both)
Clade C (Mcvaughioid clade)	Linear 1,3-diarylpropanoids (neg)	Dithioles (neg); Indoles and derivatives (both)
Clade D (Barnebyoid clade)	Diarylheptanoids (neg); Keto acids and derivatives (neg); Oxazinanes (pos)	Benzopyrans (both); Furanoid lignans (both); Glycerophospholipids (both)[§], Lignan glycosides (both); Naphthalenes (both); Naphthopyrans (pos); Propargyl-type 1,3-dipolar organic compounds (pos); Pteridines and derivatives (both); Tetrahydrofurans (both)
Clade E (Ptilochaetoid clade)	2-arylbenzofuran flavonoids (neg); Dibenzylbutane lignans (neg); Isoflavonoids (both); Oxacyclic compounds (both); Oxanes (both); Pyrrolidines (neg); Thiocarbonyl compounds (pos)	Organothiophosphorus compounds (pos); Thiophenes (both)
Clade F (Bunchosioid clade)	Azolidines (pos); Benzodioxoles (both); Organochlorides (both); Quinolizines (pos)	Organic carbonic acids and derivatives (both); Organic phosphoric acids and derivatives (both)
Clade G (Hiraeoid clade)	Piperidines (pos)	Benzofurans (neg); Benzopyrans (both); Dithioles (neg); Furanoid lignans (both); Hydroxy acids and derivatives (neg); Naphthopyrans (pos); Pteridines and derivatives (both); Pyrimidine nucleosides (both)
Clade H (Tetrapteroid clade)	–	Organic phosphonic acids and derivatives (pos)
Clade I (Malpighioid clade)	(3′➔5′)-dinucleotides and analogues (neg); Imidolactams (both); Piperidines (pos)	Benzofurans (neg); Furanoid lignans (both); Lignan glycosides (both)
Clade J (Stigmaphylloid clade)	Macrolactams (both)	Biotin and derivatives (both); Sulfenyl compounds (pos)

The ionization mode in which each classification was obtained is described (pos, positive ionization mode; neg, negative ionization mode; and both, both ionization modes).
[†]Classes retrieved as present in all Malpighiaceae clades: Alkyl halides (both); Allyl-type 1,3-dipolar organic compounds (both); Aryl halides (both); Azacyclic compounds (both); Azoles (both); Benzene and substituted derivatives (both); Boronic acid derivatives (both); Carboxylic acids and derivatives (both); Cinnamic acids and derivatives (both); Diazinanes (pos); Diazines (both); Fatty Acyls (both); Flavonoids (both); Glycerolipids (both); Heteroaromatic compounds (both); Imidazopyrimidines (both); Macrolides and analogues (pos); Organic metal salts (pos); Organic sulfonic acids and derivatives (both); Organonitrogen compounds (both); Organooxygen compounds (both); Phenol ethers (both); Phenols (both); Prenol lipids (both); Purine nucleosides (both); Purine nucleotides (both); Pyridines and derivatives (both); Sphingolipids (both); Steroids and steroid derivatives (both); Sulfonyls (pos); and Thioethers (both).
[§]Synapomorphy.

times). In addition, it also clearly shows which classes were important to circumscribe specific clades/genera.

Our chemotaxonomic approach based on MS/MS analyses of Malpighiaceae plant samples allowed us to obtain a comprehensive overview of the classes of secondary metabolites produced by this taxon. Plant secondary metabolites are known to show patterns of occurrence in certain taxa (Wink, 2003). However, it is important to emphasize that the chemical diversity of a sample is highly influenced by many factors, such as genetic variation and environmental influences (soil nutrients, humidity, herbivory, and ecological interactions, among others; Isah, 2019). Studies have shown that even the same species collected in different biomes (Bueno et al., 2021) or different seasons (Zanatta et al., 2021) can produce different relative amounts of specific metabolites. Therefore, even for a single species, a range of factors can be explored to understand how the metabolites are affected to infer their possible ecological roles. However, it is important to emphasize that the ancestral character reconstructions obtained only focus on qualitative characters (presence and absence) rather than quantitative ones.

Our study provides a starting point for follow-up and systematic evaluation of such factors, and in-depth studies must be conducted to confirm and expand these chemotaxonomic conclusions for both Malpighiaceae and Malpighiales. In fact, there is a big jump from a large (but limited) sampling, as the one we present here, toward a chemotaxonomic investigation of the entire plant family. Nonetheless, as shown by several recent studies mentioned above, using phylogenetic methods with chemistry of natural products data is a promising and revolutionary new line of research that aims to elucidate the evolution of specialized metabolites in living organisms. Therefore, future investigations must be conducted to confirm the hypotheses raised in our study, especially for species that do not present any previous phytochemical study. In addition, our efforts were directed to obtain chemical information at a major clades and generic levels. There might be discrepancies if the ancestral character state reconstructions are obtained at different taxonomic levels, such as intrageneric and species levels.

Finally, the relevance of our evolutionary approach to the study of secondary metabolites can be evidenced if we consider as examples three relevant classes of plant secondary metabolites (**Table 1**: furanoid lignans, isoflavonoids, and piperidines). The absence of the phenylpropanoids furanoid lignans in clades D, G, and I is regarded as an informative homoplasy in our analyzes since this information can be used to point which clades in the family one must focus on in future studies to search for this class of metabolites. This information can also be used in chemosystematic studies of Malpighiaceae to chemically characterize these clades and help establish a new classification system based on chemical compounds and morphology, following de Almeida et al. (2017). On the other hand, the presence of the phenolic isoflavonoids in clades B and E, and the alkaloid class of piperidines in clades G and I are also informative homoplasies that can be interpreted in the same light as the

furanoid lignans. In addition, it is important to emphasize that the *in silico* classifications obtained rely on structural databases populated with compounds from diverse sources, including plants and microorganisms metabolites. Therefore, some *in silico* classifications of metabolites more usually found in microorganisms may occur in plant datasets (for instance, the features classified as lactams). A deeper investigation in each taxonomic group should be performed in the future to confirm their presence. In fact, the population of such databases with more compounds derived from plants is necessary to have more accurate information regarding plant species in the future.

The chemical characterization of all Malpighiaceae clades is the first step toward enabling a new research line on the evolution of secondary metabolites in this plant family since this family already has a dated and calibrated molecular phylogeny available in the literature (Wink, 2003; Davis et al., 2014). Merging both analyses would allow us to infer the geological time in which all classes of secondary metabolites have arisen in Malpighiaceae and correlate these dates with past biogeographic events, such as the colonization of different biomes by the most recent common ancestor of all analyzed clades (de Almeida et al., 2018). Additionally, it would also be possible to evaluate which Malpighiaceae lineages experienced a higher diversification throughout the geological time and identify, and which classes of secondary metabolites are correlated with these diversification events (Xi et al., 2012).

CONCLUSION

Metabolomics analyses based on tandem mass spectrometry and bioinformatics tools have enabled a more comprehensive investigation of the metabolites produced by organisms, and have been increasingly used for this purpose. In fact, due to the low amount of material necessary, it is possible to investigate entire families for chemoevolutionary studies based on the remaining samples from molecular phylogenies used for total DNA extraction. Even though there was a significant advance in this field, many caveats must be considered for proper use. The ionization mode and extraction protocols must be carefully evaluated since these factors influence the results, especially when aiming at chemotaxonomic investigations. Our results showed that positive and negative ionization modes lead to complementary results both in library searches and *in silico* classification tools. However, as the public libraries are more populated with data acquired in the positive ionization mode, less extensive information can be retrieved from analyses performed in the negative ionization mode. More complete and precise results will certainly be obtained for chemotaxonomic studies as these databases get more populated and new bioinformatic tools are developed. Similarly, conclusions obtained from *in silico* approaches must be confirmed with complementary techniques, and the classical methodologies are of great value for deeper investigations. In addition, the population of structural databases with more plant-derived compounds will be of great value to have more accurate results for these organisms.

Our study explored several Malpighiaceae plant species, genera, and clades for the first time, which greatly improved the chemical knowledge of this family. There are several challenges in performing chemotaxonomic investigations at a plant family level, and the evolutionary conclusions retrieved must be carefully inspected; however, they can be of great value to underpin interesting features in the chemodiversity of a certain taxon. We hope that our findings guide future studies in Malpighiaceae as we reported evidence of specific classes of compounds that most likely occur in specific clades or genera. Therefore, if a particular chemical class is of interest for presenting specific biological activities, one can focus their search on specific groups pointed out in this study. All the software and libraries used in this study are publicly available, making this workflow accessible to be reproduced in other taxa. In addition, we expect that the workflow followed in this study will be used in future studies in several fields, such as chemotaxonomy, metabolomics, chemical ecology, and for the discovery of new natural products.

DATA AVAILABILITY STATEMENT

The mass spectrometry data can be accessed on the Mass spectrometry Interactive Virtual Environment (MassIVE) as the dataset MSV000085119, which is publicly available. The Feature-Based Molecular Networking jobs on GNPS can be accessed online at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2c5f11403ac847a298e4d7866a491143 and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=501c16500476451f978311057266fbdf for positive and negative ionization modes, respectively.

The Cytoscape visualization files, the feature tables (.csv) and ion information (.mgf) files exported from MZmine2, and the files generated in the Qemistree workflow in Qiime2 are also available in MSV000085119. The chemical hierarchy tree for interactive visualization and scripts used in this project are available in https://github.com/helenamrusso/Malpighiaceae_supplementary.

AUTHOR CONTRIBUTIONS

HM-R designed the study, performed the extraction and LC–MS/MS analyses, processed and analyzed the MS/MS data, performed the ancestral character reconstructions, inspected the results, wrote the manuscript, and revised the manuscript. RA designed the study, collected and/or identified the species, performed the ancestral character reconstructions, inspected the results, wrote the manuscript, and revised the manuscript. WDGN assisted in the MS/MS data analysis, inspected the results, and revised the manuscript. PCPB inspected the results and revised the manuscript. AC-R performed the extraction and LC–MS/MS analyses, inspected the results, and revised the manuscript. AB inspected the results and revised the manuscript. PD designed the study, acquired funding, and revised the manuscript. VB designed the study, acquired funding, and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by São Paulo Research Foundation (FAPESP-CEPID, #2013/07600-3 and FAPESP-INCT, #2014/50926-0) and the Brazilian Council for Scientific and Technological Development (CNPq-INCT, #2014/465637-0). This research was supported by resources supplied by the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP). HM-R acknowledges CNPq (#142014/2018-4) and the Brazilian Fulbright Commission for the scholarships provided. PCPB and AB acknowledge FAPESP (grants #2017/19702-6, #2019/08477-7, and #2018/24865-4) for the research grants and scholarships provided. AC-R and PD were supported by the Gordon and Betty Moore Foundation through grant GBMF7622, the U.S. National Institutes of Health for the Center (P41 GM103484 and R01 GM107550), and Federal Award DE-SC0021340 subaward 1070261-436503.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.854842/full#supplementary-material

REFERENCES

Aguiar, R. M., David, J. P., and David, J. M. (2005). Unusual naphthoquinones, catechin and triterpene from *Byrsonima microphylla*. *Phytochemistry* 66, 2388–2392. doi: 10.1016/j.phytochem.2005.07.011

Allevato, D. M., Groppo, M., Kiyota, E., Mazzafera, P., and Nixon, K. C. (2019). Evolution of phytochemical diversity in *Pilocarpus* (Rutaceae). *Phytochemistry* 163, 132–146. doi: 10.1016/j.phytochem.2019.03.027

Anderson, W. R. (1990). The origin of the Malpighiaceae-The evidence from morphology. *Mem. N. Y. Bot. Gard.* 64, 210–224.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Anderson, W. R. (2006). Eight segregates from the Neotropical genus *Mascagnia* (Malpighiaceae). *Novon J. Bot. Nomenc.* 16, 168–204. doi: 10.3417/1055-3177(2006)16[168:ESFTNG]2.0.CO;2

Anderson, C. (2011). Revision of *Ryssopterys* and transfer to *Stigmaphyllon* (Malpighiaceae). *Blumea* 56, 73–104. doi: 10.3767/000651911X573444

Anderson, W. R., and Davis, C. C. (2006). Expansion of *Diplopterys* at the expense of *Banisteriopsis* (Malpighiaceae). *Harv. Pap. Bot.* 11, 1–16. doi: 10.3100/1043-4534(2006)95[1:EODATE]2.0.CO;2

Angiosperm Phylogeny Group (1998). An ordinal classification for the families of flowering plants. *Ann. Mo. Bot. Gard.* 85, 531–553. doi: 10.2307/2992015

Angiosperm Phylogeny Group (2003). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141, 399–436. doi: 10.1046/j.1095-8339.2003.t01-1-00158.x

Angiosperm Phylogeny Group (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Angiosperm Phylogeny GroupChase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

Aron, A. T., Gentry, E. C., McPhail, K. L., Nothias, L.-F., Nothias-Esposito, M., Bouslimani, A., et al. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* 15, 1954–1991. doi: 10.1038/s41596-020-0317-5

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M.International Natural Product Sciences Taskforce, and Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* 20, 200–216. doi: 10.1038/s41573-020-00114-z

Barrère, C., Hubert-Roux, M., Afonso, C., Rejaibi, M., Kebir, N., Désilles, N., et al. (2014). Tandem mass spectrometry of low solubility polyamides. *Anal. Chim. Acta* 808, 3–9. doi: 10.1016/j.aca.2013.03.064

Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K., and Dorrestein, P. C. (2022). Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* 20, 143–160. doi: 10.1038/s41579-021-00621-9

Beaulieu, W. T., Panaccione, D. G., Quach, Q. N., Smoot, K. L., and Clay, K. (2021). Diversification of ergot alkaloids and heritable fungal symbionts in morning glories. *Commun. Biol.* 4:1362. doi: 10.1038/s42003-021-02870-z

Belwal, T., Devkota, H. P., Hassan, H. A., Ahluwalia, S., Ramadan, M. F., Mocan, A., et al. (2018). Phytopharmacology of Acerola (*Malpighia* spp.) and its potential as functional food. *Trends Food Sci. Technol.* 74, 99–106. doi: 10.1016/j.tifs.2018.01.014

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bondoc, K. G. V., Lee, H., Cruz, L. J., Lebrilla, C. B., and Juinio-Meñez, M. A. (2013). Chemical fingerprinting and phylogenetic mapping of saponin congeners from three tropical holothurian sea cucumbers. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 166, 182–193. doi: 10.1016/j.cbpb.2013.09.002

Bueno, P. C. P., Abarca, L. F. S., Anhesine, N. B., Giffoni, M. S., Pereira, F. M. V., Torres, R. B., et al. (2021). Infraspecific chemical variability and biological activity of *Casearia sylvestris* from different Brazilian biomes. *Planta Med.* 87, 148–159. doi: 10.1055/a-1301-0183

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* 52, 98–111. doi: 10.1111/j.1744-7909.2010.00905.x

Cai, L., Xi, Z., Amorim, A. M., Sugumaran, M., Rest, J. S., Liu, L., et al. (2019). Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* 221, 565–576. doi: 10.1111/nph.15357

Cameron, K. M., Chase, M. W., Anderson, W. R., and Hills, H. G. (2001). Molecular systematics of Malpighiaceae: evidence from plastid rbcL and matK sequences. *Am. J. Bot.* 88, 1847–1862. doi: 10.2307/3558361

Cantrell, K., Fedarko, M. W., Rahman, G., McDonald, D., Yang, Y., Zaw, T., et al. (2021). EMPress enables tree-guided, interactive, and exploratory analyses of multi-omic data sets. *mSystems* 6, e01216–e01220. doi: 10.1128/mSystems.01216-20

Cao, S., Guza, R. C., Miller, J. S., Andriantsiferana, R., Rasamison, V. E., and Kingston, D. G. I. (2004). Cytotoxic triterpenoids from *Acridocarpus vivy* from the Madagascar rain forest. *J. Nat. Prod.* 67, 986–989. doi: 10.1021/np040058h

Cardoso Taketa, A. T., Lozada-Lechuga, J., Fragoso-Serrano, M., Villarreal, M. L., and Pereda-Miranda, R. (2004). Isolation of nor-secofriedelanes from the sedative extracts of *Galphimia glauca*. *J. Nat. Prod.* 67, 644–649. doi: 10.1021/np0304666

Chen, Y., Mulder, P. P. J., Schaap, O., Memelink, J., Klinkhamer, P. G. L., and Vrieling, K. (2020). The evolution of pyrrolizidine alkaloid diversity among and within Jacobaea species. *J. Syst. Evol.* 60, 361–376. doi: 10.1111/jse.12671

Coley, P. D., Endara, M.-J., Ghabash, G., Kidner, C. A., Nicholls, J. A., Pennington, R. T., et al. (2019). Macroevolutionary patterns in overexpression of tyrosine: an anti-herbivore defence in a speciose tropical tree genus, *Inga* (Fabaceae). *J. Ecol.* 107, 1620–1632. doi: 10.1111/1365-2745.13208

Creydt, M., and Fischer, M. (2017). Plant metabolomics: maximizing metabolome coverage by optimizing mobile phase additives for nontargeted mass spectrometry in positive and negative electrospray ionization mode. *Anal. Chem.* 89, 10474–10486. doi: 10.1021/acs.analchem.7b02592

Crüsemann, M., O'Neill, E. C., Larson, C. B., Melnik, A. V., Floros, D. J., da Silva, R. R., et al. (2017). Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J. Nat. Prod.* 80, 588–597. doi: 10.1021/acs.jnatprod.6b00722

Davioud, E., Bailleul, F., Delaveau, P., and Jacquemin, H. (1985). Iridoids of guyanese species of *Stigmaphyllon*. *Planta Med.* 51:78. doi: 10.1055/s-2007-969406

Davis, C. C., and Anderson, W. R. (2010). A complete generic phylogeny of Malpighiaceae inferred from nucleotide sequence data and morphology. *Am. J. Bot.* 97, 2031–2048. doi: 10.3732/ajb.1000146

Davis, C. C., Anderson, W. R., and Donoghue, M. J. (2001). Phylogeny of Malpighiaceae: evidence from chloroplast ndhF and trnl-F nucleotide sequences. *Am. J. Bot.* 88, 1830–1846. doi: 10.2307/3558360

Davis, C. C., Marinho, L. C., and Amorim, A. M. (2020). *Andersoniodoxa*, a replacement name for *Andersoniella* (Malpighiaceae). *Phytotaxa* 470, 121–122. doi: 10.11646/phytotaxa.470.1.9

Davis, C. C., Schaefer, H., Xi, Z., Baum, D. A., Donoghue, M. J., and Harmon, L. J. (2014). Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5914–5919. doi: 10.1073/pnas.1403157111

de Almeida, R. F., Amorim, A. M., da Silva Corrêa, A. M., and Van den Berg, C. (2017). A new infrageneric classification for *Amorimia* (Malpighiaceae) based on morphological, phytochemical and molecular evidence. *Phytotaxa* 313, 231–248. doi: 10.11646/phytotaxa.313.3.1

de Almeida, R. F., Amorim, A. M. A., and van den Berg, C. (2018). Timing the origin and past connections between Andean and Atlantic seasonally dry tropical forests in South America: insights from the biogeographical history of *Amorimia* (Malpighiaceae). *Taxon* 67, 739–751. doi: 10.12705/674.4

de Almeida, R. F., Francener, A., Pessoa, C., Sebastiani, R., Oliveira, Y. R., Amorim, A. M. A., et al. (2021). Malpighiaceae. Flora do Brasil 2020 em construção. Available at: http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB155 (Accessed January 13, 2022).

de Almeida, R. F., and van den Berg, C. (2020). Biogeography of stigmaphyllon (Malpighiaceae) and a meta-analysis of vascular plant lineages diversified in the Brazilian Atlantic rainforests point to the late Eocene origins of this megadiverse biome. *Plan. Theory* 9:1569. doi: 10.3390/plants9111569

de Almeida, R. F., and van den Berg, C. (2021). Molecular phylogeny and character mapping support generic adjustments in the Tetrapteroid clade (Malpighiaceae). *Nord. J. Bot.* 39:2876. doi: 10.1111/njb.02876

de Pinna, M. C. C. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7, 367–394. doi: 10.1111/j.1096-0031.1991.tb00045.x

De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778–791. doi: 10.1038/nprot.2007.95

Demarque, D. P., Crotti, A. E. M., Vessecchi, R., Lopes, J. L. C., and Lopes, N. P. (2016). Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep.* 33, 432–455. doi: 10.1039/c5np00073d

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302. doi: 10.1038/s41592-019-0344-8

Dührkop, K., Nothias, L.-F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M. A., et al. (2021). Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* 39, 462–471. doi: 10.1038/s41587-020-0740-8

Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12580–12585. doi: 10.1073/pnas.1509788112

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Edler, D., Klein, J., Antonelli, A., and Silvestro, D. (2021). raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* 12, 373–377. doi: 10.1111/2041-210X.13512

Ernst, M., Nothias, L.-F., van der Hooft, J. J. J., Silva, R. R., Saslis-Lagoudakis, C. H., Grace, O. M., et al. (2019). Assessing specialized metabolite diversity in the cosmopolitan plant genus *Euphorbia* L. *Front. Plant Sci.* 10:846. doi: 10.3389/fpls.2019.00846

Floros, D. J., Jensen, P. R., Dorrestein, P. C., and Koyama, N. (2016). A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* 12:145. doi: 10.1007/s11306-016-1087-5

Fraige, K., Dametto, A. C., Zeraik, M. L., de Freitas, L., Saraiva, A. C., Medeiros, A. I., et al. (2018). Dereplication by HPLC-DAD-ESI-MS/MS and screening for biological activities of *Byrsonima* species (Malpighiaceae). *Phytochem. Anal.* 29, 196–204. doi: 10.1002/pca.2734

Gallon, M. E., Monge, M., Casoti, R., Da Costa, F. B., Semir, J., and Gobbo-Neto, L. (2018). Metabolomic analysis applied to chemosystematics and evolution of megadiverse Brazilian *Vernonieae* (Asteraceae). *Phytochemistry* 150, 93–105. doi: 10.1016/j.phytochem.2018.03.007

Gemperline, E., Keller, C., and Li, L. (2016). Mass spectrometry in plant-omics. *Anal. Chem.* 88, 3422–3434. doi: 10.1021/acs.analchem.5b02938

Gong, Y.-X., Jiang, N., Zhang, Z.-R., Tian, Y.-H., and Yu, W.-B. (2020). Complete plastome sequence of *Aspidopterys obcordata* (Malpighiaceae), a traditional Dai medicinal plant from Xishuangbanna, Yunnan, China. *Mitochondrial DNA B Resour.* 5, 3533–3535. doi: 10.1080/23802359.2020.1827998

Huerta-Reyes, M., Herrera-Ruiz, M., González-Cortazar, M., Zamilpa, A., León, E., Reyes-Chilpa, R., et al. (2013). Neuropharmacological in vivo effects and phytochemical profile of the extract from the aerial parts of *Heteropterys brachiata* (L.) DC. (Malpighiaceae). *J. Ethnopharmacol.* 146, 311–317. doi: 10.1016/j.jep.2012.12.049

Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52:39. doi: 10.1186/s40659-019-0246-3

Ivanova, P. T., Milne, S. B., Byrne, M. O., Xiang, Y., and Brown, H. A. (2007). Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry. *Methods Enzymol.* 432, 21–57. doi: 10.1016/S0076-6879(07)32002-8

Jo, S., Kim, Y.-K., Cheon, S.-H., and Kim, K.-J. (2019). The complete plastome sequence from the family Malpighiaceae, *Bunchosia argentea* (Jacq.) DC. *Mitochondrial DNA B Resour.* 4, 1027–1029. doi: 10.1080/23802359.2019.1584065

Kang, K. B., Ernst, M., Hooft, J. J. J., Silva, R. R., Park, J., Medema, M. H., et al. (2019). Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. *Plant J.* 98, 1134–1144. doi: 10.1111/tpj.14292

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

Lavrynenko, O., Nedielkov, R., Möller, H. M., and Shevchenko, A. (2013). Girard derivatization for LC-MS/MS profiling of endogenous ecdysteroids in *Drosophila*. *J. Lipid Res.* 54, 2265–2272. doi: 10.1194/jlr.D035949

Lee, S. T., Cook, D., Pfister, J. A., Allen, J. G., Colegate, S. M., Riet-Correa, F., et al. (2014). Monofluoroacetate-containing plants that are potentially toxic to livestock. *J. Agric. Food Chem.* 62, 7345–7354. doi: 10.1021/jf500563h

Li, D., and Gaquerel, E. (2021). Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. *Annu. Rev. Plant Biol.* 72, 867–891. doi: 10.1146/annurev-arplant-071720-114836

Ludwig, M., Nothias, L.-F., Dührkop, K., Koester, I., Fleischauer, M., Hoffmann, M. A., et al. (2020). Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat. Mach. Intell.* 2, 629–641. doi: 10.1038/s42256-020-00234-6

Lumbsch, H. T., Schmitt, I., Barker, D., and Pagel, M. (2006). Evolution of micromorphological and chemical characters in the lichen-forming fungal family Pertusariaceae. *Biol. J. Linn. Soc. Lond.* 89, 615–626. doi: 10.1111/j.1095-8312.2006.00696.x

Maddison, W. P., and Maddison, D. R. (2006). Mesquite: a modular system for evolutionary analysis. Available at: https://www.mesquiteproject.org/ (Accessed March 30, 2020).

Mamode Cassim, A., Gouguet, P., Gronnier, J., Laurent, N., Germain, V., Grison, M., et al. (2019). Plant lipids: key players of plasma membrane organization and function. *Prog. Lipid Res.* 73, 1–27. doi: 10.1016/j.plipres.2018.11.002

Mannochio Russo, H., Ferreira Queiroz, E., Marcourt, L., Rutz, A., Allard, P.-M., de Almeida, R. F., et al. (2020). Phytochemical analysis of the methanolic leaves extract of *Niedenzuella multiglandulosa* (Malpighiaceae), a plant species toxic to cattle in Brazil. *Phytochem. Lett.* 37, 10–16. doi: 10.1016/j.phytol.2020.02.005

Mannochio-Russo, H., Bueno, P. C. P., Bauermeister, A., de Almeida, R. F., Dorrestein, P. C., Cavalheiro, A. J., et al. (2020). Can statistical evaluation tools for chromatographic method development assist in the natural products workflow? A case study on selected species of the plant family Malpighiaceae. *J. Nat. Prod.* 83, 3239–3249. doi: 10.1021/acs.jnatprod.0c00495

Martucci, M. E. P., Loeuille, B., Pirani, J. R., and Gobbo-Neto, L. (2018). Comprehensive untargeted metabolomics of Lychnophorinae subtribe (Asteraceae: Vernonieae) in a phylogenetic context. *PLoS One* 13:e0190104. doi: 10.1371/journal.pone.0190104

Medema, M. H. (2021). The year 2020 in natural product bioinformatics: an overview of the latest tools and databases. *Nat. Prod. Rep.* 38, 301–306. doi: 10.1039/d0np00090f

Menezes, A. P. A., Resende-Moreira, L. C., Buzatti, R. S. O., Nazareno, A. G., Carlsen, M., Lobo, F. P., et al. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci. Rep.* 8:2210. doi: 10.1038/s41598-018-20189-4

Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Sci.* 43, 1235–1248. doi: 10.2135/cropsci2003.1235

Neri-Numa, I. A., Soriano Sancho, R. A., Pereira, A. P. A., and Pastore, G. M. (2018). Small Brazilian wild fruits: nutrients, bioactive compounds, health-promotion properties and commercial interest. *Food Res. Int.* 103, 345–360. doi: 10.1016/j.foodres.2017.10.053

Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285

Nixon, K. C. (1999). Winclada 1.0. Available at: http://www.diversityoflife.org/winclada/ (Accessed March 30, 2020).

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* 17, 905–908. doi: 10.1038/s41592-020-0933-6

Paula-Freire, L. I. G., Mendes, F. R., Molska, G. R., Duarte-Almeida, J. M., and Carlini, E. A. (2013). Comparison of the chemical composition and biological effects of the roots, branches and leaves of *Heteropterys tomentosa* A Juss. *J. Ethnopharmacol.* 145, 647–652. doi: 10.1016/j.jep.2012.12.004

Pilon, A., Selegato, D., Fernandes, R., Bueno, P., Pinho, D., Carnevale Neto, F., et al. (2020). Metabolômica de plantas: métodos e desafios. *Quim Nova.* 43, 329–354. doi: 10.21577/0100-4042.20170499

Possobom, C. C. F., Guimarães, E., and Machado, S. R. (2010). Leaf glands act as nectaries in *Diplopterys pubipetala* (Malpighiaceae). *Plant Biol.* 12, 863–870. doi: 10.1111/j.1438-8677.2009.00304.x

Qing, Z., Xu, Y., Yu, L., Liu, J., Huang, X., Tang, Z., et al. (2020). Investigation of fragmentation behaviours of isoquinoline alkaloids by mass spectrometry combined with computational chemistry. *Sci. Rep.* 10:733. doi: 10.1038/s41598-019-57406-7

Queiroz, M. M. F., Queiroz, E. F., Zeraik, M. L., Ebrahimi, S. N., Marcourt, L., Cuendet, M., et al. (2014). Chemical composition of the bark of *Tetrapterys 36ucronate* and identification of acetylcholinesterase inhibitory constituents. *J. Nat. Prod.* 77, 650–656. doi: 10.1021/np401003p

Ramachandran, P., Zhang, N., McLaughlin, W. B., Luo, Y., Handy, S., Duke, J. A., et al. (2018). Sequencing the vine of the soul: full chloroplast genome sequence of *Banisteriopsis caapi*. *Genome Announc.* 6, e00203–e00218. doi: 10.1128/genomeA.00203-18

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). Tracer v1.6. Available at: http://beast.bio.ed.ac.uk/Tracer (Accessed March 30, 2020).

Renner, S. S., and Schaefer, H. (2010). The evolution and loss of oil-offering flowers: new insights from dated phylogenies for angiosperms and bees. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 423–435. doi: 10.1098/rstb.2009.0229

Riet-Correa, F., Medeiros, R. M. T., and Schild, A. L. (2012). A review of poisonous plants that cause reproductive failure and malformations in the ruminants of Brazil. *J. Appl. Toxicol.* 32, 245–254. doi: 10.1002/jat.1754

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Graham, J., et al. (2021). Open natural products research: Curation and dissemination of biological occurrences of chemical structures through Wikidata. bioRxiv [Preprint]. doi: 10.1101/2021.02.28.433265

Sainty, D., Bailleul, F., Delaveau, P., and Jacquemin, H. (1981). Malpighiacees: Nouvelle Famille a Iridoides Etude du *Stigmaphyllon sagittatum*. *J. Nat. Prod.* 44, 576–578. doi: 10.1021/np50017a012

Samoylenko, V., Rahman, M. M., Tekwani, B. L., Tripathi, L. M., Wang, Y.-H., Khan, S. I., et al. (2010). *Banisteriopsis caapi*, a unique combination of MAO inhibitory and antioxidative constituents for the activities relevant to neurodegenerative disorders and Parkinson's disease. *J. Ethnopharmacol.* 127, 357–367. doi: 10.1016/j.jep.2009.10.030

Schmidt, J., Raith, K., Boettcher, C., and Zenk, M. H. (2005). Analysis of benzylisoquinoline-type alkaloids by electrospray tandem mass spectrometry and atmospheric pressure photoionization. *Eur. J. Mass Spectrom.* 11, 325–333. doi: 10.1255/ejms.745

Schmitt, I., and Barker, F. K. (2009). Phylogenetic methods in natural product research. *Nat. Prod. Rep.* 26, 1585–1602. doi: 10.1039/b910458p

Sereno, P. C. (2007). Logical basis for morphological characters in phylogenetics. *Cladistics* 23, 565–587. doi: 10.1111/j.1096-0031.2007.00161.x

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

Steckel, A., and Schlosser, G. (2019). An organic chemist's guide to electrospray mass spectrometric structure elucidation. *Molecules* 24:611. doi: 10.3390/molecules24030611

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Tripathi, A., Vázquez-Baeza, Y., Gauglitz, J. M., Wang, M., Dührkop, K., Nothias-Esposito, M., et al. (2021). Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat. Chem. Biol.* 17, 146–151. doi: 10.1038/s41589-020-00677-3

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience* 2:16. doi: 10.1186/2047-217x-2-16

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64, 3–19. doi: 10.1016/S0031-9422(03)00300-5

Wolfender, J.-L., Nuzillard, J.-M., van der Hooft, J. J. J., Renault, J.-H., and Bertrand, S. (2019). Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry and NMR profiling, in Silico databases, and Chemometrics. *Anal. Chem.* 91, 704–742. doi: 10.1021/acs.analchem.8b05112

Wu, Q., Yuan, Q., Liu, E.-H., Qi, L.-W., Bi, Z.-M., and Li, P. (2010). Fragmentation study of iridoid glycosides and phenylpropanoid glycosides in radix Scrophulariae by rapid resolution liquid chromatography with diode-array detection and electrospray ionization time-of-flight mass spectrometry. *Biomed. Chromatogr.* 24, 808–819. doi: 10.1002/bmc.1368

Xi, Z., Ruhfel, B. R., Schaefer, H., Amorim, A. M., Sugumaran, M., Wurdack, K. J., et al. (2012). Phylogenomics and a posteriori data partitioning resolve the cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17519–17524. doi: 10.1073/pnas.1205818109

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., et al. (2013). Molecular networking as a dereplication strategy. *J. Nat. Prod.* 76, 1686–1699. doi: 10.1021/np400413s

Zanatta, A. C., Vilegas, W., and Edrada-Ebel, R. (2021). UHPLC-(ESI)-HRMS and NMR-based metabolomics approach to access the seasonality of *Byrsonima intermedia* and *Serjania marginata* From Brazilian Cerrado Flora diversity. *Front. Chem.* 9:710025. doi: 10.3389/fchem.2021.710025

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpls.2022.892907

OPEN ACCESS

Edited by:
Sheng-Hong Li,
Kunming Institute of Botany (CAS), China

Reviewed by:
Shan Lu,
Nanjing University, China
Sibongile Mafu,
University of Massachusetts Amherst, United States

***Correspondence:**
Axel Schmidt
aschmidt@ice.mpg.de

†Present addresses:
Raimund Nagel,
Department of Plant Physiology, Institute of Biology, Leipzig University, Leipzig, Germany

Almuth Hammerbacher,
Department of Zoology and Entomology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

Michael A. Phillips,
Department of Biology, University of Toronto Mississauga, Mississauga, ON, Canada

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

Received: *09 March 2022*
Accepted: *11 April 2022*
Published: *06 May 2022*

Citation:
Nagel R, Hammerbacher A, Kunert G, Phillips MA, Gershenzon J and Schmidt A (2022) Bark Beetle Attack History Does Not Influence the Induction of Terpene and Phenolic Defenses in Mature Norway Spruce (Picea abies) Trees by the Bark Beetle-Associated Fungus Endoconidiophora polonica. Front. Plant Sci. 13:892907. doi: 10.3389/fpls.2022.892907

Bark Beetle Attack History Does Not Influence the Induction of Terpene and Phenolic Defenses in Mature Norway Spruce (*Picea abies*) Trees by the Bark Beetle-Associated Fungus *Endoconidiophora polonica*

Raimund Nagel†, Almuth Hammerbacher†, Grit Kunert, Michael A. Phillips†, Jonathan Gershenzon and Axel Schmidt*

Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

Terpenes and phenolics are important constitutive and inducible conifer defenses against bark beetles and their associated fungi. In this study, the inducible defenses of mature Norway spruce (*Picea abies*) trees with different histories of attack by the spruce bark beetle, *Ips typographus* were tested by inoculation with the *I. typographus*-associated fungus *Endoconidiophora polonica*. We compared trees that had been under previous attack with those under current attack and those that had no record of attack. After fungal inoculation, the concentrations of mono-, sesqui-, and diterpenes in bark increased 3- to 9-fold. For the phenolics, the flavan-3-ols, catechin, and gallocatechin, increased significantly by 2- and 5-fold, respectively, while other flavonoids and stilbenes did not. The magnitudes of these inductions were not influenced by prior bark beetle attack history for all the major compounds and compound classes measured. Before fungal inoculation, the total amounts of monoterpenes, diterpenes, and phenolics (constitutive defenses) were greater in trees that had been previously attacked compared to those under current attack, possibly a result of previous induction. The transcript levels of many genes involved in terpene formation (isoprenyl diphosphate synthases and terpene synthases) and phenolic formation (chalcone synthases) were significantly enhanced by fungal inoculation suggesting *de novo* biosynthesis. Similar inductions were found for the enzymatic activity of isoprenyl diphosphate synthases and the concentration of their prenyl diphosphate products after fungal inoculation. Quantification of defense hormones revealed a significant induction of the jasmonate pathway, but not the salicylic acid pathway after fungal inoculation. Our data highlight the coordinated induction of terpenes and phenolics in spruce upon infection by *E. polonica*, a fungal associate of the bark beetle *I. typographus*, but provide no evidence for the priming of these defense responses by prior beetle attack.

Keywords: terpenes, stilbenes, flavonoids, flavan-3-ols, isoprenyl diphosphate synthase, defense priming

INTRODUCTION

Conifers throughout the world have come under increasing attack from bark beetles in recent years as a result of global warming. Rising temperatures have promoted beetle reproduction while increasing stress on the trees themselves due to heat and drought (Biedermann et al., 2019). In central Europe, the dominant conifer Norway spruce (*Picea abies*) is attacked by the Eurasian spruce bark beetle, *Ips typographus*, which vectors various fungi that contribute to its successful attack. The ascomycete *Ceratocystis polonica*, recently renamed as *Endoconidiophora polonica* (De Beer et al., 2014), is considered the most virulent (Wermelinger, 2004). The partnership between beetles and their fungal symbionts is lethal for many hectares of spruce forest each year.

Norway spruce possesses various modes of defense against beetle-fungal attack. Besides the physical barrier of the bark consisting of lignified and suberized cells, chemical barriers in the bark include the presence of terpene-based oleoresin accumulating in axial and radial resin ducts and phenolic compounds stored in phloem parenchyma cells (**Figure 1**). Although terpenes and phenolics are present constitutively in Norway spruce, their accumulation increases upon bark beetle attack or artificial inoculation with *E. polonica* (Viiri et al., 2001; Lieutier et al., 2003; Franceschi et al., 2005; Zhao et al., 2010; Hammerbacher et al., 2011, 2014; Wadke et al., 2016). In addition, application of the defense hormone methyl jasmonate results in significant accumulation of terpene and phenolic defenses (Martin et al., 2002; Erbilgin et al., 2006; Zeneli et al., 2006; Schmidt et al., 2011; Schiebe et al., 2012).

Spruce oleoresin terpenes include monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpenes (C_{20}). These compounds are synthesized from the C_5 building blocks, isopentenyl diphosphate (IDP), and dimethylallyl diphosphate (DMADP), produced either by the cytosolic mevalonate pathway (MVA) or the plastidial methylerythritol phosphate (MEP) pathway. The C_5 building blocks are assembled by isoprenyl diphosphate synthases (IDS), members of an enzyme class known as prenyltransferases, which catalyze the sequential head-to-tail condensation of 1–3 IDP molecules with DMADP producing geranyl diphosphate (GDP, C_{10}), farnesyl diphosphate (FDP, C_{15}), and geranylgeranyl diphosphate (GGDP, C_{20}) (**Figure 1**; Wang and Ohnuma, 2000; Liang et al., 2002; Schmidt and Gershenzon, 2007, 2008; Schmidt et al., 2010, 2011; Nagel et al., 2019). GDP, FDP, and GGDP are substrates for terpene synthases (TPS), which produce an array of terpene skeletons that can be further modified, for example, *via* the formation of copalyl diphosphate (CDP) and by cytochrome P450s generating the diverse chemical composition of spruce oleoresin (Keeling and Bohlmann, 2006; Bohlmann and Keeling, 2008; Chen et al., 2011; Hall et al., 2013). In order to gain a deeper understanding how terpene biosynthesis is regulated under different conditions, such as after bark beetle attack, it is useful to study the expression of genes encoding IDS and TPS, measure enzyme activity and determine the *in vivo* concentrations of the key intermediates GDP, FDP, and GGDP (Nagel et al., 2012, 2014).

In addition to terpenoids, phenolics also constitute an important group of defensive metabolites in spruce. The major phenolic compounds accumulating in Norway spruce are stilbenes and flavonoids (**Figure 1**). These compounds are both biosynthesized by condensation of *para*-coumaroyl coenzyme A (CoA), formed from phenylalanine *via* the phenylpropanoid pathway, with malonyl CoA units formed from the polyketide pathway. These condensation reactions are catalyzed by stilbene synthase (STS) and chalcone synthase (CHS) enzymes to form a related tetraketide. However, STS enzymes cyclize the tetraketide to the tetrahydroxystilbene, resveratrol, which is further modified *via* glycosylation, oxidation, and methylation to the compounds astringin and isorhapontin, the predominant stilbenes in the bark of *P. abies* (Hammerbacher et al., 2011). Meanwhile, cyclization of the tetraketide by CHS yields the flavanone naringenin (Cheynier et al., 2013) which is further modified *via* two oxidation reactions to form the dihydroflavonol taxifolin that accumulates in spruce bark and needles. Further modification of taxifolin *via* two consecutive reductions yields the flavan-3-ol catechin, the precursor for proanthocyanidins (condensed tannins) (Hammerbacher et al., 2014). However, our knowledge of how phenolic biosynthesis is regulated in spruce after insect or pathogen attack is still limited.

As only mature spruce trees with a minimum bark thickness are attacked by bark beetles, large trees growing under natural conditions represent the most relevant study system in which to investigate defensive responses to beetles and their associated fungi. However, most previous experiments on spruce defenses have been performed on young saplings after application of methyl jasmonate (MJ) to induce defense responses (Lieutier et al., 2003; Erbilgin et al., 2006; Hammerbacher et al., 2011, 2013, 2014). Among studies on mature trees (Viiri et al., 2001; Zeneli et al., 2006; Zhao et al., 2010; Schiebe et al., 2012), few have tested the direct inoculation of a bark beetle-associated fungus, and no study has comprehensively quantified oleoresin terpenes and phenolic compounds in parallel while linking compound accumulation to levels of defense hormones and biosynthetic genes, enzymes and intermediates.

Previous research on plant-insect interactions in many systems has indicated that initial herbivore or pathogen attack may lead to greater resistance to subsequent attack by inducing the accumulation of defense compounds or priming plants to respond more vigorously on subsequent attack (Aranega-Bou et al., 2014). In Norway spruce, pre-treatment with the defense hormone MJ or a sublethal dose of bark beetle-associated fungus can lead to enhanced defensive responses to the subsequent invasion of bark beetles (Erbilgin et al., 2006; Zhao et al., 2011; Mageroy et al., 2020). However, it is still unclear if trees from previously attacked stands are also more resistant to future attacks and whether this is due to priming of defenses because of greater accumulation of oleoresin terpenes and phenolics or priming of defense responses. Thus we investigated defense responses in Norway spruce trees with different histories of bark beetle attack: (a) trees from a stand that were the sole survivors of a previous attack, (b) trees under actual, current attack, and (c) trees without any previous or current attack. All trees studied showed a similar magnitude of terpene and

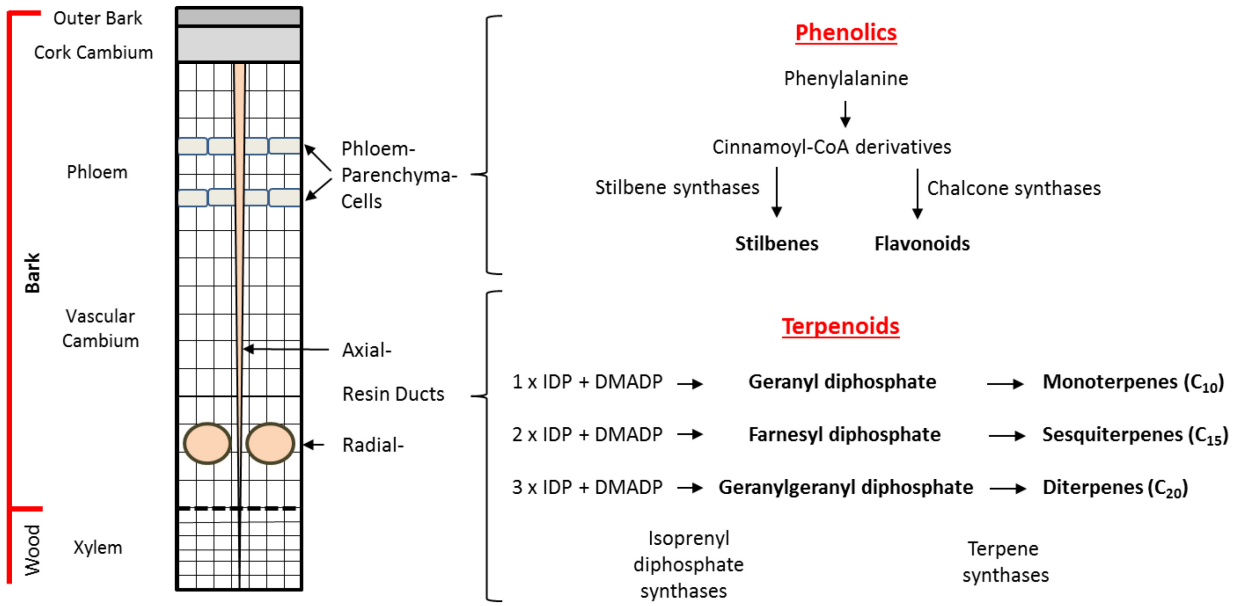


FIGURE 1 | Biosynthesis and storage of defense compounds in Norway spruce bark. Depicted on the left is a schematic cross section of the cambial region of the trunk of a mature tree. Phenolic compounds are localized in bands of phloem parenchyma cells in the phloem. Terpenoid resin compounds are localized in radial resin ducts in the xylem and in axial resin channels that traverse both the xylem and phloem. An outline of the biosynthetic pathways to terpene and phenolic defense compounds is given on the right. The major phenolic compounds (stilbenes, flavonoids) are synthesized in the phloem parenchyma cells from phenylalanine with stilbene synthases (STSs) and chalcone synthases (CHSs) being the key branchpoint enzymes. The terpenoid resin compounds are synthesized in the epithelial cells lining the resin ducts from the C_5 units isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP). The isoprenyl diphosphate synthases catalyze the formation of the intermediates geranyl diphosphate (GDP), farnesyl diphosphate (FDP), and geranylgeranyl diphosphate (GGDP). These intermediates are converted into monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpenes (C_{20}), respectively, by the action of terpene synthases (TPSs).

phenolic induction after fungal inoculation regardless of their attack history. Induction was accompanied by an up-regulation of genes, enzymes, and metabolic intermediates involved in their biosynthesis.

MATERIALS AND METHODS

Chemicals

Isopentenyl diphosphate (IDP), dimethylallyl diphosphate (DMADP), geranyl diphosphate (GDP), farnesyl diphosphate (FDP), geranylgeranyl diphosphate (GGDP), 1,9-decadiene, ammonium bicarbonate, chlorogenic acid, formic acid, tert-butyl methyl ether (TBME), methanol and acetonitrile (LC-MS grade) were purchased from Sigma-Aldrich (St. Louis, MO, United States). Trimethylsulfonium hydroxide (TMSH) was ordered from Macherey-Nagel (Düren, Germany).

Biological Material, Fungal Inoculation, and Sampling

Bark from 15 mature (approximately 80-year-old, diameter of 70 cm at the height of 1.30 m) spruce trees (*P. abies*), which were sole survivors of a massive bark beetle attack in 2008, was collected in the Bohemian Forest National Park near Plöckenstein, Austria (48.46°N–13.50°E) in June 2012. These trees were still surrounded by dead trees, were not subject to any bark beetle attacks between 2008 and 2012 (personal communication from the local forester), and had no apparent above-ground signs or symptoms of biotic and physical injury. They were designated as trees with a history of **p**revious **a**ttack (PA). At the same time, 19 trees of approximately the same age were sampled in the Kalkalpen National Park close to Bodinggraben, Austria (47.47°N–14.24°E) that were being attacked by bark beetles at the time of sampling. These trees are designated as trees under **c**urrent **a**ttack (CA). In the same stand, six trees not affected by the current outbreak were sampled. These trees had never been recorded to be attacked by bark beetles (personal communication from the local forester), and also had no apparent above-ground signs or symptoms of biotic and physical injury. They were designated as trees with **n**o **a**ttack (NA). All trees selected for the experiment were found in stands dominated by Norway spruce, typical for a mountain habitat in Central Europe at an elevation of 1300 m (4265 ft) above sea level, and did not show any obvious phenotypical differences.

At the onset of the experiment, bark samples including outer bark, cork cambium, phloem, and vascular cambium (for details see **Figure 1**) with a size of 1.5 × 10 × 0.7 cm (W × H × D) were cut from the stem with a knife at a height of approximately 1.50 m above the soil. At the same time, an 8-mm bark plug was removed on the opposite side of the sampled stem with a cork borer and inoculated with *E. polonica* culture 1993-208/115 using the method of Hammerbacher et al. (2011). Bark samples were then taken from these inoculated trees 14 d afterward. All sampled material was frozen immediately after harvest in dry ice

and stored at −80°C. The experimental approach is illustrated in **Supplementary Figure 1**.

Resin Terpene Quantification

The protocol was adapted from Lewinsohn et al. (1993). In brief, 100 mg of frozen ground plant material was extracted with 1 mL TBME containing 50 μg mL^{-1} 1,9-decadiene and 47.3 μg mL^{-1} dichlorodehydroabietic acid as internal standards under continuous shaking for 24 h. The extract was removed, washed with 0.3 mL of 0.1 M $(NH_4)_2CO_3$, pH 8.0, and dried by using a Pasteur pipette filled with 100 mg of Na_2SO_4. The Na_2SO_4 column was further washed with 0.5 mL of TBME. To 0.4 mL of extract, 50 μL of *N*-trimethylsulfonium hydroxide in MeOH was added for methylation of diterpenoid resin acids, while the rest of the extract was used for mono- and sesquiterpene analysis. Both extracts were subsequently analyzed by GC-MS and GC-FID. An Agilent 6890 series coupled to either an Agilent 5973 mass spectrometer or and FID detector were used. The column for both instruments was an Phenomenex ZB-5MSi (30 m, 0.25 mm, 0.25 μm), and the injector temperature was set at 280°C in split less mode with injections of 1 μL. Starting temperature for mono- and sesquiterpene analysis was 40°C, held for 4 min and increased by 5°C min^{-1} to 180°C, followed by a bake out at 280°C. Starting temperature for diterpene analysis was 140°C, held for 4 min and then increased by 5°C min^{-1} to 250°C followed by a bake out at 280°C. The extracted plant material was dried after removal of the liquid phase and the dry weight determined, with the fresh weight generally being 1.9–2.2 times the dry weight.

Extraction and Quantification of Phenolic Compounds From Spruce

For extraction of phenolic compounds, Norway spruce tissue was ground to a fine powder in liquid nitrogen and lyophilized at 0.34 mbar pressure using an Alpha 1–4 LD plus freeze dryer (Martin Christ GmbH, Osterode, Germany). Approximately 20 mg dried tissue was extracted with 1.2 mL analytical grade methanol containing 80 μg mL^{-1} chlorogenic acid as internal standard for 4 h at 4°C. The extract was centrifuged at 13,000 × *g* and the supernatant was recovered. Insoluble material was re-extracted with 1 mL methanol for 16 h. Supernatants were combined and evaporated to dryness under a stream of nitrogen. Dried samples were re-dissolved in 1.2 mL methanol and diluted 20 times (v/v) with water.

Chromatography was performed on an Agilent 1200 HPLC system (Agilent). Separation was achieved on a XDB C18 column (1.8 μm, 50 mm × 4.6 mm; Agilent Technologies, Waldbrunn, Germany). Formic acid (0.05%) in water and acetonitrile were employed as mobile phases A and B, respectively. The elution profile was: 0–1 min, 100% A; 1–7 min, 0–65% B in A; 7–8 min 65–100% B in A; 8–9 min 100% B, and 9–10 min 100% A. The total mobile phase flow rate was 1.1 mL min^{-1} and the column temperature was maintained at 25°C.

An API 3200 tandem mass spectrometer (AB Sciex) equipped with a turbospray ion source was operated in the negative ionization mode. The instrument parameters were optimized by infusion experiments with pure standards of catechin, gallocatechin, proanthocyanidin B1, astringin, taxifolin, and quercetin glucoside. For isorhapontin, piceid, and taxifolin glucoside, partially purified plant extracts were used for optimization. The ion spray voltage was maintained at −4500 V. The turbo gas temperature was set at 700°C. Nebulizing gas was set at 70 psi, curtain gas at 25 psi, heating gas at 60 psi, and collision gas at 10 psi. For each target analyte, multiple reaction monitoring (MRM) was used to monitor parent ion-to-product ion formation: for catechin- *m/z* 299.9/109.1 [collision energy (CE)-34 V; declustering potential (DP) −30 V]; for gallocatechin- *m/z* 304.8/179 (CE −28 V; DP −39 V); for proanthocyanidin B1- *m/z* 576.9/289.1 (CE −30 V; DP −50 V); for astringin- *m/z* 404.9/243.1 (CE −38 V; DP −50 V); for isorhapontin- *m/z* 419/257.1 (CE −18 V; DP −25 V); for piceid- *m/z* 389/227 (CE −38 V; DP −50 V); for taxifolin glucoside- *m/z* 465/285 (CE −44 V; DP −55 V); for quercetin glucoside- *m/z* 463/301 (CE −40 V; DP −55 V); for taxifolin- *m/z* 303/125 (CE −28 V; DP −40 V). Both Q1 and Q3 quadrupoles were maintained at unit resolution. Analyst 1.5 software (AB Sciex) was used for data acquisition and processing. Linearity of compound detection for quantification was verified by external calibration curves for catechin, astringin, quercetin glucoside, and taxifolin. Flavan-3-ol concentrations were determined relative to the catechin calibration curve, stilbenes relative to the astringin calibration curve, flavonol glucosides relative to the quercetin glucoside calibration curve and taxifolin relative to the taxifolin calibration curve.

Quantitative Real-Time PCR

RNA isolation and complementary DNA synthesis from a subset of trees (PA = 6, CA = 12, NA = 6) were carried out as described in Schmidt et al. (2011). Quantitative real-time PCR was done with Brilliant SYBR Green QPCR Master Mix (Stratagene) and 10 pmol forward and 10 pmol reverse primer. Primer sequences for *Pa*IDS1, *Pa*IDS4, and *Pa*IDS5 are given in Schmidt and Gershenzon (2007) and Schmidt et al. (2010). For all other genes the primer sequences are shown in **Supplementary Table 1**. PCR was performed using a Stratagene MX3000P thermocycler according to the instruction manual. Transcript abundance was normalized to the transcript abundance of ubiquitin as described in Schmidt et al. (2010).

Protein Extraction and Quantification of Isoprenyl Diphosphate Synthase Enzyme Activity

Total protein extracts from bark tissue and protein quantification were done as described in Nagel et al. (2012). Enzyme assays were carried out using 10 μg of total protein in 200 μL of 25 mM 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid (MOPSO) buffer at pH 7.2 with 10% (v/v) glycerol, 10 mM $MgCl_2$ and 50 μM IDP and DMADP each (Sigma-Aldrich, St. Louis, MO, United States) for 2 h at 30°C. Quantification was done as described under *in vivo prenyl diphosphate quantification* below but with an injection volume of 1 μL and without MRMs for the internal standards GSDP and FSDP.

In vivo Prenyl Diphosphate Quantification

A 1 g portion of plant material was extracted three times with 5 mL of methanol:water (7:3, v/v) containing 0.3 μg each of geranyl S-thiolodiphosphate (GSDP) and farnesyl S-thiolodiphosphate (FSDP) each (Echelon Biosciences Incorporated, Salt Lake City, UT, United States). Extracts were combined and purified using 150 mg (6 mL) Chromabond HR-XA columns (Macherey-Nagel) that had been conditioned with 5 mL methanol and 5 mL water. After application of the extract, the column was washed with 4 mL water followed by 5 mL methanol. Prenyl diphosphates were eluted with 3 mL 1 M ammonium formate in methanol, evaporated under a stream of nitrogen to dryness and dissolved in 300 μL water:methanol (1:1). Quantification was done using an Agilent 1260 HPLC system (Agilent Technologies, Waldbrunn, Germany) coupled to an API 5000 triple-quadrupole mass spectrometer (AB Sciex Instruments). For separation, a Zorbax Extended C-18 column (1.8 μm, 50 mm × 4.6 mm; Agilent Technologies, Waldbrunn, Germany) was used. The mobile phase consisted of 5 mM ammonium bicarbonate in water as solvent A and acetonitrile as solvent B, with the flow rate set at 1.2 mL min^{-1} and the column temperature kept at 20°C. Separation was achieved by using a gradient starting at 5% B, increasing to 70% B in 5 min and 100% B in 1 min (0.5-min hold), followed by a re-equilbration to 0% B for 1.5 min (1-min hold) before the next injection. The injection volume for samples and standards was 2 μl; autosampler temperature was 4°C. The mass spectrometer was used in the negative electrospray ionization (EI) mode. Optimal settings were determined using standards, except for CDP for which no commercial standard is available; instead the settings for GGDP were used. Levels of ion source gases 1 and 2 were set at 60 and 70 psi, respectively, with a temperature of 700°C. Curtain gas was set at 30 psi and collision gas was set at 7 psi, with all gases being nitrogen. Ion spray voltage was maintained at −4200 V. For each target analyte, MRM was used to monitor parent ion-to-product ion formation (DP, declustering potential; EP, entrance potential; CE, collision energy; CXP, collision cell exit potential): for GDP- *m/z* 312.9/79 (DP −40, EP −6, CE −38, CXP 0), for FDP- *m/z* 380.9/79 (DP −40, EP −3, CE −42, CXP 0), for GGDP and CDP- *m/z* 449/79 (DP −45, EP −10, CE −50, CXP 0), for GSDP- *m/z* 329/79 (DP −45, EP −6, CE −20, CXP −13) and 329/159 (DP −45, EP −6, CE −20, CXP −19) and for FSDP- *m/z* 379/79 (DP −75, EP −6, CE −68, CXP −15) and 379/159 (DP −75, EP −6, CE −24, CXP −17). Data analysis was performed using Analyst Software 1.6 Build 3773 (AB Sciex).

Identification of Copalyl Diphosphate as a Product When Total Protein Extract Was Assayed With Geranylgeranyl Diphosphate

Protein isolation from spruce bark was done as described in Nagel et al. (2012). An enzyme assay was performed in 4 mL of 25 mM MOPSO buffer, pH 7.2, with 10% (v/v) glycerol, 10 mM $MgCl_2$ with 60 μg GGDP and incubated for 10 h at 30°C. Then 1 mL of 5 N HCl was added to cleave the diphosphate, and the assay was extracted three times with 2 mL pentane each. For GC-MS analysis the combined phases were evaporated to 200 μL.

An Agilent 6890 series GC with an Agilent 5973 mass spectrometer and a Zebron ZB-5MS column (30 m × 0.25 m × 0.25 μm) (Phenomenex, Aschaffenburg, Germany) were used for detection. Two μL of sample were injected in splitless mode with an injector temperature of 250°C and a flow rate of 1 mL min^{-1} helium. Oven temperature was initially 40°C, then raised by 7°C min^{-1} to 280°C and held there for 5 min. ChemStation G1701 was used for data analysis.

To identify CDP and determine its stereochemistry, standards were obtained from *Oryza sativa* copalyl diphosphate synthase 4 (*Os*4; syn-CDP) (Xu et al., 2004), *Arabidopsis thaliana* copalyl diphosphate synthase (*At*CPS; ent-CDP) (Prisic et al., 2004) and *Abies grandis* abietadiene synthase D621A mutant (*Ag*AS:D621A; normal-CDP) (Prisic et al., 2004). These genes were expressed as described in Wu et al. (2012). Bacterial pellets were resuspended in 3 mL of buffer containing 20 mM MOPSO, pH 7.0, 10% (v/v) glycerol, and 10 mM $MgCl_2$ and sonicated using a Sonopuls HD 2070 (Bandelin, Berlin, Germany) for 4 min, cycle 2, power 60%. The lysate was centrifuged at 14,000 × *g* for 30 min. A 200 μL portion of the supernatant was used for enzyme assays carried out in 1 mL of 25 mM MOPSO buffer, pH 7.2, with 10% (v/v) glycerol, 10 mM $MgCl_2$ and 20 μg GGDP as substrate, and incubated for 10 h at 30°C.

The enzyme assays were either injected directly into an LC-MS/MS as described under Protein extraction and quantification of isoprenyl diphosphate synthase enzyme activity or 200 μL 5 N HCl were added to cleave the diphosphate. The assay was extracted three times with 1 mL pentane each. The pentane phase was combined and evaporated to 100 μL for GC-MS analysis as described above.

An all *cis*-GGDP standard was obtained from *Solanum lycopersicum* *cis*-prenyltransferase 2 *Sl*CPT2 expressed and purified as described in Schilmiller et al. (2009) and Akhtar et al. (2013).

Phytohormone Analysis

A subset of trees (PA = 6, CA = 12, NA = 6) was used to analyze phytohormone levels in order to obtain a better understanding of the signaling pathways involved in the defense response to *E. polonica* inoculation. The phytohormone analysis was based on the procedure described by Schmidt et al. (2011). Approximately 0.15 g of ground bark was homogenized in 1 mL methanol spiked with 40 ng of [$^2H_2$]JA, 40 ng [$^2H_4$]SA, 40 ng [$^2H_6$]ABA, and 8 ng of JA-[$^{13}C_6$]Ile by shaking for 60 min. Homogenates were centrifuged at 20,000 × *g* for 20 min at 4°C, the methanol phase collected, and the homogenate re-extracted with 1.0 mL methanol. The organic phases were combined and the samples evaporated to dryness in a vacuum concentrator at 30°C. The dry residue was reconstituted in 0.5 mL of 70% (v/v) methanol/water and analyzed by LC-MS/MS. Chromatography was performed on an Agilent 1200 HPLC system (Agilent). Separation was achieved on a XDB C18 column (1.8 μm, 50 mm × 4.6 mm; Agilent Technologies, Waldbrunn, Germany). The mobile phase, comprised of solvent A (0.05% formic acid) and solvent B

(acetonitrile), was used in a gradient of 0–0.5 min, 5% B; 0.5–9.5 min, 0–58% B; 9.5–9.52 min, 58–100% B; 9.52–11 min, 100% B; 11–11.1 min, 5% B, and 11.1–14 min, 5% B with a flow rate of 1.1 mL min^{-1}. The column temperature was maintained at 25°C. An injection volume of 2 μL was used for all samples. An API 5000 tandem mass spectrometer (AB Sciex) equipped with a turbospray ion source was operated in negative ionization mode. The ion spray voltage was maintained at −4500 V. The turbo gas temperature was set at 700°C. Nebulizing gas was set at 60 psi, curtain gas at 25 psi, heating gas at 60 psi, and collision gas at 7 psi. For each analyte, MRM was used to monitor parent ion-to-product ion formation (DP, declustering potential; EP, entrance potential; CE, collision energy; CXP, collision cell exit potential): for [$^2H_4$]SA- *m/z* 141/97 (DP −35, EP −8, CE −22, CXP 0); for SA- *m/z* 137/93 (DP −35, EP −8, CE −22, CXP 0); for [$^2H_2$]JA- *m/z* 213/59 (DP −35, EP −9, CE −24, CXP 0); for JA- *m/z* 209/59 (DP −35, EP −9, CE −24, CXP 0); for [$^2H_6$]ABA- *m/z* 269/159 (DP −35, EP −12, CE −22, CXP −2); for ABA- *m/z* 263/153 (DP −35, EP −12, CE −22, CXP −2); for JA-[$^{13}C_6$]Ile- *m/z* 328/136 (DP −50, EP −4, CE −30, CXP −4); for JA-Ile- *m/z* 322/130 (DP −50, EP −4, CE −30, CXP −4); for OPDA- *m/z* 291/165 (DP −45, EP −12, CE −24, CXP −2); for IAA- *m/z* 174/130 (DP −25, EP −9, CE −14, CXP −2); for hydroxy-JA-Ile- *m/z* 338/130 (DP −50, EP −4, CE −30, CXP −4); for hydroxy-JA- *m/z* 225/59 (DP −35, EP −9, CE −24, CXP 0) and for carboxy-JA-Ile- *m/z* 352/130 (DP −50, EP −4, CE −30, CXP −4). Data analysis was performed using Analyst Software 1.6 Build 3773 (AB Sciex). JA, JA-Ile, ABA, and SA were quantified according to the labeled standards, while hydroxy-JA and OPDA were quantified using [$^2H_2$]JA with response factors of 1.0 and 0.5, respectively; IAA was quantified using [$^2H_6$]ABA and a response factor of 3.4, and hydroxy-JA-Ile and carboxy-JA-Ile were quantified using JA-[$^{13}C_6$]Ile and response factors of 1.

Statistical Analysis

Isoprenyl diphosphate synthases enzyme activity, prenyl-diphospate abundances and relative expression of biosynthetic genes were analyzed using mixed effects models lme of the nlme library (Pinheiro et al., 2018) with bark beetle attack history and fungal infection (before and after fungal inoculation) as fixed effects and tree individuals as a random intercept. To account for the variance heterogeneity of the residuals, the varIdent variance structure was applied if necessary. Whether the different variance of bark beetle attack, fungal infection, or the combination of both factors should be incorporated into the model, was determined by comparing models with different variance structures with a likelihood ratio test and choosing the model with the smallest AIC. If necessary, data were transformed in order to achieve normality of residuals. The influences of the fixed effects were obtained by removing fixed effects one after another and comparing the simpler with the more complex model with a likelihood ratio test (Zuur et al., 2009). Differences between factor levels were determined by factor level reduction (Crawley, 2013). CDP abundance and CDP enzyme activity was only detectable after fungal infection. Whether the CDP abundance and enzyme activity differed between trees with different bark beetle attack histories was analyzed using a one-way anova followed by the Tukey HSD test. If necessary to achieve variance homogeneity and normality of residuals, data were log-transformed.

The influence of the bark beetle attack history on initial terpene concentrations was analyzed with one way analysis of variance (anova) or if variance homogeneity or normality of residuals (inspected graphically) were violated with the non-parametric Kruskal–Wallis rank sum test. In case of significant differences, the Dunn's test (package "dunn.test," Dinno, 2017) was applied to reveal group differences. Terpenes with concentrations below the detection limit in more than half of all samples in each bark beetle attack history group were defined as not present and were not analyzed. To analyze the changes in terpene, phenolic and phytohormone concentrations due to fungal inoculation, the respective concentrations before fungal inoculation were subtracted from the concentrations 14 days after fungal inoculation of the respective tree. Positive values indicated an increase and negative values indicated a decrease in concentration due to fungal inoculation. Values around 0 suggest no changes in concentration. Whether the changes in concentration differed between trees with different bark beetle attack histories was analyzed with one-way analyses of variance (anova) followed by a Tukey HSD test in case of significant differences. If either variance homogeneity or normality of residuals (inspected graphically) were violated, the non-parametric Kruskal–Wallis rank sum test was applied, followed by the Dunn's test (package "dunn.test," Dinno, 2017) in case of significant differences. Afterward, we analyzed whether the differences in terpene and phytohormone concentration were significantly different from 0 (indicating a significant increase or decrease of terpenes or phytohormones due to fungal infection). This was done for the concentrations of all trees, regardless of their bark beetle attack history, if there were non-significant differences among the tree groups in the previous analysis. If there were significant differences in concentration among the tree groups, this analysis was done separately for the three tree groups. A one-sample *t*-test was used in case data were normally distributed; otherwise the non-parametric Wilcoxon signed rank test with continuity correction was performed. When the three tree groups were analyzed separately, the Bonferroni correction was applied to control the family-wise error rate. For all analyzes the type I error due to multiple comparisons was controlled using the false discovery rate (FDR). Data were analyzed using R 3.4.4 (R Core Team, 2018).

RESULTS

All Classes of Resin Terpenes in Bark Increased Several-Fold Upon Experimental Fungal Infection, but Increase Was Not Influenced by Bark Beetle Attack History

Resin terpenes were quantified to assess whether *E. polonica* inoculation of *P. abies* bark elicited a defense reaction. The

total concentration of mono-, sesqui-, and diterpenes in the bark increased from approximately 0.65–3.3% of fresh weight (**Figure 2**). This increase was independent of the bark beetle attack history of the trees (**Supplementary Table 2**). However, total terpene concentrations prior to fungal infestation did differ significantly between trees with different bark beetle attack histories. Trees which survived a previous bark beetle attack (PA) had higher terpene concentrations than trees that were under current attack (CA), but neither PA nor CA trees were significantly different from non-attacked-trees (NA) (**Supplementary Table 3**).

Regarding monoterpenes, the concentrations of the four most abundant compounds, α-pinene, β-pinene, myrcene, and limonene, increased significantly by more than 4-fold 14 days after *E. polonica* infection, but the degree of increase was independent of bark beetle attack history (**Figure 2** and **Supplementary Table 2**). The concentration before inoculation was significantly higher in PA trees than in CA trees for β-pinene, myrcene and limonene (**Figure 2** and **Supplementary Table 3**). For minor terpenes, nearly all concentrations also increased significantly after *E. polonica* infection. Only α-phellandrene concentrations in CA trees and NA trees and *cis*-sabinene hydrate in NA trees did not change (**Supplementary Figure 2** and **Supplementary Table 2**). Related to bark beetle history, the increase of concentrations between trees with different histories was only significantly different for the minor monoterpene *cis*-sabinene hydrate. Its increase was higher in PA trees than in CA trees (**Supplementary Figure 2** and **Supplementary Table 2**). Similarly, the concentration of nearly all minor monoterpenes before inoculation was similar among trees with different histories, except that the concentrations of terpinolene and bornyl acetate were higher in PA trees than CA trees (**Supplementary Figure 2** and **Supplementary Table 3**).

Out of the 12 sesquiterpenes, the concentrations of all but two, including the most abundant five, longifolene, (*E*)-β-caryophyllene, germacrene D, δ-cadinene, and germacrene D-4-ol, increased significantly after infection between 3- and 7-fold for samples from all attack histories (**Figure 2** and **Supplementary Table 2**). The only compound showing a significantly different response in trees with different attack histories was *trans*-α-bergamotene, which did not increase at all in PA trees but slightly in NA trees after induction with *E. polonica*. The concentrations of sesquiterpenes before fungal inoculation were almost similar among trees with different attack histories, except that *trans*-α-bergamotene was higher in PA than in CA trees, while the opposite was true for sesquisabinene (**Figure 2** and **Supplementary Table 3**).

The diterpene levels of all trees were significantly higher after *E. polonica* infection, increasing between 4— and 9-fold, but these increases were independent of the attack history of the trees (**Figure 2** and **Supplementary Table 2**). However, the concentrations of diterpenes before fungal inoculation were similar among trees with different attack histories, except that abienol and neoabietic acid were significantly higher in PA trees than in CA trees (**Figure 2** and **Supplementary Table 3**).

Total Phenolic Levels Increased Upon Fungal Infection, but There Were Differences Among Phenolic Groups and Previously Attacked Trees Had Higher Levels of Total Phenolics

Spruce contains several groups of phenolic compounds, including flavan-3-ols, other flavonoids and stilbenes. In response to fungal inoculation, the total amount of phenolics plus the amount of flavan-3-ols increased significantly, but there was no influence of bark beetle attack history on the level of increase (**Figure 3** and **Supplementary Table 4**). Prior to fungal inoculation, PA trees had higher levels of total phenolics than CA or NA trees. The same pattern was found for the total stilbene level (**Figure 3** and **Supplementary Table 5**). Total flavan-3-ols and the amounts of gallocatechin, but not the amounts of the flavan-3-ol dimers, were higher in PA than NA trees, but neither PA nor NA trees were significantly different from CA trees.

Transcripts of Most Terpene and Phenolic Biosynthetic Genes Were Elevated in Response to Fungal Infection

To determine if terpene biosynthesis in *P. abies* bark was activated by *E. polonica* inoculation and how this was affected by bark beetle attack history, genes encoding three isoprenyl diphosphate synthases (*IDS*) and five terpene synthases (*TPS*) were selected for transcriptional studies. These genes encode proteins that catalyze the initial dedicated steps of resin formation (Martin et al., 2004; Schmidt and Gershenzon, 2007, 2008; Schmidt et al., 2010, 2011).

Transcript levels of all *P. abies IDS* measured were significantly altered by *E. polonica* infection in the investigated bark tissue (**Figure 4** and **Supplementary Table 6**). Expression of *IDS1*, a bi-functional geranyl diphosphate and geranylgeranyl diphosphate synthase, *IDS4*, a farnesyl diphosphate synthase, and *IDS5*, a geranyl diphosphate synthase, increased significantly upon infection except for *IDS1* and *IDS4* in PA trees. When comparing the initial transcript level of *IDS* genes prior to fungal inoculation among trees with different attack histories, *IDS1* and *IDS5*, the genes involved in the biosynthesis of the most abundant resin constituents, the monoterpenes and diterpenes, had significantly greater transcript abundance in PA trees than in CA or NA trees. On the other hand, *IDS4*, which is involved in sesquiterpene biosynthesis, had significantly greater transcript levels in CA trees followed by PA trees and NA trees.

Expression of monoterpene and diterpene synthase genes was also significantly induced by fungal infection (**Figure 4** and **Supplementary Table 6**). For the monoterpene synthase gene α/β-pinene synthase, transcript levels increased significantly by amounts ranging from 2- to 10-fold, depending on attack history, and δ-3-carene synthase transcripts increased by amounts ranging from 600- to over 900-fold, depending on attack history. Genes encoding the two diterpene synthase genes, isopimara-7,15-diene synthase and levopimaradiene/abietadiene synthase, were quantified together because of the high nucleotide sequence identity (∼ 95%) in the open reading frame. Transcripts increased

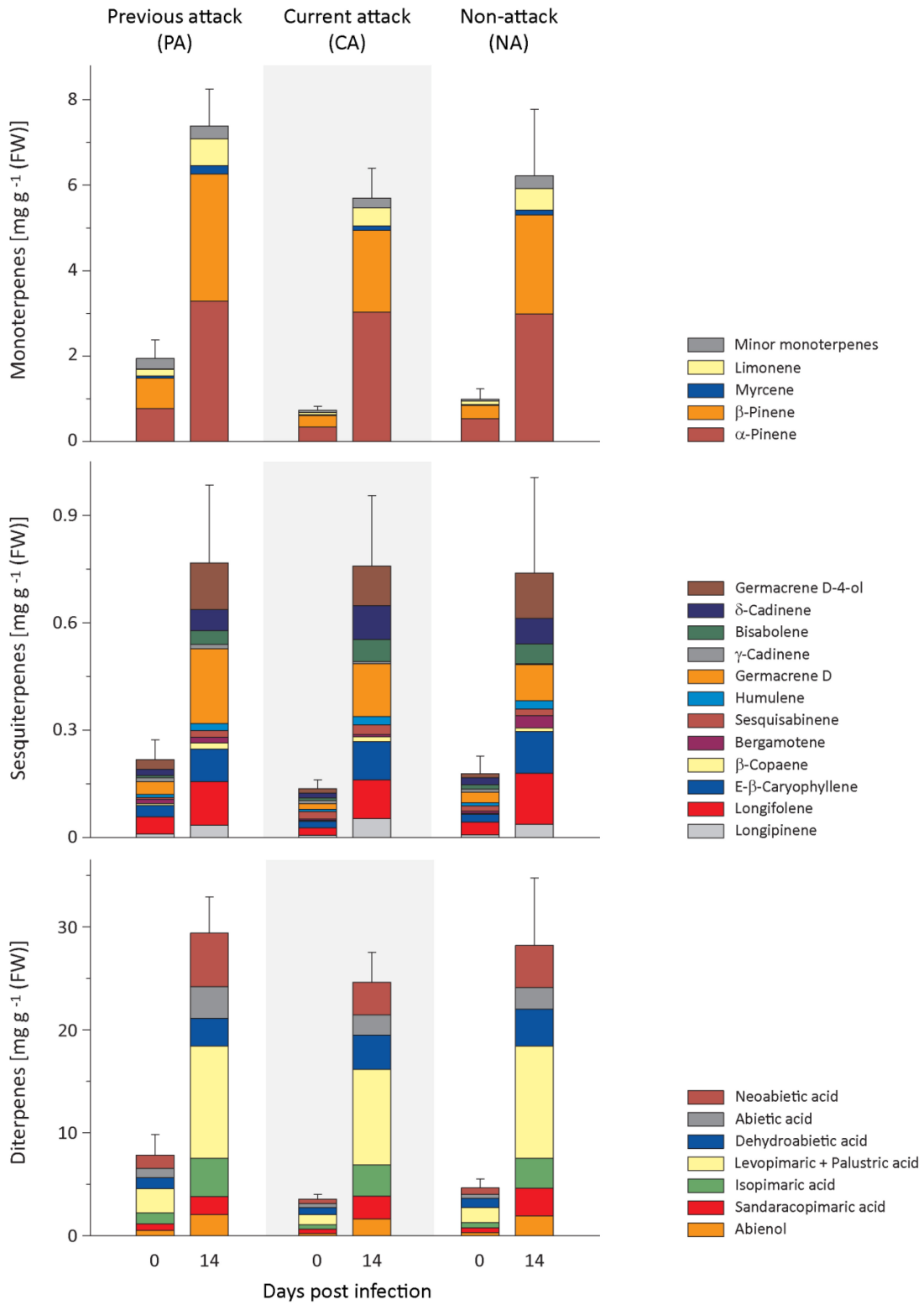


FIGURE 2 | Induction of resin terpenes in spruce bark after *Endoconidiophora polonica* infection. The concentrations of the major resin monoterpenes **(upper panel)**, sesquiterpenes **(middle panel)**, and diterpenes **(lower panel)** were quantified by GC-FID analysis of bark tissue harvested before and 14 days after infection with *E. polonica*. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Data are means ± SE of measurements from 6 to 19 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Tables 2, 3**.

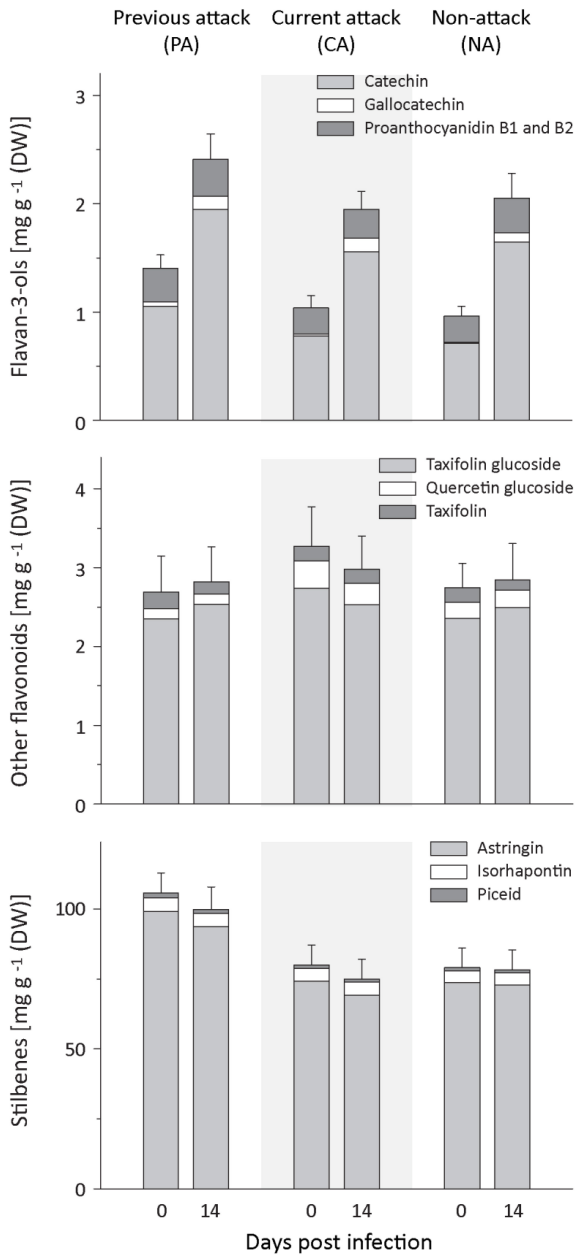


FIGURE 3 | Induction of phenolic compounds in spruce bark after *Endoconidiophora polonica* infection. The concentration of flavan-3-ols **(upper panel)**, other flavonoids **(middle panel)**, and stilbenes **(lower panel)** were quantified by LC-MS/MS analysis from bark tissue harvested before and 14 days after infection with *E. polonica*. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Data are means ± SD of measurements from 6 to 19 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Tables 4, 5**.

significantly by amounts ranging from 2 to 5-fold, depending on attack history. Among the sesquiterpene synthase genes, expression of α-bisabolene synthase decreased significantly after *E. polonica* inoculation by about 2-fold in PA and CA trees, but more than 4-fold in NA trees. Expression of longifolene synthase increased in PA and CA trees by less than 2-fold, but decreased for NA trees by nearly 50% (**Figure 4** and **Supplementary Table 6**).

For phenolic biosynthesis, we investigated genes encoding chalcone synthase, the first step in flavonoid formation, and genes encoding stilbene synthase, the first step in the biosynthesis of this type of phenolic compound. Expression of the chalcone synthases *CHS2*, *CHS6*, and *CHS8*, on the other hand, and the stilbene synthases *STS1* and *STS2*, on the other hand, were quantified together because of the high identity of the gene sequences at the nucleotide level. Expression of the three CHS genes increased significantly after inoculation with *E. polonica* by 2. 9−, 3. 8−, and 5.6-fold for PA, CA, and NA trees, respectively. In contrast, expression of the two *STS* genes was more variable; it was unaltered by fungal inoculation in PA trees, but decreased significantly by 1.5-fold in CA trees and increased significantly by 3.5-fold in NA trees (**Figure 5** and **Supplementary Table 6**).

Isoprenyl Diphosphate Synthase Activities and Levels of Isoprenyl Diphosphate Intermediates Increased Upon Fungal Infection

To understand the regulation of terpenoid biosynthesis after fungal infection in Norway spruce bark, we followed up on the activation of *IDS* genes by looking for changes in IDS enzyme activity by assaying crude protein extracts *in vitro*. IDS activities showed significant increases after *E. polonica* infection with a 1. 8−, 1. 6−, and 2.2-fold increase in the rate of formation of GDP in PA, CA, and NA trees, respectively, a 2. 9−, 2. 5−, and 3.2-fold increase in FDP formation in these tree classes, respectively, and a 3. 3−, 8. 5−, and 22.6-fold increase of GGDP formation (**Figure 6** and **Supplementary Table 7**). We also measured changes in the *in vivo* abundance of the prenyl diphosphate intermediates themselves by analysis of methanol-water extracts by LC-MS/MS. The *in vivo* levels of GDP and FDP increased significantly in response to infection with a 6−, 10. 9−, and 15.8-fold increase in GDP and 11. 5−, 15. 1−, and 15.8-fold increases in FDP in PA, CA, and NA trees, respectively, with a significant role of attack history. Increase in CA and NA trees was higher than in PA trees (**Figure 6** and **Supplementary Table 8**). However, the *in vivo* level of GGDP did not change significantly after fungal infection.

In the isoprenyl diphosphate assays and analyses of plant extracts, we also identified another prenyl diphosphate in addition to GDP, FDP, and GGDP. This metabolite had the same mass and precursor-to-product ion formation as GGDP, but an earlier elution time (**Figure 7**). It was eventually identified as CDP, an intermediate in the formation of gibberellins and other diterpenes by comparison of retention time and mass spectrum with those of an authentic standard (**Figure 7**) as described previously (Nagel and Peters, 2017). Since separation of the stereoisomers of CDP *via* HPLC was not possible, this compound was analyzed by GC-MS after dephosphorylation and determined to be either the *normal*- or *ent*-isomer, which coelute (Wu et al., 2012) rather than the *syn*-isomer, which elutes separately (**Supplementary Figure 3**). CDP and

CDP-synthesizing activity were only detected after *E. polonica* inoculation (**Figure 6**) and were significantly higher in CA trees than in PA and NA trees. CDP activity was lower in these PA than in CA trees, but both were similar to NA trees (**Supplementary Tables 7, 8**). The compound was presumably formed from GGDP in the total protein extract *via* a class II terpene synthase, and the changes observed indicate an elevated rate of diterpene biosynthesis.

Jasmonates Increased Upon Fungal Infection, but Salicylic Acid, Indole-3-Acetic Acid, and Abscisic Acid Did Not

The levels of several defense-related phytohormones in *P. abies* bark tissue were quantified to gain a better understanding of the signaling pathways involved in the defense response to *E. polonica* infection. The levels of abscisic acid (ABA) and indole-3-acetic acid (IAA) decreased significantly after inoculation with *E. polonica* (**Figure 8** and **Supplementary Table 9**). Salicylic acid (SA) was not detected before or after inoculation of spruce bark. On the other hand, the concentration of jasmonic acid and the active (+)-7 jasmonic acid isoleucine conjugate increased significantly 14 days after induction after *E. polonica* inoculation (**Figure 8** and **Supplementary Table 9**). For the jasmonic acid pathway, five additional metabolites were quantified (**Supplementary Figure 4**). The concentration of the jasmonic acid precursor 12-oxophytodienoic acid was significantly reduced after *E. polonica* inoculation but concentrations of the jasmonic acid degradation product, 12-hydroxyjasmonic acid, were not significantly altered after fungal inoculation (**Supplementary Figure 5**). However, the (−)-7 jasmonic acid isoleucine conjugate, considered to be a catabolite of the active (+)-7-conjugate, increased significantly after inoculation, while two degradation products of the active (+)-7-conjugate, the 12-hydroxyjasmonic acid isoleucine and the 12-carboxyjasmonic acid isoleucine conjugates, were only detectable after *E. polonica* inoculation (**Supplementary Figure 4** and **Supplementary Table 9**). Due to the fact that not all phytohormones, their precursors and metabolites were detectable before fungal inoculation, differences between trees of different attack history prior to infection were not statistically evaluated.

DISCUSSION

Bark Beetle Attack History Did Not Affect the Accumulation of Chemical Defenses After Fungal Infection

Inoculation of mature Norway spruce trees with the bark beetle-associated ascomycete *E. polonica* induced the jasmonate signaling pathway, leading to the increased accumulation of terpenes and phenolic compounds. The induction of chemical defenses in this conifer has been previously reported upon bark beetle attack and after wounding, methyl jasmonate treatment or inoculation with bark beetle fungal associates (Franceschi et al., 2002; Martin et al., 2002; Erbilgin et al., 2006; Heijari et al., 2008; Schmidt et al., 2011).

The extent of induction in this study was not dependent on bark beetle attack history. Trees that had been previously attacked did not show a greater degree of induction from those under current attack or those that had never been attacked. Thus, there was no priming effect of prior attack on the accumulation of terpene or phenolic defenses after the fungal inoculation we performed. Prior studies on Norway spruce have detected some evidence for defense priming, although the interval between the initial attack and subsequent challenge was much less than the 4 years in this study. For example, mechanical wounding inflicted one month after methyl jasmonate treatment gave an over 20-fold induction in terpenoids (Zhao et al., 2011). Curiously, in another study bark beetle attack 35 days after methyl jasmonate treatment did not lead to any significant increase in terpene content, yet trees were much more resistant to the beetles (Mageroy et al., 2020). Additional chemical and morphological defenses need to be screened to determine the basis of this resistance.

Trees From Previously Attacked Stands Had Higher Levels of Some Terpenes and Phenolics

Independent of priming, trees that had been previously attacked by bark beetles may have had greater defense levels upon our fungal inoculation trial due to prior induction. Once formed, resin ducts persist in conifers for many years, while the phenolic-rich cells of the phloem remain intact for several years after formation and defend the tree (Nagy et al., 2000). In the present study trees with different bark beetle attack histories did not show significant differences in chemical defenses in most cases prior to our fungal inoculation. However, for the major monoterpenes, diterpenes and some phenolic compounds, previously attacked trees had higher levels than at least one of the other two categories (currently attacked or non-attacked trees) or both of them. These higher levels of defenses could have been induced by the original bark beetle attack and may help explain the survival of these trees. On the other hand, "previously attacked" trees may simply be survivors of the original attack that had never themselves suffered beetle damage. Constitutively high defense levels may have ensured their survival compared to less defended neighbors. We did not note any observable signs of prior beetle attack at sampling, and no anatomical studies were conducted.

Induction of Terpene Biosynthesis and Accumulation Are Common Spruce Responses to Actual or Simulated Enemy Attack

After inoculation with the blue stain fungus *E. polonica*, the total terpene concentration in the bark of mature Norway spruce (*P. abies*) increased 3- to 9-fold with increases being generally larger for diterpenes (C_{20}) and sesquiterpenes (C_{15}) than monoterpenes (C_{10}). Elevated terpene accumulation after *E. polonica* infection or methyl jasmonate (MJ) treatment (often employed as a surrogate for bark beetle attack) has been recorded

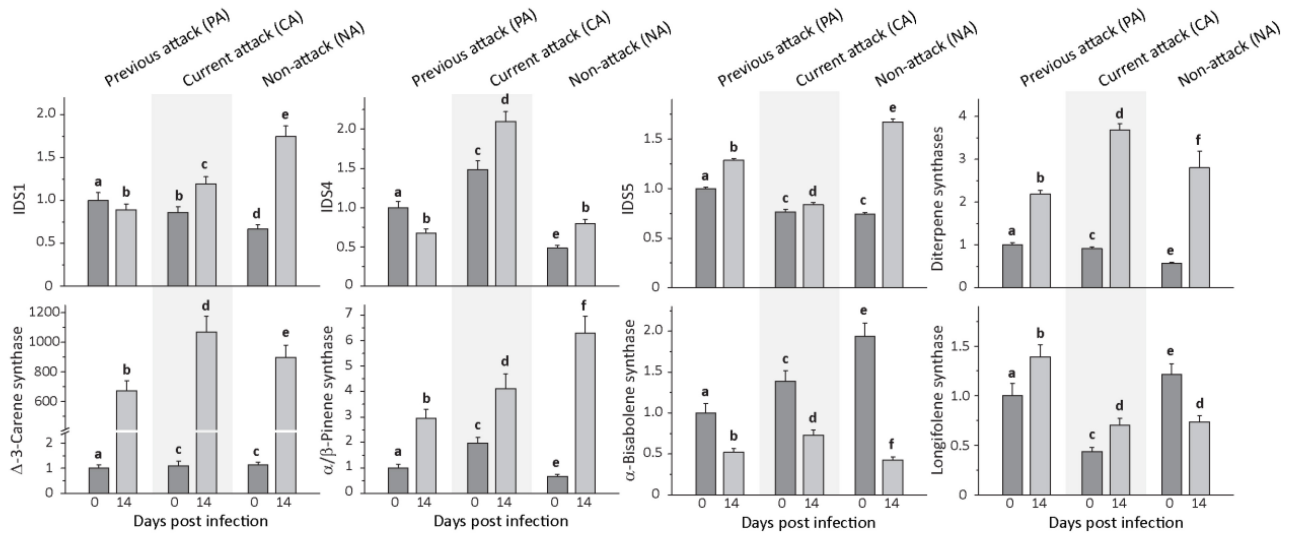


FIGURE 4 | Relative expression of terpene biosynthetic genes before and after *Endoconidiophora polonica* infection. The expression of isoprenyl diphosphate synthase 1 (IDS1), IDS 4, IDS5, and various terpene synthases (diterpene synthases, δ-3-carene synthase, α/β-pinene synthase, α-bisabolene synthase, and longifolene synthase) were measured by quantitative real-time PCR analysis from bark tissue harvested before and 14 days after infection with *E. polonica*. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Different letters indicate significant differences. Data are means ± SD of measurements from 6 to 12 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Table 6**.

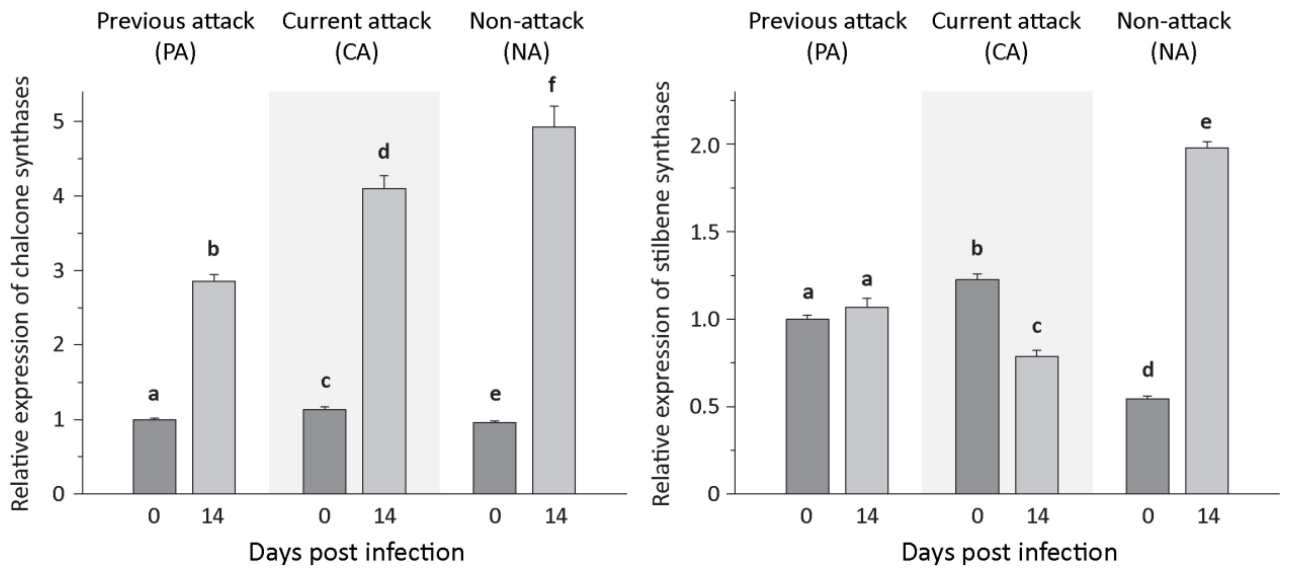


FIGURE 5 | Relative expression of phenolic biosynthetic genes before and after *Endoconidiophora polonica* infection. The relative expression of chalcone synthases **(left panel)** and stilbene synthases **(right panel)** were measured by quantitative real-time PCR analysis from bark tissue harvested before and 14 days after infection with *E. polonica*. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Different letters indicate significant differences. Data are means ± SD of measurements from 6 to 12 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Table 6**.

in several previous studies (Viiri et al., 2001; Martin et al., 2002; Erbilgin et al., 2006; Zeneli et al., 2006; Zhao et al., 2010; Schmidt et al., 2011; Schiebe et al., 2012). However, the magnitudes of increase were variable with mature trees tending to exhibit greater accumulation than saplings. The induction of terpenoid in conifer stems is ascribed to the formation of additional resin ducts, known as traumatic ducts, in the developing xylem (Franceschi et al., 2002; Martin et al., 2002; Schmidt et al., 2010).

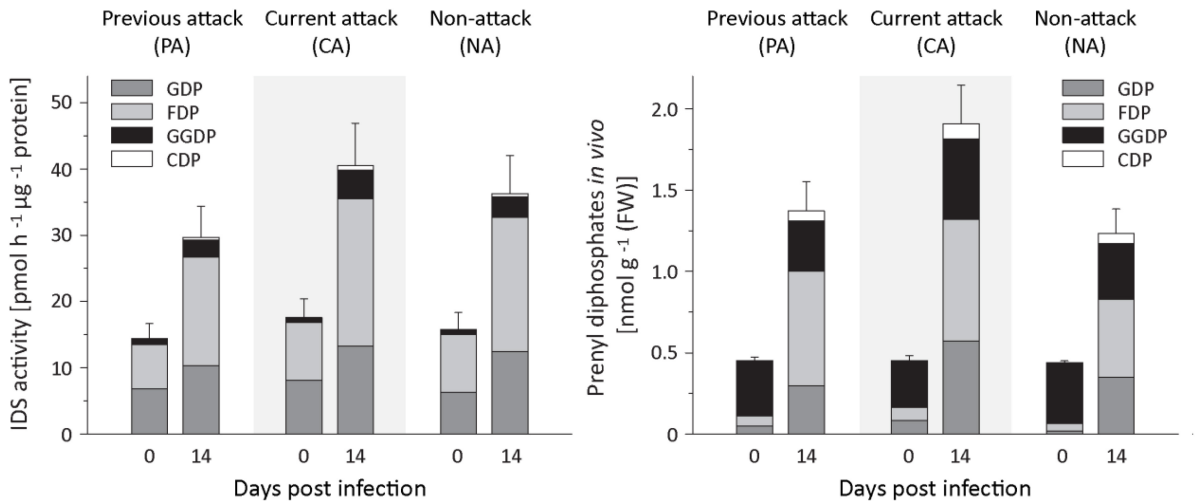


FIGURE 6 | Quantification of IDS enzyme activity and prenyl diphosphate levels in spruce bark before and after *Endoconidiophora polonica* infection. The IDS enzyme activity of total protein extracts was assayed *in vitro* with IDP and DMADP as substrates (left panel). The abundance of the prenyl diphosphates GDP, FDP, GGDP, and CDP in spruce bark tissue *in vivo* was measured by LC-MS/MS. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Data are means ± SD of measurements from 6 to 12 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Tables 7, 8**.

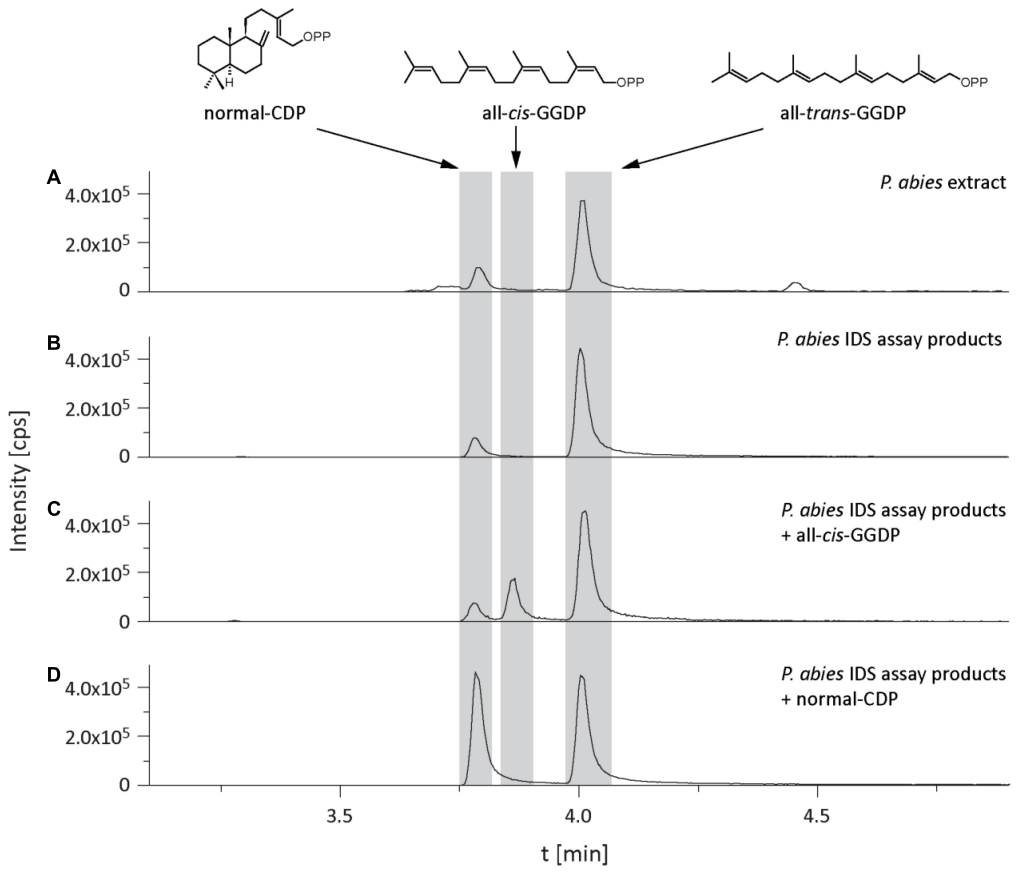


FIGURE 7 | Identification of copalyl diphosphate (CDP) in *Picea abies* bark extracts and in IDS assays of total protein extracts with IDP and DMADP as substrates. LC-MS/MS chromatograms [m/z: 449/79] of extracts of **(A)** bark tissue, **(B)** IDS assay, **(C)** IDS assay plus all-*cis*-GGDP, **(D)** IDS assay plus *normal*-CDP.

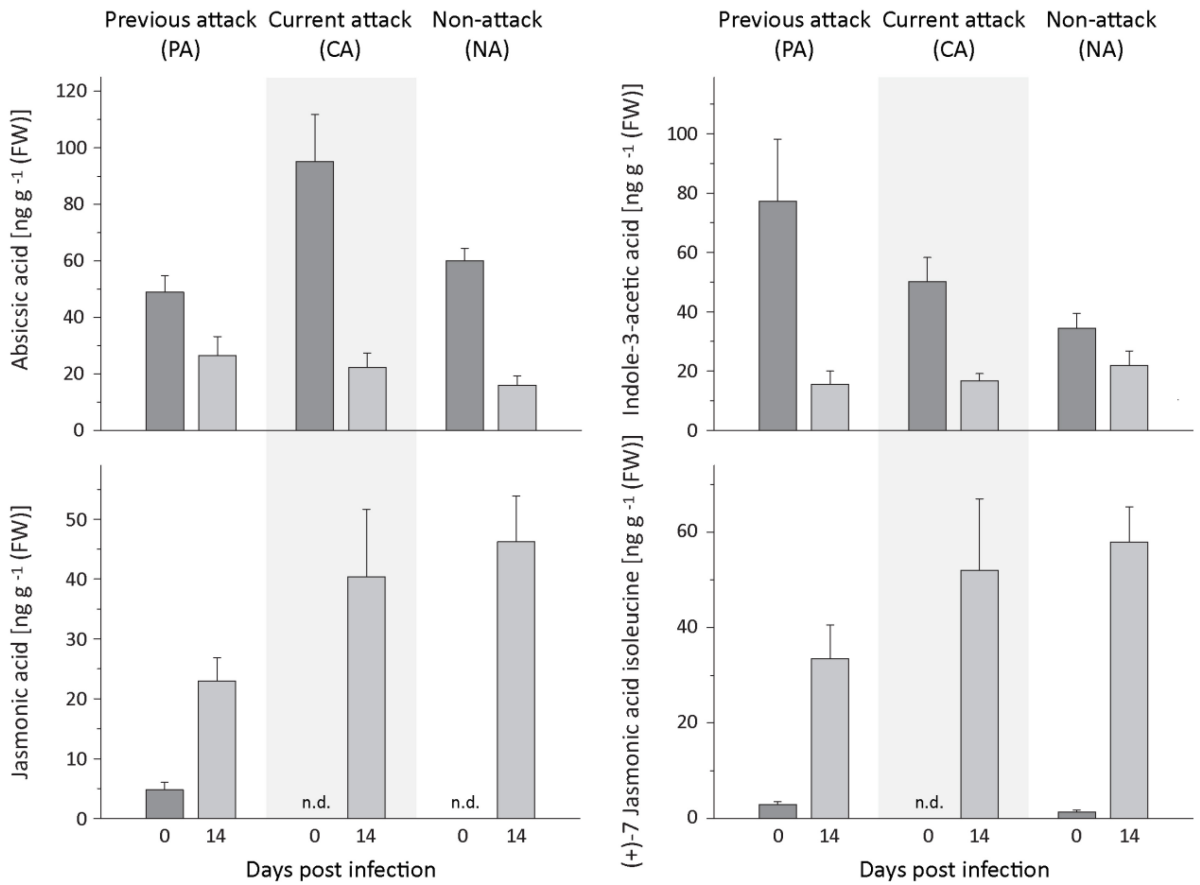


FIGURE 8 | Quantification of phytohormones in spruce bark before and after *Endoconidiophora polonica* infection. The concentration of abscisic acid **(upper left panel)**, indole-3-acetic acid **(upper right panel)**, jasmonic acid **(lower left panel)**, and the (+)-7 jasmonic acid-isoleucine conjugate **(lower right panel)** were quantified by LC-MS/MS analysis from bark tissue harvested before and 14 days after infection with *E. polonica*. Trees investigated: (1) were from a stand with a history of previous bark beetle attack, but no current attack (PA), (2) were under current attack, but had no previous history of attack (CA), or (3) had no previous nor current attack (NA). Data are means ± SD of measurements from 6 to 12 biological replicates; exact numbers are given in the Section "Materials and Methods." Statistical values for individual compounds are given in **Supplementary Table 9**.

To determine if terpene accumulation after *E. polonica* fungal infection results from *de novo* biosynthesis as previously demonstrated after MJ treatment (Schmidt et al., 2010, 2011), the expression of biosynthetic genes was measured. Among the isoprenyl diphosphate synthases, the transcript levels of *IDS1*, encoding the formation of a joint GDP (C_{10}) and GGDP (C_{20}) synthase, *IDS4*, encoding the formation of an FDP (C_{15}) synthase, and *IDS5*, encoding a GGDP (C_{20}) synthase, were elevated on fungal infection in trees with all attack histories, except for *IDS1* and *IDS4* in previously attacked trees. In earlier studies on young Norway spruce saplings, *IDS1* and *IDS5* expression both increased after MJ application (Schmidt and Gershenzon, 2007; Schmidt et al., 2010), being elevated by more than 6-fold 22 days after treatment (Schmidt et al., 2011), whereas the expression of *IDS4* was not increased at all. Among the terpene synthase genes measured, the transcript level of monoterpene and diterpene synthases increased even more dramatically than those of the isoprenyl diphosphate synthases, but this was not true for the sesquiterpene synthases. Monoterpenes and diterpenes typically make up the bulk of conifer resin, with sesquiterpenes present in only minor amounts.

Along with the increase in *IDS* gene expression, there was an elevation in IDS enzyme activity producing GDP, FDP and GGDP and the *in vivo* concentration of GDP and FDP, but not GGDP, significantly increased, all consistent with a rise in terpene biosynthetic rate. After fungal inoculation, there was also a significant increase in the concentration of CDP, an intermediate in the conversion of GGDP to diterpene olefins by various conifer diterpene synthases (Peters et al., 2001). This increase paralleled the increases measured in diterpene synthase gene expression and terpene accumulation. Although *in vitro* studies with a recombinant conifer abietadiene synthase showed CDP to be present at a maximum of 2% of GGDP, we found CDP *in vivo* at levels of almost 20% of the GGDP concentration. This suggests that bifunctional diterpene synthases in *P. abies* (Martin et al., 2004) are not efficient in channeling CDP from the active site of the first reaction (conversion of GGDP to normal-CDP) to the site of the second reaction (conversion of normal-CDP to

various diterpene olefins). The only diterpene synthases that are not bifunctional in spruce are *ent*-copalyl diphosphate synthase and *ent*-kaurene synthase, which are involved in gibberellin biosynthesis (Keeling et al., 2009). In contrast to *Taxus brevifolia* or *Ricinus communis* no diterpene synthases in spruce are known that use GGDP as a substrate and directly produce the diterpene product without the involvement of CDP (Zi et al., 2014).

In this study, prior bark beetle attack was not shown to prime the induction of terpene accumulation following fungal infection. Yet trees that had been previously attacked showed significantly higher expression of *IDS* genes involved in the formation of monoterpene and diterpenes than non-attacked trees or trees under current attack. Whether elevated transcript expression could be responsible for the priming of terpene formation under other conditions remains to be determined. It would be especially interesting to investigate the consequences of a genuine bark beetle attack, rather than artificial inoculation with a bark beetle-associated fungus.

Induction of Phenolic Compounds by Fungal Infection May Be Influenced by Fungal Catabolism

Phenolic compounds also accumulated upon fungal infection in this study, but the increase was not as pronounced as for terpenes. Moreover, there was no effect of bark beetle attack history on the magnitude of accumulation. The increase in total phenolics was due to an increase of ∼2-fold in flavan-3-ol content, but no significant changes were detected in the levels of other flavonoids and stilbenes upon fungal infection, in agreement with earlier studies on fungus-infected *P. abies* saplings and mature trees (Viiri et al., 2001; Lieutier et al., 2003; Hammerbacher et al., 2014). Phenolics have also been shown to increase after MJ treatment (Erbilgin et al., 2006; Zeneli et al., 2006), but the pattern of accumulation is different compared to fungal infection. MJ applied to mature trees induced significant increases in all classes of phenolics measured (Schiebe et al., 2012). The lack of a broad increase after fungal infection may be ascribed to the ability of *E. polonica* to degrade certain host tree phenolics during infection (Hammerbacher et al., 2013; Wadke et al., 2016). The 3- to 5-fold increase of chalcone synthase transcript levels after fungal infection indicates that flavonoid biosynthesis was significantly enhanced by infection. Hence the lack of a measurable increase in flavonoids other than flavan-3-ols could well have resulted from fungal catabolism.

Induction of Jasmonate but Not Salicylic Acid by *Endoconidiophora polonica* Is Consistent With Its Necrotrophic Lifestyle

The patterns of hormone signaling in response to fungal pathogens have long been known to depend on the lifestyle of the fungus. The response to necrotrophic pathogens typically involves both jasmonate and ethylene signaling, while the response to biotrophic and hemi-biotrophic pathogens is usually mediated by salicylic acid signaling (Glazebrook, 2005). In this study, infection of the bark of mature spruce trees with *E. polonica* induced the jasmonate pathway, but not the salicylic acid pathway, corresponding to the necrotrophic lifestyle of this fungus (Wadke et al., 2016). Jasmonic acid and several other jasmonic acid metabolites increased many-fold after *E. polonica* inoculation in all categories of tree samples regardless of their bark beetle attack history. However, no salicylic acid was detected either before or after inoculation in contrast to an earlier study involving the MJ treatment of 15-year-old trees, which used similar analytical methods (Schmidt et al., 2011). Both salicylic acid and jasmonic acid were present in significantly lower amounts than in woody angiosperms, such as in a poplar species attacked by insect herbivores (Clavijo McCormick et al., 2014), but the concentration of the active jasmonic acid isoleucine conjugate was comparable. Whether this difference reflects a true difference in jasmonate signaling between gymnosperms and angiosperms requires further study. In contrast to the jasmonates, the amount of abscisic acid declined significantly after fungal induction regardless of attack history. Unfortunately, our knowledge about the link between abscisic acid and defense in woody plants is still very limited. In poplar, abscisic acid was reported be induced by infection with a biotrophic fungus (Ullah et al., 2019).

CONCLUSION

Infection of Norway spruce by the bark beetle-associated *E. polonica* resulted in the induction of a complex mixture of monoterpenes, sesquiterpenes, diterpene acids and flavan-3-ols in bark formed largely by *de novo* biosynthesis. These substances are believed to represent principal conifer defenses against the fungus and its bark beetle vector, serving as toxins, repellents, or physical barriers to invasion (Franceschi et al., 2005; Hammerbacher et al., 2011, 2013). No evidence was found that previous bark beetle attack primes the extent of defense induction although the 4-year-period between initial attack and our experimental fungal inoculation may have been too long for the maintenance of the priming signal. Prior to our induction, many of the terpene and phenolic defenses measured were higher in previously attacked trees than unattacked trees or those under current attack, likely a remnant of previous induction. Further research is required to understand the role of constitutive versus induced chemical defenses in protecting conifers against bark beetle invasion.

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

AH and AS designed the experiments and did the sampling. RN and AH performed the experiments and analyzed the data.

GK performed the statistics. MP contributed to the molecular methods. JG and AS supervised the study. RN wrote the manuscript with contribution from AH, GK, MP, JG, and AS. All authors approved the submitted version.

FUNDING

This work was funded by the Zwillenberg-Tietz Foundation and the Max Planck Society.

ACKNOWLEDGMENTS

We would like to thank Johann Reisenberger from the Government of Upper Austria, Directorate Regional Planning, Economical and Rural Development, Department of Agriculture and Forestry and Hans Kammleitner from the Kalkalpen National Park for permission to enter the park and for their help during the sampling procedure. We also thank Marcel Alsleben for help in GC-MS analysis, Sandra Irmisch for her help in expressing CDP synthases, Reuben J. Peters for sharing *At*CPS1, AgAS:D621A, and *Os*CPS4 constructs, and Eran Pichersky for providing the *Sl*CPT2 construct.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.892907/full#supplementary-material

REFERENCES

Akhtar, T. A., Matsuba, Y., Schauvinhold, I., Yu, G., Lees, H. A., Klein, S. E., et al. (2013). The tomato *cis*–prenyltransferase gene family. *Plant J.* 73, 640–652. doi: 10.1111/tpj.12063

Aranega-Bou, P., de la O Leyva, M., Finiti, I., García-Agustín, P., and González-Bosch, C. (2014). Priming of plant resistance by natural compounds. Hexanoic acid as a model. *Front. Plant Sci.* 5:488. doi: 10.3389/fpls.2014.00488

Biedermann, P. H. W., Müller, J., Grégoire, J.-G., Gruppe, A., Hagge, J., Hammerbacher, A., et al. (2019). Bark beetle population dynamics in the Anthropocene: challenges and solutions. *Trends Ecol. Evol.* 34, 914–924. doi: 10.1016/j.tree.2019.06.002

Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313x.2008.03449.x

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

Cheynier, V., Comte, G., Davies, K. M., Lattanzio, V., and Martens, S. (2013). Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* 72, 1–20. doi: 10.1016/j.plaphy.2013.05.009

Clavijo McCormick, A., Irmisch, S., Reinecke, A., Böckler, G. A., Veit, D., Reichelt, M., et al. (2014). Herbivore-induced volatile emission in black poplar: regulation and role in attracting herbivore enemies. *Plant Cell Environ.* 37, 1909–1923. doi: 10.1111/pce.12287

Crawley, M. J. (2013). *The R Book*. West Sussex: John Wiley & Sons Ltd.

De Beer, Z. W., Duong, T. A., Barnes, I., Wingfield, B. D., and Wingfield, M. J. (2014). Redefining ceratocystis and allied genera. *Stud. Mycol.* 79, 187–219. doi: 10.1016/j.simyco.2014.10.001

Dinno, A. (2017). *Dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.5*. Available Online at: https://CRAN.R-project.org/package=dunn.test

Erbilgin, N., Krokene, P., Christiansen, E., Zeneli, G., and Gershenzon, J. (2006). Exogenous application of methyl jasmonate elicits defenses in Norway spruce (*Picea abies*) and reduces host colonization by the bark beetle *Ips typographus*. *Oecologia* 148, 426–436. doi: 10.1007/s00442-006-0394-3

Franceschi, V. R., Krekling, T., and Christiansen, E. (2002). Application of methyl jasmonate on *Picea abies* (Pinaceae) stems induces defense-related responses in phloem and xylem. *Am. J. Bot.* 89, 602–610. doi: 10.3732/ajb.89.4.578

Franceschi, V. R., Krokene, P., Christiansen, E., and Krekling, T. (2005). Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytol.* 167, 353–376. doi: 10.1111/j.1469-8137.2005.01436.x

Frost, C. J., Mescher, M. C., Carlson, J. E., and De Moraes, C. M. (2008). Plant defense priming against herbivores: getting ready for a different battle. *Plant Physiol.* 146, 818–824. doi: 10.1104/pp.107.113027

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 42, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923

Hall, D. E., Zerbe, P., Jancsik, S., Quesada, A. L., Dullat, H., Madilao, L. L., et al. (2013). Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in Lodgepole pine and Jack pine involves monofunctional and bifunctional diterpene synthases. *Plant Physiol.* 161, 600–616. doi: 10.1104/pp.112.208546

Hammerbacher, A., Paetz, C., Wright, L. P., Fischer, T. C., Bohlmann, J., Davis, A. J., et al. (2014). Flavan-3-ols in Norway spruce: biosynthesis, accumulation, and function in response to attack by the bark beetle-associated fungus *Ceratocystis polonica*. *Plant Physiol.* 164, 2107–2122. doi: 10.1104/pp.113.232389

Hammerbacher, A., Ralph, S. G., Bohlmann, J., Fenning, T. M., Gershenzon, J., and Schmidt, A. (2011). Biosynthesis of the major tetrahydroxystilbenes in spruce, astringin and isorhapontin, proceeds via resveratrol and is enhanced by fungal infection. *Plant Physiol.* 157, 876–890. doi: 10.1104/pp.111.181420

Hammerbacher, A., Schmidt, A., Wadke, N., Wright, L. P., Schneider, B., Bohlmann, J., et al. (2013). A common fungal associate of the spruce bark beetle metabolizes the stilbene defenses of Norway spruce. *Plant Physiol.* 162, 1324–1336. doi: 10.1104/pp.113.218610

Heijari, J., Nerg, A. M., Kainulainen, P., Vuorinen, M., and Holopainen, J. K. (2008). Long-term effects of exogenous methyl jasmonate application on Scots pine (*Pinus sylvestris*) needle chemical defense and diprionid sawfly performance. *Entomol. Exp. Appl.* 128, 162–171. doi: 10.1111/j.1570-7458.2008.00708.x

Keeling, C. I., and Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry* 67, 2415–2423. doi: 10.1016/j.phytochem.2006.08.019

Keeling, C. I., Dullat, H. K., Yuen, M., Ralph, S. G., Jancsik, S., and Bohlmann, J. (2009). Identification and functional characterization of monofunctional *ent*-copalyl diphosphate and *ent*-kaurene synthases in White Spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiol.* 152, 1197–1208. doi: 10.1104/pp.109.151456

Krokene, P. (2015). "Conifer defense and resistance to bark beetles," in *Bark Beetles: Biology and Ecology of Native and Invasive Species*, eds R. W. Hofstetter and F. E. Vega (Amsterdam: Elsevier), 177–207. doi: 10.1016/b978-0-12-417156-5.00005-8

Krokene, P., Solheim, H., Krekling, T., and Christiansen, E. (2003). Inducible anatomical defense responses in Norway spruce stems and their possible role in induced resistance. *Tree Physiol.* 23, 191–197. doi: 10.1093/treephys/23.3.191

Lewinsohn, E., Savage, T. J., Gijzen, M., and Croteau, R. B. (1993). Simultaneous analysis of monoterpenes and diterpenoids of conifer oleoresin. *Phytochem. Anal.* 4, 220–225. doi: 10.1002/pca.2800040506

Liang, P. H., Ko, T. P., and Wang, A. H. J. (2002). Structure, mechanism and function of prenyltransferases. *Eur. J. Biochem.* 269, 3339–3354. doi: 10.1046/j.1432-1033.2002.03014.x

Lieutier, F., Brignolas, F., Sauvard, D., Yart, A., Galet, C., Brunet, M., et al. (2003). Intra- and inter-provenance variability in phloem phenols of *Picea abies* and relationship to a bark beetle-associated fungus. *Tree Physiol.* 23, 247–256. doi: 10.1093/treephys/23.4.247

Mageroy, M. H., Christiansen, E., Långström, B., Borg-Karlson, A.-K., Solheim, H., Björklund, N., et al. (2020). Priming of inducible defenses protects Norway spruce against tree-killing bark beetles. *Plant Cell Environ.* 43, 420–430. doi: 10.1111/pce.13661

Martin, D., Tholl, D., Gershenzon, J., and Bohlmann, J. (2002). Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiol.* 129, 1003–1018. doi: 10.1104/pp.011001

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

Nagel, R., Berasategui, A., Paetz, C., Gershenzon, J., and Schmidt, A. (2014). Overexpression of an isoprenyl diphosphate synthase in spruce leads to unexpected terpene diversion products that function in plant defense. *Plant Physiol.* 164, 555–569. doi: 10.1104/pp.113.228940

Nagel, R., Gershenzon, J., and Schmidt, A. (2012). Nonradioactive assay for detecting isoprenyl diphosphate synthase activity in crude plant extracts using liquid chromatography coupled with tandem mass spectrometry. *Anal. Biochem.* 422, 33–38. doi: 10.1016/j.ab.2011.12.037

Nagel, R., and Peters, R. J. (2017). Investigating the phylogenetic range of gibberellin biosynthesis in bacteria. *Mol. Plant Microbe Interact.* 30, 343–349. doi: 10.1094/MPMI-01-17-0001-R

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

Nagy, N., Franceschi, V. R., Solheim, H., Krekling, T., and Christiansen, E. (2000). Wound-induced traumatic resin duct development in stems of Norway spruce (Pinaceae): anatomy and cytochemical traits. *Am. J. Bot.* 87, 302–313. doi: 10.2307/2656626

Peters, R. J., Ravn, M. M., Coates, R. M., and Croteau, R. B. (2001). Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* 123, 8974–8978. doi: 10.1021/ja010670k

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2018). *nlme: Linear and Nonlinear Mixed Effects Models. R Package version 3.1-131.1.* Available Online at: https://CRAN.R-project.org/package=nlme

Prisic, S., Xu, M., Wilderman, P. R., and Peters, R. J. (2004). Rice contains two disparate *ent*-copalyl diphosphate synthases with distinct metabolic functions. *Plant Physiol.* 136, 4228–4236. doi: 10.1104/pp.104.050567

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for statistical computing.

Robert-Seilaniantz, A., Navarro, L., Bari, R., and Jones, J. D. G. (2007). Pathological hormone imbalances. *Curr. Opin. Plant Biol.* 10, 372–379.

Schiebe, C., Hammerbacher, A., Birgersson, G., Witzell, J., Brodelius, P. E., Gershenzon, J., et al. (2012). Inducibility of chemical defenses in Norway spruce bark is correlated with unsuccessful mass attacks by the spruce bark beetle. *Oecologia* 170, 183–198. doi: 10.1007/s00442-012-2298-8

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

Schmidt, A., and Gershenzon, J. (2007). Cloning and characterization of isoprenyl diphosphate synthases with farnesyl diphosphate and geranylgeranyl diphosphate synthase activity from Norway spruce (*Picea abies*) and their relation to induced oleoresin formation. *Phytochemistry* 68, 2649–2659. doi: 10.1016/j.phytochem.2007.05.037

Schmidt, A., and Gershenzon, J. (2008). Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). *Phytochemistry* 69, 49–57. doi: 10.1016/j.phytochem.2007.06.022

Schmidt, A., Nagel, R., Krekling, T., Christiansen, E., Gershenzon, J., and Krokene, P. (2011). Induction of isoprenyl diphosphate synthases, plant hormones and defense signalling genes correlates with traumatic resin duct formation in Norway spruce (*Picea abies*). *Plant Mol. Biol.* 77, 577–590. doi: 10.1007/s11103-011-9832-7

Schmidt, A., Waechtler, B., Temp, U., Krekling, T., Seguin, A., and Gershenzon, J. (2010). A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiol.* 152, 639–655. doi: 10.1104/pp.109.144691

Ullah, C., Tsai, C.-J., Unsicker, S. B., Xue, L., Reichelt, M., Gershenzon, J., et al. (2019). Salicylic acid activates poplar defense against the biotrophic rust fungus *Melampsora larici-populina* via increased biosynthesis of catechin and proanthocyanidins. *New Phytol.* 221, 960–975. doi: 10.1111/nph.15396

Viiri, H., Annila, E., Kitunen, V., and Niemela, P. (2001). Induced responses in stilbenes and terpenes in fertilized Norway spruce after inoculation with blue-stain fungus, *Ceratocystis polonica*. *Trees* 15, 112–122. doi: 10.1007/s004680000082

Wadke, N., Kandasamy, D., Vogel, H., Lah, L., Wingfield, B. D., Paetz, C., et al. (2016). Catechol dioxygenases catalyzing the first step in Norway spruce phenolic degradation are key virulence factors in the bark beetle-vectored fungus *Endoconidiophora polonica*. *Plant Physiol.* 171, 914–931.

Wang, K. C., and Ohnuma, S. (2000). Isoprenyl diphosphate synthases. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1529, 33–48. doi: 10.1016/s1388-1981(00)00136-0

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

Wermelinger, B. (2004). Ecology and management of the spruce bark beetle *Ips typographus* - a review of recent research. *For. Ecol. Manag.* 202, 67–82. doi: 10.1016/j.foreco.2004.07.018

Wu, Y., Zhou, K., Toyomasu, T., Sugawara, C., Oku, M., Abe, S., et al. (2012). Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* 84, 40–46. doi: 10.1016/j.phytochem.2012.08.022

Xu, M., Hillwig, M. L., Prisic, S., Coates, R. M., and Peters, R. J. (2004). Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. *Plant J.* 39, 309–318. doi: 10.1111/j.1365-313X.2004.02137.x

Zeneli, G., Krokene, P., Christiansen, E., Krekling, T., and Gershenzon, J. (2006). Methyl jasmonate treatment of mature Norway spruce (*Picea abies*) trees increases the accumulation of terpenoid resin components and protects against infection by *Ceratocystis polonica*, a bark beetle-associated fungus. *Tree Physiol.* 26, 977–988. doi: 10.1093/treephys/26.8.977

Zhao, T., Krokene, P., Bjorklund, N., Langstrom, B., Solheim, H., Christiansen, E., et al. (2010). The influence of *Ceratocystis polonica* inoculation and methyl jasmonate application on terpene chemistry of Norway spruce, *Picea abies*. *Phytochemistry* 71, 1332–1341. doi: 10.1016/j.phytochem.2010.05.017

Zhao, T., Krokene, P., Hu, J., Christiansen, E., Björklund, N., Langstrom, B., et al. (2011). Induced terpene accumulation in Norway spruce inhibits bark beetle colonization in a dose-dependent manner. *PLoS One* 6:e26649. doi: 10.1371/journal.pone.0026649

Zi, J., Mafu, S., and Peters, R. J. (2014). To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annu. Rev. Plant Biol.* 65, 259–286. doi: 10.1146/annurev-arplant-050213-035705

Zuur, A. F., Leno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 June 2022
doi: 10.3389/fpls.2022.862915

Jasmonate-Responsive Transcription Factors NnWRKY70a and NnWRKY70b Positively Regulate Benzylisoquinoline Alkaloid Biosynthesis in Lotus (*Nelumbo nucifera*)

*Jing Li[1], Yi Li[1], Mingjing Dang[1], Shang Li[1], Simeng Chen[1], Ruizhen Liu[1], Zeyu Zhang[1], Guoqian Li[1], Minghua Zhang[2], Dong Yang[2], Mei Yang[2], Yanling Liu[2], Daike Tian[3] and Xianbao Deng[2,3]**

[1] School of Chemistry, Chemical Engineering and Life Sciences, Wuhan University of Technology, Wuhan, China, [2] Aquatic Plant Research Center, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, [3] Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai, China

OPEN ACCESS

Edited by:
Sumit Ghosh,
Council of Scientific and Industrial Research (CSIR), India

Reviewed by:
Barunava Patra,
University of Kentucky, United States
Yasuyuki Yamada,
Kobe Pharmaceutical University, Japan

***Correspondence:**
Xianbao Deng
dengxianbao@wbgcas.cn

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

Received: 26 January 2022
Accepted: 17 May 2022
Published: 15 June 2022

Citation:
Li J, Li Y, Dang M, Li S, Chen S, Liu R, Zhang Z, Li G, Zhang M, Yang D, Yang M, Liu Y, Tian D and Deng X (2022) Jasmonate-Responsive Transcription Factors NnWRKY70a and NnWRKY70b Positively Regulate Benzylisoquinoline Alkaloid Biosynthesis in Lotus (Nelumbo nucifera). Front. Plant Sci. 13:862915. doi: 10.3389/fpls.2022.862915

Lotus (*Nelumbo nucifera*) is a large aquatic plant that accumulates pharmacologically significant benzylisoquinoline alkaloids (BIAs). However, little is known about their biosynthesis and regulation. Here, we show that the two group III WRKY transcription factors (TFs), NnWRKY70a and NnWRKY70b, positively regulate the BIA biosynthesis in lotus. Both NnWRKY70s are jasmonic acid (JA) responsive, with their expression profiles highly correlated to the BIA concentration and BIA pathway gene expression. A dual-luciferase assay showed that NnWRKY70a could transactivate the *NnTYDC* promoter, whereas NnWRKY70b could activate promoters of the three BIA structural genes, including *NnTYDC*, *NnCYP80G*, and *Nn7OMT*. In addition, the transient overexpression of *NnWRKY70a* and *NnWRKY70b* in lotus petals significantly elevated the BIA alkaloid concentrations. Notably, NnWRKY70b seems to be a stronger BIA biosynthesis regulator, because it dramatically induced more BIA structural gene expressions and BIA accumulation than NnWRKY70a. A yeast two-hybrid assay further revealed that NnWRKY70b physically interacted with NnJAZ1 and two other group III WRKY TFs (NnWRKY53b and NnWRKY70a), suggesting that it may cooperate with the other group III WRKYs to adjust the lotus BIA biosynthesis via the JA-signaling pathway. To illustrate the mechanism underlying NnWRKY70b-mediated BIA regulation in the lotus, a simplified model is proposed. Our study provides useful insights into the regulatory roles of WRKY TFs in the biosynthesis of secondary metabolites.

Keywords: NnWRKY70, transcriptional regulation, lotus, benzylisoquinoline alkaloid biosynthesis, secondary metabolite

INTRODUCTION

Benzylisoquinoline alkaloids (BIA) are a diverse group of nitrogen-containing secondary metabolites in plants, with over 2,500 known structures to date (Facchini, 2001). Unlike other secondary metabolites that are found in most of the higher plants, BIAs only occur in limited plant families, such as Magnoliaceae, Ranunculaceae, Papaveraceae, and Berberidaceae, and are mostly pharmacologically significant. The typical medicinal BIAs include morphine and codeine (narcotic analgesics), sanguinarine and berberine (anti-microbial agents), tubocurarine and papaverine (muscle relaxants), and noscapine (cough suppressant and anti-cancer agent) (Hagel and Facchini, 2013).

Despite their high-structural diversity, the biosynthesis of all BIAs in plants is conserved to a single L-tyrosine substrate origin. Tyrosine/DOPA decarboxylase (TYDC) converts L-tyrosine into dopamine and 4-hydroxyphenylacetaldehyde, which are subsequently condensed by (*S*)-norcoclaurine synthase (NCS) to yield (*S*)-norcoclaurine, the common precursor of all plant BIAs (**Figure 1**) (Facchini and De Luca, 1994; Maldonado-Mendoza et al., 1996; Samanani et al., 2004; Minami et al., 2007). Four additional enzymatic steps, catalyzed by (*S*)-norcoclaurine 6-*O*-methyltransferase (6OMT), (*S*)-coclaurine *N*-methyltransferase (CNMT), (*S*)-*N*-methylcoclaurine 3′-hydroxylase (CYP80B), and (*S*)-3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase (4′OMT), convert (*S*)-norcoclaurine to (*S*)-reticuline, a common intermediate branch point of most BIAs (Sato et al., 1994; Pauli and Kutchan, 1998; Morishige et al., 2000; Samanani et al., 2004). In the past decades, efforts have been devoted to elucidate the biosynthetic pathways of morphine, berberine, and sanguinarine because of their pharmacological significance. As a result, nearly all enzymes involved in the biosynthesis of these BIAs have been characterized, and *Papaver somniferum*, *Coptis japonica*, and *Eschscholzia californica* became model species for the study of BIA biosynthesis.

In plants, the biosynthesis and proper accumulation of BIAs are strictly controlled in a spatial and temporal manner, and may be influenced by numerous internal and external signals. Jasmonic acid (JA) and its derivatives are plant stress hormones that regulate various stress responses, such as microbial infections, herbivores, wounding, and UV-irradiation (Zhou and Memelink, 2016). The external application of JAs can induce the biosynthesis of most secondary metabolites by activating JA-responsive transcription factors (TFs), including AP2/ERFs, bHLHs, MYBs, and WRKYs (Patra et al., 2013; Wasternack and Strnad, 2019). These TFs in turn bind to specific *cis*-regulatory sequences of the biosynthetic gene promoters, leading to their transcriptional activation. In contrast, jasmonate ZIM-domain proteins (JAZ) are key negative regulators of JA signaling by binding to JA-responsive TFs and repressing their transcription. Exogenous JA treatment induces JAZ protein degradation, which sets the JAZ-repressed TFs free and activates the JA-responsive metabolic pathways (Santner and Estelle, 2007; Song et al., 2014; Wasternack and Strnad, 2019).

WRKY-domain-containing genes form one of the largest TF families in plants and play essential roles in the plant JA-signaling cascade (Eulgem et al., 2000; Rushton et al., 2010; Chen et al., 2018). The polypeptide sequences of these TFs characteristically comprise a WRKY DNA binding domain of approximately 60 amino acids, with one or two highly conserved WRKYGQK heptapeptides in the N-terminals and a C2H2 or C2HC zinc finger motif in the C-terminals (Eulgem et al., 2000; Li et al., 2019). Studies have shown that WRKY TFs, especially members of group III, are involved in the regulation of secondary metabolite biosynthesis in various medicinal plants. For example, CrWRKY1, a group III WRKY protein of *Catharanthus roseus*, positively regulates the monoterpenoid indole alkaloid (MIA) biosynthesis through binding and activating the *CrTDC* gene in the MIA biosynthetic pathway (Suttipanta et al., 2011). CjWRKY1, a *Coptis japonica* WRKY in the IIc group, has been reported to be JA responsive and specifically bound to the W-boxes in the structural gene promoters to regulate the berberine biosynthesis (Kato et al., 2007; Yamada and Sato, 2016). More recently, three group III WRKYs in *Ophiorrhiza pumila* were shown to regulate the biosynthesis of the anticancer drug camptothecin. Of them, OpWRKY1 represses the expression of *OpCPR*, and negatively regulates camptothecin accumulation (Xu et al., 2020), whereas both OpWRKY2 and OpWRKY3 could activate the camptothecin pathway gene expression and positively regulate the camptothecin biosynthesis (Wang et al., 2019; Hao et al., 2021).

Lotus (*Nelumbo nucifera*) is an aquatic plant species widely cultivated in Asian countries (Deng et al., 2022). In addition to their attractive flowers and nutritious rhizome and seeds, lotuses are rich in valuable medicinal BIAs, including nuciferine, *N*-nornuciferine, *O*-nornuciferine, roemerine, and anonaine in lotus leaves, as well as Liensinine, Isoliensinine, and Neferine in lotus embryos (Deng et al., 2016, 2018; Lin et al., 2019). Lotus contains 65 WRKY encoding genes, 34 of which are JA responsive and are deemed to be potential BIA biosynthesis regulators (Li et al., 2019). Previously, we have shown that two JA-responsive lotus WRKYs, NnWRKY40a and NnWRKY40b of the IIa group, could activate the promoters of two BIA biosynthetic genes (Li et al., 2019). Here, we present an extensive functional evaluation of two other lotus group III WRKY proteins: NnWRKY70a and NnWRKY70b. Our results showed that both NnWRKY70s positively regulate the BIA accumulation in the lotus through the transcriptional modulation of structural gene expressions.

MATERIALS AND METHODS

Plant Material, JA Treatment, and Alkaloid Quantification

The lotus variety "Qiuxing" was rhizome propagated in pots of 40 cm diameter and 40 cm height in late April. All pots were set outdoor on a flat ground at the Wuhan Botanical Garden of the Chinese Academy of Sciences (Wuhan, Hubei province, China). Routine water and fertilization management were applied to all the plants during the growth season. *Nicotiana benthamiana* plants were cultivated in the growth room

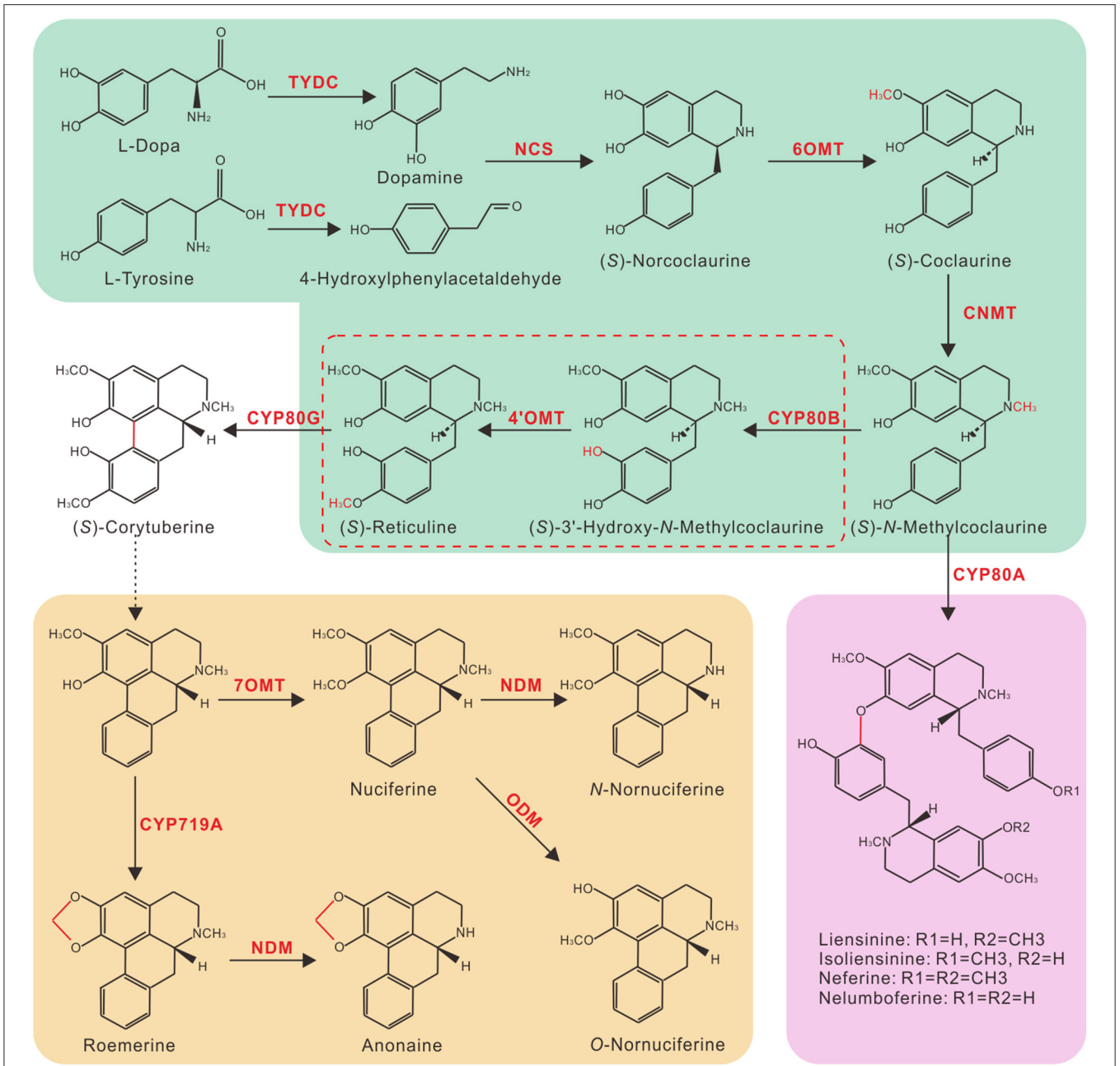


FIGURE 1 | Putative benzylisoquinoline alkaloid biosynthesis (BIA) pathway in lotus. Branch pathways are shown in colored boxes: green, common steps to (*S*)-Reticuline; brown, aporphine alkaloids in lotus leaves; and purple, bis-benzylisoquinoline alkaloids in lotus embryos. Solid arrows indicate single-step reactions, dotted arrow indicates multiple steps, and the dotted box indicates possible redundant enzymatic steps in lotus. TYDC, tyrosine/DOPA decarboxylase; NCS, (*S*)-norcoclaurine synthase; 6OMT, (*S*)-norcoclaurine 6-*O*-methyltransferase; CNMT, (*S*)-coclaurine *N*-methyltransferase; CYP80B, (*S*)-*N*-methylcoclaurine 3′-hydroxylase; 4′OMT, (*S*)-3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase; NDM, *N*-demethylase; ODM, *O*- demethylase.

under controlled conditions: day/night temperature, 24/22°C; day/night length, 16/8 h; light intensity, 250 μmol $m^{-2}s^{-1}$; and relative humidity, 60%.

For JA treatment, methyl jasmonate (MeJA, 100 μM) was exogenously applied on lotus leaves at developmental stage 3 (S3) (Deng et al., 2016) and then wrapped with transparent plastic bags. A final concentration of 100 μM MeJA was prepared by first dissolving 11.2 μL MeJA (4.4 M) in 488.8 μL pure ethanol to make a 100 mM (1000×) MeJA stock solution, which was then diluted with deionized water to the desired concentration. Leaf samples were collected at 0, 3, 6, and 24 h time intervals after MeJA treatment. For a tissue-specific expression analysis, eight lotus organ samples, including, root, rhizome, leaf, petal, embryo, seedpod, petiole, and stamen, as well as leaves at seven

developmental stages (S1–S7) were harvested in July. All the samples were frozen immediately in liquid nitrogen and then stored at −80°C until use. The BIA alkaloids in the lotus were extracted and quantified as previously described (Deng et al., 2016).

Bioinformatic and Phylogenetic Analysis

The protein sequences of all WRKYs previously reported to be involved in secondary metabolism were obtained from the National Center for Biotechnology Information (NCBI) GenBank. The genomic and coding sequences of *AtWRKY70* were retrieved from the TAIR database (https://www.arabidopsis.org/). The sequences of lotus genes were PCR amplified from the lotus variety "Qiuxing" and have been deposited in the GenBank with the following accession numbers: *NnWRKY70a*, OL469000; *NnWRKY70b*, OL469001; *NnWRKY53b*, OL468999; and *NnJAZ1*, OL469002. The full-length protein sequences were aligned with MUSCLE (Edgar, 2004), followed by a phylogenetic analysis in MEGA7.0 (Kumar et al., 2016) with a neighbor-joining method and 1,000 bootstrap replicates. The tree was finally viewed and modified with FigTree V1.4.2.

Quantitative Real-Time PCR

Total RNA was extracted using the RNAprep Pure Plant Kit (Tiangen Biotec, Beijing, China), and cDNA was synthesized with the cDNA Synthesis SuperMix (TransGen, Beijing, China). The quantitative real-time PCR was carried out using SYBR® Premix Ex Taq™ II (Takara, Dalian, China) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative gene expression (Pfaffl, 2001), and the *NnACTIN* gene was used as an internal reference for the normalization of gene expression levels (Gu et al., 2013). The primers used for real-time PCR are listed in **Supplementary Table S1**.

Subcellular Localization Assay

The full-length coding sequences of *NnWRKY70a* and *NnWRKY70b* without the stop codon were amplified and inserted into the pMDC83 vector to generate CaMV 35S:NnWRKY70a-GFP and 35S:NnWRKY70b-GFP constructs. The empty pMDC83-GFP vector was used as a control. The recombinant constructs were transformed into the *Agrobacterium* GV3101 strain. Agro-infiltration was conducted on fully expanded leaves of 5–6-weeks-old *N. benthamiana* plants. The subcellular localization of fluorescent protein fusions was checked 2 days after the inoculation. A 5 μg/mL DAPI nuclear marker was infiltrated into the same area of agro-infiltration 30 min before sampling. The images were taken with the Zeiss Confocal Fluorescence Microscope (LSM710 Meta, Carl Zeiss) as previously reported (Li et al., 2019). The primers used for cloning subcellular localization vectors are listed in **Supplementary Table S1**.

Dual-Luciferase Reporter Assay

The promoter regions spanning about 1.5–2 kb of *NnNCS1, NnTYDC1, NnCYP80G,* and *Nn7OMT* genes were PCR amplified and inserted into the dual-luciferase (LUC) reporter gene expression vector pGreen 0800-LUC to form *NnNCS1*pro::LUC, *NnTYDC1*pro::LUC, *NnCYP80G*pro::LUC, and *Nn7OMT*pro::LUC constructs. The complete coding sequence (CDS) of *NnWRKY70a* and *NnWRKY70b* genes were cloned into the pSAK277 vector under the 35S promoter to form the TF expression vectors. *Agrobacterium* transformation, agro-infiltration, and the measurements of Firefly luciferase (F-Luc) and Renilla luciferase (R-Luc) activities were performed as previously described (Deng et al., 2018). The primers used in the dual-luciferase reporter assay are listed in **Supplementary Table S1**.

Yeast Hybrid Assays

The yeast one-hybrid assay was conducted according to the instructions of Matchmaker® Gold Yeast One-Hybrid Library Screening System User Manual (Clonteck, USA). The promoter sequences of approximately 1.5-kb upstream of the BIA pathway gene start codon were PCR amplified and cloned into pAbAi vectors to generate bait-reporter yeast strains, while whole coding sequences of *NnWRKY70a* and *NnWRKY70b* were cloned into the pGADT7 vector to generate pGADT7-TF constructs. The minimal inhibitory concentration of Aureobasidin A (100 ng/mL) was determined with different bait-reporter yeast strains transformed with empty pGADT7 AD vectors. To evaluate whether NnWRKY70a and NnWRKY70b interact with BIA pathway gene promoters, pGADT7-TF plasmids were transformed into Y1HGold bait strains carrying different promoters and cultured on SD/-Leu containing 100 ng/mL AbA. Y1HGold bait strains transformed with empty pGADT7 were set as negative controls, while Y1HGold (p53-AbAi) transformed with pGADT7-p53 was set as a positive control. To check the specific binding of NnWRKY70a and NnWRKY70b to the W-box *cis*-elements, the promoter mutants were generated with overlap extension PCR to remove all W-box *cis*-elements.

For the yeast two-hybrid library screening, a lotus "Mate & Plate" library was constructed in Y187 yeast strain according to the "Make Your own Mate & Plate Library System User Manual" (Clonteck, USA). cDNA for the library was generated from a tissue mixture including leaves, petals, and embryos of the lotus variety "China Antique". The library screening and yeast two-hybrid assays (Y2H) were conducted as described in the "Matchmaker Gold Yeast Two-Hybrid System User Manual" (Clonteck, USA). Truncated *NnWRKY70b* coding sequence harboring the WRKY domain was PCR amplified and cloned into the pGBKT7 (BD) vector as a bait to avoid autoactivation, while the full-length CDS of *NnWRKY70a*, *NnWRKY70b*, *NnWRKY53b,* and *NnJAZ1* were cloned into the pGADT7(AD) vector as a prey. Combinations of the different prey vectors and BD-NnWRKY70b bait were co-transformed into the Y2H Gold yeast strain. Meanwhile, the combinations of BD-53 + AD-T and BD-Lam + AD-T were co-transformed as positive and negative controls, respectively. The yeast cells were cultured on DDO (SD/-Leu/-Trp) selective medium and on TDO (SD/-Leu/-Trp/-His) and QDO (SD/-Leu/-Trp/-His/-Ade) media for interaction assays. The primers used for cloning Y1H and Y2H vectors are listed in **Supplementary Table S1**.

Transient Overexpression of *NnWRKY70s* in Lotus Petals

Transient overexpression of *NnWRKY70s* was conducted in the lotus petals of the "Qiuxing" variety 2 days before blooming. *Agrobacterium* GV3101 strain carrying the pMDC83:NnWRKY70a and pMDC83:NnWRKY70b vectors were infiltrated on the lower side of petals at OD600 = 0.5, using a 1 mL needleless syringe. The lotus petals were sampled 2 days after infiltration for subsequent gene expression analysis and determination of alkaloid content.

Bimolecular Fluorescence Complementation (BiFC) Assay

BiFC assay was conducted with a 2-in-1 system based on the splitting enhanced yellow fluorescent protein (EYFP) as previously described (Grefen and Blatt, 2012). Coding DNA sequences of two independent genes were gateway cloned into two different expression cassettes on the pBiFCt-2in1-CN vector. To construct the NnWRKY70b:nYFP fusion protein expression cassette, a termination code of NnWRKY70b was dropped out. *Agrobacterium tumefaciens* strain GV3101 carrying different constructs was infiltrated in *N. benthamiana* leaves. Yellow and red fluorescence was observed with Confocal Fluorescence Microscope as described in the subcellular localization assays.

RESULTS

Nucleotide and Protein Sequence Characteristics of Lotus WRKY70s

Previously, two independent studies have identified NnWRKY70a (Nnu_24385) and NnWRKY70b (Nnu_12194) as key candidate regulators of the lotus BIA biosynthesis based on a positive correlation between their expression and BIA accumulation in lotus (Deng et al., 2018; Meelaph et al., 2018). To evaluate the possible roles of these two WRKYs in regulating the lotus BIA biosynthesis, we first analyzed their nucleotide and amino acid sequence characteristics. *NnWRKY70a* was found to be located on the lotus chromosome 5, which contains a 981 nucleotide (nt) open reading frame (ORF), encoding a protein with a calculated molecular weight of 36.5 kDa. *NnWRKY70b* was located on chromosome 2, which has a 990 nt ORF, encoding a 37.06 kDa protein. Pairwise sequence alignment showed that the two NnWRKY70s shared high-sequence identity, with nucleotide and amino acid sequence similarity of 70.6 and 74.4%, respectively.

A phylogenetic analysis was conducted using two NnWRKY70s and other previously characterized WRKYs involved in the regulation of plant secondary metabolism (**Supplementary Table S2**). Two NnWRKYs clustered together with group III members (**Figure 2A**). NnWRKY70s showed the highest sequence similarity with the AtWRKY70, followed by CrWRKY1 and OpWRKY1. Both CrWRKY1 and OpWRKY1 are group III WRKYs involved in regulating MIA biosynthesis in *Catharanthus roseus* and *Ophiorrhiza pumila*, respectively (Suttipanta et al., 2011; Xu et al., 2020). Similar to the other group III WRKYs, both NnWRKY70s contained a core WRKYGQK heptapeptide and a conserved $CX_7CX_{23}HXC$ zinc-finger motif in their WRKY domains (**Figure 2B**). Therefore, NnWRKY70a and NnWRKY70b were both typical group III WRKY proteins similar to AtWRKY70.

Expression Profiling and Subcellular Localization of NnWRKY70s

Next, we analyzed the spatial expression patterns of *NnWRKY70* genes in the lotus variety "Qiuxing" by quantitative real-time PCR. The two *WRKY70*s showed quite distinct expression patterns (**Figure 3A**). Overall, *NnWRKY70b* was predominantly expressed in the lotus root, followed by the rhizome and leaf. In contrast, *NnWRKY70a* had the highest transcript levels in the lotus leaf, where mono-BIA are primarily accumulated.

Their expression profiling was also evaluated in seven lotus leaf developmental stages (**Figure 3B**). Two lotus *NnWRKY70* genes showed quite similar expression patterns, with both consistently increasing throughout the tested developmental stages. This expression pattern followed well with our previously reported BIA accumulation pattern in lotus leaves (Deng et al., 2016). In addition, we carried out a subcellular localization assay by fusing the coding regions of *NnWRKY70s* with a GFP reporter (**Figure 3C**). The GFP fluorescence of both NnWRKY70a:GFP and NnWRKY70b:GFP fused proteins co-localized with the DAPI-stained nucleus, demonstrating an obvious nuclear localization of both NnWRKY70a and NnWRKY70b proteins.

NnWRKY70s Are JA-Responsive TFs That Transactivate BIA Pathway Gene Promoters

JA is known to trigger most biosynthetic pathways leading to secondary metabolites by activating the JA-responsive TFs (Wasternack and Strnad, 2019; Yamada et al., 2020). To verify whether *NnWRKY70*s are JA responsive, we checked their expression in the MeJA-treated samples, wherein the lotus leaves were treated with 100 μM MeJA at the developmental stage 4 (Li et al., 2019). MeJA significantly induced both *NnWRKY70* genes, with their expression increasing approximately 5-fold at 24 h after treatment (**Figure 4A**). This was consistent with the increased BIA content in the lotus leaves, a 15–30% increase between 3 and 6 h, and an approximately 50% increase at 24 h after the MeJA treatment (Li et al., 2019).

Next, we screened the promoter regions of the BIA biosynthetic genes, including *NnTYDC1*, *NnNCS1*, *NnCYP80G,* and *Nn7OMT*, for potential binding sites of WRKY TFs. Conserved W-box (TTGACT) was found in the 2 kb promoter regions of *NnTYDC1*, *NnCYP80G,* and *Nn7OMT*, but not of the *NnNCS1* gene. To further confirm whether NnWRKY70a and NnWRKY70b could activate the three BIA gene promoters, a dual-luciferase assay was conducted. Transient overexpression of NnWRKY70a significantly enhanced the pNnTYDC1-driven firefly luciferase transcription but had no effect on the other three promoters (**Figure 4B**). In contrast, transient overexpression of NnWRKY70b was able to activate three promoters, including *pNnTYDC1*, *pNnCYP80G*, and *pNn7OMT2* (**Figure 4C**).

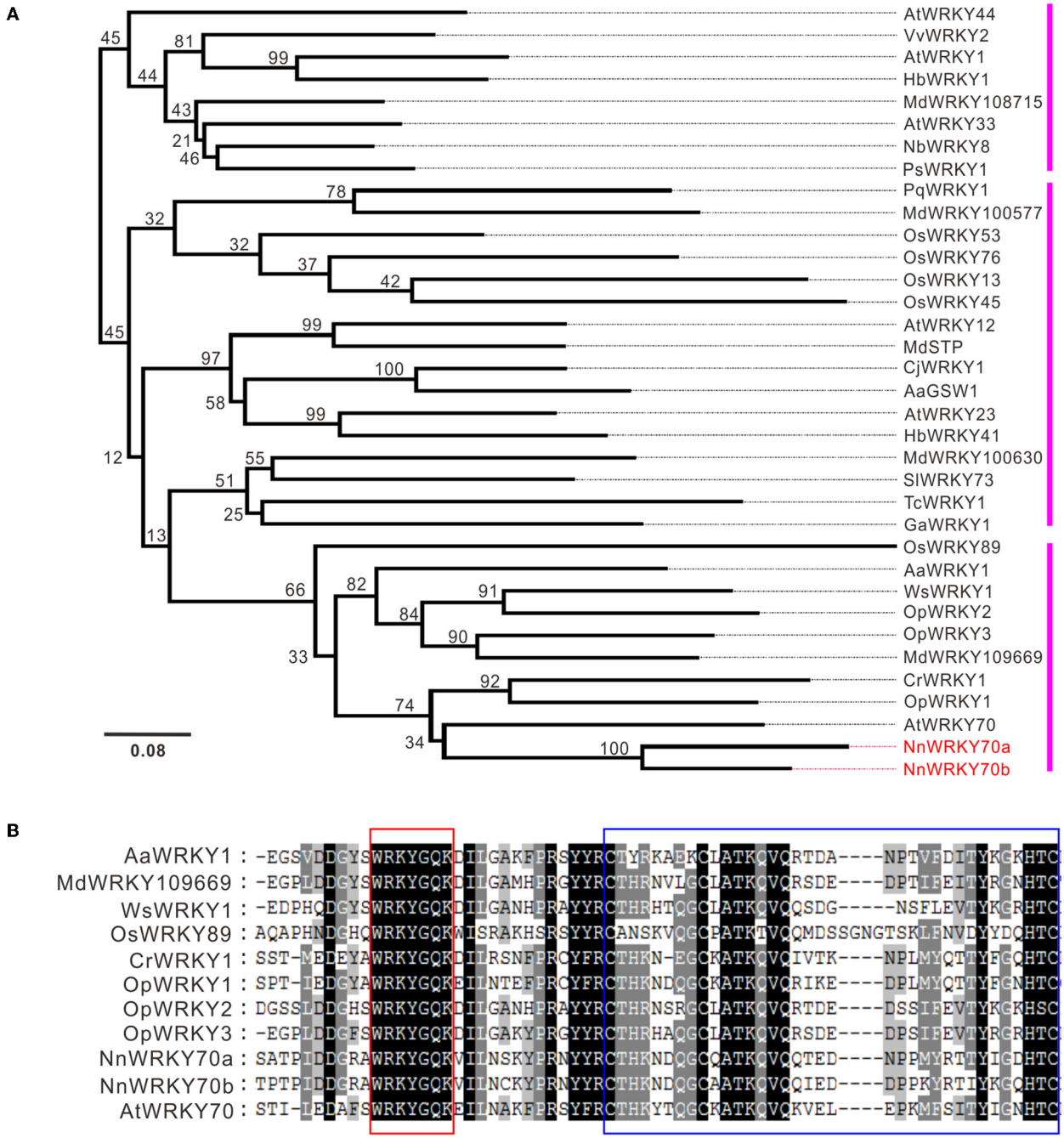


FIGURE 2 | Nucleotide and amino acid sequence characteristics of lotus WRKY70s. **(A)** Phylogenetic analysis of NnWRKY70a and NnWRKY70b with other WRKYs involved in the biosynthesis of plant secondary metabolites. Protein names, species, and GenBank accession numbers are listed in **Supplementary Table S2**. **(B)** Protein sequence alignment of the selected group III WRKY members. Conserved WRKY and zinc-finger domains are highlighted in red and blue boxes, respectively.

To further test whether NnWRKY70a and NnWRKY70b bind directly to these promoters, we conducted the yeast one-hybrid assays for putative BIA structural gene promoters and their corresponding mutants (with W-box removed). NnWRKY70b was bound to *pNnTYDC1* and *pNnCYP719A* and weakly bound to *pNnCYP80G* and *pNn7OMT* (**Figure 4D**). In contrast, NnWRKY70a was bound obviously to the *NnCYP80G* and *NnCYP719A* promoters, while weakly bound to the *NnTYDC1* and *Nn7OMT* promoters. However, neither NnWRKY70b nor NnWRKY70a bound to the *NnNCS1* promoter. Markedly, when the W-boxes were removed from these promoters, none of them were bound by NnWRKY70a or NnWRKY70b, indicating

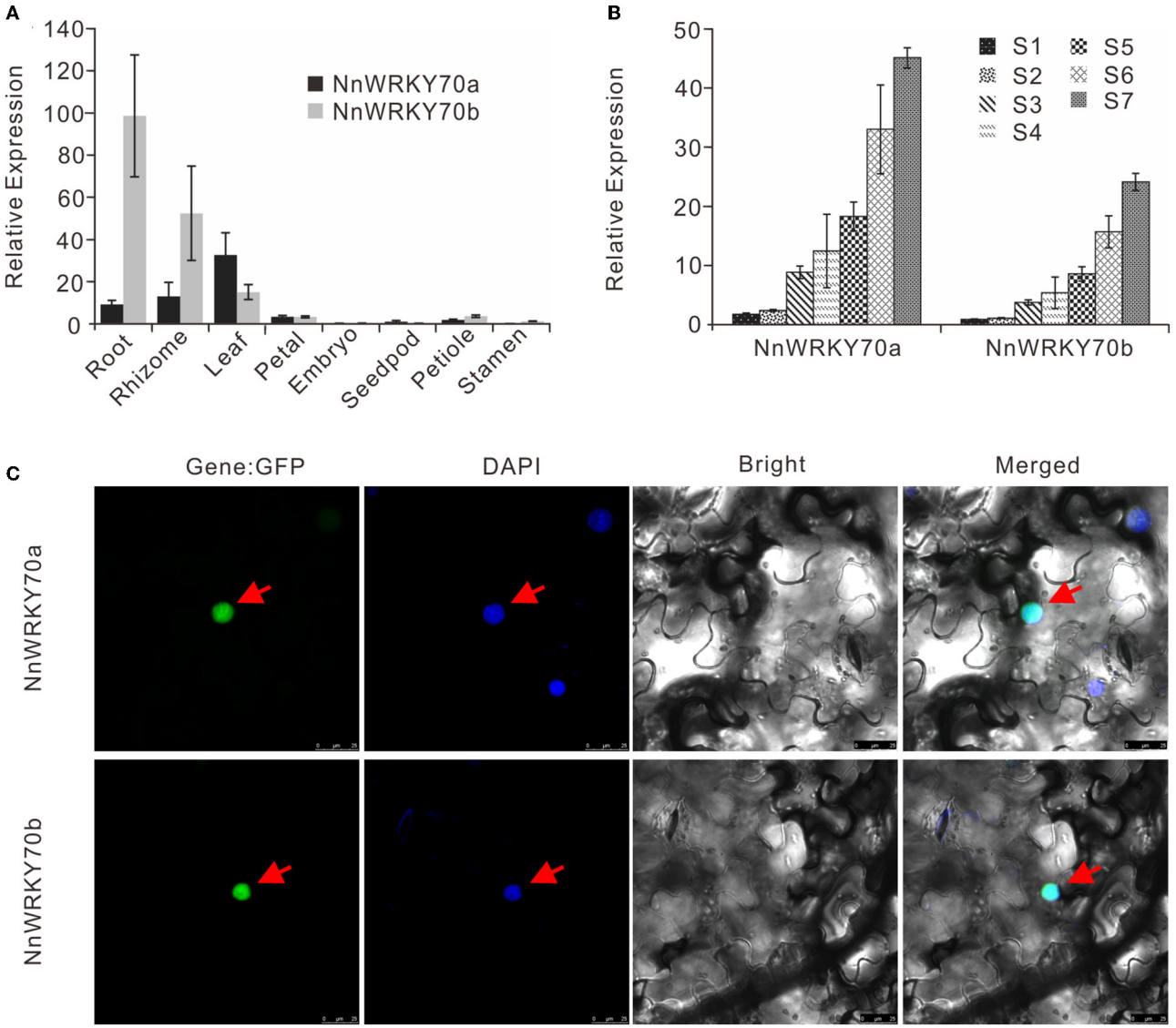


FIGURE 3 | Expression profiling and subcellular localization of NnWRKY70s. **(A)** Spatial expression profiling of *NnWRKY70*s in different lotus organs. **(B)** Temporal expression profiling of *NnWRKY70*s in seven developmental stages of lotus leaf (S1–S7). Data of **(A,B)** are mean ± SE ($n = 3$). **(C)** Subcellular localization of NnWRKY70a and NnWRKY70b in *N. benthamiana* leaves. Panels from left to right refer to GFP, DAPI, bright, and merged images, respectively. Protein fusions co-localized with DAPI markers are marked with red arrows.

a specific binding of NnWRKY70a or NnWRKY70b to the W-box *cis*-elements in the BIA biosynthetic gene promoters. Taken together, these results suggest that both NnWRKY70a and NnWRKY70b can bind and activate the BIA pathway gene promoters.

Overexpression of *NnWRKY70s* Enhances BIA Accumulation in Lotus

To determine *in planta* whether NnWRKY70a and NnWRKY70b activate the BIA biosynthesis in lotus, an *Agrobacterium*-mediated transient overexpression assay was conducted. Given that the lotus leaves were almost unable to be infiltrated, we performed agroinfiltration in the lotus petals in the "Qiuxing" variety, which is known to accumulate moderate levels of BIAs (Deng et al., 2016, 2022). As a result, a significant 4.2-fold increase in the transcript levels of *NnWRKY70a* was observed in the petals infiltrated with *Agrobacterium* carrying an overexpression vector (**Figure 5A**). Similarly, an approximately 2-fold increase in the expression of *NnWRKY70b* was observed in the *NnWRKY70b* overexpressing petals (**Figure 5B**). Interestingly, the overexpression of either *NnWRKY70a* or *NnWRKY70b* significantly promoted the transcription of the other members (**Figures 5A,B**), indicating a possible inter-activation ability between the two NnWRKY70s.

Moreover, the overexpression of the two *NnWRKY70*s enhanced the BIA contents in the lotus petals (**Figures 5A,B**).

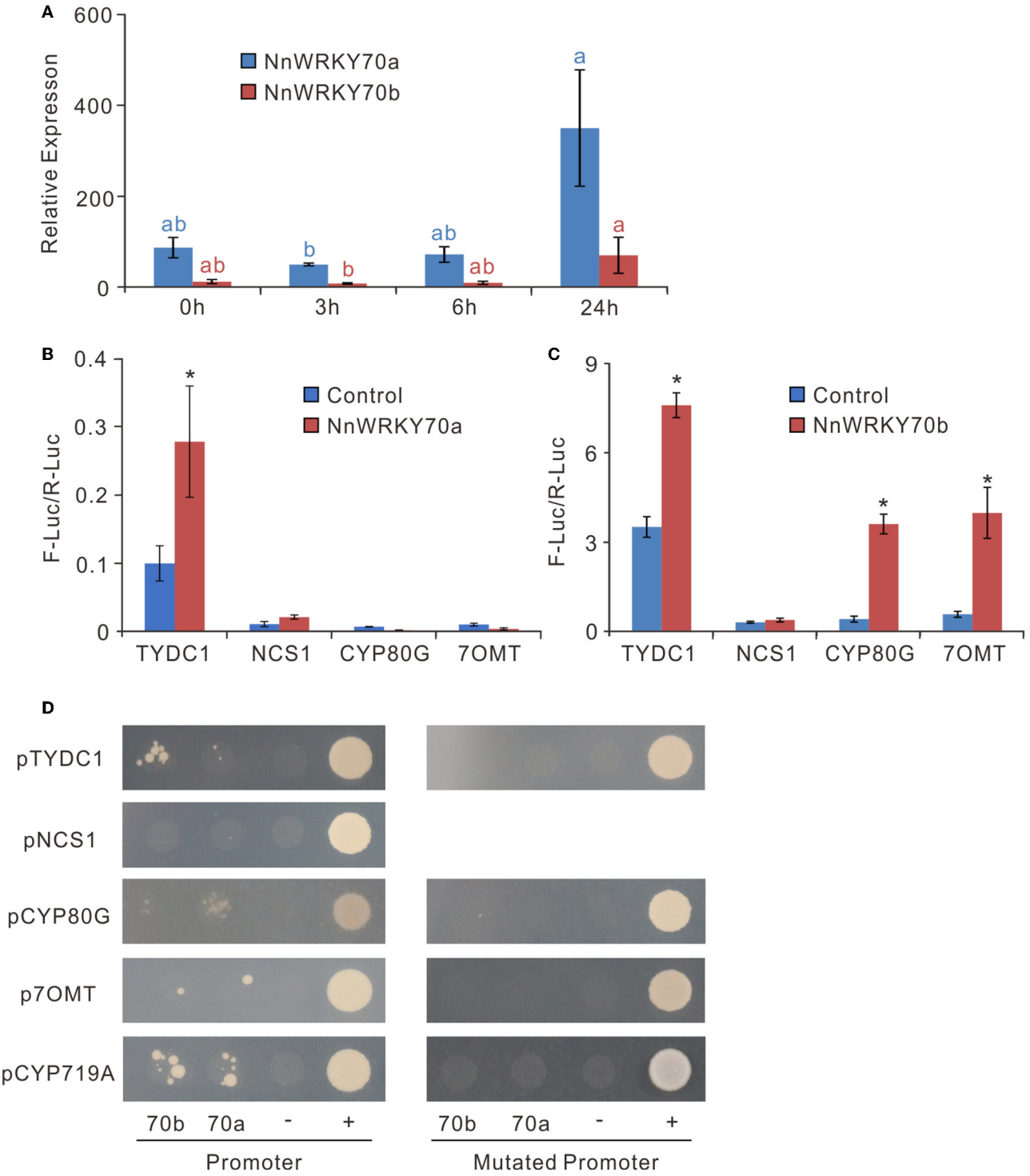


FIGURE 4 | NnWRKY70s are JA responsive and can activate the transcription of BIA biosynthetic genes. **(A)** Expression of *NnWRKY70a* and *NnWRKY70b* in lotus leaves in response to MeJA treatment. Lotus leaves were treated with MeJA at the developmental stage 4. **(B,C)** Dual-luciferase assays assessing the effect of NnWRKY70a and NnWRKY70b on activating the BIA pathway gene promoters. **(D)** Yeast one-hybrid assays checking the interactions between NnWRKY70a/NnWRKY70b transcription factors and BIA pathway gene promoters. Y1HGold bait strains transformed with pGADT7 prey vectors were cultured on SD/-Leu medium supplied with 100 ng/mL AbA. "70b", NnWRKY70b; "70a", NnWRKY70a; "-", negative control transformed with empty AD vector; "+", positive control (p53-AbAi+AD-p53); Mutated promoters, BIA pathway gene promoters with W-boxes removed. Data of **(A–C)** are mean ± SE ($n = 3$). Lower case letters and asterisks (*) represent significant differences at $P < 0.05$.

Petals overexpressing *NnWRKY70a* had approximately 60% increase in the total BIA content, whereas those overexpressing *NnWRKY70b* exhibited an even higher increase of about 126%. Consistently, increased BIA contents in petals overexpressing *NnWRKY70*s were accompanied by significant elevation of BIA pathway genes. Of the six tested BIA biosynthetic genes, three, including *NnCYP80G*, *NnCNMT*, and *Nn7OMT*, were markedly enhanced in the petals overexpressing *NnWRKY70a*, with approximately 8.1-, 4.4-, and 11.4-fold increase in the expression, respectively (**Figure 5A**). In contrast, a more dramatic increase in the gene expression was observed in the petals overexpressing *NnWRKY70b*, with *NnTYDC1*, *NnCYP80G*, *NnCNMT*, *Nn7OMT*, and *NnCYP719A* genes exhibiting approximately 151.4-, 2.2-, 3.9-, 9.7-, and 309.3-fold increase in their expression, respectively (**Figure 5B**). This further revealed that *NnWRKY70b* overexpression could induce a more dramatic increase in the BIA accumulation.

It was known that lotus petals mainly accumulate five types of BIAs, nuciferine, *N*-nornuciferine, *O*-nornuciferine, roemerine, and anonaine, with nuciferine as a predominant one (Deng et al., 2016). Of the six tested BIA pathway enzymes, NnTYDC1, NnNCS1, NnCYP80G, and NnCNMT are located upstream and are involved in the biosynthesis of all the five BIAs in lotus petals (**Figure 1**). In contrast, Nn7OMT is especially involved in the production of nuciferine, *N*-, and *O*-nornuciferine, whereas NnCYP719A mainly catalyzes the production of roemerine and anonaine. An HPLC analysis detected three BIA peaks corresponding to *N*-nornuciferine (peak 1), *O*-nornuciferine (peak 2), and nuciferine (peak 4) in the uninfiltrated "Qiuxing" petals (**Figure 5C**). The overexpression of *NnWRKY70*s significantly increased the accumulation of all the three detected BIAs (**Supplementary Figure S1**). Interestingly, the two additional peaks, representing anonaine (peak 3) and roemerine (peak 5), respectively, were evidently detected (**Figure 5C**; **Supplementary Figures S1A,B**). This is obviously due to the significant elevation of *NnCYP719A* expression in the lotus petals. Taken together, our results demonstrate that both NnWRKY70a and NnWRKY70b are involved in the BIA biosynthesis *in planta*.

NnWRKY70b Physically Interacts With NnWRKY70a, NnWRKY53b, and NnJAZ1 Proteins

As shown above, NnWRKY70a and NnWRKY70b might interregulate the expression of each other. Previous studies also showed that the group II and III WRKY members in *Arabidopsis* often function by forming homodimers or heterodimers with other WRKY TFs (Xu et al., 2006; Besseau et al., 2012). To investigate the possible interaction between NnWRKY70s and other proteins in regulating the lotus BIA biosynthesis, we performed a yeast two-hybrid cDNA library screen assay. NnWRKY70b was selected as a bait due to its significant ability to regulate the BIA biosynthesis. The bait construct harboring full-length NnWRKY70b displayed a strong autoactivation activity. Thus, truncated NnWRKY70b bait was used to screen the library. In this bait, the WRKY and zinc-finger domains were kept, while its activation domain was dropped out (**Figure 6A**). As a result, a total of 32 proteins potentially interacting with NnWRKY70b were identified (**Supplementary Table S3**). Of these interactions, NnWRKY70b vs. NnWRKY70a, NnWRKY70b vs. NnWRKY53b, and NnWRKY70b vs. NnJAZ1 were proved to be genuine during the subsequent yeast two-hybrid assays.

The truncated NnWRKY70b bait could not interact with NnWRKY70b itself but positively interacted with NnWRKY70a, and another group III type protein NnWRKY53b (**Figure 6B**). NnWRKY70b also interacted with NnJAZ1, a lotus JAZ repressor protein. Both NnWRKY53b and NnJAZ1 were also JA responsive and displayed similar expression patterns as NnWRKY70a and NnWRKY70b under the JA treatment (**Figures 6C,D**). These interactions were also confirmed *in vivo* by the BiFC assays. The EYFP fluorescence was detected in the living *N. benthamiana* epidermal cells infiltrated with pBiFCt-2in1 vectors carrying coding sequence of *NnWRKY70b* and *NnWRKY70a*, *NnWRKY70b* and *NnWRKY53b*, and *NnWRKY70b* and *NnJAZ1* (**Figure 7**).

Next, we evaluated whether NnWRKY53b interacts with NnJAZ1 and the capacity of NnWRKY53b to activate BIA pathway gene promoters. The yeast two-hybrid assay showed that NnWRKY53b did not physically interact with NnJAZ1 (**Supplementary Figure S2**). A transient trans-activation assay in tobacco indicated that NnWRKY53b alone can activate *pNnTYDC1* (**Supplementary Figure S3**). A combination of NnWRKY70b and NnWRKY53b, however, did not show enhanced activation abilities on BIA gene promoters, in compared to either of the singletons. These results suggest NnWRKY70b may regulate the lotus BIA biosynthesis through the JA-signaling pathway. The interaction between NnWRKY70b and NnWRKY53b, however, did not enhance the BIA biosynthesis.

DISCUSSION

Lotus is a large aquatic plant that has long been used as a traditional herb (Mukherjee et al., 2009). BIAs, including nuciferine and liensinine, are the major bioactive components in lotus tissues and have been proved to have significant pharmacological properties, such as antioxidant, anti-cancer, anti-HIV, anti-inflammatory, anti-obesity, and hepatoprotective effects (Sharma et al., 2017). To date, the studies on the lotus BIAs have been mainly focused on their structural elucidation, component isolation, and pharmacological function characterization, but little is still known about the regulation of their biosynthesis, with only a few available reports on the BIA biosynthesis (Vimolmangkang et al., 2016; Deng et al., 2018; Menendez-Perdomo and Facchini, 2020). This study presents a comprehensive functional analysis of the two JA-responsive WRKY TFs, NnWRKY70a and NnWRKY70b, and demonstrates their roles in the positive regulation of the BIA biosynthesis in lotus through transcriptional activation of BIA structural genes.

A phylogenetic analysis revealed that NnWRKY70a and NnWRKY70b are typical group III WRKYs, clustered closely

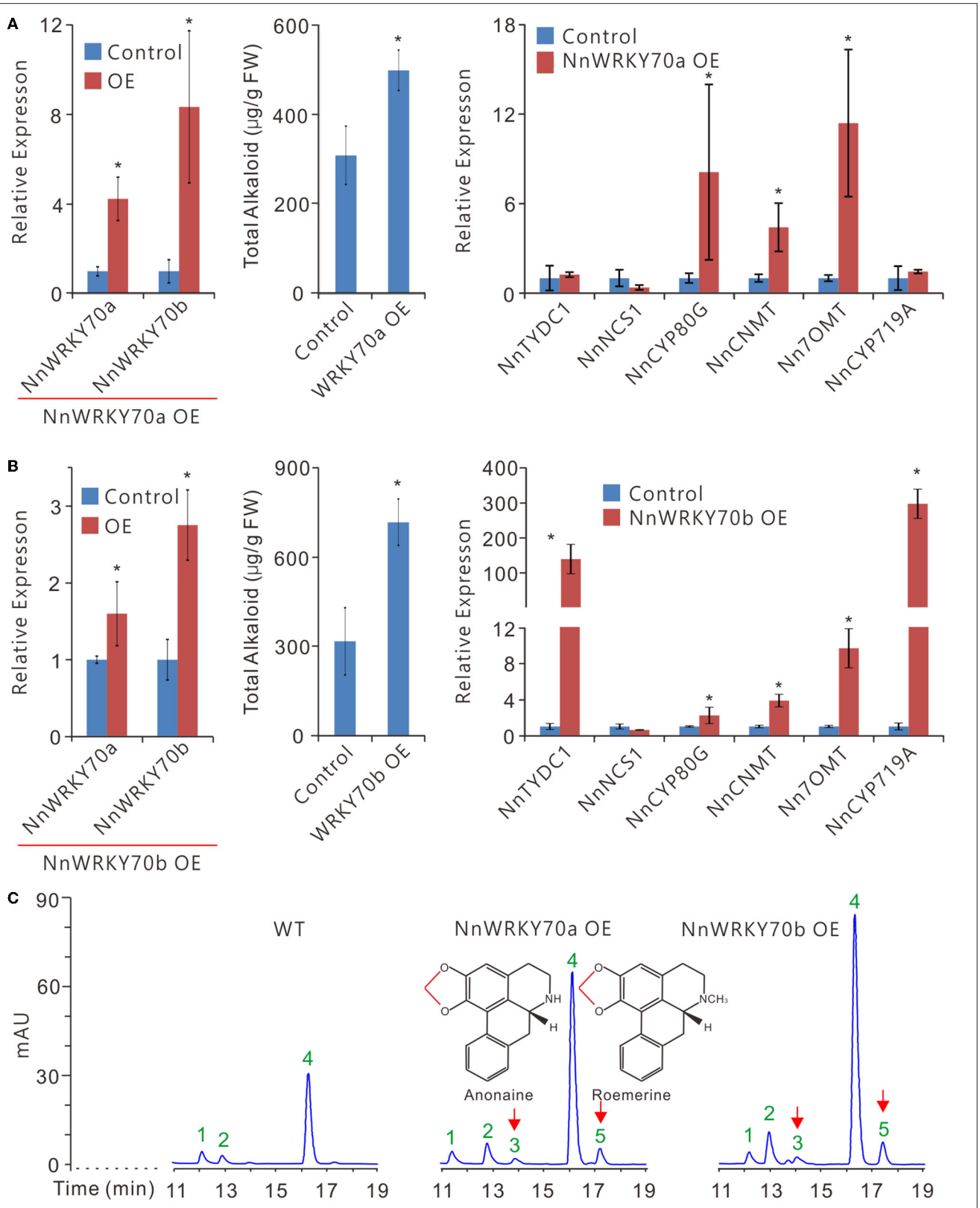


FIGURE 5 | Transient overexpression of *NnWRKY70a* and *NnWRKY70b* significantly enhances BIA accumulation in lotus petals. **(A,B)** Effects of NnWRKY70 overexpression on BIA biosynthesis in lotus. Left panel, expression of *NnWRKY70a* and *NnWRKY70b*; Middle panel, BIA concentration; Right panel, expression of BIA pathway genes in lotus petals overexpressing *NnWRKY70*s. **(C)** HPLC scanning of BIA constituents and their abundance in lotus petals. Data are means ± SE ($n = 3$). Asterisks (*) represents significant differences at $P < 0.05$. Peak signals 1-5 represent *N*-nornuciferine, *O*-nornuciferine, anonaine, nuciferine, and roemerine alkaloids, respectively. Chemical bonds marked in red are methylenedioxy bridges formed by the catalytic activity of NnCYP719A.

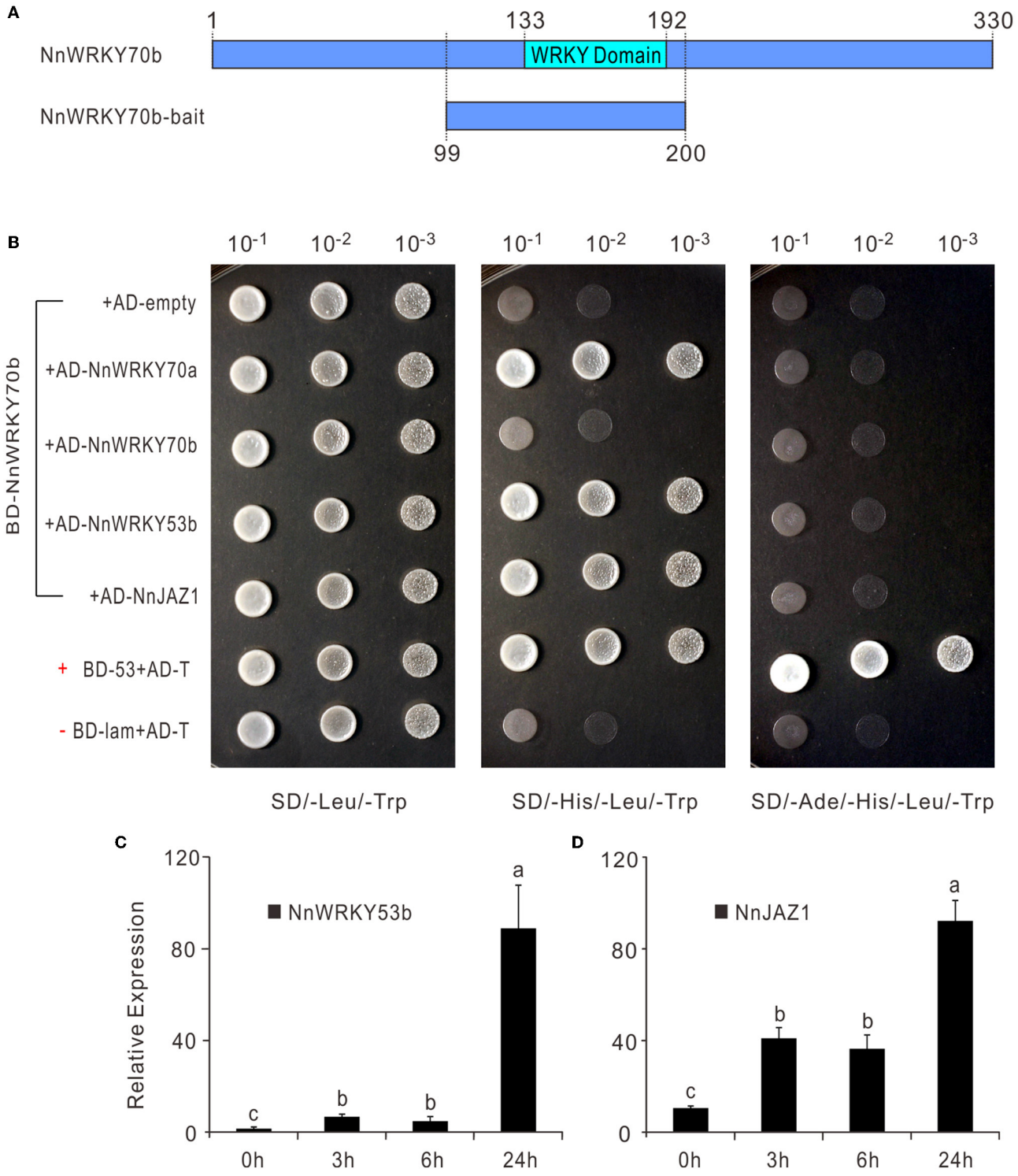


FIGURE 6 | NnWRKY70b regulates the lotus BIA biosynthesis by interacting with NnWRKY70a, NnWRKY53b, and NnJAZ1 proteins. **(A)** Schematic showing the truncated NnWRKY70b bait used in the yeast two-hybrid assay. **(B)** Yeast two-hybrid assay showing the interaction between NnWRKY70b and other proteins. AD and BD represent the two yeast vectors of pGADT7 and pGBKT7, respectively. Control vectors BD-53 and BD-lam carry the murine p53 and the lamin respectively, while AD-T carries SV40 larger T-antigen. p53 and T-antigen are known to interact in the yeast two-hybrid assay. Thus, a combination of BD-53+AD-T was used as a positive control, while the un-interactional BD-Lam+ AD-T was set as a negative control. **(C,D)** Expression of *NnWRKY53b* and *NnJAZ1* in response to MeJA treatment. MeJA treatment was conducted in lotus leaves at the developmental stage 4. Data are means ± SE ($n = 3$). Lower case letters represent significant differences at $P < 0.05$.

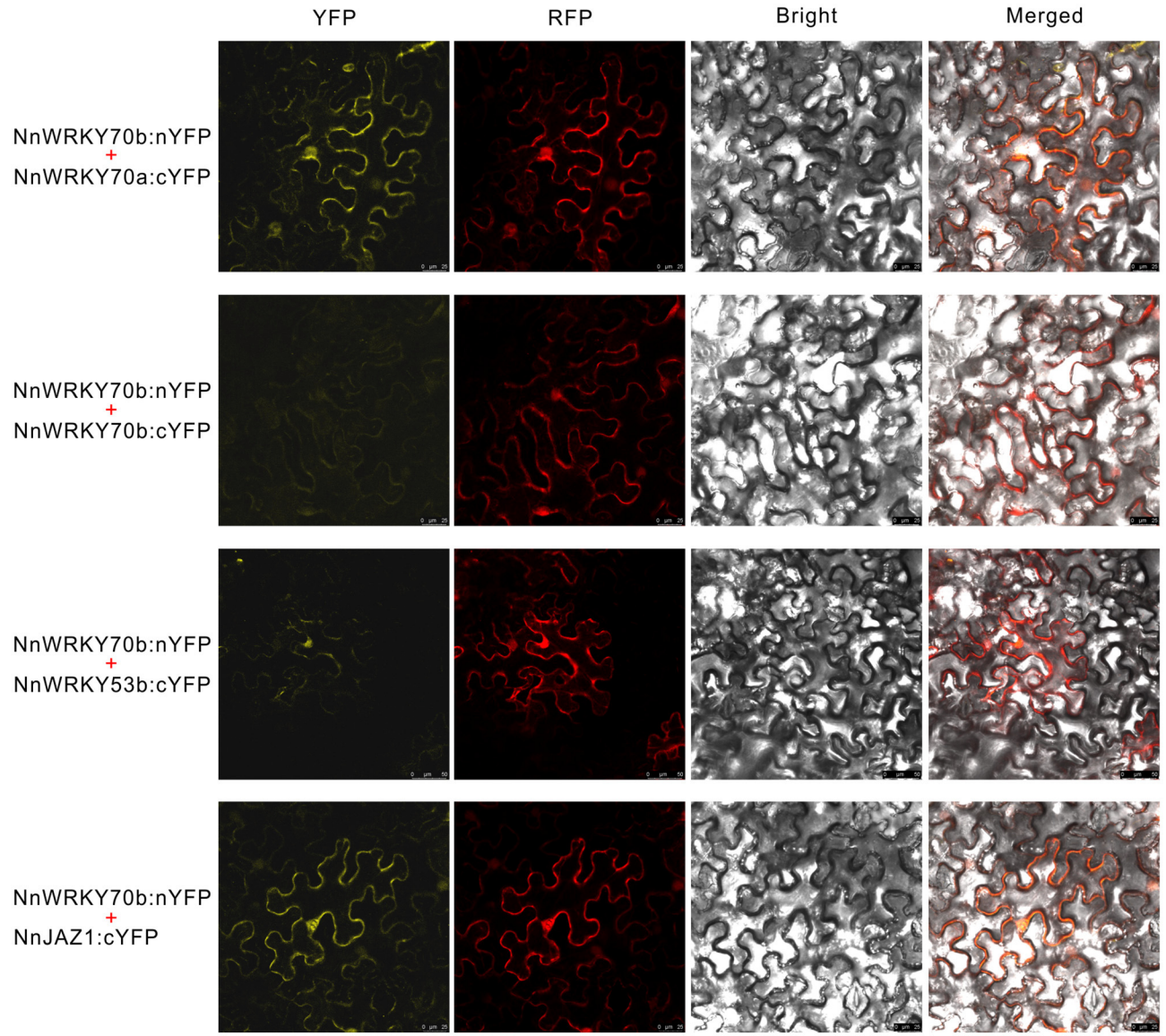


FIGURE 7 | Bimolecular Fluorescence Complementation (BiFC) assays showed that NnWRKY70b could interact with NnWRKY70a, NnWRKY53b, and NnJAZ1 *in vivo*. BiFC assays were conducted with a 2in1 system based on the splitting of enhanced yellow fluorescent protein (EYFP). Two fusion proteins were expressed with the same pBiFCt-2in1-CN vector, and a soluble monomeric red fluorescent protein (mRFP) was expressed in the same vector as an internal marker for transformation and expression control.

with AtWRKY70 and harboring a conserved WRKY DNA binding domain, with a WRKYGQK heptapeptide in the N-terminal and a C2HC type zinc-finger motif in the C-terminal (**Figure 2**). Two NnWRKY70s shared over 70% amino acid sequence identity and were highly syntenic, suggesting that they probably evolved from a common ancestor during a recent whole genome duplication event in lotus (Ming et al., 2013; Gui et al., 2018; Li et al., 2019).

Previous studies have demonstrated a strong correlation in their expression and the accumulation of BIA in lotus leaves (Deng et al., 2018; Meelaph et al., 2018). Here, we further showed that exogenous application of MeJA in the lotus leaves significantly induced their expression and the accumulation of BIA in the lotus leaves. The dual-luciferase assays showed that NnWRKY70a could activate the *NnTYDC2* promoter, while NnWRKY70b activated three BIA pathway gene promoters (**Figure 4**). Furthermore, the overexpression of the two WRKYs in lotus petals significantly elevated the BIA biosynthetic gene expression and the BIA accumulation. The yeast one-hybrid assays showed that NnWRKY70a and

NnWRKY70b can specifically bind to the W-box *cis*-elements of BIA pathway gene promoters. These observations suggest that both NnWRKY70a and NnWRKY70b transcriptionally activate the BIA structural genes and positively regulate the lotus BIA biosynthesis.

Notably, the two NnWRKY70 TFs are unevenly involved in the regulation of BIA biosynthesis in the lotus. A stronger activation of BIA biosynthesis was observed for NnWRKY70b than NnWRKY70a. For example, NnWRKY70b significantly activated the promoters of the three structural genes, including *NnTYDC*, *NnCYP80G*, and *Nn7OMT*, whereas NnWRKY70a could only activate the promoter of *NnTYDC*. In addition, more BIA structural genes were upregulated in the lotus petals overexpressing NnWRKY70b than in those overexpressing NnWRKY70a (**Figure 5**). In addition to the three genes (*NnCYP80G*, *NnCNMT*, and *Nn7OMT*) activated by the NnWRKY70a overexpression, the NnWRKY70b overexpression also strongly boosted the expression of *NnTYDC1* and *NnCYP719a*, with their expression increased over 100 folds. Moreover, approximately 120% increase in the accumulation of BIA was observed in the petals overexpressing NnWRKY70b, while about 60% increase was observed for the petals overexpressing NnWRKY70a. This is similar to the previous observations: two closely related group III WRKY proteins in *O. Pumila*, OpWRKY2 and OpWRKY3, showed varied regulatory effects on the biosynthesis of an anticancer drug camptothecin (Wang et al., 2019; Hao et al., 2021). It has been reported that WRKY70 in *Arabidopsis* acted as a convergent node for salicylic acid (SA) and JA-mediated defense signal pathways, and was intensively involved both in biotic- and abiotic-stress responses (Li et al., 2004, 2013, 2017). Thus, besides their role in regulating the BIA biosynthesis, NnWRKY70s in lotus could also potentially be associated with lotus's innate immunity.

Interestingly, *NnNCS1*, a crucial BIA biosynthetic gene in the lotus (Vimolmangkang et al., 2016), was not activated by either of the two NnWRKY70 TFs. This could be attributed to the lack of a W-box *cis*-element binding site in its promoter (Eulgem et al., 2000). W-box was identified within 2 kb promoter regions of the four BIA structural genes, including *NnTYPC*, *NnCYP80G*, *Nn7OMT*, and *NnCYP719A*, but not in the promoters of *NnNCS1* and *NnCNMT*. The yeast one-hybrid assays showed that NnWRKY70s can bind to the four tested BIA gene promoters that contain W-boxes (**Figure 4**). However, *NnCNMT*, despite the lacking of W-box in its promoter, was significantly induced by the overexpression of both NnWRKY70s. Thus, the two NnWRKY70 TFs may have binding preferences other than W-box; otherwise, they may have activated other TFs that can positively regulate *NnCNMT*. Previous studies have shown that some WRKY TFs had indeed altered the binding preference of WK (TTTCCAC) and WT (GGACTTTC) boxes (Machens et al., 2014; Kanofsky et al., 2017). Further studies are still needed to explore the exact binding site of NnWRKY70s. Intriguingly, NnWRKY70b overexpression significantly also enhanced the expression of *NnCYP719A*. As a result, the accumulation of roemerine and anonaine alkaloids was significantly elevated in lotus petals (**Figure 5**). This further verified the role of NnCYP719A in catalyzing the formation of methylenedioxy bridges in roemerine and anonaine biosynthesis (Ikezawa et al., 2003).

Both the yeast two-hybrid and BiFC assays further revealed possible interactions between NnWRKY70b and other proteins (**Figures 6**, **7**). NnWRKY70b did not physically interact with itself, but interacted with two group III WRKY TFs, NnWRKY70a and NnWRKY53b, suggesting that NnWRKY70b may activate BIA pathway genes through forming protein complexes. Notably, although NnWRKY53b activated two BIA pathway gene promoters, it did not physically interact with NnJAZ1. The combination of NnWRKY70a and NnWRKY53b did not show gained effects on the activation of BIA structural gene promoters. AtWRKY70 and AtWRKY53 are known to negatively regulate the leaf senescence and drought tolerance in *Arabidopsis* (Li et al., 2004; Miao and Zentgraf, 2007; Sun and Yu, 2015). In contrast to the JA-inducible characteristic of NnWRKY70s and NnWRKY53b, the homologous *Arabidopsis* genes were reported to be suppressed by the MeJA treatment. Thus, functions of WRKY70 and WRKY53 may have diverged in the two species.

In addition, NnWRKY70b also interacted with NnJAZ1. JAZ family proteins are the key regulators in the JA signaling pathway, which interact with and repress TFs that regulate plant secondary metabolism (Pauwels and Goossens, 2011; Nagels Durand et al., 2016). Bioactive JAs can be sensed by the F-box protein Coronatine Insensive1 (COI1), the recognition component of the E3 ubiquitin ligase complex SCF^{COI1}. As a result, JAZ proteins are ubiquitinated and degraded *via* the ubiquitin-26S proteasome pathway (Thines et al., 2007). Given the positive interaction between NnWRKY70b and NnJAZ1, it is reasonable to speculate that the JA-responsive NnWRKY70b also regulates the lotus BIA biosynthesis via the JA-signaling pathway. It is however should be noticed that the EYFP fluorescence raised from the interacted protein complex was not located in the nucleus, where the interaction was supposed to happen. This is probably due to the slow maturation property of EYFP and the irreversible BiFC complex, which resulted in the discrepancy between the bimolecular fluorescent and the site of protein interactions, as has been reviewed and pointed out previously (Miller et al., 2015).

Based on the above findings, we propose a summary model depicting the possible action mode of NnWRKY70 TFs in regulating the lotus BIA biosynthesis. In the absence of JAs, JAZ proteins bind to NnWRKY70b and suppress its activity (**Figure 8A**), whereas, in the presence of bioactive JAs, JAZ proteins are degraded *via* the SCF^{COI1}-mediated ubiquitin-26S proteasome pathway (**Figure 8B**). The unbound NnWRKY70b then forms protein complexes with other co-factors such as NnWRKY70a, and transactivate BIA biosynthetic genes by binding to the W-box *cis*-elements in their promoters. Transcription activators can usually activate the expression of a range of pathway genes, and massively increase the secondary metabolite production. Our present findings demonstrate the positive regulatory role of NnWRKY70 TFs in activating the biosynthesis of BIA in the lotus and providing a feasible strategy for improving the BIA production through TF-based genetic engineering.

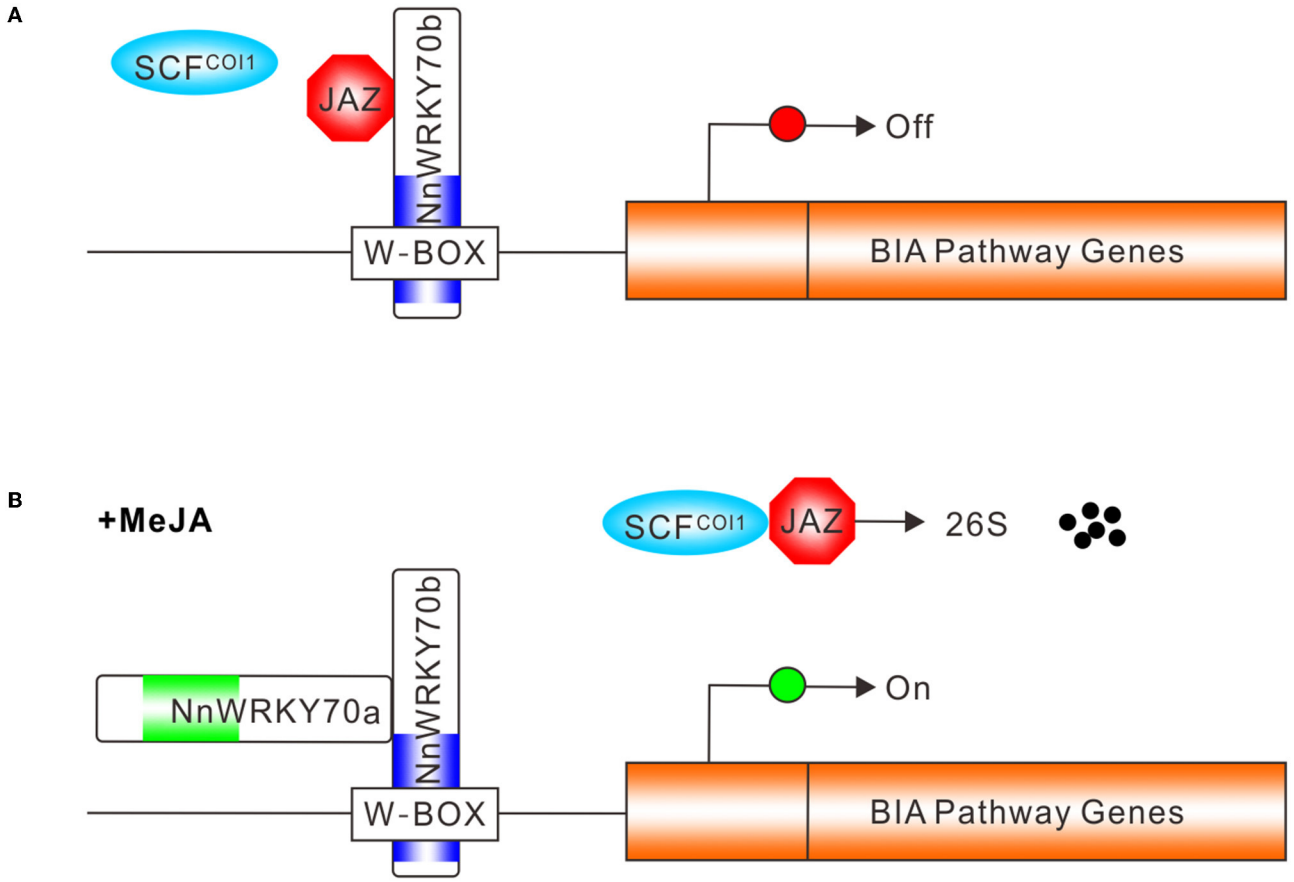


FIGURE 8 | Simplified model of NnWRKY70b TF action in regulating lotus BIA biosynthesis. **(A)** In the absence of JA, JAZ proteins bind to NnWRKY70b and repress its activity, resulting in the inactivation of BIA pathway genes in the lotus. **(B)** In the presence of JA, the SCF^{COI1} complex facilitates the ubiquitin-26S mediated JAZ degradation. The NnWRKY70b protein is then released and forms protein complexes with another group III WRKY proteins to positively regulate the BIA pathway genes, through binding the W-box *cis*-element in their promoters. Blue and green boxes represent WRKY DNA binding domains in NnWRKY70a and NnWRKY70b, respectively.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JL and XD conceived and designed the experiments. JL, YLi, XD, MD, SL, SC, and MZ performed the experiments. JL wrote the paper. XD, MY, DY, YLiu, and DT revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by funds received from the National Natural Science Foundation of China (Grant Nos. 31700262 and 32070336), the Fundamental Research Funds for the Central Universities (WUT: 2020IB031), the National Innovation and Entrepreneurship Training Program for College Students (Grant No. S202010497060), the Open Fund of Shanghai Key Laboratory of Plant Functional Genomics and Resources, and the Natural Science Foundation of Shandong Province (No. ZR2021MC163).

ACKNOWLEDGMENTS

We thank Wuhan GeneCreate Biological Engineering Co., Ltd., for their assistance in the yeast two-hybrid assays, and the Public Laboratory Platform of the Wuhan Botanical Garden for their excellent services.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.862915/full#supplementary-material

REFERENCES

Besseau, S., Li, J., and Palva, E. T. (2012). WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana. *J. Exp. Bot.* 63, 2667–2679. doi: 10.1093/jxb/err450

Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., et al. (2018). The WRKY transcription factor family in model plants and crops. *Crit. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103

Deng, X., Yang, D., Sun, H., Liu, J., Song, H., Xiong, Y., et al. (2022). Time-course analysis and transcriptomic identification of key response strategies to complete submergence in Nelumbo nucifera. *Hortic Res.* 9:uhac001. doi: 10.1093/hr/uhac001

Deng, X., Zhao, L., Fang, T., Xiong, Y., Ogutu, C., Yang, D., et al. (2018). Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. *Hortic. Res.* 5, 29. doi: 10.1038/s41438-018-0035-0

Deng, X., Zhu, L., Fang, T., Vimolmangkang, S., Yang, D., Ogutu, C., et al. (2016). Analysis of isoquinoline alkaloid composition and wound-induced variation in nelumbo using HPLC-MS/MS. *J. Agric. Food Chem.* 64, 1130–1136. doi: 10.1021/acs.jafc.5b06099

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5, 113. doi: 10.1186/1471-2105-5-113

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

Facchini, P. J. (2001). Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 29–66. doi: 10.1146/annurev.arplant.52.1.29

Facchini, P. J., and De Luca, V. (1994). Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. *J. Biol. Chem.* 269, 26684–26690. doi: 10.1016/S0021-9258(18)47073-1

Grefen, C., and Blatt, M. R. (2012). A 2in1 cloning system enables ratiometric bimolecular fluorescence complementation (rBiFC). *Biotechniques* 53, 311–314. doi: 10.2144/000113941

Gu, C., Wang, L., Zhang, L., Liu, Y., Yang, M., Yuan, Z., et al. (2013). Characterization of genes encoding granule-bound starch synthase in sacred lotus reveals phylogenetic affinity of nelumbo to proteales. *Plant Molec. Biol. Report.* 31, 1157–1165. doi: 10.1007/s11105-013-0605-0

Gui, S., Peng, J., Wang, X., Wu, Z., Cao, R., Salse, J., et al. (2018). Improving Nelumbo nucifera genome assemblies using high-resolution genetic maps and BioNano genome mapping reveals ancient chromosome rearrangements. *Plant J.* 94, 721–734. doi: 10.1111/tpj.13894

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

Hao, X., Xie, C., Ruan, Q., Zhang, X., Wu, C., Han, B., et al. (2021). The transcription factor OpWRKY2 positively regulates the biosynthesis of the anticancer drug camptothecin in Ophiorrhiza pumila. *Hortic Res.* 8, 7. doi: 10.1038/s41438-020-00437-3

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

Kanofsky, K., Bahlmann, A. K., Hehl, R., and Dong, D. X. (2017). Combinatorial requirement of W- and WT-boxes in microbe-associated molecular pattern-responsive synthetic promoters. *Plant Cell Rep.* 36, 971–986. doi: 10.1007/s00299-017-2130-3

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

Li, J., Besseau, S., Törönen, P., Sipari, N., Kollist, H., Holm, L., et al. (2013). Defense-related transcription factors WRKY70 and WRKY54 modulate osmotic stress tolerance by regulating stomatal aperture in Arabidopsis. *New Phytol.* 200, 457–472. doi: 10.1111/nph.12378

Li, J., Brader, G., and Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell* 16, 319–331. doi: 10.1105/tpc.016980

Li, J., Xiong, Y., Li, Y., Ye, S., Yin, Q., Gao, S., et al. (2019). Comprehensive analysis and functional studies of WRKY Transcription Factors in Nelumbo nucifera. *Int. J. Mol. Sci.* 20, 5006. doi: 10.3390/ijms20205006

Li, J., Zhong, R., and Palva, E. T. (2017). WRKY70 and its homolog WRKY54 negatively modulate the cell wall-associated defenses to necrotrophic pathogens in Arabidopsis. *PLoS ONE* 12, e0183731. doi: 10.1371/journal.pone.0183731

Lin, Z., Zhang, C., Cao, D., Damaris, R. N., and Yang, P. (2019). The latest studies on lotus (Nelumbo nucifera)-an emerging horticultural model plant. *Int. J. Mol. Sci.* 20, 3680. doi: 10.3390/ijms20153680

Machens, F., Becker, M., Umrath, F., and Hehl, R. (2014). Identification of a novel type of WRKY transcription factor binding site in elicitor-responsive cis-sequences from Arabidopsis thaliana. *Plant Mol. Biol.* 84, 371–385. doi: 10.1007/s11103-013-0136-y

Maldonado-Mendoza, I. E., López-Meyer, M., Galef, J. R., Burnett, R. J., and Nessler, C. L. (1996). Molecular analysis of a new member of the opium poppy tyrosine/3,4-dihydroxyphenylalanine decarboxylase gene family. *Plant Physiol.* 110, 43–49. doi: 10.1104/pp.110.1.43

Meelaph, T., Kobtrakul, K., Chansilpa, N. N., Han, Y., Rani, D., De-Eknamkul, W., et al. (2018). Coregulation of biosynthetic genes and transcription factors for aporphine-type alkaloid production in wounded lotus provides insight into the biosynthetic pathway of nuciferine. *ACS Omega.* 3, 8794–8802. doi: 10.1021/acsomega.8b00827

Menendez-Perdomo, I. M., and Facchini, P. J. (2020). Isolation and characterization of two O-methyltransferases involved in benzylisoquinoline alkaloid biosynthesis in sacred lotus (Nelumbo nucifera). *J. Biol. Chem.* 295, 1598–1612. doi: 10.1074/jbc.RA119.011547

Miao, Y., and Zentgraf, U. (2007). The antagonist function of Arabidopsis WRKY53 and ESR/ESP in leaf senescence is modulated by the jasmonic and salicylic acid equilibrium. *Plant Cell* 19, 819–830. doi: 10.1105/tpc.106.042705

Miller, K. E., Kim, Y., Huh, W. K., and Park, H. O. (2015). Bimolecular fluorescence complementation (BiFC) analysis: advances and recent applications for genome-wide interaction studies. *J. Mol. Biol.* 427, 2039–2055. doi: 10.1016/j.jmb.2015.03.005

Minami, H., Dubouzet, E., Iwasa, K., and Sato, F. (2007). Functional analysis of norcoclaurine synthase in Coptis japonica. *J. Biol. Chem.* 282, 6274–6282. doi: 10.1074/jbc.M608933200

Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L. T., et al. (2013). Genome of the long-living sacred lotus (Nelumbo nucifera Gaertn.). *Genome Biol.* 14, R41. doi: 10.1186/gb-2013-14-5-r41

Morishige, T., Tsujita, T., Yamada, Y., and Sato, F. (2000). Molecular characterization of the S-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase involved in isoquinoline alkaloid biosynthesis in Coptis japonica. *J. Biol. Chem.* 275, 23398–23405. doi: 10.1074/jbc.M002439200

Mukherjee, P. K., Mukherjee, D., Maji, A. K., Rai, S., and Heinrich, M. (2009). The sacred lotus (Nelumbo nucifera) - phytochemical and therapeutic profile. *J. Pharm. Pharmacol.* 61, 407–422. doi: 10.1211/jpp/61.04.0001

Nagels Durand, A., Pauwels, L., and Goossens, A. (2016). The ubiquitin system and jasmonate signaling. *Plants* 5, 6. doi: 10.3390/plants5010006

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta.* 1829, 1236–1247. doi: 10.1016/j.bbagrm.2013.09.006

Pauli, H. H., and Kutchan, T. M. (1998). Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3'-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. *Plant J.* 13, 793–801. doi: 10.1046/j.1365-313X.1998.00085.x

Pauwels, L., and Goossens, A. (2011). The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *Plant Cell.* 23, 3089–3100. doi: 10.1105/tpc.111.089300

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic. Acids Res.* 29, e45. doi: 10.1093/nar/29.9.e45

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

Samanani, N., Liscombe, D. K., and Facchini, P. J. (2004). Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* 40, 302–313. doi: 10.1111/j.1365-313X.2004.02210.x

Santner, A., and Estelle, M. (2007). The JAZ proteins link jasmonate perception with transcriptional changes. *Plant Cell.* 19, 3839–3842. doi: 10.1105/tpc.107.056960

Sato, F., Tsujita, T., Katagiri, Y., Yoshida, S., and Yamada, Y. (1994). Purification and characterization of S-adenosyl-L-methionine: norcoclaurine 6-O-methyltransferase from cultured Coptis japonica cells. *Eur. J. Biochem.* 225, 125–131. doi: 10.1111/j.1432-1033.1994.00125.x

Sharma, B. R., Gautam, L. N., Adhikari, D., and Karki, R. (2017). A comprehensive review on chemical profiling of nelumbo nucifera: potential for drug development. *Phytother. Res.* 31, 3–26. doi: 10.1002/ptr.5732

Song, S., Qi, T., Wasternack, C., and Xie, D. (2014). Jasmonate signaling and crosstalk with gibberellin and ethylene. *Curr. Opin. Plant Biol.* 21, 112–119. doi: 10.1016/j.pbi.2014.07.005

Sun, Y., and Yu, D. (2015). Activated expression of AtWRKY53 negatively regulates drought tolerance by mediating stomatal movement. *Plant Cell Rep.* 34, 1295–1306. doi: 10.1007/s00299-015-1787-8

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

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

Vimolmangkang, S., Deng, X., Owiti, A., Meelaph, T., Ogutu, C., and Han, Y. (2016). Evolutionary origin of the NCSI gene subfamily encoding norcoclaurine synthase is associated with the biosynthesis of benzylisoquinoline alkaloids in plants. *Sci. Rep.* 6, 26323. doi: 10.1038/srep 26323

Wang, C., Wu, C., Wang, Y., Xie, C., Shi, M., Nile, S., et al. (2019). Transcription factor OpWRKY3 is involved in the development and biosynthesis of camptothecin and its precursors in ophiorrhiza pumila hairy roots. *Int. J. Mol. Sci.* 20, 3996. doi: 10.3390/ijms20163996

Wasternack, C., and Strnad, M. (2019). Jasmonates are signals in the biosynthesis of secondary metabolites - Pathways, transcription factors and applied aspects - A brief review. *N Biotechnol.* 48, 1–11. doi: 10.1016/j.nbt.2017.09.007

Xu, M., Wu, C., Zhao, L., Wang, Y., Wang, C., Zhou, W., et al. (2020). WRKY transcription factor OpWRKY1 acts as a negative regulator of camptothecin biosynthesis in Ophiorrhiza pumila hairy roots. *Plant Cell, Tissue Organ. Cult.* 142, 69–78. doi: 10.1007/s11240-020-01833-2

Xu, X., Chen, C., Fan, B., and Chen, Z. (2006). Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell.* 18, 1310–1326. doi: 10.1105/tpc.105.037523

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. (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

Zhou, M., and Memelink, J. (2016). Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol. Adv.* 34, 441–449. doi: 10.1016/j.biotechadv.2016.02.004

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fpls.2022.914138

Functional Characterization of Genes Coding for Novel β-D-Glucosidases Involved in the Initial Step of Secoiridoid Glucosides Catabolism in *Centaurium erythraea* Rafn

Jelena Božunović, Milica Milutinović, Neda Aničić, Marijana Skorić, Dragana Matekalo, Suzana Živković, Milan Dragićević, Biljana Filipović, Tijana Banjanac, Luka Petrović and Danijela Mišić**

Department of Plant Physiology, Institute for Biological Research "Siniša Stanković"- National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia

OPEN ACCESS

Edited by:
Supaart Sirikantaramas, Chulalongkorn University, Thailand

Reviewed by:
Praveen Awasthi, Institute of Bioorganic Chemistry (PAS), Poland
Ramesha Thimmappa, Amity University, India

***Correspondence:**
Dragana Matekalo
dragana.bozic@ibiss.bg.ac.rs
Danijela Mišić
dmisic@ibiss.bg.ac.rs

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

Received: *06 April 2022*
Accepted: *23 May 2022*
Published: *23 June 2022*

Citation:
Božunović J, Milutinović M, Aničić N, Skorić M, Matekalo D, Živković S, Dragićević M, Filipović B, Banjanac T, Petrović L and Mišić D (2022) Functional Characterization of Genes Coding for Novel β-D-Glucosidases Involved in the Initial Step of Secoiridoid Glucosides Catabolism in Centaurium erythraea Rafn. Front. Plant Sci. 13:914138. doi: 10.3389/fpls.2022.914138

Secoiridoid glucosides (SGs) are monoterpenoids derived from the iridoid cyclopentane-C-pyran skeleton with β-D glucose linked at C1 position. Coordinated metabolic processes, such as biosynthesis and catabolism of SGs, ensure constitutive presence of these bitter tasting compounds in plant tissues, which plays a decisive role in the defense against pathogens and herbivores. These compounds are susceptible to hydrolysis mediated by enzymes β-glucosidases, and the resulting aglycones are subsequently directed toward different metabolic pathways in plants. Function of two β-D-glucosidases (named *Ce*BGlu1 and *Ce*BGlu2) from centaury (*Centaurium erythraea* Rafn; fam. Gentianaceae), belonging to the glycoside hydrolase 1 (GH1) family, was confirmed using *in vitro* assays with recombinant proteins, following their heterologous expression in *E. coli* and His-tag affinity purification. Although they show slightly differential substrate preference, both isoforms display high specificity toward SGs and the organ-specific distribution of transcripts was positively correlated with the content of SGs in diploid and tetraploid *C. erythraea* plants. Transient overexpression of *Ce*BGlu1 and *Ce*BGlu2 in *C. erythraea* leaves induced changes in metabolite profiles. The effectiveness of transgene overexpression has been altered by plant ploidy. UHPLC/DAD/(±)HESI – MS^2 profiling of leaves of diploid and tetraploid *C. erythraea* genotypes revealed that the amounts of major SGs; sweroside, swertiamarin, and gentiopicrin was decreased in agroinfiltrated leaves, especially when *Ce*BGlu1 and *Ce*BGlu2 were co-expressed with transgene silencing suppressor p19. The work demonstrates that *in planta* metabolic engineering adopting transient overexpression of *Ce*BGlu1 and *Ce*BGlu2 is a suitable tool for the modulation of SGs content and glucosides/aglycones ratio, which might have substantial effects on overall phytochemistry of *C. erythraea*.

Keywords: beta glucosidase, secoiridoid glucosides, functional characterization, agroinfiltration, *Centaurium erythraea*

INTRODUCTION

Secoiridoid glucosides (SGs) are a group of plant-derived natural compounds widely present in species belonging to orders Gentianales, Dipsacales, Cornales, and Lamiales (family Oleaceae) (Ghisalberti, 1998; Jensen et al., 2002). Species of the genus *Centaurium* Hill (fam. Gentianaceae) are a rich source of SGs, among which sweroside (SW), swertiamarin (SWM), and gentiopicrin (GP) predominate. Leaves are the major site of SGs biosynthesis and accumulation in *C. erythraea* (Matekalo et al., 2018). Secoiridoids are named after secologanin (SEC), derived after the cleavage of the cyclopentane ring of iridoids (between C-7 and C-8). Secologanin is considered the common precursor of all secoiridoids in plant sources (Jensen and Schripsema, 2002), but is also the building block for monoterpene indole alkaloids (MIAs). *Centaurium* species are presumed to share a part of the secoiridoid biosynthetic pathway up to SEC with MIAs-producing *Catharanthus roseus*. The biosynthetic pathway of SEC starts with the synthesis of geranyl diphosphate (GPP), which is converted to geraniol and continues via a number of intermediates such are 8-hydroxygeraniol, 8-oxogeraniol, nepetalactol, iridotrial, 7-deoxyloganetic acid, 7-deoxyloganic acid, loganic acid and loganin (**Figure 1A**). The biosynthetic pathway involves a series of reactions, catalyzed by enzymes geranyl diphosphate synthase (GPPS), geraniol-8-oxidase (G8O), 8-hidroxygenaniol oxidoreductase (8HGO), iridoid synthase (IS), iridoid oxidase (IO), 7-deoxyloganetic acid glucosyltransferase (7DLGT), 7-deoxyloganic acid hydrolase (7DLH), loganic acid O-methyltransferase (LAMT), and secologanin synthase (SLS). In *C. roseus*, SEC and tryptamine are precursors of monoterpenoid indole alkaloids (MIAs) such as strictosidine, while in *C. erythraea* SEC is metabolized into secoiridoid glucosides SW, SWM and GP, as well as their derivatives. Enzymes responsible for the conversion of SEC to SW, and further to SWM and GP are not elucidated yet (**Figure 1A**).

The content of SGs in *C. erythraea* tissues is a net result of the two dynamic metabolic processes, their biosynthesis and their catabolism (**Figure 1A**). Catabolism of SGs starts with their deglycosylation catalyzed by β-glucosidases. Cleavage of the glucose (Glu) residue from SW, SWM, and GP results in the formation of aglycones (**Figures 1A,B**), which are further metabolized in plant tissues through isomerization, reduction and oxidation reactions. Gentiopicral and erythrocentaurin are the common products of SWM and GP hydrolysis (Ishiguro et al., 1983; Zeng et al., 2013; Božunović et al., 2018), while SW is metabolized into naucledal and epinaucledal (Purdy and McLean, 1977; El-Sedawy et al., 1990) (**Figure 1C**). In order to better understand the catabolism of SGs in *C. erythraea*, we focused our attention on β-glucosidases (β-D-glucoside glucohydrolases, E.C. 3.2.1.21), categorized into the glycoside hydrolase family 1 (GH1), the largest GH family in plants. GH1 β-glucosidases are enzymes that hydrolyze glycosidic bonds to release non-reducing terminal glucosyl residues from various compounds – benzoxazinoid, cyanogenic, iridoid and phenolic glucosides – as well as glucosinolates (Morant et al., 2008a). Genes coding for β-glucosidase (β-Glu) enzymes have been previously characterized in some members of the order Gentianales, including *C. roseus* and *Rauvolfia serpentina* (Geerlings et al., 2000; Warzecha et al., 2000; Barleben et al., 2007). Isolation and heterologous expression of *C. roseus* β-glucosidase with high specificity toward strictosidine which directly derives from SEC, provided valuable information on the function of this β-glucosidase involved in metabolic pathway of indole alkaloids (Geerlings et al., 2000). The aim of the present study was to isolate and functionally characterize *C. erythraea* β-glucosidase with high substrate specificity toward SGs using *in vitro* and *in planta* assays. Our presumption was that the action of *C. erythraea* β-glucosidase with high specificity toward SGs is essential for the catabolism of these compounds, and is indirectly related to defense against herbivores and pathogens. Secoiridoids in the form of glucosides possess remarkable antimicrobial effects (Božunović et al., 2018), and their hydrolysis/deglycosylation mediated by β-D-glucosidases release biologically active aglycones which provide more efficient antioxidant protection (Božunović et al., 2018). Moreover, during hydrolysis of SGs, glucose is also released, which may serve as an alternative source of energy under stress conditions.

MATERIALS AND METHODS

Plant Material

Plants used in experiments *in vitro* were obtained as previously described by Filipović et al. (2019). Briefly, mother stock shoot cultures of *C. erythraea* diploids and tetraploids originating from seeds collected at the locality Tjentište (Sutjeska National Park, Bosnia and Herzegovina), in 2007 and 2016, respectively, were maintained *in vitro* on half–strength MS medium ($^1/_2$ MS, Murashige and Skoog, 1962) in 370 ml glass jars. Root cultures, established from root segments of 3-month-old diploid and tetraploid plants, were grown on solid 1/2 MS medium in Petri dishes. Spontaneously regenerated shoots, formed on root explants, were further transferred on fresh 1/2 MS medium in 370 ml glass jars for rooting. After 10 weeks of culturing, shoots and roots of the obtained regenerated plants were harvested and used in experiments to determine the expression patterns of *Ce*BGlu candidates and their activities. All *in vitro* cultures were maintained at a temperature of 25 ± 2°C under fluorescent light of 47 μmol s^{-1} m^2 and a 16 h/8 h light/dark photoperiod.

Agroinfiltration experiments were performed under greenhouse conditions. *C. erythraea* seedlings were established in greenhouse, in pots filled with Floradur B seed substrate for multiplication (Floragard Vertriebs-GmbH, Oldenburg, Germany). Two-month-old seedlings were individually transferred into pots with Floragard growth medium (Floragard Vertriebs-GmbH, Oldenburg, Germany), and grown under greenhouse conditions at 50–85% humidity. Five-month-old diploid and tetraploid plants, displaying rosette phenotype, were used in agroinfiltration experiments. Seeds of tetraploid *C. erythraea*, collected in 2006, were obtained from Ecological-Botanical Garden of the University of Bayreuth (Germany). Diploid individuals were of the same origin as those used in *in vitro* experiments. All the *C. erythraea* accessions used in the present study are deposited within the seed collection at the

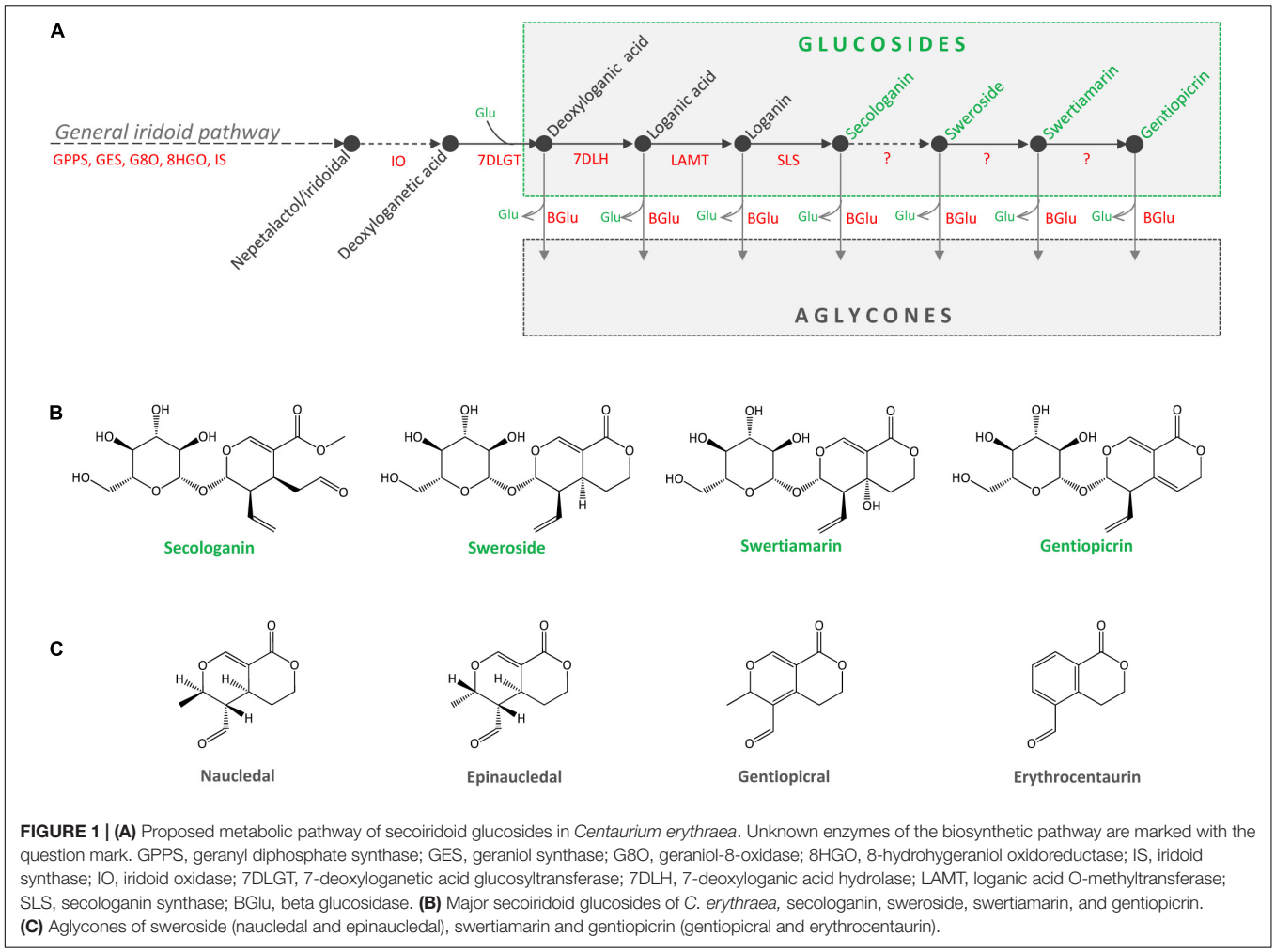


FIGURE 1 | (A) Proposed metabolic pathway of secoiridoid glucosides in *Centaurium erythraea*. Unknown enzymes of the biosynthetic pathway are marked with the question mark. GPPS, geranyl diphosphate synthase; GES, geraniol synthase; G8O, geraniol-8-oxidase; 8HGO, 8-hydrohygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glucosyltransferase; 7DLH, 7-deoxyloganic acid hydrolase; LAMT, loganic acid O-methyltransferase; SLS, secologanin synthase; BGlu, beta glucosidase. **(B)** Major secoiridoid glucosides of *C. erythraea,* secologanin, sweroside, swertiamarin, and gentiopicrin. **(C)** Aglycones of sweroside (naucledal and epinaucledal), swertiamarin and gentiopicrin (gentiopicral and erythrocentaurin).

Department of Plant Physiology, Institute for Biological Research "Siniša Stanković" – National Institute of the Republic of Serbia, University of Belgrade (Serbia).

RNA Isolation and cDNA Synthesis

Total RNA from approximately 150 mg of *C. erythraea* shoots and roots was isolated using modified CTAB method (Gasic et al., 2004). Isolated RNA was quantified using Qubit 3.0 Fluorometer (Thermo Fisher Scientific, United States), and its integrity was confirmed using gel electrophoresis. Prior to RT-PCR, isolated RNA was treated with DNAse I (Thermo Fisher Scientific, United States) to deplete contaminating genomic DNA. First strand cDNA was synthesized from 300 μg RNA using the RevertAid Premium First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, United States).

Isolation and Cloning of β-Glucosidase Candidate Genes

Centaurium erythraea leaf transcriptome databases (Malkov and Simonović, 2011; Ćuković et al., 2020) was searched for homologs of *Catharanthus roseus* strictosidine β-glucosidase (Geerlings et al., 2000). After selecting the most promising candidate, gene specific primers were designed using Primer3Plus software[1] in order to amplify the full length of the *Ce*BGlu coding sequence. A list of primers used for cloning and expression analysis is given in **Supplementary Table 1**.

The full length amplification of the candidate gene was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs, United States) and gene specific forward and reverse primers (**Supplementary Table 1**) following cycling conditions: one cycle of 98°C for 3 min, 35 cycles of 98°C for 30 s, 62°C for 2 min, and 72°C for 5 min followed by a final extension of 72°C for 10 min in a thermal cycler (Eppendorf, Austria). The amplicon was separated by 1% agarose gel electrophoresis and then purified by GeneJET Gel Extraction Kit (Thermo Fisher Scientific, United States) according to manufacturer's instructions. The purified amplicon was cloned into pTZ57R/T cloning vector using T/A PCR product cloning kit (Thermo Fisher Scientific, United States). In summary, 20 μl of purified PCR product was mixed with 3 μl of pTZ57R plasmid vector (600 ng), 6 μl of 5× ligation buffer, 1 μl of T4 DNA Ligase (5 U μl^{-1}), in 1.5 ml microtube. A total of 30 μl reaction mixture

[1]http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

was incubated for 1 h at room temperature to let the ligation reaction take place.

For transformation, 5 μl of the ligation product was added into 100 μl of Mach1 *E. coli* competent cells, incubated on ice for 20 min and then heat-shocked for 45 s in water bath at 42°C. The mixtures were immediately placed on ice and subsequently cultivated with 250 μl liquid Luria-Bertani (LB) media followed by 1 h incubation at 37°C. The transformed cells were cultivated on LB agar plate containing ampicillin (100 $\mu g\ ml^{-1}$), which was followed by overnight incubation at 37°C. Positive transformants were verified by colony PCR with *Ce*BGlu gene-specific primers (**Supplementary Table 1**). Cells harboring the recombinant plasmid were cultured in ampicillin containing liquid LB medium overnight at 37°C in a shaker incubator at 220 rpm. GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, United States) was used to purify plasmids from 2 ml of Mach1 *E. coli* overnight culture according to the manufacturer's instructions. Recombinant plasmids were confirmed by restriction digestion using *Xho*I and *Kpn*I restriction enzymes (Thermo Fisher Scientific, United States). Subsequent sequencing has confirmed the isolation of two highly similar gene variants named *CeBGlu1* and *CeBGlu2*.

Organ-Specific Profiling of *CeBGlu* Expression

For qRT-PCR, due to the very high similarity between *CeBGlu1* and *CeBGlu2* sequences, it was possible to design only one pair of primers common to both isoforms (**Supplementary Table 1**). For *CeBGlu* expression analysis, SYBR Green I (Maxima SYBR Green/ROX Kit, Thermo Scientific, United States) was used. Amplification was conducted in Light cycler QuantStudio 3 (Thermo Fisher Scientific, United States), according to the manufacturer's instructions. General thermocycler conditions were 95°C for 10 min; 40 cycles of 95°C for 15 s; 60°C for 30 s; 72°C for 30 s and final extension at 72°C for 10 min. Expression levels of targeted genes were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). *EF1* gene expression was used as endogenous control to normalize all data (primer sequences are presented within **Supplementary Table 1**). Presented results are obtained from three biological replicates.

Organ-Specific Profiling of Total Hydrolytic Activity Against Secoiridoid Glycosides

Shoots and roots of diploid and tetraploid *C. erythraea* plants were ground to a fine powder using liquid nitrogen, and proteins were isolated in 100 mM potassium-phosphate buffer (pH 6.5) supplemented with phenylmethylsulfonyl fluoride (PMSF) and 5 mM ascorbate. Protein content was determined according to Bradford (1976) using bovine serum albumin as a standard.

Activity of protein extracts to reduce the content of secoiridoid glycosides was evaluated using standards of SW, SWM and GP as substrates. The reduction in quantity of the mentioned substrates is the net result of combined enzymatic activities present in the extract including glucosyl hydrolases which produce the respective aglycones and unknown biosynthetic enzymes involved in transformations of secoiridoids (**Figure 1A**).

All reference compounds were diluted in dH2O (w:v = 1:1) and kept as a stock solutions. Aliquots of 30 μl were vacuum evaporated and diluted in 300 μl 50 mM citrate phosphate buffer pH 5.5 containing 10 μg of protein extracts. Reactions were incubated for 24 h at 37°C. Subsequently, 700 μl of methanol was added, and samples were centrifuged at 10,000 *g* for 10 min. Supernatants were filtered through 0.2 μm cellulose filters and subsequently analyzed for the content of SW, SWM and GP.

Phylogenetic Analysis

For phylogenetic tree creation, multiple sequence alignments were generated using the Muscle algorithm. A neighbor-joining tree was constructed using MEGA X, Version 10.2.6 (Kumar et al., 2018). Cluster stability was estimated with 1,000 bootstrap replicates. The evolutionary distances were computed using the Poisson correction method. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 766 positions in the final dataset. The data were converted into Newick format and transferred to Dendroscope (Huson and Scornavacca, 2012) for creating the final phylogenetic tree. Refer to **Supplementary Table 2** for a list of plant protein sequences used for phylogenetic analysis and their corresponding accession numbers, including those of *Ce*BGlu1 and *Ce*BGlu2 functionally characterized within the present study.

Sequence Analysis, 3D Modeling and Ligand Docking

Multiple sequence alignments were generated with the DECIPHER R package (Wright, 2015) using 5 iterations and 5 refinements. Subcellular localization based on primary sequence was estimated using Light attention (Stärk et al., 2021).

Tertiary protein structure was estimated using AlphaFold2.1 (Jumper et al., 2021) queried via UCSF ChimeraX 1.4rc (Pettersen et al., 2021). The generated structures were assessed using MolProbity 4.4 (Williams et al., 2018) via SWISS-MODEL Workspace (Waterhouse et al., 2018). Obtained models were compared to each other and with experimental PDB structures using pairwise structure alignment[2] and the jFATCAT rigid model (Li et al., 2020), as well as using ChimeraX after superposition using the matchmaker command with default parameters (Pettersen et al., 2021). Protonation state for the predicted structures was estimated using PlayMolecule ProteinPrepare (Martínez-Rosell et al., 2017). Ligand docking was performed using AlphaFold predicted protein structure of *Ce*BGlu1. Ligand preparation was performed starting from tridimensional sdf files obtained from PubChem for SWM (PubChem id: 442435), SW (PubChem id: 161036) and GP (PubChem id: 88708). Conformers for these compounds were generated using the ETKDG version 3 method with imposing small ring torsion angle preferences (Wang et al., 2020) using RDKit 2021.09.5 (Landrum et al., 2022). A 100 conformers were generated per compound and subsequently filtered using

[2]https://www.rcsb.org/alignment

an RMSD threshold of 0.5 Å so that only those conformations that are at least 0.5 Å RMSD away from all retained are kept. The geometry of the resulting conformers was optimized using Merck molecular force field (MMFF94s) as implemented in RDKit 2021.09.5 (Landrum et al., 2022). The resulting ligand conformations were docked into the predicted enzyme structure using AutoDock Vina 1.2.3 (Eberhardt et al., 2021). The docking box was defined based on the coordinates of the malo secologanin ligand in the experimental structure of the raucaffricine β-D glucosidase from *Rauvolfia serpentina* (PDB: 3U5Y, Xia et al., 2012) after superposition of the structure to the AlphaFold model of *Ce*BGlu1. The docking box center was defined as $x = -1.964$, $y = 0.374$, $z = -6.148$, while the size of the box was 18, 18.75, and 21 Å in each direction, respectively. The docking was performed using autogrid4 (Morris et al., 2009) precalculated affinity maps, with autodock4 scoring function. Flexible docking was performed where the Glu476 side chain was allowed to change conformations. The exhaustiveness of the algorithm was set to 64. The highest estimated affinity poses, as reported by AutoDock Vina, were inspected and compared to the position of secologanin in 3U5Y, and the best pose based on these two criteria for each ligand was analyzed using UCSF ChimeraX 1.4rc (Pettersen et al., 2021) and protein–ligand interaction profiler (PLIP, Adasme et al., 2021).

Accession Numbers

Sequence data from this article can be found in the NCBI database under the following accession numbers: ON060690 (*Ce*BGlu1); ON060691 (*Ce*BGlu2).

Heterologous Expression and Purification of 6His-Tagged Proteins

For heterologous expression, *CeBGlu1* and *CeBGlu2* were cloned into bacterial expression vector pRSETA, yielding the final plasmids pRSETA:*CeBGlu1* and pRSETA:*CeBGlu2* (**Figure 2A**). Amplification of the gene sequences by PCR was performed using Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, United States) and a pair of *Ce*BGlu1/*Ce*BGlu2 gene specific primers with added *Xho*I/*Kpn*I restriction sites (**Supplementary Table 1**). The constructs were heat-shock transformed, as described above, into Mach1 *E. coli* competent cells which were further cultured on LB medium containing ampicillin. Polymerase chain reaction using gene specific primers, *Xho*I/*Kpn*I double restriction digestion and sequencing were applied to verify the constructs containing transgenes of interest.

pRSETA vector contains a N-terminal polyhistidine (6×His) tag, which can be used as an affinity ligand for protein purification. The pRSETA:*Ce*BGlu1 and pRSETA:*Ce*BGlu2 recombinant constructs were used for heat-shock transformation of BL21 (DE3) CodonPlus-RIL cells (Stratagene, United States). Cultures were grown overnight in 20 ml 2 × Yeast Extract Tryptone (2 × YT) media broth supplemented with ampicillin (100 μg ml^{-1}), 50 μg ml^{-1} kanamycin and 17 μg ml^{-1} chloramphenicol at 37°C with shaking at 220 rpm. On the next day 200 ml culture was initiated by inoculating above mentioned overnight cultures, with the initial OD600 set to 0.1. Cultures were incubated at 37°C with shaking at 230 rpm. After reaching an OD600 of 0.4–0.5, protein expression was induced using 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Subsequently, incubation continued for 4 h at 21°C. Following centrifugation for 10 min at 5,000 *g* and removal of supernatant, the cell pellets were harvested. All of the purification steps were performed at 4°C. For each purification batch, Ni-NTA agarose (Qiagen, United States) beads were equilibrated with equal volumes of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole pH 8.0). Bacterial cells were collected by centrifugation and re-suspended in 500 μl of lysis buffer. Following lysozyme (Sigma Aldrich, Germany) addition (1 mg ml^{-1} final concentration), samples were incubated for 30 min on ice. After 10 min of centrifugation (12,000 *g* at 4°C) supernatants were loaded into columns containing Ni–NTA resin and incubated at 4°C using FALC F205 rotary tube mixer (Falc Instruments, Treviglio, Italy). After 1 h, samples were spun at 12,000 g for 5 min at 4°C. Supernatant was discarded and resin was rinsed for 3 times using washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 50 mM imidazole, pH 8.0). The recombinant protein was eluted with increased imidazole concentration in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 10.2).

Eluted fractions were analyzed using 5–10% SDS-PAGE using Mini-PROTEAN II Electrophoresis Cell (Bio-Rad, United States) followed by Coomassie blue staining and immunoblotting. Proteins were electro-transferred to PVDF membrane (Amersham Biosciences, Germany), using Mini Trans-Blot Module electric transfer system (Bio-Rad, United States). Transfer on the membrane was performed at 4°C for 90 min at a constant voltage of 60 V. Membrane was blocked with 10% (w/v) non-fat dry milk (NFDM; Nestle, United States) in phosphate-buffered saline containing 0.05% Tween-20 overnight at 4°C. The presence of 6×His labeled proteins was confirmed using His-probe antibody in 1:100 dilution (Acc. No. sc-53073, Santa Cruz Biotechnology, United States), which was followed by incubation with goat anti-mouse IgG-HPR (1:5,000, Agrisera Antibodies, Sweden). The bound antibodies were visualized by enhanced chemiluminescence (ECL). The membrane was incubated for 5 min at room temperature using ECL solution containing 100 mM Tris–HCl pH 8.5, 0.2 mM *p*-coumaric acid, 1.25 mM 3-aminophthalhydrazide and 1.7 μl of 30% H_2O_2. Detection was performed by exposure to radiographic film (Medical X-ray Green/MXG Film, Carestream Health, United States) for 10 min.

Confirmation of *Ce*BGlu1 or *Ce*BGlu2 Hydrolytic Activity Using *in vitro* Enzymatic Assays

Beta-glucosidase activity of *Ce*BGlu1 and *Ce*BGlu2 was determined in a reaction with 20 μg of recombinant proteins and 10 mM 4-nitrophenyl β-D-glucopyranoside (*p*NPG) as a substrate, in 50 mM citric phosphate buffer pH 5.5 at 40°C for 48 h. The reaction was stopped by the addition of ice cold 1 M Na_2CO_3 (1:1 = v:v) and colorimetric detection of *p*-nitrophenol,

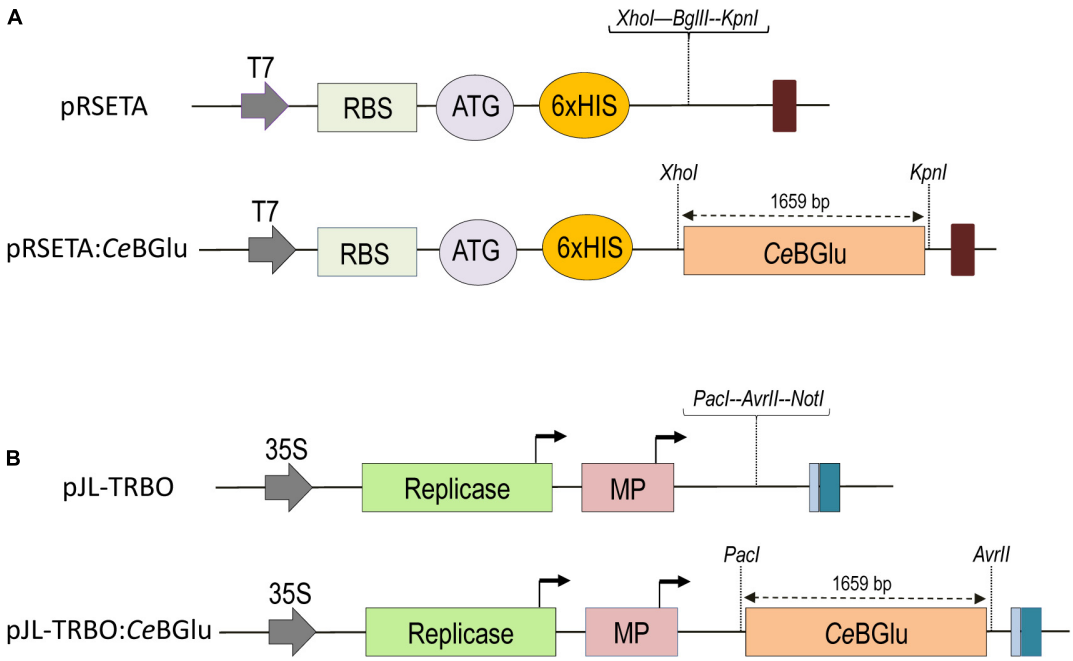


FIGURE 2 | Maps of vectors used in this study, pRSETA for the expression in *E.coli* **(A)** and pJL-TRBO for overexpression in *C. erythraea* **(B)**. CeBGlu, *β*-D-glucosidase from *C. erythraea*, was inserted as a XhoI-KpnI fragment into pRSETA expression vector **(A)**, or as PacI-AvrII fragment into pJL-TRBO **(B)** to produce pRSETA:CeBGlu and pJL-TRBO:CeBGlu constructs, respectively. The T-DNA regions of binary plasmids are presented. Block arrow, CaMV duplicated 35S promoter and T7 promoter. Dark blue box, CaMV polyA signal sequence/terminator. Light blue box, Ribozyme. Bent arrows, Subgenomic promoters. ORFs are represented by colored boxes and their identities are labeled in boxes. Replicase, TMV 126K/183K ORF; MP, movement protein; RBS, ribosome binding site; Colored oval boxes present start codon (ATG) and polyhistidine-tag sequence (6xHIS).

the product of *p*NPG hydrolysis, was spectrophotometrically measured at 410 nm (Esen and Blanchard, 2000).

Aerial parts of *C. erythraea* (100 mg) were ground into homogeneous powder using liquid nitrogen and extracted with 1 ml of 99.8% methanol (AppliChem GmbH, Germany) by vortexing for 30 s and subsequent sonication for 10 min using an ultrasonic bath (RK100, Bandelin, Berlin, Germany). After centrifugation at 10,000 *g* for 10 min supernatants were filtered using 0.2 μm syringe filters (Agilent Technologies, Santa Clara, CA, United States). *C. erythrae*a methanol extract (10 μl) was evaporated in a Vacuum Rotor Evaporator (Eppendorf Concentrator 5301, Germany) at room temperature. Subsequently, dried extracts were dissolved in 300 μl 50 mM citrate phosphate buffer (pH = 5.5) containing 20 μg of purified recombinant enzyme (*Ce*BGlu1 or *Ce*BGlu2) and incubated for 48 h at 37°C. Following incubation, 700 μl of methanol was added to stop the reaction, and reaction mixtures were centrifuged for 10 min at 10,000 *g*. Supernatants were filtered through 0.2 μm cellulose filters (Agilent Technologies, United States) and subsequently subjected to UHPLC/DAD/(±)HESI−MS^2 quantification of SGs. Control samples were prepared by replacing recombinant protein with the elution buffer used for protein purification.

Hydrolytic activity of *Ce*BGlu1 and *Ce*BGlu2 was further tested using standards of epideoxyloganic acid, loganin, secologanin, sweroside, swertiamarin, gentiopicrin, apigetrin, isoquercitrin, and vitexin as substrates. All reference compounds were diluted in dH_2O (w:v = 1:1) and kept as a stock solutions. Aliquots of 30 μl were dried in vacuum evaporator and diluted in 300 μl 50 mM citrate phosphate buffer pH = 5.5. Reaction mixture containing 30 μl of previously diluted standard compound and 20 μg of purified recombinant enzyme in a final volume of 300 μl 50 mM citrate phosphate buffer pH = 5.5 was incubated for 48 h at 37°C. Subsequently, 700 μl of methanol was added, and samples were centrifuged at 10,000 *g* for 10 min. Control samples contained elution buffer instead of purified recombinant protein, and the final concentration of standards was 3 μg ml^{-1}. Supernatants were filtered through 0.2 μm cellulose filters and injected into UHPLC/DAD/(±)HESI−MS^2 instrument.

Construction of *Ce*BGlu Expression Plasmids and Agrobacterium-Mediated Transformation

CeBGlu1 and *CeBGlu2* sequences were PCR-amplified from pTZ57R/T plasmids with the addition of *Pac*I/*Avr*II restriction sites. Amplicons were ligated to transient expression vector pJL-TRBO (Lindbo, 2007), and positive colonies were identified by colony PCR using vector- and gene-specific primers. All PCR reactions were performed with Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, United States) and a pair of *Ce*BGlu1/*Ce*BGlu2 gene specific primers (**Supplementary Table 1**). After heat-shock transformation

with pJL-TRBO:*Ce*BGlu1 and pJL-TRBO:*Ce*BGlu2 (**Figure 2B**), Mach1 *E. coli* competent cells were cultured overnight in LB liquid medium containing kanamycin (50 μg ml^{-1}) at 37°C in a shaker incubator at 230 rpm (IKA KS 4000 ic control, China). GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, United States) was used to purify plasmids from 5 ml of Mach1 *E. coli* cultures according to the manufacturer's instructions. After plasmid isolation, each *CeBGlu* sequence was confirmed by sequencing.

The recombinant pJL-TRBO:*Ce*BGlu1 and pJL-TRBO:*Ce*BGlu2 plasmids extracted from Mach1 cells were transferred to *A. tumefaciens GV3101* strain by electroporation using "Gene Pulser" (Bio-Rad, United States) and subsequently cultured in LB plates containing kanamycin (50 μg ml^{-1}), gentamicine (25 μg ml^{-1}) and rifampicin (10 μg ml^{-1}). A single colony of recombinant bacteria was inoculated into 5 ml liquid LB media containing antibiotics and incubated overnight at 28°C with shaking at 230 rpm. On the next day, 1 ml of bacterial suspension was sub-cultivated in 10 ml of liquid LB media containing 10 mM MES-KOH (pH 5.5) and 20 μM acetosyringone. Cultures were incubated overnight at 28°C with shaking at 230 rpm. The agrobacterial cells were harvested by centrifugation for 20 min at 3,000 *g*, and the pellet was resuspended in the infiltration medium (10 mM MES, 10 mM $MgCl_2$ and 100 μM acetosyringone) to a final OD600 of 1.0. After an incubation at room temperature for 4 h, cultures were introduced into abaxial surface of leaves of five-month-old *C. erythraea* plantlets using a blunt tipped plastic syringe and applying gentle pressure. Additionally, to prevent the silencing of transgene expression in *C. erythraea*, pJL-TRBO:*Ce*BGlu1 and pJL-TRBO:*Ce*BGlu2 were co-infiltrated in a 1:1 ratio with pBIN vector expressing the p19 silencing-suppressor gene from TBSV (pBIN:p19). After agroinfiltration, plants continued to grow for 5 days in the greenhouse. Leaves were harvested from diploid and tetraploid *C. erythraea* plants and immediately frozen at liquid nitrogen. Samples were stored at −80°C until further use.

Plant Methanol Extract Preparation

Plant material (shoots and roots) of *C. erythraea* was manually ground in liquid nitrogen into fine powder and diluted in 96% methanol (w:v = 10:1). Following vortexing for 1 min, extraction was performed overnight at 4°C. The next day, extraction was continued for 20 min in an ultrasonic bath (RK100, Bandelin, Germany) maintained at room temperature. Samples were centrifuged at 8,000 *g* for 20 min and supernatants were filtered using 15 mm RC filters with 0.22 μm pore size (Agilent Technologies, United States). Samples were stored at 4°C until use. All extractions were performed in biological triplicates.

UHPLC/DAD/(±)HESI-MS^2 Quantification of Targeted β-D Glucosides

Samples were analyzed using Dionex Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Germany) equipped with a DAD detector and connected to a triple quadrupole mass spectrometer (TSQ Quantum Access MAX, Thermo Fisher Scientific, Switzerland). Samples were chromatographically separated on a Hypersil gold C18 column (50 × 2.1 mm) with 1.9 μm particle size (Thermo Fisher Scientific, United States) thermostated at 40°C. Mobile phase, consisting of water + 0.01% acetic acid (A) and MS grade acetonitrile (B), was eluted at flow rate of 0.4 ml min^{-1} according to Banjanac et al. (2017). Injection volume was set to 10 μl. DAD absorption was acquired at λmax = 260 and 320 nm. A triple quadrupole mass spectrometer with a heated electrospray ionization (HESI) was operated with a following parameters: vaporizer temperature 300°C, spray voltage 4,000 V, sheet gas (N_2) pressure 27 AU, ion sweep gas (N_2) pressure 1.0 AU and auxiliary gas (N_2) pressure at 10 AU, capillary temperature 275°C, skimmer offset 0 V. Argon was used as the collision gas in the collision-induced fragmentation of the selected reaction monitoring (SRM) mode of the instrument, and collision energies (cE) were set as shown in **Supplementary Table 3**. Calibration curves of targeted compounds showed excellent linearity with correlation coefficients $r = 0.999$, $p < 0.001$. Total concentrations of targeted secoiridoids were obtained by calculating their peak areas, and were expressed as μg per 100 mg of plant fresh weight (μg 100 mg^{-1} FW). Xcalibur software (version 2.2) was used for the instrument control, data acquisition and analysis. All analyses were performed using three biological replicates.

Statistical Analysis

Statistical significance was determined by using Minitab Statistical Software (Minitab, State College, PA, United States). For statistical analysis of relative gene expression and compound quantification in shoots and roots of diploid and tetraploid plants one-way ANOVA was performed followed by Fisher LSD test ($p < 0.05$). For comparison of hydrolytic activities, as well as for comparing SGs content in leaves after agroinfiltration with *Ce*BGlu1 and *Ce*BGlu2 cloned into pJL-TRBO expression vector (without or with silencing inhibitor p19) Student's *t*-tests were used for data analysis ($p < 0.05$).

RESULTS AND DISCUSSION

Specific hydrolytic enzymes activate many glycosylated compounds in plants (e.g., glucosinolates, alkaloids, benzoxazinoids, cyanogenic, and (seco)iridoid glucosides), which often defines their dual-defense system against herbivores. Highly active and unstable aglycones released from iridoid and secoiridoid glycosides and monoterpenoid indole alkaloids usually display more prominent biological activities toward herbivores and pathogens, by adversely affecting their enzymatic machinery. Aglycones react with nucleophilic side chains of amino acids to form covalent protein complexes (Bartholomaeus and Ahokas, 1995; Konno et al., 1999; Kim et al., 2000; Guirimand et al., 2010), and act as unspecific enzyme inhibitors (Bartholomaeus and Ahokas, 1995; Ling et al., 2003). The typical defense compounds of Plantaginaceae are iridoid glycosides, which retard growth and/or enhance mortality of non-adapted herbivores. As a part of the dual defense system, *Plantago lanceolata* and *P. major* possess β-glucosidases that hydrolyze aucubin, one of the two major iridoid glycosides in these species,

and thereby release protein-denaturing aglycones (Pankoke et al., 2013, 2015). Oleaceae species like *Ligustrum obtusifolium* and *Olea europaea* are rich sources of SGs that after tissue disruption are metabolized by endogenous plant β-glucosidases (Konno et al., 1999; Mazzuca et al., 2006). Oleuropein β-glucosidase (OeGlu) from *O. europea* is reported to have a defensive role in young organs and meristem tissues as well as in mature tissues, where it activates oleuropein into a potent protein cross-linking agent during de-compartmentalization caused by pests (Koudounas et al., 2015). Oleuropein β-glucosidase (OeGlu) is highlighted as a molecular target of high biotechnology interest to regulate qualitative and quantitative content of bioactive secoiridoids in olive oils, and thus their organoleptic properties (Koudounas et al., 2021). Likewise, in MIAs-rich *C. roseus* and *R. serpentina* (fam. Apocynaceae), strictosidine is hydrolyzed by β-glucosidases (Guirimand et al., 2010). Following the disruption of strictosidine and strictosidine β-D-glucosidase (SGD) compartmentalization in *C. roseus*, initiated by cellular breakup after tissue wounding, the highly reactive strictosidine aglycone with prominent feeding-deterrent/toxic properties is released, and further conducted toward the production of cytotoxic MIAs (Guirimand et al., 2010, 2011).

Secoiridoid glucosides of *C. erythraea* are defense compounds constitutively present in tissues that, due to their extremely bitter taste, pre-ingestively deter feeding and thus reduce consumption rates of various non-adapted herbivores. The presumption of the present study was that *C. erythraea* might also possess β-glucosidases displaying high specificity toward SGs. By activating aglycones of SEC, SW, SWM, and GP, these hydrolytic enzymes could be involved in initial steps of their catabolism. Furthermore, resulting aglycones might offer multi-level protective features to plants, and, as previously shown, shape their antioxidant properties (Božunović et al., 2018).

Isolation of Full Length of *Ce*BGlu, Comparison With Homologs, and Phylogenetic Analysis

By exploring transcriptomic resources of *C. erythraea* leaves, we identified a candidate for β-D- glucosidase (*Ce*BGlu) displaying high homology with strictosidine β-D-glucosidase previously characterized from *C. roseus* (Geerlings et al., 2000). Following PCR, cloning of amplified product into *pTZ57R/T* vector and sequencing, two highly similar *Ce*BGlu candidates (*Ce*BGlu1 and *Ce*BGlu2) were revealed. Full lengths of isolated *CeBGlu1* and *CeBGlu2* comprised open reading frames of 1,659 bp encoding polypeptides of 552 amino acids with calculated molecular masses of 62.06 kDa. The two *Ce*BGlu gene candidates shared a high level of similarity (99.3%) and differed at 12 nucleotide sites, i.e., 4 amino acids (**Figure 3** and **Supplementary Table 4**). Based on primary structure both enzymes were predicted to be localized in the lysosome/vacuole.

Amino acid sequences of *Ce*BGlu1 and *Ce*BGlu2 were aligned with previously investigated β-D-glucosidase amino acid sequences of plant species belonging to the Gentianales order (**Figure 3**): strictosidine glucosidase from *Gelsemium sempervirens* (*Gs*SGD) (Franke et al., 2019), raucaffricine-*O*-β-D-glucosidase (*Rs*RG) (Warzecha et al., 2000) and strictosidine β-D-glucosidase (*Rs*SGR1) from *Rauvolfia serpentina* (Xia et al., 2012), strictosidine β-D-glucosidase from *R. verticillata* (*Rv*SGD) (Chen et al., 2008), strictosidine β-D-glucosidase from *C. roseus* (*Cr*STR) (Geerlings et al., 2000), and beta-glucosidase from *Carapichea ipecacuanha* (*Ci*Ipeglu1) (Nomura et al., 2008). The highest amino acid sequence identity for *Ce*BGlu1 and *Ce*BGlu2 was observed with *GsSGD* (69.2 and 69.8%, respectively) (**Supplementary Table 4**).

Phylogenetic tree was constructed incorporating β-D-glucosidase amino acid sequences of diverse plant species to narrow the prediction of the function of both *Ce*BGlu candidates (**Figure 4**). The neighbor-joining tree grouped the enzymes into several clusters based on their role in plants: defense response, lignification, hormone deglycosylation, β-mannosidase or myrosinase activities. Defense-related enzymes included a separate cluster that belonged to monocots, and four clusters belonging to dicots, having different types of substrates: cyanogenic glucosides, isoflavonoid conjugates, alkaloid glucosides, and terpenoid glucosides. Interestingly, even though their substrates belong to the group of terpenoids, the two *C. erythraea* candidates (*Ce*BGlu1 and *Ce*BGlu2) showed the highest homology with β-D-glucosidases of the order Gentianales that prefer alkaloid glucosides as substrates (*Gs*SGD, *Rs*RG, *Rs*SGR1, *Rv*SGD, *Cr*STR, and *Ci*Ipeglu1) (**Figure 4**).

Comparative Analysis of Organ-Specific Secoiridoid Glucosides Content and *Ce*BGlu Expression and Activity

Plant glucosidase genes are developmentally regulated (Morant et al., 2008a), and exhibit different spatial expression patterns depending on their physiological functions. In this sense, we analyzed the organ-specific distribution of transcript levels and activities of the two *C. erythraea* β-glucosidase candidates, in parallel with the content of major SGs, in both diploid and tetraploid genotypes. The presence of *Ce*BGlu1 and *Ce*BGlu2 transcripts in *C. erythraea* was validated by amplifying their specific fragments using qRT-PCR. Due to the high sequence similarity between the two gene candidates, a combination of primers common to both transcripts were employed (**Figure 5A**). *CeBGlu* was amplified at low level when cDNA from roots was used as a template in qRT-PCR analysis. Expression analysis revealed that the transcription of *CeBGlu* is regulated in an organ-specific manner (**Figure 5A**). The *CeBGlu* expression pattern revealed significantly higher transcript levels in shoots compared to roots of *C. erythraea*, in both diploid and tetraploid genotypes. However, tetraploid plants displayed significantly higher *CeBGlu* transcript levels in shoots and roots compared to diploids.

In parallel, content of major secoiridoids (SW, SWM, and GP) was significantly higher in shoots than in roots of both diploid and tetraploid *C. erythraea* genotypes (**Figure 5B**). It is well documented that aerial parts are the major site of SGs biosynthesis and accumulation in common centaury (Šiler et al., 2012, 2014; Matekalo et al., 2018; Božunović et al., 2019; Filipović et al., 2019). Similarly, as in our previous study (Božunović et al., 2019), the dominant SG in leaves of diploid and tetraploid plants was SW, followed by GP and SWM. Diploid plants displayed slightly higher GP content in shoots than tetraploid ones, which

FIGURE 3 | Multiple sequence alignment of *Ce*BGlu1 and *Ce*BGlu2 with *β*-glucosidases from species belonging to the Gentianales order. The two Glu catalytic residues are colored blue, active site residues involved in sugar specificity are colored olive, and residues differing between *Ce*BGlu1 and *Ce*BGlu2 are indicated in red. The secondary structure diagram shown below the alignment is based on the *Ce*BGlu1 AlphaFold model. The pLDDT score (predicted lDDT-Cα) shown below the alignment is a per-residue estimate of AlphaFold confidence on a scale from 0 to 100. Lower values are often associated with disordered regions. It is shown for *Ce*BGlu1 AlphaFold model. RsRG sequence corresponds to the PDB:3U5Y glucosidase from *Rauvolfia serpentina*.

is in accordance with our previous study (Filipović et al., 2019), and higher content of SW in roots compared to tetraploids.

To test the total hydrolytic activity of *C. erythraea* organs, pure SW, SWM and GP were subjected to an *in vitro* enzymatic assay using crude protein extract of shoots or roots, and the decrease in SG content was analyzed using UHPLC/DAD/(±)HESI−MS^2 analysis. SWM and GP were efficiently hydrolyzed, while no significant decrease in SW content was recorded (**Figure 5C**). Shoots of both diploid and tetraploid genotypes displayed more intensive hydrolysis of SWM than corresponding roots. Both shoots and roots were efficient in hydrolyzing GP. These results generally indicate higher SG-related *β*-D-glucosidase activity in shoots than in roots, which corresponds to higher expression level of the two candidates identified within the present study.

Lower content of SGs in shoots and roots of tetraploid individuals when compared to diploids might be, at least partially, ascribed to higher expression level of SGs-related *β*-D-glucosidases and higher activity of these enzymes in tetraploids. Tetraploids most likely display more intense SGs catabolism which is reflected through lower SGs content. As tetraploid plants used in the present study are not a direct offspring of the diploid ones, the observed divergence between them is most likely also influenced by differences in their genomes.

Cloning and Heterologous Expression of *Ce*BGlu in *E. coli*

Isolated BGlu candidates were sub-cloned into the pRSETA vector; pRSETA:*CeBGlu1* and pRSETA:*CeBGlu2* constructs were transformed into *E. coli* BL21 (DE3) competent cells. To achieve simple and efficient purification of the enzyme, *CeBGlu1* and *CeBGlu2* were expressed in fusion with the His-tag.

Following expression of recombinant proteins in *E. coli*, they were isolated using His-tag affinity purification and subsequently resolved in SDS-PAGE (**Figure 6Aa** and

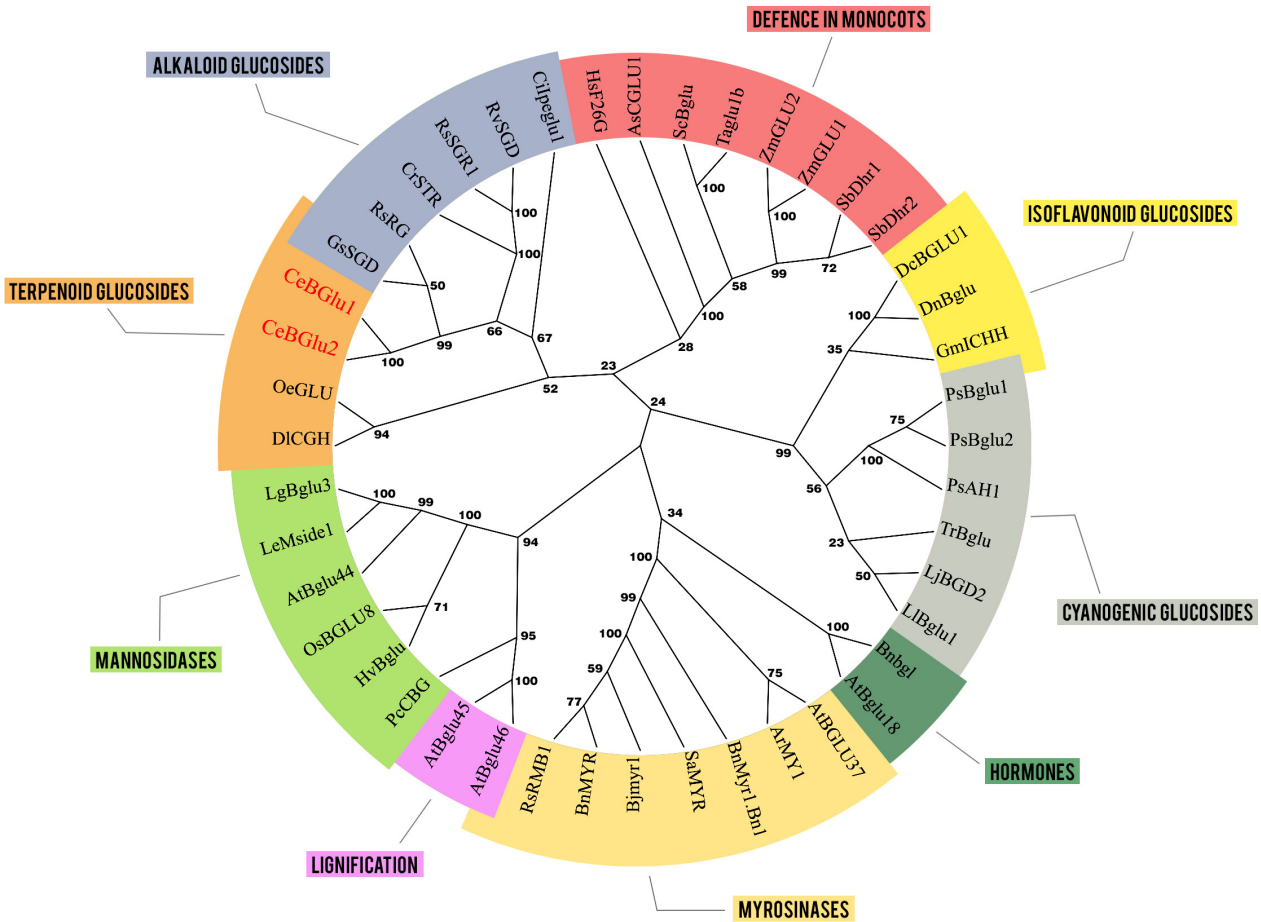


FIGURE 4 | The neighbor-joining tree generated using MEGA X (Version 10.2.6) with a bootstrap of 1,000 replicates, illustrating the phylogenetic relationship of the *Ce*BGlu1 and *Ce*BGlu2 candidates (red letters) relatively to homologous genes from various plant species (refer to **Supplementary Table 2** for a list of plant protein sequences used for phylogenetic analysis and their corresponding accession numbers). The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. Groups of enzymes with experimentally defined functions are labeled with different colors.

Supplementary Figure 1). SDS-PAGE analysis of the purified recombinant proteins detected bands with a slightly higher *Ce*BGlu1 and *Ce*BGlu2 molecular masses of ∼70 kDa than expected (62.06 kDa) (could be, at least partially, ascribed to a His-tail of recombinant proteins). Purified preparations of recombinant *Ce*BGlu1 and *Ce*BGlu2 were also analyzed by western blot using the anti-6×His antibody (**Figure 6Ab**). In both cases, a protein band with the same molecular mass as deduced from the SDS-gel was observed.

In vitro Functional Characterization of *Ce*BGlu1 and *Ce*BGlu2

Substrates such as 4-nitrophenyl-β-D-galactopyranoside, 4-nitrophenyl-β-D-thioglucoside, and 4-nitrophenyl-β-D-glucopyranoside (*p*NPG), are most commonly used artificial substrates in enzymatic assays for confirmation of β-glucosidase enzymes function (Czjzek et al., 2000; Geerlings et al., 2000; Warzecha et al., 2000; Morant et al., 2008b; Shaik et al., 2013; Tiwari et al., 2016). In order to examine functionality of *Ce*BGlu1 and *Ce*BGlu2, their hydrolytic activity against commercial substrate 4-nitrophenyl-β-D-glucopyranoside (*p*NPG) was analyzed. Both enzymes efficiently hydrolyze synthetic glucoside *p*NPG, which confirms their glucosidase activity (**Figure 6B**). *Ce*BGlu2 exhibits a higher affinity for the commercial *p*NPG substrate (4.029 μmol pNPG min^{-1} mg^{-1} protein) compared to *Ce*BGlu1 (2.267 μmol *p*NPG min^{-1} mg^{-1} protein). Similarly, enzymes isolated from *Lamium galeobdolon* (*Lg*Glu2 and *Lg*Glu4) hydrolyze *p*NPG, but the efficiency of the two tested enzymes differs (Hannemann et al., 2018). Purified β-glucosidases of *Lotus japonicus* (*Lj*BGD2 and *Lj*BGD4) hydrolyze *p*NPG, but show higher affinity for plant-synthesized substrates, such as prunasin, lotaustralin, and rhodiocyanoside (Morant et al., 2008b). However, plant β-glucosidases are not always effective in reaction with synthetic substrates. Thus, β-glucosidase of *O. europaea* (*Oe*Glu) does not show hydrolytic activity against tested *p*NPG substrate, and its function was assigned based on a reaction in which

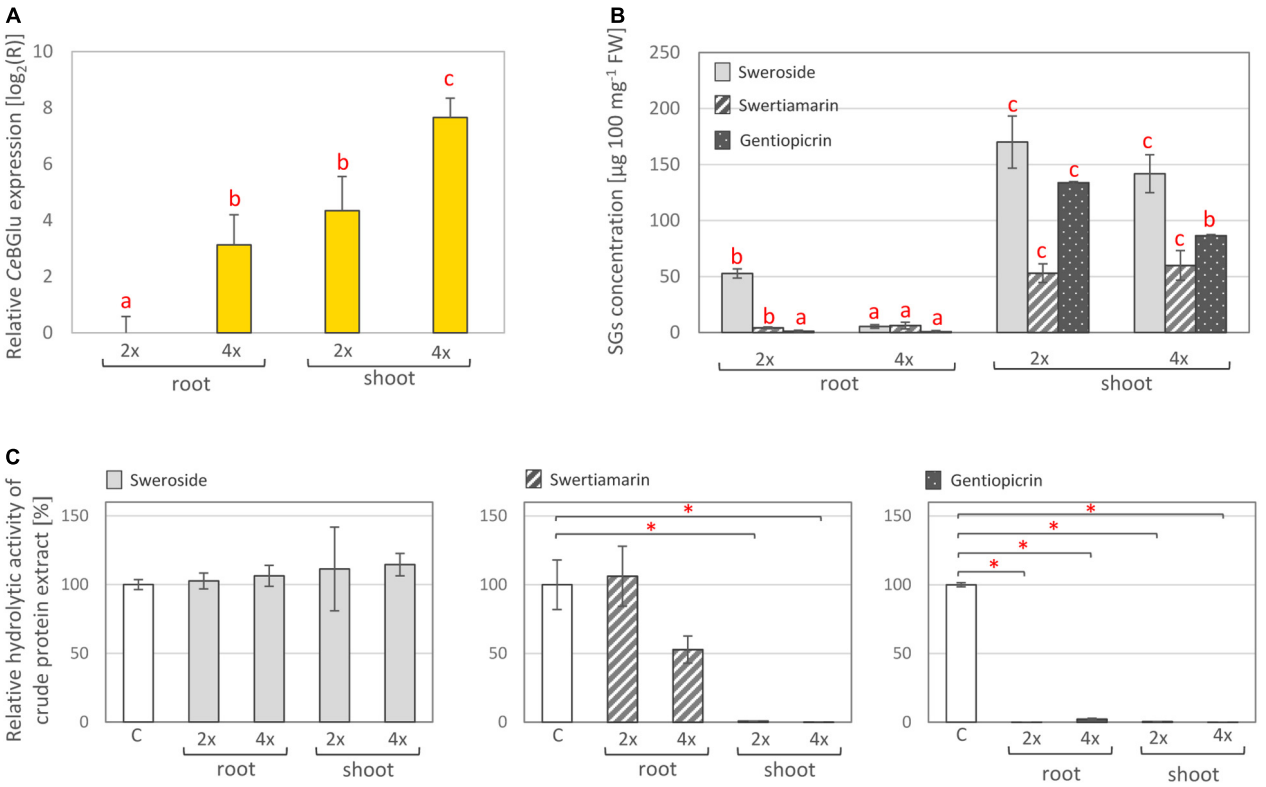


FIGURE 5 | (A) Relative expression of *CeBGlu* in shoots and roots of *in vitro* grown *C. erythraea* diploid (2×) and tetraploid (4×) plants. Values are normalized to the *CeBGlu* expression level in roots of diploid plants. Letters above the bars denote significant differences according to Fisher LSD test at $p < 0.05$. **(B)** Content of major secoiridoids in shoots and roots of diploid and tetraploid *C. erythraea* plants. Values represent means ± SE ($n = 3$). Letters above the bars denote significant differences according to Fisher LSD *post hoc* test at $p < 0.05$, for each compound independently. **(C)** Hydrolytic activity of total proteins isolated from shoots and roots of diploid and tetraploid *C. erythraea* genotypes, using sweroside, swertiamarin and gentiopicrin as substrates. The bars indicate the amount of substrates (%) not enzymatically hydrolyzed in an *in vitro* enzymatic assay. Control group (white bars), where no protein extract was added, was set to 100%, and all the values are presented relatively, in respect to control. Data are mean ± SD ($n = 3$). Red asterisks denote significantly different adjusted *p*-values, $^*p < 0.001$.

*Oe*Glu efficiently hydrolyzes a specific substrate available *in planta* – secoiridoid oleuropein (Koudounas et al., 2015). Also, *β*-glucosidases of *Arabidopsis thaliana* (AtBGlu21, AtBGlu22, and AtBGlu23) hydrolyze the tropan alkaloid scopolin, but do not exhibit hydrolytic activity against a synthetic substrate (Ahn et al., 2010).

*Ce*BGlu1 and *Ce*BGlu2 hydrolytic activity was subsequently tested in enzymatic reactions with SG-rich methanol extract of *C. erythraea* as a substrate, and changes in the content of major SGs (SW, SWM, and GP) were recorded (**Figure 6C**). *Ce*BGlu2 was more efficient than *Ce*BGlu1 in hydrolyzing SWM, which was the most abundant SG in *C. erythraea* methanol extract used in experiments. *Ce*BGlu1 and *Ce*BGlu2 were equally efficient against SW and GP. As the amounts of SW (2.0 µg ml^{-1} extract), SWM (3.5 µg ml^{-1} extract) and GP (2.1 µg ml^{-1} extract) in *C. erythraea* methanol extract were not equal, it was not possible to make conclusions about the preferable substrates of the two *Ce*BGlu.

To test the substrate specificity, *β*-D-glucosidase activity of recombinant *Ce*BGlu1 and *Ce*BGlu2 enzymes was determined against several plant-derived glucosides of the *β*-D-type (**Figure 6D**). Hydrolytic activity of *Ce*BGlu1 and *Ce*BGlu2 was tested against iridoid glucosides (1,5,9-epideoxyloganic acid-1,5,9-eDLA and loganin- LOG), secoiridoid glucosides (SEC, SW, SWM and GP), and flavonoid glucosides (apigetrin- AP, isoquercitrin- IQ, and vitexin- VI). Pure 1,5,9-eDLA was isolated from methanol extracts of *Nepeta rtanjensis* as described by Aničić et al. (2021), while AP and VI, as well as standards of iridoids and secoiridoids, were commercially purchased. As standards for aglycones of secoiridoids were not available, the *β*-D-glucosidase activity of recombinant proteins was evaluated by UHPLC/DAD/(±)HESI−MS^2 quantification of targeted glucosides (**Supplementary Figure 2**). Their content in reaction mixtures were compared to those of the negative controls where recombinant proteins were excluded. Although two recombinant proteins share 99.3% similarity of amino acid sequences, *Ce*BGlu1 and *Ce*BGlu2 display slightly differential hydrolytic activity and specificity against tested substrates (**Figure 6D**). The highest activity of *Ce*BGlu1 was recorded when using GP and AP as substrates, which were hydrolyzed by 95.4 and 97.1%, respectively. This enzyme was also efficient in hydrolyzing LOG (76.0%). The highest hydrolytic activity of *Ce*BGlu2 was recorded for GP (99.7%) and SW (98.3%), followed by AP (63.5%) and SWM (52.2%). Generally, both *Ce*BGlu1 and

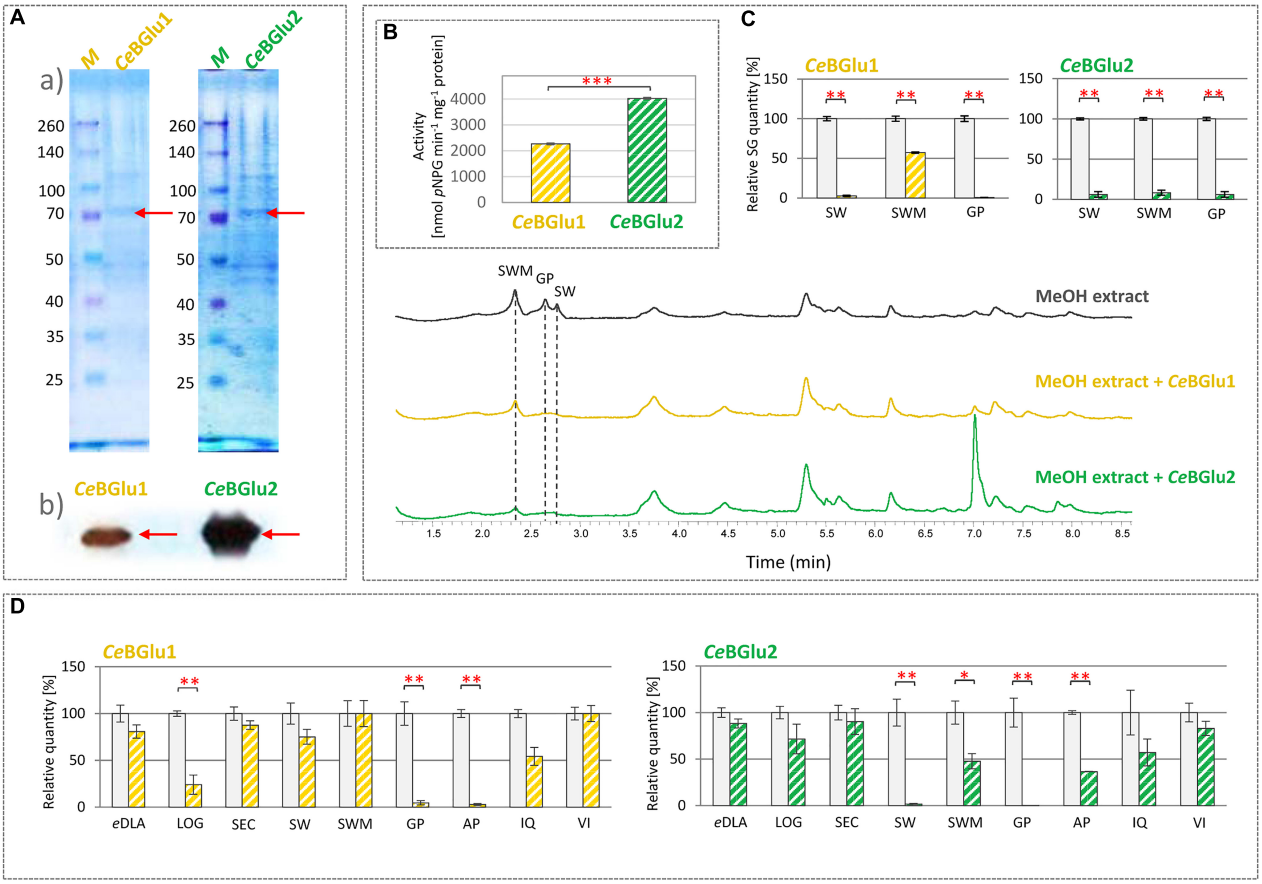


FIGURE 6 | (A) Characterization of recombinant *Ce*BGlu1 and *Ce*BGlu2 produced in *E. coli* BL21: **(a)** SDS-PAGE electropherogram; M- pre-stained SDS-PAGE size marker; purified recombinant *Ce*BGlu1 and *Ce*BGlu2 eluted after His-tag affinity purification, **(b)** Immunodetection of recombinant *Ce*BGlu1 and *Ce*BGlu2 using anti-His antibodies. **(B)** Hydrolytic activity of recombinant *Ce*BGlu1 and *Ce*BGlu2 against synthetic *p*NPG substrate. **(C)** Hydrolytic activity of *Ce*BGlu1 and *Ce*BGlu2 toward secoiridoids present in methanol extracts of *C. erythraea*. Presented are representative UHPLC/DAD chromatograms of *C. erythraea* methanol extracts at λ = 260 nm, before (black line) and after the hydrolysis with *Ce*BGlu1 (orange line) and *Ce*BGlu2 (green line). **(D)** Hydrolytic activity of recombinant *Ce*BGlu1 and *Ce*BGlu2 against 1,5,9-epideoxyloganic acid (eDLA), loganin (LOG), secologanin (SEC), sweroside (SW), swertiamarin (SWM), gentiopicrin (GP), apigetrin (AP), isoquercitrin (IQ), vitexin (VI). The bars indicate the quantity of substrates (%) not enzymatically hydrolyzed in an *in vitro* enzymatic assay. Control group, where no protein extract was added, was set to 100%, and all the values are presented relatively, in respect to control. Data are mean ± SD ($n = 3$). Red asterisks denote significantly different values according to the *t*-test, *p*-values, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

*Ce*BGlu2 display higher affinity for GP, than for SW and SWM. Although significant activity of *Ce*BGlu1 and *Ce*BGlu2 was recorded when using AP (apigenin 7-O-glucoside) as a substrate, one should bear in mind that, up to the best of our knowledge, this compound was not previously detected in *C. erythraea* methanol extracts (Šiler and Mišić, 2016; Banjanac et al., 2017; Božunović et al., 2018). Slight hydrolytic activity of *Ce*BGlu1 and *Ce*BGlu2 with 1,5,9-*e*DLA, SEC and IQ was recorded, but it was not statistically significant (**Figure 6D**). The two enzymes displayed no hydrolytic activity against VI, an apigenin 8-C-glycoside, which was also not previously detected in *C. erythraea*. *β*-D-glucosidases display substantial substrate promiscuity (Hannemann et al., 2018) so the hydrolysis of several substrates is not uncommon. Many beta-glucosidases have transglucosidase activities in addition to their hydrolase activity (Opassiri et al., 2004). However, the number of glucoconjugates in plants is likely larger than the number of beta-glucosidases and the enzymes tend to have overlapping specificities, which complicates determination of their exact functions. These cases of multiple functions have been categorized into "multitasking," where the enzyme carries out multiple functions at the same time, or "moonlighting," where the enzyme has two different functions in divided situations, according to Ketudat Cairns et al. (2015).

Transient Overexpression of *Ce*BGlu1 and *Ce*BGlu2 in Leaves of Diploid and Tetraploid *Centaurium erythraea*

Transient expression of recombinant proteins in plants using tobacco mosaic virus (TMV) based viral vectors have been well documented (Chiong et al., 2021). TMV-based overexpression pJL-TRBO vector can be used for the insertion of a target gene by agroinfiltration into plants. *Agrobacterium*-mediated delivery of genes, results in highly efficient transient expression of the foreign

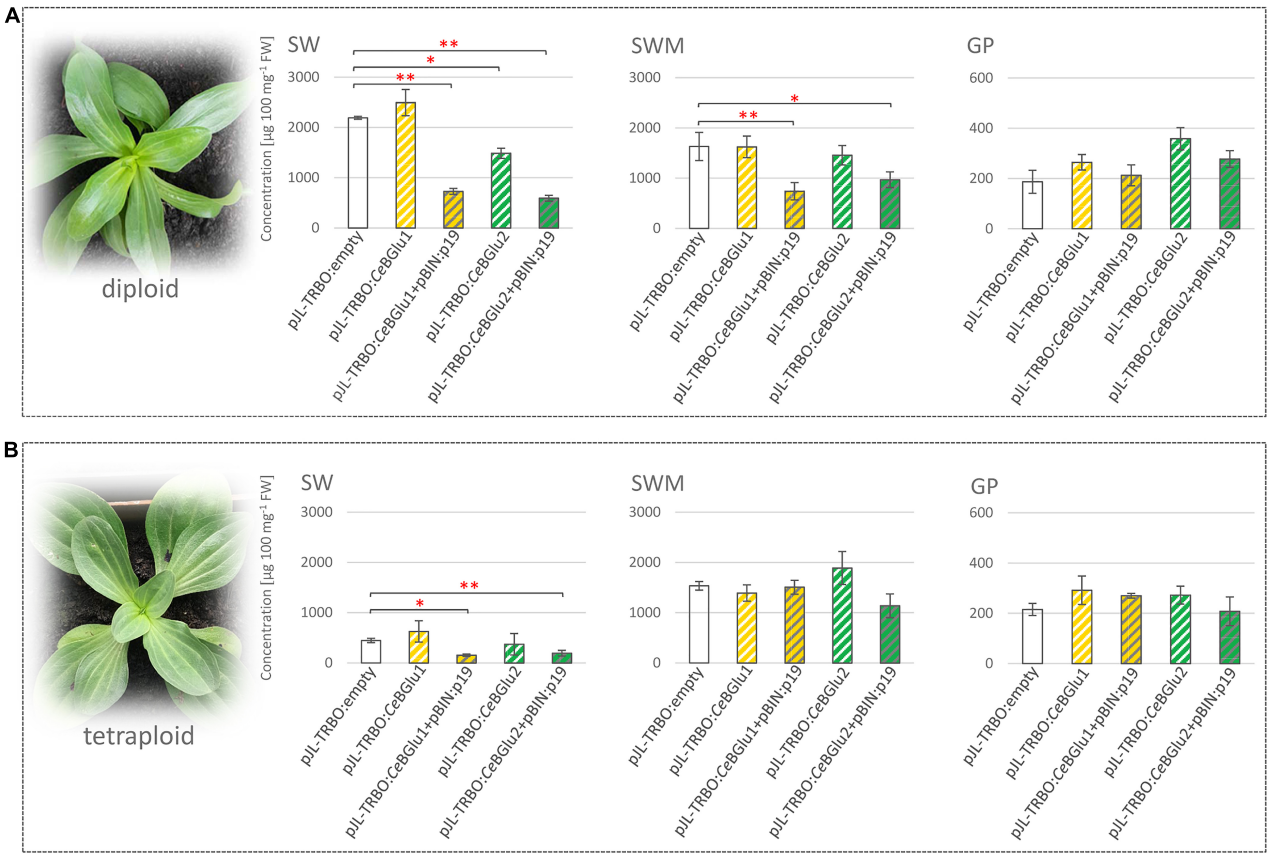


FIGURE 7 | Agroinfiltration of leaves of diploid **(A)** and tetraploid **(B)** *C. erythraea* plants. Changes in the content of secoiridoid glucosides (SGs) swertiamarin (SWM), sweroside (SW) and gentiopicrin (GP) 5 days following agroinfiltration were recorded: *Ce*BGlu1 and *Ce*BGlu2 cloned into pJL-TRBO expression vector, were co-expressed either without or with silencing inhibitor p19 (pBIN:p19 construct). Results are expressed as means ± SE. Red asterisks denote significantly different values according to the *t*-test, *p*-values, $*p < 0.05$, $**p < 0.01$.

protein, producing up to 100-fold more recombinant protein compared to a non-viral system (Lindbo, 2007).

To verify the functional identity of the *Ce*BGlu gene, *Agrobacterium*-mediated transient expression in centaury leaves was carried out. To that end, the *Ce*BGlu coding region was subcloned into the pJL-TRBO vector. The resultant plasmids designated pJL-TRBO:*Ce*BGlu1 and pJL-TRBO:*Ce*BGlu2 were used to transform *A. tumefaciens* GV3101 for the transient *in planta* expression. The two *Ce*BGLU candidates were overexpressed in leaves of five-month-old diploid and tetraploid *C. erythraea* plants, alone or in combination with p19 (**Figure 7**). Phytochemical profiling of SGs in leaves of diploid and tetraploid *C. erythraea* plants, harvested 5 days after agroinfiltration, was performed. Overexpression of *Ce*BGlu1 and *Ce*BGlu2 induced changes of SGs content in leaves of diploid and tetraploid *C. erythraea* plants when co-expressed with p19 gene silencing-suppressor (**Figure 7**). In general, phytochemical changes were more pronounced in leaves of diploid plants. Significant decrease in SWM content was recorded in diploid plants jointly over-expressing *Ce*BGlu1 and p19, or *Ce*BGlu2 and p19. The content of SW in diploid plants was significantly decreased following overexpression of *Ce*BGlu1 and *Ce*BGlu2 in combination with p19, but also when *Ce*BGlu2 was over-expressed alone. In tetraploid plants, SW content was decreased following agroinfiltration with *Ce*BGlu1 and *Ce*BGlu2 in combination with p19. No statistically significant alterations in GP content in response to agroinfiltration with *Ce*BGlu1 and *Ce*BGlu2 was observed, in both diploid and tetraploid plants. Viral-encoded suppressor of gene silencing, the p19 protein of tomato bushy stunt virus (TBSV), which is believed to prevent the onset of post-transcriptional gene silencing in the infiltrated tissues (Voinnet et al., 2003; Shah et al., 2013), obviously allowed high level of *Ce*BGlu1 and *Ce*BGlu2 transient expression, which resulted in pronounced phytochemical changes in *C. erythraea* leaves. Previous studies have suggested that transgene expression or RNA silencing in plants can be affected by ploidy, and is significantly less efficient in tetraploids than in diploids (Finn et al., 2011). In other words, transgenes are more prone to transcriptional inactivation in polyploids than in diploids (Mittelsten Scheid et al., 1996, 2003), which often poses a major limitation to polyploids improvement via biotechnology (Gao et al., 2013). Thus, it is not surprising that the use of p19 more efficiently increases the transient expression of recombinant *Ce*BGlu1 and *Ce*BGlu2 proteins in leaves of *C. erythraea* diploids

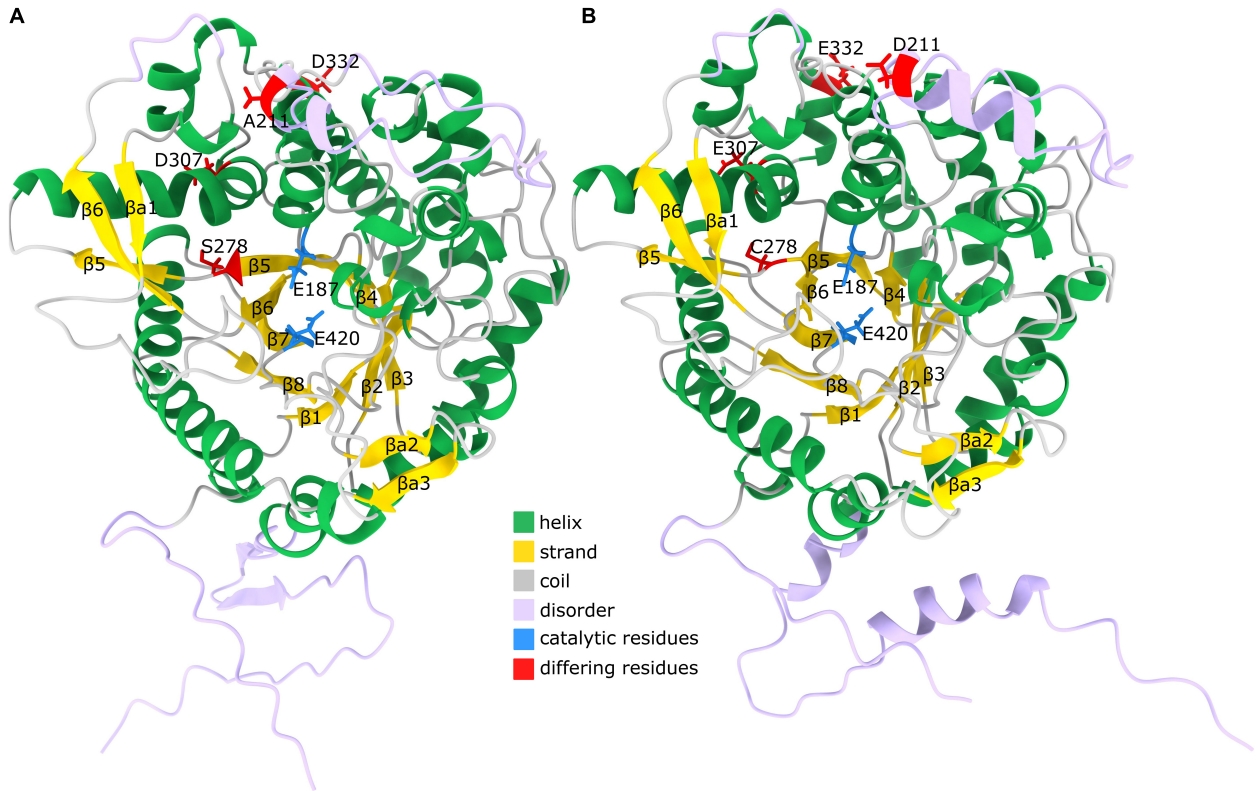


FIGURE 8 | Predicted structure of *Ce*BGlu. **(A)** AlphaFold model of *Ce*BGlu1. Helices are shown in green, strands in gold and coils in gray. Regions with pLDDT < 80 are indicated with lavender color. The two active site residues Glu187 and Glu420 are indicated with blue color, while residues differing between *Ce*BGlu1 and *Ce*BGlu2 are indicated in red **(B)** AlphaFold model of *Ce*BGlu2. Coloring is as under **(B)**.

than in tetraploids, which is accompanied with the more pronounced decrease in SGs content (**Figure 7**). Taken together, overexpression of the two *CeBGlu* isogenes decreased the content of SGs in *C. erythraea* leaves, which additionally confirmed the β-glucosidase function of these two genes. The entire centaury genome sequence is still unknown, yet it probably contains many other β-glucosidases of diverse as well as similar functions and substrate specificities, which could also contribute to shaping the secoiridoid profiles of these remarkable plants.

Comparative 3D Modeling of *Ce*BGlu1 and *Ce*BGlu2

To gain insights into sequence-structure relations of the two centaury β-glucosidases their 3D structure was inferred using AlphaFold (models provided as **Supplementary Material 1**). The obtained models had a low percentage of Ramachandran outliers and a low clash score (**Supplementary Table 5**). The percentage of Ramachandran favored Cα angles was between 92 and 93% for the two models. The majority of residues not within the favored parts of the Ramachandran diagram had a low pLDDT value (a per-residue estimate of AlphaFold confidence on a scale from 0 to 100) suggesting they are in disordered regions. The two predicted models corresponding to *Ce*BGlu1 to *Ce*BGlu2 are highly structurally similar, especially if the likely disordered regions are not taken into account – the RMSD of the superimposed pruned 475 Cα atoms in the two structures is 0.216 Å. Both obtained models showed a high degree of structural similarity to the experimental structure of the raucaffricine β-D-glucosidase from *R. serpentina* (PDB: 3U5Y, **Supplementary Table 5**; Xia et al., 2012). The models have a (β/α) 8-barrel (**Figures 8A,B**) fold characteristic for the CAZy GH1, consisting of eight parallel β-strands forming a barrel-like structure. These strands are connected via intricate helical regions and loops (**Figure 8**). Strands β5 and β6 are longer compared to other, and protrude the barrel (**Figures 8A,B**). Based on AlphaFold secondary structure annotation these are split into two strand regions each. Together with an antiparallel strand βa1 (**Figure 8**) which is not a part of the barrel, β5 and β6 form a sheet. After the (β/α) 8-barrel structure is a sheet formed from two antiparallel strands βa2 and βa3. The regions with low pLDDT are on the termini, between β4 and β5, as well as between β6 and βa1, matching closely the regions with missing density in the experimental structure of raucaffricine β-D-glucosidase from *R. serpentina* which strengthens the conclusion these regions are disordered.

The enzymes from GH1 family have a retaining double displacement mechanism in which one Glu residue acts as a proton donor for the leaving group, while another Glu residue acts as a nucleophile (Ketudat Cairns and Esen, 2010).

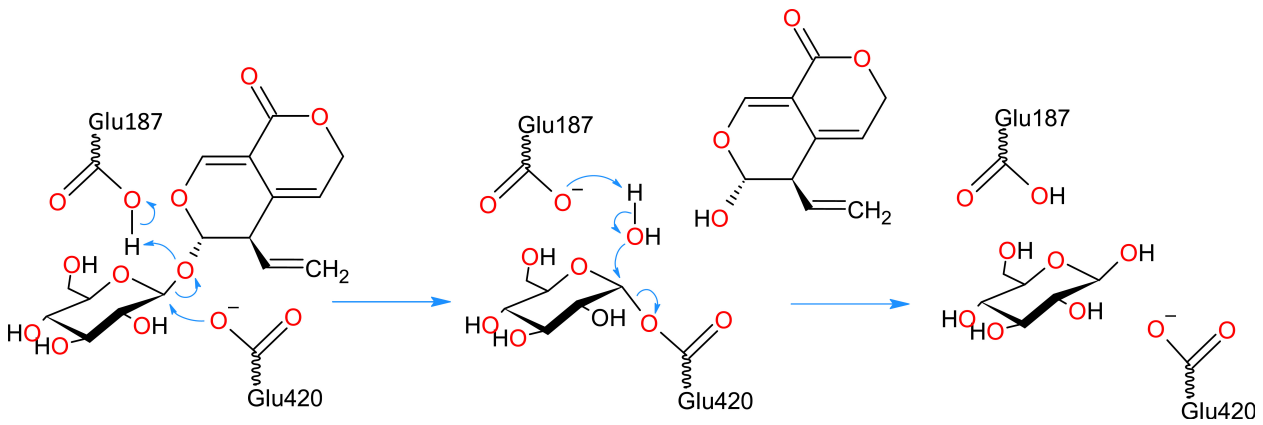


FIGURE 9 | Retaining double displacement reaction mechanism of GH1 β-glucosidase enzymes. Glu187 is the catalytic proton donor which activates the leaving group while Glu420 residue is the nucleophile. The reaction proceeds via a covalent glycosyl enzyme intermediate with an inverted bond which is hydrolyzed by nucleophilic attack of a water molecule, activated by the proton donor residue, resulting in second bond inversion and in overall anomeric retention.

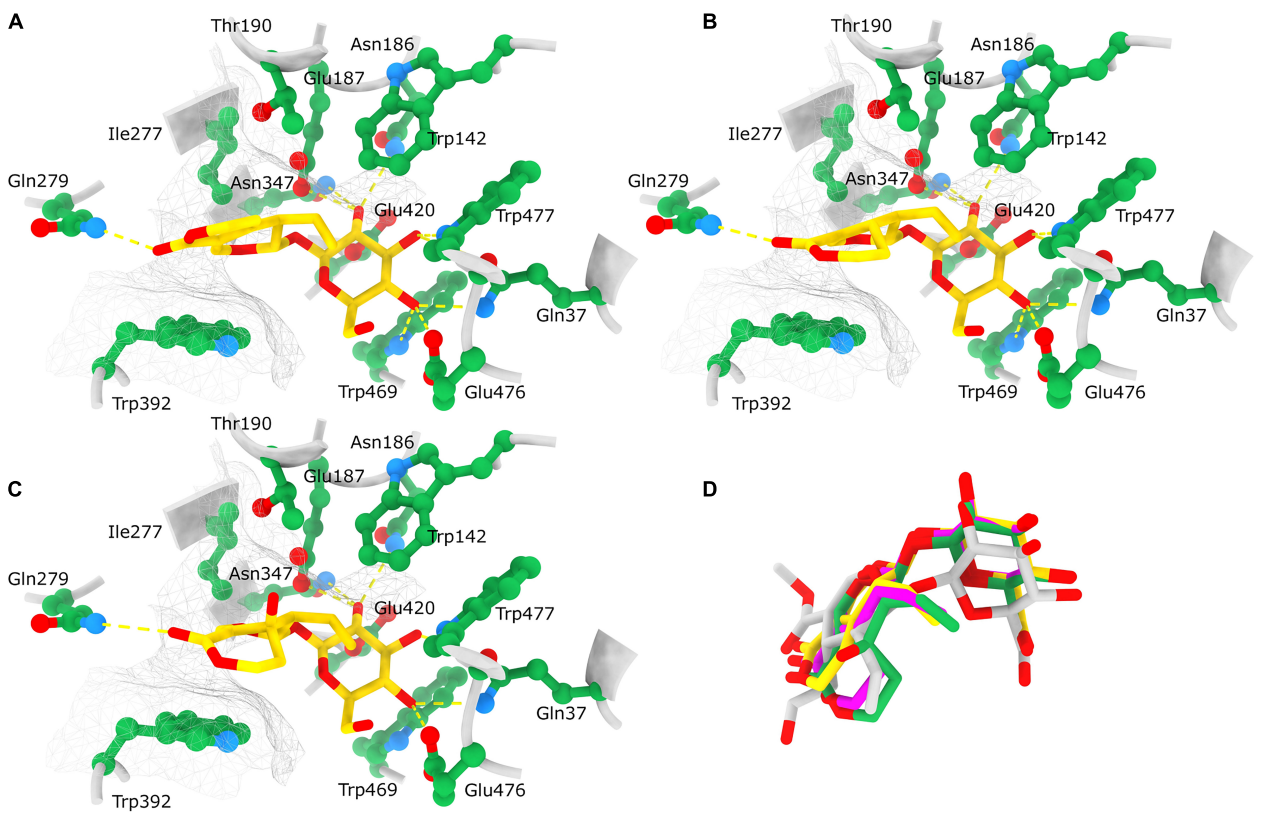


FIGURE 10 | Position of docked ligands in the *Ce*BGlu1 active site. Interactions of docked ligands: gentiopicrin **(A)**, sweroside **(B)** and swertiamarin **(C)** with the *Ce*BGlu1 active site residues. Ligands are colored yellow with oxygen shown in red; active site residues are colored green with oxygen shown in red and nitrogen in blue. H-bonds are indicated with yellow dashed lines while the hydrophobic surface encompassing the aglycone is indicated with a silver mesh. **(D)** Superposition of ligands docked to the *Ce*BGlu1 active: gentiopicrin (yellow), sweroside (pink) and swertiamarin (green) with secologanin (gray) from the experimental structure of glucosidase from *Rauvolfia serpentina* (PDB: 3U5Y). Superposition of *Ce*BGlu1 with 3U5Y was performed using matchmaker command in ChimeraX using default options.

The nucleophilic attack forms a covalent glycosyl enzyme intermediate with an inverted bond (**Figure 9**). This covalent intermediate is hydrolyzed by nucleophilic attack of a water molecule, activated by the proton donor residue, resulting in second bond inversion and in overall anomeric retention. The catalytic proton donor and nucleophile in the two centaury

β-glucosidases correspond to Glu187 and Glu420, respectively (**Figure 9**). Glu187 is located in the coil just after β4, while Glu420 is situated at the end of β7, while their Cα atoms are positioned at a distance of 10.3 Å in the models. The pKa values of respective carboxylic groups were estimated to be ∼8 for Glu187 and ∼5.5 for Glu420 which is in concordance to their proposed role of proton donor and nucleophile, respectively.

The two centaury β-glucosidases differ in four amino acids which include, from *Ce*BGlu1 to *Ce*BGlu2: A211D, S278C, D307E and D332E; the first two being radical replacements, while the latter two are conservative replacements. Position 211 is located in the disordered region between β4 and β5; 278 is located in β5, proximate to, but with a side chain facing away from the active site, while 307 and 332 are located in the helical regions connecting β5 to β6 (**Figure 8**). Based on the type, position and orientation of the differing amino acids at positions 211, 307, and 332 no impact of the mentioned amino acids replacements on the overall enzyme performance can be expected. It should be noted that these amino acids are not at the dimer interface based on comparison with 3U5Y (result not shown), so not even a long-distance interaction, due to differential multimer binding, is to be expected. However, one cannot rule out the possible indirect roles of these amino acids on the overall enzyme performance. On the other hand, the position 278 is close to the substrate binding site, and the amino acids around it – Ile277 and Gln279 are directly involved in interactions with the aglycone (**Figure 10**). Thus, even though the predicted structures of *Ce*BGlu1 to *Ce*BGlu2 are highly similar, including the active site, and the orientation of Ile277 and Gln279 side chains, S278C could have an impact on substrate recognition and overall catalytic performance. It should be noted there are no nearby Cys with which potential disulfide bonds can be formed by Cys287. Thus, the observed differences in the substrate preferences of the two *Ce*BGlu enzymes (**Figure 6**), might, at least partially, be ascribed to the differences in their amino acid sequences, especially at position 287.

To explore and visualize the binding of substrates to the active site ligand docking was analyzed only for *Ce*BGlu1, because of the very high overall similarity of the predicted models for the two β-glucosidases. To paraphrase quantitatively, the RMSD between the 236 atom pairs belonging to residues Gln37, Trp142, Asn186, Glu187, Thr190, Ile277, Gln279, Asn347, Trp392, Glu420, Trp469, Glu476, and Trp477 which form the active sites including the catalytic residues between the two predicted structures is 0.625 Å. Most open-source software for ligand docking is able to perform some sort of conformer search during docking, usually limited to rotatable bonds. The new AutoDock-Vina version implements a macrocycle conformer sampling method (Eberhardt et al., 2021), however, an analogs method appropriate for small aliphatic rings is lacking. Due to this, sets of conformers were pre generated by the ETKDG3 method with small ring torsion angle preferences (Wang et al., 2020) for GP, SW and SWM and used for docking. Prior to docking, while comparing the centaury β-glucosidase models with similar experimental structures from the PDB it was noticed *Ce*BGlu1 and *Ce*BGlu2 Glu476 side chain clashes with the glucose in the active center of superimposed experimental structures. Therefore, a flexible docking procedure was employed where Glu476 was allowed to change conformations, while all other enzyme residues were rigid. For GP the highest scoring docked pose (**Table 1** and **Figure 10A**) seems to be very close to what could be expected *in vivo*, at least based on comparison to the pose of the secologanin ligand in the experimental structure of raucaffricine β-D-glucosidase from *R. serpentina* (**Figure 10D**, 3U5Y, Xia et al., 2012). The specificity of the active site for β-glucosides is achieved by numerous H-bonds with the sugar OH groups: Asn186, Glu187 and Asn347 form H-bonds with O2, Gln37 and Trp477 with O3, Gln37, Trp469 and Glu476 with O4 on β-glucose (**Figure 10A**). The aglycone binding pocket is tight and hydrophobic, enclosed from one side by Trp392 which is parallel to the aglycone, and from the others by Trp142, Thr190 and Ile277. The only polar interaction with the ligand is via an H-bond between Gln279 and the keto oxygen. The relatively tight space in the aglycone binding site allows mostly planar aglycones, while Gln279 position favors aglycons with H-acceptors or H-donors in the appropriate position. The positions of the highest scoring poses for SW and SWM were not suitable for hydrolysis, because the aglycone occupied the position of glucose. However, the 2nd best pose for SWM and 3rd best pose for SW (**Table 1** and **Figures 10B,C**), which scored only slightly worse, were very similar both to the previously mentioned docked GP, as well as to secologanin ligand in 3U5Y (**Figure 10D**). The suboptimal highest scoring poses for SWM and SW might be the result of limitations of the used docking procedure, and imperfections in the scoring function.

TABLE 1 | Ligand docking summary into *Ce*BGlu1 AlphaFold model.

Ligand	Chosen pose affinity (kcal/mol)	Best pose affinity (kcal/mol)	Rank of chosen pose
Gentiopicrin	−9.561	−9.561	1
Swertiamarin	−9.105	−9.169	2
Sweroside	−9.359	−9.644	3

CONCLUSION

The present study highlights the direct link between β-D-glucosidases characterized within the present study (*Ce*BGlu1 and *Ce*BGlu2) and the content of SGs in analyzed *C. erythraea* tissues. Thus, it provides the first evidence of SGs-related β-glucosidases in centaury plants, and advances our understanding of SGs deglycosylation, as a part of their catabolism. These enzymes could serve as a molecular target of high biotechnological interest, in order to produce centaury plants with an optimal composition of secoiridoids, which are essential defense compounds in plants with beneficial effects in human and animal health. One of the major challenges ahead is to gain better understanding of how plants regulate the function of *Ce*BGlu1 and *Ce*BGlu2 in distinct tissues and developmental stages, and in response to biotic and abiotic stress stimuli, which is the course of our further work. Furthermore, it is imperative to assess their biological function and confirm their localization at the cellular and organelle level and define conditions under

which these enzymes come into contact with their physiological substrates. Alternatively, gene silencing of the *Ce*BGlu1 and *Ce*BGlu2 combined with metabolic profiling of silenced plants might provide more information on the role and function of these genes, and could also result in elevated amounts of SGs in tissues. In parallel, unrevealing the remaining unknown steps of SGs biosynthetic pathway in *C. erythraea* is of essential importance for establishing the biotechnology-based production of these valuable bioactive compounds. Accumulating knowledge will, in the future, enable manipulation of SGs biosynthesis and catabolism through multi-target metabolic engineering, which will further enable large-scale production of desired secoiridoids in *C. erythraea*.

DATA AVAILABILITY STATEMENT

The data presented in this study are deposited in the National Center for Biotechnology Information (nih.gov) repository (NCBI; https://www.ncbi.nlm.nih.gov/) under accession numbers ON060690 and ON060691.

AUTHOR CONTRIBUTIONS

JB, DMa, SŽ, MM, MS, and DMi conceived and designed the experiments. JB, MM, NA, MS, DMa, SŽ, BF, and TB performed the experiments. DMi performed the phytochemical characterization of samples. MD was responsible for the 3D modeling and ligand docking analysis. JB, MM, DMa, and DMi organized and wrote the manuscript with editing from all the authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, grant numbers OI173024, 451-03-9/2021-14/200007, and 451-03-68/2022-14/200007.

ACKNOWLEDGMENTS

We want to thank Milica Bogdanović (Institute for Biological Research "Siniša Stanković"- National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia) for providing the pBIN:p19 plasmid.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.914138/full#supplementary-material

REFERENCES

Adasme, M. F., Linnemann, K. L., Bolz, S. N., Kaiser, F., Salentin, S., Haupt, V. J., et al. (2021). PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.* 49, W530–W534. doi: 10.1093/nar/gkab294

Ahn, Y. O., Shimizu, B. I., Sakata, K., Gantulga, D., Zhou, Z., Bevan, D. R., et al. (2010). Scopolin-hydrolyzing β-glucosidases in roots of Arabidopsis. *Plant Cell Physiol.* 51, 132–143. doi: 10.1093/pcp/pcp174

Aničić, N., Gašić, U., Lu, F., Ćirić, A., Ivanov, M., Jevtić, B., et al. (2021). Antimicrobial and immunomodulating activities of two endemic Nepeta species and their major iridoids isolated from natural sources. *Pharmaceuticals* 14:414. doi: 10.3390/ph14050414

Banjanac, T., Dragicevic, M., Siler, B., Gasic, U., Bohanec, B., Nestorovic Zivkovic, J., et al. (2017). Chemodiversity of two closely related tetraploid *Centaurium* species and their hexaploid hybrid: metabolomic search for high-resolution taxonomic classifiers. *Phytochemistry* 140, 27–44. doi: 10.1016/j.phytochem.2017.04.005

Barleben, L., Panjikar, S., Ruppert, M., Koepke, J., and Stöckigta, J. (2007). Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell* 19, 2886–2897. doi: 10.1105/tpc.106.045682

Bartholomaeus, A., and Ahokas, J. (1995). Inhibition of P-450 by aucubin: is the biological activity of aucubin due to its glutaraldehyde-like aglycone? *Toxicol. Lett.* 80, 75–83. doi: 10.1016/0378-4274(95)03339-M

Božunović, J., Skorić, M., Matekalo, D., Živković, S., Dragićević, M., Aničić, N., et al. (2019). Secoiridoids metabolism response to wounding in common centaury (*Centaurium erythraea* Rafn) leaves. *Plants* 8:589. doi: 10.3390/plants8120589

Božunović, J., Živković, S., Gašić, U., Glamočlija, J., Ćirić, A., Matekalo, D., et al. (2018). In vitro and in vivo transformations of *Centaurium erythraea* secoiridoid glucosides alternate their antioxidant and antimicrobial capacity. *Ind. Crops Prod.* 111, 705–721. doi: 10.1016/j.indcrop.2017.11.040

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Chen, R., Liao, Z. H., Chen, M., Wang, Q., Yang, C. X., and Yang, Y. J. (2008). Molecular cloning and characterization of the Strictosidine synthase Gene from *Rauwolfia verticillata*. *Russ. J. Plant Physiol.* 55, 670–675. doi: 10.1134/S1021443708050117

Chiong, K. T., Cody, W. B., and Scholthof, H. B. (2021). RNA silencing suppressor-influenced performance of a virus vector delivering both guide RNA and Cas9 for CRISPR gene editing. *Sci. Rep.* 11:6769. doi: 10.1038/s41598-021-85366-4

Ćuković, K., Dragićević, M., Bogdanović, M., Paunović, D., Giurato, G., Filipović, B., et al. (2020). Plant regeneration in leaf culture of *Centaurium erythraea* Rafn. Part 3: de novo transcriptome assembly and validation of housekeeping genes for studies of in vitro morphogenesis. *Plant Cell Tissue Organ Cult.* 141, 417–433. doi: 10.1007/s11240-020-01801-w

Czjzek, M., Cicek, M., Zamboni, V., Bevan, D. R., Henrissat, B., and Esen, A. (2000). The mechanism of substrate (aglycone) specificity in β-glucosidases is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA,-DIMBOAGlc, and-dhurrin complexes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13555–13560. doi: 10.1073/pnas.97.25.13555

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). AutoDock Vina 1.2. 0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* 61, 3891–3898. doi: 10.1021/acs.jcim.1c00203

El-Sedawy, A. I., Hattori Kobashi, K., and Namba, T. (1990). Metabolism of sweroside from *Swertia japonica* by human intestinal bacteria. *Shoyakugaku Zasshi* 44, 122–126. doi: 10.1055/s-2006-961909

Esen, A., and Blanchard, D. J. (2000). A specific β-glucosidase-aggregating factor is responsible for the β-glucosidase null phenotype in maize. *Plant Physiol.* 122, 563–572. doi: 10.1104/pp.122.2.563

Filipović, B., Šiler, B., Nestorović Živković, J., Banjanac, T., Skorić, M., Božunović, J., et al. (2019). Diploid vs. tetraploid *Centaurium erythraea* Rafn: a comparative study of regenerative in vitro potential and biosynthetic capacity. *Lekovite Sirovine* 39, 52–59. doi: 10.5937/leksir1939053F

Finn, T. E., Wang, L., Smolilo, D., Smith, N. A., White, R., Chaudhury, A., et al. (2011). Transgene expression and transgene-induced silencing in diploid and autotetraploid Arabidopsis. *Genetics* 187, 409–423. doi: 10.1534/genetics.110.124370

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

Gao, S. J., Damaj, M. B., Park, J. W., Beyene, G., Buenrostro-Nava, M. T., Molina, J., et al. (2013). Enhanced transgene expression in sugarcane by co-expression of virus-encoded RNA silencing suppressors. *PLoS One* 8:e66046. doi: 10.1371/journal.pone.0066046

Gasic, K., Hernandez, A., and Korban, S. S. (2004). RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol. Biol. Rep.* 22, 437–438. doi: 10.1007/BF02772687

Geerlings, A., Martinez-Lozano Ibañez, M., Memelink, J., Van Der Heijden, R., and Verpoorte, R. (2000). Molecular cloning and analysis of strictosidine β-D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *J. Biol. Chem.* 275, 3051–3056. doi: 10.1074/jbc.275.5.3051

Ghisalberti, E. L. (1998). Biological and pharmacological activity of naturally occurring iridoids and secoiridoids. *Phytomedicine* 5, 147–163. doi: 10.1016/S0944-7113(98)80012-3

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

Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., et al. (2011). Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* 168, 549–557. doi: 10.1016/j.jplph.2010.08.018

Hannemann, L., Lucaciu, C. R., Sharma, S., Rattei, T., Mayer, K. F. X., Gierl, A., et al. (2018). A promiscuous beta-glucosidase is involved in benzoxazinoid deglycosylation in *Lamium galeobdolon*. *Phytochemistry* 156, 224–233. doi: 10.1016/j.phytochem.2018.10.012

Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062

Ishiguro, K., Yamaki, M., and Takagi, S. (1983). Studies on Iridoid-related compounds III: gentiopicral, the aglucone of gentiopicroside. *Planta Med.* 49, 208–210. doi: 10.1055/s-2007-969852

Jensen, S. R., Franzyk, H., and Wallander, E. (2002). Chemotaxonomy of the Oleaceae: iridoids as taxonomic markers. *Phytochemistry* 60, 213–231. doi: 10.1016/S0031-9422(02)00102-4

Jensen, S. R., and Schripsema, J. (2002). Chemotaxonomy and pharmacology of Gentianaceae. *Gentianaceae Syst. Nat. Hist.* 5, 574–631. doi: 10.1017/CBO9780511541865.007

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Ketudat Cairns, J. R., and Esen, A. (2010). β-glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2

Ketudat Cairns, J. R., Mahong, B., Baiya, S., and Jeon, J. S. (2015). β-Glucosidases: multitasking, moonlighting or simply misunderstood? *Plant Sci.* 241, 246–259. doi: 10.1016/j.plantsci.2015.10.014

Kim, D. H., Kim, B. R., Kim, J. Y., and Jeong, Y. C. (2000). Mechanism of covalent adduct formation of aucubin to proteins. *Toxicol. Lett.* 114, 181–188. doi: 10.1016/S0378-4274(99)00295-7

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. Natl. 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., Rigakou, A., Angeli, E., Melliou, E., Magiatis, P., et al. (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

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Landrum, G., Tosco, P., Kelley, B., Ric, Gedeck, S., Vianello, R., et al. (2022). *rdkit/rdkit: 2021_09_5 (Q3 2021) Release*. doi: 10.5281/zenodo.6330241

Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., and Godzik, A. (2020). FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res.* 48, W60–W64. doi: 10.1093/nar/gkaa443

Lindbo, J. A. (2007). TRBO: a high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiol.* 145, 1232–1240. doi: 10.1104/pp.107.106377

Ling, S. K., Tanaka, T., and Kouno, I. (2003). Effects of iridoids on lipoxygenase and hyaluronidase activities and their activation by β-glucosidase in the presence of amino acids. *Biol. Pharm. Bull.* 26, 352–356. doi: 10.1248/bpb.26.352

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

Malkov, S., and Simonović, A. (2011). "Shotgun assembly of *Centaurium erythraea* transcriptome," in *Proceedings of the 19th Symposium of the Serbian Plant Physiology Society*, Banja Vrujci, 13–15.

Martínez-Rosell, G., Giorgino, T., and De Fabritiis, G. (2017). PlayMolecule ProteinPrepare: a web application for protein preparation for molecular dynamics simulations. *J. Chem. Inf. Model.* 57, 1511–1516. doi: 10.1021/acs.jcim.7b00190

Matekalo, D., Skorić, M., Nikolić, T., Novaković, L., Lukić, M., Božunović, J., et al. (2018). Organ-specific and genotype-dependent constitutive biosynthesis of secoiridoid glucosides in *Centaurium erythraea* Rafn, and its elicitation with methyl jasmonate. *Phytochemistry* 155, 69–82. doi: 10.1016/j.phytochem.2018.07.015

Mazzuca, S., Spadafora, A., and Innocenti, A. M. (2006). Cell and tissue localization of β-glucosidase during the ripening of olive fruit (*Olea europaea*) by in situ activity assay. *Plant Sci.* 171, 726–733. doi: 10.1016/j.plantsci.2006.07.006

Mittelsten Scheid, O., Afsar, K., and Paszkowski, J. (2003). Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. *Nat. Genet.* 34, 450–454. doi: 10.1038/ng1210

Mittelsten Scheid, O., Jakovleva, L., Afsar, K., Maluszynska, J., and Paszkowski, J. (1996). A change of ploidy can modify epigenetic silencing. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7114–7119. doi: 10.1073/pnas.93.14.7114

Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., et al. (2008a). β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69, 1795–1813. doi: 10.1016/j.phytochem.2008.03.006

Morant, A. V., Bjarnholt, N., Kragh, M. E., Kjærgaard, C. H., Jørgensen, K., Paquette, S. M., et al. (2008b). The β-glucosidases responsible for bioactivation of hydroxynitrile glucosides in *Lotus japonicus*. *Plant Physiol.* 147, 1072–1091. doi: 10.1104/pp.107.109512

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and autodocktools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497.

Nomura, T., Quesada, A. L., and Kutchan, T. M. (2008). The new β-D-glucosidase in terpenoid-isoquinoline alkaloid biosynthesis in *Psychotria ipecacuanha*. *J. Biol. Chem.* 283, 34650–34659. doi: 10.1074/jbc.M806953200

Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A., et al. (2004). Beta-glucosidase, exo-beta-glucanase and pyridoxine transglucosylase activities of rice BGlu1. *Biochem. J.* 379, 125–131. doi: 10.1042/bj20031485

Pankoke, H., Buschmann, T., and Müller, C. (2013). Role of plant β-glucosidases in the dual defense system of iridoid glycosides and their hydrolyzing enzymes in *Plantago lanceolata* and *Plantago major*. *Phytochemistry* 94, 99–107. doi: 10.1016/j.phytochem.2013.04.016

Pankoke, H., Gehring, R., and Müller, C. (2015). Impact of the dual defence system of Plantago lanceolata (Plantaginaceae) on performance, nutrient utilisation and feeding choice behaviour of *Amata mogadorensis* larvae (Lepidoptera, Erebidae). *J. Insect Physiol.* 82, 99–108. doi: 10.1016/j.jinsphys.2015.08.006

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943

Purdy, J., and McLean, S. (1977). Constituents of *Nauclea diderrichii*. Part IX. Conversion of sweroside to naucledal and 3-epinaucledal. *Can. J. Chem.* 55, 4233–4237. doi: 10.1139/v77-600

Shah, K. H., Almaghrabi, B., and Bohlmann, H. (2013). Comparison of expression vectors for transient expression of recombinant proteins in plants. *Plant Mol. Biol. Rep.* 31, 1529–1538. doi: 10.1007/s11105-013-0614-z

Shaik, N. M., Misra, A., Singh, S., Fatangare, A. B., Ramakumar, S., Rawal, S. K., et al. (2013). Functional characterization, homology modeling and docking studies of β-glucosidase responsible for bioactivation of cyanogenic hydroxynitrile glucosides from *Leucaena leucocephala* (subabul). *Mol. Biol. Rep.* 40, 1351–1363. doi: 10.1007/s11033-012-2179-6

Šiler, B., Avramov, S., Banjanac, T., Cvetković, J., Nestorović Živković, J., Patenković, A., et al. (2012). Secoiridoid glycosides as a marker system in chemical variability estimation and chemotype assignment of *Centaurium erythraea* Rafn from the Balkan Peninsula. *Ind. Crops Prod.* 40, 336–344. doi: 10.1016/j.indcrop.2012.03.026

Šiler, B., and Mišić, D. (2016). Biologically active compounds from the genus *Centaurium* s.l. (*Gentianaceae*): current knowledge and future prospects in medicine. *Stud. Nat. Prod. Chem.* 49, 363–397. doi: 10.1016/B978-0-444-63601-0.00011-9

Šiler, B., Živković, S., Banjanac, T., Cvetković, J., Nestorović Živković, J., Ćirić, A., et al. (2014). Centauries as underestimated food additives: antioxidant and antimicrobial potential. *Food Chem.* 147, 367–376. doi: 10.1016/j.foodchem.2013.10.007

Stärk, H., Dallago, C., Heinzinger, M., and Rost, B. (2021). Light attention predicts protein location from the language of life. *Bioinform. Adv.* 1:vbab035. doi: 10.1093/bioadv/vbab035

Tiwari, R., Kumar, K., Singh, S., Nain, L., and Shukla, P. (2016). Molecular detection and environment-specific diversity of glycosyl hydrolase family 1 β-glucosidase in different habitats. *Front. Microbiol.* 7:1597. doi: 10.3389/fmicb.2016.01597

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33, 949–956. doi: 10.1046/j.1365-313X.2003.01676.x

Wang, S., Witek, J., Landrum, G. A., and Riniker, S. (2020). Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *J. Chem. Inf. Model.* 60, 2044–2058. doi: 10.1021/acs.jcim.0c00025

Warzecha, H., Gerasimenko, I., Kutchan, T. M., and Stöckigt, J. (2000). Molecular cloning and functional bacterial expression of a plant glucosidase specifically involved in alkaloid biosynthesis. *Phytochemistry* 54, 657–666. doi: 10.1016/S0031-9422(00)00175-8

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* 27, 293–315. doi: 10.1002/pro.3330

Wright, E. S. (2015). DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16:322. doi: 10.1186/s12859-015-0749-z

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

Zeng, W., Han, H., Tao, Y., Yang, L., Wang, Z., and Chen, K. (2013). Identification of bio-active metabolites of gentiopicroside by UPLC/Q-TOF MS and NMR. *Biomedi. Chromatogr.* 27, 1129–1136. doi: 10.1002/bmc.2917

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

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

frontiers | Frontiers in Plant Science

MINI REVIEW
published: 27 June 2022
doi: 10.3389/fpls.2022.938480

Metabolomics for Agricultural Waste Valorization: Shifting Toward a Sustainable Bioeconomy

*Gholamreza Khaksar[1], Mongkon Sirijan[2], Nithiwat Suntichaikamolkul[1] and Supaart Sirikantaramas[1,3]**

[1]Center of Excellence for Molecular Crop, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, [2]Faculty of Agriculture Natural Resources and Environment, Naresuan University, Phitsanulok, Thailand, [3]Omics Sciences and Bioinformatics Center, Chulalongkorn University, Bangkok, Thailand

Agriculture has been considered as a fundamental industry for human survival since ancient times. Local and traditional agriculture are based on circular sustainability models, which produce practically no waste. However, owing to population growth and current market demands, modern agriculture is based on linear and large-scale production systems, generating tons of organic agricultural waste (OAW), such as rejected or inedible plant tissues (shells, peels, stalks, etc.). Generally, this waste accumulates in landfills and creates negative environmental impacts. The plant kingdom is rich in metabolic diversity, harboring over 200,000 structurally distinct metabolites that are naturally present in plants. Hence, OAW is considered to be a rich source of bioactive compounds, including phenolic compounds and secondary metabolites that exert a wide range of health benefits. Accordingly, OAW can be used as extraction material for the discovery and recovery of novel functional compounds that can be reinserted into the production system. This approach would alleviate the undesired environmental impacts of OAW accumulation in landfills, while providing added value to food, pharmaceutical, cosmetic, and nutraceutical products and introducing a circular economic model in the modern agricultural industry. In this regard, metabolomics-based approaches have gained increasing interest in the agri-food sector for a variety of applications, including the rediscovery of bioactive compounds, owing to advances in analytical instrumentation and data analytics platforms. This mini review summarizes the major aspects regarding the identification of novel bioactive compounds from agricultural waste, focusing on metabolomics as the main tool.

Keywords: bioactive compound, metabolomics, organic agricultural waste, valorization, value-added product

OPEN ACCESS

Edited by:
Laigeng Li,
Center for Excellence in Molecular Plant Sciences (CAS), China

Reviewed by:
Rocio Rodriguez Arcos,
Institute for Fats (CSIC), Spain

****Correspondence:***
Supaart Sirikantaramas
supaart.s@chula.ac.th

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

Received: *07 May 2022*
Accepted: *02 June 2022*
Published: *27 June 2022*

Citation:
Khaksar G, Sirijan M, Suntichaikamolkul N and Sirikantaramas S (2022) Metabolomics for Agricultural Waste Valorization: Shifting Toward a Sustainable Bioeconomy. Front. Plant Sci. 13:938480. doi: 10.3389/fpls.2022.938480

INTRODUCTION

For centuries, the agricultural industry has been vital for providing food and materials to humankind. Traditional and local agriculture utilize available plants and resources (water, soil, land, etc.) in a sustainable way, ensuring the subsistence of the local community. Practically no waste is produced in these approaches as waste and unexploited products are utilized further as fertilizers, which are easily absorbed by soils (Harris and Hillman, 2014). In the

wine production industry, the generated grape marc would be further utilized in the production of other alcohols, and the final product would be used to fertilize the soil (Nerantzis and Tataridis, 2006). However, population growth and its consequent need to produce large amounts of food, as well as globalization and the pursuit of individual economic benefits have promoted the emergence of a linear-producing modern agricultural system. Unlike traditional and local agriculture, which are based on circular sustainability models, the modern system aims to meet global demand by increasing the profitability of production. To achieve this aim, modern agriculture overexploits natural resources by using the soil extensively along with large amounts of water and energy, applying pesticides to eliminate insects, and choosing monocultures over mixed production (Rockström et al., 2017; Ramankutty et al., 2018; Duque-Acevedo et al., 2020). The world population is predicted to reach ~10 billion by 2050 (data from Department of Economic and Social Affairs, United Nations (UN), n.d., Food and Agriculture Organization (FAO), n.d.). Accordingly, agricultural production also needs to grow, albeit in a sustainable way (Ramankutty et al., 2018). However, a major drawback of this modern system is the increased production of organic agricultural waste (OAW) from crops (Gustavsson et al., 2011), including rejected or inedible plant tissues such as pruning, fruit trimming, shelling or forestall residues, and food processing wastes such as rice husk and wheat straw. Notably, fruit pulp is also considered a major OAW because fruit juice production generates tons of squeezed pulp. In addition, huge quantities of fruit pulp may be rejected owing to post-harvest loss, as is commonly seen for climacteric fruits that possess a strikingly limited shelf life after harvesting. Generally, OAWs that are not further utilized accumulate uncontrollably in landfills. If poorly managed, the accumulated OAW would generate various biotic and abiotic by-products that would negatively impact the environment, health, and economy (El-Haggar, 2007; Nagendran, 2011; He et al., 2019). Moreover, such landfills generate considerable amounts of methane, nitrous oxide, sulfur dioxide, and smoke when the OAW is burned in open air, a practice which is common in many landfills and leads to atmospheric pollution with significant emissions of carbon dioxide (Wang et al., 2019).

TOWARD A SUSTAINABLE BIOECONOMY USING ORGANIC AGRICULTURAL WASTE

The question arises as to why OAW is not reutilized. Different reutilization approaches for these residues have been described, including as animal feed and subjection to anaerobic digestion and composting. Although the effectiveness of these strategies has been described by several studies, such as those demonstrating the beneficial effects of using OAW as fertilizer (Sud et al., 2008; Meng et al., 2017), OAWs continue to accumulate. The reason for this is profitability. For instance, farmers are not willing to risk replacing synthetic fertilizers, which deliver a precise quantity of nutrients, with OAW, which provides an imprecise quantity (Innes, 2013). Moreover, these strategies do not typically generate significant economic value (Garcia-Garcia et al., 2019). Therefore, a shift toward a more sustainable approach is vital. Over the last decades, the development of novel value-added products based on the exploitation of bioactive compounds from OAW has gained considerable interest, which makes OAW a suitable feedstock for valorization. The plant kingdom is extensively rich in metabolic diversity, harboring over 200,000 structurally distinct metabolites (Wurtzel and Kutchan, 2016) that are naturally present in plants, especially under stress and/or damage conditions, which also magnifies their presence in OAWs. Hence, OAW is a rich source of bioactive compounds, including phenolic compounds (PCs) and secondary metabolites, which exert a wide range of health benefits such as antioxidant, anti-cancer, anti-inflammatory, cardioprotective, anti-microbial, and anti-allergenic activities (Coman et al., 2020; Jimenez-Lopez et al., 2020). PCs are a large group of secondary metabolites generated by plants in response to multiple environmental stimuli. Owing to the numerous health-beneficial properties associated with PCs and their abundance in OAW, PCs from OAW are increasingly attracting industrial interest. Moreover, since some of these compounds are difficult and/or expensive to synthesize, their availability from OAW makes chemical synthesis unnecessary (Burri et al., 2017; Jimenez-Lopez et al., 2020).

The importance of converting OAW into value-added products has been incorporated into various market sectors. According to a recent study, the market value of agricultural waste products peaked at USD 63.3 billion in the beverage industry, followed by USD 48 billion in the medical industry and approximately USD 46 billion in the food and consumer goods sector (Beltrán-Ramírez et al., 2019). Taken together, the valorization of OAW generates a significant economic value by increasing income per harvest and improving the livelihood of the local communities, while reducing the excessive costs for waste disposal and minimizing the carbon footprint (Lucarini et al., 2018; Singh et al., 2019).

SHIFT TOWARD SUSTAINABLE AGRICULTURE: VALORIZATION OF OAW AS A POTENTIAL SOURCE OF BIOACTIVE COMPOUNDS

Moving toward a systemic, circular model of "reuse, recycle, and regenerate" is vital for developing a sustainable agricultural industry. In this context, OAW biomass should be considered a sustainable resource rather than a waste product. OAW valorization is based on the concept that any residual material or by-product can be used as an extraction material, and the recovered bioactive compounds be reinserted into the production chain. The reutilization of these functional compounds not only represents various potential applications, including in the preparation of functional foods, food and feed additives, and nutraceutical and cosmeceutical products, but also alleviates certain negative effects of OAW accumulation in landfills, thus representing a favorable measure for the environment.

The extraction of these functional bioactive compounds is an important aspect of OAW valorization both in the context of economic benefit, owing to the recovery of valuable compounds, and in the context of waste detoxification, owing to the removal of some compounds which could be undesirable in subsequent biological post-treatments (Serrano et al., 2017; Negro et al., 2018). Over the last few decades, researchers have focused on optimizing the extraction processes. Different parameters have been investigated to optimize the extraction yields of bioactive compounds available in OAW (Kareem and Rahman, 2013; Dorta et al., 2014; Wong et al., 2014). **Table 1** summarizes the different OAWs from various crops, the amounts generated in Asia (tons/year) as of 2020, the bioactive compounds extracted from them, and the main analytical platforms utilized according to the literature published in the last 5 years (2017 until now).

METABOLOMICS IN AGRI-FOOD SECTOR: CURRENT PRACTICES FOR VALORIZATION OF OAW

Metabolomics is the comprehensive characterization of small molecules or metabolites present in a biological sample. Owing to the development of chemometrics and advanced analytical platforms, metabolomics has deepened our understanding of various metabolomic and pathway networks (Hollywood et al., 2006). Numerous high-throughput analytical platforms, including liquid chromatography and gas chromatography–mass spectrometry (LC–MS and GC–MS), and nuclear magnetic resonance (NMR) spectroscopy, have been extensively utilized for this purpose (Johanningsmeier et al., 2016). Metabolomics studies use untargeted or targeted approaches, and the selection of the analytical approach depends mainly on the research question and expected outcomes. Targeted analyses focus on a class of metabolites of interest based on our pre-existing knowledge. However, untargeted analyses utilize unbiased metabolite fingerprinting to profile the global metabolome of diverse chemical classes of metabolites associated with various known and/or unknown pathways (Scalbert et al., 2009; Patti et al., 2012). As shown in **Table 1**, high-performance liquid chromatography (HPLC) has been extensively utilized as the main analytical platform for the identification and/or discovery of various bioactive compounds from OAW using a targeted approach. However, over the past few years, a combination of both targeted and untargeted approaches (HPLC coupled with LC–MS and/or GC–MS) has been utilized to obtain a complete profile of the metabolites present in OAWs (as seen in **Table 1**).

Over the past few decades, metabolomics has been extensively applied to the valorization of different OAWs from various crops. Owing to recent advances in analytical instrumentation and data analytics platforms (Putri et al., 2013; Rubert et al., 2015), metabolomics-based approaches have gained significant interest in the agri-food sector for the identification and/or rediscovery of diverse high-value bioactive compounds, especially PCs, from OAWs. Over 10,000 different PC structures with diverse natures are currently known, the most well-known of which include phenolic acids, flavonoids, and tannins (Kennedy and Wightman, 2011). These exist naturally in various concentrations in different plant parts, from roots to shoots, as well as in fruits. Accordingly, they are also present in OAWs. Recently, numerous studies have focused on the research and development of natural compounds as substitutes for synthetic additives because synthetic substances are strongly associated with various health risks, such as the appearance of allergies or even carcinogenesis (Zheng and Wang, 2001). In this context, OAWs are considered as suitable source materials for the extraction of numerous natural bioactive compounds, such as PCs, and metabolomics can be considered as an ideal approach for the identification and/or rediscovery of these compounds from OAWs.

Numerous value-added products have been generated in the food, pharmaceutical, cosmetic, and nutraceutical industries by reinserting these high-value compounds into the production chain, as in a circular economic model. Typical examples include thickening, gelling, and food stabilizing agents from tomato and citrus waste (John et al., 2017; Morales-Contreras et al., 2017); food preservatives (meat and oil product preservatives), food stabilizers, and bactericidal agents from potato peel (Sampaio et al., 2020); essential oils with anti-cancer agents from orange waste (Yang et al., 2017); biobutanol from rice husk (Quispe et al., 2017); hydrogel from durian rind (Cui et al., 2021); and single-cell protein (SCP) from corn stover and orange peel (Diwan et al., 2018).

These value-added products, produced through OAW valorization, generate significant economic value. However, further research and development is vital to fully convert the still-evolving valorization process into a sustainable approach. By integrating metabolomics into this process, we can gain a deeper understanding of the metabolic profiles of OAWs, and this can further promote the valorization process and add greater value to such products. To the best of our knowledge, only a few studies have utilized metabolomics to fully profile the metabolome of OAW and enhance the value of such products. These include studies on the production of pineapple wine and vinegar from pineapple peel and pulp (Roda et al., 2017), essential oils from the aerial parts of plants belonging to the genus *Lavandula*, mainly *L. angustifolia* (LA) and *L.* × *intermedia* (LI; Truzzi et al., 2022), supplements with therapeutic applications from *Passiflora mollissima* seeds (Ballesteros-Vivas et al., 2020), and functional foods and nutraceuticals from bean (*Vicia faba* L.) by-products (Abu-Reidah et al., 2017).

Taken together, the recovery of functional bioactive compounds can be achieved by obtaining products that can be reinserted into the economy as new raw materials within a circular and sustainable bioeconomy. **Figure 1** presents a schematic overview of the application of metabolomics in OAW valorization using a circular economy concept. OAW was considered as the input material for the valorization process in this overview. To gain a deeper understanding of the metabolic profile of OAW, we can take advantage of metabolomics to identify and/or rediscover bioactive compounds that can then be reinserted into the production chain to generate value-added products. Notably, for non-edible wastes such as peels and seeds, more studies, including toxicity tests and/or animal model studies, are needed to ensure their safety.

TABLE 1 | List of organic agricultural wastes (OAWs) from various crops, the major bioactive compounds identified and/or extracted from them, main analytical platform utilized, and recent research studies (2017–present) in this area.

Crop	Agricultural waste	Amount of waste produced in Asia (tons/year) in 2020*	Bioactive compounds	Analytical platform	References
Almond	Skin from seed	648,111	Catechin, kaempferol, isorhamnetin, naringenin, quercetin	HPLC	Chen et al., 2019
Apple	Pomace, seed, peel	55,707,264	Anthocyanins, catechin, caffeic acid, phloretin glycosides, quercetin glycosides, rutin	HPLC	Călinoiu et al., 2017; Gunes et al., 2019; Nile et al., 2019
Avocado	Peel, seed	943,327	Catechin, chlorogenic acid, cyanidin, epicatechin, gallic acid, hydroxybenzoic acid, procyanidins, 1-caffeoylquinic acid, 3-glucosidecitric acid, 3-O-p-coumaroylquinic acid, 4-caffeoylquinic acid	LC–MS/MS, HPLC	Tremocoldi et al., 2018
Banana	Peel, stalk, pulp	64,730,743	Anthocyanins, auroxanthin, catecholamine, cyaniding, delphinidin, gallocatechin, hydroxycinnamic, flavonoids, isolutein, lutein, neoxanthin, α-carotene, β-carotene, β-cryptoxanthin	LC–MS/MS	Kraithong and Issara, 2021
Barley	Husk	25,516,523	Catechins, flavonoids, gallocatechin, cis-ferulate, trans-ferulate	HPLC	Nigam, 2017
Carrot	Peel	26,126,853	Anthocyanidin, carotenoids, α-carotene, β-carotene	HPLC	Gulsunoglu et al., 2019
Cauliflower	Stem, leaves	No data	Caffeic acid, ferulate, glucoiberin	HPLC	Xu et al., 2017
Cocoa	Skin, husk, shell	777,259	Apigenin, catechin, epicatechin	LC–MS/MS, HPLC	Campos-Vega et al., 2018
Coffee	Cherry pulp	No data	Anthocyanins, caffeic acid, chlorogenic acid, di-caffeoylquinic acid	HPLC	Heeger et al., 2017
Corn	Bran	365,305,747	Anthocyanins, caffeic acid, ferulate, *p*-coumaric acid	LC–MS/MS, HPLC	Luna-Vital et al., 2017
Date	Pulp, seed	370,583,855	Phenolic acids, flavonols, fatty acids, sphingolipids, steroids	LC–MS, NMR	Otify et al., 2019
Durian	Peel, pulp, rind, seed	1,111,928 (in Thailand)	Glutathione, γ-glutamylcysteine, pyridoxamine, cysteine, leucine	CE-MS, HPLC, GC–MS, HPAEC-PAD	Pinsorn et al., 2018; Cui et al., 2021; Panpetch and Sirikantaramas, 2021; Ramli et al., 2021; Sangpong et al., 2021
Grape	Stalk, seed, pulp	29,824,812	Anthocyanins, caffeic acid, catechins, coumarate, epicatechin	HPLC–MS/MS, HPLC	Mattos et al., 2017
Grapefruit	Peel, pulp, seed	No data	Neohesperidosides, naringenin	HPIEC, LC–MS, GC–MS	Ahmed et al., 2019; Fernandez-Fernandez et al., 2020; Dorado et al., 2021
Lemon	Seed, peel, pulp	920,592	Apigenin-6, caffeic acid, coumarate, ferulate	LC–MS, HPLC	Sharma et al., 2017; Long and Mohan, 2021
Mango	Skin, pulp, seed	39,742,461	Flavonoids, gallates, hydrolysable tannins, methyl gallate, phenolics	LC–MS	Baddi et al., 2018; Bernal-Mercado et al., 2018; Wall-Medrano et al., 2020
Orange	Peel, seed, pulp	28,366,264	Caffeic acid, chrologenic acid, cinnamic, ferulate, *p*-Coumaric acid, heperetin, hesperidin, hesperetin-7-O-rutinoside, naringenin-7-O-rutinoside	GC	Pacheco et al., 2018
Papaya	Seed, peel	7,814,260	Carotene, cryptoxanthin, lutein	HPLC	Siddique et al., 2018
Pineapple	Stem, pulp, peel	12,500,507	Catechin, epicatechin, ferulate, gallic acid, phenolics	LC–MS, HPLC	Campos et al., 2020
Pomegranate	Pulp, seed, peel	No data	Anthocyanins, flavonoids, gallic acid, punicalagin	HPLC	Sandhya et al., 2018; Meselhy et al., 2020
Potato	Peel, tuber, leaf	178,599,864	Anthocyanin, caffeic acid, carotenoid, lutein, 5-O-caffeoylquinic acid, 3-O-caffeoylquinic acid, 4-O-caffeoylquinic acid, 4,5-di-O-caffeoylquinic acid, 3,5-di-O-caffeoylquinic acid, 3,4-di-O-caffeoylquinic acid, 3,4,5-tri-O-caffeoylquinic acid	HPAEC-PAD	Scharf et al., 2020
Rice	Husk, straw, bran	676,610,485	Anthocyanins, caffeic acid, ferulate, niacin, pantothenic, pyridoxine, phytosterols, tricin, tocopherols, tocotrienols, thiamine	HPLC, LC–MS	Perez-Ternero et al., 2017; Bodie et al., 2019; Peanparkdee and Iwamoto, 2019
Soybean	Husk	33,560,440	chlorogenic acid, ferulate, gallic acid	HPLC	Carneiro et al., 2020

(Continued)

TABLE 1 | Continued

Crop	Agricultural waste	Amount of waste produced in Asia (tons/year) in 2020*	Bioactive compounds	Analytical platform	References
Sweet potato	Peel, tuber, leaf	55,979,599	Anthocyanin, caffeic, lutein, 5-O-caffeoylquinic acid, 3-O-caffeoylquinic acid, 4-O-caffeoylquinic acid, 4,5-di-O-caffeoylquinic acid, 3,5-di-O-caffeoylquinic acid, 3,4-di-O-caffeoylquinic acid, 3,4,5-tri-O-caffeoylquinic acid	HPLC	Akoetey et al., 2017
Tomato	Peel, pulp, seed	116,993,632	Caffeic acid, chlorogenic acid, ferulate, β-carotene, lycopene	HPLC, LC–MS/MS	Szabo et al., 2018; Coelho et al., 2019; Lu et al., 2019
Wheat	Bran	347,921,349	Caffeic acid, ferulate, gallic acid, *p*-coumaric acid	HPLC, LC–MS/MS	Seifdavati et al., 2021; Sisti et al., 2021

**retrieved from https://www.fao.org/faostat/en/#home.*

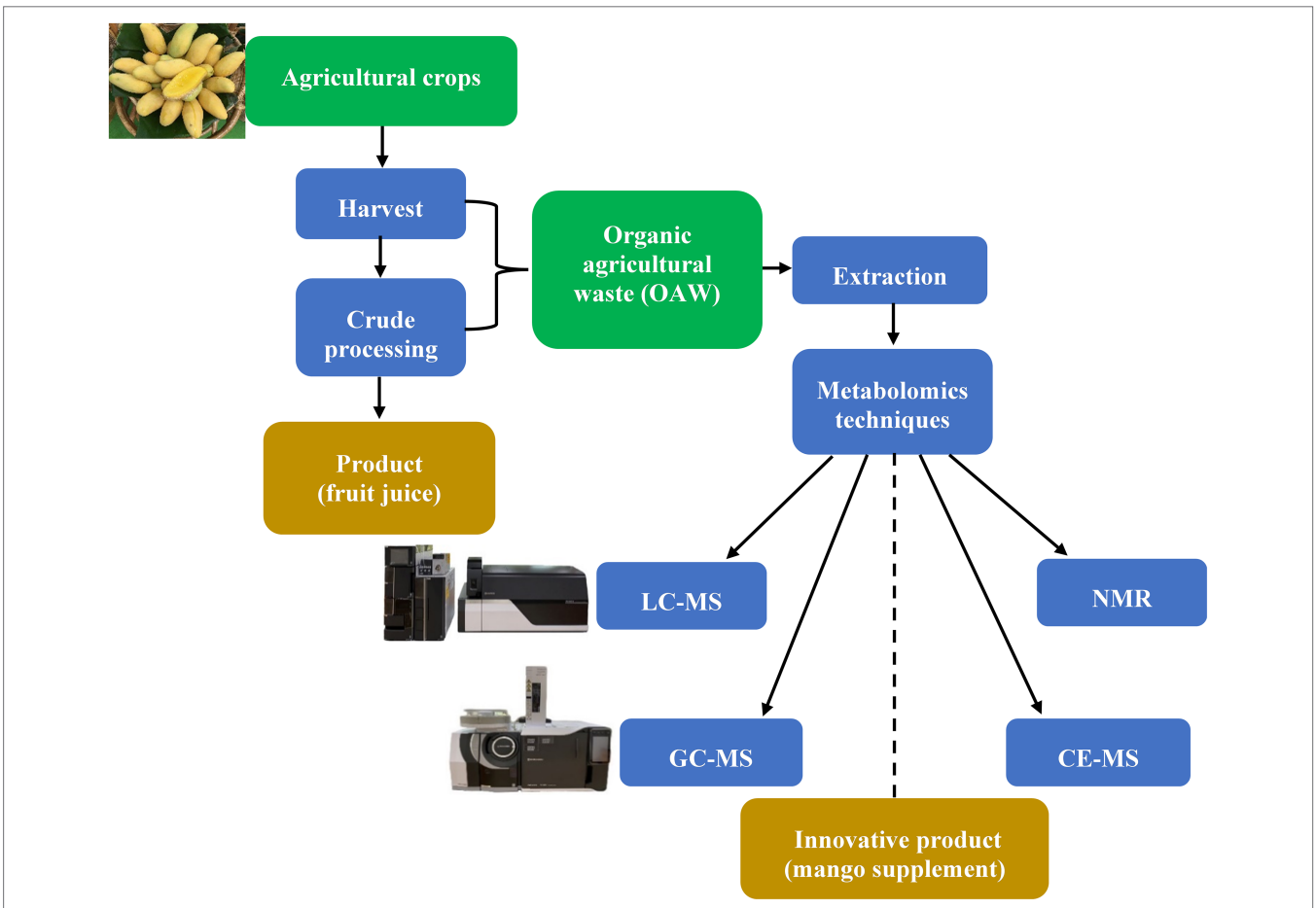


FIGURE 1 | A schematic overview of the application of metabolomics in OAW valorization in a circular economy concept. The dotted line represents the multiple steps that might be needed to generate the final product.

CONCLUSION AND FUTURE PERSPECTIVE

Owing to population growth and current market demands, modern agricultural systems are linear in nature and generate millions of tons of OAW. These wastes accumulate in landfills and create adverse environmental impacts. Since OAWs are rich in bioactive compounds, including secondary metabolites and PCs, which have various health benefits, their valorization will provide us with numerous exploitable economic, environmental, and social opportunities. To develop a circular and sustainable bioeconomy, OAW can be used as an extraction material, and the recovered

bioactive compounds can be reinserted into the production chain. In this regard, metabolomics-based approaches have gained increasing interest in the agri-food sector for the identification and/or rediscovery of these bioactive compounds. Additionally, OAW valorization can be used as a powerful and effective approach for tackling current global issues, including food shortages, waste disposal, and landfill reserves. However, further investigation is still vital to optimize extraction techniques to obtain increased product yields in an eco-friendly and economical manner. Moreover, further developments are required to fully integrate the currently evolving valorization system into a sustainable and efficient industrial tool. In this context, metabolomics can be utilized as a powerful tool to obtain a complete metabolic profile of OAWs. An important risk factor in this process could be the presence of contamination from chemicals in the crop residues owing to excessive use of pesticides and synthetic fertilizers, which should be taken into consideration. There is a dilemma among gardeners and/or orchard owners whether to use pesticides and chemical fertilizers as much as needed to avoid any yield loss and gain profit from higher quality products, or to minimize or, if possible, avoid the use of pesticides and provide the valorization industry with non-contaminated OAWs and gain profit from the OAW valorization process. For the latter to occur, we need to increase public awareness regarding the importance and need to shift toward a circular and sustainable bioeconomy in which OAW is considered a natural resource for the valorization process. In addition, more companies should dedicate themselves to the valorization of OAW and production of value-added products. Taken together, it can be concluded that, although metabolomics can be used as an effective tool to improve the valorization potential of OAWs, the question as to which approach to follow remains open-ended.

AUTHOR CONTRIBUTIONS

SS: conceptualization and supervision. GK and SS: writing—original draft preparation. GK, MS, NS, and SS: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

Research in our laboratory was supported by Chulalongkorn University Fund (GRU 6407023008-1).

ACKNOWLEDGMENTS

GK is grateful to the Graduate School, Chulalongkorn University, for supporting his Postdoctoral Fellowship from the Ratchadapisek Somphot Fund. We also thank Center of Excellence for Molecular Crop for partly supporting this work.

REFERENCES

Abu-Reidah, I. M., Arráez-Román, D., Warad, I., Fernández-Gutiérrez, A., and Segura-Carretero, A. (2017). UHPLC/MS_2-based approach for the comprehensive metabolite profiling of bean (*Vicia faba L.*) by-products: A promising source of bioactive constituents. *Food Res. Int.* 93, 87–96. doi: 10.1016/j.gecco.2020.e00902

Ahmed, S., Rattanpal, H. S., Gul, K., Dar, R. A., and Sharma, A. (2019). Chemical composition, antioxidant activity and GC-MS analysis of juice and peel oil of grapefruit varieties cultivated in India. *J. Integr. Agric.* 18, 1634–1642. doi: 10.1016/S2095-3119(19)62602-X

Akoetey, W., Britain, M. M., and Morawicki, R. O. (2017). Potential use of byproducts from cultivation and processing of sweet potatoes. *Cienc. Rural* 47:e20160610. doi: 10.1590/0103-8478cr20160610

Baddi, J., Vijayalakshmi, D., and Kapale, M. (2018). Extraction of total polyphenols and dietary fiber from mango peel - as potential sources of natural phytonutrients. *Int. J. Curr. Microbiol. App. Sci.* 7, 1196–1205. doi: 10.20546/ijcmas.2018.705.146

Ballesteros-Vivas, D., Alvarez-Rivera, G., León, C., Morantes, S. J., Ibánez, E., Parada-Alfonso, F., et al. (2020). Foodomics evaluation of the anti-proliferative potential of Passiflora mollissima seeds. *Food Res. Int.* 130:108938. doi: 10.1016/j.foodres.2019.108938

Beltrán-Ramírez, F., Orona-Tamayo, D., Cornejo-Corona, I., González-Cervantes, J. L. N., de Jesús Esparza-Claudio, J., and Quintana-Rodríguez, E. (2019). "Agro-industrial waste revalorization: the growing biorefinery," in *Biomass for Bioenergy-Recent Trends and Future Challenges*. ed. A. E.-F. Abomohra (London, UK: IntechOpen), 1–20.

Bernal-Mercado, A. T., Acevedo-Hernandez, C., Silva-Espinoza, B. A., Cruz-Valenzuela, M. R., Gonzalez-Aguilar, G. A., Nazzaro, F., et al. (2018). Antioxidant and antimicrobial capacity of phenolic compounds of mango (*Mangifera indica* L.) seed depending upon the extraction process. *J. Med. Plants Prod.* 7, 209–219. doi: 10.22092/jmpb.2018.118149

Bodie, A. R., Micciche, A. C., Atungulu, G. G., Rothrock, M. J. Jr., and Ricke, S. C. (2019). Current trends of rice milling byproducts for agricultural applications and alternative food production systems. *Front. Sustain. Food Syst.* 3:47. doi: 10.3389/fsufs.2019.00047

Burri, S. C. M., Ekholm, A., Håkansson, Å., Tornberg, E., and Rumpunen, K. (2017). Antioxidant capacity and major phenol compounds of horticultural plant materials not usually used. *J. Funct. Foods* 38, 119–127. doi: 10.1016/j.jff.2017.09.003

Călinoiu, L. F., Mitrea, L., Precup, G., Bindea, M., Rusu, B., Dulf, F. V., et al. (2017). Characterization of grape and apple peel wastes' bioactive compounds and their increased bioavailability after exposure to thermal process. *Bull. Univ. Agric. Sci. Veter-Med. Cluj-Napoca. Food Sci. Technol.* 74:80. doi: 10.15835/buasvmcn-fst:0028

Campos, D. A., Ribeiro, T. B., Teixeira, J. A., Pastrana, L., and Pintado, M. M. (2020). Integral valorization of pineapple (*Ananas comosus* L.) by-products through a green chemistry approach towards added value ingredients. *Foods.* 9:60. doi: 10.3390/foods9010060

Campos-Vega, R., Nieto-Figueroa, K. H., and Oomah, B. D. (2018). Cocoa (*Theobroma Cacao* L.) pod husk: renewable source bioactive compounds. *Trends Food Sci. Technol.* 81, 172–184. doi: 10.1016/j.tifs.2018.09.022

Carneiro, A. M., Moreira, E. A., Bragagnolo, F. S., Borges, M. S., Pilon, A. C., Rinaldo, D., et al. (2020). Soya agricultural waste as a rich source of isoflavones. *Food Res. Int.* 130:108949. doi: 10.1016/j.foodres.2019.108949

Chen, C. Y. O., Milbury, P. E., and Blumberg, J. B. (2019). Polyphenols in almond skins after blanching modulate plasma biomarkers of oxidative stress in healthy humans. *Antioxidants.* 8:95. doi: 10.3390/antiox8040095

Coelho, M., Pereira, R., Rodrigues, A. S., Teixeira, J. A., and Pintado, M. E. (2019). Extraction of tomato by-products' bioactive compounds using ohmic technology. *Food Bioprod. Process.* 117, 329–339. doi: 10.1016/j.fbp.2019.08.005

Coman, V., Teleky, B. E., Mitrea, L., Martau, G. A., Szabo, K., Calinoiu, L. F., et al. (2020). "Bioactive potential of fruit and vegetable wastes," in *Advances in Food and Nutrition Research*. ed. F. Toldra (Cambridge: Academic Press), 157–225.

Cui, X., Lee, J., Ng, K. R., and Chen, W. N. (2021). Food waste durian rind-derived cellulose organohydrogels: toward anti-freezing and antimicrobial wound dressing. *ACS Sustain. Chem. Eng.* 9, 1304–1312. doi: 10.1021/acssuschemeng.0c07705

Department of Economic and Social Affairs, United Nations (UN) (n.d.) Available at: https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100 (Accessed 5 May 2022).

Diwan, B., Parkhey, P., and Gupta, P. (2018). From agro-industrial wastes to single cell oils: a step towards prospective biorefinery. *Folia Microbiol.* 63, 547–568. doi: 10.1007/s12223-018-0602-7

Dorado, C., Cameron, R. G., Manthey, J. A., Bai, J., and Ferguson, K. L. (2021). Analysis and potential value of compounds extracted from star ruby, rio red, and ruby red grapefruit, and grapefruit juice processing residues *via* steam explosion. *Front. Nutr.* 8:691663. doi: 10.3389/fnut.2021.691663

Dorta, E., González, M., Lobo, M. G., Sánchez-Moreno, C., and de Ancos, B. (2014). Screening of phenolic compounds in by-products extracts from mangoes (*Mangifera indica* L.) by HPLC-ESI-QTOF-MS and multivariate analysis for use as food ingredients. *Food Res. Int.* 57, 51–60. doi: 10.1016/j.foodres.2014.01.012

Duque-Acevedo, M., Belmonte-Ureña, L. J., Cortés-García, F. J., and Camacho-Ferre, F. (2020). Agricultural waste: Review of the evolution, approaches and perspectives on alternative uses. *Glob. Ecol. Conserv.*, 22:e00902.

El-Haggar, S. M. (2007). "Sustainable Development and Environmental Reform," in *Sustainable Industrial Design and Waste Management*. eds. S. M. B. T.-S. I. D. and W. M. El-Haggar (Oxford: Elsevier), 125–148.

Fernandez-Fernandez, A. M., Dellacassa, E., Medrano-Fernandez, A., and Del Castillo, M. D. (2020). "Citrus waste recovery for sustainable nutrition and health," in *Food Wastes and By-Products: Nutraceutical and Health Potential*. eds. R. Campos-Vega, B. D. Oomah and H. A. Vergara-Castaneda (Hoboken, New Jersey, U.S.: Blackwell Publishing), 193–222.

Food and Agriculture Organization (FAO) (n.d.). Available at: https://www.fao.org/faostat/en/#home (Accessed 7 February 2022).

Garcia-Garcia, G., Stone, J., and Rahimifard, S. (2019). Opportunities for waste valorisation in the food industry – a case study with four UK food manufacturers. *J. Clean. Prod.* 211, 1339–1356. doi: 10.1016/j.jclepro.2018.11.269

Gulsunoglu, Z., Karbancioglu-Guler, F., Raes, K., and Kilic-Akyilmaz, M. (2019). Soluble and insoluble-bound phenolics and antioxidant activity of various industrial plant wastes. *Int. J. Food Prop.* 22, 1501–1510. doi: 10.1080/10942912.2019.1656233

Gunes, R., Palabiyik, I., Toker, O. S., Konar, N., and Kurultay, S. (2019). Incorporation of defatted apple seeds in chewing gum system and phloridzin dissolution kinetics. *J. Food Eng.* 255, 9–14. doi: 10.1016/j.jfoodeng.2019.03.010

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., and Meybeck, A. (2011). "Global food losses and food waste." in *Proceedings of the Save Food Congress FAO, 2011*. Rome, Italy.

Harris, D. R., and Hillman, G. C. (2014). *Foraging and Farming: The Evolution of Plant Exploitation*. London, UK: Routledge.

He, K., Zhang, J., and Zeng, Y. (2019). Knowledge domain and emerging trends of agricultural waste management in the field of social science: a scientometric review. *Sci. Total Environ.* 670, 236–244. doi: 10.1016/j.scitotenv.2019.03.184

Heeger, A., Kosińska-Cagnazzo, A., Cantergiani, E., and Andlauer, W. (2017). Bioactives of coffee cherry pulp and its utilisation for production of cascara beverage. *Food Chem.* 221, 969–975. doi: 10.1016/j.foodchem.2016.11.067

Hollywood, K., Brison, D. R., and Goodacre, R. (2006). Metabolomics: current technologies and future trends. *Proteomics* 6, 4716–4723. doi: 10.1002/pmic.200600106

Innes, R. (2013). "Economics of agricultural residuals and overfertilization: chemical fertilizer use, livestock waste, manure management, and environmental impacts," in *Encyclopedia of Energy, Natural Resource, and Environmental Economics*. ed. J. F. Shogren (Hoboken, New Jersey, U.S.: Elsevier Inc.), 50–57.

Jimenez-Lopez, C., Fraga-Corral, M., Carpena, M., García-Oliveira, P., Echave, J., Pereira, A. G., et al. (2020). Agriculture waste valorisation as a source of antioxidant phenolic compounds within a circular and sustainable bioeconomy. *Food Funct.* 11, 4853–4877. doi: 10.1039/D0FO00937G

Johanningsmeier, S. D., Harris, G. K., and Klevorn, C. M. (2016). Metabolomic technologies for improving the quality of food: practice and promise. *Annu. Rev. Food Sci. Technol.* 7, 413–438. doi: 10.1146/annurev-food-022814-015721

John, I., Muthukumar, K., and Arunagiri, A. (2017). A review on the potential of citrus waste for D-limonene, pectin, and bioethanol production. *Int. J. Green Energy*, 14, 599–612. doi: 10.1080/15435075.2017.1307753

Kareem, S., and Rahman, R. (2013). Utilization of banana peels for citric acid production by *Aspergillus Niger*. *Agric. Biol. J. North Am.* 4, 384–387. doi: 10.5251/abjna.2013.4.4.384.387

Kennedy, D. O., and Wightman, E. L. (2011). Herbal extracts and phytochemicals: plant secondary metabolites and the enhancement of human brain function. *Adv. Nutr.* 2, 32–50. doi: 10.3945/an.110.000117

Kraithong, S., and Issara, U. (2021). A strategic review on plant by-product from banana harvesting: A potentially bio-based ingredient for approaching novel food and agro-industry sustainability. *J. Saudi Soc. Agric. Sci.* 20, 530–543. doi: 10.1016/j.jssas.2021.06.004

Long, J. M., and Mohan, A. (2021). Food flavoring prepared with lemon by-product. *J. Food Process. Preserv.* 45:e15462. doi: 10.1111/jfpp.15462

Lu, Z., Wang, J., Gao, R., Ye, F., and Zhao, G. (2019). Sustainable valorisation of tomato pomace: a comprehensive review. *Trends Food Sci. Technol.* 86, 172–187. doi: 10.1016/j.tifs.2019.02.020

Lucarini, M., Durazzo, A., Romani, A., Campo, M., Lombardi-Boccia, G., and Cecchini, F. (2018). Bio-based compounds from grape seeds: A biorefinery approach. *Molecules* 23:1888. doi: 10.3390/molecules23081888

Luna-Vital, D., Li, Q., West, L., West, M., and De Mejia, E. G. (2017). Anthocyanin condensed forms do not affect color or chemical stability of purple corn pericarp extracts stored under different pHs. *Food Chem.* 232, 639–647. doi: 10.1016/j.foodchem.2017.03.169

Mattos, G. N., Tonon, R. V., Furtado, A. A. L., and Cabral, L. M. C. (2017). Grape by-product extracts against microbial proliferation and lipid oxidation: a review. *J. Sci. Food Agric.* 97, 1055–1064. doi: 10.1002/jsfa.8062

Meng, F., Dungait, J. A. J., Xu, X., Bol, R., Zhang, X., and Wu, W. (2017). Coupled incorporation of maize (*Zea mays* L.) straw with nitrogen fertilizer increased soil organic carbon in Fluvic Cambisol. *Geoderma* 304, 19–27. doi: 10.1016/j.geoderma.2016.09.010

Meselhy, K. M., Shams, M. M., Sherif, N. H., and El-Sonbaty, S. M. (2020). Phytochemical study, potential cytotoxic and antioxidant activities of selected food byproducts (pomegranate peel, Rice bran, Rice straw & mulberry bark). *Nat. Prod. Res.* 34, 530–533. doi: 10.1080/14786419.2018.1488708

Morales-Contreras, B. E., Contreras-Esquivel, J. C., Wicker, L., Ochoa-Martínez, L. A., and Morales-Castro, J. (2017). Husk Tomato (*Physalis ixocarpa Brot.*). Waste as a Promising Source of Pectin: Extraction and Physicochemical Characterization. *J. Food Sci.* 82, 1594–1601. doi: 10.1111/1750-3841.13768

Nagendran, R. (2011). "Agricultural waste and pollution," in *Waste*. eds. T. M. Letcher and D. A. Vallero (Amsterdam, The Netherlands: Elsevier, Inc.), 341–355.

Negro, V., Ruggeri, B., and Fino, D. (2018). Recovery of energy from orange peels through anaerobic digestion and pyrolysis processes after d-limonene extraction. *Waste Biomass Valorization.* 9, 1331–1337. doi: 10.1007/s12649-017-9915-z

Nerantzis, E. T., and Tataridis, P. (2006). Integrated enology-utilization of winery by-products into high added value products. *E J. Sci. Technol* 1, 79–89.

Nigam, P. S. (2017). An overview: recycling of solid barley waste generated as a by-product in distillery and brewery. *Waste Manag.* 62, 255–261. doi: 10.1016/j.wasman.2017.02.018

Nile, S. H., Nile, A., Liu, J., Kim, D. H., and Kai, G. (2019). Exploitation of apple pomace towards extraction of triterpenic acids, antioxidant potential, cytotoxic effects, and inhibition of clinically important enzymes. *Food Chem. Toxicol.* 131:110563. doi: 10.1016/j.fct.2019.110563

Otify, A. M., El-Sayed, A. M., Michel, C. G., and Farag, M. A. (2019). Metabolites profiling of date palm (*Phoenix dactylifera* L.) commercial by-products (pits and pollen) in relation to its antioxidant effect: a multiplex approach of MS and NMR metabolomics. *Metabolomics* 15:119. doi: 10.1007/s11306-019-1581-7

Pacheco, M. T., Moreno, F. J., and Villamiel, M. (2018). Chemical and physicochemical characterization of orange by-products derived from industry. *J. Sci. Food Agric.* 99, 868–876. doi: 10.1002/jsfa.9257

Patti, G. J., Yanes, O., and Siuzdak, G. (2012). Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi: 10.1038/nrm3314

Panpetch, P., and Sirikantaramas, S. (2021). Fruit ripening-associated leucylaminopeptidase with cysteinylglycine dipeptidase activity from durian suggests its involvement in glutathione recycling. *BMC Plant Biology*. 21, 1–14. doi: 10.1186/s12870-021-02845-6

Peanparkdee, M., and Iwamoto, S. (2019). Bioactive compounds from by-products of rice cultivation and rice processing: extraction and application in the food

and pharmaceutical industries. *Trends Food Sci. Technol.* 86, 109–117. doi: 10.1016/j.tifs.2019.02.041

Perez-Ternero, C., De Sotomayor, M. A., and Herrera, M. D. (2017). Contribution of ferulic acid, γ-oryzanol and tocotrienols to the cardiometabolic protective effects of rice bran. *J. Funct. Foods* 32, 58–71. doi: 10.1016/j.jff.2017.02.014

Pinsorn, P., Oikawa, A., Watanabe, M., Sasaki, R., Ngamchuachit, P., Hoefgen, R., et al. (2018). Metabolic variation in the pulps of two durian cultivars: Unraveling the metabolites that contribute to the flavor. *Food Chem.* 268, 118–125. doi: 10.1016/j.foodchem.2018.06.066

Putri, S. P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., et al. (2013). Current metabolomics: practical applications. *J. Biosci. Bioeng.* 115, 579–589. doi: 10.1016/j.jbiosc.2012.12.007

Quispe, I., Navia, R., and Kahhat, R. (2017). Energy potential from rice husk through direct combustion and fast pyrolysis: a review. *Waste Manag.* 59, 200–210. doi: 10.1016/j.wasman.2016.10.001

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., et al. (2018). Trends in global agricultural land use: implications for environmental health and food security. *Annu. Rev. Plant Biol.* 69, 789–815. doi: 10.1146/annurev-arplant-042817-040256

Ramli, A. N. M., Sukri, B. M., Azelee, N. A. W., and Bhuyar, P. (2021). Exploration of antibacterial and antioxidative activity of seed/peel extracts of south-east Asian fruit durian (*Durio zibethinus*) for effective shelf-life enhancement of preserved meat. *J. Food Process. Preserv.* 45:e15662. doi: 10.1111/jfpp.15662

Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., et al. (2017). Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio* 46, 4–17. doi: 10.1007/s13280-016-0793-6

Roda, A., Lucini, L., Torchio, F., Dordoni, R., De Faveri, D. M., and Lambri, M. (2017). Metabolite profiling and volatiles of pineapple wine and vinegar obtained from pineapple waste. *Food Chem.* 229, 734–742. doi: 10.1016/j.foodchem.2017.02.111

Rubert, J., Zachariasova, M., and Hajslova, J. (2015). Advances in high-resolution mass spectrometry based on metabolomics studies for food-a review. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* 32, 1685–1708. doi: 10.1080/19440049.2015.1084539

Sandhya, S., Khamrui, K., Prasad, W., Kumar, M., and Kumar, C. M. (2018). Preparation of pomegranate peel extract powder and evaluation of its effect on functional properties and shelf life of curd. *LWT.* 92, 416–421. doi: 10.1016/j.lwt.2018.02.057

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., and Van Ommen, B. (2009). Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 5, 435–458. doi: 10.1007/s11306-009-0168-0

Sampaio, S. L., Petropoulos, S. A., Alexopoulos, A., Heleno, S. A., Santos-Buelga, C., Barros, L., et al. (2020). Potato peels as sources of functional compounds for the food industry: a review. *Trends Food Sci. Technol.* 103, 118–129. doi: 10.1016/j.tifs.2020.07.015

Sangpong, L., Khaksar, G., Pinsorn, P., Oikawa, A., Sasaki, R., Erban, A., et al. (2021). Assessing dynamic changes of taste-related primary metabolism during ripening of durian pulp using metabolomic and transcriptomic analyses. *Front. Plant Sci.* 12:687799. doi: 10.3389/fpls.2021.687799

Scharf, R., Wang, R., Maycock, J., Ho, P., Chen, S., and Orfila, C. (2020). Valorisation of potato (*Solanum tuberosum*) peel waste: extraction of fibre, monosaccharides, and uronic acids. *Waste Biomass Valor.* 11, 2123–2128. doi: 10.1007/s12649-018-0532-2

Seifdavati, J., Seifzadeh, S., Ramezani, M., Mashak, R. B., Seyedsharifi, R., Elghandour, M. M. M. Y., et al. (2021). Wastes valorization of wheat straw and wheat bran treated with urea, probiotic or organic acids to enhance ruminal gas production and digestibility of pumpkin by-product. *Waste Biomass Valor.* 12, 5979–5989. doi: 10.1007/s12649-021-01432-y

Serrano, A., Fermoso, F. G., Alonso-Fariñas, B., Rodríguez-Gutierrez, G., Fernandez-Bolaños, J., and Borja, R. (2017). Olive mill solid waste biorefinery: high-temperature thermal pretreatment for phenol recovery and biomethanization. *J. Clean. Prod.* 148, 314–323. doi: 10.1016/j.jclepro.2017.01.152

Sharma, K., Mahato, N., Cho, M. H., and Lee, Y. R. (2017). Converting citrus wastes into value-added products: economic and environmently friendly approaches. *Nutrition* 34, 29–46. doi: 10.1016/j.nut.2016.09.006

Siddique, S., Nawaz, S., Muhammad, F., Akhtar, B., and Aslam, B. (2018). Phytochemical screening and *in-vitro* evaluation of pharmacological activities of peels of *Musa sapientum* and *Carica papaya* fruit. *Nat. Prod. Res.* 32, 1333–1336. doi: 10.1080/14786419.2017.1342089

Singh, K., Kumar, T., Prince, V. K., Sharma, S., and Rani, J. (2019). A review on conversion of food wastes and by-products into value added products. *IJCS* 7, 2068–2073.

Sisti, L., Gioia, C., Totaro, G., Verstichel, S., Cartabia, M., Camere, S., et al. (2021). Valorization of wheat bran agro-industrial byproduct as an upgrading filler for mycelium-based composite materials. *Ind. Crop. Prod.* 170:113742. doi: 10.1016/j.indcrop.2021.113742

Sud, D., Mahajan, G., and Kaur, M. P. (2008). Agricultural waste material as potential adsorbent for sequestering heavy metal ions from aqueous solutions - a review. *Bioresour. Technol.* 99, 6017–6027. doi: 10.1016/j.biortech.2007.11.064

Szabo, K., Cătoi, A. F., and Vodnar, D. C. (2018). Bioactive compounds extracted from tomato processing by-products as a source of valuable nutrients. *Plant Foods Hum. Nutr.* 73, 268–277. doi: 10.1007/s11130-018-0691-0

Tremocoldi, M. A., Rosalen, P. L., Franchin, M., Massarioli, A. P., Denny, C., Daiuto, E. R., et al. (2018). Exploration of avocado by-products as natural sources of bioactive compounds. *PLoS One* 13:e0192577. doi: 10.1371/journal.pone.0192577

Truzzi, E., Chaouch, M. A., Rossi, G., Tagliazucchi, L., Bertelli, D., Benvenuti, S., et al. (2022). Characterization and Valorization of the Agricultural Waste Obtained from Lavandula Steam Distillation for Its Reuse in the Food and Pharmaceutical. *Fields. Mol.* 27:1613. doi: 10.3390/molecules27051613

Wall-Medrano, A., Olivas-Aguirre, F. J., Ayala-Zavala, J. F., Domínguez-Avila, J. A., Gonzalez Aguilar, G. A., Herrera-Cazares, L. A., et al. (2020). "Health benefits of mango by-products," in *Food Wastes and By-Products: Nutraceutical and Health Potential*. eds. R. Campos-Vega, B. D. Oomah and H. A. Vergara-Castaneda (Hoboken, New Jersey, U.S.: Blackwell Publishing), 159–191.

Wang, W., Akhtar, K., Ren, G., Yang, G., Feng, Y., and Yuan, L. (2019). Impact of straw management on seasonal soil carbon dioxide emissions, soil water content, and temperature in a semi-arid region of China. *Sci. Total Environ.* 652, 471–482. doi: 10.1016/j.scitotenv.2018.10.207

Wong, Y. S., Sia, C. M., Khoo, H. E., Ang, Y. K., Chang, S. K., and Yim, H. S. (2014). Influence of extraction conditions on antioxidant properties of passion fruit (*Passiflora Edulis*) peel. *Acta. Sci. Pol. Technol. Aliment.* 13, 257–265. doi: 10.17306/J.AFS.2014.3.4

Wurtzel, E. T., and Kutchan, T. M. (2016). Plant metabolism, the diverse chemistry set of the future. *Science* 353, 1232–1236. doi: 10.1126/science.aad2062

Xu, Y., Li, Y., Bao, T., Zheng, X., Chen, W., and Wang, J. (2017). A recyclable protein resource derived from cauliflower by-products: potential biological activities of protein hydrolysates. *Food Chem.* 221, 114–122. doi: 10.1016/j.foodchem.2016.10.053

Yang, C., Chen, H., Chen, H., Zhong, B., Luo, X., and Chun, J. (2017). Antioxidant and anticancer activities of essential oil from Gannan navel orange peel. *Molecules* 22:1391. doi: 10.3390/molecules22081391

Zheng, W., and Wang, S. Y. (2001). Antioxidant activity and phenolic compounds in selected herbs. *J. Agric. Food Chem.* 49, 5165–5170. doi: 10.1021/jf010697n

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fpls.2022.923717

Features of Activity of the Phenylpropanoid Biosynthesis Pathway in Melanin-Accumulating Barley Grains

Anastasiia Y. Glagoleva[1,2], Alexander V. Vikhorev[1], Nikolay A. Shmakov[1,2], Sergey V. Morozov[3], Elena I. Chernyak[3], Gennady V. Vasiliev[1,2], Natalia V. Shatskaya[1,2], Elena K. Khlestkina[1,4] and Olesya Y. Shoeva[1,2]*

[1] Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia, [2] Kurchatov Genomics Center, ICG, SB RAS, Novosibirsk, Russia, [3] N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia, [4] N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources, Saint Petersburg, Russia

OPEN ACCESS

Edited by:
Sezai Ercisli,
Atatürk University, Turkey

Reviewed by:
Ahmad Arzani,
Isfahan University of Technology, Iran
Victoria Kuznetsova,
N.I. Vavilov Research Institute of Plant Industry, Russia
Umakanta Sarker,
Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

***Correspondence:**
Anastasiia Y. Glagoleva
glagoleva@bionet.nsc.ru

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

Received: 19 April 2022
Accepted: 09 June 2022
Published: 11 July 2022

Citation:
Glagoleva AY, Vikhorev AV, Shmakov NA, Morozov SV, Chernyak EI, Vasiliev GV, Shatskaya NV, Khlestkina EK and Shoeva OY (2022) Features of Activity of the Phenylpropanoid Biosynthesis Pathway in Melanin-Accumulating Barley Grains. Front. Plant Sci. 13:923717. doi: 10.3389/fpls.2022.923717

Barley (*Hordeum vulgare* L.) grain pigmentation is caused by two types of phenolic compounds: anthocyanins (which are flavonoids) give a blue or purple color, and melanins (which are products of enzymatic oxidation and polymerization of phenolic compounds) give a black or brown color. Genes *Ant1* and *Ant2* determine the synthesis of purple anthocyanins in the grain pericarp, whereas melanins are formed under the control of the *Blp1* gene in hulls and pericarp tissues. Unlike anthocyanin synthesis, melanin synthesis is poorly understood. The objective of the current work was to reveal features of the phenylpropanoid biosynthesis pathway functioning in melanin-accumulating barley grains. For this purpose, comparative transcriptomic and metabolomic analyses of three barley near-isogenic lines accumulating anthocyanins, melanins, or both in the grain, were performed. A comparative analysis of mRNA libraries constructed for three stages of spike development (booting, late milk, and early dough) showed transcriptional activation of genes encoding enzymes of the general phenylpropanoid pathway in all the lines regardless of pigmentation; however, as the spike matured, unique transcriptomic patterns associated with melanin and anthocyanin synthesis stood out. Secondary activation of transcription of the genes encoding enzymes of the general phenylpropanoid pathway together with genes of monolignol synthesis was revealed in the line accumulating only melanin. This pattern differs from the one observed in the anthocyanin-accumulating lines, where — together with the genes of general phenylpropanoid and monolignol synthesis pathways — flavonoid biosynthesis genes were found to be upregulated, with earlier activation of these genes in the line accumulating both types of pigments. These transcriptomic shifts may underlie the observed differences in concentrations of phenylpropanoid metabolites analyzed in the grain at a late developmental stage by high-performance liquid chromatography. Both melanin-accumulating lines showed an increased total level of benzoic acids. By contrast, anthocyanin-accumulating lines showed higher concentrations of flavonoids and *p*-coumaric and ferulic acids. A possible negative

effect of melanogenesis on the total flavonoid content and a positive influence on the anthocyanin content were noted in the line accumulating both types of pigments. As a conclusion, redirection of metabolic fluxes in the phenylpropanoid biosynthesis pathway occurs when melanin is synthesized.

Keywords: benzoic acid, hydroxycinnamic acid, *Hordeum vulgare*, metabolome, phenylpropanoids, transcriptome

INTRODUCTION

Phenylpropanoids are naturally occurring phenolic compounds. They represent a very diverse class of plant secondary metabolites, derived from phenylalanine, and play a vital role in plant physiology. These compounds are essential structural components of cell walls, protect plants against various biotic and abiotic environmental factors, act as phytoalexins against herbivores and pathogens, and—because some of them are pigments—mediate interactions of plants with pollinators and seed dispersers (Treutter, 2006; Vogt, 2010; Kiani et al., 2021; Gharaghanipor et al., 2022). Synthesis of all phenylpropanoids starts with the conversion of phenylalanine (or tyrosine) originating from the shikimate pathway by phenylalanine ammonia lyase (PAL; or tyrosine ammonia lyase, TAL), cinnamate 4-hydroxylase (C4H), and 4-coumaroyl coenzyme A (CoA) ligase (4CL) into *p*-coumaroyl-CoA, which is a precursor for all subsequent pathway branches and resulting metabolites including benzoic acids, flavonoids, monolignols, stilbenes, and coumarins (Vogt, 2010; Deng and Lu, 2017; **Figure 1**). Discovery of *p*-coumaric acid as a monomer of the melanin accumulating in oat bracts, allows to categorize melanin as a phenylpropanoid too (Varga et al., 2016).

Melanin is a widespread pigment in plant kingdom. It accumulates mostly in covering tissues of seeds, where it defends embryo against excess solar radiation, mechanical injury, and dehydration (Riley, 1997; Pandey and Dhakal, 2001; Fritz and Saukel, 2011; Lusa et al., 2018). Due to the unique features of melanin that were revealed by studies of animal, fungal and synthetic melanins, perspectives to use it in a broad range of biomedical and technological applications has been demonstrated (d'Ischia et al., 2015; Di Mauro et al., 2017; Vahidzadeh et al., 2018). Recently sunflower husks melanin was successfully used as a sorbent with high enterosorption efficiency and as an antiaging agent in elastomer compositions (Gracheva and Zheltobryukhov, 2019; Kablov et al., 2019). Besides technological application, melanin is considered as a food supplement that can increase nutrition value of stable food, for example, the melanin isolated from black garlic (Wang and Rhim, 2019). The implementation of melanin powder obtained from buckwheat husks into the desserts increased their antioxidant activity (Korpacheva et al., 2021). Despite the important functions of this pigment for plants and perspectives of its broad applications, it remains one of the poorly studied pigment in plants.

Melanin from different sources is a high-molecular-weight black-and-brown pigment that is formed by a common mechanism, which includes two stages: the oxidation of phenolic compounds by polyphenol oxidases (PPOs) and subsequent polymerization of the resultant intermediates (Britton, 1983). In plants, this process is also known as the enzymatic browning reaction, which takes place in damaged plant tissues (Nicolas et al., 1994). Participation of PPOs in melanogenesis in intact plant tissues was proposed recently (Boeckx et al., 2017), and genes encoding these enzymes have been shown to be involved in melanin synthesis in the seeds of rice, peanut, sesame, and watermelon (Fukuda et al., 2012; Wan et al., 2016; Li et al., 2020; Wang et al., 2020).

Plant melanin represents a heterogeneous group of compounds because different precursors can participate in its synthesis, among which catechol and caffeic, chlorogenic, gallic, *p*-coumaric, and protocatechuic acids have been reported (Solano, 2014; Varga et al., 2016; Jukanti, 2017). Identification of substrates for PPOs in a given plant species is considered the main difficulty with assigning melanin synthesis to biosynthetic pathways. Nevertheless, given that the same substrates can be used for the synthesis of melanin and of other phenylpropanoids, redirection of a biosynthetic pathway and as a consequence some metabolic shifts in the metabolism of phenolic compounds can be hypothesized when melanin is being produced in an organism. Research into such metabolic shifts makes it possible to trace the pathways involved in melanin formation. As a genetic model for such study the barley (*Hordeum vulgare* L.) was chosen.

The grain of this plant species can be colored by pigments of a phenolic nature: anthocyanins (which are flavonoids) give a purple and blue color, and melanin bestows a black or brown color to the grain. Synthesis of these pigments is inherited independently: genes *Ant1* and *Ant2* determine the biosynthesis of purple anthocyanins in the pericarp (Gordeeva et al., 2019), *Blx1–5* control the biosynthesis of blue anthocyanins in the aleurone layer (Strygina et al., 2017), whereas melanin is produced under the control of the *Blp1* gene in hulls and the pericarp (Costa et al., 2001; Long et al., 2019). Simple genetic control of anthocyanin and melanin synthesis in barley and the availability of backcrossed near-isogenic lines (NILs) differing in dominant alleles of *Ant1*, *Ant2*, and *Blp1* (Druka et al., 2011), make barley a convenient genetic model for exploring metabolic and transcriptomic differences underlying melanogenesis.

Although precursors of barley melanin have not been identified yet, transcriptional activation of the gene encoding PAL, but not the genes encoding enzymes of the flavonoid branch of the pathway, in a melanin-accumulating NIL (Glagoleva et al., 2017) suggests that phenolic substrates synthesized prior to the branching of the phenylpropanoid pathway may participate in melanin synthesis. In particular, an effect of the *Blp1* locus

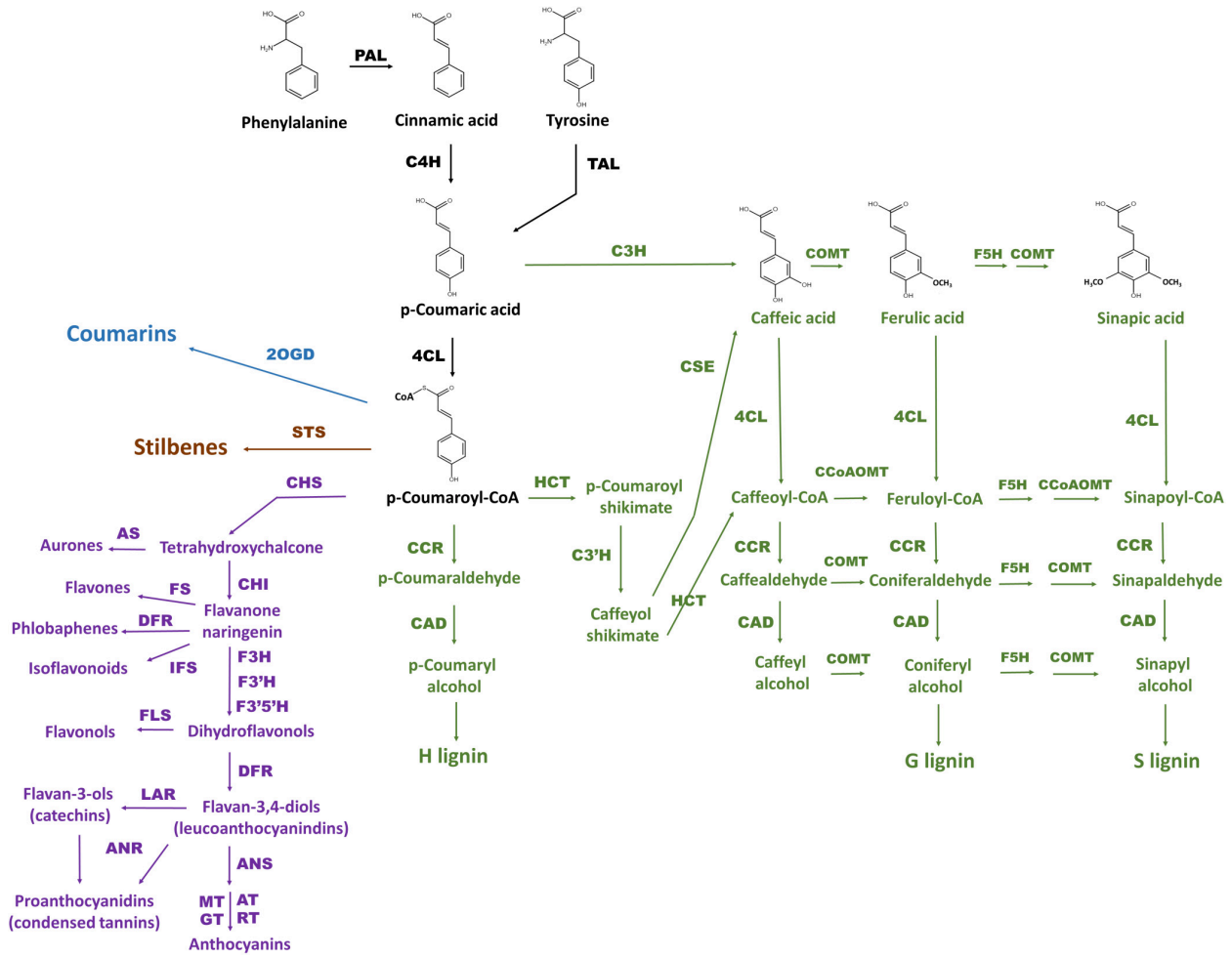


FIGURE 1 | The scheme of phenylpropanoid biosynthesis including pathway branches for flavonoids (purple), lignins (green), stilbenes (brown), and coumarins (blue) according to Deng and Lu (2017) and Marchiosi et al. (2020). General phenylpropanoid pathway reactions are black. The enzymes are PAL: phenylalanine ammonia lyase, C4H: cinnamate 4-hydroxylase, 4CL: 4-coumarate CoA ligase, HCT: hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase, C3H: *p*-coumaroyl shikimate 3-hydroxylase, C3′H: 4-coumaroylshikimate/quinate 3′-hydroxylase, CSE: caffeoyl shikimate esterase, CCoAOMT: caffeoyl CoA O-methyltransferase, F5H: ferulate 5-hydroxylase, COMT: caffeic acid O-methyltransferase, CCR: cinnamoyl CoA reductase, CAD: cinnamyl alcohol dehydrogenase, CHS: chalcone synthase, CHI: chalcone-flavanone isomerase, AS: aureusidin synthase, F3H: flavanone 3-hydroxylase, FLS: flavonol synthase, FS: flavone synthase, F3′H: flavonoid 3¢-hydroxylase, F3′5′H: flavonoid 3¢,5¢-hydroxylase, IFS: isoflavonoid synthase, DFR: dihydroflavonol 4-reductase, ANS: anthocyanidin synthase, AT: acetyltransferase, GT: glycosyltransferase, MT: methyltransferase, RT: rhamnosyltransferase, LAR: leucoanthocyanidin reductase, ANR: anthocyanidin reductase, STS: stilbene synthase, 2OGD: 2-oxoglutarate-dependent dioxygenase.

on phenylpropanoid metabolism was revealed in a barley NIL accumulating anthocyanin and melanin pigments in the grain (and carrying dominant alleles of *Ant1*, *Ant2*, and *Blp1*). In this line, earlier activation of anthocyanin synthesis and a higher content of anthocyanins in the grain were registered in comparison to the line that accumulates anthocyanins only (*Ant1*, *Ant2*) (Glagoleva et al., 2022).

The objective of the current work was to reveal further features of the phenylpropanoid biosynthesis pathway functioning in melanin-accumulating barley grains and the effects of the *Blp1* locus on phenylpropanoid metabolism. For this purpose, comparative transcriptomic and metabolomic analyses of three barley NILs accumulating anthocyanins, melanins, or both in the grain were performed.

MATERIALS AND METHODS

Plant Material and Phenotyping

Three NILs obtained in the spring *cv*. Bowman ("Bowman From Fargo," NGB22812)[1] genetic background was used in the study. The first one is the i:Bw*Blp1* line (NGB20470; hereafter: BLP – black lemma and pericarp) carrying the *Blp1* locus mapped to chromosome 1HL (Druka et al., 2011) and characterized by melanin accumulation in hulls and the grain pericarp. The second NIL is i:Bw*Ant1Ant2* (NGB22213; hereafter: PLP – purple lemma and pericarp), which is characterized by anthocyanin accumulation in the grain pericarp and leaf sheath and carries

[1]www.nordgen.org

complementary genes *Ant1* and *Ant2* located on chromosomes 7HS and 2HL, respectively (Druka et al., 2011). The third NIL is the i:Bw*Ant1Ant2Blp1* line (hereafter: BP – black and purple), which was developed previously via marker-assisted selection on the basis of lines BLP and PLP and is characterized by the accumulation of both anthocyanins and melanins in the grain (Glagoleva et al., 2022). The plants were grown in a greenhouse of the Institute of Cytology and Genetics (ICG) SB RAS (Novosibirsk, Russia) under a 12 h photoperiod in a temperature range of 20–25°C.

Extraction and Chromatographic Analysis of Phenylpropanoids

To analyze the phenylpropanoid content of the grains, three NILs (BLP, PLP, and BP) and *cv.* Bowman (parental line) were collected at the early dough stage (BBCH-83) of spike maturity. The whole grains were frozen in liquid nitrogen and dried in vacuum. Then, the dried grains were ground up and used for phenolic-compound extraction. Phenolic compounds (phenolic acids, flavonoids, and anthocyanins) were extracted from 300 mg of mature seeds with 2.5 mL of a methanol:2% formic acid mixture (7:3, v/v) (Shoeva et al., 2016) with ultrasonication at 40°C. Three independent extraction procedures were performed for each genotype under study. After centrifugation, the extracts were analyzed with Agilent 1100 liquid chromatography system (Agilent Technologies, Santa Clara, CA, United States) equipped with a quaternary pump, online degasser, autosampler, and diode array detector. Separation was implemented on an Eclipse SB-C18 (4.6 mm × 150 mm, 5 μm) column in a binary solvent system consisting of 0.1% CF_3COOH in water and methanol. The gradient of methanol was linear from 2 to 98% for 40 min, the injection volume was 4 μL, the column temperature was 23 ± 2°C, the flow rate was 0.8 mL/min. The high-performance liquid chromatography (HPLC) runs were monitored at 254 nm for benzoic acids, at 320 nm for cinnamic acids, at 360 nm for flavonoids, and at 520 nm for anthocyanins. Identification and quantification of benzoic and cinnamic acids, flavonoids, and anthocyanins were based on comparisons of the UV data and HPLC retention times of standard compounds. The limit of detection (LOD) was 0.2 μg/g in sample and the limit of quantification (LOQ) was 0.5 μg/g in sample.

Chemicals and Reagents

The following reagents were used as a standard compounds: 3,4-dihydroxybenzoic acid, catechin, dihydromyricetin, vanillic acid, *trans*-caffeic acid, 4-hydroxycinnamic acid, dihydroquercetin, *trans*-ferulic acid, dimethoxybenzoic acid, 3-hydroxycinnamic acid, dihydrokaempferol, hesperidin, rutin, trihydrate, 2-hydroxycinnamic acid, cinnamic acid, luteolin, apigenin, quercetin, naringenin, 3-*O*-glucoside cyaniding, 3-glucoside malvidin (Merck, Darmstadt, Germany).

RNA Extraction and Sequencing

For a comparative high-throughput RNA sequencing (RNA-seq) analysis, total RNA was isolated from the lines BLP, PLP, BP, and parent *cv.* Bowman at three stages of spike development. The stage 1 is the booting stage, when first awns are visible (BBCH-49); stage 2 is the late milk stage (BBCH-77), when the start of anthocyanin pigmentation is observed; and the stage 3 is the early dough stage (BBCH-83), when melanin pigmentation appears. At the stage 1, RNA was extracted from fresh whole flowers, which was separated from the spike and homogenized in liquid nitrogen. At the stages 2 and 3, pericarp and hulls were peeled from fresh grains of the NILs and homogenized in liquid nitrogen. Spikes of near-isogenic lines under study are presented in **Figure 2**. RNA was extracted in three biological replicates using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany); each replicate combined material from three plants (i.e., nine plants of each line were analyzed in total as three biological replicates). After the removal of DNA traces with the Turbo DNA free kit, RNA quality was assessed using Bioanalyzer 2100 with the RNA Nano kit. An mRNA fraction was isolated, and barcoded RNA-seq libraries for the Illumina system were generated by means of TruSeq Stranded mRNA Library Preparation Kit A. The resulting 36 barcoded libraries (3 biological replicates × 4 barley lines × 3 stages of development) were sequenced on an Illumina NextSeq 550 instrument with the NextSeq 500 High Output v2 Kit (75 cycles).

RNA-Seq Data Analysis

FASTQC v.0.11.9 was employed to evaluate the libraries' quality (Andrews, 2010). Filtering of the libraries was performed in Trimmomatic software v.0.39 with the following parameters: "LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50" (Bolger et al., 2014). This approach results in the removal of unidentified bases (N) or bases with a Phred quality score below 20 from both 3′ and 5′ ends of a read, in the removal of reads with length less than 50, and in the trimming of 5′ ends of the read when mean quality in the sliding window of length 4 drops below 30. DART tool v.1.4.2 (Lin and Hsu, 2018) was used to align the filtered reads to barley genome assembly IBSC v.2 (Mascher et al., 2017) release 47 from the Ensembl Plants database[2]. The reads aligned to each gene were counted with the help of the featureCounts function in the Subread software (Liao et al., 2014). Based on the obtained counts, principal component analysis was conducted using the DESeq2 function of plotPCA (Love et al., 2014). Raw read counts data before normalization and genes expression after CPM (counts per million mapped reads) normalization are presented in **Supplementary Data Sheets 1, 2**.

Differential gene expression analysis was performed by means of the edgeR package for R (Robinson et al., 2010). Weakly expressed genes were eliminated using the "filterByExpr" function. Differential expression between samples was detected via the generalized linear model approach. Differential expression was calculated between time stages within each barley line and between barley lines within each time stage. Thus, 26 comparisons were performed. Genes with a false discovery rate (FDR) of <0.05 and an absolute value of a $\log_2$-transformed fold change greater than 2 ($|\log_2(FC)| > 2$) were considered differentially expressed.

[2]http://plants.ensembl.org

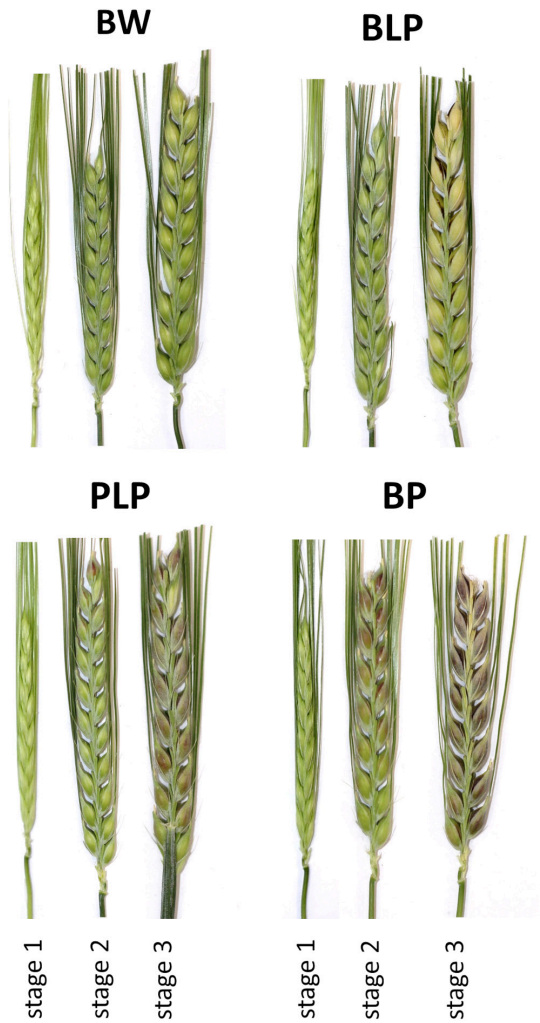


FIGURE 2 | Spikes of barley NILs BLP, PLP, BP, and of cv. Bowman (BW) used in the study at the booting stage (1), late milk stage (2), and early dough stage (3).

For functional annotation of differentially expressed genes (DEGs), Gene Ontology (GO) enrichment analysis and metabolic pathway involvement analysis were undertaken. Singular enrichment analysis of DEG lists was performed using the AgriGO v.2 toolkit (Tian et al., 2017). For each comparison performed for the differential expression analysis, lists of upregulated and downregulated genes were examined separately. GO terms with adjusted p-value < 0.05 were regarded as significantly enriched. KEGG pathways enriched in the DEG lists were identified in BlastKoala v.2.2 (Kanehisa et al., 2016). In this case, amino acid sequences of DEGs were subjected to a BLAST search against a non-redundant dataset of pangenome sequences at the genus level and then annotated in KEGG Ontology (KO) terms by the KOALA algorithm. Barley genes involved in the phenylpropanoid biosynthesis pathway were detected via a homology search in the BioMart software using *Arabidopsis thaliana* genes participating in the phenylpropanoid biosynthesis pathway (KEGG id: KO00940) as a query. Heatmap figures were generated using the seaborn v.0.11.2 library in Python.

Verification of the RNA-Seq Data by Quantitative Reverse-Transcription PCR

Nine genes with different expression patterns were chosen randomly for the qRT-PCR verification of the RNA-seq data: *HORVU1Hr1G011930, HORVU2Hr1G103040, HORVU2Hr1G103000, HORVU3Hr1G077790, HORVU4Hr 1G090870, HORVU6Hr1G075900, HORVU3Hr1G077960, HORVU7Hr1G098280,* and *HORVU4Hr1G014010*. The qRT-PCR was performed on a QuantStudio 5 machine (Applied Biosystems, Waltham, MA, United States)[3] with the HS-qPCR Lo-ROX SYBR kit (Biolabmix, Novosibirsk, Russia) in a 15 μL reaction mixture. The number of amplification cycles and annealing temperature were optimized for each primer pair (**Supplementary Table 1**). Three technical replicates of each reaction were run. Gene expression levels were calculated by the relative-standard-curve method and normalized to the geometrical mean of *actin* and *ubiquitin* expression (Himi et al., 2005; von Zitzewitz et al., 2005).

RESULTS

Transcriptome Sequencing and Identification of Differentially Expressed Genes

A total of 36 mRNA libraries were sequenced that were derived from grain envelopes of the three NILs (BLP, PLP, and BP) and parental *cv.* Bowman at three stages of spike development: stage 1, booting; stage 2, late milk; and stage 3, early dough (three biological replicates for each type of sample). In total, 801,184,069 single-end 75 bp raw reads were obtained. After filtering, 741,311,996 (92.51%) reads were retained. On average, 92.43% (71.11% unique) reads were successfully mapped to the reference barley genome in the DART software (**Supplementary Table 2**).

After the removal of genes with low expression and CPM (counts per million mapped reads) normalization, there were 30,969 genes with an expression level higher than the significance threshold. Then, differential expression was calculated between stages within each line and between lines within each stage (**Supplementary Table 3**). An average, the number of upregulated and downregulated DEGs between the NILs at the stage 1 were 23 and 54, respectively, at the stage 2 – 283 upregulated and 130 downregulated DEGs, and at the stage 3 – 677 upregulated and 304 downregulated DEGs between the NILs under study were revealed (**Supplementary Figures 1–3**). In a comparison of numbers of DEGs between stages 1 and 2 within the lines, on average 1,788 upregulated and 2,415 downregulated DEGs that are common to all lines were found. The average number of line-specific DEGs was approximately an order of magnitude less than the number of common genes (**Figure 3**). In contrast, 76 upregulated and 15 downregulated

[3]http://www.lifetechnologies.com

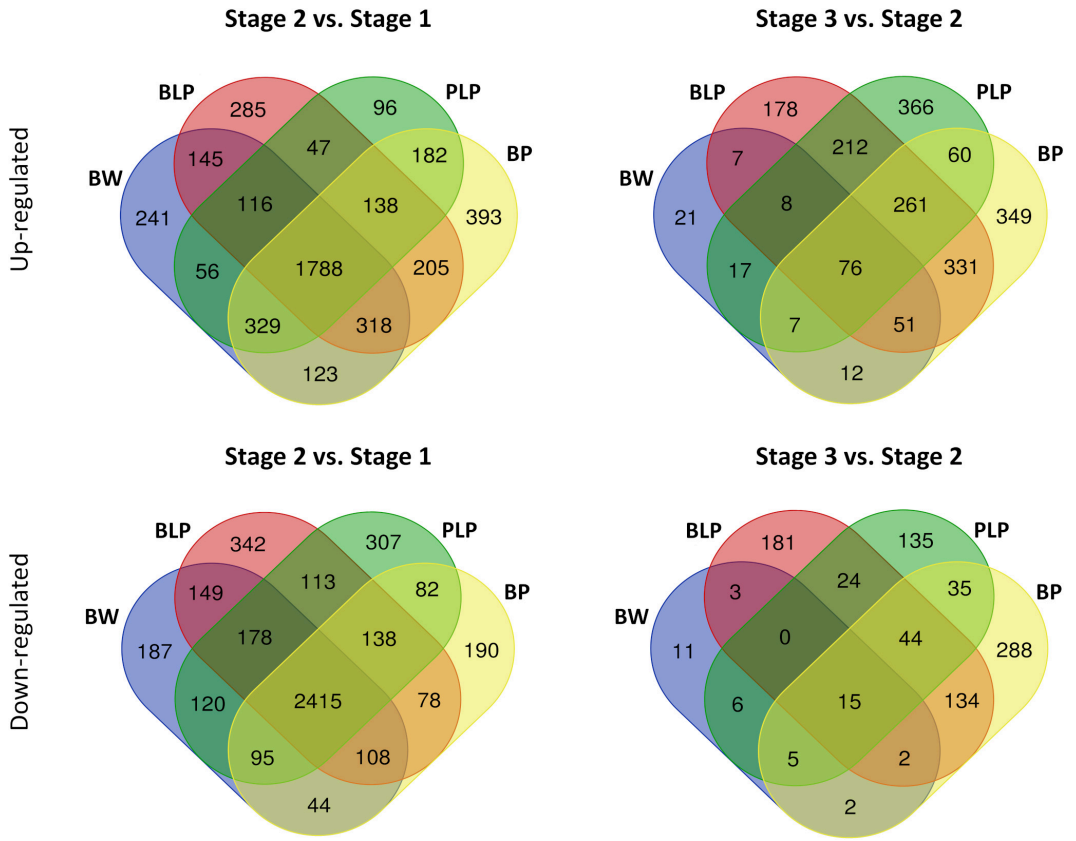


FIGURE 3 | Venn diagrams of upregulated and downregulated DEGs between stages 1 and 2 and between stages 2 and 3 within the lines.

DEGs that are common for all lines were revealed between stages 2 and 3, whereas the number of line-specific DEGs was substantially greater in all lines characterized by pigment accumulation but not in the uncolored parental line (Bowman). It is possible that at the early stages of spike development, genes responsible for plant primary metabolism that are common to all the tested lines were activated regardless of grain pigmentation; on the contrary, during spike maturation, genes of special or secondary metabolism may be activated, which is associated with pigment synthesis in the barley grain.

Gene Ontology Term Enrichment in Lists of Differentially Expressed Genes

The lists of DEGs between different lines and different stages were annotated by singular enrichment analysis of GO terms to gain a better understanding of the DEGs functions. In total, 49.4% of upregulated genes and 52.2% of downregulated genes were annotated successfully. Full lists of DEGs along with GO term enrichment are presented in **Supplementary Table 4**.

At the booting stage (stage 1), no significantly enriched GO terms were found between the non-pigmented Bowman cultivar and the lines characterized by accumulation of melanins and/or anthocyanins. In contrast, there were differences between the lines in metabolic processes in conjunction with pigmentation at the later stages. In the lines accumulating melanins (BLP and BP), the lists of upregulated DEGs at stages 2 and 3 were found to be enriched with GO terms related to fatty acid biosynthesis, lipid biosynthetic process, and organic acid biosynthesis when compared with Bowman. On the other hand, the upregulation of genes involved in the L-phenylalanine metabolic process and ammonia-lyase activity was detectable only in the BLP line at the later stage of spike maturation (stage 3) in comparison with Bowman (**Figure 4**). In the list of downregulated DEGs, GO terms "photosynthesis," "thylakoid," and "photosystem" were found for the BLP line at stage 2, but they were absent in the other pigmented lines in comparison with Bowman. Moreover, the genes involved in cell wall biogenesis and assembly, lignin metabolic process, and cell growth turned out to be downregulated in lines BLP and BP when compared with Bowman at stage 3 but not in the PLP line. In the comparison of the PLP line with Bowman, there were no line-specific GO terms that were not detectable in the other lines.

Identification of Genes Involved in Phenylpropanoid Biosynthesis

Barley genes encoding the main enzymes of the phenylpropanoid biosynthesis pathway were identified in the KEGG database via homology with *A. thaliana* genes (**Supplementary Table 5**). Most of these enzymes are encoded by gene families in the barley genome; therefore, we used mean levels of gene expression for

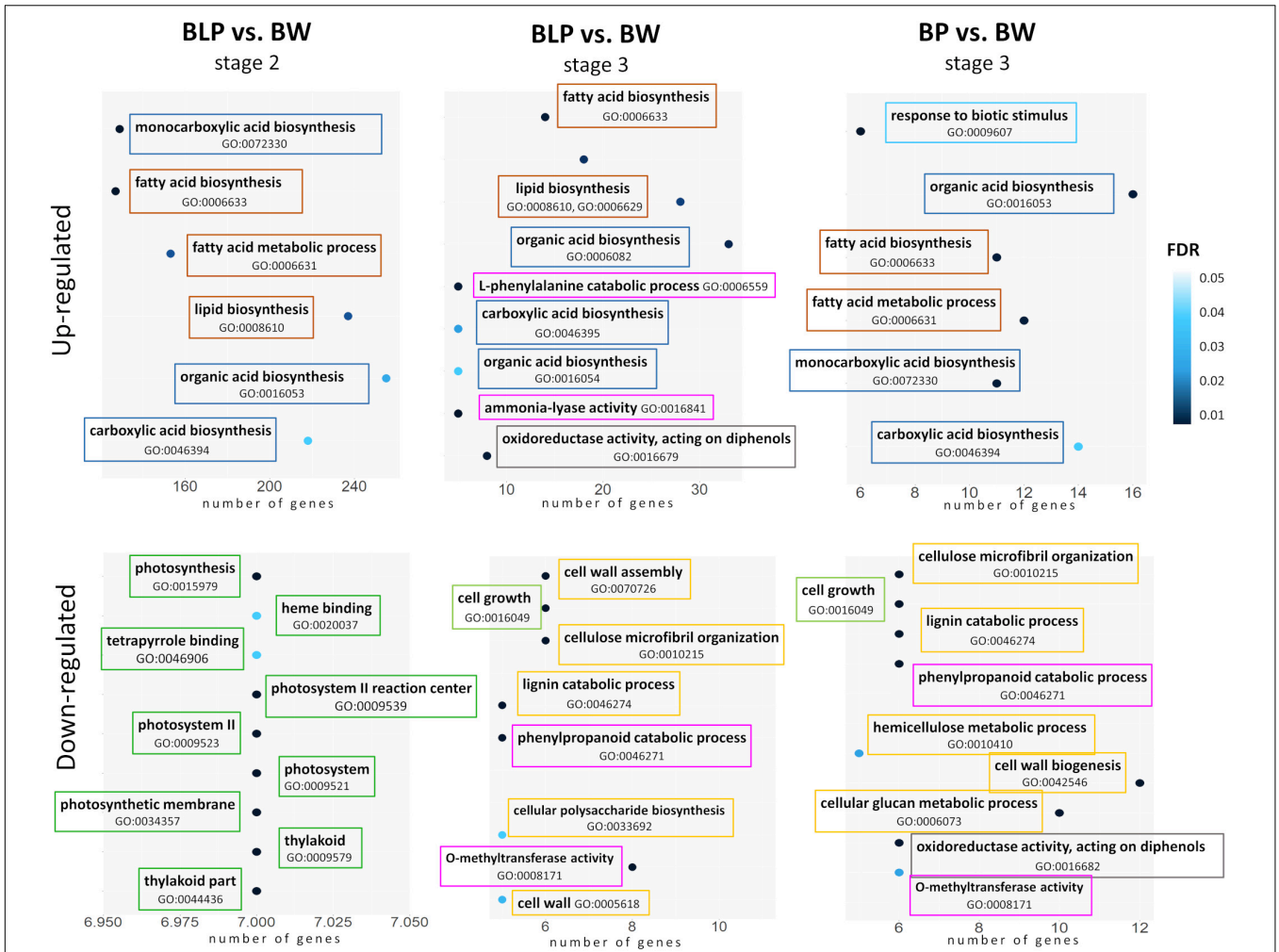


FIGURE 4 | GO term enrichment within lists of upregulated and downregulated DEGs in lines BLP and BP in comparison with Bowman (BW). Related metabolic processes, functions, and compartments are highlighted in the same color.

heatmap construction (**Figure 5A**). According the expression patterns, the genes can be subdivided into two groups: early and late genes. The early genes are activated at the booting stage (stage 1) in all four lines regardless of pigmentation. These include the following genes: *PAL*, *C4H*, *4CL*, *C3′H*, *CSE*, *CCoAMT*, *COMT*, *F5H*, *CCR*, *CHS*, *CHI*, and *FLS*. The late genes are transcribed at later stages of spike development (stages 2 and 3), and their expression proved to be line-specific and depended on the pigment type accumulated. In lines characterized by the presence of anthocyanins (PLP and BP), specific activation of the flavonoid biosynthesis pathway at the later stages was observed (**Figure 5A**). The pathway includes enzymes F3H, F3′H, and ANS. Furthermore, secondary activation of genes encoding CHS and CHI, which are implicated in the key steps of flavonoid biosynthesis, was registered in lines PLP and BP. It should be noted that in the BP line, these genes were found to be activated earlier, at stage 2, whereas in BP's parental line PLP, they are highly transcribed only at stage 3. Moreover, activation of some specific genes from the monolignol branch of the biosynthesis pathway was observed in lines PLP and BP, including gene *HORVU3Hr1G056560* encoding CCR, whereas the expression of gene *HORVU3Hr1G099990* encoding CCoAMT was noticed specifically in the PLP line (**Figure 5B**).

The expression of regulatory genes *Ant1*, *Ant2*, and *WD40* was investigated too; they are necessary for MBW complex assembly and anthocyanin biosynthesis in the grain pericarp. Specific activation of *Ant1* and *Ant2* was detectable in anthocyanin-accumulating lines; moreover, the expression pattern of the *Ant2* gene was the same as that of the late genes of flavonoid biosynthesis pathway, and activation timing was different between lines PLP and BP. By contrast, the expression of *WD40* was constant in all the lines at all examined stages (**Figure 5A**).

Similarly, the genes with specific activation in the presence of melanin pigments at stages 2 and 3 were identified in the phenylpropanoid pathway (**Figure 5B**). They are represented by specific genes from gene families encoding enzymes of the general phenylpropanoid pathway such as PAL (gene *HORVU2Hr1G089440*), and enzymes participating in

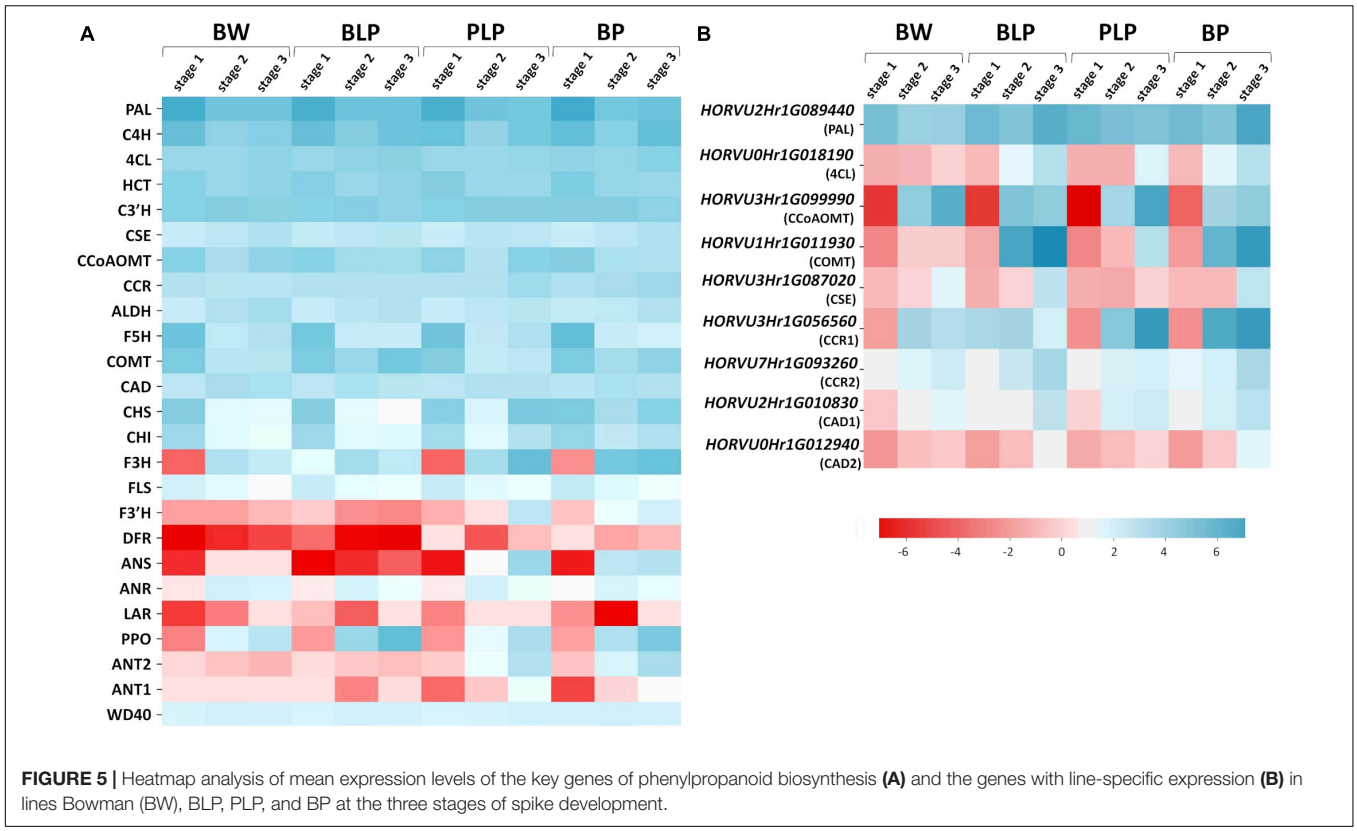


FIGURE 5 | Heatmap analysis of mean expression levels of the key genes of phenylpropanoid biosynthesis **(A)** and the genes with line-specific expression **(B)** in lines Bowman (BW), BLP, PLP, and BP at the three stages of spike development.

monolignol synthesis such as COMT (*HORVU1Hr1G011930*), CSE (*HORVU3Hr1G087020*), CCR (*HORVU3Hr1G056560* and *HORVU7Hr1G093260*), and CAD (*HORVU2Hr1G010830* and *HORVU0Hr1G012940*). In addition, genes encoding C4H and 4CL proved to be activated at stage 3 in all pigmented lines; what is more, in the case of *C4H*, there is secondary activation of this gene in response to *Ant1*, *Ant2*, and *Blp1* presence.

Given that melanin is a product of phenolic-compound polymerization, the expression of PPOs was also analyzed. It was observed that PPO genes are inactive at stage 1, and their expression increases during spike maturation in all the analyzed lines, whereas in BLP, these genes are most transcriptionally active (**Figure 5A**).

Quantitative Reverse-Transcription PCR Verification of Gene Expression

Expression of nine genes was confirmed by qRT-PCR. These genes code for o-methyltransferase (*HORVU1Hr1G011930*), polyphenol oxidases (*HORVU2Hr1G103040*, *HORVU2Hr1G103000*, *HORVU3Hr1G077790*, and *HORVU4Hr1G090870*), chloroplastic geranylgeranyl diphosphate reductase (*HORVU6Hr1G075900*), translation initiation factor SUI1 (*HORVU3Hr1G077960*), xyloglucan endotransglucosylase/hydrolase (*HORVU7Hr1G098280*), and chloroplastic photosynthetic NDH subunit of subcomplex B1 (*HORVU4Hr1G014010*). *In vitro* expression analysis of these genes was highly consistent with the *in silico* data (Spearman's correlation was 0.93), thus indicating the reliability of our RNA-seq data (**Figure 6**).

Phenylpropanoid Compounds' Levels in the Near-Isogenic Lines

Extraction of phenolic compounds with subsequent HPLC analysis was carried out at the 3rd stage of spike development, when the greatest of line-specific differences in gene expression were observed. The quantities of identified phenolic compounds and chromatographic profiles are presented in **Supplementary Table 6** and **Supplementary Figures 4–7**, respectively. The total contents of benzoic acids, hydroxycinnamic acids, flavonoids and anthocyanins are presented in **Figure 7A**. The total benzoic-acid content was significantly higher in melanin-containing NILs BLP (25.2 ± 0.3 μg/g) and BP (38.6 ± 4.6 μg/g) as compared with PLP (20.6 ± 0.7 μg/g) and *cv.* Bowman (21.5 ± 0.2 μg/g), whereas in the BP line, it was statistically significantly the highest among all the lines. Among the compounds in this family, protocatechuic, 4-hydroxybenzoic, and vanillic acids were identified. It is noteworthy that an elevated concentration of protocatechuic acid was found to be characteristic of lines PLP and BP, which accumulate anthocyanins; a significantly increased content of vanillic acid was also documented in the BP line, whereas in BLP, the increase was not significant (**Figure 7B**).

The highest level of hydroxycinnamic acids was registered in PLP (118.2 ± 5.9 μg/g); however, this level was also significantly higher in BP (108.7 ± 3.4 μg/g) than in BLP (78.8 ± 3.9 μg/g)

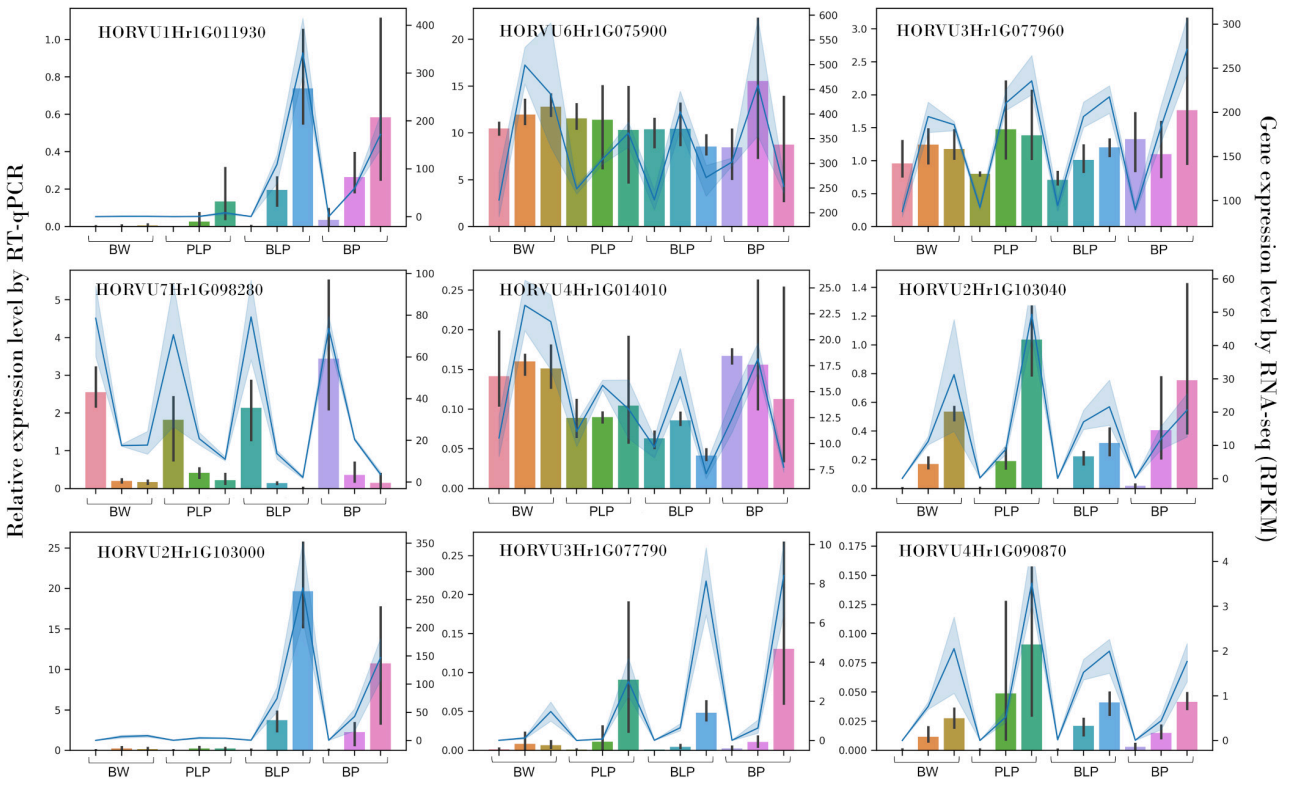


FIGURE 6 | qRT-PCR verification of nine DEGs in NILs (columns) at stages 1, 2, and 3 (left to right) and a comparison with our RNA-seq data (blue line).

and Bowman (82.2 ± 7.8 μg/g). In this family of compounds, caffeic acid, *p*-coumaric acid, *m*-coumaric acid, and ferulic acid were identified. Furthermore, four additional compounds were assigned to the family of cinnamic acids according to their characteristic UV spectra. Two patterns of these acids in the lines under study can be distinguished. The first one concerns caffeic acid accumulation and represents its significantly increased content in NILs with melanin. The second pattern concerns *p*-coumaric and ferulic acids: greater accumulation in lines PLP and BP (**Figure 7B**).

The flavonoid content in Bowman and BLP was 41.2 ± 2.3 and 25.3 ± 1.7 μg/g, respectively: dramatically different from the flavonoid content in PLP and BP (404.5 ± 16.5 and 213.8 ± 10.3 μg/g, respectively). In this family of compounds, luteolin, apigenin, and quercetin 3-rhamnoglucoside (rutin) were identified. Besides, a series of glycosides of luteolin and apigenin was identified by means of characteristic UV spectra. In the subfamily of luteolin derivatives, a flavonoid was found, which, according to the UV spectrum and retention time, can be luteolin methyl ester. This compound was not found in Bowman and BLP and is one of the major phenolic compounds in lines PLP and BP. Anthocyanins were not detectable in lines Bowman and BLP, but their presence in lines PLP and BP was confirmed. In the family of anthocyanins, cyanidin-3-glucoside and malvidin-3-glucoside were identified. It was noticed that the total concentration of flavonoids was ∼1.5 times higher in PLP than in BP; however, the total anthocyanin content was significantly greater in BP than in its parental line PLP (26.2 ± 1.7 and 18.0 ± 0.4 μg/g, respectively).

DISCUSSION

Phenolic compounds, including anthocyanin and melanin pigments, that can by synthesized in hulls and pericarp of barley grain, play diverse biochemical and molecular functions in plant physiology (Treutter, 2006; Vogt, 2010; Kiani et al., 2021; Gharaghanipor et al., 2022). They are not only protect the seeds from damage, deterioration, and pest but also are beneficial to human health (Rasouli et al., 2017). Unlike anthocyanins, melanins are a less studied family of pigments. One of obvious reasons for that is the difficulty with their extraction and chemical analysis. To date, to our knowledge, *p*-coumaric acid as a precursor of melanin synthesis has been quantified only in oat bracts (Varga et al., 2016). Here, comparative metabolic and transcriptomic assays of three barley NILs that differ in profiles of anthocyanin and melanin compounds in the grain were carried out to detect metabolic shifts in phenolic compounds including phenylpropanoids and thus identify the pathways that could be involved in melanin synthesis in barley.

Earlier, an increasing total phenolic content was documented in black-grained sesame in comparison with white-grained one (Zhou et al., 2016). By a comparative LC-MS/MS-based metabolome profiles analysis of NaOH and methanol

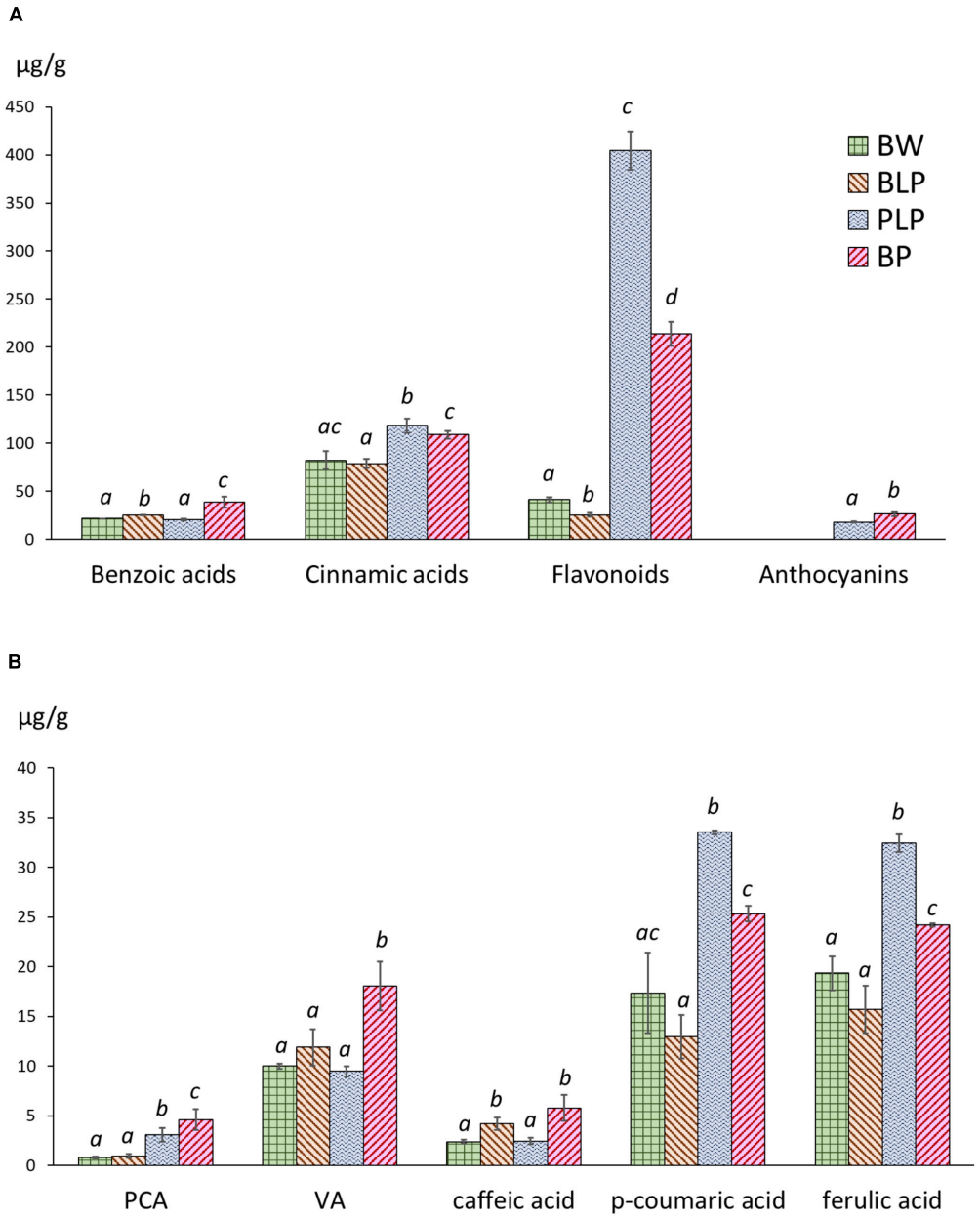


FIGURE 7 | (A) The total contents of benzoic acids, hydroxycinnamic acids, flavonoids, and anthocyanins in NILs at the early dough stage (stage 3) of spike maturity. **(B)** Concentrations of individual phenylpropanoids at stage 3. PCA, protocatechuic acid; VA, vanillic acid. The different letters denote statistically significant differences between the lines (*t* test, $p \leq 0.05$).

extracts of black and white sesame seeds, benzoic, caffeic, ferulic, homogentisic, indole-carboxylic, protocatechuic, and vanillic acids were identified as the potential precursors of the sesame melanin (Dossou et al., 2022). In the current work taking advantage of precise genetic models, one may state with more confidence that melanogenesis is associated with metabolic shifts of phenolic compounds and with significant alterations in phenylpropanoid metabolism. In barley line BLP, the total benzoic acid content of grain at the soft dough stage of ripeness proved to be higher in comparison with unpigmented-grain Bowman, and addition of the genome fragments of chromosomes 7HS and 2HL carrying, respectively, anthocyanin-specific genes *Ant1* and *Ant2* seemed to increase the benzoic-acid content in the BP line. Among benzoic acids, protocatechuic and vanillic acids were identified. Although the concentration of these compounds was not elevated in the BLP line in comparison with Bowman, the genome fragment of chromosome 1HL carrying *Blp1* probably significantly influences the levels of these compounds in the BP line in comparison with PLP.

Benzoic acids — whose content is elevated in BLP and BP — can be synthesized from intermediates of shikimate or phenylpropanoid pathways (Marchiosi et al., 2020). However, since both melanogenesis and the enzyme which catalyze the enzymatic browning reaction in damaged tissues and are believed to participate in melanin synthesis in intact tissues — PPO — are present in plastids (Boeckx et al., 2015, 2017; Shoeva et al., 2020), the benzoic acids originating from the shikimate pathway that occurs in plastids too are more probable precursors of melanin than compounds originating from the phenylpropanoid pathway that takes place in the cytoplasm (Widhalm and Dudareva, 2015). In barley, substrates for PPOs that could be precursors for melanin synthesis have been identified neither in the grain nor in damaged leaf tissue (Claussen and Pepper, 1968). However, in species related to barley — bread and durum wheat (belonging just as barley to the *Poaceae* family) —, 4-methylcatechol and catechol, respectively, were identified as substrates for PPOs (Jukanti, 2017). Since the nucleotide sequences of the wheats' PPOs are almost identical to the barley ones (Taketa et al., 2010), one may assume that barley PPOs can use catechol or its derivatives as substrates for melanin synthesis too. Whether barley melanin is consisted of these compounds should be clarified in the future.

Besides changes observed in the content of the simple phenolic compounds (that contain one phenol unit or a derivative of it) in grain at the soft dough stage, the metabolic shifts in the general phenylpropanoid metabolism at transcriptomic level were revealed. At the booting stage, genes of general phenylpropanoid metabolism are activated that are common for all the tested lines regardless of their pigmentation; however, as the spike matures, genes responsible for special metabolism are induced, and the number of DEGs specific for each pigmented line increases. Secondary activation of transcription of genes of the general phenylpropanoid pathway (*Pal*, *C4h*, and *4Cl)* together with monolignol synthesis genes (*Comt*, *Cse*, *Ccr*, and *Cad)*, but not flavonoid biosynthesis genes, was observed in the BLP line. This transcription pattern can explain the increased content of caffeic acid in this line, while the total content of flavonoids and anthocyanins turned out to be comparable to these values in Bowman. In contrast, activation of general phenylpropanoid genes *C4h* and *4Cl* together with anthocyanin biosynthesis genes *Chs*, *Chi*, *F3h*, *F3′h*, and *Ans* (occurring earlier in the BP line than in PLP), coincided with upregulation of hydroxycinnamic acids (ferulic and *p*-coumaric), total flavonoids, and anthocyanins in the *Ant1*/*Ant2*-carrying lines in comparison with uncolored Bowman. Given that the total flavonoid content was found to be lower in the BP line than in its parental line PLP, while the anthocyanin content in this line is greater in comparison with PLP, there is possible redirection of the flavonoid biosynthesis pathway when melanogenesis occurs.

Because the lines accumulating either anthocyanin or melanin manifested upregulation of the genes taking part in monolignol biosynthesis at late stages of spike development accompanied by increased concentrations of ferulic and caffeic acids, respectively, one may expect differences in lignin levels between these lines and Bowman. Analysis of the lignin content of barley grains of different colors was performed earlier (Choo et al., 2005). Using diverse genetic materials such as 96 doubled-haploid lines, 40 landraces, four F_3 bulked populations, and 10 NILs (including those investigated here: Bowman, PLP, and BLP), those authors showed that purple-grained lines do not differ in the lignin content from yellow-grained ones, whereas the black barley is characterized by more protein and a higher lignin content than yellow barley; the three barley samples with the highest concentration of lignin were black seeded (Choo et al., 2005).

Based on these results, we can hypothesize that greater production of benzoic acid in the melanin-accumulating lines may be caused by activation of the shikimate pathway during melanogenesis, which takes place in plastids too. Next, intermediates are introduced into that phenylpropanoid pathway, which is probably redirected either into lignin synthesis in melanin-accumulating lines or into anthocyanin synthesis if anthocyanin-specific regulatory genes are in a dominant state. Whether *Blp1* activates the shikimate pathway (which provides precursors for melanin synthesis and phenylpropanoid pathways) should be clarified in the future.

Besides effect of melanogenesis on phenylpropanoid metabolism, downregulation of genes associated with photosynthesis in stage 2 was revealed in the BLP line compared to Bowman. This finding is consistent with a recent report on oat bracts where the inhibition of photosynthesis was demonstrated at the flowering stage during melanin pigmentation development (Liu et al., 2021). In barley, underexpression of the genes related to photosynthesis was observed in the study at the late milk stage, before appearance of melanins in the grain. Because melanins are synthesized in aging chloroplasts (Shoeva et al., 2020), the effect of melanogenesis on photosynthesis was expected, and one can assume that plastid reorganization at the late milk stage (when chloroplasts turn into melanoplasts) causes the observed suppression of expression of the genes related to photosynthetic activity in the melanin-accumulating grain. At the same time, in the BP line, which accumulates anthocyanins and melanins, such underexpression of photosynthesis genes was not detectable at any stages. What may be related to the protective effect of anthocyanins on photosynthetic apparatus is their antioxidant capacity, which can defend tissues against the reactive oxygen species generated in the chloroplast (Guidi et al., 2016; Mushtaq et al., 2016). In the melanin-accumulating lines downregulation of the photosynthesis-related genes was accompanied by intensification of the fatty acid biosynthesis and lipid biosynthetic process at later developmental stages. These processes may occur in melanoplasts where synthesis of melanin takes place. However, in opposite to another type of aging plastids — gerontoplasts —, where degradation of thylakoid lipids and releasing the fatty acids have been observed (Biswal et al., 2003), in melanoplasts, active biogenesis of these chemicals that are components of cell membranes, including thylakoids, may take place for melanin sequestration in cells. Indeed, the presence of membrane-delimited structures in melanin-accumulating lines was observed up to the hard-dough stage when the plastids were degraded totally in non-pigmented lines (Mursalimov et al., 2022).

CONCLUSION

Here, with the help of NILs as a genetic model, a significant impact of melanogenesis on metabolic processes that take place in barley grain was revealed. Downregulation of the photosynthesis-related genes, accompanied by intensification of the fatty acid biosynthesis, lipid biosynthetic processes and significant changes in the phenolic metabolism was showed in the melanin-accumulating lines. These lines showed an increased total level of benzoic acids, among them, protocatechuic and vanillic acids were identified. By contrast, anthocyanin-accumulating lines showed higher concentrations of flavonoids and *p*-coumaric and ferulic acids. Based on the data the simple phenolic compounds that synthesized in plastids where melanin synthesis occurs too were suggested as likely precursors of melanin in barley.

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/, PRJNA832149.

AUTHOR CONTRIBUTIONS

AG contributed to interpretation of the data and study conception and wrote the original draft of the manuscript. AV and NAS performed the analysis of the transcriptomic data and participated in the drafting of the manuscript. SM and EC performed the extraction and analysis of phenolic compounds and participated in the drafting of the manuscript. GV and NVS prepared the RNA-seq libraries and carried out the sequencing of the libraries on the Illumina platform. EK participated in interpretation of the data and revised the manuscript critically. OS conceived the study and contributed to its design and coordination, interpretation of the data, and manuscript writing. All authors reviewed and edited the manuscript.

FUNDING

This study was funded by the Russian Science Foundation, grant number 21-76-10024. The cultivation of the barley plants at the Greenhouse Facility was supported by ICG project FWNR-2022-0017.

ACKNOWLEDGMENTS

The English language was corrected and certified by shevchuk-editing.com.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.923717/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Biswal, U. C., Biswal, B., and Raval, M. K. (2003). "Transformation of Chloroplast to Gerontoplast," in *Chloroplast Biogenesis*, eds U. C. Biswal, B. Biswal, and M. K. Raval (Dordrecht: Springer), doi: 10.1007/978-94-017-0247-8_4

Boeckx, T., Winters, A., Webb, K. J., and Kingston-Smith, A. H. (2017). Detection of potential chloroplastic substrates for polyphenol oxidase suggests a role in undamaged leaves. *Front. Plant Sci.* 8:237. doi: 10.3389/fpls.2017.00237

Boeckx, T., Winters, A. L., Webb, K. J., and Kingston-Smith, A. H. (2015). Polyphenol oxidase in leaves: is there any significance to the chloroplastic localization?. *J. Exp. Bot.* 66, 3571–3579. doi: 10.1093/jxb/erv141

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/BIOINFORMATICS/BTU170

Britton, G. (1983). *The Biochemistry of Natural Pigments*. Cambridge: Cambridge University Press, 366.

Choo, T. M., Vigier, B., Ho, K. M., Ceccarelli, S., Grando, S., and Franckowiak, J. D. (2005). Comparison of black, purple, and yellow barleys. *Genet. Resour. Crop Evol.* 52, 121–126. doi: 10.1007/s10722-003-3086-4

Claussen, K. A., and Pepper, E. H. (1968). An examination of the brown pigments from barley leaves. *Cereal Chem.* 45, 124–132.

Costa, J. M., Corey, A., Hayes, P. M., Jobet, C., Kleinhofs, A., Kopisch-Obusch, A., et al. (2001). Molecular mapping of the Oregon Wolfe Barleys: a phenotypically polymorphic doubled-haploid population. *Theor. Appl. Genet.* 103, 415–424. doi: 10.1007/s001220100622

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

Di Mauro, E., Xu, R., Soliveri, G., and Santato, C. (2017). Natural melanin pigments and their interfaces with metal ions and oxides: emerging concepts and technologies. *MRS Commun.* 7, 141–151. doi: 10.1557/mrc.2017.33

d'Ischia, M., Wakamatsu, K., Cicoira, F., Di Mauro, E., Garcia-Borron, J. C., Commo, S., et al. (2015). Melanins and melanogenesis: from pigment cells to human health and technological applications. *Pigment. Cell Melanoma Res.* 28, 520–544. doi: 10.1111/pcmr.12393

Dossou, S. S. K., Luo, Z., Wang, Z., Zhou, W., Zhou, R., Zhang, Y., et al. (2022). The dark pigment in the sesame (*Sesamum indicum* L.) seed coat: isolation, characterization, and its potential precursors. *Front. Nutr.* 9:858673. doi: 10.3389/fnut.2022.858673

Druka, A., Franckowiak, J., Lundqvist, U., Bonar, N., Alexander, J., Houston, K., et al. (2011). Genetic dissection of barley morphology and development. *Plant Physiol.* 155, 617–627. doi: 10.1104/pp.110.166249

Fritz, E., and Saukel, J. (2011). Secretory structures of subterranean organs of some species of the Cardueae and their diagnostic value. *Acta Biol. Crac. Ser. Bot.* 53, 62–72. doi: 10.2478/v10182-011-0009-0

Fukuda, A., Shimizu, H., Shiratsuchi, H., Yamaguchi, H., Ohdaira, Y., and Mochida, H. (2012). Complementary genes that cause black ripening hulls in F 1 plants of crosses between indica and japonica rice cultivars. *Plant Prod. Sci.* 15, 270–273. doi: 10.1626/pps.15.270

Gharaghanipor, N., Arzani, A., Rahimmalek, M., and Ravash, R. (2022). Physiological and transcriptome indicators of salt tolerance in wild and cultivated barley. *Front. Plant Sci.* 13:819282. doi: 10.3389/fpls.2022.819282

Glagoleva, A., Kukoeva, T., Mursalimov, S., Khlestkina, E., and Shoeva, O. (2022). Effects of combining the genes controlling anthocyanin and melanin synthesis in the barley grain on pigment accumulation and plant development. *Agronomy* 12:112. doi: 10.3390/agronomy12010112

Glagoleva, A. Y., Shmakov, N. A., Shoeva, O. Y., Vasiliev, G. V., Shatskaya, N. V., Börner, A., et al. (2017). Metabolic pathways and genes identified by RNA-seq analysis of barley near-isogenic lines differing by allelic state of the Black lemma and pericarp (Blp) gene. *BMC Plant Biol.* 17:182. doi: 10.1186/s12870-017-1124-1

Gordeeva, E. I., Glagoleva, A. Y., Kukoeva, T. V., Khlestkina, E. K., and Shoeva, O. Y. (2019). Purple-grained barley (*Hordeum vulgare* L.): marker-assisted development of NILs for investigating peculiarities of the anthocyanin biosynthesis regulatory network. *BMC Plant Biol.* 19:52. doi: 10.1186/s12870-019-1638-9

Gracheva, N. V., and Zheltobryukhov, V. F. (2019). Sorption properties of sunflower husk melanins. *Pharm. Chem. J.* 53, 337–341. doi: 10.1007/s11094-019-02002-2

Guidi, L., Brunetti, C., Fini, A., Agati, G., Ferrini, F., Gori, A., et al. (2016). UV radiation promotes flavonoid biosynthesis, while negatively affecting the biosynthesis and the de-epoxidation of xanthophylls: consequence for photoprotection?. *Environ. Exp. Bot.* 127, 14–25. doi: 10.1016/J.ENVEXPBOT.2016.03.002

Himi, E., Nisar, A., and Noda, K. (2005). Colour genes (*R* and *Rc*) for grain and coleoptile upregulate flavonoid biosynthesis genes in wheat. *Genome* 48, 747–754. doi: 10.1139/g05-026

Jukanti, A. (2017). *Polyphenol Oxidases (PPOs) in Plants*. Berlin: Springer, doi: 10.1007/978-981-10-5747-2

Kablov, V. F., Novopoltseva, O. M., Gracheva, N. V., Zheltobryukho, V. F., and Dao, P. K. (2019). Prospects of application of melanins as antiaging agents in elastomer compositions. *Vietnam J. Chem.* 57, 255–260. doi: 10.1002/vjch.201960024

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/J.JMB.2015.11.006

Kiani, R., Arzani, A., and Mirmohammady Maibody, S. A. M. (2021). Polyphenols, flavonoids, and antioxidant activity involved in salt tolerance in wheat, *Aegilops cylindrica* and their amphidiploids. *Front. Plant Sci.* 12:646221. doi: 10.3389/fpls.2021.646221

Korpacheva, S., Serasutdinova, K., Lomovsky, I., and Chugunova, O. (2021). Technological aspects of obtaining melanin and powder from buckwheat hull and their use in food technology. *E3S Web Conf.* 296:07007. doi: 10.1051/e3sconf/202129607007

Li, B., Lu, X., Gebremeskel, H., Zhao, S., He, N., Yuan, P., et al. (2020). Genetic mapping and discovery of the candidate gene for black seed coat color in watermelon (*Citrullus lanatus*). *Front Plant Sci.* 10:1689. doi: 10.3389/fpls.2019.01689

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/BIOINFORMATICS/BTT656

Lin, H.-N., and Hsu, W.-L. (2018). DART: a fast and accurate RNA-seq mapper with a partitioning strategy. *Bioinformatics* 34, 190–197. doi: 10.1093/BIOINFORMATICS/BTX558

Liu, B., Zhang, D., Sun, M., Li, M., Ma, X., Jia, S., et al. (2021). PSII activity was inhibited at flowering stage with developing black bracts of oat. *Int. J. Mol. Sci.* 22:5258. doi: 10.3390/ijms22105258

Long, Z., Jia, Y., Tan, C., Zhang, X.-Q., Angessa, T., Broughton, S., et al. (2019). Genetic mapping and evolutionary analyses of the black grain trait in barley. *Front. Plant Sci.* 9:1921. doi: 10.3389/fpls.2018.01921

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/S13059-014-0550-8

Lusa, M. G., Loeuille, B. F. P., and Appezzato-da-Glória, B. (2018). First record of phytomelanin in aerial vegetative organs and its evolutionary implications in Lychnophorinae (*Vernonieae*: *Asteraceae*). *Perspect Plant Ecol. Evol. Syst.* 33, 18–33. doi: 10.1016/j.ppees.2018.04.006

Marchiosi, R., dos Santos, W. D., Constantin, R. P., de Lima, R. B., Soares, A. R., Finger-Teixeira, A., et al. (2020). Biosynthesis and metabolic actions of simple phenolic acids in plants. *Phytochem. Rev.* 19, 865–906. doi: 10.1007/s11101-020-09689-2

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043

Mursalimov, S., Glagoleva, A., Khlestkina, E., and Shoeva, O. (2022). Chlorophyll deficiency delays but does not prevent melanogenesis in barley seed melanoplasts. *Protoplasma* 259, 317–326. doi: 10.1007/s00709-021-01669-3

Mushtaq, M. A., Pan, Q., Chen, D., Zhang, Q., Ge, X., and Li, Z. (2016). Comparative leaves transcriptome analysis emphasizing on accumulation of anthocyanins in brassica: molecular regulation and potential interaction with photosynthesis. *Front. Plant Sci.* 7:311. doi: 10.3389/fpls.2016.00311

Nicolas, J. J., Richard-Forget, F. C., Goupy, P. M., Amiot, M., and Aubert, S. Y. (1994). Enzymatic browning reactions in apple and apple products. *Crit. Rev. Food Sci. Nutr.* 34, 109–157. doi: 10.1080/10408399409527653

Pandey, A. K., and Dhakal, M. R. (2001). Phytomelanin in Compositae. *Curr. Sci.* 80, 933–940.

Rasouli, H., Farzaei, M. H., and Khodarahmi, R. (2017). Polyphenols and their benefits: a review. *Int. J. Food Prop.* 20, 1700–1741. doi: 10.1080/10942912.2017.1354017

Riley, P. (1997). Melanin. *Int. J. Biochem. Cell Biol.* 29, 1235–1239. doi: 10.1016/s1357-2725(97)00013-7

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Shoeva, O. Y., Mock, H.-P., Kukoeva, T. V., Börner, A., and Khlestkina, E. K. (2016). Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of *Hordeum vulgare*. *PLoS One* 11:e0163782. doi: 10.1371/journal.pone.0163782

Shoeva, O. Y., Mursalimov, S. R., Gracheva, N. V., Glagoleva, A. Y., Börner, A., and Khlestkina, E. K. (2020). Melanin formation in barley grain occurs within plastids of pericarp and husk cells. *Sci. Rep.* 10:179. doi: 10.1038/s41598-019-56982-y

Solano, F. (2014). Melanins: skin Pigments and Much More—Types, Structural Models, Biological Functions, and Formation Routes. *New J. Sci.* 2014, 1–28. doi: 10.1155/2014/498276

Strygina, K. V., Börner, A., and Khlestkina, E. K. (2017). Identification and characterization of regulatory network components for anthocyanin synthesis in barley aleurone. *BMC Plant Biol.* 17:184. doi: 10.1186/s12870-017-1122-3

Taketa, S., Matsuki, K., Amano, S., Saisho, D., Himi, E., Shitsukawa, N., et al. (2010). Duplicate polyphenol oxidase genes on barley chromosome 2H and their functional differentiation in the phenol reaction of spikes and grains. *J. Exp. Bot.* 61, 3983–3993. doi: 10.1093/jxb/erq211

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/NAR/GKX382

Treutter, D. (2006). Significance of flavonoids in plant resistance: a review. *Environ. Chem. Lett.* 4:147. doi: 10.1007/s10311-006-0068-8

Vahidzadeh, E., Kalra, A. P., and Shankar, K. (2018). Melanin-based electronics: from proton conductors to photovoltaics and beyond. *Biosens. Bioelectron.* 122, 127–139. doi: 10.1016/J.BIOS.2018.09.026

Varga, M., Berkesi, O., Darula, Z., May, N. V., and Palágyi, A. (2016). Structural characterization of allomelanin from black oat. *Phytochemistry* 130, 313–320. doi: 10.1016/j.phytochem.2016.07.002

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

von Zitzewitz, J., Szûcs, P., Dubcovsky, J., Yan, L., Francia, E., Pecchioni, N., et al. (2005). Molecular and structural characterization of barley vernalization genes. *Plant Mol. Biol.* 59, 449–467. doi: 10.1007/s11103-005-0351-2

Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front. Plant Sci.* 7:1491. doi: 10.3389/fpls.2016.01491

Wang, L., Dossou, S. S. K., Wei, X., Zhang, Y., Li, D., Yu, J., et al. (2020). Transcriptome dynamics during black and white sesame (*Sesamum indicum* L.) seed development and identification of candidate genes associated with black pigmentation. *Genes* 11:1399. doi: 10.3390/genes11121399

Wang, L.-F., and Rhim, J.-W. (2019). Isolation and characterization of melanin from black garlic and sepia ink. *LWT* 99, 17–23. doi: 10.1016/j.lwt.2018.09.033

Widhalm, J. R., and Dudareva, N. (2015). A familiar ring to it: biosynthesis of plant benzoic acids. *Mol. Plant* 8, 83–97. doi: 10.1016/j.molp.2014.12.001

Zhou, L., Lin, X., Abbasi, A. M., and Zheng, B. (2016). Phytochemical contents and antioxidant and antiproliferative activities of selected black and white sesame seeds. *Biomed. Res. Int.* 2016:8495630. doi: 10.1155/2016/8495630

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

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

frontiers | Frontiers in Plant Science

REVIEW
published: 14 July 2022
doi: 10.3389/fpls.2022.900035

Application of High-Throughput Sequencing on the Chinese Herbal Medicine for the Data-Mining of the Bioactive Compounds

Xiaoyan Liu[1], Xun Gong[2], Yi Liu[1,3], Junlin Liu[1], Hantao Zhang[1], Sen Qiao[1], Gang Li[4] and Min Tang[1]**

[1]School of Life Sciences, Jiangsu University, Zhenjiang, China, [2]Department of Rheumatology and Immunology, Affiliated Hospital of Jiangsu University, Zhenjiang, China, [3]Institute of Animal Husbandry, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [4]Department of Vascular Surgery, The Second Affiliated Hospital of Shandong First Medical University, Taian, China

OPEN ACCESS

Edited by:
Jian Li Yang, Zhejiang University, China

Reviewed by:
Niveshika, Banaras University, India
Yanqiang Li, Boston Children's Hospital and Harvard Medical School, United States

***Correspondence:**
Min Tang
mt3138@ujs.edu.cn
Gang Li
ligang111666@163.com

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

Received: *21 March 2022*
Accepted: *10 June 2022*
Published: *14 July 2022*

Citation:
Liu X, Gong X, Liu Y, Liu J, Zhang H, Qiao S, Li G and Tang M (2022) Application of High-Throughput Sequencing on the Chinese Herbal Medicine for the Data-Mining of the Bioactive Compounds. Front. Plant Sci. 13:900035. doi: 10.3389/fpls.2022.900035

The Chinese Herbal Medicine (CHM) has been used worldwide in clinic to treat the vast majority of human diseases, and the healing effect is remarkable. However, the functional components and the corresponding pharmacological mechanism of the herbs are unclear. As one of the main means, the high-throughput sequencing (HTS) technologies have been employed to discover and parse the active ingredients of CHM. Moreover, a tremendous amount of effort is made to uncover the pharmacodynamic genes associated with the synthesis of active substances. Here, based on the genome-assembly and the downstream bioinformatics analysis, we present a comprehensive summary of the application of HTS on CHM for the synthesis pathways of active ingredients from two aspects: active ingredient properties and disease classification, which are important for pharmacological, herb molecular breeding, and synthetic biology studies.

Keywords: Chinese herbal medicine, high-throughput sequencing technologies, genome assembly, bioactive compounds, functional genes

INTRODUCTION

The Chinese herbal medicine (CHM) makes a great contribution to the human healthcare and clinical therapy due to its remarkable efficacy and fewer side effects (Flower et al., 2012; Lu et al., 2019; Xiang et al., 2019; Zheng et al., 2020). Since the Qin and Han dynasties, the Chinese ancestors had made the natural plants to cure patients without the knowledge of the chemical constituents, which gradually formed the later mature system of CHM with clarified the properties (He et al., 2015). *Artemisia annua*, an ancient medicine, grows broadly in China, the province such as, Jiangsu, Shanxi, Guangdong. Artemisinin, the main medicinal ingredient of *A. annua*, is world-famous for its treatment of malaria. Moreover, both traditional and modern pharmaceutical research imply it has anti-inflammatory, anti-viral, and anti-cancer effects (Nair et al., 2019; Feng et al., 2020; Liu H. et al., 2021). Accumulated evidence suggests that the bioactive components originated from CHM play a non-negligible role in the treatment of diseases (Li et al., 2016; Ji et al., 2019). However, the low-abundance is insufficient to meet the clinical requirements, such as paclitaxel, a well-known natural anti-cancer drug (Sabzehzari

et al., 2020; Xiong et al., 2021). Nowadays, with the continuous advances in sequencing technologies for the fine genome assembly, the synthesis of bioactive components, or called cell-Bio-fluid sync, can be well elucidated, which provides abundant genetic resources for life and pharmaceutical sciences.

Direct application of the conventional Sanger sequencing (called first-generation sequencing) on the herbs with large and complicated genomes is grudging because of the low throughput and accuracy (Lo and Shaw, 2019). Instead of it, the next-generation sequencing (NGS, also called second-generation sequencing) was gradually applied since 2010 (Cheng Q. et al., 2021). NGS can perform millions to billions of independent sequencing processes, dramatically increasing sequence output, which includes Illumina Solexa, Roche454, and ABI SOLiD platform. Based on the principle of reversible termination and fluorescently labeled dNTP, Illumina Solexa is sequencing while synthesizing (Guo et al., 2021a). It also has certain drawbacks, such as short read length (usually 200-800 bp), base mismatches, GC preference, and template migration during PCR amplification, which might influence the accuracy and integrity of sequencing data (Cheng Q. et al., 2021; Guo et al., 2021a). Subsequently, the third-generation sequencing (TGS) stands out for the high-throughput sequencing (HTS) technologies as a routine method. Oxford Nanopore and PacBio single-molecule real-time (SMRT) sequencing technology are now the main TGS platforms. Although SMRT can achieve read lengths of 100 kb, additional factors like template breakage, enzyme denaturation, and short library sequences can affect read lengths and accuracy. Unlike SMART sequencing, the Nanopore read length is determined by the length of the DNA molecules to be sequenced rather than the sequencing technique (van Dijk et al., 2018). Here, we mainly reviewed relevant articles from the recent decades and presented a comprehensive summary of the application of HTS on CHM for the synthesis pathways of active ingredients from two aspects: active ingredient properties and disease classification.

DE NOVO GENOME ASSEMBLY OF HERBS AND BIOINFORMATICS ANALYSIS

The last 5 years have been an era of expansion in medicinal plants genome sequencing, with 2nd and 3rd generation technologies combining to assure long read length, high throughput, and reasonable sequencing price for medicinal plant genomes assembled to the chromosome level. The employment of HTS technologies on data-mining bioactive compounds is shown in **Figure 1**. This strategy integrates genomics, transcriptomics, and metabolomics data to analyze genomic properties, synthetic and metabolic pathways of bioactive constituents, the overall transcriptional activity of organisms, and pathway regulatory mechanisms that will be revealed to uncover functional genes (see **Table 1**). The *Panax notoginseng* genome, for example, has been assembled in five versions. Chen et al. (2017) and Zhang D. et al. (2017) used Illumina technology to assemble a sketch of the *P. notoginseng* genome, but the assembly was highly fragmented. Fan G. et al. (2020) employed Pacbio and Oxford Nanopore technologies to assemble the genome to the chromosome level in 2020, with significantly improved assembly continuity. The last two versions were assembled more accurately by using a combination of 2nd and 3rd generation sequencing (Jiang Z. et al., 2021; Yang Z. et al., 2021). In contrast to the study by Jiang Z. et al. (2021) and Yang Z. et al. (2021) illustrated dencichine biosynthesis, the other major bioactive compounds derived from *P. notoginseng*. *Vaccinium darrowi*, the diploid blueberry, benefits cardiovascular, neural, and retinal. Cui et al. (2022) obtained a *de novo* genome assembly for *V. darrowi* according to Oxford Nanopore, Illumina short reads, and Hi-C data. *Sapindus mukorossi*, an environmentally herb, has been used for treating inflammatory conditions owing to its abundant active compounds. Xue et al. (2022) revealed the first reference genome sequence of *S. mukorossi*. Li et al. (2022) assembled the high-quality reference genome of *Gentiana dahurica* using Nanopore long reads, Illumina short reads, and Hi-C technologies, which is the first chromosome-level genome of Gentianaceae. Based on comparative genomic and transcriptome analyses, *cytochrome P450* candidate genes related to gentiopicroside biosynthesis were identified. Noteworthy, Wang et al. (2022) employed genome-wide association studies (GWAS) and identified six quantitative trait loci (QTLs) related to fruit traits according to the latest version genome of *Dimocarpus longan* at the chromosome level with 455.5 Mb assembled into 15 chromosomes. Based on the genome-assembly and the downstream bioinformatics analysis, these articles weremainly focusing on providing new insight for the discovery of novel drug candidates in CHM.

APPLICATION OF HIGH THROUGHPUT SEQUENCING IN BIOACTIVE COMPOUNDS DISCOVERY

Based on the accumulation in the sequencing field, many useful bioactive compounds and their varieties have been screened out from the complex mixtures and the clinic effects have been validated (see **Figure 2**; **Supplementary Figure S1**). Specially, the secondary metabolites constitute the backbone of many drugs as the active ingredients of the medicinal plants and are widely used in pharmaceutical products. In recent years, due to the innovation of sequencing technology, the HTS accelerates the study of secondary metabolites biosynthetic in the medicinal plants, which indirectly expands the global commercial market of the herb products (Barbosa et al., 2019). In general, the secondary metabolites are divided into seven major groups, namely flavonoid, terpenoid, alkaloid, phenylpropanoid, quinone, tannin, and steroid (Lo and Shaw, 2019; Erb and Kliebenstein, 2020).

Terpenoids

Terpenoids are hydrocarbon compounds consisting of isoprenoid as structural units (Chen et al., 2011). There are two synthetic

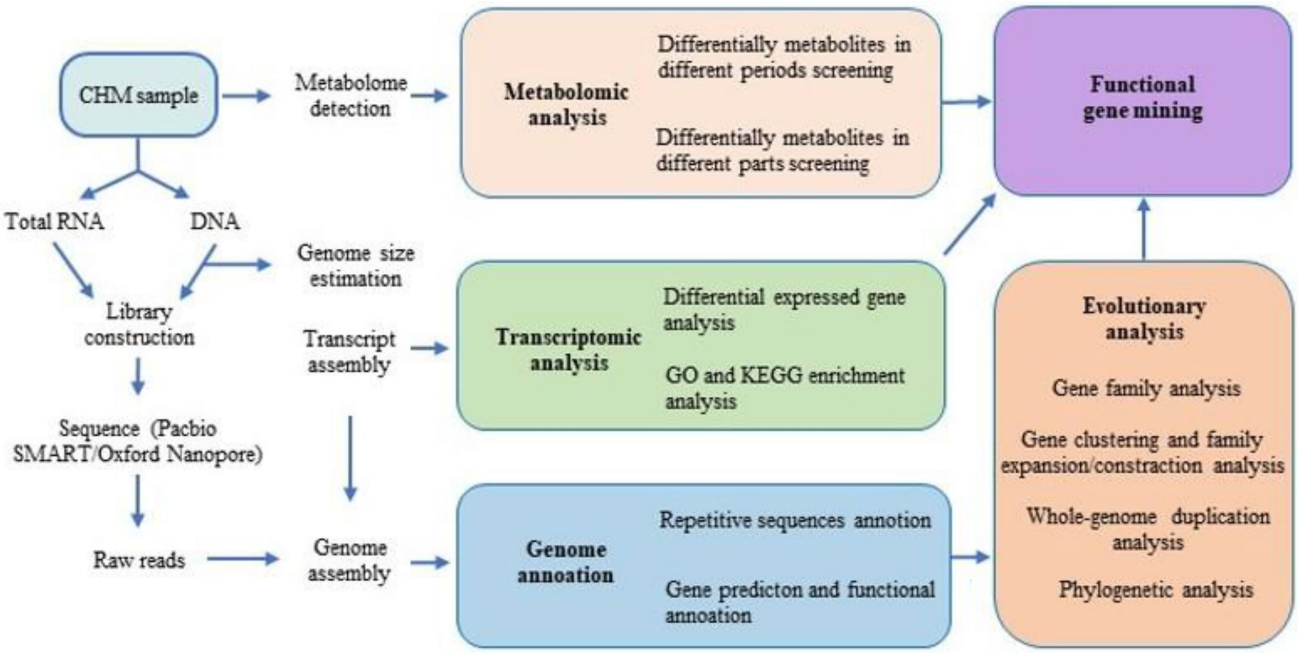


FIGURE 1 | Flow chart of bioactive compound related functional genes discovering based on high-throughput sequencing (HTS).

pathways for terpenoid in medicinal plants: MVA pathway and DOXP/MEP pathway. In the MVA pathway, IPP is condensed to generate DAMP by IDI, and FDPS converts IPP and DMAPP to FPP, which are the common precursors of sesquiterpenes, triterpene, and sterol (Lichtenthaler, 1999). Monoterpene, diterpene, and tetraterpene produce in the MEP pathway (Rohmer, 1999). The first rate-limiting enzyme in the MVA pathway, 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), is critical for its regulation (Korth et al., 1997). Cytochrome P450 hydroxylases (CYP450) are used to structurally modify terpene end-products to eventually become a large number of natural products. Tu et al. (2020) reported a reference genome of *Tripterygium wilfordii* using PacBio long reads, 10X Genomics, and high throughput chromosome conformation capture (Hi-C) data and annotated 28,321 protein-coding genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses showed that 951 *T. wilfordii*-specific genes were especially enriched in terpene synthases. According to the gene-to-metabolite network, 57 *CYP* genes may be involved in the triptolide biosynthesis pathway were identified, of which 10 candidate genes with tissue-specific were selected for further functional validation. RNAi studies found that the transcript levels of four candidate genes (CYP728B70, TW011445.1, TW012149.1, and TW006625.1) were decreased and triptolide accumulation showed the same results. Additionally, based on substrate feeding and overexpression studies, the findings showed that CYP728B70 was involved in triptolide biosynthesis. This finding provides new insight into triptolide biosynthesis and the theoretical foundation for heterologous bioproduction (Tu et al., 2020). Xu et al. (2017) used the Illumina Hiseq X-Ten platform to sequence the genome and the transcriptome of *Panax ginseng* and predicted 42,006 protein-coding genes. Weighted gene coexpression network analysis (WGCNA) obtained 15,762 genes positively associated with ginsenosides, which are produced by the precursor IPP *via* the MVA pathway. BLAST search discovered 31 genes were linked to 10 upstream enzymes. Finally, eight genes encoding HMGRs found that they may perform different roles during ginseng development. This is of great significance for future studies on ginseng breeding and synthetic biology. Xia et al. (2018) reported a high-quality draft genome of *Siraitia grosvenorii* using SMRT sequencing *via* PacBio platform and Illumina paired-end reads. According to the genomic information, 127 candidate genes were found in the mogrosides biosynthesis pathway, including SQEs, EPHs, CYP450s, and UGTs. In addition, mogrosides are accumulated during the development of *S. grosvenorii* fruit. Up-regulated genes in fruit development were found significantly enriched in the sesquiterpenoid and triterpenoid biosynthesis pathways through RNA-seq data and KEGG analysis. This study for *S. grosvenorii* genome assembly and annotation will contribute to the discovery of new functional genes. Chen et al. performed PacBio sequencing technologies to construct the first full-length transcriptome of *Pogostemon cablin* and annotated 102 transcripts related to patchoulol biosynthesis. Patchoulo, the main bioactive compounds, among the 39 chemical compositions in *P. cablin* were detected by GS-MS analysis. Patchoulol synthase (*PatPTS*) converts famesy1-pp (FPP) to patchouli alcohol (PA). Furthermore, based on the *P. cablin* full-length transcriptome and transcriptome profiling under MeJA treatment, 427 DEGs were up-regulated in leaves after MeJA treatment, of which *HMGR*, *DXS*, *HDR*, *IDI*, *FDPS*, *PatPTS* genes related to patchouli biosynthesis were up-regulated under MeJA treatment and

TABLE 1 | Summary of the bioactive compounds related functional genes from the assembly Chinese herbal medicines (CHM) genome.

Name	Chinese name	Functional genes	Validation methods	Download*	References
Coptis chinensis	黄连	*Cch00017825*	Enzymatic assays	PRJNA662860	Chen D. et al., 2021; Liu Y. et al., 2021
Panax ginseng	人参	*UGT94Q2*, *HMGRs*	/	PRJNA385956	Chen et al., 2017; Xu et al., 2017; Zhang D. et al., 2017; Xue et al., 2019; Fan G. et al., 2020; Zhang et al., 2020; Jiang Z. et al., 2021; Yang Z. et al., 2021
Taxus chinensis	红豆杉	*CYP725As*	Enzymatic assays, qRT-PCR	PRJNA730337	Xiong et al., 2021
Tripterygium wilfordii	雷公藤	*CYP728B70*	Construct RNAi lines	PRJNA542587	Tu et al., 2020
Dioscorea zingiberensis	山药	*CYP90G6*, *CYP94D144*	Metabolic engineering in yeast	PRJNA541739	Cheng J. et al., 2021
Salvia miltiorrhiza	丹参	*CYP71D375*	Construct RNAi lines, enzymatic assays	PRJNA682867	Ma Y. et al., 2021
Carthamus tinctorius	红花	*CtAH08G60157600*	/	PRJNA642978	Wu et al., 2021
Lithospermum erythrorhizon	紫草	*PGT1*	Construct RNAi lines	PRJNA596998	Auber et al., 2020
Senna tora	决明	*CHS-L9*	Enzymatic assays	PRJNA605066	Kang et al., 2020b
Dendrobium officinale	铁皮石斛	*CYP71D*, *CYP75A*, *CYP75B*, *MWL1*	/	PRJNA662181	Niu et al., 2021
Glycyrrhiza uralensis	甘草	*UGT73B1*, *CYP72A9*	/	PRJDB3943	Mochida et al., 2017
Platycodon grandiflorus	桔梗	*CYP716A*, *bAS*	/	PRJNA656905	Kim et al., 2020; Yu H. et al., 2021
Artemisia annua	青蒿	*HMGR*(AA201470), *FPS*(AA043570, AA174930), *DBR2*(AA04970,AA049710)	Construct the overexpressing transgenic lines	PRJNA416223	Shen et al., 2018
Cannabis sativa	大麻	*THCAS*, *CBDAS*	/	PRJNA562042	van Bakel et al., 2011; Hurgobin et al., 2021
Ophiorrhiza pumila	短小蛇根草	*OG0015245*	/	PRJDB8685	Rai et al., 2021
Zingiber officinale	生姜	*C3OMT2*, *C3OMT3*, *C3OMT13*	/	PRJNA647255	Li H. et al., 2021
Camptotheca acuminata	喜树	*CarGene13172*, *CarGene10888*	/	PRJNA639006	Kang et al., 2021
Andrographis paniculata	穿心莲	*UGT73AU1*	Enzymatic assays	PRJNA549104	Li L. et al., 2017, 2019; Sun et al., 2019; Yue et al., 2019

The fifth column is the Bioproject accession number in the NCBI database.

positively correlated with patchouli content. Although this study did not validate the identified genes using qRT-PCR, it provides a valuable genetic resource for further research in patchouli. Dong et al. (2021) constructed the genome of *Magnolia biondii* using SMRT *via* PacBio long read, 10X Genomics and Hi-C data. The chromosome-level reference genome of *M. biondii* is approximately 2.22 Gb long and predicted 47,547 protein-coding genes. The volatile oil extracted from the flower buds of *M. biondii* has many pharmacological properties such as anti-inflammatory and is rich in terpenoids, which are catalyzed by terpene synthase (TPS). Based on genomic information and RNA-seq data, 102 TPS genes were identified and the expression profiles showed 33 TPS genes were higher expressed in flowers than in leaves. These findings will improve the understanding of the molecular breeding of *M. biondii*. Kim et al. (2020) assembled a draft genome of *Platycodon grandiflorus* through PacBio platform and contained 40,017 protein-coding genes. Gene family expansion and contraction analysis found that CYP76C, CYP72, and CYP716 families in *P. grandiflorus* underwent expansion. Based on orthologous gene annotation, *β-amyrin synthases* (*bASs*) were found that underwent expansion in *P. grandifloras*. Previous research has revealed that the CYP716 gene family was involved in the platycodon saponins biosynthesis. Additionally, among the expanded gene families, *CYP716* and *bAS* genes were highest expression in roots than other tissues. To investigate the terpenoid biosynthesis pathway in *Artemisia argyi*, gene expression analysis was performed and found 36,820 non-redundant transcripts, of which 187 transcripts relevant to terpenoid biosynthesis were discovered *via* KEGG analysis. Among them, eight diterpenoid biosynthesis genes were identified and highly expressed compared to other tissues. Finally, qRT-PCR verified 12 genes that were highest expressed in leaves were consistent with RNA-seq data (Kim et al., 2020). Shen et al. (2018) reported a draft *A. annua* genome sequence of 1.74 Gb that is assembled by Illumina and PacBio sequencing platform. They annotated 63, 226 protein-coding genes based on expression evidence. Gene expansion and contraction analysis revealed that 7,286 expanded and 3,950 contracted gene families in *A. annua*. Among the *A. annua* expanded gene families, TPS families expanded significantly and 122 TPS genes were identified. Subsequently, according to genomic and transcriptomic analyses, sesquiterpenoid synthesis-related genes were found and expressed in specific tissues. Previous studies were conducted to increase

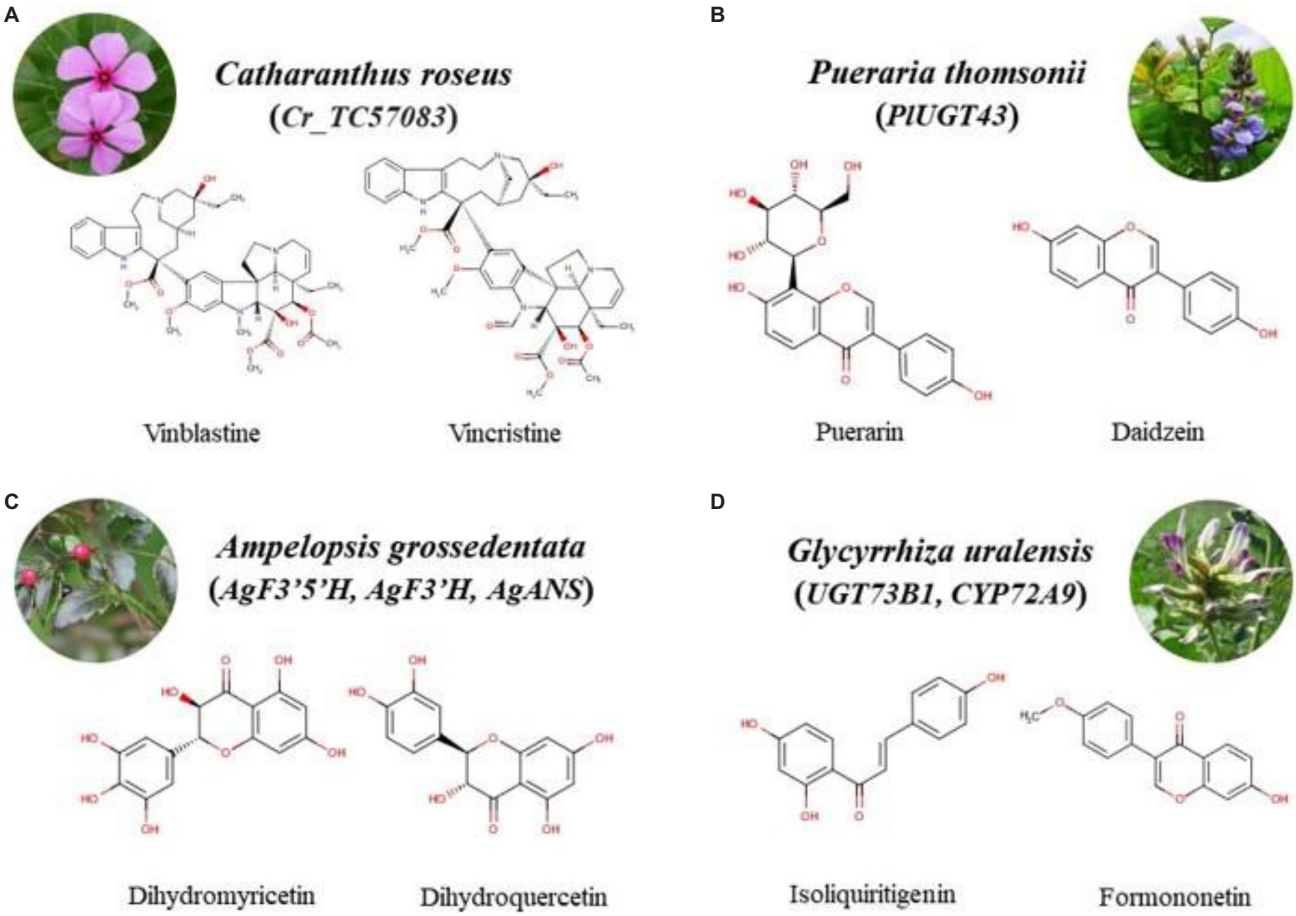


FIGURE 2 | Chemical structure formula of the bioactive compounds from four species. *Bioactive compound-related functional genes are shown in the brackets. **(A)** Cr_TC57083, the functional gene related to synthesis of vinblastine and vincristine in *Catharanthus roseus*; **(B)** PlUGT43, the functional gene involved in synthesis of puerarin and daidzein in *Pueraria thomsonii*; **(C)** AgF3'5'H, AgF3'H, AgANS, the functional genes associated with biosynthesis of dihydromyricetin and dihydroquercetin in *Ampelopsis grossdentata*; **(D)** UGT73B1, CYP72A9, the functional genes correlation with synthesis of isoliquiritigrnin and formononetin in *Glycyrrhiza uralensis*.

the yield of artemisinin by overexpressing the upstream or downstream enzymes, such as *FPS* and *HMGR*, but the artemisinin content did not increase significantly. Therefore, this study overexpressed the upstream, midstream and downstream enzymes [*HMGR (AA201470)*, *FPS*, and *DBR2*] simultaneously to examine the content of artemisinin, and the revealed that the transgenic plants produced more artemisinin. The present study provides a large number of candidate genes for further enhancement of artemisinin content. Xiong et al. (2021) revealed the *Taxus* genome with 10.23 Gb assembled into 12 chromosomes using Illumina HiSeq 2500, PacBio Sequel II, and Hi-C data. Based on genomic information and previous literature, 649 CYP450 genes were identified, of which CYP750 and CYP725 families in *Taxus* underwent significant expansion. Moreover, expression levels of four gene clusters on chromosome 9 where most CYP725A genes located on were significantly up-regulated by jasmonic acid induction. These results suggested that the gene cluster probably contains the majority of the paclitaxel synthesis genes that originated from the evolution of *T. chinensis*. PlantiSMASH analysis further indicated a gene cluster in group 9.2 that may be associated with terpenoid biosynthesis, including two TS genes, two T5αH genes, and two unknown CYP725As. Biochemical assays demonstrated that TS and T5αH were mainly responsible for catalyzing the first two steps of paclitaxel biosynthesis. Finally, 17 CYP725A genes which were closely related to known paclitaxel biosynthesis genes were identified through a gene-to-gene coregulation network. This article helps to accelerate paclitaxel biosynthesis and *Taxus* biotechnology applications.

Flavonoids

Flavonoids are natural products with a C6-C3-C6 carbon skeleton structure, usually combined with sugars to form glycosides present in medicinal plants, which classified into flavonols, flavones, isoflavones, and anthocyanidins (Ciumarnean et al., 2020). Phenylalanine, the biosynthetic precursor of flavonoids, which is then converted to rutin through 10 enzymes, including phenylalanine deaminase (PAL), cinnamic acid-4-hydroxylase (C4H), 4-coumarate coenzyme A ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanones-3′-hydroxylase

(F3'H), flavanones-3′5'-hydroxylase (F3'5'H), flavonol synthase (FLS), glucose/witch hazel transferases (UGT/GT; Zhang L. et al., 2017). CHS is the first key enzyme in the flavonoid synthesis pathway and has been studied more in medicinal plants. In addition, O-methyltransferase (OMT) catalyzes the conversion of root-specific norbornin to other flavones, such as wogonin, isowogonin, and moslosooflavone (Liu W. et al., 2021). Zhao et al. (2019) reported the high-quality reference genome sequence of *Scutellaria baicalensis* at the chromosome level with 408.14 Mb (93%) assembled into 9 pseudochromosomes using Illumina, PacBio, and Hi-C technology. The article further elucidated the biosynthesis of norwogonin to wogonin catalyzed by 8-O-methyltransferase. Hence, 28,930 genes were annotated by bioinformatics tools based on *de novo* predictions, homology-based prediction, and RNA-seq data. On the basis of the genomic and transcriptome information, six genes encoding OMTs were found in the tandem repeat region unique to *S. baicalensis*. Moreover, enzyme activation and RNAi experiments confirmed that *SbPFOMT*s were involved in the synthesis of Wogonin, and the Wogonin synthesis pathway was resolved. It provides a basis for synthetic biology to obtain baicalein and a reference for genetic analysis of other Labiatae plants (Zhao et al., 2019). Wu et al. (2021) reported the first chromosome-scale reference genome of *Carthamus tinctorius* through combined PacBio platform and Hi-C mapping and predicted 33,343 protein-coding genes. Among them, seven CHS genes were identified based on the modified flavonoid biosynthetic pathway combined with the KEGG database and related literature. *CarCHS5* and *CarCHS6* which are unique to *C. tinctorius* were revealed by Collinearity analysis. These results provide evolutionary insights into the flavonoid biosynthesis in *C. tinctorius*. Qing et al. (2021) generated the *Hemerocallis citrina* genome of 3.77 Gb that was assembled by PacBio long reads and Hi-C data. Gene family expansion and contraction analysis revealed that 10,375 gene families in *H. citrina* underwent expansion, whereas a significant number of gene families (6707) underwent contraction. In addition, the expanded gene families were mainly enriched in flavonoids biosynthesis, which may have contributed to the content of rutin in *H. citrina*. Among them, 108 genes were identified by homology search. In addition, High-performance liquid chromatography/quadrupole time-of-flight (HPLC-Q-TOF) data revealed that rutin was mainly accumulated in flower buds, 20 candidate genes which mainly expressed in flower buds were identified in combination with transcriptome data. Zhang L. et al. (2017) reported a high-quality assembly of the 489.3 Mb genome of *Fagopyrum tataricum* at chromosome-scale, and contained 33, 366 protein-coding genes. In *F. tataricum* genome, nine rutin biosynthesis enzymes were identified, such as CHI (FtPinG0002790600) and F3'H (FtPinG0002353900). Although this study did not further validate the candidate genes, it provided some insights into the biosynthetic pathway of rutin.

Alkaloids

Alkaloids are basic nitrogen-containing organic compounds with complex ring structures. It is mainly synthesized with precursors such as tryptophan, tyrosine, phenylalanine, lysine, and arginine. According to the structures, it can be classified into pyridines, indoles, terpenoids, isoquinolines, steroids, etc. (Liu et al., 2019). The previous report showed that CYP80 and CYP719 families catalyze many different reactions in the biosynthesis of BIAs, including hydroxylation, C-C, or C-O coupling, and formation of methylenedioxy bridges (Deng X. et al., 2018). Hu et al. (2021) presented the draft genome sequence of *Strobilanthes cusia*, constructed using PacBio long reads and Hi-C sequencing data. The draft genome assembly had a size of 913.74 Mb and contained 2,974 coding gene sequences, of which 2,975 DEGs were identified by transcriptome analysis and enriched in phenylpropanoid, flavonoid, and triterpenoid biosynthesis. Based on gene family expansion and contraction analysis, 60 gene families expanded in *S. cusia*, while 16 gene families contracted. In addition, these expanded gene families were mainly enriched in secondary metabolism pathway according to GO and KEGG analysis. On the basic of homology searching, 18 genes coding IA-related key enzymes were identified as DEGs in the *S. cusia* genome, such as *UGT*, *CYP450*, *ASA*, *TSB*, *BGL*, *CS*, and *EPSPS* genes. This study reveals the molecular basis of the accumulated indole alkaloids of *S. cusia*. Yang et al. (2021a) presented a reference chromosome level genome of *Areca catechu* by Illumina and PacBio data. The assembled genome was 2.51 Gb with a N50 scaffold size of 1.7 Mb and predicted 31,571 protein-coding genes. Among them, 904 expanded gene families were enriched in secondary metabolites pathways, including flavonoid, terpenoid, and isoquinoline alkaloid biosynthesis. Although this article was not mining deeper into functional genes related to secondary metabolite synthesis, it provides basic insights into areca alkaloid biosynthesis. Liu Y. et al. (2021) assembled a high-quality genome of *Coptis chinensis* at the chromosome level through integrating Nanopore sequencing, Illumina short reads, and Hi-C technology. Forty-one thousand four protein-coding genes were annotated. In addition, 1,083 gene families underwent expansion were revealed by gene evolution analysis. Among them, two new P450 families are found in early-diverging eudicots: CYP719 and CYP749. According to a comparison of these CYP719 genes, *Cch00017825* is clustered on chromosome 3 and expressed significantly in the *C. chinensis* rhizomes. Liu X. et al. (2017) reported the *Macleaya cordata* genome with 378 Mb using Illumina HiSeq 2000. Based on the previous report, 39 candidate genes related to SAN and CHE biosynthesis were identified. They performed liquid chromatography coupled with mass spectrometry (LC/MS) to detect the metabolites in *M. cordata* and obtained the SAN and CHE biosynthesis. Based on the reference genome of *M. cordata*, 39 genes were identified involved in SAN and CHE biosynthesis pathway. Additionally, SAN and CHE were not accumulated in the stem, 16 candidate genes were selected in combination with RNA-seq data. Finally, Metabolic engineering verified 14 candidate genes involved in catalytic reactions of SAN and CHE biosynthesis. This research could help medicinal plants produce more SAN and CHE. Rai et al. (2021) used PacBio long reads, Illumina short reads, Bionano optical mapping and Hi-C sequencing to assemble the *Ophiorrhiza pumila* genome. Based on metabolome annotation datasets, 273 nitrogen containing metabolites were annotated,

most of them were indole alkaloids (IAs). According to a previous study and comparative genomic analysis, they found that monoterpene indole alkaloids (MIAs) biosynthesis are originated from strictosidine biosynthesis. In the late secoiridoid pathway, OG0014621 (*LAMT*) was found specifically expanded in *O. pumila*. Additionally, genes encoding MIA-related enzymes were replicated and obtained, such as *STR*, *SLS*, *7-DLH*, and *7-DLGT*. These suggested that gene clusters include several functional signature genes. The study presents a pangenome model of MIA biosynthesis that will help establish a sustainable supply of camptothecin. Kang et al. (2021) presented a high-quality genome sequence of *Camptotheca acuminata*, constructed using SMRT sequencing technology from PacBio, Illumina platform, and Hi-C techniques. In *C. acuminata*, gene expansion and contraction analysis indicated that 2,951 gene families expanded, while 1,733 gene families contracted. Previous research on *C. roseus* illustrated that CYP72A219 (SLS) and CYP72A224 (7-DLH) play a vital role in indole alkaloids biosynthesis in periwinkle. Hence, on the basic of homology searching and phylogenetic tress, two *SLS* genes (*CacGene13172*, *CacGene10833*) and two *7-DLH* genes (*CacGene13171*, *CacGene10832*) were identified in *C. acuminata* genome. This study reveals candidate genes that may play a role in the camptothecin biosynthesis in *C. acuminata*, providing a basis for future high-yield artificial biosynthesis.

Quinone

Kang et al. (2020b) presented the chromosome-level genome of *Senna tora* with 526.4 Mb by PacBio long read sequencing and Illumina data and were assembled into 13 chromosomes using Hi-C data. Metabolite profile analyzed the content of 10 anthraquinone in different developmental stages of seeds and found that aurantio-obtusin was the major bioactive compounds in mature seeds. Furthermore, they predicted 45,268 protein-coding genes. Comparative genomics revealed that 2,874 gene families in *S. tora* underwent expansion, while 3,371 gene families underwent contraction. Interestingly, the expanded gene families were mainly enriched in phenylpropanoid, isoflavonoid, and terpene biosynthesis. Among the expanded gene families, 16 CHS-L genes were identified due to its rapidly expanded in in *S. tora*. At stage 4 when anthraquinones began to accumulate, two CHS-L genes (STO07G228250 and STO07G228220) presented high expression levels. Further phylogenetic tree analysis showed that STO07G228250 was more similar to HpPKS and ArOKS, octaketide synthases. Finally, according to ESI-MS spectrum and enzyme activity experiment, STO07G228250 was demonstrated to be involved in the first step of anthraquinone biosynthesis. The study provides a platform for medicinal plant *S. tora* with high bioactive molecular content.

Phenylpropanoid

Yang et al. (2021b) constructed the assembly of 1.79 Gb genome sequence *of Arctium lappa* and obtained 32,771 protein-coding genes. Based on the genomic information, 616 positively selected candidate genes were discovered in *A. lappa*. Transcriptome analysis revealed that genes related to lignan biosynthesis in five different stages of *A. annua* (4CL), dirigent protein (DIR), and hydroxycinnamoyl transferase (HCT) were highly related to arctiin biosynthesis.

PHARMACODYNAMIC GENES MINING FOR THE CHRONIC COMPLEX DISEASES

Herbs have been demonstrated in studies to help with the treatment of rheumatism, diabetes, cancer, Alzheimer's disease, and cardiovascular disease (Li X. et al., 2019). Take rheumatism as an example, they treat rheumatism by the following effects: dispelling wind, eliminating the body moisture, removing coldness, clearing heart, dredging the channel, expectorant and diffusing impediment, benefiting Qi and nourishing the blood, and invigorating the kidney and strengthening the spleen.

Rheumatism

Callerya speciosa made contributions to treat rheumatism *via* dredging the channel effects, due to its main medicinal ingredients-isoflavonoids, such as maackiain and formononetin. Yao et al. (2021) used NGS technology to sequence the transcriptome of *C. speciosa* and identified 4,337 DEGs during the tuberous root development. Among them, 15 genes related to isoflavonoids biosynthesis were found. These results indicated that these genes may be promoted the accumulation of isoflavonoids in the tuberous toot. In addition, qRT-PCR validated the expression pattern of candidate DEGs were consistent with the RNA-seq data. The study provides new insights into the potential mechanisms of isoflavonoid biosynthesis in *C. speciosa*. Flavonoid, a bioactive compound derived from *Fritillaria hupehensis*, has often been used as an expectorant to treat rheumatic diseases. Guo K. et al. (2021) performed SMART analysis from the PacBio platform to sequence the full-length cDNA of *F. hupehensis*. Thirty-four flavonoid biosynthesis unigenes were found using the KEGG pathway, and divided into five branches by blast against model plants. The study provides a valued resource for herb breeding and bioactive compounds for pharmacological application. *Asarum sieboldii* has abundant medicinal properties, such as anti-inflammatory, antiallergic, and removing coldness. Asarinin and aristolochic acid are bioactive compounds that originated from *A. sieboldii*. Chen C. et al. (2021) used full-length transcriptome analysis to uncover genes involved in asarinin and aristolochic acid biosynthesis. The result found 63, 023 transcriptional sequences, of which 41 asarinin biosynthesis candidate genes and 56 aristolochic acid biosynthesis candidate genes were identified, including *AsCOMT*, *AsEPI*, *AsCYP81Q2*, *AsCYP81Q4*, *AsCYP81Q7*, *AsCYP81Q29*, *AsTYR*, *AsTYDC*, *AsNCS*, *AsNOMT*, *AsCNMT*, and *AsCYP80B1*. Finally, qRT-PCR data verified 4 genes were significantly expressed in the root, including *AsCCR*, *AsPAL*, *AsCOMT*, and *AsCYP81Q*. The study will provide a good basis for the production of a valuable, low toxicity active ingredient. *Gynostemma pentaphyllum* has

the pharmacological effect of eliminating the body moisture and clearing the heart. Gypenosides, triterpene saponins, are the main active compounds in *G. pentaphyllum*. Liang et al. (2019) obtained 140,157 unigenes using PacBio standard analysis pipeline and Illumina data. Among them, 404 gypenoside biosynthetic genes were detected and annotated. GpOSC1, GpCYP89, and GpUGT35 were demonstrated the leading candidate genes for gypenoside biosynthesis by qRT-PCR technology. These findings will lay a new foundation for gypenosides biosynthesis. *Akebia trifoliata* possesses the properties of strengthening the spleen and relieving pain. Bioactive compounds contribute medicinal effects to *A. trifoliata*, including triterpenoid saponins, triterpenes, and flavonoids. Huang H. et al. (2021) presented the chromosome level genome sequence of *A. trifoliata* using Illumina HiSeq X-Ten sequencing technology, SMRT platform from PacBio and Hi-C technique. The genome assembly had a size of 682.14 Mb and predicted 25,598 protein-coding genes. Two hundred forty-six expanded and 473 contracted gene families in *A. trifoliata* were discovered, according to the phylogenetic tree. Interestingly, the expanded gene families were mainly enriched in terpenoid biosynthesis *via* KEGG enrichment analysis. Among the expanded gene families, 24 *Atrβ-AS* genes, 12 UDP-glucoronosyl, three UDP-glucosyltransferase and seven cytochromes P450 gene families were involved in sesquiterpenoid and triterpenoid biosynthesis pathways. In addition, three UDP-glucosyltransferase, 14 cytochrome P450, and two TPS gene families were constructed in the *A. trifoliata* genome. The findings suggested that these gene families were quickly changing to synthesis varies of triterpenes in *A. trifoliata*. The study provides a useful genetic resource for pharmacological applications of *A. trifoliata*. *Trillium govanianum* contributes to the treatment of rheumatism disease, due to its ability to benefit Qi and nourish the blood. Diosgenin, as steroidal saponins, is considered to be the main bioactive component of *T. govanianum*. Singh et al. (2017) performed spatial transcriptome analysis of the leaf, fruit, stem, and rhizome tissues of *T. govanianum* and obtained 69,174 transcripts. As a result, 108 CYP genes and 58 UGTs were identified, of which 87 CYP genes and 49 UGTs were differentially expressed in four tissues. Based on KEGG classification, genes involved in steroidal saponin biosynthesis were divided into three pathways: terpenoid backbone (16 genes), sesquiterpenoid and triterpenoid (two genes), and steroid biosynthesis (14 genes). In addition, qRT-PCR was employed to confirm the expression pattern of 29 genes, which was consistent with the RNA-seq results. Thus, compared with transcriptome sequencing, *de novo* genome assembly of herbs mines more functional genes involved in the biosynthesis pathways of active ingredients, which is more beneficial to the development of herb molecular breeding.

Diabetes

For many years in India, *Gymnema sylvestre*, a well-known and valuable medicinal plant, has been used to treat diabetes (Hossain et al., 2016; Tiwari et al., 2017; Pham et al., 2018). Ayachit et al. (2019) performed RNA-seq data to uncover terpenoid biosynthesis genes and identified 111 transcripts involved in the terpenoid biosynthetic pathway, such as monoterpenes, di-terpenes, tri-terpenes, and ubiquinones. Finally, qRT-PCR verified six transcripts involved in the MEP pathway were a positive correlation to terpenoid biosynthesis. This study provides insights for future functional genomics studies of *G. sylvestre*. *Eriobotrya japonica*, a traditional medicine, is beneficial in the treatment of diabetes, due to its variety of active compounds, such as flavonoids and terpenoids (Mogole et al., 2020). Wang (2021) constructed a draft genome of *E. japonica* to discover medicinal bioactive compounds using HiSeq 4,000 sequencing platform, PacBio long-read sequencing technology, and Hi-C data and obtained 45,492 protein-coding genes. According to gene family expansion and contraction analysis, 483 gene families in *E. japonica* expanded significantly and were mainly enriched for metabolic pathways in combination with KEGG analysis. Metabolite profiles showed that phenolic acids, flavonoids, and terpenoids were detected abundantly in the *E. japonica*. Based on genomic information, 71 flavonoid biosynthesis genes were annotated, of which 3 genes encoding key enzyme were identified in the quercetin biosynthesis pathway. In addition, 286 predicted protein-coding genes in phenylpropanoid biosynthesis were identified, only five genes underwent in the expansion family. According to KEGG analysis, 92, 32, 56, and 37 candidate genes were identified involved in terpenoid backbones, monoterpenoids, diterpenoids, and sesquiterpenoid-triterpenoids biosynthesis pathways, respectively. The study provides a valuable introduction for further molecular pharmacological studies of *E. japonica*. Thousands of years ago, *Pueraria thomsonii* was used to treat diabetes in the East. Puerarin as the bioactive isoflavones is mainly accumulated in the root of *P. thomsoni* and has antioxidant and anti-inflammatory properties (Chen et al., 2018; Yang et al., 2019; Luo et al., 2021). He et al. (2019) performed PacBio and Illumina sequencing technology to sequence the *P. thomsoni* transcriptome and acquired 44,339 transcripts. They discovered 9,225 differentially expressed transcripts (DETs). Among them, 32 genes might be involved in isoflavone production, of which the expression profile of eight genes were confirmed by qRT-PCR which consistent with RNA-Seq data. *Glycyrrhiza uralensis*, an important medicinal plant of the genus *Glycyrrhiza*, has been used as TCM. Flavonoids and Glycyrrhizin originated from liquorice possess antioxidative, antidiabetic, and anti-inflammatory effects (Lee et al., 2010; Feng et al., 2013; Lin et al., 2022). Mochida et al. (2017) reported a draft genome of *Glycyrrhiza uralensis*, based on Illumina short reads and PacBio long reads. The assembled genome is 379 Mb with scaffold N50 of 109 kb, encoding 34,445 predicted genes. On the basis of genomic information, *CYP93C*, *HI4OMT*, and *7-IOMT* some of which are involved in isoflavonoid biosynthesis were observed and generated a cluster. Based on homolog searching and functional annotation, P450 and UDP-dependent glycosyltransferase (UGT) families involved in triterpenoid saponin biosynthesis were predicted and the expression of *bAS*, *CYP88D6* and *CYP72A154* were consistent with the glycyrrhizin yield of the *G. uralensis* samples. In addition, RNA-seq data revealed that Glyur002597s00038051 and Glyur002597s00038050 which closely homologous to CYP72A9 and UGT73B1 in

Arabidopsis thaliana were high correlation of expression patterns. Hence, P450 and UGT genes might be involved in triterpene saponin biosynthesis in *G. uralensis*. These findings help researchers use genomic resources combined with biosynthetic approaches to create a library of rare natural or novel bioactive compounds to facilitate drug discovery. *Sophora flavescens* are important traditional medicinal plants with pharmacological properties effective in the treatment of inflammatory disorders including diabetes complications (Guo et al., 2021b). Alkaloids and flavonoids are the major bioactive compounds in root tissues. Wei et al. (2021) performed transcriptome analysis of the periderm, phloem, and xylem tissues of *S. flavescens* and obtained 58,327 unigenes. High-performance liquid chromatography (HPLC) detected metabolite contents in the root tissues and the results showed that alkaloids contents were highest in the phloem, while flavonoids contents were highest in the xylem. Fifty-two and one hundred thirty-seven CYP transcripts involved in alkaloid biosynthesis were identified and expressed highest in the xylem. Additionally, 37 transcripts were found in flavonoid biosynthesis and expressed highest in the xylem. Correlation analysis found LYSA, AO, PMT transcripts were markedly and positively correlated with alkaloids contents and 4CL, 2'OH, CHI5, CHRI transcripts were markedly and positively correlated with flavonoids contents. These results provide a basis for the molecular breeding of *S. flavescens*.

Alzheimer's Disease

Curcumin, the bioactive compound of *Curcuma longa*, may prevent or reverse Alzheimer's disease due to its anti-inflammatory and antioxidant activity (Kim et al., 2022). Chakraborty et al. (2021) assembled the draft genome sequence of *C. longa* using Oxford Nanopore long reads and obtained 50,401 coding gene sequences. Among 10 enzymes involved in the curcuminoid biosynthesis pathway, gene family evolution analysis revealed that two gene families (*HCT* and *OMT* genes) appear to be undergoing contraction, while a significant number of gene families (8) appear to be expanding in *C. longa*. The result suggests that genes related to the curcumin biosynthesis pathway have evolved, providing a genetic basis for its pharmacological properties. *Corydalis yanhusuo* have been used to relieve neuropathic pain, due to its bioactive compounds-tetrahydropalmatine, a member of BIAs. Xu D. et al. (2021) performed SMRT sequencing from the PacBio platform to sequence the cDNA library from leaves and tubers of *C. yanhusuo*. Based on the tblastn results, 101 unigenes involved in BIA biosynthetic pathway were founded, of which 36 unigenes were identified as DEGs *via* expression analysis, owing to the majority of them expressed at a higher degree in tubers which consistent with the metabolome data. In addition, phylogenetic analysis showed that 10 OMT unigenes involved in the final step for tetrahydropalmatine synthesis were identified. This study provides the basis for the subsequent molecular cloning and activity validation of the enzyme. Diosgenin is an anti-inflammatory and antioxidant compound found in the rhizome of *Dioscorea zingiberensis*. Cheng J. et al. (2021) used Illumina HiSeq and PacBio SMART technologies to sequence the *D. zingiberensis* genome and annotated 26,022 protein-coding genes. Based on previous study and genomic information, they found *DzinCYP90G6* and *DzinCYP94D144*, two P450 genes, were related to diosgenin biosynthesis using blastp queries. Finally, metabolic engineering verified that coexpression of *DzinCYP90G6* and *DzinCYP94D144* in yeast was able to produce diosgenin. This study helps to decode the evolutionary trajectory of the biosynthetic pathway of diosgenin, but also provides insights into the enhancement of diosgenin production through biochemical synthesis. Chlorogenic acid (CGA) accumulates in the leaves and bark of *Eucommia ulmoides*, and it can reduce the concentration of glucose in the blood after the meal and lower blood pressure. Li et al. (2020) used PacBio Sequel platform, Illumina NovaSeq platform, and Hi-C data to assemble a high-quality *E. ulmoides* genome. Twenty-six thousand one protein-coding genes were predicted. Based on homologous gene comparison in the *E. ulmoides* genome, 23 candidate genes encoding six key enzymes involved in the CGA biosynthesis pathway were identified, including PAL, 4CL, C4H, C3′H, HCT, and HQT genes. In addition, gene expression profile analysis showed that the expression levels of these genes were higher than other genes. This work will accelerate the understanding of the molecular mechanisms of biosynthesis of other valuable secondary metabolites, such as rutin and quercetin. Dihydroquercetin (DHQ) is a pharmacologically active, which can be converted to dihydromyricetin (DHM). Yu Z. et al. (2021) used HPLC to detect the DHM content in *Ampelopsis grossedentata* from different geographical locations and divided into two groups: B group (low DHM) and D group (high DHM). Then, Illumina HiSeq 2000 sequencing platform was performed to sequence the transcriptome of *A. grossedentata* from B and D groups and annotated 57,016 unigenes in D group using seven public protein databases. The differentially expressed gene analysis revealed 926 DEGs in B vs. D, of which 446 up-regulated genes and 480 down-regulated genes. DEGs of 10 structural enzyme genes associated with flavonoids biosynthesis were identified, including *PALs*, *CLs*, *CHSs*, *F3'H*, *F3'5'H*, *ANS*. In addition, qRT-PCR verified the expression level of selected genes which was matched the transcriptome data, including *CHSs*, *F3'H*, and *F3'5'H*. This work will stimulate further genetic research on *A. grossedentata* and may ultimately lead to genetic improvements in the plant's DHQ content.

Cancers

Catharanthus roseus accumulates vinblastine and vincristine, which have long been known to be an anti-cancer drug to cure glioma disease (Skubnik et al., 2021). Verma et al. (2014) performed Illumina platform to sequence the *C. roseus* transcriptome from leaf, flower, and root and obtained 59,220 unique transcripts. Next, using the CathaCyc database analysis, 30 well-known genes involved in TIA biosynthetic pathways were identified and they are conserved in the sequenced reference genome of tomato, potato, and Arabidopsis at the protein level. Based on the RNA-seq data, most of TIA biosynthesis genes were up-regulated in leaf and root tissues, while Cr_TC35206 and Cr_TC35622 were highly down-regulated in roots, Cr_TC04217 were highly down-regulated

in leaves. Finally, they performed qRT-PCR to validate 10 TIA biosynthetic pathway genes and their expression pattern consistent with RNA-seq data (see **Figure 2**). This research adds to our knowledge of the regulatory systems that control alkaloid production. *Andrographis paniculata* produces a large number of diterpenoid lactones with antitumor and immunomodulatory effects which are usually used to treat esophageal cancer, including andrographolide and neoandrographolide (Li L. et al., 2017, 2019; Sun et al., 2019; Yue et al., 2019). Sun et al. (2019) assembled *A. paniculata* genome sequence of 269 Mb at chromosome-scale using Illumina short reads, PacBio long reads, and Hi-C data and predicted 25,428 protein-coding genes. Based on the phylogenetic tree, 1,290 expanded and 5,383 contracted gene families in *A. paniculate* were discovered. Interestingly, the expand gene families were mainly enriched in the secondary metabolism pathway, such as TPS and CYP gene families, which may have contributed to the synthesis of andrographolide and neoandrographolide. Previous reports suggested that CYPs and 2OGDs may be involved in turing the diterpene backbone into diterpenes. Therefore, 278 CYP genes and 112 putative 2OGDs in the *A. paniculata* genome. Moreover, diterpene lactones need to be glycosylated to become neandrographolide, and 120 putative UGT genes were identified. Transcriptome data revealed that 18 CYP71and 6 CYP76 family members, 17 2OGDs, and 29 ApUGTs were significantly elevated after MeJA treatment. Finally, UGT73AU1 was confirmed that it converted the andrograpanin to neandrographolide *via* enzymatic assays. The study provides further understanding of the production of bioactive diterpene lactone components. The antitumor and anticancer medicinal properties of *Lantana camara* are attributed to its multiple bioactive substances, such as triterpenod which can be used to treat papilloma (Sharma et al., 2007). Shah et al. (2020) performed Illumina sequencing platform to analyze *L. camara* transcriptome in leaf and root and found 72,877 and 513,985 unigenes from leaf and root tissues, respectively. Moreover, 229 and 943 genes involved in the phenylpropanoid biosynthesis in leaf and root tissues, respectively were identified by pathway analysis. Twenty thousand forty-four DEGs were identified. Among them, 11,496 genes were up-regulated and 8,548 genes were down-regulated. The transcriptome analysis provides the basis for further molecular studies of *L. camara*. *Salvia officinalis* is used in medication for treating liver cancer, leukemia, colon cancer through inhibiting cell proliferation and apoptosis (Jantová et al., 2014; Pedro et al., 2016; Jiang et al., 2017). Ali et al. (2017) used Illumina HiSeq 2000 to sequence the transcriptome of *S. officinalis* leaves and assembled 48,671 unigenes. Among them, 65 unigenes involved in terpene synthase were identified, according to the TPS sequence in the reference database. Furthermore, 11 DEGs were selected to be validated by qRT-PCR. The results showed that eight candidate genes were most highly expressed in young leaves, except for *SoGGPS*, *SoLINS*, and *SoHUMS* genes were most highly expressed in stems. qRT-PCR data are well in agreement with GC–MS analysis data, major terpene groups present in young leaves. In addition, five terpene synthase genes (*SoNEOD*, *SoCINS*, *SoSABS*, *SoLINS*, and *SoTPS6*) were selected to overexpress in tobacco, respectively. The results indicate that more terpenoids were produced in transgenic tobacco. 2-methoxy-1,4-naphthoquinone (MNQ) is an active component of *Impatiens balsamina* with anticancer pharmacological properties. Shikimate and 1,4-dihydroxy-2-naphthoate (DHNA) pathways are responsible for the synthesis of MNQ, which lawsone is first synthesized by oxidative decarboxylation of DHNA and then possibly catalyzed by S-adenosylmethionine-dependent O-methyltransferase (SAM-dependent O-MT), NADH-quinone oxidoreductase, and UDP glycosyltransferases. Foong et al. (2020) reported the transcriptome sequencing of *I. balsamina* through Illumina HiSeq4000 paired-end sequencing technology and obtained 50,786 unigenes, of which 27, 104, 82, and 122 unigenes related to DHNA pathway, SAM-dependent O-MT activity, NADH-quinone oxidoreductase, and UDP glycosyltransferases, respectively. Among them, five unigenes related to DHNA pathway were highest expressed in early-stage capsule, gene expression of six unigenes were substantially and positively connected with MNQ content, 3 NADH-quinone oxidoreductase and 5 UDP glycosyltransferases were connected with lawsone content significantly. In addition, HPLC analysis indicated that lawsone was highest expressed in mature leaves, whereas MNQ was expressed only in capsules pericarps. Finally, qRT-PCR verified 20 candidate genes were consistent with transcriptome data. *Dendrobium officinale* is a widely used medicinal plant that creates a variety of bioactive compounds which resist cancer by inhibiting cell proliferation (Luo et al., 2019; Tao et al., 2021). Niu et al. (2021) reported its 1.23 Gb genome, encoding 27,631 predicted genes using PacBio long-reads, Illumina short-reads, and Hi-C data. Additionally, gene family evolution analysis found that 820 gene families expanded in *D. officinale*, while 975 gene families contracted. Interestingly, the expanded gene families were mainly enriched in polysaccharides, alkaloids and flavonoids biosynthesis according to KEGG enrichment analysis. Based on homolog searching and functional annotation, 268, 98, and 52 genes related to three main bioactive ingredients biosynthesis were identified, including polysaccharides, alkaloids, and flavonoids biosynthesis, respectively. Differential expression profiling revealed 1,677 DEGs, of which 51 DEGs related to polysaccharides, alkaloids, and flavonoids biosynthesis were identified. Additionally, 218 *CYP450* genes were found in the *D. officinale* genome according to homologous search from *A. thaliana*, of which 29 DEGs were identified using comparative transcriptome analysis and most of them are CYP71 family members which were up-regulated. The results suggested that CYP71 groups make a great contribution in regulating the synthesis pathway of active ingredients. Therefore, the high-quality reference genome reported in this study can contribute to the functional genomics study and molecular breeding of *D. officinale*. Chen et al. (2012) reported a reference genome assembly of *Ganoderma lucidum* using Illumina next-generation sequencing and predicted 16,113 genes. Terpenoids and polysaccharides are the main active components of *G. lucidum*. One hundred ninety-seven

CYP functional genes were identified, of which 78 CYP genes expression levels were consistent with terpenoids content in *G. lucidum*. According to previous literatures, 15 CYP512 and one CYP5144 genes were discovered in the *G. lucidum* genome, which may be related to triterpenoid biosynthesis. In addition, two polysaccharide genes, two LZ-8 genes, 12 TPS genes were identified. This study will pave the way for the role of *G. lucidum* in pharmacological and industrial applications (see **Figure 3**). *Zingiber officinale*, a gingerol-producing medicinal plant, possesses anti-cancer properties and treats breast cancer by inhibiting apoptosis (Huang P. et al., 2021). Li H. et al. (2021) reported chromosome-level reference genome assembly of *Z. officinale* using PacBio long reads, Illumina short reads, and Hi-C reads. Based on the assembly *Z. officinale* genome, they investigated the gene family expansion and contraction and revealed that 1,098 gene families underwent expansion, while 20 gene families underwent contraction. In addition, two expanded gene families (*PKS* and *AOR*) were mainly enriched in the biosynthesis of secondary metabolites. UHPLC–MS/MS detected rhizomes of *Z. officinale* at different stages and revealed that the concentration of active compounds were the lowest in R1, whereas were the highest in R5, such as gingerol analogs-6-gingerol and tetrahydrocurcumin. According to the metabolomic data and previous study, 12 gene families involved in the gingerol biosynthesis pathway were identified, and the expression level of 10 gene families were compatible with the concentration of gingerol. Additionally, *C3OMT2*, *C3OMT3*, and *C3OMT13* formed a unique clade in the *Z. officinale* genome which suggested that C3OMTs were likely related to feruloyl-CoA biosynthesis.

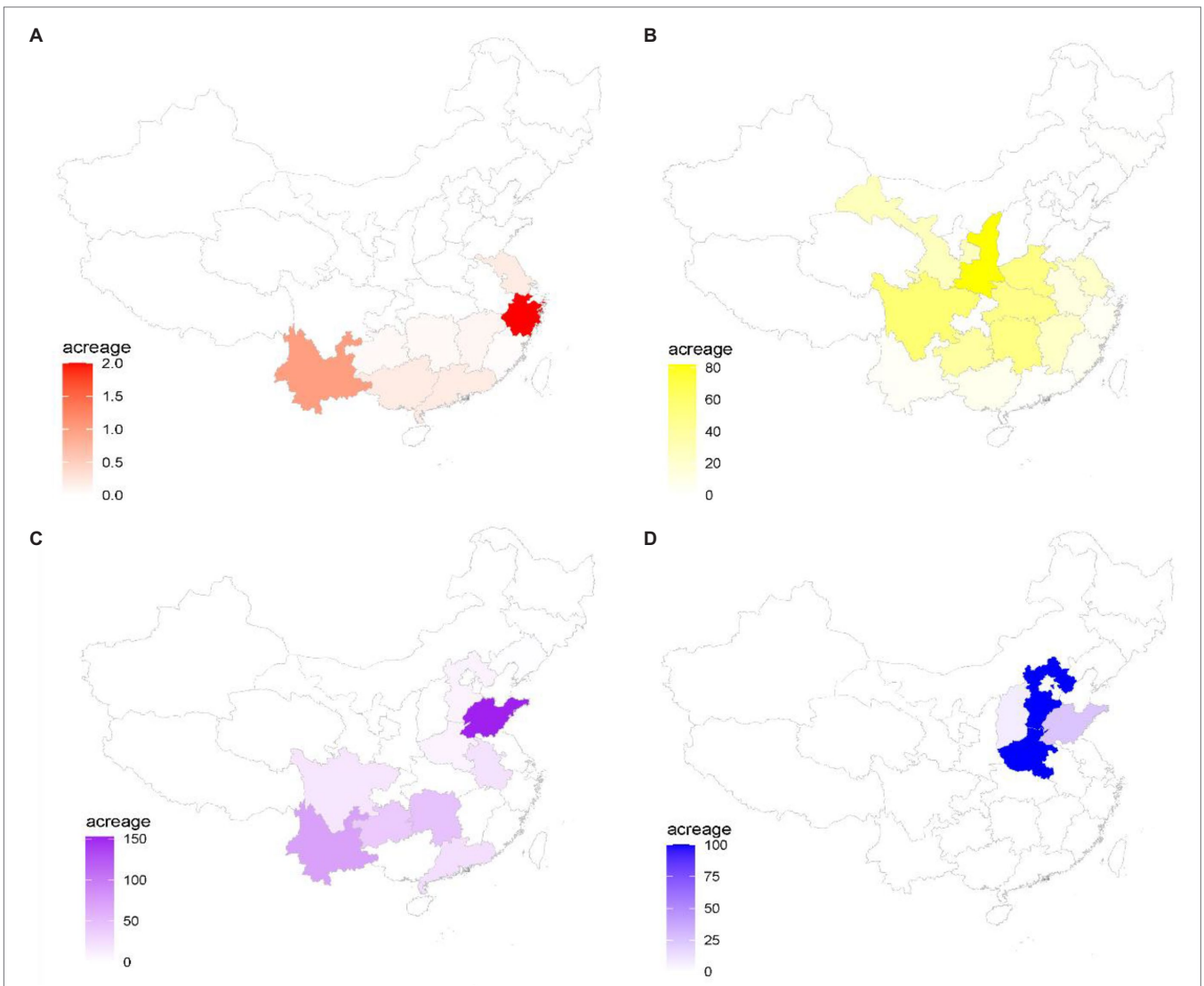


FIGURE 3 | Industry distribution and covered areas of four representative species in China. **(A)** *Dendrobium officinale*; **(B)** *Eucommia ulmoides*; **(C)** *Zingiber officinale*; and **(D)** *Dioscorea zingiberensis*. A color scale bar is shown the acreage size of species in each province of China.

Cardiovascular Disease

The tubers of *Ophiopogon japonicus* are accumulated in many active substances which contribute significantly to the treatment of cardiovascular diseases through their antioxidant, anti-inflammatory, and cardioprotective pharmacological properties, such as flavonoids, saponins, and polysaccharides (He et al., 2016; Wu et al., 2019; Fan S. et al., 2020). Liu H. et al. (2017) employed the Illumina platform to sequence the transcriptome of different ages tubers (Y1, Y2, Y3) from *O. japonicus* and generated 96,738 unigenes with *de novo* assembly. According to searching against the five databases, 77,409 unigenes were annotated among the 96,738 unigenes. Based on the result of gene annotation, 245, 135, and 236 unigenes encoding key enzymes in flavonoid, saponin, polysaccharide biosynthesis were identified, respectively. Most genes related to polysaccharide biosynthesis have the highest expression in Y2, while genes involved in flavonoid and saponin biosynthesis have the highest expression in Y1. qRT-PCR verified the expression level of 17 unigenes were consistent with RNA-Seq data, which were selected from flavonoid, saponin, polysaccharide biosynthesis, respectively. This study can accelerate the understanding of the physiological process of active ingredient synthesis at the molecular level and promote the development of natural drugs. Chrysanthemums have significant protective functions for cardiovascular (Song et al., 2018). Song et al. (2018) performed Oxford Nanopore long reads and Illumina short reads to assemble the genome of *Chrysanthemum nankingense*. In *C. nankingense*, gene evolution analysis revealed that 1965 gene families underwent expansion, while a significant number of gene families (1777) underwent contraction. Surprisingly, the expanded gene families were mainly enriched in terpene synthase activity. Among them, 219 TS genes and 708 CYP genes were identified, including seven SQSs, 158 TPSs, and 54 TCCs. Due to, the genes that make up the terpene synthesis pathway are organized into gene clusters, TPS-a/CYP99 and TPS-g/CYP79/CYP76, a new combination, were found in the *C. nankingense* genome. Further analysis revealed CHR00048430 and CHR00048432 genes from the TPS-a/CYP99 cluster were mainly expressed in roots, whereas CHR00011805 and CHR0001181 from TPS-g/CYP79/CYP76 cluster were mainly expressed in flowers. Additionally, two key flavonoid biosynthesis genes (*CHS* and *CHI*) from expansion families were significantly highly expressed in flowers. The study gives genomic data to help researchers better understand the evolutionary history of chrysanthemums. *Rehmannia glutinosa* possesses a significant effect on treating cardiovascular disease accumulation of iridoids in root tissues is thought to be responsible for its health advantages (Ma L. et al., 2021). Ma L. et al. (2021) reported a chromosome level of 2.49 Gb genome sequence of *R. glutinosa* and were assembled into 14 chromosomes using Illumina NovaSeq reads, Oxford Nanopore technology, and Hi-C data. Most iridoids exist mainly as glycosides, due to glycosylation being the final step in terpene biosynthesis. In addition, gene family evolution analysis revealed that 6,237 gene families in *R. glutinosa* underwent expansion, whereas 848 gene families underwent contraction. Eighty-seven TPS and 333 UGT genes were discovered in the expanded gene families, with the majority of them being significantly expressed in roots. The genomic resources provided by this study are essential for the molecular breeding of medicinal plants. Previous research reported that flavonoids derived from *Ziziphora bungeana* possess the ability to treat cardiovascular disease (Li et al., 2018). He et al. (2020) performed Illumina technology to sequence the transcriptome of *Z. bungeana* from different tissues and assembled 397,182 unigenes. Firstly, LC/MS detected 12 flavonoid components and linarin was one of the main bioactive compounds which highly accumulated in the inflorescence. Secondly, based on the transcriptome data, 18 candidate genes were identified from assembled unigenes that encode four key enzymes, including PAL, C4H, 4CL, and FNSII. Finally, qRT-PCR data revealed that *ZbPAL3*, *Zb4CL3*, *ZbCHS1*, *ZbFNSII*, and *ZbANS* had the highest expression levels in inflorescence, which suggested that they were likely related to the lignin biosynthesis.

CONCLUSION AND OUTLOOK

Research on medicinal plants has focused on the genetic mechanism of active ingredients synthesis and the subsequent pharmacological effects on diseases treatment. Thus, when researchers collect samples, they need to pay attention to the geographical distribution, different tissues and different developmental stages of the herb, due to these aspects play an important role in the efficacy of the pharmacological components (Duan et al., 2014; Zhang X. et al., 2017). The characteristics summary of the CHM is shown in **Table 2**. NGS had been widely applied in transcriptome sequencing of medicinal plants to get the expression profile and mined the genes related to bioactive compounds before the advent of TGS. However, due to the short read length, it was not discovering more candidate genes involved in the synthesis of active ingredients. With the aid of matured sequencing technologies, TGS is used to sequence the full-length transcriptome of medicinal plants, making a name for itself in herbal medicine research. *Salvia miltiorrhiza*, a well-known herbal medicine, is commonly used to treat diabetes. It has various biological functions, such as anti-oxidative stress and anti-inflammatory (Jia et al., 2019; Guo Y. et al., 2021; Yin et al., 2021). For instance, Xu et al. (2015) used SMRT *via* PacBio long reads to sequence RNA mixes from three root tissues of *S. miltiorrhiza* and obtained high-quality full-length transcriptome sequences. Furthermore, using Illumina short reads to quantify gene expression in three different tissues and obtained DEGs which elucidated the molecular mechanism of salvinone accumulation in the periderm. What's more, alternative splicing analysis on the full-length transcripts revealed that many terpenoids and isoprene biosynthesis genes underwent alternative splicing. Based on the above findings, the results suggest that these genes may regulate the diterpene metabolic pathway. Nevertheless, comprehensive analysis of multi-omics data is required for mining pharmacodynamic

TABLE 2 | Characteristics summary of the CHMs.

Name	Chinese name	Bioactive compounds	Representative medicinal tissue	Medical applications	Industrialized	References
Aconitum carmichaelii	乌头	Atisine, napelline, and aconitine	Roots	Treat pain, rheumatics, heart failure, diarrhea, beriberi, and edema	Yes	Mao et al., 2021
Akebia trifoliata	白木通	Oeanolic acid	Stems	Anti-inflammatory, antimicrobial, antioxidative, and anticancer	Yes	Huang H. et al., 2021
Ampelopsis grossedentata	显齿蛇葡萄	Dihydroquercetin and dihydromyricetin	Immature leaves	Antioxidants, anti-inflammatory, anti-microbial, anti-cancer, anti-Alzheimer, and anti-toxoplasmosis	Yes	Yu Z. et al., 2021
Andrographis paniculata	穿心莲	Neoandrographolide	Stems and leaves	Treat infectious, lung diseases, diabetes, and cancer	Yes	Li L. et al., 2017, 2019; Sun et al., 2019; Yue et al., 2019
Angelica sinensis	当归	Ferulicacid and soluble sugar	Roots	Treat irregular menstruation	Yes	Gao et al., 2021
Arctium lappa	牛蒡	Lignans	Fruits and roots	Treat common cold, measles, rubella, and mumps	Yes	Yang et al., 2021b
Areca catechu	槟榔	Arecoline	Seeds and nut	Treat nasal ulcers, Alzheimer's disease and obesity	Yes	Yang et al., 2021a
Arisaema heterophyllum	天南星	Terpineol, linoleic acid, schaftoside, and isoschaftoside	Tuber	Anti-tumor, antibacterial, anticonvulsant, analgesic, and anti-inflammatory	Yes	Wang C. et al., 2018
Artemisia annua	青蒿	Artemisinin	Whole plant	Treat uncomplicated malaria, tuberculosis, diabetes, and cancer	Yes	Shen et al., 2018
Artemisia argyi	艾蒿	Lupenone, β-amyrin, and simiarenol	Leaves	Antibacterial, anti-malarial, cough relief, anti-cancer, anti-diabetes, and anti-diarrhea	Yes	Cui et al., 2021
Asarum heterotropoides	库叶细辛	Aristolochic acid	Roots and rhizomes	Treat wind chill, headache, and cough	No	Wang X. et al., 2018
Asarum sieboldii	细辛	Asarinin and aristolochic acid	Whole plant	Treat cough, ache, inflammation, and cancer	Yes	Chen C. et al., 2021
Astragalus membranaceus	黄芪	Astragalosides, calycosin, and calycosin-7-O-β-D-glucoside	Roots	Treat cardiovascular disease, type 2 diabetes, nephritis, and cancers	Yes	Li J. et al., 2017
Berberis koreana	掌刺小檗	Berberine	Stems and roots	Antibacterial, anti-convulsion, sedative, anti-cholinergic, cholagogic, hepatoprotective, and anticancer	Yes	Roy et al., 2021
Callerya speciosa	鸡血藤	Maackiain and formononetin	Tuberous roots	Enhance immunity, hepatoprotection, arresting cough, and expectorant and anti-asthmatic	Yes	Yao et al., 2021
Camptotheca acuminata	喜树	Camptothecin	Leaves and fruits	Treat malignant tumors	Yes	Kang et al., 2021
Carthamus tinctorius	红花	Hydroxysafflor yellow A and linoleic acid	Flowers and seeds	Treat gynecological, cerebrovascular, and cardiovascular	Yes	Wu et al., 2021
Cassia obtusifolia	钝叶决明	Aurantio-obtusin	Seeds	Treat dizziness, headache, and diabetes	Yes	Deng Y. et al., 2018
Catharanthus roseus	长春花	Vinblastine and vincristine	Whole plant	Treat diabetes, cardio-vascular disorders, and anti-cancer	Yes	Verma et al., 2014
Chrysanthemum nankingense	菊花脑	Linalool, β-farnesene and lemon oil	Stems and leaves	Antioxidative, cardiovascular protective, vasorelaxant, and antivirus	Yes	Song et al., 2018

(Continued)

TABLE 2 | Continued

Name	Chinese name	Bioactive compounds	Representative medicinal tissue	Medical applications	Industrialized	References
Coptis chinensis	黄连	Berberine, coptisine, and worenine	Rhizomes	Treat cardiovascular, diabetes, cancer, and the nervous system diseases	Yes	Chen D. et al., 2021; Liu Y. et al., 2021
Cordyceps guangdongensis	冬虫夏草	Cordycepic acid and adenosine	Whole plant	Cure chronic renal failure	Yes	Zhang et al., 2018; Chen Y. et al., 2019
Corydalis yanhusuo	延胡索	Sanguinarine and noscapine	Tubers	Treat cardiovascular diseases, cancer, enhance energy, and provide analgesia	Yes	Xu D. et al., 2021
Crocus sativus	番红花	Apocarotenoid	Stigmas	Alleviate cramps, depression, anxiety, cardiovascular diseases, nervous disorders, and cancer	Yes	Yue et al., 2020
Curcuma longa	姜黄	Curcumin	Roots	Anti-inflammatory, antioxidant, and anti-cancer activities	Yes	Kim et al., 2022
Dendrobium officinale	铁皮石斛	Cyanidin, dihydroquercetin, fomononetin, biochanin A, and chorismate	Stems	Benefit the stomach, resist cancer, and enhance the body's immunity	Yes	Luo et al., 2019; Niu et al., 2021; Tao et al., 2021
Entada phaseoloides	榼藤	Palmitic acid and entada saponin	Stems	Dispel wind and body moisture, exhibit anti-inflammatory activity	No	Liao et al., 2020
Eriobotrya japonica	枇杷	Citric acid, nerolidol, and crategolic acid	Leaves	Treat inflammation, diabetes, cancer, bacterial infection, aging, pain, and allergy	Yes	Mogole et al., 2020; Wang, 2021
Eucommia ulmoides	杜仲	Chlorogenic acid	Leaves and bark	Reduce blood pressure and antibacterial, antiviral, antitumor, and antioxidant	Yes	Li et al., 2020
Euphorbia tirucalli	光棍树	Steroid	Stems	Treat sexual impotence, warts, epilepsy, toothache, hemorrhoids, and snake bites	Yes	Qiao et al., 2018
Fagopyrum tataricum	苦荞	Rutin	Tuberous roots	Reduce cholesterol levels, blood clots, and high blood pressure	Yes	Zhang L. et al., 2017
Fritillaria hupehensis	湖北贝母	Fritillaria	Bulbs	Treat hot-type bronchitis with dry cough and heart diseases, antitussive, anti-asthmatic, and expectorant	Yes	Guo K. et al., 2021
Ganoderma tsugae	灵芝	Ganoenic acid	Sub-entities	Antioxidant, anti-inflammation, anti-tumor, lipid lowering, and immunity enhancement	Yes	Jiang N. et al., 2021
Glycyrrhiza uralensis	甘草	Glycyrrhizin	Rhizomes	Anti-inflammatory, anticancer, antiallergic, and antiviral	Yes	Mochida et al., 2017; Lin et al., 2022
Gymnema sylvestre	匙羹藤	Gymnemic acid	Leaves	Treat diabetes, hypolipidemic3, stomachic, diuretic, refrigerant, and astringent properties	Yes	Hossain et al., 2016; Tiwari et al., 2017; Pham et al., 2018; Ayachit et al., 2019
Gynostemma pentaphyllum	绞股蓝	Gypenoside	Whole plant	Antitumor, hypoglycemic, hypolipidemic, cardiovascular and cerebrovascular protection, and immunoprotection	Yes	Liang et al., 2019

(Continued)

TABLE 2 | Continued

Name	Chinese name	Bioactive compounds	Representative medicinal tissue	Medical applications	Industrialized	References
Hemerocallis citrina	黄花菜	Rutin	Flower buds	Relieve depression and promote lactation, and anti-inflammatory	Yes	Liang et al., 2019
Impatiens balsamina	凤仙花	2-methoxy-1,4-naphthoquinone	Pods	Treat breast cancer, lung adenocarcinoma, and gastric adenocarcinoma	Yes	Foong et al., 2020
Lantana camara	马缨丹	Phenylpropanoid glycosides		Treat cancer, ulcers, asthma, and fever	Yes	Shah et al., 2020
Lithospermum erythrorhizon	紫草	Shikonin	Roots	Treat breast cancer and anti-HIV	Yes	Auber et al., 2020
Macleaya cordata	博落回	Sanguinarine and chelerythrine	Whole plant	Antimicrobial and animal growth-promoting	Yes	Liu X. et al., 2017
Magnolia biondii	望春花	Magnolin and fargesin	Flowers	Anesthesia, anti-inflammatory, analgesic, blood pressure-decreasing, and anti-allergic	Yes	Dong et al., 2021
Melia toosendan	川楝	Toosendanin	Fruits	Treat disease caused by a parasite	No	Lian et al., 2020
Murraya koenigii	咖喱树	Girinimbine, mahanimbine, murrayafoline, pyrafoline-D, sabinene, α-farnesene, and α-pinene	Leaves and roots	Treat asthma, prostate cancer, obesity,diarrhoea, dysentery, diabetes, and skin eruptions	No	Meena et al., 2017
Ophiopogon japonicus	浙麦冬	Methylophiopogonanone A and sprengerinin A	Tubers	Treat cardiovascular disease, cancer, and diabetes	Yes	He et al., 2016; Liu H. et al., 2017; Wu et al., 2019; Fan S. et al., 2020
Ophiorrhiza pumila	短小蛇根草	Camptothecin and pumiloside	Roots	Antitumor drug	No	Rai et al., 2021
Panax ginseng	人参	Ginsenosides	Roots and rhizomes	Cure central nervous system diseases, chronic diseases, and cancer	Yes	Xu et al., 2017; Xue et al., 2019
Panax notoginseng	三七	Ginsinoside and dencichine	Roots	Treat injury-induced trauma and cancer	Yes	Chen et al., 2017; Zhang D. et al., 2017; Fan G. et al., 2020; Zhang et al., 2020; Jiang Z. et al., 2021; Yang Z. et al., 2021
Papaver somniferum	罂粟	Morphine and codeine	Fruits	Painkilling drugs	Yes	Guo et al., 2018
Phellinus gilvus	桑黄	Agaricic acid, veratrine, and ergosterol	Sub entities	Treat stomachache, inflammation, tumor, and diabetes	Yes	Huo et al., 2020
Piper nigrum	黑胡椒	Piperine	Fruits	Treat diabetes and cancer	Yes	Hu et al., 2019
Platycodon grandiflorus	桔梗	Platycodin D and E and polygalacin D	Roots	Anti-inflammatory, antiobesity, anticancer, antiviral, and antiallergy	Yes	Kim et al., 2020; Yu H. et al., 2021
Pogostemon cablin	广藿香	Patchouli alcohol	Leaves	Remove moisture, prevent vomiting, and stimulate the appetite	Yes	Chen X. et al., 2019
Pueraria thomsonii	葛根	Puerarin and daidzein	Roots	Treat influenza, body stiffness, diabetes, and vascular hypertension	Yes	He et al., 2019
Rehmannia glutinosa	地黄	Catalpol, rehmaglutosides, and myobontiosideA	Tuberous roots	Strengthen cardiovascular, central nervous, immune, and visceral system	Yes	Ma L. et al., 2021
Salvia miltiorrhiza	丹参	Tanshinone IIA, tanshinone I, cryptotanshinone, and 15,16-dihydrotanshinone I	Roots	Treat coronary artery disease and cancer, anti-HIV	Yes	Jia et al., 2019; Guo Y. et al., 2021; Yin et al., 2021

(Continued)

TABLE 2 | Continued

Name	Chinese name	Bioactive compounds	Representative medicinal tissue	Medical applications	Industrialized	References
Salvia officinalis	鼠尾草	Tujone, camphor, and sugiol	Leaves	Treat choleretic and cancer	Yes	Jantová et al., 2014; Jiang et al., 2017
Scutellaria baicalensis	黄芩	Wogonin, baicalein, and baicalin	Roots	Treat liver, lung complaints, and cancer	Yes	Zhao et al., 2019
Senna tora	决明	Emodin, glucoaurantio-obtusin, physcion, chrysophanol, gluco-obtusifolin, aurantio-obtusin, obtusifolin, and obtusin	Seeds	Inhibit microbial and parasitic infections. Prevent neurodegenerative diseases and diabetes	No	Kang et al., 2020a,b
Siraitia grosvenorii	罗汉果	Mogroside	Fruits	Treat lung congestion, sore throat, and constipation	Yes	Xia et al., 2018
Sophora flavescens	苦参	Matrine, oxymatrine, sophorine, oxysophoridine, trifolirhizin, maackiain, kushenol I, kurarinone, and sophorafavanone G	Roots	Antiviral	Yes	He et al., 2015; Wei et al., 2021
Strobilanthes cusia	板蓝	Indigo and indirubin	Rhizomesand roots	Treat oral ulcers, skin diseases, influenza A infection, and mumps	Yes	Hu et al., 2021
Swertia mussotii	獐牙菜	Secoiridoids, swertiamarin, mangiferin, gentiopicroside, sweroside, oleanolic acid, and xanthones	Whole plant	Treat hepatitis, diabetes, and hyperlipidaemic	Yes	Liu Y. et al., 2017
Taxus chinensis	红豆杉	Paclitaxel	Leaves	Anticancer drug	Yes	Xiong et al., 2021
Trachyspermum ammi	阿育魏实	Thymol, γ-terpinene, and p-cymene	Seeds	Antifungal, antibacterial, antivirus, and anti-inflammatory	No	Soltani Howyzeh et al., 2018
Trillium govanianum	西藏延龄草	Steroidal saponin	Rhizomes	Cure menstrual, dysentery, headache, and fever	No	Singh et al., 2017
Tripterygium wilfordii	雷公藤	Triptolide and celastrol	Leaves	Treat central nervous system diseases and cancer	Yes	Tu et al., 2020
Zingiber officinale	生姜	Gingerols	Rhizomes	Inhibit hyperproliferation, inflammation, and carcinogenesis	Yes	Li H. et al., 2021
Ziziphora bungeana	白枣	Linarin, cafeic acid, and rosmarinic acid	Whole plant	Treat fever, headache, insomnia, and cardiovascular disease	Yes	He et al., 2020

genes more accurately, such as genomics, transcriptomics, metabolomics, and so on (Ma et al., 2020). Ma Y. et al. (2021) revealed the *S. miltiorrhiza* genome with 622 Mb using Illumina Hiseq2000 and PacBio RS platform and obtained 33,760 protein-coding genes. CYP450 genes were discovered to be present as a gene cluster after a thorough study of genomic data. Based on this, a CYP71D subfamily, which is significantly expanded in the *S. miltiorrhiza* genome, was identified using gene expansion and contraction analysis, and four candidate genes were targeted by co-expression analysis for enzyme activity and RNAi studies. The results showed that three of these genes play important roles in the tanshinone biosynthetic pathway, two of which can catalyze the generation of tanshinone characteristic furan rings and one is associated with the hydroxylation process at the C20 position of tanshinones. Based on previous literature, the researchers used the genome-assembly and the downstream bioinformatics analysis to uncover key enzymes genes involved in the biosynthesis pathway of active ingredient. Furthermore, the validation methods for functional genes are shown in **Table 1**.

Regarding future practical applications, on one hand, combining 2nd and 3rd generation sequencing technologies makes full use of the strengths of each, such as long read length, high throughput, and acceptable sequencing costs. On the other hand, single-cell sequencing technology has great application prospects in the mining of active ingredients of Chinese traditional medicinal herbs. Traditional RNA-Seq technologies sequence RNA extracted from a variety of tissues and cells, ignoring intercellular differences. Single-cell RNA sequencing isolates the target cells from the sample and then sequences them, allowing for the unique characteristics of individual cells. Additionally, scRNA-seq has yielded rich results in the fields of tumors, microorganisms, and so on (Li S. et al., 2021). Xu X. et al. (2021) used scRNA-seq to mine trait genes during maize development. It is more accurate and efficient than traditional RNA-seq and uncovers more information.

AUTHOR CONTRIBUTIONS

This study was designed by MT and GL. Data analysis was performed by JLL. The figures were organized by YL and HTZ. The manuscript was written by XYL and XG. The manuscript was revised by SQ. All authors made a direct and intellectual contribution to this topic and approved the article for publication.

FUNDING

This work was financially supported by grants from the National Natural Science Foundation of China (32002235).

ACKNOWLEDGMENTS

We thank all the individuals who have helped us in this study. We acknowledge the valuable work of the many investigators whose published articles we were unable to cite owing to space limitations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.900035/full#supplementary-material

REFERENCES

Ali, M., Li, P., She, G., Chen, D., Wan, X., and Zhao, J. (2017). Transcriptome and metabolite analyses reveal the complex metabolic genes involved in volatile terpenoid biosynthesis in garden sage (*Salvia officinalis*). *Sci. Rep.* 7:16074. doi: 10.1038/s41598-017-15478-3

Auber, R. P., Suttiyut, T., McCoy, R. M., Ghaste, M., Crook, J. W., Pendleton, A. L., et al. (2020). Hybrid de novo genome assembly of red gromwell (*Lithospermum erythrorhizon*) reveals evolutionary insight into shikonin biosynthesis. *Hortic. Res.* 7:82. doi: 10.1038/s41438-020-0301-9

Ayachit, G., Shaikh, I., Sharma, P., Jani, B., Shukla, L., Sharma, P., et al. (2019). De novo transcriptome of *Gymnema sylvestre* identified putative lncRNA and genes regulating terpenoid biosynthesis pathway. *Sci. Rep.* 9:14876. doi: 10.1038/s41598-019-51355-x

Barbosa, C., Nogueira, S., Gadanho, M., and Chaves, S. (2019). Study on commercial spice and herb products using next-generation sequencing (NGS). *J. AOAC Int.* 102, 369–375. doi: 10.5740/jaoacint.18-0407

Chakraborty, A., Mahajan, S., Jaiswal, S. K., and Sharma, V. K. (2021). Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* 4:1193. doi: 10.1038/s42003-021-02720-y

Chen, W., Kui, L., Zhang, G., Zhu, S., Zhang, J., Wang, X., et al. (2017). Whole-genome sequencing and analysis of the Chinese herbal plant Panax notoginseng. *Mol. Plant* 10, 899–902. doi: 10.1016/j.molp.2017.02.010

Chen, X., Li, J., Wang, X., Zhong, L., Tang, Y., Zhou, X., et al. (2019). Full-length transcriptome sequencing and methyl jasmonate-induced expression profile analysis of genes related to patchoulol biosynthesis and regulation in *Pogostemon cablin*. *BMC Plant Biol.* 19:266. doi: 10.1186/s12870-019-1884-x

Chen, D. X., Pan, Y., Wang, Y., Cui, Y. Z., Zhang, Y. J., Mo, R. Y., et al. (2021). The chromosome-level reference genome of *Coptis chinensis* provides insights into genomic evolution and berberine biosynthesis. *Hortic. Res.* 8:121. doi: 10.1038/s41438-021-00559-2

Chen, C., Shi, X., Zhou, T., Li, W., Li, S., and Bai, G. (2021). Full-length transcriptome analysis and identification of genes involved in asarinin and aristolochic acid biosynthesis in medicinal plant *Asarum sieboldii*. *Genome* 64, 639–653. doi: 10.1139/gen-2020-0095

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

Chen, Y., Wu, Y., Liu, L., Feng, J., Zhang, T., Qin, S., et al. (2019). Study of the whole genome, methylome and transcriptome of *Cordyceps militaris*. *Sci. Rep.* 9:898. doi: 10.1038/s41598-018-38021-4

Chen, S., Xu, J., Liu, C., Zhu, Y., Nelson, D. R., Zhou, S., et al. (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat. Commun.* 3:913. doi: 10.1038/ncomms1923

Chen, X., Yu, J., and Shi, J. (2018). Management of diabetes mellitus with puerarin, a natural isoflavone from *Pueraria lobata*. *Am. J. Chin. Med.* 46, 1771–1789. doi: 10.1142/S0192415X18500891

Cheng, J., Chen, J., Liu, X., Li, X., Zhang, W., Dai, Z., et al. (2021). The origin and evolution of the diosgenin biosynthetic pathway in yam. *Plant Commun.* 2:100079. doi: 10.1016/j.xplc.2020.100079

Cheng, Q. Q., Ouyang, Y., Tang, Z. Y., Lao, C. C., Zhang, Y. Y., Cheng, C. S., et al. (2021). Review on the development and applications of medicinal plant genomes. *Front. Plant Sci.* 12:791219. doi: 10.3389/fpls.2021.791219

Ciumarnean, L., Milaciu, M. V., Runcan, O., Vesa, S. C., Rachisan, A. L., Negrean, V., et al. (2020). The effects of flavonoids in cardiovascular diseases. *Molecules* 25:4320. doi: 10.3390/molecules25184320

Cui, Y., Gao, X., Wang, J., Shang, Z., Zhang, Z., Zhou, Z., et al. (2021). Full-length transcriptome analysis reveals candidate genes involved in terpenoid biosynthesis in *Artemisia argyi*. *Front. Genet.* 12:659962. doi: 10.3389/fgene.2021.659962

Cui, F., Ye, X., Li, X., Yang, Y., Hu, Z., Overmyer, K., et al. (2022). Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii* provides insights into its subtropical adaptation and cuticle synthesis. *Plant Commun.*:100307. doi: 10.1016/j.xplc.2022.100307

Deng, X., Zhao, L., Fang, T., Xiong, Y., Ogutu, C., Yang, D., et al. (2018). Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. *Hortic. Res.* 5:29. doi: 10.1038/s41438-018-0035-0

Deng, Y., Zheng, H., Yan, Z., Liao, D., Li, C., Zhou, J., et al. (2018). Full-length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin biosynthesis, seed formation and development, and stress response. *Int. J. Mol. Sci.* 19:2476. doi: 10.3390/ijms19092476

Dong, S., Liu, M., Liu, Y., Chen, F., Yang, T., Chen, L., et al. (2021). The genome of *Magnolia biondii* Pamp. Provides insights into the evolution of magnoliales and biosynthesis of terpenoids. *Hortic. Res.* 8:38. doi: 10.1038/s41438-021-00471-9

Duan, X., Zhang, D., Nie, L., and Zang, H. (2014). Rapid discrimination of geographical origin and evaluation of antioxidant activity of *Salvia miltiorrhiza* var. alba by Fourier transform near infrared spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 122, 751–757. doi: 10.1016/j.saa.2013.12.003

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

Fan, G., Liu, X., Sun, S., Shi, C., Du, X., Han, K., et al. (2020). The chromosome level genome and genome-wide association study for the agronomic traits of *Panax notoginseng*. *iScience* 23:101538. doi: 10.1016/j.isci.2020.101538

Fan, S., Zhang, J., Xiao, Q., Liu, P., Zhang, Y., Yao, E., et al. (2020). Cardioprotective effect of the polysaccharide from *Ophiopogon japonicus* on isoproterenol-induced myocardial ischemia in rats. *Int. J. Biol. Macromol.* 147, 233–240. doi: 10.1016/j.ijbiomac.2020.01.068

Feng, X., Cao, S., Qiu, F., and Zhang, B. (2020). Traditional application and modern pharmacological research of *Artemisia annua* L. *Pharmacol. Ther.* 216:107650. doi: 10.1016/j.pharmthera.2020.107650

Feng, L., Zhu, M. M., Zhang, M. H., Wang, R. S., Tan, X. B., Song, J., et al. (2013). Protection of glycyrrhizic acid against AGEs-induced endothelial dysfunction through inhibiting RAGE/NF-κB pathway activation in human

umbilical vein endothelial cells. *J. Ethnopharmacol.* 148, 27–36. doi: 10.1016/j.jep.2013.03.035

Flower, A., Liu, J. P., Lewith, G., Little, P., and Li, Q. (2012). Chinese herbal medicine for endometriosis. *Cochrane Database Syst. Rev.*:CD006568. doi: 10.1002/14651858.CD006568.pub3

Foong, L. C., Chai, J. Y., Ho, A. S. H., Yeo, B. P. H., Lim, Y. M., and Tam, S. M. (2020). Comparative transcriptome analysis to identify candidate genes involved in 2-methoxy-1,4-naphthoquinone (MNQ) biosynthesis in *Impatiens balsamina* L. *Sci. Rep.* 10:16123. doi: 10.1038/s41598-020-72997-2

Gao, X., Guo, F., Chen, Y., Bai, G., Liu, Y., Jin, J., et al. (2021). Full-length transcriptome analysis provides new insights into the early bolting occurrence in medicinal *Angelica sinensis*. *Sci. Rep.* 11:13000. doi: 10.1038/s41598-021-92494-4

Guo, K., Chen, J., Niu, Y., and Lin, X. (2021). Full-length transcriptome sequencing provides insights into flavonoid biosynthesis in *Fritillaria hupehensis*. *Life* 11:287. doi: 10.3390/life11040287

Guo, J., Huang, Z., Sun, J., Cui, X., and Liu, Y. (2021a). Research Progress and future development trends in medicinal plant Transcriptomics. *Front. Plant Sci.* 12:691838. doi: 10.3389/fpls.2021.691838

Guo, J., Li, J., Wei, H., and Liang, Z. (2021b). Maackiain protects the kidneys of type 2 diabetic rats via modulating the Nrf2/HO-1 and TLR4/NF-κB/Caspase-3 pathways. *Drug Des. Devel. Ther.* 15, 4339–4358. doi: 10.2147/DDDT.S326975

Guo, Y., Sun, J., Zhang, R., Yang, P., Zhang, S., and Wu, Z. (2021). *Salvia miltiorrhiza* improves type 2 diabetes: a protocol for systematic review and meta-analysis. *Medicine* 100:e23843. doi: 10.1097/MD.0000000000023843

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

He, X., Fang, J., Huang, L., Wang, J., and Huang, X. (2015). Sophora flavescens Ait.: traditional usage, phytochemistry and pharmacology of an important traditional Chinese medicine. *J. Ethnopharmacol.* 172, 10–29. doi: 10.1016/j.jep.2015.06.010

He, F., Xu, B. L., Chen, C., Jia, H. J., Wu, J. X., Wang, X. C., et al. (2016). Methylophiopogonanone A suppresses ischemia/reperfusion-induced myocardial apoptosis in mice via activating PI3K/Akt/eNOS signaling pathway. *Acta Pharmacol. Sin.* 37, 763–771. doi: 10.1038/aps.2016.14

He, J., Yang, W., Cheng, B., Ma, L., Tursunjiang, D., Ding, Z., et al. (2020). Integrated metabolomic and transcriptomic profiling reveals the tissue-specific flavonoid compositions and their biosynthesis pathways in *Ziziphora bungeana*. *Chin. Med.* 15:73. doi: 10.1186/s13020-020-00354-6

He, M., Yao, Y., Li, Y., Yang, M., Li, Y., Wu, B., et al. (2019). Comprehensive transcriptome analysis reveals genes potentially involved in isoflavone biosynthesis in *Pueraria thomsonii* benth. *PLoS One* 14:e0217593. doi: 10.1371/journal.pone.0217593

Hossain, M. U., Khan, M. A., Rakib-Uz-Zaman, S. M., Ali, M. T., Islam, M. S., Keya, C. A., et al. (2016). Treating diabetes mellitus: pharmacophore based designing of potential drugs from *Gymnema sylvestre* against insulin receptor protein. *Biomed. Res. Int.* 2016:3187647. doi: 10.1155/2016/3187647

Hu, Y., Ma, D., Ning, S., Ye, Q., Zhao, X., Ding, Q., et al. (2021). High-quality genome of the medicinal plant *Strobilanthes cusia* provides insights Into the biosynthesis of Indole alkaloids. *Front. Plant Sci.* 12:742420. doi: 10.3389/fpls.2021.742420

Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat. Commun.* 10:4702. doi: 10.1038/s41467-019-12607-6

Huang, H., Liang, J., Tan, Q., Ou, L., Li, X., Zhong, C., et al. (2021). Insights into triterpene synthesis and unsaturated fatty-acid accumulation provided by chromosomal-level genome analysis of Akebia trifoliata subsp. australis. *Hortic. Res.* 8:33. doi: 10.1038/s41438-020-00458-y

Huang, P., Zhou, P., Liang, Y., Wu, J., Wu, G., Xu, R., et al. (2021). Exploring the molecular targets and mechanisms of [10]-gingerol for treating triple-negative breast cancer using bioinformatics approaches, molecular docking, and in vivo experiments. *Transl. Cancer Res.* 10, 4680–4693. doi: 10.21037/tcr-21-1138

Huo, J., Zhong, S., Du, X., Cao, Y., Wang, W., Sun, Y., et al. (2020). Whole-genome sequence of *Phellinus gilvus* (mulberry Sanghuang) reveals its unique medicinal values. *J. Adv. Res.* 24, 325–335. doi: 10.1016/j.jare.2020.04.011

Hurgobin, B., Tamiru-Oli, M., Welling, M. T., Doblin, M. S., Bacic, A., Whelan, J., et al. (2021). Recent advances in *Cannabis sativa* genomics research. *New Phytol.* 230, 73–89. doi: 10.1111/nph.17140

Jantová, S., Hudec, R., Sekretár, S., Kučerák, J., and Melušová, M. (2014). *Salvia officinalis* L. extract and its new food antioxidant formulations induce apoptosis through mitochondrial/caspase pathway in leukemia L1210 cells. *Interdiscip. Toxicol.* 7, 146–153. doi: 10.2478/intox-2014-0020

Ji, X., Shi, S., Liu, B., Shan, M., Tang, D., Zhang, W., et al. (2019). Bioactive compounds from herbal medicines to manage dyslipidemia. *Biomed. Pharmacother.* 118:109338. doi: 10.1016/j.biopha.2019.109338

Jia, Q., Zhu, R., Tian, Y., Chen, B., Li, R., Li, L., et al. (2019). *Salvia miltiorrhiza* in diabetes: a review of its pharmacology, phytochemistry, and safety. *Phytomedicine* 58:152871. doi: 10.1016/j.phymed.2019.152871

Jiang, N., Hu, S., Peng, B., Li, Z., Yuan, X., Xiao, S., et al. (2021). Genome of Ganoderma species provides insights Into the evolution, conifers substrate utilization, and terpene synthesis for *Ganoderma tsugae*. *Front. Microbiol.* 12:724451. doi: 10.3389/fmicb.2021.724451

Jiang, Z., Tu, L., Yang, W., Zhang, Y., Hu, T., Ma, B., et al. (2021). The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Commun.* 2:100113. doi: 10.1016/j.xplc.2020.100113

Jiang, Y., Zhang, L., and Rupasinghe, H. P. (2017). Antiproliferative effects of extracts from *Salvia officinalis* L. and saliva miltiorrhiza Bunge on hepatocellular carcinoma cells. *Biomed. Pharmacother.* 85, 57–67. doi: 10.1016/j.biopha.2016.11.113

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., Lee, W. H., Lee, C. M., Sim, J. S., Won, S. Y., Han, S. R., et al. (2020a). De novo transcriptome sequence of *Senna tora* provides insights into anthraquinone biosynthesis. *PLoS One* 15:e0225564. doi: 10.1371/journal.pone.0225564

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

Kim, J., Kang, S. H., Park, S. G., Yang, T. J., Lee, Y., Kim, O. T., et al. (2020). Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in *Platycodon grandiflorus*, a medicinal plant. *Hortic. Res.* 7:112. doi: 10.1038/s41438-020-0329-x

Kim, M. J., Park, S. Y., Kim, Y., Jeon, S., Cha, M. S., Kim, Y. J., et al. (2022). Beneficial effects of a combination of *Curcuma longa* L. and citrus junos against beta-amyloid peptide-induced neurodegeneration in mice. *J. Med. Food* 25, 12–23. doi: 10.1089/jmf.2021.K.0104

Korth, K. L., Stermer, B. A., Bhattacharyya, M. K., and Dixon, R. A. (1997). HMG-CoA reductase gene families that differentially accumulate transcripts in potato tubers are developmentally expressed in floral tissues. *Plant Mol. Biol.* 33, 545–551. doi: 10.1023/A:1005743011651

Lee, Y. S., Kim, S. H., Jung, S. H., Kim, J. K., Pan, C. H., and Lim, S. S. (2010). Aldose reductase inhibitory compounds from *Glycyrrhiza uralensis*. *Biol. Pharm. Bull.* 33, 917–921. doi: 10.1248/bpb.33.917

Li, J. H., Chen, Z. X., Zhang, X. G., Li, Y., Yang, W. T., Zheng, X. W., et al. (2016). Bioactive components of Chinese herbal medicine enhance endogenous neurogenesis in animal models of ischemic stroke: a systematic analysis. *Medicine (Baltimore)* 95:e4904. doi: 10.1097/MD.0000000000004904

Li, J., Harata-Lee, Y., Denton, M. D., Feng, Q., Rathjen, J. R., Qu, Z., et al. (2017). Long read reference genome-free reconstruction of a full-length transcriptome from astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov.* 3:17031. doi: 10.1038/celldisc.2017.31

Li, X., Liang, Z., Du, J., Wang, Z., Mei, S., Li, Z., et al. (2019). Herbal decoctosome is a novel form of medicine. *Sci. China Life Sci.* 62, 333–348. doi: 10.1007/s11427-018-9508-0

Li, Q., Tursun, D., Shi, C., Heyrulla, M., Zhang, X., and Yang, W. (2018). Ziziphora clinopodioides flavonoids protect myocardial cell damage from myocardial ischemia-reperfusion injury. *Evid. Based Complement. Alternat. Med.* 2018:8495010. doi: 10.1155/2018/8495010

Li, Y., Wei, H., Yang, J., Du, K., Li, J., Zhang, Y., et al. (2020). High-quality de novo assembly of the *Eucommia ulmoides* haploid genome provides new

insights into evolution and rubber biosynthesis. *Hortic. Res.* 7:183. doi: 10.1038/s41438-020-00406-w

Li, H. L., Wu, L., Dong, Z., Jiang, Y., Jiang, S., Xing, H., et al. (2021). Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic. Res.* 8:189. doi: 10.1038/s41438-021-00627-7

Li, S., Yan, H., and Lee, J. (2021). Identification of gene regulatory networks from single-cell expression data. *Methods Mol. Biol.* 2328, 153–170. doi: 10.1007/978-1-0716-1534-8_9

Li, T., Yu, X., Ren, Y., Kang, M., Yang, W., Feng, L., et al. (2022). The chromosome-level genome assembly of Gentiana dahurica (Gentianaceae) provides insights into gentiopicroside biosynthesis. *DNA Res.* 29. doi: 10.1093/dnares/dsac008

Li, L., Yue, G. G. L., Fung, K. P., Yu, J., Lau, C. B. S., and Chiu, P. W. Y. (2019). Combined use of Andrographis paniculata and chemotherapeutics for metastatic oesophageal cancer: a pre-clinical study. *Hong Kong Med. J.* 25, 43–46.

Li, L., Yue, G. G., Lee, J. K., Wong, E. C., Fung, K. P., Yu, J., et al. (2017). The adjuvant value of *Andrographis paniculata* in metastatic esophageal cancer treatment - from preclinical perspectives. *Sci. Rep.* 7:854. doi: 10.1038/s41598-017-00934-x

Lian, X., Zhang, X., Wang, F., Wang, X., Xue, Z., and Qi, X. (2020). Characterization of a 2,3-oxidosqualene cyclase in the toosendanin biosynthetic pathway of *Melia toosendan*. *Physiol. Plant.* 170, 528–536. doi: 10.1111/ppl.13189

Liang, T., Zou, L., Sun, S., Kuang, X., Wei, J., Wang, L., et al. (2019). Hybrid sequencing of the *Gynostemma pentaphyllum* transcriptome provides new insights into gypenoside biosynthesis. *BMC Genomics* 20:632. doi: 10.1186/s12864-019-6000-y

Liao, W., Mei, Z., Miao, L., Liu, P., and Gao, R. (2020). Comparative transcriptome analysis of root, stem, and leaf tissues of *Entada phaseoloides* reveals potential genes involved in triterpenoid saponin biosynthesis. *BMC Genomics* 21:639. doi: 10.1186/s12864-020-07056-1

Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 47–65. doi: 10.1146/annurev.arplant.50.1.47

Lin, H. C., Paul, C. R., Kuo, C. H., Chang, Y. H., Chen, W. S., Ho, T. J., et al. (2022). *Glycyrrhiza uralensis* root extract ameliorates high glucose-induced renal proximal tubular fibrosis by attenuating tubular epithelial-myofibroblast transdifferentiation by targeting TGF-β1/Smad/Stat3 pathway. *J. Food Biochem.* 46:e14041. doi: 10.1111/jfbc.14041

Liu, W., Feng, Y., Yu, S., Fan, Z., Li, X., Li, J., et al. (2021). The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* 22:12824. doi: 10.3390/ijms222312824

Liu, H., Guo, S. S., Lu, L., Li, D., Liang, J., Huang, Z. H., et al. (2021). Essential oil from *Artemisia annua* aerial parts: composition and repellent activity against two storage pests. *Nat. Prod. Res.* 35, 822–825. doi: 10.1080/14786419.2019.1599887

Liu, X., Liu, Y., Huang, P., Ma, Y., Qing, Z., Tang, Q., et al. (2017). The genome of medicinal plant *Macleaya cordata* provides new insights into benzylisoquinoline alkaloids metabolism. *Mol. Plant* 10, 975–989. doi: 10.1016/j.molp.2017.05.007

Liu, Y., Wang, Y., Guo, F., Zhan, L., Mohr, T., Cheng, P., et al. (2017). Deep sequencing and transcriptome analyses to identify genes involved in secoiridoid biosynthesis in the tibetan medicinal plant *Swertia mussotii*. *Sci. Rep.* 7:43108. doi: 10.1038/srep43108

Liu, Y., Wang, B., Shu, S., Li, Z., Song, C., Liu, D., et al. (2021). Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. *Nat. Commun.* 12:3276. doi: 10.1038/s41467-021-23611-0

Liu, H., Wang, Y., Wang, T., Ying, X., Wu, R., and Chen, H. (2017). De novo assembly and annotation of the Zhe-Maidong (*Ophiopogon japonicus* (L.f.) Ker-Gawl) transcriptome in different growth stages. *Sci. Rep.* 7:3616. doi: 10.1038/s41598-017-03937-w

Liu, C., Yang, S., Wang, K., Bao, X., Liu, Y., Zhou, S., et al. (2019). Alkaloids from traditional Chinese medicine against hepatocellular carcinoma. *Biomed. Pharmacother.* 120:109543. doi: 10.1016/j.biopha.2019.109543

Lo, Y. T., and Shaw, P. C. (2019). Application of next-generation sequencing for the identification of herbal products. *Biotechnol. Adv.* 37:107450. doi: 10.1016/j.biotechadv.2019.107450

Lu, Z., Zhong, Y., Liu, W., Xiang, L., and Deng, Y. (2019). The efficacy and mechanism of Chinese herbal medicine on diabetic kidney disease. *J. Diabetes Res.* 2019:2697672. doi: 10.1155/2019/2697672

Luo, D., Dong, X., Huang, J., Huang, C., Fang, G., and Huang, Y. (2021). *Pueraria lobata* root polysaccharide alleviates glucose and lipid metabolic dysfunction in diabetic db/db mice. *Pharm. Biol.* 59, 382–390. doi: 10.1080/13880209.2021.1898648

Luo, Y., Ren, Z., Du, B., Xing, S., Huang, S., Li, Y., et al. (2019). Structure identification of ViceninII extracted from *Dendrobium officinale* and the reversal of TGF-β1-induced epithelial-Mesenchymal transition in lung adenocarcinoma cells through TGF-β/Smad and PI3K/Akt/mTOR signaling pathways. *Molecules* 24:144. doi: 10.3390/molecules24010144

Ma, Y., Cui, G., Chen, T., Ma, X., Wang, R., Jin, B., et al. (2021). Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. *Nat. Commun.* 12:685. doi: 10.1038/s41467-021-20959-1

Ma, L., Dong, C., Song, C., Wang, X., Zheng, X., Niu, Y., et al. (2021). De novo genome assembly of the potent medicinal plant *Rehmannia glutinosa* using nanopore technology. *Comput. Struct. Biotechnol. J.* 19, 3954–3963. doi: 10.1016/j.csbj.2021.07.006

Ma, X., Meng, Y., Wang, P., Tang, Z., Wang, H., and Xie, T. (2020). Bioinformatics-assisted, integrated omics studies on medicinal plants. *Brief. Bioinform.* 21, 1857–1874. doi: 10.1093/bib/bbz132

Mao, L., Jin, B., Chen, L., Tian, M., Ma, R., Yin, B., et al. (2021). Functional identification of the terpene synthase family involved in diterpenoid alkaloids biosynthesis in *Aconitum carmichaelii*. *Acta Pharm. Sin. B* 11, 3310–3321. doi: 10.1016/j.apsb.2021.04.008

Meena, S., Rajeev Kumar, S., Dwivedi, V., Kumar Singh, A., Chanotiya, C. S., Akhtar, 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

Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., et al. (2017). Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *Plant J.* 89, 181–194. doi: 10.1111/tpj.13385

Mogole, L., Omwoyo, W., and Mtunzi, F. (2020). Phytochemical screening, anti-oxidant activity and α-amylase inhibition study using different extracts of loquat (*Eriobotrya japonica*) leaves. *Heliyon* 6:e04736. doi: 10.1016/j.heliyon.2020.e04736

Nair, P., Mall, M., Sharma, P., Khan, F., Nagegowda, D. A., Rout, P. K., et al. (2019). Characterization of a class III peroxidase from *Artemisia annua*: relevance to artemisinin metabolism and beyond. *Plant Mol. Biol.* 100, 527–541. doi: 10.1007/s11103-019-00879-x

Niu, Z., Zhu, F., Fan, Y., Li, C., Zhang, B., Zhu, S., et al. (2021). The chromosome-level reference genome assembly for *Dendrobium officinale* and its utility of functional genomics research and molecular breeding study. *Acta Pharm. Sin. B* 11, 2080–2092. doi: 10.1016/j.apsb.2021.01.019

Pedro, D. F., Ramos, A. A., Lima, C. F., Baltazar, F., and Pereira-Wilson, C. (2016). Colon Cancer chemoprevention by sage tea drinking: decreased DNA damage and cell proliferation. *Phytother. Res.* 30, 298–305. doi: 10.1002/ptr.5531

Pham, H. T. T., Hoang, M. C., Ha, T. K. Q., Dang, L. H., Tran, V. O., Nguyen, T. B. T., et al. (2018). Discrimination of different geographic varieties of *Gymnema sylvestre*, an anti-sweet plant used for the treatment of type 2 diabetes. *Phytochemistry* 150, 12–22. doi: 10.1016/j.phytochem.2018.02.013

Qiao, W., Li, C., Mosongo, I., Liang, Q., Liu, M., and Wang, X. (2018). Comparative transcriptome analysis identifies putative genes involved in steroid biosynthesis in *Euphorbia tirucalli*. *Gene* 9:38. doi: 10.3390/genes9010038

Qing, Z., Liu, J., Yi, X., Liu, X., Hu, G., Lao, J., et al. (2021). The chromosome-level *Hemerocallis citrina* Borani genome provides new insights into the rutin biosynthesis and the lack of colchicine. *Hortic. Res.* 8:89. doi: 10.1038/s41438-021-00539-6

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

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565–574. doi: 10.1039/a709175c

Roy, N. S., Choi, I. Y., Um, T., Jeon, M. J., Kim, B. Y., Kim, Y. D., et al. (2021). Gene expression and isoform identification of PacBio full-length cDNA sequences for berberine biosynthesis in *Berberis koreana*. *Plants* 10:1314. doi: 10.3390/plants10071314

Sabzehzari, M., Zeinali, M., and Naghavi, M. R. (2020). Alternative sources and metabolic engineering of Taxol: advances and future perspectives. *Biotechnol. Adv.* 43:107569. doi: 10.1016/j.biotechadv.2020.107569

Shah, M., Alharby, H. F., Hakeem, K. R., Ali, N., Rahman, I. U., Munawar, M., et al. (2020). De novo transcriptome analysis of *Lantana camara* L. revealed candidate genes involved in phenylpropanoid biosynthesis pathway. *Sci. Rep.* 10:13726. doi: 10.1038/s41598-020-70635-5

Sharma, M., Sharma, P. D., Bansal, M. P., and Singh, J. (2007). Synthesis, cytotoxicity, and antitumor activity of lantadene: a congeners. *Chem. Biodivers.* 4, 932–939. doi: 10.1002/cbdv.200790082

Shen, Q., Zhang, L., Liao, Z., Wang, S., Yan, T., Shi, P., et al. (2018). The genome of *Artemisia annua* provides insight into the evolution of Asteraceae Family and Artemisinin biosynthesis. *Mol. Plant* 11, 776–788. doi: 10.1016/j.molp.2018.03.015

Singh, P., Singh, G., Bhandawat, A., Singh, G., Parmar, R., Seth, R., et al. (2017). Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*. *Sci. Rep.* 7:45295. doi: 10.1038/srep45295

Skubnik, J., Pavlickova, V. S., Ruml, T., and Rimpelova, S. (2021). Vincristine in combination therapy of cancer: emerging trends in clinics. *Biology* 10:849. doi: 10.3390/biology10090849

Soltani Howyzeh, M., Sadat Noori, S. A., Shariati, J. V., and Amiripour, M. (2018). Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. *Sci. Rep.* 8:13405. doi: 10.1038/s41598-018-31618-9

Song, C., Liu, Y., Song, A., Dong, G., Zhao, H., Sun, W., et al. (2018). The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of Chrysanthemum flowers and medicinal traits. *Mol. Plant* 11, 1482–1491. doi: 10.1016/j.molp.2018.10.003

Sun, W., Leng, L., Yin, Q., Xu, M., Huang, M., Xu, Z., et al. (2019). The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J.* 97, 841–857. doi: 10.1111/tpj.14162

Tao, S., Ren, Z., Yang, Z., Duan, S., Wan, Z., Huang, J., et al. (2021). Effects of different molecular weight polysaccharides from *Dendrobium officinale* Kimura & Migo on human colorectal cancer and transcriptome analysis of differentially expressed genes. *Front. Pharmacol.* 12:704486. doi: 10.3389/fphar.2021.704486

Tiwari, P., Ahmad, K., and Baig, M. H. (2017). *Gymnema sylvestre* for diabetes: from traditional herb to future's therapeutic. *Curr. Pharm. Des.* 23, 1667–1676. doi: 10.2174/1381612823666161108162048

Tu, L., Su, P., Zhang, Z., Gao, L., Wang, J., Hu, T., et al. (2020). Genome of *Tripterygium wilfordii* and identification of cytochrome P450 involved in triptolide biosynthesis. *Nat. Commun.* 11:971. doi: 10.1038/s41467-020-14776-1

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-r102

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008

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

Wang, Y. (2021). A draft genome, resequencing, and metabolomes reveal the genetic background and molecular basis of the nutritional and medicinal properties of loquat (*Eriobotrya japonica* (Thunb.) Lindl). *Hortic. Res.* 8:231. doi: 10.1038/s41438-021-00657-1

Wang, X., Hui, F., Yang, Y., and Yang, S. (2018). Deep sequencing and transcriptome analysis to identify genes related to biosynthesis of aristolochic acid in *Asarum heterotropoides*. *Sci. Rep.* 8:17850. doi: 10.1038/s41598-018-36316-0

Wang, J., Li, J., Li, Z., Liu, B., Zhang, L., Guo, D., et al. (2022). Genomic insights into longan evolution from a chromosome-level genome assembly and population genomics of longan accessions. *Hortic. Res.* 9:uhac021. doi: 10.1093/hr/uhac021

Wang, C., Zhu, J., Liu, M., Yang, Q., Wu, J., and Li, Z. (2018). De novo sequencing and transcriptome assembly of *Arisaema heterophyllum* Blume and identification of genes involved in isoflavonoid biosynthesis. *Sci. Rep.* 8:17643. doi: 10.1038/s41598-018-35664-1

Wei, G., Chen, Y., Guo, X., Wei, J., Dong, L., and Chen, S. (2021). Biosyntheses characterization of alkaloids and flavonoids in *Sophora flavescens* by combining metabolome and transcriptome. *Sci. Rep.* 11:7388. doi: 10.1038/s41598-021-86970-0

Wu, Z., Liu, H., Zhan, W., Yu, Z., Qin, E., Liu, S., et al. (2021). The chromosome-scale reference genome of safflower (*Carthamus tinctorius*) provides insights into linoleic acid and flavonoid biosynthesis. *Plant Biotechnol. J.* 19, 1725–1742. doi: 10.1111/pbi.13586

Wu, Z., Zhao, X., Miyamoto, A., Zhao, S., Liu, C., Zheng, W., et al. (2019). Effects of steroidal saponins extract from *Ophiopogon japonicus* root ameliorates doxorubicin-induced chronic heart failure by inhibiting oxidative stress and inflammatory response. *Pharm. Biol.* 57, 176–183. doi: 10.1080/13880209.2019.1577467

Xia, M., Han, X., He, H., Yu, R., Zhen, G., Jia, X., et al. (2018). Improved de novo genome assembly and analysis of the Chinese cucurbit *Siraitia grosvenorii*, also known as monk fruit or luo-han-guo. *Gigascience* 7:giy067. doi: 10.1093/gigascience/giy067

Xiang, Y., Guo, Z., Zhu, P., Chen, J., and Huang, Y. (2019). Traditional Chinese medicine as a cancer treatment: modern perspectives of ancient but advanced science. *Cancer Med.* 8, 1958–1975. doi: 10.1002/cam4.2108

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* 7, 1026–1036. doi: 10.1038/s41477-021-00963-5

Xu, J., Chu, Y., Liao, B., Xiao, S., Yin, Q., Bai, R., et al. (2017). *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience* 6, 1–15. doi: 10.1093/gigascience/gix093

Xu, X., Crow, M., Rice, B. R., Li, F., Harris, B., Liu, L., et al. (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* 56, 557.e6–568.e6. doi: 10.1016/j.devcel.2020.12.015

Xu, D., Lin, H., Tang, Y., Huang, L., Xu, J., Nian, S., et al. (2021). Integration of full-length transcriptomics and targeted metabolomics to identify benzylisoquinoline alkaloid biosynthetic genes in *Corydalis yanhusuo*. *Hortic. Res.* 8:16. doi: 10.1038/s41438-020-00450-6

Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., 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

Xue, T., Chen, D., Zhang, T., Chen, Y., Fan, H., Huang, Y., et al. (2022). Chromosome-scale assembly and population diversity analyses provide insights into the evolution of *Sapindus mukorossi*. *Hortic. Res.* 9:uhac012. doi: 10.1093/hr/uhac012

Xue, L., He, Z., Bi, X., Xu, W., Wei, T., Wu, S., et al. (2019). Transcriptomic profiling reveals MEP pathway contributing to ginsenoside biosynthesis in *Panax ginseng*. *BMC Genomics* 20:383. doi: 10.1186/s12864-019-5718-x

Yang, L., Chen, J., Lu, H., Lai, J., He, Y., Liu, S., et al. (2019). *Pueraria lobata* for diabetes mellitus: past, present and future. *Am. J. Chin. Med.* 47, 1419–1444. doi: 10.1142/S0192415X19500733

Yang, Y., Huang, L., Xu, C., Qi, L., Wu, Z., Li, J., et al. (2021a). Chromosome-scale genome assembly of areca palm (*Areca catechu*). *Mol. Ecol. Resour.* 21, 2504–2519. doi: 10.1111/1755-0998.13446

Yang, Y., Li, S., Xing, Y., Zhang, Z., Liu, T., Ao, W., et al. (2021b). The first high-quality chromosomal genome assembly of a medicinal and edible plant *Arctium lappa*. *Mol. Ecol. Resour.* 22, 1493–1507. doi: 10.1111/1755-0998.13547

Yang, Z., Liu, G., Zhang, G., Yan, J., Dong, Y., Lu, Y., et al. (2021). The chromosome-scale high-quality genome assembly of *Panax notoginseng* provides insight into dencichine biosynthesis. *Plant Biotechnol. J.* 19, 869–871. doi: 10.1111/pbi.13558

Yao, S., Lan, Z., Huang, R., Tan, Y., Huang, D., Gu, J., et al. (2021). Hormonal and transcriptional analyses provides new insights into the molecular mechanisms underlying root thickening and isoflavonoid biosynthesis in *Callerya speciosa* (champ. Ex Benth.) Schot. *Sci. Rep.* 11:9. doi: 10.1038/s41598-020-76633-x

Yin, Z., Wang, X., Yang, X., Chen, Y., Duan, Y., and Han, J. (2021). *Salvia miltiorrhiza* in anti-diabetic angiopathy. *Curr. Mol. Pharmacol.* 14, 960–974. doi: 10.2174/1874467214999210111222918

Yu, H., Liu, M., Yin, M., Shan, T., Peng, H., Wang, J., et al. (2021). Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandiflorus*. *Planta* 254:34. doi: 10.1007/s00425-021-03677-2

Yu, Z. W., Zhang, N., Jiang, C. Y., Wu, S. X., Feng, X. Y., and Feng, X. Y. (2021). Exploring the genes involved in biosynthesis of dihydroquercetin and dihydromyricetin in Ampelopsis grossedentata. *Sci. Rep.* 11:15596. doi: 10.1038/s41598-021-95071-x

Yue, G. G., Li, L., Lee, J. K., Kwok, H. F., Wong, E. C., Li, M., et al. (2019). Multiple modulatory activities of *Andrographis paniculata* on immune responses and xenograft growth in esophageal cancer preclinical models. *Phytomedicine* 60:152886. doi: 10.1016/j.phymed.2019.152886

Yue, J., Wang, R., Ma, X., Liu, J., Lu, X., Balaso Thakar, S., et al. (2020). Full-length transcriptome sequencing provides insights into the evolution of apocarotenoid biosynthesis in *Crocus sativus*. *Comput. Struct. Biotechnol. J.* 18, 774–783. doi: 10.1016/j.csbj.2020.03.022

Zhang, C., Deng, W., Yan, W., and Li, T. (2018). Whole genome sequence of an edible and potential medicinal fungus, *Cordyceps guangdongensis*. *G3* 8, 1863–1870. doi: 10.1534/g3.118.200287

Zhang, D., Li, W., Chen, Z. J., Wei, F. G., Liu, Y. L., and Gao, L. Z. (2020). SMRT- and Illumina-based RNA-seq analyses unveil the ginsinoside biosynthesis and transcriptomic complexity in *Panax notoginseng*. *Sci. Rep.* 10:15310. doi: 10.1038/s41598-020-72291-1

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

Zhang, D., Li, W., Xia, E. H., Zhang, Q. J., Liu, Y., Zhang, Y., et al. (2017). The medicinal herb Panax notoginseng genome provides insights into Ginsenoside biosynthesis and genome evolution. *Mol. Plant* 10, 903–907. doi: 10.1016/j.molp.2017.02.011

Zhang, X., Zhao, Y., Guo, L., Qiu, Z., Huang, L., and Qu, X. (2017). Differences in chemical constituents of *Artemisia annua* L from different geographical regions in China. *PLoS One* 12:e0183047. doi: 10.1371/journal.pone.0183047

Zhao, Q., Yang, J., Cui, M. Y., Liu, J., Fang, Y., Yan, M., et al. (2019). The reference genome sequence of *Scutellaria baicalensis* provides insights into the evolution of Wogonin biosynthesis. *Mol. Plant* 12, 935–950. doi: 10.1016/j.molp.2019.04.002

Zheng, W., Wu, J., Gu, J., Weng, H., Wang, J., Wang, T., et al. (2020). Modular characteristics and mechanism of action of herbs for endometriosis treatment in Chinese medicine: a data mining and network pharmacology-based identification. *Front. Pharmacol.* 11:147. doi: 10.3389/fphar.2020.00147

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

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

frontiers | Frontiers in Plant Science

TYPE Mini Review
PUBLISHED 10 August 2022
DOI 10.3389/fpls.2022.954083

OPEN ACCESS

EDITED BY
Laigeng Li,
Center for Excellence in Molecular Plant Sciences (CAS), China

REVIEWED BY
Jean-Etienne Bassard,
UPR2357 Institut de Biologie Moléculaire des Plantes (IBMP), France

*CORRESPONDENCE
Michael Gutensohn
michael.gutensohn@mail.wvu.edu

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

RECEIVED 26 May 2022
ACCEPTED 01 July 2022
PUBLISHED 10 August 2022

CITATION
Gutensohn M, Hartzell E and Dudareva N (2022) Another level of complex-ity: The role of metabolic channeling and metabolons in plant terpenoid metabolism.
Front. Plant Sci. 13:954083.
doi: 10.3389/fpls.2022.954083

Another level of complex-ity: The role of metabolic channeling and metabolons in plant terpenoid metabolism

Michael Gutensohn[1]*, Erin Hartzell[1] and Natalia Dudareva[2,3,4]

[1]Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV, United States, [2]Department of Biochemistry, Purdue University, West Lafayette, IN, United States, [3]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, United States, [4]Purdue Center for Plant Biology, Purdue University, West Lafayette, IN, United States

Terpenoids constitute one of the largest and most diverse classes of plant metabolites. While some terpenoids are involved in essential plant processes such as photosynthesis, respiration, growth, and development, others are specialized metabolites playing roles in the interaction of plants with their biotic and abiotic environment. Due to the distinct functions and properties of specific terpenoid compounds, there is a growing interest to introduce or modify their production in plants by metabolic engineering for agricultural, pharmaceutical, or industrial applications. The MVA and MEP pathways and the prenyltransferases providing the general precursors for terpenoid formation, as well as the enzymes of the various downstream metabolic pathways leading to the formation of different groups of terpenoid compounds have been characterized in detail in plants. In contrast, the molecular mechanisms directing the metabolic flux of precursors specifically toward one of several potentially competing terpenoid biosynthetic pathways are still not well understood. The formation of metabolons, multi-protein complexes composed of enzymes catalyzing sequential reactions of a metabolic pathway, provides a promising concept to explain the metabolic channeling that appears to occur in the complex terpenoid biosynthetic network of plants. Here we provide an overview about examples of potential metabolons involved in plant terpenoid metabolism that have been recently characterized and the first attempts to utilize metabolic channeling in terpenoid metabolic engineering. In addition, we discuss the gaps in our current knowledge and in consequence the need for future basic and applied research.

KEYWORDS

terpenoids, metabolic channeling, metabolons, prenyltransferases, membrane complexes, chlorophyll, sterols, metabolic engineering

Introduction

Terpenoids represent a prominent class of metabolites present in all living organisms. An exceptionally large and structurally diverse number of terpenoids is produced by plants, which employ multiple biosynthetic pathways frequently acting in parallel to create a plethora of these metabolites (Chen et al., 2011; Pichersky and Raguso, 2018). Some terpenoids found in all plants serve essential roles and include pigments (chlorophylls, carotenoids), electron carriers (plastoquinone, ubiquinone), membrane components (sterols), and hormones (gibberellins, abscisic acid, steroids, strigolactones). In contrast, the majority of plant terpenoids, in particular mono-, sesqui- and di-terpenes, are species-specific specialized metabolites that are involved in antagonistic and beneficial interactions with the environment. Due to their roles in defense against pests and pathogens, attraction of beneficial organisms, and contribution to the aromas of fruits and other edible parts of plants, these terpenoids are valuable agronomic traits. In addition, terpenoids are widely used by humans as flavors, fragrances, preservatives, pharmaceuticals and biofuels (Ajikumar et al., 2008; Immethun et al., 2013; Tippmann et al., 2013). Despite the tremendous diversity of terpenoids, they are all derived from the same precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In plants these isomers are synthesized by two alternative pathways (**Figure 1**), the mevalonic acid (MVA) pathway operating in the cytosol and peroxisomes, and the methylerythritol phosphate (MEP) pathway in plastids (Ashour et al., 2010; Hemmerlin et al., 2012). Although these pathways act independently, there is an IPP and DMAPP exchange between them *via* yet unidentified transporters in the plastid envelope membranes (Soler et al., 1993; Bick and Lange, 2003; Flügge and Gao, 2005). IPP and DMAPP are subsequently utilized by short-chain prenyltransferases that join the isoprene units in the *trans* configuration to form larger prenyl diphosphate intermediates, which ultimately serve as precursors for the downstream terpenoid biosynthesis (**Figure 1**). Farnesyl diphosphate synthase (FPPS) forms *trans*-farnesyl diphosphate (*E*,*E*-FPP) in the cytosol, while geranyl diphosphate synthase (GPPS) and geranylgeranyl diphosphate synthase (GGPPS) are responsible for respective geranyl diphosphate (GPP) and geranylgeranyl diphosphate (GGPP) formation primarily in plastids (Gutensohn et al., 2013). These prenyl diphosphates are then utilized by a large family of terpene synthases (TPSs) to produce the variety of mono-, sesqui- and di-terpenes in plants (Chen et al., 2011; Karunanithi and Zerbe, 2019). In parallel *E*,*E*-FPP and GGPP are used for head-to-head condensations by squalene synthase (SQS) and phytoene synthase (PSY) to form the backbones of tri- and tetra-terpenes such as sterols and carotenoids, respectively. In plastids GGPP also serves as precursor for the formation of gibberellins and the side chains of chlorophylls, tocopherols and plastoquinones (Ruiz-Sola et al., 2016). An additional class of species-specific *cis*-prenyltransferases has been identified (Oh et al., 2000; Asawatreratanakul et al., 2003; Akhtar et al., 2013), which link isoprene units in the *cis* configuration to synthesize neryl diphosphate (NPP) and *cis*-farnesyl diphosphate (*Z*,*Z*-FPP), the precursors for mono- and sesquiterpenes and some long-chain terpenoids, like rubber and dolichols.

The diversity of terpenoid metabolites in plants and the presence of multiple biosynthetic pathways competing for the same prenyl diphosphate substrates suggest that molecular mechanisms must exist to sufficiently direct the metabolic flux toward the formation of specific terpenoid compounds. Recent studies also indicated that some *cis*-prenyl diphosphates could inhibit other terpenoid biosynthetic enzymes (Gutensohn et al., 2014), further highlighting the need for containment and channeling of certain intermediates in the plant terpenoid biosynthetic network. Due to their hydrophobicity, terpenoid metabolites could partition to any cellular membrane, therefore biosynthetic pathways should also involve mechanisms that target respective terpenoid compounds toward specific membranes in the plant cell.

The concept of metabolons, supramolecular complexes of sequential metabolic enzymes and cellular structural elements that mediate substrate channeling, was proposed already more than three decades ago (Srere, 1985). While these multi-protein complexes can be either transient and dynamic, or more stable associations of enzymes (Winkel, 2004; Zhang and Fernie, 2021), a clear picture has emerged about their potential functions (Obata, 2020). They (i) sequester pathway intermediates thus increasing their local concentrations and subsequently enhancing reaction rates, (ii) prevent the release of pathway intermediates thus precluding their potential degradation or harmful reactions with other cellular components, (iii) protect from the access of compounds acting as enzyme inhibitors, (iv) restrict the consumption of intermediates by competing enzymes thus determining the direction of metabolic fluxes, and (v) frequently interact with membranes or cytoskeleton elements thus controlling the localization of enzymes. An increasing number of metabolons has recently been characterized that are involved in plant central and specialized metabolism (summarized in Winkel, 2004; Obata, 2019; Zhang and Fernie, 2021). Here, we summarize the current knowledge and discuss future challenges in studying the role of metabolic channeling and potential metabolons in plant terpenoid metabolism as well as respective applications in the metabolic engineering of terpenoid production.

GGPPS containing complexes in plastids

In metabolic networks the same precursor is often shared by several biosynthetic pathways localized in the same subcellular compartment, which requires the regulation of its allocation toward each pathway. In the plant

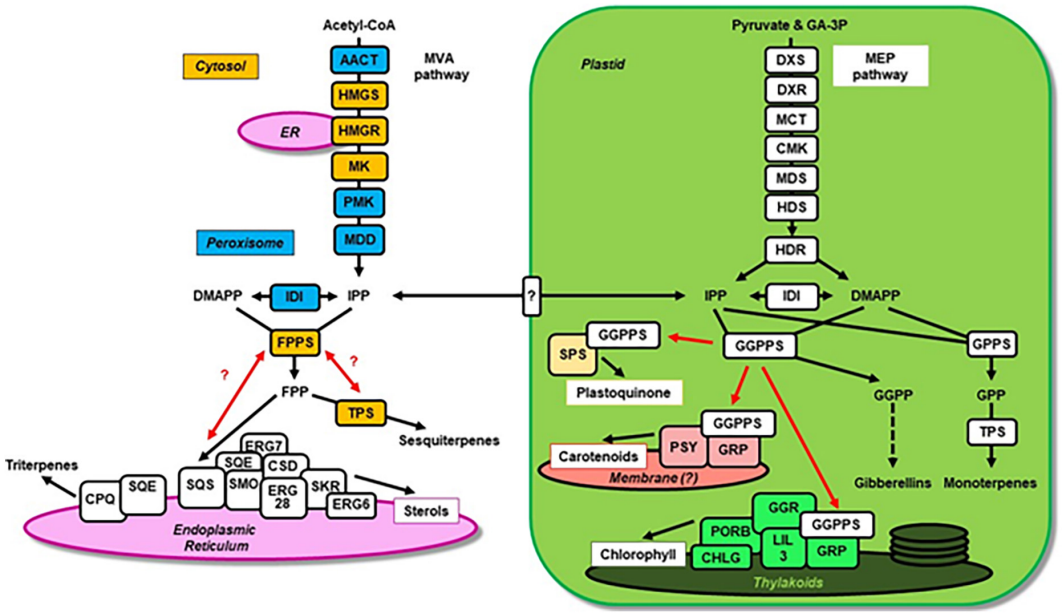


FIGURE 1
The plant terpenoid metabolic network and potential metabolons involved in the biosynthesis of specific compounds. The plastid and endoplasmic reticulum are highlighted in green and purple, respectively. The MVA pathway enzymes localized in peroxisomes and the cytosol are labeled in blue and orange, respectively. Individual enzymes are depicted as boxes and black arrows indicate metabolic fluxes. Confirmed and putative (with question mark) interactions of prenyltransferases with downstream enzymes forming metabolons are indicated by red arrows. The unknown transporter involved in IPP and DMAPP exchange between cytosol and plastid is shown in the plastid envelope membrane. Abbreviations: AACT, aceto-acetyl-CoA thiolase; CHLG, chlorophyll synthase; CMK, 4-(cytidine 5-diphospho)-2-C-methyl-D-erythritol kinase; CPQ, cucurbitadienol synthase; CSD, sterol C-3 dehydrogenase/C-4 decarboxylase; DMAPP, dimethylallyl diphosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; FPPS, farnesyl diphosphate synthase; GA-3P, D-glyceraldehyde 3-phosphate; GGPPS, geranylgeranyl diphosphate synthase; GGR, geranylgeranyl reductase; GPPS, geranyl diphosphate synthase; GRP, GGPPS recruiting protein; HDR, (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HDS, (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; IDI, isopentenyl diphosphate isomerase; IPP, isopentenyl diphosphate; LIL3; light-harvesting-like protein 3; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; MDD, mevalonate diphosphate decarboxylase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; PORB, protochlorophyllide oxidoreductase; PSY, phytoene synthase; SKR, sterol C-3 keto-reductase; SMO, sterol C-4 methyl oxidase; SPS, solanesyl diphosphate synthase; SQE, squalene epoxidase; SQS, squalene synthase; TPS, terpene synthases (including mono- and sesquiterpene synthases).

terpenoid metabolic network, prenyltransferases catalyze key reactions controlling carbon flux toward the biosynthesis of distinct classes of terpenoid compounds. In plastids, for example, GGPP biosynthesis represents such an essential branch-point as this precursor is shared by several vital pathways including the formation of gibberellins, carotenoids, tocopherols, phylloquinons, plastoquinones, and chlorophylls (**Figure 1**). Indeed, the Arabidopsis knock-down mutant of a plastidic GGPPS (*ggpps11*), a hub isozyme required for the production of photosynthesis-related metabolites, contains reduced levels of chlorophylls, carotenoids, tocopherols, phylloquinons, and plastoquinones (Ruiz-Sola et al., 2016). It was found that GGPPS11 interacts with three GGPP-utilizing enzymes: geranylgeranyl reductase (GGR) involved in the formation of phytyl chains for chlorophyll, tocopherol and phylloquinone; PSY involved in carotenoid formation; and solanesyl diphosphate synthase (SPS) involved in plastoquinone formation (Ruiz-Sola et al., 2016). These results suggested that multi-enzyme complexes containing GGPPS and GGPP-consuming enzymes could present a mechanism for channeling of metabolic flux toward specific downstream biosynthetic pathways. Characterization of GGPPS in rice chloroplasts provided further details about the molecular mechanisms regulating the allocation of prenyltransferases to such complexes (Zhou et al., 2017). Rice contains only one functionally active GGPPS, OsGGPPS1, which exists as a homodimer in the chloroplast stroma. However, a thylakoid localized GGPPS recruiting protein (GRP) forms a heterodimer with OsGGPPS1 and recruits the latter toward thylakoid membranes. The catalytically inactive GRP not only enhances the catalytic efficiency of OsGGPPS1, but also directs it to a multi-enzyme complex consisting of GGR, light-harvesting-like protein 3 (LIL3), protochlorophyllide oxidoreductase (PORB) and chlorophyll synthase (CHLG). These results suggest that the interaction of GRP with OsGGPPS1 provides a mechanism for recruiting GGPPS toward a thylakoid localized metabolon thus increasing metabolic flux toward chlorophyll biosynthesis and reducing flux toward competing terpenoid pathways.

Remarkably, a similar interaction between GGPPS and a GRP homolog has also been demonstrated in ripening fruits of red pepper (*Capsicum annuum*) that accumulate carotenoids (Wang et al., 2018). In *C. annuum* both, GGPPS and GRP, interact with PSY, which previously has been found in a protein complex involved in carotenoid formation (Dogbo and Camara, 1987). Moreover, phytoene desaturase (PDS), the enzyme acting downstream of PSY in the carotenoid biosynthesis, was likewise found to be present in a large protein complex in plastid membranes (Lopez et al., 2008). Thus, it appears that in plastids GGPPS can also be recruited into a membrane-bound protein complex that contains PSY, PDS and potentially other enzymes, thereby directing the metabolic flux toward carotenoid formation.

Complexes involved in sterol and triterpenoid biosynthesis

Similar to the situation in plastids where multiple biosynthetic pathways compete for GGPP, FPPS resides at a branch-point of the plant terpenoid metabolic network in the cytosol (**Figure 1**). The synthesized *E*,*E*-FPP is utilized by competing biosynthetic pathways leading to the formation of sesquiterpenes, sterols, brassinosteroids, and triterpenes. Earlier studies showed that the exposure of tobacco and potato to pathogens or elicitors increased the biosynthesis of sesquiterpenes while concomitantly decreasing sterol formation, indicating an altered allocation of *E*,*E*-FPP between these competing pathways (Threfall and Whitehead, 1988; Vögeli and Chappell, 1988; Zook and Kuc, 1991). Squalene synthase (SQS) catalyzing the first committed step of sterol biosynthesis contains an N-terminal catalytic domain and a C-terminal domain tethering the enzyme to the endoplasmic reticulum membrane. A 26-amino acid hinge region linking these two domains was found to be unique to SQSs in different kingdoms of life (e.g., plants and fungi). Expression of Arabidopsis SQS1 in the yeast Δ*erg9* background could only partially complement this SQS deletion mutant, while a chimeric *Arabidopsis* SQS carrying the C-terminus of the yeast enzyme fully restored the fungal sterol formation (Kribii et al., 1997; Linscott et al., 2016). Remarkably, Δ*erg9* yeast lines expressing the plant SQS produced presqualene diphosphate, a toxic intermediate normally not released from the enzyme, suggesting that the SQS hinge region is involved in the interaction with downstream sterol biosynthetic enzyme(s) and crucial for metabolic channeling in this pathway. Moreover, exogenously supplied squalene was a poor substrate for the second pathway enzyme squalene epoxidase (SQE) when analyzed in a yeast microsomal fraction, while *E*,*E*-FPP was efficiently converted to the SQE product 2,3-oxidosqualene (M'Baya and Karst, 1987), supporting the existence of SQS and SQE proximity that allows metabolite channeling.

A complex network of protein–protein interactions among sterol biosynthetic enzymes was also discovered in yeast using yeast two-hybrid analysis (Mo and Bard, 2005) further confirming the involvement of a multi-enzyme complex in this pathway. The core of this complex is found in fungi, humans and plants, and appeared to be formed by three enzymes involved in the sterol C-4 demethylation, namely sterol C-4 methyl oxidase (SMO), sterol C-3 dehydrogenase/C-4 decarboxylase (CSD) and sterol C-3 keto-reductase (SKR), *via* their interaction with the non-catalytic protein Erg28 (Gachotte et al., 2001; Mialoundama et al., 2013). This core complex likely acts as a hub with which enzymes acting earlier in the pathway, such as oxidosqualene cyclase (Erg7p) (Taramino et al., 2010), as well as later pathway enzymes such as Erg6p can interact (Mo and Bard, 2005). Such an organization also allows the variability in the sterol biosynthesis found in different organisms that results in sterol compounds with diverse positions of double bonds and side-chain alkyl groups introduced in the downstream enzymatic steps (Mo and Bard, 2005). The role of the core complex in metabolic channeling in the sterol biosynthesis in plants was highlighted by the fact that RNAi and mutant lines for *Arabidopsis thaliana ERG28* accumulated 4-carboxy-4-methyl-24-methylenecycloartanol (CMMC), a sterol biosynthetic intermediate (Mialoundama et al., 2013). CMMC is channeled within the complex to produce membrane sterols and brassinosteroids under normal conditions, but is released once the complex is deregulated and inhibits polar auxin transport (Mialoundama et al., 2013).

cis-Prenyltransferase complex involved in rubber formation

While over 2500 plant species produce natural rubber, a *cis*-1,4-polyisoprene, only a few have been used to study its biosynthesis in detail (summarized in Cherian et al., 2019) including the para rubber tree, guayule, and rubber dandelion. The basic biochemical mechanism of rubber formation appears to be relatively conserved among these species. It involves *cis*-prenyltransferases (CPTs), also called rubber transferases, that catalyze the sequential *cis*-1,4-condensation of IPP, initially onto *trans*-short chain prenyl diphosphates serving as priming substrates, and then to the α-terminus of polyprenyl pyrophosphates and the partially polymerized rubber molecules. Rubber biosynthesis takes place at the surface of rubber particles that are surrounded by a lipid monolayer membrane with integrated and associated proteins, and contain a hydrophobic core of rubber polymers. While some of the proteins in rubber particles only have structural functions, CPTs form a multi-protein complex, which is responsible for rubber biosynthesis. This rubber transferase complex contains two small substrate-binding proteins (1.6–1.8 and 3.6–3.9 kDa) that bind IPP and FPP (Cornish et al., 2008) and likely initiate the biosynthetic

process. The CPTs were found to directly interact with a non-catalytic CPT-binding protein (CBP), which links them to the rubber particle to ensure efficient rubber biosynthesis (Yamashita et al., 2016; Lakusta et al., 2019; Niephaus et al., 2019). Rubber elongation factor (REF) is another protein interacting with CPT (Yamashita et al., 2016), however, its role in rubber biosynthesis is not fully resolved. A small rubber particle protein (SRPP) interacts with REF, thus also representing a subunit of this multi-protein complex, and appears to play a role in regulating the molecular weight of the rubber polymer (Collins-Silva et al., 2012). Overall, the rubber transferase complex represents a metabolon that is involved in substrate binding, catalysis, molecular weight regulation, and particularly in channeling of the hydrophobic polyisoprene product into the interior of the rubber particle.

Metabolic channeling in the engineering of terpenoid formation

The above examples suggest that metabolons are an integral part of terpenoid metabolism in plants that reduce metabolic competition while increasing the flux toward the formation of distinct terpenoid products, prevent the release of potentially harmful pathway intermediates, and direct terpenoid products toward specific membranes and cellular compartments. Thus, it is not surprising that this emerging knowledge was already applied in the metabolic engineering of terpenoid production. Three different strategies have been explored so far to achieve metabolic channeling in engineered terpenoid biosynthetic pathways (**Figure 2**). The simplest approach to enhance the metabolic flux is to utilize fusions of metabolic enzymes catalyzing successive reactions. To engineer increased carotenoid formation in plants the co-expression of GGPPS and PSY as individual enzymes was compared to the expression of a fusion construct with GGPPS linked to the C-terminus of PSY (Camagna et al., 2019). While the individual enzymes converted only 60% of the IPP substrate to phytoene suggesting significant GGPP leakage, the overexpression of the PSY-GGPPS construct in Arabidopsis (**Figure 2A**) resulted in almost quantitative conversion of IPP implying efficient metabolite channeling within the enzyme fusion. Likewise, fusion constructs containing FPPS linked *via* a short Gly-Ser-Gly peptide linker to sesquiterpene synthases (**Figure 2A**), *epi*-aristolochene synthase (EAS) or amorpha-4, 11-diene synthase (ADS), were designed and tested *in vitro* and *in planta* (Brodelius et al., 2002; Han et al., 2016). The fusion of enzymes did not affect their folding and affinity for substrates as the K_m values were the same for the single FPPS and EAS, as well as EAS-FPPS linked enzymes (Brodelius et al., 2002), but the amount of produced *epi*-aristolochene was significantly

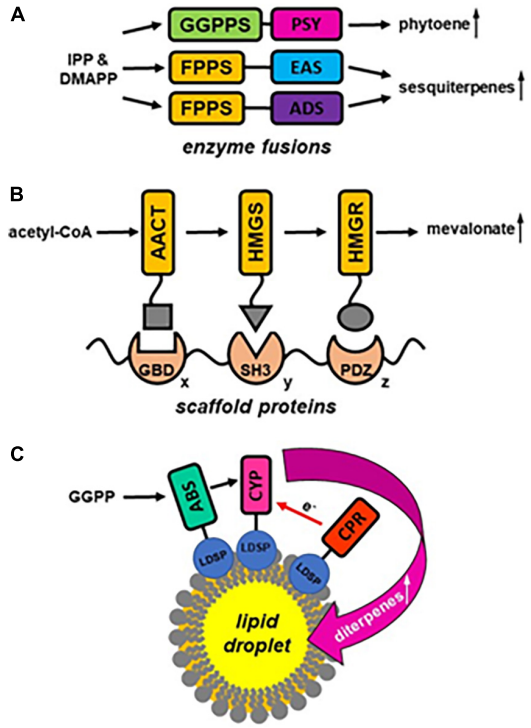


FIGURE 2
Engineering approaches utilizing metabolic channeling for improved terpenoid formation. **(A)** Enzyme fusion strategy: fusion of prenyltransferases (GGPPS, FPPS) with phytoene synthase (PSY) and terpene synthases (*epi*-aristolochene synthase EAS, amorpha-4,11-diene synthase ADS), respectively, leads to increased substrate (IPP, DMAPP) conversion to product formation. **(B)** Synthetic scaffold protein strategy: scaffold proteins carry multiple protein-protein interaction domains (GBD, SH3, PDZ) in various ratios (x: y: z) to organize multiple sequential pathway enzymes (e.g., AACT, HMGS, HMGR) tagged with specific small peptide ligands (indicated by a square, triangle and circle) to increase metabolic flux through a pathway. **(C)** Lipid droplet strategy: engineered cytosolic lipid droplets serve as scaffold to co-localize terpenoid biosynthetic enzymes, e.g., diterpene synthase abietadiene synthase (ABS), cytochrome P450 (CYP), and cytochrome P450 reductase (CPR), that are expressed as fusion constructs with a lipid droplet surface protein (LDSP). It will result in increased production of terpenoid compounds (e.g., diterpenes) and additionally in their sequestration in the engineered lipid droplets.

higher in case of fused proteins (Brodelius et al., 2002). Similarly, plants expressing the FPPS-ADS fusion construct displayed a higher flux toward artemisinin formation (Han et al., 2016) suggesting that the close proximity of the fused enzymes reduced the diffusion of FPP substrate from the engineered metabolons. Although enzyme fusions can result in substrate channeling, this approach suffers from several restrictions including (i) the potential negative effects on protein folding, (ii) the limitation in the number of enzymes that can be linked, and (iii) the fact that only a fixed 1:1 ratio of enzymes can be achieved due to the fusion. An alternative metabolic engineering approach addresses these issues by utilizing synthetic scaffold proteins to

organize multiple enzymes into complexes. For this purpose, scaffold proteins carrying multiple protein-protein interaction domains are used to co-localize multiple sequential pathway enzymes (**Figure 2B**), each tagged with a specific small peptide ligand that will bind to the domains on the scaffold and is less likely to interfere with protein folding (Lee et al., 2012). The modular architecture of the scaffold not only allows combining more than two enzymes, but also to control the relative ratio of enzymes within the metabolon by varying the number of interaction domains in the scaffold. The feasibility of this approach was tested for three sequential MVA pathway enzymes (**Figure 1**), acetoacetyl-CoA transferase (AACT), hydroxy-methylglutaryl-CoA synthase (HMGS) and hydroxy-methylglutaryl-CoA reductase (HMGR) (Dueber et al., 2009). Upon optimizing the stoichiometry, this multi-enzyme complex organized by a scaffold resulted in a 77-fold improvement in mevalonate formation and reduced the accumulation of intermediates. A third approach to create a scaffold for multiple terpenoid biosynthetic enzymes involves the engineering of ER-derived cytosolic lipid droplets in plants (Sadre et al., 2019). When the diterpene synthase abietadiene synthase (ABS), cytochrome P450 (CYP), and cytochrome P450 reductase (CPR) were expressed as fusion constructs with a lipid droplet surface protein (LDSP) in *Nicotiana benthamiana* leaves with engineered lipid droplets these proteins all co-localized at the surface of the artificial organelles (**Figure 2C**). Co-expression of cytosolic HMGR and GGPPS in these leaves resulted not only in the efficient formation of diterpenoids and diterpenoid acids, but also in the sequestration of these products in the engineered lipid droplets which might limit negative feedback on the pathway.

Conclusion

Metabolons by definition are transient multi-protein complexes of sequential enzymes that mediate substrate channeling (Srere, 1985; Zhang and Fernie, 2021). To date, the presence of many metabolons has been proposed in plants based on observed protein-protein interactions, however, only in a few cases channeling of metabolites has been proven yet (Achnine et al., 2004; Graham et al., 2007; Laursen et al., 2016; Zhang et al., 2017). Here we focused on potential metabolons in the plant terpenoid metabolism, for which just protein interactions and complexes have been demonstrated so far, and only indirect evidence for substrate channeling exists from genetic studies. The formation of metabolons and their dynamic nature could provide an additional level of regulation in the complex and highly branched plant terpenoid metabolic network. Such type of regulation allows to rapidly redirect metabolic fluxes depending on changing demands during a plant's development or in response to its abiotic and biotic environment. However, to understand the role of metabolons in terpenoid biosynthetic pathways in particular a further detailed analysis of substrate channeling is required and could be achieved by (i) isotope dilution and enrichment studies, (ii) transient time analysis, (iii) evaluating resistance to competing reactions and inhibitors, and (iv) enzyme-buffering analysis (Fernie et al., 2018; Sweetlove and Fernie, 2018). Moreover, future investigation of interactions of terpenoid biosynthetic enzymes at different developmental stages and under changing environmental conditions will uncover potential additional players of novel or currently proposed metabolons. While GGPPS was found to form complexes with various downstream terpenoid biosynthetic enzymes as summarized above, no interactions, for example, were identified yet between FPPS and enzymes of the sterol biosynthesis pathway or cytosolic terpene synthases, despite indications for changing metabolic fluxes between these competing pathways (Threfall and Whitehead, 1988; Vögeli and Chappell, 1988; Zook and Kuc, 1991).

Author contributions

MG conceived the concept for this manuscript. All authors wrote the manuscript and have read and approved the final manuscript.

Funding

The work in the Gutensohn lab was supported by the USDA National Institute of Food and Agriculture - Agricultural and Food Research Initiative competitive grants: 2018-67014-28092 and 2020-67014-30901, the West Virginia Agricultural and Forestry Experiment Station Hatch Project WVA00730, and the Ray Marsh and Arthur Pingree Dye Professorship. The work in the Dudareva lab was supported by a grant from the National Science Foundation IOS-1655438 and by the USDA National Institute of Food and Agriculture Hatch Project number: 177845.

Conflict of interest

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

Publisher's note

References

Achnine, L., Blancaflor, E. B., Rasmussen, S., and Dixon, R. A. (2004). Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 16, 3098–3109. doi: 10.1105/tpc.104.024406

Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H., and Stephanopoulos, G. (2008). Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* 5, 167–190. doi: 10.1021/mp700151b

Akhtar, T. A., Matsuba, Y., Schauvinhold, I., Yu, G., Lees, H. A., Klein, S. E., et al. (2013). The tomato cis-prenyltransferase gene family. *Plant J.* 73, 640–652. doi: 10.1111/tpj.12063

Asawatreratanakul, K., Zhang, Y. W., Wititsuwannakul, D., Wititsuwannakul, R., Takahashi, S., Rattanapittayaporn, A., et al. (2003). Molecular cloning, expression and characterization of cDNA encoding cis-prenyltransferases from *Hevea brasiliensis*. *Eur. J. Biochem.* 270, 4671–4680. doi: 10.1046/j.1432-1033.2003.03863.x

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

Bick, J. A., and Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* 415, 146–154. doi: 10.1016/S0003-9861(03)00233-9

Brodelius, M., Lundgren, A., Mercke, P., and Brodelius, P. E. (2002). Fusion of farnesyldiphosphate synthase and epi-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis in *Nicotiana tabacum*. *Eur. J. Biochem.* 269, 3570–3577. doi: 10.1046/j.1432-1033.2002.03044.x

Camagna, M., Grundmann, A., Bär, C., Koschmieder, J., Beyer, P., and Welsch, R. (2019). Enzyme fusion removes competition for geranylgeranyl diphosphate in carotenogenesis. *Plant Physiol.* 179, 1013–1027. doi: 10.1104/pp.18.01026

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

Cherian, S., Ryu, S. B., and Cornish, K. (2019). Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineering progress and prospects. *Plant Biotechnol. J.* 17, 2041–2061. doi: 10.1111/pbi.13181

Collins-Silva, J., Nural, A. T., Skaggs, A., Scott, D., Hathwaik, U., Woolsey, R., et al. (2012). Altered levels of the *Taraxacum kok-saghyz* (Russian dandelion) small rubber particle protein, TkSRPP3, result in qualitative and quantitative changes in rubber metabolism. *Phytochemistry* 79, 46–56. doi: 10.1016/j.phytochem.2012.04.015

Cornish, K., Scott, D. J., Xie, W., Mau, C. J. D., Zheng, Y. F., Liu, X., et al. (2008). Unusual subunits are directly involved in binding substrates for natural rubber biosynthesis in multiple plant species. *Phytochemistry* 156, 55–72. doi: 10.1016/j.phytochem.2018.08.014

Dogbo, O., and Camara, B. (1987). Purification of isopentenyl pyrophosphate isomerase and geranylgeranyl pyrophosphate synthase from *Capsicum* chromoplasts by affinity chromatography. *Biochim. Biophys. Acta* 920, 140–148. doi: 10.1016/0005-2760(87)90253-0

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759. doi: 10.1038/nbt.1557

Fernie, A. R., Zhang, Y., and Sweetlove, L. J. (2018). Passing the baton: substrate channelling in respiratory metabolism. *Research* 2018:1539325. doi: 10.1155/2018/1539325

Flügge, U. I., and Gao, W. (2005). Transport of isoprenoid intermediates across chloroplast envelope membranes. *Plant Biol.* 7, 91–97. doi: 10.1055/s-2004-830446

Gachotte, D., Eckstein, J., Barbuch, R., Hughes, T., Roberts, C., and Bard, M. (2001). A novel gene conserved from yeast to humans is involved in sterol biosynthesis. *J. Lipid Res.* 42, 150–154. doi: 10.1016/S0022-2275(20)32347-6

Graham, J. W., Williams, T. C., Morgan, M., Fernie, A. R., Ratcliffe, R. G., and Sweetlove, L. J. (2007). Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. *Plant Cell* 19, 3723–3738. doi: 10.1105/tpc.107.053371

Gutensohn, M., Nagegowda, D. A., and Dudareva, N. (2013). "Involvement of compartmentalization in monoterpene and sesquiterpene biosynthesis in plants," in *Isoprenoid Synthesis in Plants and Microorganisms*, eds T. J. Bach and M. Rohmer (New York, NY: Springer), 155–169. doi: 10.1007/978-1-4614-4063-5_11

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

Han, J., Wang, H., Kanagarajan, S., Hao, M., Lundgren, A., and Brodelius, P. E. (2016). Promoting artemisinin biosynthesis in *Artemisia annua* plants by substrate channeling. *Mol. Plant* 9, 946–948. doi: 10.1016/j.molp.2016.03.004

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

Immethun, C. M., Hoynes-O'Connor, A. G., Balassy, A., and Moon, T. S. (2013). Microbial production of isoprenoids enabled by synthetic biology. *Front. Microbiol.* 4:75. doi: 10.3389/fmicb.2013.00075

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

Kribii, R., Arro, M., Del Arco, A., Gonzalez, V., Balcells, L., Delourme, D., et al. (1997). Cloning and characterization of the *Arabidopsis thaliana* SQS1 gene encoding squalene synthase. *Eur. J. Biochem.* 249, 61–69. doi: 10.1111/j.1432-1033.1997.00061.x

Lakusta, A. M., Kwon, M., Kwon, E. J. G., Stonebloom, S., Scheller, H. V., and Ro, D. K. (2019). Molecular studies of the protein complexes involving cis-prenyltransferase in guayule (*Parthenium argentatum*), an alternative rubber-producing plant. *Front. Plant Sci.* 10:165. doi: 10.3389/fpls.2019.00165

Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H. J., et al. (2016). Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* 354, 890–893. doi: 10.1126/science.aag2347

Lee, H., DeLoache, W. C., and Dueber, J. E. (2012). Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14, 242–251. doi: 10.1016/j.ymben.2011.09.003

Linscott, K. B., Niehaus, T. D., Zhuang, X., Bell, S. A., and Chappell, J. (2016). Mapping a kingdom-specific functional domain of squalene synthase. *Biochim. Biophys. Acta* 1861, 1049–1057. doi: 10.1016/j.bbalip.2016.06.008

Lopez, A. B., Yang, Y., Thannhauser, T. W., and Li, L. (2008). Phytoene desaturase is present in a large protein complex in the plastid membrane. *Physiol. Plant.* 133, 190–198. doi: 10.1111/j.1399-3054.2008.01058.x

M'Baya, B., and Karst, F. (1987). *In vitro* assay of squalene epoxidase of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 147, 556–564. doi: 10.1016/0006-291X(87)90967-3

Mialoundama, A. S., Jadid, N., Brunel, J., Di Pascoli, T., Heintz, D., Erhardt, M., et al. (2013). *Arabidopsis* ERG28 tethers the sterol C4-demethylation complex to prevent accumulation of a biosynthetic intermediate that interferes with polar auxin transport. *Plant Cell* 25, 4879–4893. doi: 10.1105/tpc.113.115576

Mo, C., and Bard, M. (2005). A systematic study of yeast sterol biosynthetic protein-protein interactions using the split-ubiquitin system. *Biochim. Biophys. Acta* 1737, 152–160. doi: 10.1016/j.bbalip.2005.11.002

Niephaus, E., Müller, B., van Deenen, N., Lassowskat, I., Bonin, M., Finkemeier, I., et al. (2019). Uncovering mechanisms of rubber biosynthesis in *Taraxacum kok-saghyz*–role of cis-prenyltransferase-like 1 protein. *Plant J.* 100, 591–609. doi: 10.1111/tpj.14471

Obata, T. (2019). Metabolons in plant primary and secondary metabolism. *Phytochem. Rev.* 18, 1483–1507. doi: 10.1007/s11101-019-09619-x

Obata, T. (2020). Toward an evaluation of metabolite channeling *in vivo*. *Curr. Opin. Biotechnol.* 64, 55–61. doi: 10.1016/j.copbio.2019.09.013

Oh, S. K., Han, K. H., Ryu, S. B., and Kang, H. (2000). Molecular cloning, expression, and functional analysis of a cis-prenyltransferase from *Arabidopsis thaliana*. *J. Biol. Chem.* 275, 18482–18488. doi: 10.1074/jbc.M002000200

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

Ruiz-Sola, M. A., Coman, D., Beck, G., Barja, M. V., Colinas, M., Graf, A., et al. (2016). *Arabidopsis* geranylgeranyl diphosphate synthase 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. *New Phytol.* 209, 252–264. doi: 10.1111/nph.13580

Sadre, R., Kuo, P., Chen, J., Yang, Y., Banerjee, A., Benning, C., et al. (2019). Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves. *Nat. Commun.* 10:853. doi: 10.1038/s41467-019-08515-4

Soler, E., Clastre, M., Bantignies, B., Marigo, G., and Ambid, C. (1993). Uptake of isopentenyl diphosphate by plastids isolated from *Vitis vinifera* L. cell suspensions. *Planta* 191, 324–329. doi: 10.1007/BF00195689

Srere, P. A. (1985). The metabolon. *Trends Biochem. Sci.* 10, 109–110. doi: 10.1016/0968-0004(85)90266-X

Sweetlove, L. J., and Fernie, A. R. (2018). The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nat. Commun.* 9:2136. doi: 10.1038/s41467-018-04543-8

Taramino, S., Valachovic, M., Oliaro-Bosso, S., Viola, F., Teske, B., Bard, M., et al. (2010). Interactions of oxidosqualene cyclase (Erg7p) with 3-keto reductase (Erg27p) and other enzymes of sterol biosynthesis in yeast. *Biochim. Biophys. Acta* 1801, 156–162. doi: 10.1016/j.bbalip.2009.10.005

Threfall, D. R., and Whitehead, I. M. (1988). Coordinated inhibition of squalene synthetase and induction of enzymes of sesquiterpenoid phytoalexin biosynthesis in cultures of *Nicotiana tabacum*. *Phytochemistry* 27, 2567–2580. doi: 10.1016/0031-9422(88)87028-6

Tippmann, S., Chen, Y., Siewers, V., and Nielsen, J. (2013). From flavors and pharmaceuticals to advanced biofuels: production of isoprenoids in *Saccharomyces cerevisiae*. *Biotechnol. J.* 8, 1435–1444. doi: 10.1002/biot.201300028

Vögeli, U., and Chappell, J. (1988). Induction of sesquiterpene cyclase and suppression of squalene synthetase activities in plant cell cultures treated with fungal elicitor. *Plant Physiol.* 88, 1291–1296. doi: 10.1104/pp.88.4.1291

Wang, Q., Huang, X. Q., Cao, T. J., Zhuang, Z., Wang, R., and Lu, S. (2018). Heteromeric geranylgeranyl diphosphate synthase contributes to carotenoid biosynthesis in ripening fruits of red pepper (*Capsicum annuum* var. *conoides*). *J. Agric. Food Chem.* 66, 11691–11700. doi: 10.1021/acs.jafc.8b04052

Winkel, B. S. J. (2004). Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 55, 85–107. doi: 10.1146/annurev.arplant.55.031903.141714

Yamashita, S., Yamaguchi, H., Waki, T., Aoki, Y., Mizuno, M., Yanbe, F., et al. (2016). Identification and reconstitution of the rubber biosynthetic machinery on rubber particles from *Hevea brasiliensis*. *Elife* 5:e19022. doi: 10.7554/eLife.19022

Zhang, Y. J., Beard, K. F. M., Swart, C., Bergmann, S., Krahnert, I., Nikoloski, Z., et al. (2017). Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat. Commun.* 8:15212. doi: 10.1038/ncomms15212

Zhang, Y., and Fernie, A. R. (2021). Metabolons, enzyme-enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. *Plant Commun.* 2:100081. doi: 10.1016/j.xplc.2020.100081

Zhou, F., Wang, C. Y., Gutensohn, M., Jiang, L., Zhang, P., Zhang, D., et al. (2017). A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6866–6871. doi: 10.1073/pnas.1705689114

Zook, M. N., and Kuc, J. A. (1991). Induction of sesquiterpene cyclase and suppression of squalene synthetase activity in elicitor-treated or fungal-infected potato tuber tissue. *Physiol. Mol. Plant Pathol.* 39, 377–390. doi: 10.1016/0885-5765(91)90018-D

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 23 September 2022
DOI 10.3389/fpls.2022.984174

OPEN ACCESS

EDITED BY
Laigeng Li,
Center for Excellence in Molecular Plant Sciences (CAS), China

REVIEWED BY
John Browse,
Washington State University, United States
Barbara A. Halkier,
University of Copenhagen, Denmark

*CORRESPONDENCE
Chang-Jun Liu
cliu@bnl.gov

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

RECEIVED 01 July 2022
ACCEPTED 19 August 2022
PUBLISHED 23 September 2022

CITATION
Liu C-J (2022) Cytochrome b_5: A versatile electron carrier and regulator for plant metabolism. *Front. Plant Sci.* 13:984174. doi: 10.3389/fpls.2022.984174

Cytochrome b_5: A versatile electron carrier and regulator for plant metabolism

Chang-Jun Liu*

Biology Department, Brookhaven National Laboratory, Upton, NY, United States

Cytochrome b_5 (CB5) is a small heme-binding protein, known as an electron donor delivering reducing power to the terminal enzymes involved in oxidative reactions. In plants, the CB5 protein family is substantially expanded both in its isoform numbers and cellular functions, compared to its yeast and mammalian counterparts. As an electron carrier, plant CB5 proteins function not only in fatty acid desaturation, hydroxylation and elongation, but also in the formation of specialized metabolites such as flavonoids, phenolic esters, and heteropolymer lignin. Furthermore, plant CB5s are found to interact with different non-catalytic proteins such as ethylene signaling regulator, cell death inhibitor, and sugar transporters, implicating their versatile regulatory roles in coordinating different metabolic and cellular processes, presumably in respect to the cellular redox status and/or carbon availability. Compared to the plentiful studies on biochemistry and cellular functions of mammalian CB5 proteins, the cellular and metabolic roles of plant CB5 proteins have received far less attention. This article summarizes the fragmentary information pertaining to the discovery of plant CB5 proteins, and discusses the conventional and peculiar functions that plant CB5s might play in different metabolic and cellular processes. Gaining comprehensive insight into the biological functions of CB5 proteins could offer effective biotechnological solutions to tailor plant chemodiversity and cellular responses to environment stimuli.

KEYWORDS

cytochrome b_5, cytochrome P450, lignin, flavonoids, unsaturated fatty acid, very long chain fatty acid, sugar transporter, ethylene signaling

Introduction

As sessile organisms, terrestrial plants have evolved remarkable metabolic capacity elaborating abundant primary metabolites to sustain their growth, development and reproductivity, and a vast variety of functionally specialized metabolites, such as fragrances, pigments, anti-fungal or anti-herbivory phytoalexins, lipidic surface polymers, and cell wall structural components, to cope with ever changing environmental challenges (Weng and Chapple, 2010). The fortification of land plant

metabolic capacity is not only achieved by recruiting and evolving catalytic enzymes, but also by inheriting, co-evolving, and expanding many non-catalytic auxiliary proteins and cofactors in different metabolic processes.

Cytochrome b_5 (CB5) is a small heme-binding protein found in all life kingdom, including bacteria, fungi, mammals/human, and plants. Typically CB5 protein is a tail-anchored membrane protein and possesses a single transmembrane domain and a tail region near its C-terminus to regulate intracellular localization to the endoplasmic reticulum (ER) and/or the outer mitochondrial membranes (Schenkman and Jansson, 2003). The key characteristic of CB5 is the possession of a highly conserved heme-binding motif (-HPGG-) that locates in the large N-terminal domain typically exposed to the cytosol (Schenkman and Jansson, 2003) (**Figure 1**). With bound heme molecule, CB5 possesses redox potential of ~20 mV, capable of accepting and transferring a single electron. It can be reduced either by NADH-dependent cytochrome b_5 reductase (CBR) or by NADPH-dependent cytochrome P450 reductase (CPR), therefore, shuttling electron(s) in either NADH-CBR-CB5 chain or NADHP-CPR-CB5 pathway at the ER membrane to the terminal acceptors (proteins or enzymes) involved in oxidation/hydroxylation reactions (Vergeres and Waskell, 1995; Porter, 2002; Schenkman and Jansson, 2003). As an electron donor protein, CB5 has been extensively studied in mammals in respect to its roles in cellular detoxification and drug metabolism. It has been found to function in anabolic metabolism of fatty acids and steroids, catabolism of xenobiotics and compounds of endogenous metabolism (Vergeres and Waskell, 1995; Porter, 2002; Schenkman and Jansson, 2003).

In contrast to the single copy of *CB5* gene in mammals and yeast (Porter, 2002; Kandel and Lampe, 2014), higher plants including Arabidopsis have evolved multiple copies of the gene (Hwang et al., 2004; Maggio et al., 2007). The significant expansion of the CB5 family in the plant kingdom implicates that this group of heme-containing proteins might play broader and more complicated cellular and metabolic functions than their mammalian/human counterparts. Similar to animal/human ones, the ER-resident CB5s in plants, together with their redox partners, have also been implicated as the electron carriers to deliver reducing equivalents from pyridine nucleotide cofactor to the processes of acyl-CoA/fatty acid desaturation, hydroxylation, and triple bond formation (Smith et al., 1992; Napier et al., 1997; Lee et al., 1998; Broadwater et al., 2002; Kumar et al., 2012), sphingolipid long-chain base hydroxylation and desaturation (Napier et al., 2003; Nagano et al., 2012), or sterol desaturation (Rahier et al., 1997). Moreover, sporadic evidence indicate that plant CB5 proteins also function in the redox reactions of broader metabolic processes, in particular, plant specialized metabolism for synthesis of flower pigment anthocyanins, UV-resistance phenolic esters, pathogen defensing phytoalexins, and the cell wall structural component lignin. Furthermore, either as the redox protein or structural component, plant CB5s likely serve as metabolic or signaling regulators, connecting hormone or sugar signals, or environmental stimulus with particular metabolic or cellular processes. This article overviews the studies pertaining to plant CB5s and discusses their biological functions in different metabolic and physiological processes, with which we expect to trigger more research interests in deciphering the metabolic and cellular roles of this group of non-catalytic proteins in plant growth, development and defense responses.

Evolutionary expansion of plant cytochrome b_5 family

Cytochrome b_5 are found in animals/human, plants, yeasts and purple phototrophic bacteria, indicative of their early emergence and evolution (Schenkman and Jansson, 2003). Yeast and human genomes contain a single copy of *CB5* gene, although the human gene generates two isoforms via alternative mRNA splicing. The isoform 1 locates to the cytoplasmic side of the ER, while isoform 2 is in cytoplasm (Porter, 2002; Schenkman and Jansson, 2003; Kandel and Lampe, 2014). By contrast, the genomes of higher plants typically evolve multiple *CB5* genes (Smith et al., 1992; Napier et al., 1995; Fukuchi-Mizutani et al., 1999; Hwang et al., 2004; Kumar et al., 2006; Maggio et al., 2007; Kumar et al., 2012). *Arabidopsis thaliana* possesses five canonical *CB5* genes encoding isoforms AtCB5-A (At1g26340), AtCB5-B (At2g32720), AtCB5-C (At2g46650), AtCB5-D (At5g48810), and AtCB5-E (At5g53560), which share amino acid identities ranging from 40 to 70% (**Figure 1A**). All five isoforms possess the conserved features of CB5 family: A N-terminal water-soluble/cytosolic heme binding domain, a highly flexible linker region that possibly provides the directional freedom required for an efficient complex formation with terminal enzymes such as P450s, and a C-terminal transmembrane domain that anchors protein to either the ER or mitochondria/chloroplast, followed by a short luminal tail (Ahuja et al., 2013; Pandey et al., 2013) (**Figure 1B**). In addition to the five canonical CB5s, a heme-binding protein encoded by At1g60660 possesses a short transmembrane domain at its N-terminus, which is defined as CB5-like protein, AtCB5F or AtCB5LP (Hwang et al., 2004; Maggio et al., 2007) (**Figure 1A**); among the annotated five canonical CB5 proteins, AtCB5-B, -C, -D and -E share more sequence similarity; each of their tail sequences carry a conserved ER-targeting motif (-R/H-x-Y/F-) and are proven or predicted to localize to the ER membrane (Hwang et al., 2004; Maggio et al., 2007), while AtCB5-A is demonstrated to localize to the chloroplast envelope (Maggio et al., 2007). Similarly, rice genome encodes eight CB5 isoforms and soybean encompasses about 11 putative CB5 members (Kumar et al., 2006). Overall, the number of plant CB5 family members is much larger than that in yeasts and animals. The

exact evolutionary significance of such large expansion of plant CB5 family remains to be determined. Given the fact that sessile plants need to produce a plethora of defense-related specialized metabolites to deal with ever changing environment, CB5 family might co-evolve with the massive expansion of plant metabolic enzymes such as cytochrome P450 superfamily enzymes to invent metabolic complexity, versatility and robustness.

It is worthwhile to note that the nomenclature of plant CB5 proteins is inconsistent and somehow confusing in the literatures. For example, on one hand, the same Arabidopsis CB5 member has a few different names; on the other hand, a same given name denotes to the different family members in different studies (see **Figure 1** Legend for more information). Attention should be paid to avoid potential misunderstanding.

Cytochrome b_5 functions as redox cofactor in biosynthesis of fatty acids

Fatty acid desaturation and hydroxylation

Analogous to animal/human counterparts, plant CB5s are widely implicated as a redox cofactor in the biosynthesis of plant specialty fatty acids, and the membrane or cuticular lipids. CB5s are commonly believed to associate with NADH-dependent CBR, and shuttle reducing equivalents from reductant NADH to the terminal acceptors. In most cases, CB5 transfers electrons to the ER-localized non-heme, iron-containing enzymes involved in fatty acid desaturation, hydroxylation, and elongation (Napier et al., 2003; Nagano et al., 2012). Hydroxyl fatty acids such as ricinoleic acid (12-hydroxyoctadec-*cis*-9-enoic acid) are the important feedstocks for industrial applications; while polyunsaturated fatty acids such as ω-3 fatty acids are the essential nutrient components beneficial to human health. Early biochemical studies discovered that CB5 proteins presented in many oilseed species. The CB5 protein in microsomal preparation is necessarily required for the conversion of oleate to polyunsaturated linoleate in safflower (*Carthamus tinctorius*) (Smith et al., 1990), or to ricinoleic acid in castor bean (Smith et al., 1992). Applying anti-CB5 antibody raised against the hydrophilic CB5 fragment from cauliflower (*Brassica oleracea*) to the microsomal enzyme assay system inhibits both Δ12-hydroxylase and Δ12-desaturase activities in the prepared microsomes from castor bean, suggesting that CB5 is the indispensable redox component for the membrane resident, non-heme fatty acid desaturase and hydroxylase systems (Smith et al., 1992). Arabidopsis CB5 is reduced by either NADH-dependent CBR or NADPH-dependent CPR (Fukuchi-Mizutani et al., 1999). Disruption of Arabidopsis *CBR* in either the wild type background or the transgenic line with overexpression of castor fatty acid 12-hydroxylase (FAH12) significantly depleted the accumulation of 18 carbon hydroxy fatty acids and unsaturated fatty acids in developing seeds, suggesting that the conventional NADH-CBR-CB5 electron transfer chain is responsible for fatty acid hydroxylation and desaturation (Kumar et al., 2006).

To discriminate the potential differential contributions of CB5 isoforms in fatty acid synthesis, Kumar et al. (2012) co-expressed Arabidopsis ER-resident CB5s, AtCB5-B, -C, -D and -E with Arabidopsis fatty acid desaturase FAD2 and FAD3, respectively, in a yeast *cb5* mutant to measure the production of ω-3 and ω-6 desaturation of C16 and C18 fatty acids. All four CB5s were able to enhance the accumulation of either di- or tri-unsaturated fatty acids, compared to the yeast cells expressing FAD2 or FAD3 alone. However, they exhibited distinct effects on the production of unsaturated fatty acid species. AtCB5-C and -D (denoted as Cb5-C and Cb5-B in the study) significantly enhanced the yield of 16:2 and 18:2 di-unsaturated fatty acids, about 1.5–2-fold higher than did AtCB5-B (Cb5-E) or AtCB5-E (Cb5-A) with FAD2; whereas AtCB5-B and AtCB5-E, when co-expressed with FAD3, yielded the better production of 18:3 tri-unsaturated fatty acids (Kumar et al., 2012). Although the data from yeast heterologous system might not necessarily reflect the *in planta* functionalities of CB5s, the study implicates that Arabidopsis CB5 isoforms exhibit differential stimulatory effects on FAD2 and FAD3 activities. Further genetic evidence is required for validating the functional differentiation of CB5s in the production of unsaturated fatty acid *in planta*.

Very long chain fatty acid elongation

Very long chain fatty acids (VLCFAs) are the compounds with an acyl chain of 18 carbons and longer. In plants, VLCFAs are incorporated into various lipid pools, including the neutral seed storage lipids triacylglycerols, the membrane constituents and signaling molecules phospholipids and sphingolipids, and the extracellular epicuticular waxes and suberin (Bach and Faure, 2010; Haslam and Kunst, 2013). The content of VLCFAs increases in response to various environmental stresses such as salt, wounding, drought, hypoxia, and pathogen infection (De Bigault Du Granrut and Cacas, 2016). VLCFAs are synthesized via sequential addition of two carbons to the C16 or C18 acyl-CoAs, catalyzed by elongase complex in the ER. This complex is composed of four enzymes, 3-ketoacyl coenzyme A synthase (KCS), 3-ketoacyl-CoA reductase (KCR), 3-hydroxyacyl-CoA dehydratase (HCD) and *trans*-2,3-enoyl-CoA reductase (ECR) that carry on sequential reactions of acyl-CoA condensation, reduction, dehydration then reduction (Haslam and Kunst, 2013). In Arabidopsis, the CoA condensing enzymes are encoded by 21 *FAE1*-like *KCS* genes and 4 *ELO*-like genes (*AtELO1, At3g06460; AtELO2, At3g06470; AtELO3, At1g75000; and AtELO4, At4g36830*) (Dunn et al., 2004; Joubes et al., 2008).

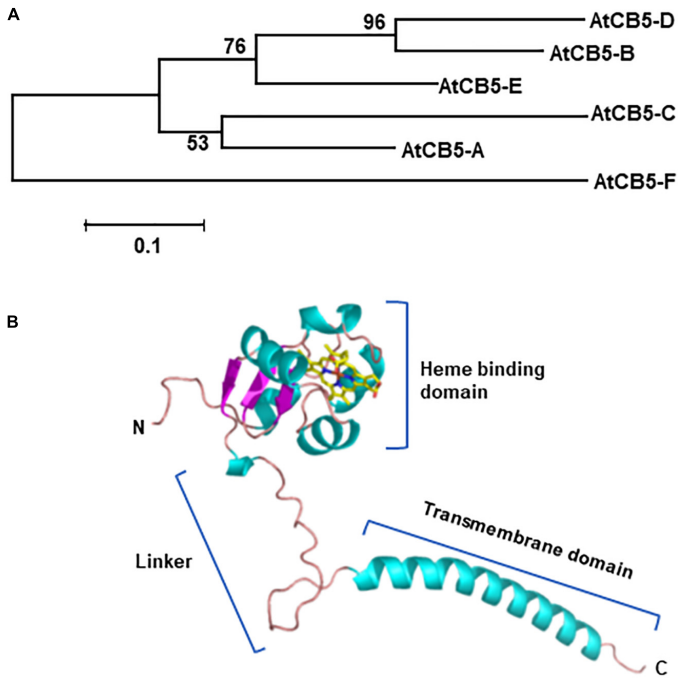


FIGURE 1
Phylogenetic relationship of Arabidopsis cytochrome b_5 proteins **(A)** and structural model of one of the cytochrome b_5 family member AtCB5-D **(B)**. Protein sequences of 6 Arabidopsis CB5 homologs were used in the analysis. The sequences were aligned with ClustalW integrated in the MEGAv.7.0 program. The evolutionary history was inferred using the neighbor joining method. The percentages (>50%) of replicate trees in which the associated genes clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The units represent the number of amino acid substitutions per site. Structural model was built with Alphafold 2. AtCB5-A: AT1G26340, cytochrome b_5 isoform A. Also named as B5#6, CB5-A. AtCB5-B: AT2G32720, cytochrome b_5 isoform B. Also named as B5#4, CB5-B, CYTB5-D. AtCB5-**C**: AT2G46650, cytochrome b_5 isoform C. Also named as B5#1, CB5-C, CYTB5-C. AtCB5-D: AT5G48810, cytochrome b_5 isoform D. Also named as B5#3, CB5-D, ATB5-B, CYTB5-B. AtCB5-E: At5G53560, cytochrome b_5 isoform E. Also named as B5#2, CB5-E, ATB5-A. AtCB5-F: AT1G60660, cytochrome b_5 like protein. Also named as ATCB5LP or CB5LP.

All four ELO homologs localize to the ER-membrane but only AtELO1 and AtELO2 contain a HXXHH motif that is reminiscent of the His-rich motif generally conserved in the integral membrane-bound desaturases, hydroxylases, and elongases (Fox et al., 1993; Mitchell and Martin, 1997; Oh et al., 1997). The histidine residues of this motif coordinate a μ-oxo diiron cluster (Fe–O–Fe) that forms part of the catalytic core. AtELO3 and AtELO4 possess neither the characteristic HXXHH motif nor the conserved amino acids, but remain their condensation activities (Quist et al., 2009; Nagano et al., 2019). Interestingly, split ubiquitin membrane yeast-two-hybrid (Y2H) and biomolecular fluorescence complementation (BiFC) assay revealed that AtCB5-B interacts with AtELO1 and AtELO2 but not AtELO3 andAtELO4 in yeasts and in plants (Quist et al., 2009; Nagano et al., 2019), suggesting that the His-rich motif might be a critical structural feature for the physical interaction between those non-heme enzymes and the CB5 proteins but such interaction might not be critical for the condensation activity. This invokes an interesting question why AtELO1 and AtELO2 need to interact CB5 protein. BiFC assay also revealed that both AtCB5-B and AtELO2, respectively, interact with or are in close proximity to the VLCFA elongase complex enzymes KCR1, PAS2/HCD, and CER10/ECR (Nagano et al., 2019), implicating that AtCB5-B and elongase components might form a large protein complex *in planta*. Within the elongase complex, KCR1 and CER10/ECR catalyze acyl-CoA reductions. Their reactions require reducing power. Therefore, CB5 integrated into the VLCFA elongase complex might serve as a necessary electron donor for the reductase activity. On the other hand, electrons are not needed for the ELO/KCS-catalyzed condensation or PAS2/HCD-catalyzed dehydration. The direct physical interactions between those proteins/enzymes and AtCB5 infer that CB5 may potentially play the roles beyond as an electron donor, for instance, as allosteric stimulator for the elongase components.

Very long chain alkane synthesis

The elongated acyl-CoAs can serve as metabolic precursors for the formation of several classes of lipids. Two distinct biosynthetic pathways direct them to the waxes: The decarbonylation pathway and the reduction pathway.

Decarbonylation process begins with the production of aldehydes from acyl-CoA by acyl-CoA reductase, followed by a decarbonylation catalyzed by an aldehyde decarbonylase to produce alkanes with odd-chains. The resulting alkanes are then hydroxylated to the secondary alcohols, which are finally oxidized to ketones (Bernard and Joubes, 2013). In Arabidopsis, the wax-associated protein ECERIFERUM1 (CER1) and ECERIFERUM3 (CER3) are the core components of very long chain alkane synthesis complex, and two proteins associate together (Bourdenx et al., 2011; Bernard et al., 2012). Both split ubiquitin Y2H and split luciferase assays revealed that CER1 and a CER1-like protein physically interact with all four ER-localized CB5 proteins (AtCB5-B, -C, -D and -E) (Bernard et al., 2012; Pascal et al., 2019). Co-expression of AtCB5-B with CER1/CER3 complex significantly increased the production of alkanes with distinct chain-length specificity in yeast. Both CER1 and CER3 possess tripartite His-rich motifs; point mutations of the His-rich motifs in CER1 diminished the alkane-forming activity of the CER1/CER3/AtCB5-B complex, but mutations of the His-rich motifs in CER3 did not compromise alkane production (Bernard et al., 2012; Pascal et al., 2019). The data suggest that AtCB5 protein likely serve as the electron donor mediating electron transfer to the catalytic site of the CER1 or CER1-like enzyme, which is necessary for the product specificity and overall productivity of very long chain alkane synthesis.

Sphingolipid fatty acid hydroxylation

Sphingolipids are the important structural components of plasma membrane (PM), accounting for more than 40% of its lipids. Sphingolipids are also found in the Golgi network and in endosomes (Bach and Faure, 2010; Haslam and Feussner, 2022). Complex sphingolipids are composed of various head groups and a ceramide formed by linking long chain base amine group to fatty acids. A major structural feature of plant sphingolipids is the hydroxylation of the C-2 position in the fatty acid molecules. More than 90% of the sphingolipids in Arabidopsis contain 2-hydroxy fatty acids. Sphingolipids are also composed of very-long-chain fatty acids (VLCFAs), which have more than 20 carbons (Haslam and Kunst, 2013; Haslam and Feussner, 2022). Similar to other fatty acid hydroxylases, e.g., yeast homolog ScFAH1, Arabidopsis sphingolipid fatty acid 2-hydroxylases AtFAH1 and AtFAH2 contain five conserved His-rich motifs [HX2-3(XH)H] that form part of di-iron cluster to receive electrons. Like CER1, both AtFAH1 and AtFAH2 physically interact with all five Arabidopsis CB5 proteins (Nagano et al., 2009), suggesting CB5 might be the indispensable electron donor in FAH-catalyzed sphingolipid 2-hydroxylation.

Connector of environmental signals and fatty acid biosynthesis

Interestingly, CB5 proteins not only interact with electron acceptors involved in fatty acid desaturation, hydroxylation and elongation, but also physically associate with the non-catalytic regulatory component relevant to the lipid biosynthesis. In Arabidopsis, the cell death suppressor Bax inhibitor-1 (BI-1), an ER membrane protein, functionally associates with various environmental stresses. Overexpression of BI-1 suppressed cell death and conferred tolerance to the oxidative, salinity and drought stresses in Arabidopsis, tobacco, rice, and sugarcane (Isbat et al., 2009; Ishikawa et al., 2010; Ishikawa et al., 2011; Ramiro et al., 2016). Moreover, overexpression of BI-1 also increased the production of the 2-hydroxylated VLCFAs in the normal condition and promoted the rapid synthesis of 2-hydroxy VLCFAs in the plants under oxidative stress, suggesting BI-1 is responsive to environmental stimuli and regulates fatty acid 2-hydroxyalse activity (Nagano et al., 2009). Notably, AtBI-1 physically interacts with the ER-resident CB5s and the yeast ScFAH1 that contains a CB5-like domain, but not with AtFAHs that lack such domain. Nevertheless, CB5 directly interacts with AtFAHs (Nagano et al., 2009). Therefore, CB5 likely acts as a linker mediating the functional association of AtBI-1 with AtFAHs. Furthermore, co-immunoprecipitation using AtBI-1 as bait pulled down VLCFA synthesizing enzymes, including AtELO2, KCS10, KCR, PAS2, CER10, and AtCB5, suggesting CB5 may also mediate the association of AtBI-1 with VLCFA elongase complex (Nagano et al., 2019). Since AtBI-1 interacts with Arabidopsis calmodulin (AtCaM) to perceive environmental simulation (Ihara-Ohori et al., 2007), it is possible that AtBI-1 transduces environmental signals from AtCaM through interacting with CB5 to regulate CB5-FAH complex activity thus enhancing 2-hydroxylation of fatty acids, and/or modulating CB5-elongase complex function. Consequently it might alter sphingolipid structures and properties thus inactivating the ceramide-mediated signal transduction in cell death or reinforcing membrane microdomain, which ultimately suppresses plant cell death process. In such processes, CB5 not only acts as an electron carrier but also as a regulator conveying environmental signals from AtBI-1.

Cytochrome b_5 acts as electron donor for cytochrome P450-catalyzed reactions

When pioneer land plants migrated from aquatic habitats to terrestrial environment ~500 million years ago, the overwhelming biotic and abiotic environmental stresses become the primary driving force to accelerate the evolution of plant

metabolic capacity. The most outstanding evolution event is the massive expansion of cytochrome P450 systems, which eventually leads to by far the largest family of enzymes in plant metabolism (e.g., 245 super family members emerging in Arabidopsis) (Bak et al., 2011; Nelson and Werck-Reichhart, 2011). The cytochrome P450 enzymes have a key function for generating the chemical diversity, which is the hallmark of plants compared with animals. Plant P450s are crucial for the biosynthesis and metabolism of fatty acids, phytosterols, plant growth regulators, and a variety of plant specialized metabolites, including phenylpropanoids, terpenoids, glucosinolates, and indoalkaloid phytoalexins (Mizutani and Ohta, 2010; Bak et al., 2011). Interestingly, although P450 enzymes play central roles in various metabolic processes, this family of heme-containing oxidases are functionally self-insufficient. Like non-heme-containing desaturases and hydroxylases in fatty acid modification, eukaryotic microsomal P450s require redox partner(s) to deliver two electrons to its catalytic center for cleavage of oxygen molecule in each catalytic cycle (Hannemann et al., 2007). Typically, microsomal P450s rely on diflavin reductase, the cytochrome P450 oxidoredcutase (CPR), to transfer electrons from cofactor NADPH to the prosthetic heme group of P450 (Urban et al., 1997; Jensen and Moller, 2010). Evidence from mammalian P450 systems also reveal that CB5 can be reduced by NADPH-dependent CPR (Enoch and Strittmatter, 1979; Vergeres et al., 1995; Porter, 2002). In such case, CB5 delivers second electron to the P450 ferrous-O_2 complex (Durr et al., 2007). CB5 augments a subset of P450-catalyzed reactions in mammalian systems. It displays either stimulation, no effect, or inhibition effects in different cases (Porter, 2002; Bart and Scott, 2017). In plants CB5s are also indispensable for some of P450-catalzyed reactions in the synthesis of specialized metabolites or cell wall structural component lignin. In petunia, disruption of a *cb5* locus, *DifF,* resulted in the discoloration of its flower petals, which coincides with the compromised activity of flavonoid 3′, 5′-hydroxylase (F3′5′H), the P450 enzyme leading to the synthesis of flavonols and the core structure of anthocyanidin delphinidin (de Vetten et al., 1999). Furthermore, a Arabidopsis CB5 family member, the AtCB5-D, was recently discovered as the key determinant for the synthesis of syringyl lignin subunits and the related soluble 5-hydroxylated phenolics, sinapoyl esters (Gou et al., 2019). Disruption of *AtCB5-D* resulted in more than 60% reduction of S-lignin deposition without impairment of guaiacyl lignin accumulation in the cell walls of Arabidopsis stem and ∼70% reduction of the wild type accumulation level of sinapoyl malate, a photoprotectant phenolic ester accumulated in the leaf epidermis of *Brassicaceae* family (Shirley et al., 2001; Milkowski and Strack, 2010). The loss of AtCB5-D primarily impaired the activity of ferulate 5-hydroxyalse 1 (AtF5H1, CYP84A1) that is a key branch point enzyme catalyzing benzene ring 5-hydroxyaltion and leading to the S-lignin monomer formation in angiosperms. The loss of the function of *AtCB5-D* also significantly suppressed the accumulation of α-pyrones, the compounds yielded from the activity of AtF5H2 (CYP84A4, At5G04330), a close paralog of AtF5H1 in Arabidopsis (Weng et al., 2012). In *cb5d* stem, the contents of the major α-pyrones arabidopyl alcohol and iso-arabidopic acid were reduced by ∼80% compared to the WT (Gou et al., 2019). However, disruption of *AtCB5-D* did not affect the accumulation of seed flavonol quercetin, the metabolite resulted from flavonoid 3′-hydroxylase (F3′H, CYP75A1) activity (Gou et al., 2019). AtF3′H is an evolutionarily close homolog of AtF5H1 and AtF5H2 (Weng et al., 2008). These data suggest that AtCB5-D possesses relatively strict specificity and supports a particular set of P450 enzymes in planta. Interestingly, both Y2H and BiFC assays revealed that AtCB5-D physically interacts with all three monolignol biosynthetic P450s, cinnamate 4-hydroxyalse (C4H), *p*-coumaryl ester 3′-hydroxyalse (C3′H) and F5H1, but genetic data clearly showed that AtCB5-D functionally only associates with F5H specific for the synthesis of S-lignin monomer and the related 5-hydroxylated phenolics (Gou et al., 2019). Therefore, it is of high interest to explore what factor(s) determine CB5's functional speciation with P450 enzymes. On the other hand, AtCB5-D shares high sequence identity with other ER-localized CB5 members (44∼68%), particularly with AtCB5-B (68%) at amino acid level (**Figure 1A**). Although those ER-resident CB5 members also appear to interact/associate with monolignol biosynthetic P450s in Y2H or BiFC assay, only AtCB5-D imposes effects on lignin biosynthesis. It is intriguing to determine how CB5s distinguish their functionalities at the structural and/or molecular levels.

Although lack of the in-depth characterization, AtCB5-C was casually linked to the glucosinolate biosynthesis in a condition-specific manner (Vik et al., 2016). Glucosinolates, also known as mustard oil glucoside, are nitrogen- and sulfur-containing small molecule bioactive compounds found in the order *Brassicales*, including Arabidopsis and oilseed rape (*Brassica napus*). They function as defense compounds against insects and pathogen infection (Halkier and Gershenzon, 2006) but are also of great interest to humans due to the flavors they impart to various condiments and for their potential anticancer effects (Traka and Mithen, 2009). The function of AtCB5-C does not seem to be essential for the overall glucosinolate biosynthesis but it influences accumulation of particular long chain aliphatic glucosinolate species, especially when induced by MeJA treatment (Vik et al., 2016), which infers the potential functional association of this CB5 protein with CYP79F enzyme that catalyzes the key step in the biosynthesis of long chain aliphatic glucosinolates.

As electron shuttle intermediate, CB5s can transfer two electrons from NADH through CBR to the terminal enzymes, independent of NADPH-CPR; or transfer the second electron to oxyferrous P450 from CPR (Porter, 2002;

Schenkman and Jansson, 2003). Arabidopsis CB5s are reduced both by CBR and CPR. However, both reductases display strict specificity toward pyridine nucleotide cofactors NADH and NADPH in reducing CB5 proteins. The recombinant AtCBR specifically utilizes NADH to reduce AtCB5; whereas AtCPR shows a sharp specificity to NADPH (Fukuchi-Mizutani et al., 1999). Notably, disrupting the ER-localized CBR in Arabidopsis resulted in no significant impairment on mature stem lignin biosynthesis, although a slight (but not statistically significant) reduction of S-lignin level occurred in the knock-out mutant (Gou et al., 2019). By contrast, disrupting Arabidopsis *CPR2*, namely *ATR2*, suppressed both G- and S-lignin synthesis up to 50% (Sundin et al., 2014; Gou et al., 2019). These data implicate that AtCB5-D in stem primarily couples with NADPH-CPR electron transport pathway for lignin synthesis. Collectively, CB5s might act as versatile electron carriers *in planta* that associate with different electron transfer chains for synthesis of distinct classes of metabolites.

Cytochrome b_5 functions more than as an electron carrier

Sugar sensing through interactions with sugar transporters

While CB5 is commonly documented as an electron shuttle intermediate, sporadic evidence also suggest some non-conventional roles it fulfils. An early study revealed that human/mammalian CB5 with its heme molecule replaced with manganese-protoporphyrin IX was unable to accept and transfer electrons from CPR or CBR but retained the ability in decreasing the *Km* values of P450 enzymes when it was included with the reconstituted P450 systems containing CYP2B4 or CYP1A2 (Morgan and Coon, 1984). Moreover, the apo form CB5 devoid of heme is still able to enhance CYP3A4-catalyzed reactions as efficiently as does the holoprotein, suggesting that CB5 might act as an allosteric stimulator to tune P450 activity (Porter, 2002). While it remains to be determined whether plant CB5s also possess allosteric stimulation effect on the oxidase-catalyzed reactions, an apple (*Malus domestica*) ER-resident CB5 protein, MdCYB5, was found to physically interact with the plasma membrane (PM) – localized, low-affinity sucrose transporter MdSUT1, and sorbitol transporter MdSOT6. The interactions tune the affinities of both sugar transporters, allowing plant cells to adapt to sugar starvation (Fan et al., 2009). MdCYB5 is the close homolog of AtCB5-E (designated as AtCYB5-2/A in the study). The pair-wise interactions of MdCYB5 with MdSUT1 and MdSOT6 were well validated through split ubiquitin Y2H assay, co-immunoprecipitation/pull down, and *in planta* BiFC assay. In the yeast system, the low sugar supply enhances the interaction of MdCYB5 with either sugar transporters. Co-expression of MdCYB5 with MdSUT1 or MdSOT6 promotes the uptake of sugars and reduces the *Km* value of either transporters toward their substrate sucrose or sorbitol, thus increasing transporter affinity to sugar (Fan et al., 2009). Hypothetically MdCYB5 with its C-terminal tail anchored to the ER membrane interacts with the PM-localized sugar transporter MdSUT1 or MdSOT6, thus forming an MdSUT1-MdCYB5 or MdSOT6-MdCYB5 bimolecular complex in response to the low sucrose or sorbitol supply. The interaction enhances the affinity of MdSUT1 or MdSOT6 to its substrate sugar, thereby stimulating sugar uptake to maintain a relatively stable sugar level inside of the cell; whereas, under high sugar supply condition, the transporter-MdCYB5 interactions are attenuated and the affinity of the transporters to sugar substrate returns to the normal level (Fan et al., 2009). Besides apple CB5, the Arabidopsis CB5s are also functionally associated with sucrose transporter AtSUT4, the homolog of MdSUT1 but localized to tonoplast membrane instead of PM. All five Arabidopsis canonical CB5s physically interact with AtSUT4 (Li et al., 2012). Similar to the observed interaction of MdCYB5 and MdSUT1, sucrose but not glucose represses the interaction between AtSUT4 and AtCB5-E (Li et al., 2012). Despite the lack of adequate genetic validation, these several lines of biochemical evidence implicate that CB5 proteins might allosterically modulate sugar transporter activity in response to sugar/carbon source availability. The interaction of CB5 with sugar transporter tunes the transport kinetics, thereby sustaining cellular sugar/carbon homeostasis or modulating sugar signaling.

Linking to ethylene signaling

Ethylene plays important roles in plant growth, development, and stress responses. It is often considered as an "aging" hormone due to its role in accelerating such developmental processes as ripening, senescence, and abscission. Ethylene is perceived by a family of receptor proteins that repress ethylene responses when ethylene is absent. Repression function of the ethylene receptor ETR1 depends on an integral membrane protein, REVERSION TO ETHYLENE SENSITIVITY1 (RTE1). The RTE1 protein acts at the upstream of ETR1 in the ER membrane and Golgi apparatus and serves as a molecular chaperone stabilizing or promoting the active signaling conformation of ETR1 (Dong et al., 2008; Resnick et al., 2008). Y2H and BiFC assays reveal that four Arabidopsis ER-localized CB5 isoforms (AtCB5-B, -C, -D, and -E) interact with RTE1 (Chang et al., 2014). Consistent with their interactions, *atcb5* mutants phenocopy Arabidopsis *rte1* line, showing partial suppression of *etr1-2* ethylene insensitivity. The AtCB5s exhibit partial functional redundancy. The single mutants of *atcb5-b*, *-c* and *-d* appear

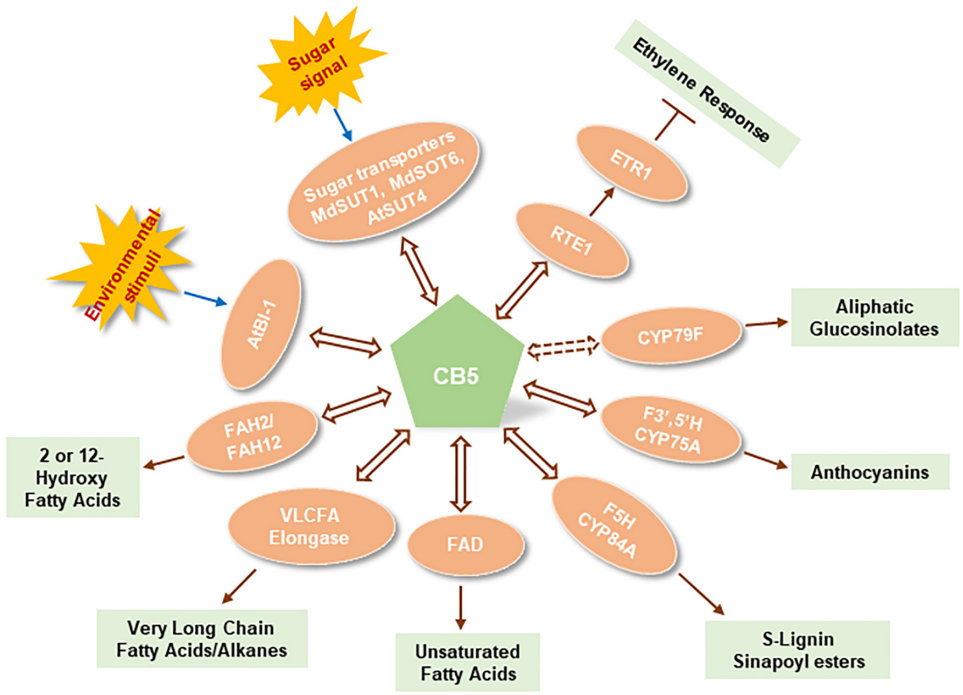


FIGURE 2
Functional association of plant cytochrome b_5s with biological factors in different metabolic and cellular processes. AtBI-1, *Arabidopsis thaliana* Bax inhibitor-1; AtSUT4, *Arabidopsis thaliana* sucrose transporter 4; MdSUTTI, *Malus domestica* sucrose transporter 1; MdSOT6, *Malus domestica* sorbitol transporter 6; *RTE1*, Reversion to ethylene sensitivity 1; *ETR1*, Ethylene receptor 1; CYP79F, Cytochrome P450 superfamily enzyme member 79F; F3′, 5′H, Flavonoid 3′, 5′-hydroxylase; F5H, Ferulate 5-hydroxylase; FAD, Fatty acid desaturase; VLCFA Elongase, Very long chain fatty acid/alkane elongase enzyme complex; FAH2, Fatty acid 2-hydroxylase; FAH12, Fatty acid 12-hydroxylase. Double head arrows with solid line indicates the experimentally recognized interactions; Double head arrow with dashed line indicates the putative functional association.

similarly to the wild type, but their double mutants display slight ethylene hypersensitivity. Conversely, over-expression of *AtCB5-D* confers a reduced ethylene sensitivity similar to that did the *RTE1* overexpression. These findings suggest an unexpected regulatory role of CB5 protein, i.e., *via* partnering with RTE1 to promote ETR1-mediated repression of ethylene signaling. In this process, AtCB5 may activate RTE1 through redox modification, thus linking cellular redox status with ethylene signaling (Chang et al., 2014). Since AtCB5-D is also an indispensable electron carrier involved in S-lignin biosynthesis, it therefore may serve as a regulatory/metabolic hub coordinating ethylene signaling and lignin synthesis in response to the cellular redox status. It is known that coordinated ethylene and auxin signaling plays an essential role in initiating and programming organ abscission. In the abscission zone, lignin is synthesized in secession cells to restrict cell wall processing enzymes to the precise area thus controlling precision process of cell walls for organ separation (Lee et al., 2018). The functional association of AtCB5-D with both ethylene signaling and lignin biosynthesis implicates that this redox protein might play exquisite regulatory roles coordinating cellular and biochemical activities in leaf, flower, or fruit abscission process.

Biotechnological applications

The collective information suggests that plant CB5s exhibit versatile cellular functions and are involved in more complicated metabolic and signaling processes relative to their yeast, and animal counterparts. This family of proteins not only serve as conventional electron donor to drive redox reactions but may also act as physiological regulator functioning in environmental stress response, sugar and hormonal signaling processes. The electron donor function of CB5s in the defined metabolisms makes them an essential target in metabolic engineering for enhancing the production of high value bioproducts. In an effort to produce artemisinic acid, the precursor of anti-malarial drug artemisinin, in heterologous yeast (*Saccharomyces cerevisiae)* system, expression of *Artimisia annua* P450 enzyme CYP71AV1 with its cognate reductase (*CPR1*) accelerated amorphadiene oxidation, producing artemisinic acid at the yield of 115 mg/L (Ro et al., 2006). However, the engineered yeast cells suffered severe oxidative stress due to the poor coupling between cytochrome P450 and its reductase thus releasing ROS (Paddon and Keasling, 2014). Integration of a *A. annua* CB5, as well as artemisinic aldehyde dehydrogenase (*ALDH1*) and NAD-dependent artemisinic alcohol dehydrogenase (*ADH1*), together

with CYP71AV1 and CPR1 significantly improved artemisinic acid yield (up to 25 g/L) (Paddon et al., 2013). Similarly, when using the engineered yeast cells to produce glycyrrhetinic acid (GA), the most essential ingredient in licorice with outstanding anti-inflammatory activity and the wide usage in medicine and cosmetics industries, introduction of the entire heterogeneous biosynthetic pathway of GA, including CYP88D6 and CYP72A154 combined with β-amyrin synthase and a *Arabidopsis thaliana* CPR into *S. cerevisiae*, only produced of 2.5 mg/L of β-amyrin and 14 μg/L of GA. However, further incorporation of a CB5 from *Glycyrrhiza uralensis* resulted in eightfold enhancement of GA production. Furthermore, combining GuCB5 with other MVA pathway genes from *S. cerevisiae,* GA concentration was improved by 40-folds during batch fermentation (Wang et al., 2019). Both cases exemplify the significance of application of redox component CB5 in metabolic engineering to enhance the production of the desired metabolites.

Conclusion and perspectives

Cytochrome b_5 is a well-known electron shuttle intermediate in yeasts, and mammals/humans. It is involved in different oxidation/reduction reactions for biosynthesis of endogenous compounds such as steroids, vitamins, and fatty acids, and for the metabolism of xenobiotics and drugs (Schenkman and Jansson, 2003). In the endomembrane P450 system, mammal/human CB5 modulates P450 catalysis with either stimulating or inhibiting effects. Functionally interacting with P450s, mammal/human CB5 either plays a pure redox role of electron delivery or acts as an allosteric modulator of P450 conformation (Porter, 2002). CB5 family is largely expanded in plants, which strongly implicates its more complicated and diverse cellular/biological functions. Nevertheless, in contrast to the extensive studies on mammalian/human counterparts, comprehensive understanding on plant CB5 functions has not been achieved so far. Sporadic evidence reveal that plant CB5s robustly interact with different types of proteins ranging from catalytic enzymes such as P450 monooxygenases, desaturases, reductases, hydroxylases, to transporters, stress response – or hormone signaling regulators. Such versatile but most likely transient physical interactions infer that plant CB5s not only simply act as electron carriers modulating enzymatic catalyses and metabolic processes but may also serve as the cellular regulators or modulators that connect and coordinate different biological processes in response to the environmental stresses, and/or to the fluctuation of cellular redox and carbon status. A summary of plant CB5 functions are depicted in **Figure 2**. Nevertheless, to date we still lack a comprehensive insight into the biochemical and biological significance of plant CB5 proteins. Many open questions remain to be addressed to deeply understanding CB5 functionalities in plants. For instance, (1) With multiple CB5 family members that plants evolved, do they play distinct biological functions or act redundantly in plant metabolisms? (2) As redox factors, what are the additional metabolic and cellular processes that CB5 proteins are involved in? (3) As an electron shuttle component, do the CB5 proteins associate with different electron transport pathways for different classes of metabolic processes? (4) What is the biochemical significance of a P450 system that necessarily recruits CB5 protein in an electron transport pathway instead employs the typical NADPH-CPR electron transfer chain? (5) What is the molecular and structural basis for a P450 system necessitating CB5 proteins? The insight gained from the exploration of this family of heme-containing proteins might offer more sophisticated and versatile molecular tools to augment our ability in engineering plant metabolism to effectively utilize the photosynthetically fixed reduced carbon and reducing power for the production of desired bioproducts.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE-SC0012704—specifically through the Physical Biosciences program of the Chemical Sciences, Geosciences and Biosciences Division (to C-JL). This work was also partially supported by the subcontract awards from the Joint BioEnergy Institute (http://www.jbei.org), the Center for Bioenergy Innovation (https://cbi.ornl.gov/), and two of Bioenergy Research Centers under Contract Nos: DE-AC02-05CH11231 and DE-AC05-00OR22725, respectively, with US DOE, Office of Science, Office of Biological and Environmental Research.

Conflict of interest

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

Publisher's note

References

Ahuja, S., Jahr, N., Im, S. C., Vivekanandan, S., Popovych, N., Le Clair, S. V., et al. (2013). A model of the membrane-bound cytochrome b5-cytochrome P450 complex from NMR and mutagenesis data. *J. Biol. Chem.* 288, 22080–22095. doi: 10.1074/jbc.M112.448225

Bach, L., and Faure, J. D. (2010). Role of very-long-chain fatty acids in plant development, when chain length does matter. *C. R. Biol.* 333, 361–370. doi: 10.1016/j.crvi.2010.01.014

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Hofer, R., Paquette, S., et al. (2011). Cytochromes p450. *Arabidopsis Book* 9:e0144.

Bart, A. G., and Scott, E. E. (2017). Structural and functional effects of cytochrome b5 interactions with human cytochrome P450 enzymes. *J. Biol. Chem.* 292, 20818–20833.

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J. D., et al. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* 24, 3106–3118. doi: 10.1105/tpc.112.099796

Bernard, A., and Joubes, J. (2013). Arabidopsis cuticular waxes: advances in synthesis, export and regulation. *Prog. Lipid Res.* 52, 110–129.

Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Leger, A., Roby, D., et al. (2011). Overexpression of Arabidopsis ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiol.* 156, 29–45. doi: 10.1104/pp.111.172320

Broadwater, J. A., Whittle, E., and Shanklin, J. (2002). Desaturation and hydroxylation. Residues 148 and 324 of Arabidopsis FAD2, in addition to substrate chain length, exert a major influence in partitioning of catalytic specificity. *J. Biol. Chem.* 277, 15613–15620. doi: 10.1074/jbc.M200231200

Chang, J., Clay, J. M., and Chang, C. (2014). Association of cytochrome b5 with ETR1 ethylene receptor signaling through RTE1 in Arabidopsis. *Plant J.* 77, 558–567. doi: 10.1111/tpj.12401

De Bigault Du Granrut, A., and Cacas, J. L. (2016). How very-long-chain fatty acids could signal stressful conditions in plants? *Front. Plant Sci.* 7:1490. doi: 10.3389/fpls.2016.01490

de Vetten, N., ter Horst, J., van Schaik, H. P., de Boer, A., Mol, J., and Koes, R. (1999). A cytochrome b5 is required for full activity of flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. *Proc. Natl. Acad. Sci. U. S. A.* 96, 778–783. doi: 10.1073/pnas.96.2.778

Dong, C. H., Rivarola, M., Resnick, J. S., Maggin, B. D., and Chang, C. (2008). Subcellular co-localization of Arabidopsis RTE1 and ETR1 supports a regulatory role for RTE1 in ETR1 ethylene signaling. *Plant J.* 53, 275–286. doi: 10.1111/j.1365-313X.2007.03339.x

Dunn, T. M., Lynch, D. V., Michaelson, L. V., and Napier, J. A. (2004). A post-genomic approach to understanding sphingolipid metabolism in *Arabidopsis thaliana*. *Ann. Bot.* 93, 483–497. doi: 10.1093/aob/mch071

Durr, U. H., Waskell, L., and Ramamoorthy, A. (2007). The cytochromes P450 and b5 and their reductases–promising targets for structural studies by advanced solid-state NMR spectroscopy. *Biochim. Biophys. Acta* 1768, 3235–3259. doi: 10.1016/j.bbamem.2007.08.007

Enoch, H. G., and Strittmatter, P. (1979). Cytochrome b5 reduction by NADPH-cytochrome P-450 reductase. *J. Biol. Chem.* 254, 8976–8981.

Fan, R. C., Peng, C. C., Xu, Y. H., Wang, X. F., Li, Y., Shang, Y., et al. (2009). Apple sucrose transporter SUT1 and sorbitol transporter SOT6 interact with cytochrome b5 to regulate their affinity for substrate sugars. *Plant Physiol.* 150, 1880–1901. doi: 10.1104/pp.109.141374

Fox, B. G., Shanklin, J., Somerville, C., and Munck, E. (1993). Stearoyl-acyl carrier protein delta 9 desaturase from *Ricinus communis* is a diiron-oxo protein. *Proc. Natl. Acad. Sci. U. S. A.* 90, 2486–2490. doi: 10.1073/pnas.90.6.2486

Fukuchi-Mizutani, M., Mizutani, M., Tanaka, Y., Kusumi, T., and Ohta, D. (1999). Microsomal electron transfer in higher plants: cloning and heterologous expression of NADH-cytochrome b5 reductase from Arabidopsis. *Plant Physiol.* 119, 353–362. doi: 10.1104/pp.119.1.353

Gou, M., Yang, X., Zhao, Y., Ran, X., Song, Y., and Liu, C. J. (2019). Cytochrome b5 is an obligate electron shuttle protein for syringyl lignin biosynthesis in Arabidopsis. *Plant Cell* 31, 1344–1366. doi: 10.1105/tpc.18.00778

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333.

Hannemann, F., Bichet, A., Ewen, K. M., and Bernhardt, R. (2007). Cytochrome P450 systems–biological variations of electron transport chains. *Biochim. Biophys. Acta* 1770, 330–344. doi: 10.1016/j.bbagen.2006.07.017

Haslam, T. M., and Feussner, I. (2022). Diversity in sphingolipid metabolism across land plants. *J. Exp. Bot.* 73, 2785–2798.

Haslam, T. M., and Kunst, L. (2013). Extending the story of very-long-chain fatty acid elongation. *Plant Sci.* 210, 93–107. doi: 10.1016/j.plantsci.2013.05.008

Hwang, Y. T., Pelitire, S. M., Henderson, M. P., Andrews, D. W., Dyer, J. M., and Mullen, R. T. (2004). Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria. *Plant Cell* 16, 3002–3019.

Ihara-Ohori, Y., Nagano, M., Muto, S., Uchimiya, H., and Kawai-Yamada, M. (2007). Cell death suppressor Arabidopsis bax inhibitor-1 is associated with calmodulin binding and ion homeostasis. *Plant Physiol.* 143, 650–660. doi: 10.1104/pp.106.090878

Isbat, M., Zeba, N., Kim, S. R., and Hong, C. B. (2009). A BAX inhibitor-1 gene in *Capsicum annuum* is induced under various abiotic stresses and endows multi-tolerance in transgenic tobacco. *J. Plant Physiol.* 166, 1685–1693. doi: 10.1016/j.jplph.2009.04.017

Ishikawa, T., Takahara, K., Hirabayashi, T., Matsumura, H., Fujisawa, S., Terauchi, R., et al. (2010). Metabolome analysis of response to oxidative stress in rice suspension cells overexpressing cell death suppressor Bax inhibitor-1. *Plant Cell Physiol.* 51, 9–20. doi: 10.1093/pcp/pcp162

Ishikawa, T., Watanabe, N., Nagano, M., Kawai-Yamada, M., and Lam, E. (2011). Bax inhibitor-1: a highly conserved endoplasmic reticulum-resident cell death suppressor. *Cell Death Differ.* 18, 1271–1278. doi: 10.1038/cdd.2011.59

Jensen, K., and Moller, B. L. (2010). Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry* 71, 132–141.

Joubes, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., et al. (2008). The VLCFA elongase gene family in *Arabidopsis thaliana*: phylogenetic analysis, 3D modelling and expression profiling. *Plant Mol. Biol.* 67, 547–566. doi: 10.1007/s11103-008-9339-z

Kandel, S. E., and Lampe, J. N. (2014). Role of protein-protein interactions in cytochrome P450-mediated drug metabolism and toxicity. *Chemi. Res. Toxicol.* 27, 1474–1486. doi: 10.1021/tx500203s

Kumar, R., Tran, L. S., Neelakandan, A. K., and Nguyen, H. T. (2012). Higher plant cytochrome b5 polypeptides modulate fatty acid desaturation. *PLoS One* 7:e31370. doi: 10.1371/journal.pone.0031370

Kumar, R., Wallis, J. G., Skidmore, C., and Browse, J. (2006). A mutation in Arabidopsis cytochrome b5 reductase identified by high-throughput screening differentially affects hydroxylation and desaturation. *Plant J.* 48, 920–932. doi: 10.1111/j.1365-313X.2006.02925.x

Lee, M., Lenman, M., Banas, A., Bafor, M., Singh, S., Schweizer, M., et al. (1998). Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science* 280, 915–918. doi: 10.1126/science.280.5365.915

Lee, Y., Yoon, T. H., Lee, J., Jeon, S. Y., Lee, J. H., Lee, M. K., et al. (2018). A lignin molecular brace controls precision processing of cell walls critical for surface integrity in Arabidopsis. *Cell* 173, 1468–1480.e9. doi: 10.1016/j.cell.2018.03.060

Li, Y., Li, L. L., Fan, R. C., Peng, C. C., Sun, H. L., Zhu, S. Y., et al. (2012). Arabidopsis sucrose transporter SUT4 interacts with cytochrome b5-2 to regulate seed germination in response to sucrose and glucose. *Mol. Plant* 5, 1029–1041. doi: 10.1093/mp/sss001

Maggio, C., Barbante, A., Ferro, F., Frigerio, L., and Pedrazzini, E. (2007). Intracellular sorting of the tail-anchored protein cytochrome b5 in plants: a comparative study using different isoforms from rabbit and Arabidopsis. *J. Exp. Bot.* 58, 1365–1379. doi: 10.1093/jxb/erl303

Milkowski, C., and Strack, D. (2010). Sinapate esters in brassicaceous plants: biochemistry, molecular biology, evolution and metabolic engineering. *Planta* 232, 19–35. doi: 10.1007/s00425-010-1168-z

Mitchell, A. G., and Martin, C. E. (1997). Fah1p, a *Saccharomyces cerevisiae* cytochrome b5 fusion protein, and its *Arabidopsis thaliana* homolog that lacks the cytochrome b5 domain both function in the alpha-hydroxylation of sphingolipid-associated very long chain fatty acids. *J. Biol. Chem.* 272, 28281–28288. doi: 10.1074/jbc.272.45.28281

Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant. Biol.* 61, 291–315.

Morgan, E. T., and Coon, M. J. (1984). Effects of cytochrome b5 on cytochrome P-450-catalyzed reactions. Studies with manganese-substituted cytochrome b5. *Drug Metab. Disposit.* 12, 358–364.

Nagano, M., Ihara-Ohori, Y., Imai, H., Inada, N., Fujimoto, M., Tsutsumi, N., et al. (2009). Functional association of cell death suppressor, Arabidopsis Bax

inhibitor-1, with fatty acid 2-hydroxylation through cytochrome b(5). *Plant J.* 58, 122–134. doi: 10.1111/j.1365-313X.2008.03765.x

Nagano, M., Kakuta, C., Fukao, Y., Fujiwara, M., Uchimiya, H., and Kawai-Yamada, M. (2019). Arabidopsis Bax inhibitor-1 interacts with enzymes related to very-long-chain fatty acid synthesis. *J. Plant Res.* 132, 131–143. doi: 10.1007/s10265-018-01081-8

Nagano, M., Takahara, K., Fujimoto, M., Tsutsumi, N., Uchimiya, H., and Kawai-Yamada, M. (2012). Arabidopsis sphingolipid fatty acid 2-hydroxylases (AtFAH1 and AtFAH2) are functionally differentiated in fatty acid 2-hydroxylation and stress responses. *Plant Physiol.* 159, 1138–1148. doi: 10.1104/pp.112.199547

Napier, J. A., Michaelson, L. V., and Sayanova, O. (2003). The role of cytochrome b5 fusion desaturases in the synthesis of polyunsaturated fatty acids. *Prostaglandins Leukot Essent Fatty Acids* 68, 135–143.

Napier, J. A., Sayanova, O., Stobart, A. K., and Shewry, P. R. (1997). A new class of cytochrome b5 fusion proteins. *Biochem. J.* 328(Pt 2), 717–718.

Napier, J. A., Smith, M. A., Stobart, A. K., and Shewry, P. R. (1995). Isolation of a cDNA encoding a cytochrome b5 specifically expressed in developing tobacco seeds. *Planta* 197, 200–202.

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

Oh, C. S., Toke, D. A., Mandala, S., and Martin, C. E. (1997). ELO2 and ELO3, homologues of the *Saccharomyces cerevisiae* ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation. *J. Biol. Chem.* 272, 17376–17384. doi: 10.1074/jbc.272.28.17376

Paddon, C. J., and Keasling, J. D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* 12, 355–367. doi: 10.1038/nrmicro3240

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532. doi: 10.1038/nature12051

Pandey, M. K., Vivekanandan, S., Ahuja, S., Huang, R., Im, S. C., Waskell, L., et al. (2013). Cytochrome-P450-cytochrome-b5 interaction in a membrane environment changes 15N chemical shift anisotropy tensors. *J. Phys. Chem. B* 117, 13851–13860. doi: 10.1021/jp4086206

Pascal, S., Bernard, A., Deslous, P., Gronnier, J., Fournier-Goss, A., Domergue, F., et al. (2019). Arabidopsis CER1-LIKE1 functions in a cuticular very-long-chain alkane-forming complex. *Plant Physiol.* 179, 415–432. doi: 10.1104/pp.18.01075

Porter, T. D. (2002). The roles of cytochrome b5 in cytochrome P450 reactions. *J. Biochem. Mol. Toxicol.* 16, 311–316.

Quist, T. M., Sokolchik, I., Shi, H., Joly, R. J., Bressan, R. A., Maggio, A., et al. (2009). HOS3, an ELO-like gene, inhibits effects of ABA and implicates a S-1-P/ceramide control system for abiotic stress responses in *Arabidopsis thaliana*. *Mol. Plant* 2, 138–151. doi: 10.1093/mp/ssn085

Rahier, A., Smith, M., and Taton, M. (1997). The role of cytochrome b5 in 4alpha-methyl-oxidation and C5(6) desaturation of plant sterol precursors. *Biochem. Biophys. Res. Commun.* 236, 434–437. doi: 10.1006/bbrc.1997.6974

Ramiro, D. A., Melotto-Passarin, D. M., Barbosa Mde, A., Santos, F. D., Gomez, S. G., Massola Junior, N. S., et al. (2016). Expression of Arabidopsis Bax Inhibitor-1 in transgenic sugarcane confers drought tolerance. *Plant Biotechnol. J.* 14, 1826–1837. doi: 10.1111/pbi.12540

Resnick, J. S., Rivarola, M., and Chang, C. (2008). Involvement of RTE1 in conformational changes promoting ETR1 ethylene receptor signaling in Arabidopsis. *Plant J.* 56, 423–431. doi: 10.1111/j.1365-313X.2008.03615.x

Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.

Schenkman, J. B., and Jansson, I. (2003). The many roles of cytochrome b5. *Pharmacol. Ther.* 97, 139–152.

Shirley, A. M., McMichael, C. M., and Chapple, C. (2001). The sng2 mutant of Arabidopsis is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. *Plant J.* 28, 83–94. doi: 10.1046/j.1365-313x.2001.01123.x

Smith, M. A., Cross, A. R., Jones, O. T., Griffiths, W. T., Stymne, S., and Stobart, K. (1990). Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine delta 12-desaturase (delta 12-desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. *Biochem. J.* 272, 23–29. doi: 10.1042/bj2720023

Smith, M. A., Jonsson, L., Stymne, S., and Stobart, K. (1992). Evidence for cytochrome b5 as an electron donor in ricinoleic acid biosynthesis in microsomal preparations from developing castor bean (*Ricinus communis* L.). *Biochem. J.* 287(Pt 1), 141–144. doi: 10.1042/bj2870141

Sundin, L., Vanholme, R., Geerinck, J., Goeminne, G., Hofer, R., Kim, H., et al. (2014). Mutation of the inducible *Arabidopsis thaliana* cytochrome p450 reductase2 alters lignin composition and improves saccharification. *Plant Physiol.* 166, 1956–1971. doi: 10.1104/pp.114.245548

Traka, M., and Mithen, R. (2009). Glucosinolates, isothiocyanates and human health. *Phytochem. Rev.* 8, 269–282.

Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., and Pompon, D. (1997). Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *J. Biol. Chem.* 272, 19176–19186. doi: 10.1074/jbc.272.31.19176

Vergeres, G., Ramsden, J., and Waskell, L. (1995). The carboxyl terminus of the membrane-binding domain of cytochrome b5 spans the bilayer of the endoplasmic reticulum. *J. Biol. Chem.* 270, 3414–3422. doi: 10.1074/jbc.270.7.3414

Vergeres, G., and Waskell, L. (1995). Cytochrome b5, its functions, structure and membrane topology. *Biochimie* 77, 604–620.

Vik, D., Crocoll, C., Andersen, T. G., Burow, M., and Halkier, B. A. (2016). CB5C affects the glucosinolate profile in *Arabidopsis thaliana*. *Plant Signal. Behav.* 11:e1160189. doi: 10.1080/15592324.2016.1160189

Wang, C., Su, X., Sun, M., Zhang, M., Wu, J., Xing, J., et al. (2019). Efficient production of glycyrrhetinic acid in metabolically engineered *Saccharomyces cerevisiae via* an integrated strategy. *Microb. Cell Fact.* 18:95. doi: 10.1186/s12934-019-1138-5

Weng, J. K., and Chapple, C. (2010). The origin and evolution of lignin biosynthesis. *New Phytol.* 187, 273–285.

Weng, J. K., Li, X., Stout, J., and Chapple, C. (2008). Independent origins of syringyl lignin in vascular plants. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7887–7892. doi: 10.1073/pnas.0801696105

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fpls.2022.903793

OPEN ACCESS

EDITED BY
Angelos K. Kanellis,
Aristotle University of Thessaloniki, Greece

REVIEWED BY
Chenggang Liu,
University of North Texas, United States
Barunava Patra,
University of Kentucky, United States
Shasha Li,
Northwest A&F University, China

*CORRESPONDENCE
Alain Goossens
alain.goossens@psb.vib-ugent.be

†PRESENT ADDRESS
Bianca Ribeiro,
Division of Crop Biotechnics, KU Leuven, Leuven, Belgium
Maite Colinas,
Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, Jena, Germany

‡These authors have contributed equally to this work

§Senior author

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

RECEIVED 24 March 2022
ACCEPTED 07 September 2022
PUBLISHED 30 September 2022

CITATION
Ribeiro B, Erffelinck M-L, Lacchini E, Ceulemans E, Colinas M, Williams C, Van Hamme E, De Clercq R, Perassolo M and Goossens A (2022) Interference between ER stress-related bZIP-type and jasmonate-inducible bHLH-type transcription factors in the regulation of triterpene saponin biosynthesis in *Medicago truncatula*. *Front. Plant Sci.* 13:903793. doi: 10.3389/fpls.2022.903793

Interference between ER stress-related bZIP-type and jasmonate-inducible bHLH-type transcription factors in the regulation of triterpene saponin biosynthesis in *Medicago truncatula*

Bianca Ribeiro [1,2†‡], Marie-Laure Erffelinck [1,2‡], Elia Lacchini [1,2‡], Evi Ceulemans [1,2], Maite Colinas [1,2†], Clara Williams [1,2], Evelien Van Hamme [3], Rebecca De Clercq [1,2], Maria Perassolo [1,2,4,5] and Alain Goossens [1,2*§]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, [2]VIB Center for Plant Systems Biology, Ghent, Belgium, [3]VIB Bio Imaging Core, Ghent, Belgium, [4]Cátedra de Biotecnología, Departamento de Microbiología, Inmunología y Biotecnología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina, [5]Instituto de Nanobiotecnología (NANOBIOTEC), Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad de Buenos Aires, Buenos Aires, Argentina

Triterpene saponins (TS) are a structurally diverse group of metabolites that are widely distributed in plants. They primarily serve as defense compounds and their production is often triggered by biotic stresses through signaling cascades that are modulated by phytohormones such as the jasmonates (JA). Two JA-modulated basic helix-loop-helix (bHLH) transcription factors (TFs), triterpene saponin biosynthesis activating regulator 1 (TSAR1) and TSAR2, have previously been identified as direct activators of TS biosynthesis in the model legume *Medicago truncatula*. Here, we report on the involvement of the core endoplasmic reticulum (ER) stress-related basic leucine zipper (bZIP) TFs bZIP17 and bZIP60 in the regulation of TS biosynthesis. Expression and processing of *M. truncatula* bZIP17 and bZIP60 proteins were altered in roots with perturbed TS biosynthesis or treated with JA. Accordingly, such roots displayed an altered ER network structure. *M. truncatula* bZIP17 and bZIP60 proteins were shown to localize in the nucleus and appeared to be capable of interfering with the TSAR-mediated transactivation of TS biosynthesis genes. Furthermore, interference between ER stress-related bZIP and JA-modulated bHLH TFs in the regulation of JA-dependent terpene biosynthetic pathways may be widespread in the plant kingdom, as we demonstrate that it also occurs in the regulation of monoterpene indole alkaloid biosynthesis in the medicinal plant *Catharanthus roseus*.

10.3389/fpls.2022.903793

KEYWORDS

triterpene, saponin, *Medicago catharanthus*, *catharanthus*, jasmonate, endoplasmic reticulum, bZIP, basic helix-loop-helix

Introduction

Plants are continuously challenged with biotic and abiotic stresses. To cope specifically with biotic stresses, such as herbivore feeding or pathogen attack, plants can trigger the biosynthesis of various classes of specialized defense metabolites. A well-known class is that of the structurally and functionally diverse triterpene saponins (TS), which are produced in distinct plant species, including legumes such as *Medicago truncatula*, and which are particularly valuable for pharmaceutical and agrochemical purposes (Chang and Keasling, 2006; Ajikumar et al., 2008; Kumari et al., 2013; Moses et al., 2014a; Netala et al., 2015). Phytohormones play an essential role in the stress-induced elicitation of these compounds, again illustrated by the fact that *M. truncatula* TS production is transcriptionally controlled by a JA signaling cascade (Suzuki et al., 2002; De Geyter et al., 2012; Pollier et al., 2013a; Mertens et al., 2016a; Goossens et al., 2017).

The early committed steps in TS biosynthesis occur mainly at the endoplasmic reticulum (ER) membrane and start from 2,3-oxidosqualene, which is the last common precursor with the phytosterols. Subsequent cyclization of 2,3-oxidosqualene by saponin-specific 2,3-oxidosqualene cyclases (OSC)s, more specifically β-amyrin synthase (BAS) in *M. truncatula*, yields the pentacyclic oleanane-type triterpene backbone β-amyrin (Supplementary Figure 1). Subsequent competitive action of two cytochrome P450-dependent monooxygenases (P450s) results in branching of the *M. truncatula* TS biosynthetic pathway, resulting in the production of two specific classes: the haemolytic and non-haemolytic TS (Gholami et al., 2014). The haemolytic TS branch is defined by three consecutive oxidations at position C-28 of β-amyrin by the P450 CYP716A12, thereby yielding oleanolic acid (Carelli et al., 2011; Fukushima et al., 2011). Positions C-2 and C-23 can be further oxidized by respectively CYP72A67 and CYP72A68v2 (Fukushima et al., 2013; Biazzi et al., 2015). The non-haemolytic branch starts with an oxidation reaction at position C-24 of β-amyrin, catalyzed by CYP93E2, and thereby precluding oxidation at position C-28 (Fukushima et al., 2013; Moses et al., 2014b). Subsequent oxidation at position C-22 by CYP72A61v2 yields soyasapogenol B (Fukushima et al., 2013). UDP-dependent glycosyltransferases (UGTs) can further decorate the triterpene aglycones through attachment of sugar moieties, additionally diversifying the TS compendium (Seki et al., 2015) (Supplementary Figure 1).

The past decades, great progress has been made in the quest for transcription factors (TFs) that are modulated by JA or other cues and that control the production of specialized metabolites in plants (De Geyter et al., 2012; Zhou and Memelink, 2016; Goossens et al., 2017; Colinas and Goossens, 2018; Shoji, 2019). Particularly relevant are the basic helix-loop-helix (bHLH) TFs (Goossens et al., 2017). MYC2 was the first bHLH TF reported to control different branches of terpene biosynthesis in *Arabidopsis thaliana*, *Solanum lycopersicum* and *Artemisia annua*, among others (Hong et al., 2012; Kazan and Manners, 2013; Spyropoulou et al., 2014; Goossens et al., 2017). Later, in the medicinal plant *Catharanthus roseus*, source of the anti-cancer drugs vinblastine and vincristine, both MYC2 as well as MYC2-unrelated bHLH TFs, such as bHLH iridoid synthesis 1 (BIS1) and BIS2, were found to elicit the monoterpenoid branch of the monoterpenoid indole alkaloid (MIA) pathway (Zhang et al., 2011; Van Moerkercke et al., 2015; Van Moerkercke et al., 2016; Goossens et al., 2017; Schweizer et al., 2018; Liu et al., 2021). Likewise, the *M. truncatula* orthologs of the BIS TFs, i.e. triterpene saponin biosynthesis activating regulator 1 (TSAR1) and TSAR2, were reported to transcriptionally regulate the non-haemolytic and haemolytic branch of TS biosynthesis, respectively (Mertens et al., 2016a).

Posttranslational regulatory mechanisms of TS biosynthesis have also been described (Hemmerlin, 2013; Erffelinck and Goossens, 2018). Particularly, the JA-inducible really interesting new gene (RING) membrane-anchor (RMA) E3 ubiquitin ligase makibishi 1 (MKB1) has been reported to control TS biosynthesis in *M. truncatula* by targeting 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), a rate-limiting enzyme in sterol and TS precursor biosynthesis, for degradation by the 26S proteasome (Pollier et al., 2013a). MKB1 forms part of the ER-associated degradation (ERAD) machinery, which monitors the correct folding of membrane and secretory proteins whose biogenesis takes place in the ER. Recently, we also identified a heat shock protein 40 that interacts with MKB1 to support its activity (Erffelinck et al., 2021).When plants are evoked with environmental stresses, a programed defense response is launched, in which the ERAD, the unfolded protein response (UPR) and other ER stress responses play an important role (Malhotra and Kaufman, 2007; Liu and Howell, 2010b).

Eukaryotic cells have developed signaling networks in response to ER stress, through ER stress sensors that are tethered at the ER membrane. In the model plant *A. thaliana*, several ER stress-specific sensors, including the RNase inositol-requiring enzyme 1 (IRE1), the basic leucine zipper (bZIP) TFs bZIP17 and bZIP28, and the NAC TFs NAC062 and NAC089, have been reported (Liu et al., 2007a; Iwata et al., 2008; Liu and Howell, 2010a; Moreno et al., 2012; Yang et al., 2014a; Yang et al., 2014b; Henriquez-Valencia et al., 2015; Kim et al., 2018; Howell, 2021). The primary target of IRE1 in response to ER stress is *bZIP60* mRNA, which is spliced, causing a frame shift and thereby the elimination of the transmembrane domain of the bZIP60 TF at translation. This truncated version is consequently translocated to the nucleus, where it can install a specific ER stress response. An analogous ER-to-nucleus translocation occurs with bZIP17 and bZIP28, but through proteolytic cleavage (Liu et al., 2007a; Liu et al., 2007b; Howell, 2021).

Here, we explored the regulatory interplay between JA and ER stress signaling in the model legume *M. truncatula*. We demonstrate that *M. truncatula* bZIP17 and bZIP60 can interfere with the transactivation of TS-specific gene promoters by the JA-responsive TSAR1 and TSAR2 TFs and thereby modulate the output of the JA response. We also provide evidence that the interplay between these two TF sets may be conserved in the plant kingdom by demonstrating that it also occurs in the regulation of terpene biosynthesis in the distinct plant *C. roseus*.

Materials and methods

DNA constructs

Sequences of the full-length ORFs of *bZIP17* (*Medtr7g088890*) and *bZIP60* (*Medtr1g050502*) were obtained from the *M. truncatula* genome version 4.0 (Tang et al., 2014) and were cloned using Gateway® technology (Invitrogen). Full-length and spliced coding sequences for *bZIP17* and *bZIP60* were PCR amplified (for primers, see Supplementary Table 1) and recombined into the donor vector pDONR221. After sequence verification, the entry clones were recombined with the destination vector p2GW7 for *Nicotiana tabacum* protoplast assays (Vanden Bossche et al., 2013). The promoter regions of *HMGR1*, *CYP93E2*, *HMGR4*, *CYP72A67*, *UGT73F3* and *BAS* recombined with the vector pGWL7 and the full-length coding sequences for *TSAR1* and *TSAR2* recombined with the destination vector p2GW7 had previously been obtained (Mertens et al., 2016a). For the generation of *M. truncatula* hairy roots, sequence-verified entry clones were recombined with the destination vector pK7WG2D for overexpression and pK7GWIWG2(II) for silencing (Karimi et al., 2007). Primers used for cloning of overexpression and silencing constructs and for quantitative reverse transcription PCR (qRT-PCR) analysis are reported in Supplementary Table 1.

The coding sequences of *C. roseus* bZIP17 and bZIP60 were amplified from *C. roseus* var. "Little bright eyes" cDNA with Q5® High-Fidelity DNA Polymerase (New England BioLabs®) and recombined into the entry vector pDONR221 (Gateway®). After sequence verification, the entry clones were recombined with the destination vector p2GW7 (Karimi et al., 2002) for *N. tabacum* protoplast assays (Vanden Bossche et al., 2013). The promoter regions of *GES, G10H* and *IS* recombined with the vector pGWL7 (Karimi et al., 2002) and the full-length coding sequence for *BIS1* recombined with the destination vector p2GW7 (Karimi et al., 2002) had previously been obtained by Van Moerkercke et al. (2015). For expression in flower petals under control of the *CaMV35S* promoter, entry clones were recombined with pK7WG2D (Karimi et al., 2002), using LR Clonase™ enzyme mix (ThermoFisher).

Flower petals of *C. roseus* var. "Little bright eyes" plants (grown under greenhouse conditions) were infiltrated with *Agrobacterium tumefaciens* C58C1 harboring the constructs for overexpression as previously described (Schweizer et al., 2018).

Generation and cultivation of *Medicago truncatula* hairy roots

Sterilization of *M. truncatula* seeds (ecotype Jemalong J5), transformation of seedlings by *A. rhizogenes* (strain LBA 9402/12), and the subsequent generation of hairy roots were carried out as described previously (Pollier et al., 2011). Hairy roots were cultivated for 21 d in liquid medium to provide proper amounts for RNA extraction.

For the elicitation of TS pathway gene expression, 100 μM methyl jasmonate (MeJA) was added to the medium. For the induction of ER stress, 300 mM NaCl, 0.5 mM SA, or 2 mM DTT was added to the medium. Given that we anticipated that the MeJA concentration that we typically employ to elicit TS biosynthesis may be too strong and possibly mimic antagonistic effects of the ER stress agents, we first determined the minimal MeJA concentration with which TS pathway gene expression could still be induced. qRT-PCR analysis of control (CTR) *M. truncatula* hairy root lines indicated that at a concentration of 5 μM MeJA, a pronounced and significant induction of TS pathway gene expression could still be observed (Supplementary Figure 2). This concentration was used for all further experimentation, except for SA, which, as a reported potent JA antagonist of the JA signaling pathway in *A. thaliana* (Van Der Does et al., 2013), was still combined with a MeJA concentration of 100 μM.

Confocal microscopy

Control and MKB1^{KD} (knock-down line of *MKB1*) (Pollier et al., 2013a) hairy roots were cultivated in nutritive liquid medium (Murashige and Skoog with vitamins supplemented

with 1% sucrose) for 2 w and treated with 100 μM MeJA or ethanol (mock treatment) for 24 h. Confocal images (16-bit) were captured with an LSM880 confocal microscope equipped with an Airyscan detector (Zeiss, Jena, Germany). Images were taken in super-resolution, FAST mode by using a Plan-Apochromat 63x/1.4 oil objective (1584 × 1584, pixel size: 43 nm × 43 nm). EGFP was excited using the 488-nm line of an Argon laser (30%) and emission was captured between 495 and 550 nm. Z-sections were made every 185 nm. Images were calculated through pixel reassignment and Wiener filtering by using the built-in "Airyscan Processing" command in the Zen software.

Subcellular localization of the bZIP proteins was determined *via Agrobacterium*-mediated transient expression in *N. benthamiana* epidermal cells. bZIP proteins, either full length or truncated, were fused at the C-terminus of eGFP using the Gateway® vector pB7WGF2. To verify subcellular localization, target proteins were co-transformed together with fluorescent markers for nucleus and ER using pB7m34GW:p35S::NLS-3XCERULEAN and pH7WG2::KDEL-RFP637, respectively. Each vector used was transformed in the *A. tumefaciens* strain C58C1 and a bacterial absorbance A600 nm of 0.8 was used for infiltrating each construct. 72 h after infiltration, leaf sections were collected and fluorescence analyzed by confocal microscopy using a Zeiss LSM710 laser scanner microscope with Plan-Apochromat 20x/0.8 M27. The nuclear marker CFP was excited using a 405-nm laser, while 518-nm and 488-nm Argon lasers were adopted for excitation of the ER RFP marker and GFP::bZIPs fusions, respectively. Z-sections were taken every 2 um. Hairy roots and *N. benthamiana* infiltration images were processed generating maximum intensity projections and adding scale bars using Fiji software.

Phylogenetic analysis

The bZIP proteins from *A. thaliana* and *M. truncatula* were selected based on Dröge-Laser et al. (2018) and Wang et al. (2015), respectively. The amino acid sequences of all selected bZIP proteins were obtained through PLAZA (Van Bel et al., 2018) and the *Catharanthus roseus* Functional Genomics Database (croFGD; http://bioinformatics.cau.edu.cn/croFGD/) (She et al., 2019). These were aligned using MAFFT. The conserved blocks were determined using GBlocks 0.91b and manual curation. IQTREE was used for model selection (Kalyaanamoorthy et al., 2017), after which the best substitution model was selected; the maximum likelihood phylogenetic tree was generated using 1000 bootstrap replicates (Nguyen et al., 2015). The tree figure was made using FigTree software.

RNA-Seq analysis

Total RNA of three independent transformant lines per construct was submitted to VIB Nucleomics Core (VIB, Leuven) for Illumina NextSeq500 RNA sequencing (75 nt, single-end read). As described (Pollier et al., 2013b) and using default parameters, the raw RNA-Seq reads were quality-trimmed and mapped on the *M. truncatula* genome v4.0 (Tang et al., 2014) with TOPHAT v2.0.6. Uniquely mapped reads were counted and FPKM values were determined with CUFFLINKS version v2.2.1 (Trapnell, 2013). Differential expression analyses were performed using Cuffdiff (Trapnell, 2013). RNA-Seq data have been deposited in the ArrayExpress database (accession E-MTAB-11668).

Semi-quantitative qRT-PCR analysis

Frozen hairy roots were ground and the material was used to prepare total RNA and first-strand complementary DNA using the RNeasy Mini Kit (Qiagen) and the iScript cDNA Synthesis Kit (Bio-Rad), respectively, according to each manufacturer's instructions. qRT-PCR primers for *bZIP17* and *bZIP60* were designed using Beacon Designer 4 (Premier Biosoft International) (Supplementary Table 1). The *M. truncatula 40S ribosomal protein S8* and *translation elongation factor 1a* were used as reference genes. The qRT-PCRs were carried out with a LightCycler 480 (Roche) and the LightCycler 480 SYBR Green I Master Kit (Roche) according to the manufacturer's guidelines. Three replicates were made for each reaction and the relative expression levels using multiple reference genes were calculated using qBase (Hellemans et al., 2007).

LC-MS and data analysis

M. truncatula hairy root samples were extracted as described (Pollier et al., 2011; Ribeiro et al., 2020) and subjected to Ultra Performance Liquid Chromatography High Resolution Mass Spectrometry (UPLC-HRMS) at the VIB Metabolomics Core Ghent (VIB-MCG). 10 μl was injected on a Waters Acquity UHPLC device connected to a Vion HDMS Q-TOF mass spectrometer (Waters, Manchester, UK). Chromatographic separation was carried out on an ACQUITY UPLC BEH C18 (150 × 2.1 mm, 1.7 μm) column (Waters, USA); column temperature was maintained at 40°C. A gradient of two buffers was used for separation: buffer A (99:1:0.1 water:acetonitrile: formic acid, pH 3) and buffer B (99:1:0.1 acetonitrile:water: formic acid, pH 3), as follows: 99% A buffer decreased to 50% A from 0 to 30 min, decreased to 30% from 30 to 35 min, and decreased to 0% from 0 to 37 min. The flow rate was set to 0.35 mL min−1. Electrospray ionization (ESI) was applied, LockSpray ion source was operated in negative ionization mode under the following specific conditions: capillary voltage, 2.5 kV; reference capillary voltage, 3 kV; source temperature, 120°C; desolvation gas temperature, 550°C; desolvation gas flow, 800 L h−1; and cone gas flow, 50 L h−1. The collision energy for full MS scan

was set at 6 eV for low energy settings, for high energy settings (HDMSe) it was ramped from 20 to 70 eV. Mass range was set from 120 to 2000 Da, scan time was set at 0.1 s. Nitrogen (greater than 99.5%) was employed as desolvation and cone gas. Leucine-enkephalin (100 pg μL−1 solubilized in water:acetonitrile 1:1 [v/v], with 0.1% formic acid) was used for the lock mass calibration, with scanning every 2 min at a scan time of 0.1 s. Profile data was recorded through Unifi Workstation v2.0 (Waters). Data processing was done with Progenesis QI v2.4 (Waters).

Data were pre-processed by removing features with constant or infinity values across samples. Zero values were replaced by 1.0e-10 for statistical processing. Subsequently, data were transformed using the Arch-sinh function and scaled using the Pareto method. Each condition was analyzed in triplicate on three independent biological replicates for both CTR and bZIP17^{KD} lines. Data were analyzed by two-way ANOVA considering only features with a p value ≤ 0.01 across samples. Ontology for selected candidate TS was determined by comparison with in-house metabolites databases integrated with ChEBI features (Chemical Entities of Biological Interest - EMBL-EBI) as well as by matching features against previously identified TS (Pollier et al., 2011; Ribeiro et al., 2020).

Transient expression assays in protoplasts

Transient expression assays in *N. tabacum* BY-2 protoplasts were carried out as described by Vanden Bossche et al. (2013). Protoplasts were transfected with a reporter, an effector, and a normalizer plasmid. The effector plasmids contained the *TSAR*, *BIS*, or *bZIP* ORFs driven by the *CaMV35S* promoter; the reporter plasmids contained the *FIREFLY LUCIFERASE* (*fLUC*) ORF under control of the target promoters. The normalization plasmid contained the *Renilla luciferase* (*rLUC*) under control of the *CaMV35S* promoter. Protoplasts were incubated overnight and lysed. fLUC and rLUC readouts were collected using the Dual-Luciferase® Reporter Assay System (Promega). Each assay incorporated eight biological repeats. Promoter activities were normalized by dividing the fLUC values with the corresponding rLUC values and the average of the normalized fLUC values was calculated and set out relatively to the control fLUC values, i.e. measured in protoplasts transfected with an effector plasmid carrying a *GUS* control gene.

Results

Confocal imaging exposes an altered ER network structure in *Medicago truncatula* MKB1^{KD} hairy roots

Previously, we have shown that silencing of *MKB1* in *M. truncatula* hairy roots (MKB1^{KD}) results in dissociated roots with caltrop-like structures and a perturbed TS profile, accompanied by a TS-specific negative transcriptional feedback (Pollier et al., 2013a). We hypothesized that perturbed ER functionality could trigger an ER-inherent mechanism to manage ER capacity and integrity, and thereby (in)directly modulate TS metabolism. Therefore, we performed confocal imaging to monitor the ER network structure of *M. truncatula* MKB1^{KD} and CTR hairy roots. Hereby, we exploited the fact that the MKB1^{KD} hairy roots also ectopically express ER-targeted GFP (GFP-KDEL) under the control of a *rolD* promoter, which is used as a visual marker for transformation. The ER network structure was notably altered in MKB1^{KD} hairy roots when compared to CTR lines, for instance exhibiting less of the characteristic three-way junctions (Figure 1). These phenotypical features were even more pronounced in MKB1^{KD} hairy roots that were elicited with MeJA compared to mock-treated MKB1^{KD} hairy roots. Interestingly, MeJA treatment of CTR roots also led to a visual alteration of the ER network structure, but the effect was distinct from or far less pronounced than the effect caused by loss of MKB1 function (Figure 1).

ER stress marker genes are transcriptionally upregulated in MKB1^{KD} hairy roots

Given that MKB1^{KD} hairy roots display an altered morphology with irregular cell shapes and the absence of intercellular spaces (Pollier et al., 2013a), the differences in ER network structure may not necessarily reflect ER stress caused by the perturbed ERAD machinery but rather an intracellular reorganization following the modifications in the cellular structure. Therefore, a transcript profiling study by RNA-sequencing (RNA-Seq) was performed on three independent *M. truncatula* CTR and MKB1^{KD} hairy root lines, either mock- or MeJA-treated. A total of 415,338,234 single-end reads of 50 nt were obtained and mapped on the *M. truncatula* genome version 4.0 (Mt4.0) (Tang et al., 2014). The resulting differential expression profiles were then mined for the closest *M. truncatula* orthologs of a list of known *A. thaliana* ER stress marker genes (Howell, 2013). As such, a set of genes encoding luminal-binding protein 1/2 (BiP1/2; Medtr8g099945), BiP3 (Medtr8g099795), stromal cell derived factor 2 (SDF2; Medtr3g106130), sorbitol dehydrogenase (SDH; Medtr1g025430), calnexin (CNX; Medtr3g098430), UDP-glucose:glycoprotein glucosyltransferase (UGGT; Medtr2g006960), heat-shock protein 70 (HSP70; Medtr3g081170), and protein disulfide isomerase-like 1-1 (PDIL1-1; Medtr3g088220) were found to be significantly upregulated in MKB1^{KD} roots compared to CTR roots, both upon mock and MeJA treatment (Figure 2A). Notably, MeJA elicitation itself was also sufficient to trigger an ER stress response in CTR roots, albeit less pronounced (Figure 2A), in accordance with the moderately altered ER network structure (Figure 1). Together, these data suggest that a transcriptome reminiscent of

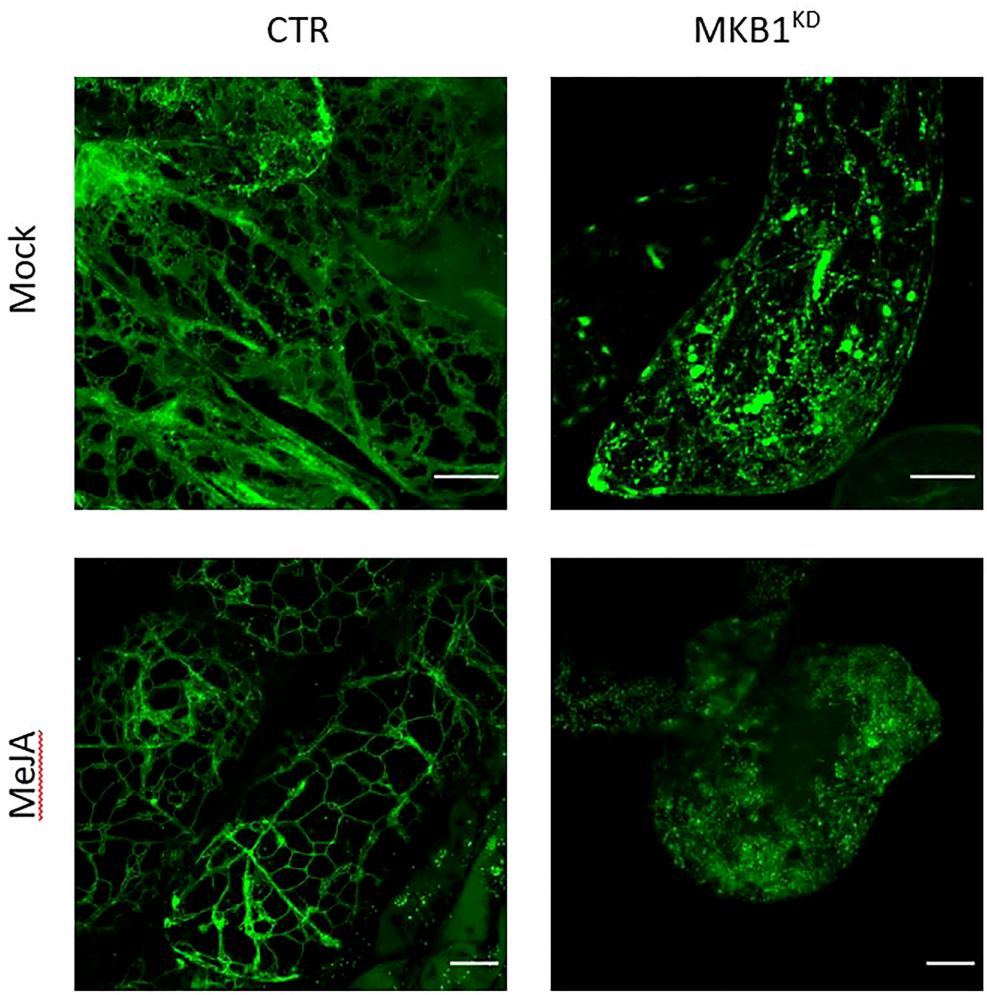


FIGURE 1
Silencing of *MKB1* and MeJA elicitation alters the ER network structure. Shown are maximum intensity projections of images obtained by Airyscan microscopy of ER-targeted GFP in stably transformed CTR and MKB1^{KD} *M. truncatula* hairy roots. Left, CTR roots elicited (24 h) with ethanol (mock) or 100 μM MeJA. Right, mock- and MeJA-treated MKB1^{KD} roots. Scale bars = 20 μm.

an ER stress response is not only triggered by loss of MKB1 function but also by JA elicitation.

We further assessed the transcript levels of the putative *M. truncatula* orthologs of the core *A. thaliana* ER stress-related bZIP TFs, *AtbZIP17*, *AtbZIP28* and *AtbZIP60*. Because the bZIP TF gene family is well represented in the *M. truncatula* genome, with at least 81 potential members (Tang et al., 2014), first a phylogenetic analysis for all potential *M. truncatula* bZIP TF genes that were annotated by Wang et al. (2015) was carried out to define the putative *M. truncatula bZIP17*, *bZIP28* and *bZIP60* orthologs. Amino acid sequences of all *M. truncatula* bZIP gene entries, as well as the previously reported *A. thaliana* bZIP TF genes (Dröge-Laser et al., 2018), were retrieved from PLAZA (Van Bel et al., 2018). As previously reported by Dröge-Laser et al. (2018), *AtbZIP17* is part of the group B that also comprises *AtbZIP28* and *AtbZIP49*, whereas *AtbZIP60* is part of the group K as a unique gene. In the *M. truncatula* genome, the bZIP TFs groups B and K are respectively and solely represented by *Medtr7g088890* and *Medtr1g050502* (Supplementary Figures 3, 4). It therefore appears that *M. truncatula* might not have other paralogs for the bZIP TF genes in either group B and K, or, alternatively, they are not annotated yet by the Mt4.0 genome, as for instance is also the case for the *MKB1* gene. However, our results are in accordance with a genome-wide analysis of the bZIP TF gene family previously carried out for six legume genomes (*Glycine max, Phaseolus vulgaris, Cicer arietinum, Cajanus cajan, Lotus japonicas*, and *M. truncatula*), where also only a single ortholog for both group B and K bZIP TFs was encountered in five of the legumes studied (Wang et al., 2015). *G. max* formed a notable exception, with two paralogs in each

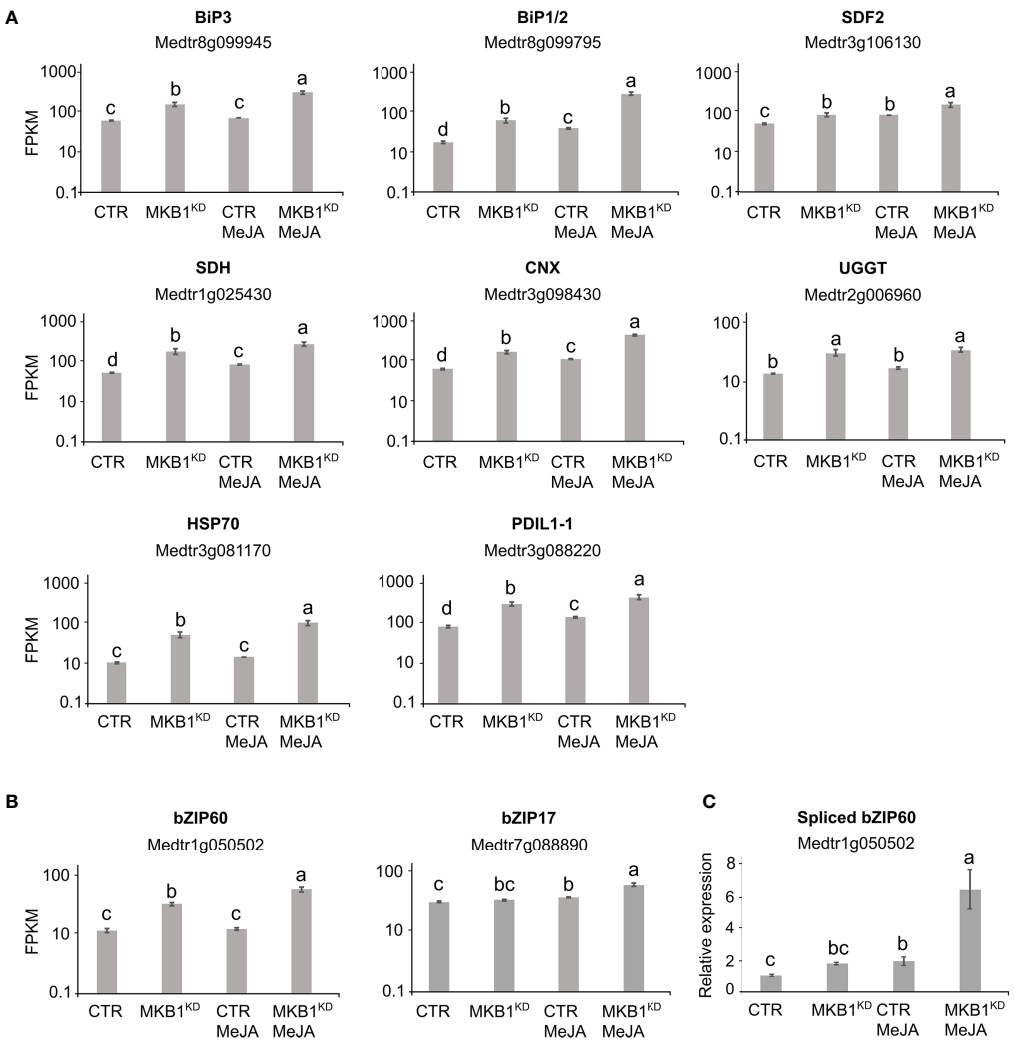


FIGURE 2
Silencing of *MKB1* and MeJA elicitation both trigger an ER stress response in *M. truncatula*. **(A, B)** RNA-Seq analysis of *M. truncatula* orthologs of *A. thaliana* ER stress marker **(A)** and bZIP TF **(B)** genes in CTR and MKB1^{KD} roots, mock- or MeJA-treated for 4 h. The Y-axis represents the normalized fragments per kb of exon per million fragments mapped (FPKM) values. Error bars designate SE (n = 3, except for MeJA-elicited CTR where n = 2). **(C)** qRT-PCR analysis for the detection of spliced *bZIP60* transcripts in CTR and MKB1^{KD} roots, mock- or MeJA-treated for 4 h. The error bars designate SE (n = 3). Different letters indicate statistically significant differences at $P < 0.05$ as determined by ANOVA, *post hoc* Tukey analysis.

group, which may be the consequence of a recent whole-genome duplication that *G. max* experienced (Wang et al., 2015).

Subsequent mining of the RNA-Seq data indicated that *bZIP17* gene transcript levels were only significantly upregulated in MKB1^{KD} hairy roots when elicited with MeJA compared to mock treatment of CTR and MBK1^{KD} (~3.2 fold) (Figure 2B). However, given that activation and translocation of AtbZIP17 occurs posttranslationally following proteolytic cleavage in the Golgi (Liu et al., 2007a; Zhou et al., 2015; Kim et al., 2018), the lack of transcriptional elicitation of *bZIP17* does not exclude that its activity could be enhanced posttranslationally by loss of MKB1 function or by MeJA elicitation. This possibility was not further investigated, given that this would demand extensive additional experimentation and that the results for *bZIP60* were more indicative of the activation of an ER stress response. Indeed, *bZIP60* transcripts accumulated to significantly higher levels in MKB1^{KD} hairy roots as compared to CTR roots, both in mock (2.6 fold) and MeJA (3.9 fold) conditions (Figure 2B). Furthermore, also MeJA treatment could elicit upregulation of *bZIP60* transcript levels particularly in the MKB1^{KD} (2.5 fold) hairy roots (Figure 2B). Contrary to bZIP17, 'activation' of bZIP60 can be assessed at the transcript level, given that it is regulated by IRE1-mediated splicing (Nagashima et al., 2011). To assess the splicing status of *bZIP60*, qRT-PCR was performed using primers

designed to detect the predicted spliced *bZIP60* amplicon. This analysis indicated that the level of spliced *bZIP60* amplicons was increased, both in MKB1^{KD} compared to CTR and following elicitation with MeJA treatment, both in CTR and MKB1^{KD} hairy roots (Figure 2C). Together, our transcriptome analysis supports the occurrence of an increased ER stress response, caused by loss of MKB1 function, and to a minor extent also by MeJA elicitation.

Medicago truncatula bZIP17 and bZIP60 can counteract transactivation of TS biosynthesis promoters by TSAR1 and TSAR2

Next, we hypothesized that the two bZIP17 and bZIP60 TFs could negatively regulate TS biosynthesis, hence explaining the TS-specific negative feedback observed in MKB1^{KD} hairy roots (Pollier et al., 2013a). Localization of some form of the bZIPs in the nucleus is expected in this hypothesis. *In silico* analysis of the full-length *M. truncatula* bZIP60 and bZIP17 sequences confirmed the presence of the expected evolutionary splicing (for bZIP60) or protease cleavage (for bZIP17) sites and the respective ER-anchoring transmembrane domains and nuclear localization signals (Supplementary Figures 5, 6), which is in line with the evolutionary conservation of the UPR pathway (Howell, 2021). We designed N-terminal GFP-tagged versions of the full-length and truncated bZIP17 and bZIP60 proteins and expressed those transiently *via* Agroinfiltration of *N. benthamiana* leaves. Though we could not unambiguously determine a single subcellular localization of either the full-length or truncated variants, nuclear localization of both truncated *M. truncatula* bZIP17 and bZIP60 variants was clearly observed (Supplementary Figure 7).

Next, we tested whether bZIP17 and bZIP60 were able to modulate the previously reported transactivation of a set of TS biosynthesis reporter constructs (promoter-fLUC) by TSAR1 or TSAR2 in a transient expression assay in *N. tabacum* Bright Yellow-2 (BY-2) protoplasts (Mertens et al., 2016a). For any of the tested reporter constructs, no effect of both the full-length bZIP17 and bZIP60 or the truncated versions lacking the transmembrane domain, bZIP17Δ and bZIP60Δ, on reporter fLUC activity was observed (exemplified with *proCYP93E2* in Supplementary Figure 8). However, the high transactivation of *proHMGR1*, *proHMGR4*, *proβAS*, and *proCYP93E2* by *TSAR1*, as compared to the *GUS* control, was significantly repressed when combined with the truncated *bZIP17Δ* or *bZIP60Δ*, but not with full-length *bZIP17* or *bZIP60* (Figure 3A). A similar trend was observed for the transactivation of *proHMGR1*, *proCYP72A67* and *proUGT73F3* by *TSAR2* (Figure 3B). Taken together, these data suggest that the truncated *M. truncatula* ER stress regulatory TFs bZIP17Δ and bZIP60Δ can counteract the transactivation of TS biosynthesis genes in *M. truncatula* by TSAR1 and TSAR2 bHLH factors and may therefore be accountable for the TS-specific transcriptional feedback observed in MKB^{KD} hairy roots (Pollier et al., 2013a).

Functional characterization of *Medicago truncatula* bZIP17 and bZIP60 *in planta*

To further evaluate the *in planta* role of bZIP17/bZIP60 in *M. truncatula*, we generated gain- and loss-of-function hairy root lines for the *bZIP17/bZIP60* genes. Three independently generated root lines expressing the *GUS* gene were used as the control. For *bZIP17*, we generated three independent *M. truncatula* hairy root lines overexpressing *bZIP17Δ* (bZIP17Δ-OE, Supplementary Figure 9A). Quantitative reverse transcription-PCR (qRT-PCR) analysis confirmed overexpression of *bZIP17Δ*, though this seemingly did not increase total *bZIP17* transcript levels in an appreciable manner (Supplementary Figure 9B). An average fourfold increase of *bZIP60* transcript levels was also observed in those lines. Accordingly, the expression level of the chaperone *BiP1/2* was significantly increased in the bZIP17Δ-OE hairy root lines, in line with previous observations in other species such as *A. thaliana* (Li et al., 2017) and maize (Yang et al., 2013). Only two of the nine tested TS biosynthesis genes, i.e. *CYP716A2* and *UGT73F3*, did show significant differential expression in the three bZIP17Δ-OE lines compared to the CTR (Supplementary Figure 9C). Likewise, constitutive ectopic *bZIP17Δ* overexpression did not appear to comprehensively affect MeJA induction of TS biosynthesis genes, because a significant decrease in the MeJA response was only observed for two of the eight tested TS biosynthesis genes, i.e. *BAS* and *CYP716A2* (Supplementary Figure 10). Unfortunately, we did not manage to generate lines overexpressing bZIP60Δ, despite several transformation rounds.

We also managed to generate three independent *bZIP17* and two independent *bZIP60* knock-down lines (bZIP17^{KD} and bZIP60^{KD}), all showing approximately a fourfold reduction in *bZIP* expression (Figure 4B and Supplementary Figure 11B). A notable growth phenotype with a callus-like morphology was observed, especially for bZIP60^{KD} hairy root lines (Figure 4A and Supplementary Figure 11A), but which is distinct from the caltrop-like MKB1^{KD} phenotype. Furthermore, *bZIP60* transcript levels were significantly increased in the bZIP17^{KD} lines (Figure 4B). Conversely, no feedback on *bZIP17* expression was observed in the bZIP60^{KD} hairy root lines (Supplementary Figure 11B). Importantly however, the transcript levels of all analyzed TS biosynthesis genes, except those of *TSAR1*, were slightly, but significantly, increased in the bZIP17^{KD} lines (Figure 4C). These observations were corroborated with metabolite profiling by liquid chromatography – mass spectrometry (LC-MS) analysis, which demonstrated significantly increased accumulation levels of several of the measured TS in the bZIP17^{KD} lines, both in mock- and MeJA-

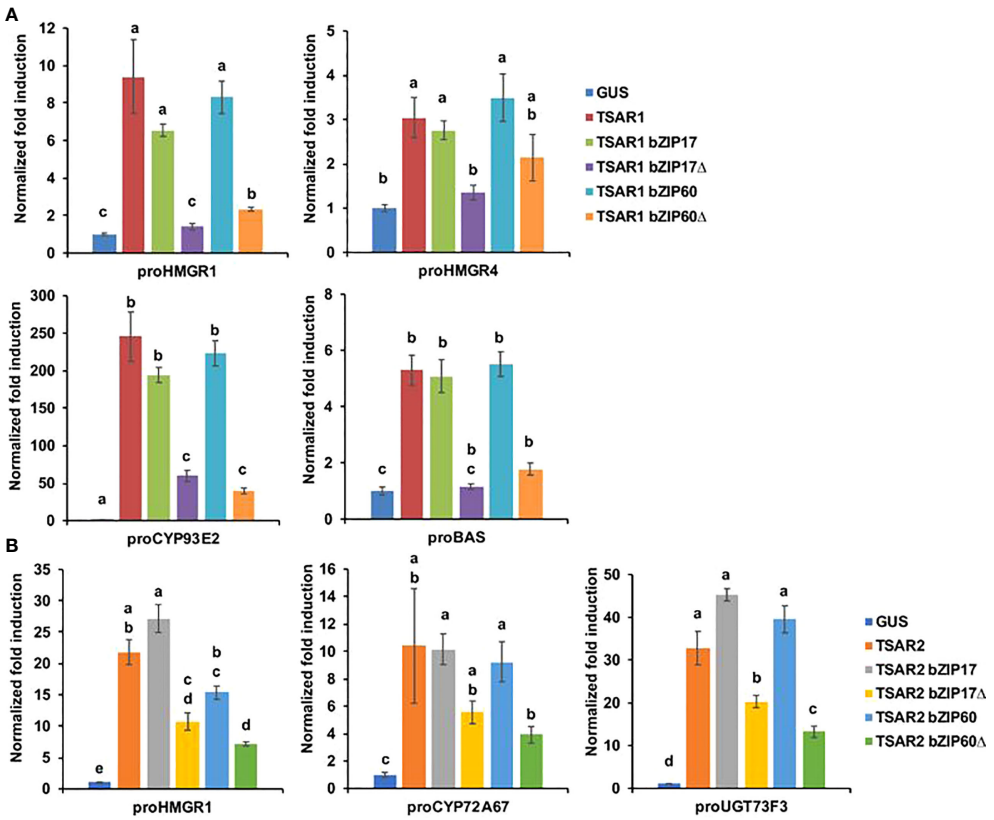


FIGURE 3
M. truncatula bZIP17Δ and *bZIP60Δ* repress transactivation of TS biosynthesis gene promoters by TSAR TFs. **(A, B)** Transient transactivation assays in BY-2 protoplasts using the indicated target promoters fused to the *fLUC* reporter gene and TSAR1 **(A)** or TSAR2 **(B)** as effectors combined with full-length or truncated bZIP17 and bZIP60. The Y-axis shows fold change in normalized fLUC activity relative to the control transfection with *proCaMV35S:GUS*. The error bars designate SE of the mean (n = 8 biological repeats). Different letters indicate statistically significant differences at $P < 0.05$, as determined by ANOVA, *post hoc* Tukey analysis.

treated conditions (Figure 5). In the bZIP17^{KD} hairy root lines, the effect was less consistent, with only a slight but significant increase in the expression level of *CYP716A12* and *UGT73F3* (Supplementary Figure 11C). Unfortunately, we did not manage to generate hairy root lines silencing both *bZIP17* and *bZIP60*, despite several transformation attempts. It is plausible to assume that because of the crucial roles of these bZIP factors for plant physiology, simultaneous loss-of-function of both is not viable. Nonetheless, the data obtained with the *M. truncatula* bZIP17^{KD} lines support the role of at least bZIP17 as a negative attenuator of the TS pathway.

ER stress inducers repress transcript levels of TS pathway genes

Often, TFs that regulate specialized metabolite biosynthetic pathways are coexpressed with the target genes that encode the enzymes of the pathways, particularly when it concerns JA-modulated pathways (Pauwels et al., 2009; De Geyter et al., 2012). In order to find plant growth or stress conditions in which *M. truncatula bZIP17* and *bZIP60* might show an opposite or correlative expression pattern to those of the TS pathway genes, we mined the *M. truncatula* Gene Expression Atlas (MtGEA; http://bioinfo.noble.org/gene-atlas/) (He et al., 2009). We particularly looked for conditions in which increased *bZIP17* and *bZIP60* expression was observed, in combination with a modulated TS gene expression pattern. Such a situation was encountered for *bZIP17* in roots of *M. truncatula* seedlings grown in the presence of 180 mM NaCl (Supplementary Figure 12). This was not unexpected, given that in *A. thaliana*, salt stress was reported to invoke ER stress, for which the action of bZIP17 is needed for the stress coping mechanism (Li et al., 2017). This observation suggested that *in planta* situations, in which altered bZIP60 and bZIP17 activity may modulate TS gene expression, can indeed be encountered.

As far as we could judge, the MtGEA did not seem to contain transcriptome data of other stress or growth conditions with a

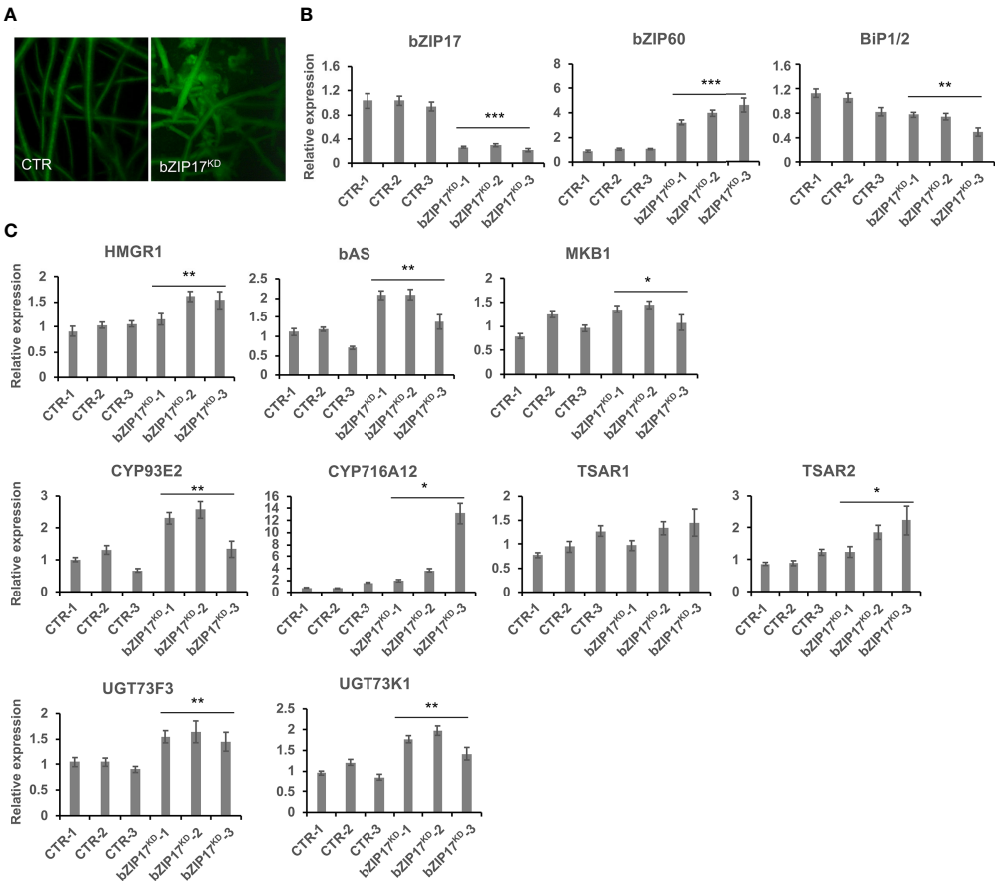


FIGURE 4
Silencing of *bZIP17* slightly increases TS biosynthesis gene expression in *M. truncatula* hairy roots. **(A)** Morphology of control (CTR) and bZIP17^{KD} hairy roots. **(B, C)** qRT-PCR analysis of *bZIP17*, *bZIP60*, *BiP1/2* genes **(B)** and TS biosynthetic genes **(C)** in three independent CTR and *bZIP17^{KD}* hairy root lines. Values in the y-axis represent the expression ratio relative to the normalized transcript levels of CTR lines. The error bars designate SE (n = 3, technical repeats). Statistical significance was determined by a Student's *t*-test (*P < 0.05, **P < 0.01, ***P < 0.001).

pronounced effect on *bZIP17/bZIP60* expression or a reported ER stress effect. Because in *A. thaliana*, the reducing agent dithiothreitol (DTT) and the stress hormone salicylic acid (SA) are known to evoke ER stress and to induce the upregulation and activation of *bZIP17* and *bZIP60* (Moreno et al., 2012; Henriquez-Valencia et al., 2015; Li et al., 2017), we decided to analyze *bZIP*, ER stress and TS pathway gene expression in the presence of NaCl, DTT, SA, all in combination or not with MeJA, in CTR *M. truncatula* hairy root lines.

As expected, *bZIP17* and *bZIP60* transcript levels were increased upon NaCl and DTT treatment in the CTR *M. truncatula* hairy roots (Supplementary Figures 13A, B). Upon SA treatment, only *bZIP60* transcript levels were significantly increased (Supplementary Figure 13C). In all cases, increased *BiP1/2* transcript levels were further indicative of successful ER stress induction by all three stress agents. In most cases, combined application with MeJA aggravated the ER stress as reflected by a further increase in the ER stress gene transcript levels (Supplementary Figure 13). Next, we assessed TS pathway gene expression in the CTR line upon different stress treatments. NaCl treatment did not affect the basal expression of TS pathway genes, nor did it interfere with the MeJA elicitation thereof (Figure 6A). Different and more interesting trends were observed with the DTT and SA treatments. First, SA had a pronounced inhibitory effect on the MeJA induction of all TS pathway genes tested (Figure 6B). Whether this is partly or entirely mediated by increased bZIP activity cannot be judged at this stage however, given that in *A. thaliana* other TFs such as ORA59 or the TF cofactor NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1) have also been implicated in the SA-mediated suppression of JA signaling (Van Der Does et al., 2013; Nomoto et al., 2021). In this regard, the results obtained with the more specific ER stress agent DTT may be considered more indicative. Indeed, also DTT application significantly suppressed the MeJA elicitation of half of the tested TS pathway genes (Figure 6C). To assess whether the DTT-

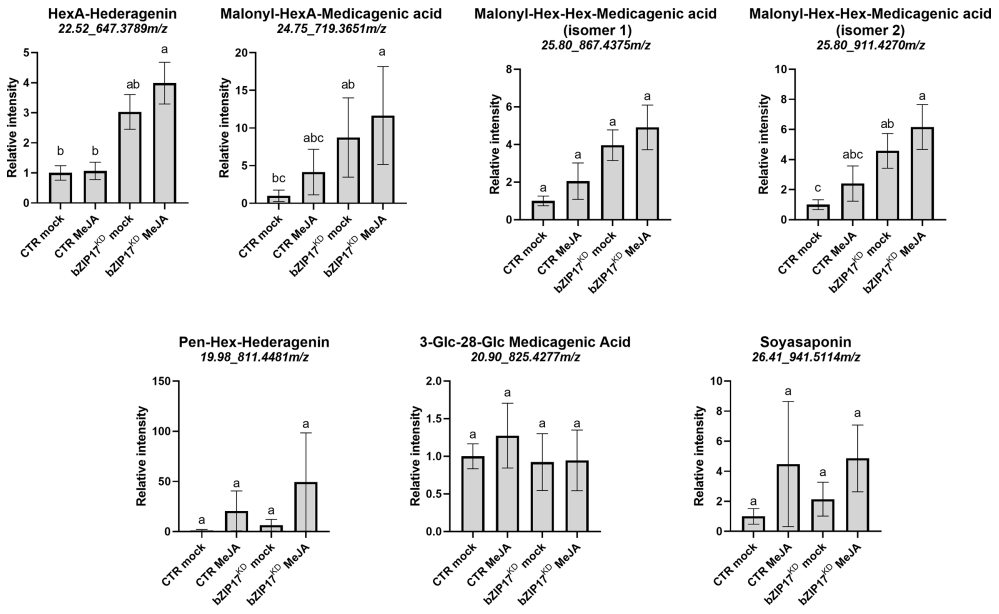


FIGURE 5
Silencing of *bZIP17* enhances the accumulation of TS in *M. truncatula* hairy roots. Relative accumulation of TS in mock- and MeJA-treated control (CTR) and bZIP17^{KD} hairy roots. Values on the y axis correspond to fold changes relative to the average of the mock-treated control lines. Intensity values were transformed, scaled and analyzed by two-way ANOVA ($p \leq 0.01$), categories refer to Tukey's *post-hoc* test ($p < 0.05$). For each of the TS, the identity, retention time and m/z value is indicated on top. The small letters represent the categories that refer to Tukey's post-hoc test ($p < 0.05$). They indicate significant differences between treatments.

suppressive effect was specific for MeJA elicitation of TS pathway genes, we also assayed expression of known early response JA genes involved in the JA amplification loop (Pauwels et al., 2009). Notably, the transcript levels of *JAZ1*, *LOX*, *MYC2a* and *MYC2b* were significantly increased by DTT treatment (Supplementary Figure 14). Moreover, DTT boosted MeJA elicitation of those JA pathway genes. The latter observations can likely be explained by the reported need for a reducing environment for the JA-Ile-induced interaction between COI1 and JAZ1 (Yan et al., 2009). As such, overall, our qRT-PCR analyses indicated that the negative effect of DTT on MeJA elicitation is specific for TS pathway genes.

Finally, we assessed the effect of combined DTT and MeJA application also in the *bZIP17^{KD}* and *bZIP60^{KD}* hairy root lines. However, no consistent or pronounced differences could be detected with regard to the antagonistic effect that was observed in the CTR line (Figures 6D, E). Only MeJA elicitation of *CYP716A12* and *TSAR2* was less pronounced in both knock-down lines, and the suppressive effect of DTT on MeJA elicitation was only attenuated for *CYP716A12*. We assume that either the MeJA effect is capable of overruling the ER stress, or alternatively that the remaining active *bZIP* gene in the knock-down lines can still account for the ER stress effect. Nonetheless, taken together, our data suggest that also *in planta*, ER stress can attenuate the JA response in *M. truncatula* roots, or at least part of it, such as the elicitation of the TS pathway.

Catharanthus roseus bZIP17 and bZIP60 counteract transactivation of monoterpenoid indole alkaloid biosynthesis gene promoters by BIS1

Since the ER stress response is a conserved mechanism in plants and many specialized metabolite biosynthesis pathways are regulated by TFs that bind G-boxes or closely related boxes, we hypothesized that the action of ER stress bZIP factors on the control of specialized metabolite pathways could be conserved across plant species. An obvious model system to explore this is the medicinal plant *C. roseus*, in which different branches of the MIA pathway are controlled by bHLH factors such as the BISs, which are functional orthologs of the TSARs (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016; Mertens et al., 2016b; Schweizer et al., 2018).

To determine the putative orthologs of the *bZIP17* and *bZIP60* genes in *C. roseus*, a BLAST analysis was performed using the AtbZIP17 and AtbZIP60 amino acid sequences as query in the *C. roseus* Functional Genomics Database (croFGD; http://bioinformatics.cau.edu.cn/croFGD/) and Medicinal Plant Genomics Resource (http://medicinalplantgenomics.msu.edu/). The highest significant hits were *CROT021933* and *CROT026761*, respectively, which were confirmed to belong to the bZIP TFs group B and K, respectively, by phylogenetic analysis (Supplementary Figure 15). Moreover, *CROT021933*

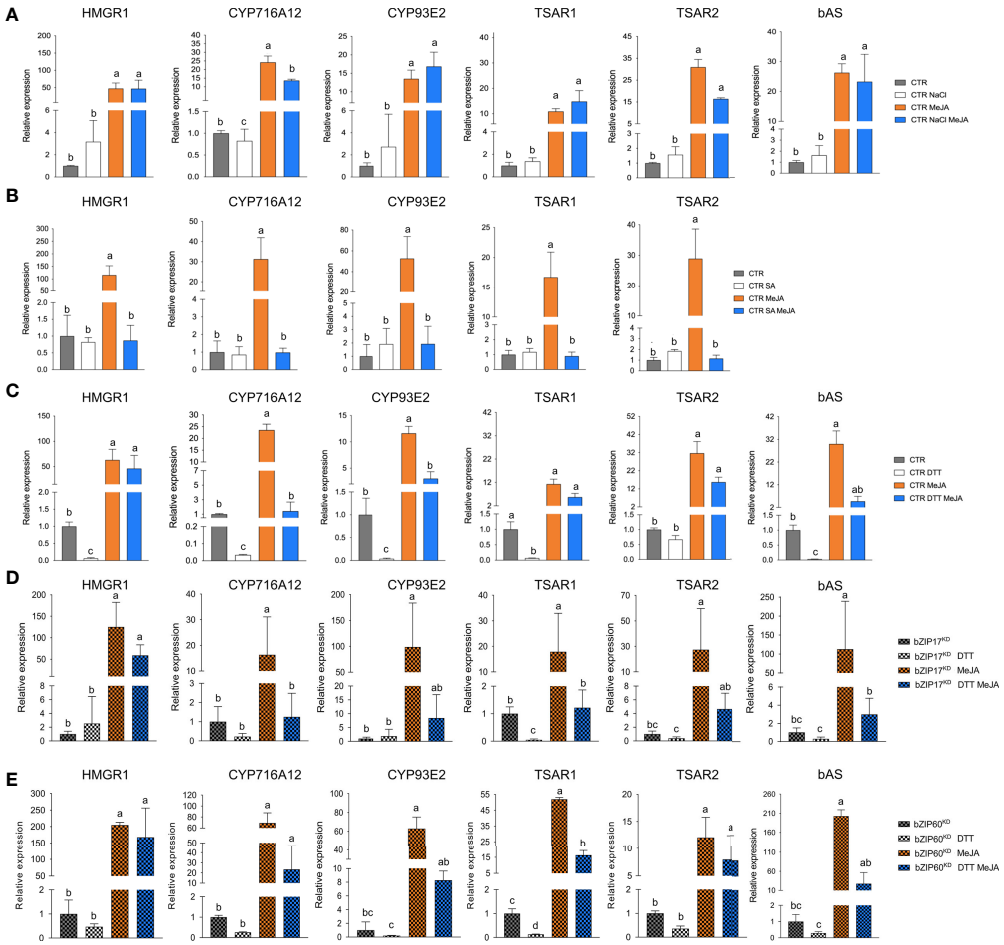


FIGURE 6
Effects of ER stress on the expression of TS pathway genes in *M. truncatula* hairy roots. **(A-E)** qRT-PCR analysis of TS pathway genes in control (CTR) **(A-C)**, *bZIP17*^{KD} **(D)** and *bZIP60*^{KD} **(E)** hairy root lines. Hairy roots were treated with 300 mM NaCl **(A)**, 0.5 mM SA **(B)** or 2 mM DTT **(C-E)** for 4 h prior to the 4-h MeJA treatment (5 μM for **(A, C-E)** and 100 μM for **(B)**). Expression ratios were plotted relative to the normalized mock-treated line. The error bars designate SE (n = 3 for **(A-D)** and n = 2 for **(E)**, corresponding to the independent transformed CTR, *bZIP17*^{KD} and *bZIP60*^{KD} lines generated). Different letters sample indicate statistically significant differences at P < 0.05, as determined by ANOVA, *post hoc* Tukey analysis.

and *CROT026761* shared the same conserved protein sequence motifs with their bZIP TF group members (Supplementary Figures 15, 16).

Next, a transient expression assay was performed in BY2 protoplasts to assess transactivation of a set of promoters from the genes encoding enzymes of the *C. roseus* MIA biosynthesis pathway that are known to be controlled by the BIS TFs (Van Moerkercke et al., 2015; Mertens et al., 2016b; Van Moerkercke et al., 2016; Schweizer et al., 2018), including *geraniol 8-oxidase* (*G8O*), *geraniol synthase* (*GES*) and *iridoid synthase* (*IS*), by the *C. roseus* BIS1 TF in combination with *C. roseus* CrbZIP17, CrbZIP60 or the truncated versions thereof (CrbZIP17Δ and CrbZIP60Δ). And indeed, the transactivation mediated by BIS1 of *proG8O* and *proIS* was compromised in the presence of the truncated *CrbZIP17Δ* and *CrbZIP60Δ*, but not of the intact *CrbZIP17* and *CrbZIP60* (Figure 7A). In the case of *proGES*, counteraction of BIS1-mediated transactivation was only observed with *CrbZIP17Δ* (Figure 7A). Together, these data suggest that translocated ER stress response bZIP TFs can counteract the BIS1-mediated transcriptional activation of MIA biosynthesis genes in *C. roseus*.

To further corroborate this, we also assessed the effect of the *C. roseus* bZIPs in *C. roseus in planta*, in particular through an *A. tumefaciens*-assisted *C. roseus* flower infiltration platform that was previously successfully used as an expression system to (co-) express TF(s) and thereby screen for novel MIA biosynthesis regulators (Schweizer et al., 2018). Here, we transiently transformed *C. roseus* flowers with overexpression cassettes for

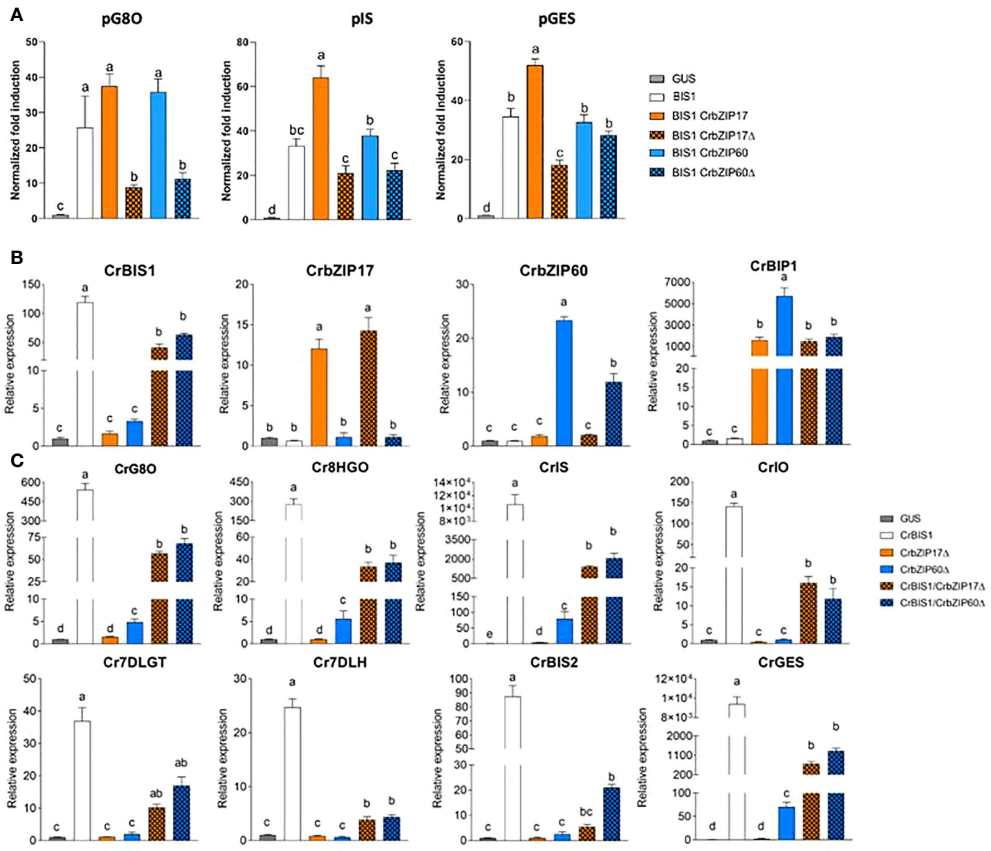


FIGURE 7
C. roseus bZIP17Δ and bZIP60Δ can repress the transactivation of MIA biosynthesis genes by BIS1. **(A)** Transient transactivation assays in BY-2 protoplasts using the indicated target promoters fused to the *fLUC* reporter gene and BIS1 as effector combined with full-length or truncated *C. roseus* bZIP17 and bZIP60. The Y-axis shows fold change in normalized fLUC activity relative to the control transfection with *proCaMV35S:GUS*. The error bars designate SE of the mean (n = 4). Different letters indicate statistically significant differences at $P < 0.05$ as determined by ANOVA, *post hoc* Tukey analysis. **(B, C)** qRT-PCR analysis of *CrBIS1, CrbZIP17, CrbZIP60* and *CrBiP1* genes **(B)** and of MIA pathway genes **(C)** in *C. roseus* flower petals transiently overexpressing *CrBIS1, CrbZIP17Δ, CrbZIP60Δ, CrBIS1/CrbZIP17Δ* or *CrBIS1/CrbZIP60Δ* under control of the *CaMV35S* promoter. Control samples were infiltrated with the *pCaMV35S:GUS* construct. The error bars designate SE of the mean (n = 4). Different letters indicate statistically significant differences at $P < 0.05$, as determined by ANOVA, *post hoc* Tukey analysis.

BIS1, *CrbZIP17Δ*, and *CrbZIP60Δ*, as well as the double combinations *BIS1/CrbZIP17Δ and BIS1/CrbZIP60Δ*. qRT-PCR analysis confirmed overexpression of all TF genes (Figure 7B). In support of the functionality of the bZIP TFs, the expression level of the chaperone *BiP1* was significantly increased in all the bZIP17Δ-OE and bZIP60Δ-OE samples (Figure 7B). Likewise, in line with previous observations (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016; Schweizer et al., 2018), overexpression of *BIS1* strongly upregulated all known BIS targets of the MIA pathway genes tested (Figure 7C). Finally, as anticipated, the combinatorial overexpression of *CrbZIP17Δ* or *CrbZIP60Δ* could significantly counteract the BIS1-mediated transcriptional activation of all MIA pathway genes tested (Figure 7C). Taken together, our findings suggest that interference by ER stress bZIP TFs in the attenuation of JA-dependent terpene biosynthetic pathways might be widespread in the plant kingdom.

Discussion

The ERAD-type E3 ubiquitin ligase MKB1 has previously been reported to manage TS biosynthesis in *M. truncatula* by controlling HMGR stability (Pollier et al., 2013a). Silencing of *MKB1* in *M. truncatula* MKB1^{KD} hairy roots results in an aberrant caltrop-like morphology, increased accumulation of monoglycosylated TS, decreased accumulation of higher glycosylated TS, and a specific downregulation of TS biosynthesis gene expression. Pollier et al. (2013a) speculated that this TS-specific transcriptional response might constitute a negative feedback loop to cope with the ectopic accumulation of bioactive monoglycosylated saponins.

Intrigued by this anomaly, we explored the MKB1^{KD} hairy root phenotype by additional microscopic and transcriptomic analyses. These combined analyses pointed to an ER stress response, reflected

by an altered ER network structure and increased transcript levels of *M. truncatula* orthologs of known *A. thaliana* ER stress marker genes. Because in *A. thaliana*, and conceivably plants in general, the ER stress response can be mediated by two signaling branches, both depending on bZIP TFs, AtbZIP17/AtbZIP28 and AtbZIP60, respectively, that translocate from the ER to the nucleus in ER stress conditions, we speculated that the TS-specific feedback in MKB1^{KD} hairy roots may be mediated by the action of the *M. truncatula* orthologs of these bZIP TFs. Indeed, subsequent functional analysis confirmed that the truncated versions of *M. truncatula* bZIP60 and bZIP17 can localize to the nucleus and interfere with TSAR1/TSAR2-mediated transactivation of TS gene promoters. We therefore speculate that this mechanism could be imposed by plants to attenuate or fine-tune the biosynthesis of TS under particular stress conditions.

What is the molecular mode of action by which the bZIP TFs repress TSAR-induced TS gene expression?

It remains to be determined how the bZIP and TSAR TFs interact to modulate TS biosynthesis. It has been demonstrated in *A. thaliana* that under non-stressed conditions, the activity of bZIP28 is inhibited by elongated hypocotyl 5 (HY5), another bZIP TF (Nawkar et al., 2017), by competition for binding to the G-box element (CACGTG) displayed within the ER stress response element (ERSE) motifs in the promoters of the UPR genes. Under ER stress conditions, HY5 undergoes proteasomal degradation, releasing the competition and enabling bZIP28 to bind to the ERSE motifs and activate the UPR (Nawkar et al., 2017). A similar scenario has been reported for JA-inducible genes by Van Der Does et al. (2013) in *A. thaliana*, where SA can suppress JA signaling downstream of the JA receptor by targeting GCC promoter motifs *via* the TF octadecanoid-responsive Arabidopsis AP2/ERF domain protein 59 (ORA 59). In *C. roseus*, the production of MIAs is regulated by the bHLH TF CrMYC2 (Zhang et al., 2011; Paul et al., 2017; Schweizer et al., 2018). Overexpression of *CrMYC2* induces expression of the genes encoding bZIP G-box binding factors (GBFs), resulting in a reduced alkaloid accumulation in *C. roseus* hairy roots (Sui et al., 2018). Given that CrGBF1 can bind the same *cis*-element (T/G-box) as CrMYC2 in MIA biosynthesis gene promoters and that CrGBFs can dimerize with CrMYC2, it has been suggested that CrGBF TFs can antagonize CrMYC2 by competitive binding to the T/G-box, and/or by forming a heterodimeric complex, preventing CrMYC2 from binding its target promoters (Sui et al., 2018). Accordingly, we hypothesize a mechanism in *M. truncatula*, in which induction of TS biosynthesis genes by the bHLH TFs TSAR1 and TSAR2 could be antagonized by bZIP17 and bZIP60, plausibly either by competitive binding to the promoters or by the formation of a protein complex that would impede TSAR TF transcriptional activity.

ERSE-like *cis*-elements, the targets of the ER stress response bZIPs, are present in several of the TS gene promoters. Notably, in the minimal *CYP93E2* promoter region that contains N-box motifs (5′-CACGAG-3′) that are necessary and sufficient for TSAR-mediated transactivation (Mertens et al., 2016a), this ERSE-like box (5′-ATTCGACCACG-3′) overlaps with one of the N-boxes, making this promoter a plausible target for inhibitory crosstalk. However, given that partial substitution of the ERSE-like motif would also affect the N-box, and that the latter reduces the capacity of TSAR1 to transactivate this promoter fragment (Mertens et al., 2016a), we could not unambiguously assess this hypothesis. Likewise, we pursued several protein–DNA and protein–protein interaction methods, including yeast one-hybrid to investigate interactions of bZIPs with TS gene promoters, and yeast two-hybrid and bimolecular fluorescence complementation in agro-infiltrated *N. benthamiana* leaves, to assess possible TSAR–bZIP protein interactions, but none of these assays yielded conclusive results. Hence, our model on how ER stress-related bZIPs hinder TSAR1 transcriptional activity, and, by consequence, transactivation of TS biosynthesis genes (Figure 8), remains speculative with regard to the exact *modus operandi*.

Physiological relevance and evolutionary conservation of the suppressive effect of ER stress-induced bZIP factors on JA-inducible and bHLH TF-mediated elicitation of terpene biosynthesis in plants

The molecular mechanisms of ER stress signaling have been well studied over the last years, at least in *A. thaliana*. Many studies made use of compounds such as tunicamycin, which blocks N-linked glycosylation, or the reducing agent DTT, as artificial ER stress inducers. Exploiting such ER stress agents, in particular DTT, we could further support the inverse correlation between TS expression and levels of ER stress, as DTT treatment repressed the expression, particularly in mock conditions and to a lesser extent in the presence of MeJA, of several TS genes encoding both pathway enzymes and the regulatory TSARs themselves in *M. truncatula* hairy root lines.

There is also increasing knowledge of the physiological conditions upon which ER stress is induced ‘naturally’ in plants. For instance, salt and heat stress have been shown to activate an overall ER stress response in *A. thaliana* (Gao et al., 2008; Deng et al., 2011; Tang et al., 2012; Li et al., 2017). Furthermore, also SA was reported to induce both IRE1/bZIP60 and bZIP17/bZIP28 branches of ER stress signaling in *A. thaliana* (Nagashima et al., 2014), although how is still under debate. It has been reported that SA treatment activates phosphatidylinositol 4-kinase (PI4K), thereby increasing PI phosphates (PIPs) in the Golgi membrane (Krinke et al., 2007). Contrarily, inhibition of PI4K activity leads

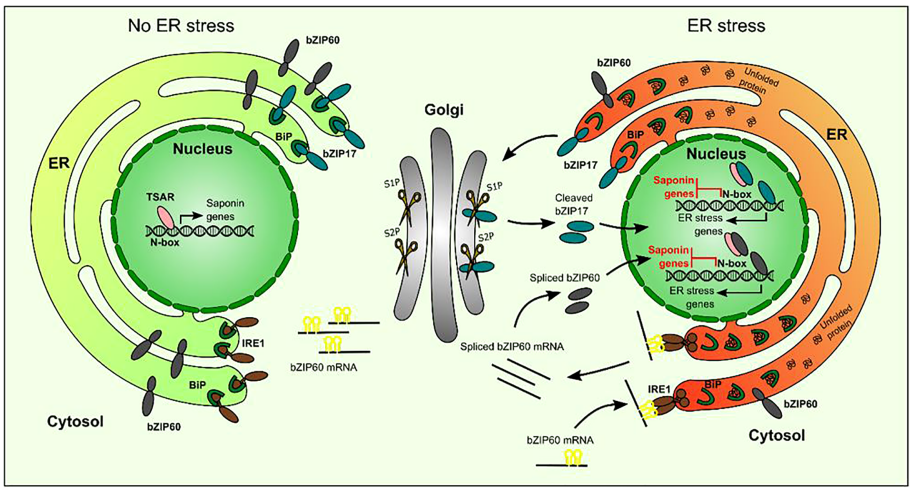


FIGURE 8
Hypothetical model of action of bZIP17 and bZIP60 and TSAR under normal and ER stress conditions. Under non-ER stress conditions, TSAR can bind to the N-box present in the promoters of TS biosynthesis genes and activate their expression. This can occur both in the absence and presence of JAs. Under ER stress conditions, unfolded proteins accumulate and bind to BiP in the ER. This triggers IRE1 to perform unconventional splicing of *bZIP60* mRNA as well as the release of bZIP17 from the ER and its cleavage by proteases in the Golgi apparatus. The resulting active forms of bZIP17 and bZIP60 translocate to the nucleus, where they can transactivate the expression of ER stress-responsive genes but also hinder the TSARs and thereby repress the transactivation of TS biosynthesis gene expression. TSAR, triterpene saponin biosynthesis activating regulator; bZIP, basic leucine zipper TF; BiP, binding protein chaperone; IRE1, inositol-requiring enzyme 1; S1P, site-1 protease; S2P, site-2 protease.

to a decrease in PIPs and inhibition of *BIP3* induction upon SA treatment (Krinke et al., 2007). Krinke et al. (2007) speculated that upon SA treatment, the phospholipid content in the ER membrane system and traffic within is changed, thereby activating ER-localized ER stress sensors in *A. thaliana*. During SA defense, the transcriptional cofactor NPR1 is converted from an oligomer into a monomer, which leads to the expression of *pathogenesis-related protein 1* (*PR1*) gene family members (Wang et al., 2005). Recently, NPR1 has also been shown to be able to interact with bZIP28 and bZIP60 and to suppress the UPR, independently from SA (Lai et al., 2018). Furthermore, Meng et al. (2017) showed that the constitutive expresser of pathogenesis-related genes 5 (CPR5), a negative modulator of SA, inhibits both the SA-dependent IRE1/bZIP60 and the ER stress-induced bZIP28/IRE1–bZIP60 branches, favoring the growth of plants. Many studies reported both on antagonistic (Rojo et al., 2003; Bostock, 2005; Beckers and Spoel, 2006) and synergistic interactions between SA and JA in plants (Schenk et al., 2000; Van Wees et al., 2000; Mur et al., 2006). SA appears to confer resistance to biotrophic pathogens, whereas JA to insect herbivory and necrotrophic pathogens (Pieterse et al., 2001; Bostock, 2005; Stout et al., 2006). This regulatory mechanism could offer the plant a way to activate specific stress response pathways, while repressing others, and, thus contribute to the homeostasis of the cell and/or the appropriate defense response. As indicated above, one mechanism by which SA could antagonize JA action is through the targeting of GCC promoter motifs *via* the TF ORA59 in *A. thaliana* (Van Der Does et al., 2013). More recently, the immune cofactor NPR1 was found to be recruited to JA-responsive promoter regions that are co-occupied by a transcription complex consisting of MYC2 and the MED25 Mediator subunit. In the presence of SA, NPR1 physically associates with MYC2 and inhibits transcriptional activation by disrupting MYC2's interaction with MED25 (Nomoto et al., 2021). Our study suggests that another mechanism of SA–JA antagonism could involve ER stress and the bZIP factors involved therein, but this hypothesis still needs further support with additional experimentation. Nonetheless, such an additional mechanism involving the ER stress machinery could offer plants a means to activate stress response or defense pathways in a stress-specific manner, allowing plants to distinguish between biotic and abiotic stress, or between different pathogens and other attackers. An SA-induced ER stress response could be a mechanism to restrain JA signaling, or at least one of its outputs, namely the elicitation of terpene and/or other specialized metabolite pathways. Given the fact that we observed an antagonism between ER stress-inducible bZIP TFs and JA-inducible bHLH factors in two distinct species, *M. truncatula* and *C. roseus*, which each produce a species-specific compendium of JA-inducible terpene metabolites, this mechanism may be widespread in the plant kingdom. Notably, activation of ER stress and the UPR by JA treatment was recently demonstrated in tomato (Czékus et al.,

2020), indicating that (complex) interplay between ER stress and JA signaling may be common in the plant kingdom. As such, the latter as well as our study may open an avenue for new research on how plants fine-tune their interaction with an ever-changing and often hostile environment.

Data availability statement

The data presented in the study are deposited in the ArrayExpress database, accession number E-MTAB-11668.

Author contributions

AG conceived the project, original screening and research plans. AG supervised the experiments. BR, M-LE, MC, EL, EC, CW, EVH, RC and MP performed the experiments. BR, M-LE, MC, EL, EC, CW, EVH and AG designed the experiments and analyzed the data. BR, M-LE and AG wrote the article with contributions of all the authors. AG agrees to serve as the author responsible for contact and ensures communication. All scientists who have contributed substantially to the conception, design or execution of the work described in the manuscript are included as authors, in accordance with the guidelines from the Committee on Publication Ethics (COPE) (http://publicationethics.org/resources/guidelines). All authors agree to the list of authors and the identified contributions. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Research Foundation Flanders by research project grants to AG (G004515N and G005915N) and a predoctoral fellowship to EC, the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 825730 (Endoscape) to AG, the Program Ciências Sem Fronteiras for a predoctoral fellowship to BR (Grant 201135/2014-0), the Special Research Fund from Ghent University (project O1J14813), the Swiss National Science Foundation for a postdoctoral fellowship to MC (P300PA_177831) and the BEC.AR program for overseas training of Argentine professionals in the fields of science, technology and productive innovation for a scholarship to MP.

Acknowledgments

The authors thank Deniz Malat and Robin Vandenbossche for excellent technical assistance, and Steven Vandersyppe, Keylla Bicalho, and Geert Goeminne for the LC-MS service; Saskia Lippens for discussions and access to the imaging facilities, and Annick Bleys for the critical reading and help in preparing the manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.903793/full#supplementary-material

References

Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H., and Stephanopoulos, G. (2008). Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* 5, 167–190. doi: 10.1021/mp700151b

Beckers, G. J. M., and Spoel, S. H. (2006). Fine-tuning plant defence signalling: salicylate versus jasmonate. *Plant Biol.* 8, 1–10. doi: 10.1055/s-2005-872705

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., et al. (2015). CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol. Plant* 8, 1493–1506. doi: 10.1016/j.molp.2015.06.003

Bostock, R. M. (2005). Signal crosstalk and induced resistance: straddling the line between cost and benefit. *Annu. Rev. Phytopathol.* 43, 545–580. doi: 10.1146/annurev.phyto.41.052002.095505

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., et al. (2011). *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* 23, 3070–3081. doi: 10.1105/tpc.111.087312

Chang, M. C. Y., and Keasling, J. D. (2006). Production of isoprenoid pharmaceuticals by engineered microbes. *Nat. Chem. Biol.* 2, 674–681. doi: 10.1038/nchembio836

Colinas, M., and Goossens, A. (2018). Combinatorial transcriptional control of plant specialized metabolism. *Trends Plant Sci.* 23, 324–336. doi: 10.1016/j.tplants.2017.12.006

Czékus, Z., Csikos, O., Ördög, A., Tari, I., and Poór, P. (2020). Effects of in ER stress and unfolded protein response in tomato plants. *Biomolecules* 10, 1031. doi: 10.3390/biom10071031

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

Deng, Y., Humbert, S., Liu, J.-X., Srivastava, R., Rothstein, S. J., and Howell, S. H. (2011). Heat induces the splicing by IRE1 of a mRNA encoding a transcription factor involved in the unfolded protein response in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7247–7252. doi: 10.1073/pnas.1102117108

Dröge-Laser, W., Snoek, B. L., Snel, B., and Weiste, C. (2018). The *Arabidopsis* bZIP transcription factor family — an update. *Curr. Opin. Plant Biol.* 45, 36–49. doi: 10.1016/j.pbi.2018.05.001

Erffelinck, M.-L., and Goossens, A. (2018). Review: Endoplasmic reticulum-associated degradation (ERAD)-dependent control of (tri)terpenoid metabolism in plants. *Planta Med.* 84, 874–880. doi: 10.1055/a-0635-8369

Erffelinck, M.-L., Ribeiro, B., Gryffroy, L., Rai, A., Pollier, J., and Goossens, A. (2021). The heat shock protein 40-type chaperone MASH supports the endoplasmic reticulum-associated degradation E3 ubiquitin ligase MAKIBISHI1 in *Medicago truncatula*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.639625

Fukushima, E. O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., et al. (2011). CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* 52, 2050–2061. doi: 10.1093/pcp/pcr146

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

Gao, H. B., Brandizzi, F., Benning, C., and Larkin, R. M. (2008). A membrane-tethered transcription factor defines a branch of the heat stress response in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16398–16403. doi: 10.1073/pnas.0808463105

Gholami, A., De Geyter, N., Pollier, J., Goormachtig, S., and Goossens, A. (2014). Natural product biosynthesis in *Medicago* species. *Nat. Prod. Rep.* 31, 356–380. doi: 10.1039/C3NP70104B

Goossens, J., Mertens, J., and Goossens, A. (2017). Role and functioning of bHLH transcription factors in jasmonate signalling. *J. Exp. Bot.* 68, 1333–1347. doi: 10.1093/jxb/erw440

He, J., Benedito, V. A., Wang, M., Murray, J. D., Zhao, P. X., Tang, Y., et al. (2009). The *Medicago truncatula* gene expression atlas web server. *BMC Bioinf.* 10, 441. doi: 10.1186/1471-2105-10-441

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8, R19. doi: 10.1186/gb-2007-8-2-r19

Hemmerlin, A. (2013). Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Sci.*, 203–204, 41-54. doi: 10.1016/j.plantsci.2012.12.008

Henriquez-Valencia, C., Moreno, A. A., Sandoval-Ibañez, O., Mitina, I., Blanco-Herrera, F., Cifuentes-Esquivel, N., et al. (2015). bZIP17 and bZIP60 regulate the expression of BiP3 and other salt stress responsive genes in an UPR-independent manner in *Arabidopsis thaliana*. *J. Cell. Biochem.* 116, 1638–1645. doi: 10.1002/jcb.25121

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

Howell, S. H. (2013). Endoplasmic reticulum stress responses in plants. *Annu. Rev. Plant Biol.* 64, 477–499. doi: 10.1146/annurev-arplant-050312-120053

Howell, S. H. (2021). Evolution of the unfolded protein response in plants. *Plant Cell Environ.* 44, 2625–2635. doi: 10.1111/pce.14063

Iwata, Y., Fedoroff, N. V., and Koizumi, N. (2008). *Arabidopsis* bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. *Plant Cell* 20, 3107–3121. doi: 10.1105/tpc.108.061002

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

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. *Plant Physiol.* 145, 1144–1154. doi: 10.1104/pp.107.106989

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Kazan, K., and Manners, J. M. (2013). MYC2: the master in action. *Mol. Plant* 6, 686–703. doi: 10.1093/Mp/Sss128

Kim, J.-S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2018). ER-anchored transcription factors bZIP17 and bZIP28 regulate root elongation. *Plant Physiol.* 176, 2221–2230. doi: 10.1104/pp.17.01414

Krinke, O., Ruelland, E., Valentová, O., Vergnolle, C., Renou, J.-P., Taconnat, L., et al. (2007). Phosphatidylinositol 4-kinase activation is an early response to salicylic acid in arabidopsis suspension cells. *Plant Physiol.* 144, 1347–1359. doi: 10.1104/pp.107.100842

Kumari, S., Priya, P., Misra, G., and Yadav, G. (2013). Structural and biochemical perspectives in plant isoprenoid biosynthesis. *Phytochem. Rev.* 12, 255–291. doi: 10.1007/s11101-013-9284-6

Lai, Y.-S., Renna, L., Yarema, J., Ruberti, C., He, S. Y., and Brandizzi, F. (2018). Salicylic acid-independent role of NPR1 is required for protection from proteotoxic stress in the plant endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5203–E5212. doi: 10.1073/pnas.1802254115

Liu, J.-X., and Howell, S. H. (2010a). bZIP28 and NF-y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in *Arabidopsis*. *Plant Cell* 22, 782–796. doi: 10.1105/tpc.109.072173

Liu, J.-X., and Howell, S. H. (2010b). Endoplasmic reticulum protein quality control and its relationship to environmental stress responses in plants. *Plant Cell* 22, 2930–2942. doi: 10.1105/tpc.110.078154

Liu, Y., Patra, B., Singh, S. K., Paul, P., Zhou, Y., Li, Y., et al. (2021). Terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*: effects and prospects of environmental factors in metabolic engineering. *Biotechnol. Lett.* 43, 2085–2103. doi: 10.1007/s10529-021-03179-x

Liu, J.-X., Srivastava, R., Che, P., and Howell, S. H. (2007a). An endoplasmic reticulum stress response in *Arabidopsis* is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. *Plant Cell* 19, 4111–4119. doi: 10.1105/tpc.106.050021

Liu, J.-X., Srivastava, R., Che, P., and Howell, S. H. (2007b). Salt stress responses in arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J.* 51, 897–909. doi: 10.1111/j.1365-313X.2007.03195.x

Li, Q., Wei, H., Liu, L., Yang, X., Zhang, X., and Xie, Q. (2017). Unfolded protein response activation compensates endoplasmic reticulum-associated degradation deficiency in *Arabidopsis*. *J. Integr. Plant Biol.* 59, 506–521. doi: 10.1111/jipb.12544

Malhotra, J. D., and Kaufman, R. J. (2007). The endoplasmic reticulum and the unfolded protein response. *Semin. Cell Dev. Biol.* 18, 716–731. doi: 10.1016/j.semcdb.2007.09.003

Meng, Z., Ruberti, C., Gong, Z., and Brandizzi, F. (2017). CPR 5 modulates salicylic acid and the unfolded protein response to manage tradeoffs between plant growth and stress responses. *Plant J.* 89, 486–501. doi: 10.1111/tpj.13397

Mertens, J., Pollier, J., Vanden Bossche, R., Lopez-Vidriero, I., Franco-Zorrilla, J. M., and Goossens, A. (2016a). The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* 170, 194–210. doi: 10.1104/pp.15.01645

Mertens, J., Van Moerkercke, A., Vanden Bossche, R., Pollier, J., and Goossens, A. (2016b). Clade IVa basic helix-loop-helix transcription factors form part of a conserved jasmonate signaling circuit for the regulation of bioactive plant terpenoid biosynthesis. *Plant Cell Physiol.* 57, 2564–2575. doi: 10.1093/pcp/pcw168

Moreno, A. A., Mukhtar, M. S., Blanco, F., Boatwright, J. L., Moreno, I., Jordan, M. R., et al. (2012). IRE1/bZIP60-mediated unfolded protein response plays distinct roles in plant immunity and abiotic stress responses. *PloS One* 7, e31944. doi: 10.1371/journal.pone.0031944

Moses, T., Papadopoulou, K. K., and Osbourn, A. (2014a). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* 49, 439–462. doi: 10.3109/10409238.2014.953628

Moses, T., Thevelein, J. M., Goossens, A., and Pollier, J. (2014b). Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry* 108, 47–56. doi: 10.1016/j.phytochem.2014.10.002

Mur, L. A. J., Kenton, P., Atzorn, R., Miersch, O., and Wasternack, C. (2006). The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. *Plant Physiol.* 140, 249–262. doi: 10.1104/pp.105.072348

Nagashima, Y., Iwata, Y., Ashida, M., Mishiba, K.-I., and Koizumi, N. (2014). Exogenous salicylic acid activates two signaling arms of the unfolded protein response in arabidopsis. *Plant Cell Physiol.* 55, 1772–1778. doi: 10.1093/pcp/pcu108

Nagashima, Y., Mishiba, K.-I., Suzuki, E., Shimada, Y., Iwata, Y., and Koizumi, N. (2011). Arabidopsis IRE1 catalyses unconventional splicing of *bZIP60* mRNA to produce the active transcription factor. *Sci. Rep.* 1, 29. doi: 10.1038/srep00029

Nawkar, G. M., Kang, C. H., Maibam, P., Park, J. H., Jung, Y. J., Chae, H. B., et al. (2017). HY5, a positive regulator of light signaling, negatively controls the unfolded

protein response in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2084–2089. doi: 10.1073/pnas.1609844114

Netala, V. R., Ghosh, S. B., Bobbu, P., Anitha, D., and Tartte, V. (2015). Triterpenoid saponins: a review on biosynthesis, applications and mechanism of their action. *Int. J. Pharm. Pharm. Sci.* 7, 24–28.

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

Nomoto, M., Skelly, M. J., Itaya, T., Mori, T., Suzuki, T., Matsushita, T., et al. (2021). Suppression of MYC transcription activators by the immune cofactor NPR1 fine-tunes plant immune responses. *Cell Rep.* 37, 110125. doi: 10.1016/j.celrep.2021.110125

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *N Phytol.* 213, 1107–1123. doi: 10.1111/nph.14252

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: what does it mean? *Trends Plant Sci.* 14, 87–91. doi: 10.1016/j.tplants.2008.11.005

Pieterse, C. M. J., Ton, J., and Van Loon, L. C. (2001). Cross-talk between plant defence signalling pathways: boost or burden? *AgBiotechNet* 3, 1–8.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Nat. Prod.* 74, 1462–1476. doi: 10.1021/np200218r

Pollier, J., Moses, T., González-Guzmán, M., De Geyter, N., Lippens, S., Vanden Bossche, R., et al. (2013a). The protein quality control system manages plant defence compound synthesis. *Nature* 504, 148–152. doi: 10.1038/nature12685

Pollier, J., Rombauts, S., and Goossens, A. (2013b). Analysis of RNA-seq data with TopHat and cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cell cultures. *Methods Mol. Biol.* 1011, 305–315. doi: 10.1007/978-1-62703-414-2_24

Ribeiro, B., Lacchini, E., Bicalho, K. U., Mertens, J., Arendt, P., Vanden Bossche, R., et al. (2020). A seed-specific regulator of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* 32, 2020–2042. doi: 10.1105/tpc.19.00609

Rojo, E., Solano, R., and Sánchez-Serrano, J. J. (2003). Interactions between signaling compounds involved in plant defense. *J. Plant Growth Regul.* 22, 82–98. doi: 10.1007/s00344-003-0027-6

Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C., et al. (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11655–11660. doi: 10.1073/pnas.97.21.11655

Schweizer, F., Colinas, M., Pollier, J., Van Moerkercke, A., Vanden Bossche, R., De Clercq, R., et al. (2018). An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in *Catharanthus roseus*. *Metab. Eng.* 48, 150–162. doi: 10.1016/j.ymben.2018.05.016

Seki, H., Tamura, K., and Muranaka, T. (2015). P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol.* 56, 1463–1471. doi: 10.1093/pcp/pcv062

She, J., Yan, H., Yang, J., Xu, W., and Su, Z. (2019). croFGD: *Catharanthus roseus* functional genomics database. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00238

Shoji, T. (2019). The recruitment model of metabolic evolution: jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00560

Spyropoulou, E. A., Haring, M. A., and Schuurink, R. C. (2014). RNA Sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics* 15, 402. doi: 10.1186/1471-2164-15-402

Stout, M. J., Thaler, J. S., and Thomma, B. P. H. J. (2006). Plant-mediated interactions between pathogenic microorganisms and herbivorous arthropods. *Annu. Rev. Entomol.* 51, 663–689. doi: 10.1146/annurev.ento.51.110104.151117

Sui, X., Singh, S. K., Patra, B., Schluttenhofer, C., Guo, W., Pattanaik, S., et al. (2018). Cross-family transcription factor interaction between MYC2 and GBFs modulates terpenoid indole alkaloid biosynthesis. *J. Exp. Bot.* 69, 4267–4281. doi: 10.1093/jxb/ery229

Suzuki, H., Achnine, L., Xu, R., Matsuda, S. P. T., and Dixon, R. A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* 32, 1033–1048. doi: 10.1046/j.1365-313X.2002.01497.x

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume medicago truncatula. *BMC Genomics* 15, 312. doi: 10.1186/1471-2164-15-312

Tang, W., Page, M., Fei, Y., Liu, L., Xu, F., Cai, X., et al. (2012). Overexpression of *AtbZIP60deltaC* gene alleviates salt-induced oxidative damage in transgenic cell cultures. *Plant Mol. Biol. Rep.* 30, 1183–1195. doi: 10.1007/s11105-012-0437-3

Trapnell, C. (2013). *Cufflinks. cuffdiff (v6). open module on GenePattern public server* (GenePattern). Available at: https://software.broadinstitute.org/cancer/software/genepattern/modules/docs/Cufflinks.cuffdiff/6.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., et al. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46, D1190–D1196. doi: 10.1093/nar/gkx1002

Vanden Bossche, R., Demedts, B., Vanderhaeghen, R., and Goossens, A. (2013). Transient expression assays in tobacco protoplasts. *Methods Mol. Biol.* 1011, 227–239. doi: 10.1007/978-1-62703-414-2_18

Van Der Does, D., Leon-Reyes, A., Koornneef, A., Van Verk, M. C., Rodenburg, N., Pauwels, L., et al. (2013). Salicylic acid suppresses jasmonic acid signaling downstream of SCF^{COI1}-JAZ by targeting GCC promoter motifs *via* transcription factor ORA59. *Plant Cell* 25, 744–761. doi: 10.1105/tpc.112.108548

Van Moerkercke, A., Steensma, P., Gariboldi, I., Espoz, J., Purnama, P. C., Schweizer, F., et al. (2016). The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant J.* 88, 3–12. doi: 10.1111/tpj.13230

Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., et al. (2015). The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8130–8135. doi: 10.1073/pnas.1504951112

Van Wees, S. C. M., De Swart, E. A. M., Van Pelt, J. A., Van Loon, L. C., and Pieterse, C. M. J. (2000). Enhancement of induced disease resistance by simultaneous activation of salicylate- and jasmonate-dependent defense pathways in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8711–8716. doi: 10.1073/pnas.130425197

Wang, Z., Cheng, K., Wan, L., Yan, L., Jiang, H., Liu, S., et al. (2015). Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genomics* 16, 1053. doi: 10.1186/s12864-015-2258-x

Wang, D., Weaver, N. D., Kesarwani, M., and Dong, X. (2005). Induction of protein secretory pathway is required for systemic acquired resistance. *Science* 308, 1036–1040. doi: 10.1126/science.1108791

Yang, Z.-T., Lu, S.-J., Wang, M.-J., Bi, D.-L., Sun, L., Zhou, S.-F., et al. (2014a). A plasma membrane-tethered transcription factor, NAC062/ANAC062/NTL6, mediates the unfolded protein response in arabidopsis. *Plant J.* 79, 1033–1043. doi: 10.1111/tpj.12604

Yang, Y.-G., Lv, W.-T., Li, M.-J., Wang, B., Sun, D.-M., and Deng, X. (2013). Maize membrane-bound transcription factor Zmbzip17 is a key regulator in the cross-talk of ER quality control and ABA signaling. *Plant Cell Physiol.* 54, 2020–2033. doi: 10.1093/pcp/pct142

Yang, Z.-T., Wang, M.-J., Sun, L., Lu, S.-J., Bi, D.-L., Sun, L., et al. (2014b). The membrane-associated transcription factor NAC089 controls ER-stress-induced programmed cell death in plants. *PloS Genet.* 10, e1004243. doi: 10.1371/journal.pgen.1004243

Yan, J., Zhang, C., Gu, M., Bai, Z., Zhang, W., Qi, T., et al. (2009). The *Arabidopsis* CORONATINE INSENSITIVE1 protein is a jasmonate receptor. *Plant Cell* 21, 2220–2236. doi: 10.1105/tpc.109.065730

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the *ORCA* genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.* 67, 61–71. doi: 10.1111/j.1365-313X.2011.04575.x

Zhou, M., and Memelink, J. (2016). Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol. Adv.* 34, 441–449. doi: 10.1016/j.biotechadv.2016.02.004

Zhou, S.-F., Sun, L., Valdés, A. E., Engström, P., Song, Z.-T., Lu, S.-J., et al. (2015). Membrane-associated transcription factor peptidase, site-2 protease, antagonizes ABA signaling in arabidopsis. *N Phytol.* 208, 188–197. doi: 10.1111/nph.13436

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fpls.2022.1039550

OPEN ACCESS

EDITED BY
Laigeng Li,
Center for Excellence in Molecular Plant Sciences (CAS), China

REVIEWED BY
Joseph Lynch,
West Virginia University, United States
Melissa Hamner Mageroy,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway

*CORRESPONDENCE
Peigao Luo
lpglab@sicau.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 08 September 2022
ACCEPTED 12 October 2022
PUBLISHED 08 November 2022

CITATION
Zhong S, Guan J, Chen C, Tan F and Luo P (2022) Multiomics analysis elucidated molecular mechanism of aromatic amino acid biosynthesis in *Akebia trifoliata* fruit. *Front. Plant Sci.* 13:1039550. doi: 10.3389/fpls.2022.1039550

Multiomics analysis elucidated molecular mechanism of aromatic amino acid biosynthesis in *Akebia trifoliata* fruit

Shengfu Zhong†, Ju Guan†, Chen Chen, Feiquan Tan and Peigao Luo*

Key Laboratory of Plant Genetics and Breeding at Sichuan Agricutural University of Sichuan Province, College of Agronomy, Sichuan Agricultural University, Chengdu, China

Akebia trifoliata is a novel edible and healthy fruit. Here, we found that this fruit had the highest content of total free amino acids and three aromatic amino acids (AAAs) compared with the other popular fruits, and there was an obvious inverse relationship between AAA and flavonoid levels in various fruit tissues. Multiomics analysis revealed that the evolutionarily strengthened synthetic pathway of all three AAAs, the largely regulating ability conferred by *ASP5* in the arogenate pathway and the complementary phenylpyruvate pathway endorsed by *ADT* of both Phe and Tyr biosynthesis provided reasonable explanations for the high AAA content in the flesh of *A. trifoliata* fruit. Gene-specific expression could be the main reason for the inverse relationship between AAAs and flavonoids. This study will help us understand the metabolic mechanism of AAAs and to develop *A. trifoliata* as a fresh fruit crop and medicinal plant by molecular breeding strategies.

KEYWORDS

***Akebia trifoliata*, aromatic amino acids, multiomics, biosynthetic mechanism, whole genome duplication**

Introduction

Amino acids are necessary structural units of all organisms and are very important nutrient materials for almost all life on earth. In addition, amino acids play a key role in various biological processes, such as ATP generation (Koopman et al., 2010), nucleotide synthesis (Laxman et al., 2013), redox balance (Buck et al., 2017), signaling pathways (Zhang et al., 2021) and cellular immunity (Kelly and Pearce, 2020). Individual amino acids can be created under non-biological chemical synthesis (Burton et al., 2012), but the ability to actively synthesize all or some amino acids is essential for all species, especially autotrophic organisms.

In plants, aromatic amino acids (AAAs), including tryptophan (Trp), phenylalanine (Phe) and tyrosine (Tyr), also serve as precursors of many secondary metabolites, such as flavonoids, alkaloids and hormones (Lynch and Dudareva, 2020). In humans, the accumulation of AAA metabolites produced by the gut microbial pathway can enhance immunity (Dodd et al., 2017). AAAs are also precursors of neurotransmitters such as dopamine, epinephrine, norepinephrine, serotonin and tryptamine (Gatellier et al., 2009). Unfortunately, most of animals have lost the ability to synthesize aromatic amino acids in the the process of evolution, especially higher animals (Costa et al., 2015). Therefore, AAAs and some of their derivatives, such as vitamins, are essential in the human diet for our survival and health, while plants are the good source of these nutrients.

All AAAs are derived from the shikimate pathway, which links carbohydrate metabolism to aromatic compound biosynthesis and is only found in microorganisms and plants (Tzin et al., 2009). At the same time, AAAs are also precursors of thousands of primary and specialized metabolites (Widhalm and Dudareva, 2015; Schenck and Maeda, 2018). For example, Trp is the precursor to auxin, alkaloids, indole glucosinolates and phytoalexins, while Tyr serves as the precursor for quinones, betalains, and isoquinoline alkaloids (Barros et al., 2016). However, the carbon flux forward to Phe is the major contributor to the high carbon flux in the shikimate pathway because it can produce more than 8000 secondary metabolites, such as anthocyanins, flavonoids, isoflavonoids, tannins, and volatiles (Maeda and Dudareva, 2012). Hence, further understanding the specific biosynthetic mechanism of AAAs in various species is valuable for both completely elucidating the AAA metabolic process in plants and finally finding new rich AAA food resources.

Akebia trifoliata, belonging to the Ranunculaceae family, is a woody perennial climbing vine (Li et al., 2010), and the fruit of *A. trifoliata*, also called augmelon, has increasingly attracted the attention of many consumers due to its abundant nutrients and health benefits (Yang et al., 2021). A recent study suggested that the largest commercial product of *A. trifoliata* should be as a new fresh fruit crop because its flesh contains a high content of free amino acids, especially AAAs and a very low content of flavonoids, while the peel has low free amino acid contents and high flavonoid contents (Guan et al., 2022). These properties also make *A. trifoliata* an ideal species for studying the molecular mechanisms of AAA metabolism. In addition, the artificial cultivation and improvement of *A. trifoliata* is still in its infancy. In order to better develop the augmelon fruit varieties, it is particularly important to analyze the main mechanism of the accumulation of aromatic amino acids in *A. trifoliata* fruit.

Multiomics, mainly consisting of genomics affording the basic information of all genes, transcriptomes containing the expression data of all functional genes and metabolomes giving the chemical traits of the final materials, is an effective and widely used method in the biological research field. In this study, we systemically characterized the profiling of genes, expressed sequences and final metabolites involved in AAA metabolism pathways of *A. trifoliata* fruit. In addition, we also compared the free amino acid content of *A. trifoliata* fruit with that of other popular fruits. Our main objectives were to understand the evolutionary mechanism at the genome level, the tissue-specific expression mechanism at the transcriptome level and the accumulation mechanism at the metabolome level and to develop strategies to further improve *A. trifoliata* as a new food resource.

Materials and methods

Plant materials

The clonal line "Shusen 1" of *A. trifoliata* was used for transcriptome sequencing and metabolome analysis in this study because the *A. trifoliata* reference genome was also produced by this line, and all the clonal lines were planted in Germplasm Nursery at Chongzhou Research Station (30°43'N, 103°65'E) of Sichuan Agricultural University. Three tissues (flesh, peel and seed) of *A. trifoliata* fruits at four developmental stages (young, enlargement, coloring and mature stages) (Yang et al., 2021) were randomly collected from 2-year-old plants, and three biological replicates were executed so that a total of 36 samples were used for RNA-seq and metabolite analysis. All fresh samples were immediately frozen in liquid nitrogen and then stored at -80°C.

Determination of *A. trifoliata* fruit metabolites

Completely ripened fruits of *A. trifoliata* were utilized to determine the free amino acid contents, and four fruit varieties, Fuji apple (*Malus pumila*), Xianren banana (*Musa nana*), Fengshui pear (*Pyrus sorotina*) and Kyoho grape (*Vitis vinifera*), purchased from Ito Yokado Supermarket, were used as control samples of current popular fruit. A total of 15 flesh samples were separated from mature fruits of these 5 varieties, with three replicates for each variety. The content of 20 common free amino acids was determined by High Performance Liquid Chromatography (HPLC) method with minor modifications (Guan et al., 2022). In briefly, approximately 2 g of homogenized pulp was mixed with 1 mL of 15% salicylsulfonic acid. Then the homogenates were separated and determined using an Agilent 1260 Infinity HPLC system (Agilent Technologies Inc., USA) at a wavelength of 550 nm for 100 min. Seventeen amino acids (MembraPure, Berlin, Germany) were used as standard samples.

The total flavonoid contents of three tissues (flesh, peel, seed) from ripe augmelon fruit with three replicates were determinded by spectrophotometry according to the previously method (Guan et al., 2022). A mixture of 2 g of pulp with 10 mL of 70% aqueous ethanol solution was sonicated for 30 min (20 kHz, KQ-300DE, ShuMei). Then 10 mL of the 5 times diluted liquid was transferred into a 25-mL volumetric flask, and 2 mL of 5% $NaNO_2$, 2 mL of 10% $AlCl_3$, 6 mL of 1 M NaOH and 5 mL of 70% aqueous ethanol solution were sequentially added. This mixture was used to detect the absorbance at 510 nm using a spectrophotometer (UV-2401, Shimadzu Co., Japan); rutin was used as the standard to make the calibration curve.

Untargeted metabolite extraction and measurement were performed by the Metware Biotechnology Co. Ltd. (Wuhan, China). Each of above 36 frozen samples was analyzed using an LC-ESI-MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS, Applied Biosystems 4500 Q TRAP). The analytical methods and conditions were as previously described (Chen et al., 2013). Briefly, each frozen sample was crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. 100 mg powder was weighted in a 2.0 mL centrifuge tube and extracted overnight at 4 °C with 1.0 mL 70% aqueous methanol. The homogenate was centrifuged at 10,000g for 10 min and then the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250mg, 3ml; ANPEL, Shanghai,China) and filtrated (SCAA-104, 0.22 μm pore size; ANPEL,Shanghai, China) before LC-ESI-MS/MS analysis. The untargeted metabolome scans were controlled by Analyst 1.6 software The peak area of all mass spectra was integrated, and the mass spectra of the same metabolite in different samples were corrected integration base on the metware database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Fraga et al., 2010).

Gene characterization and coexpression analysis of AAA and flavonoid metabolism

The *A. trifoliata* genome assembly (Shusen 1) was downloaded from the Genome Warehouse in the National Genomics Data Center under accession number GWHBISH00000000, which is publicly accessible at https://ngdc.cncb.ac.cn/gwh. Each total ribonucleic acid (RNA) of above 36 frozen samples was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol, respectively. The transcriptome library with 240 bp inserts was prepared using an Illumina Kit and sequenced on an Illumina HiSeq™ 2500 (Illumina, Inc. 9885 Towne Centre Drive, San Diego, CA, USA), using paired-end technology. FASTX-Toolkit (v 0.0.13) (Gordon, Cold Spring Harbor, NY, USA) and Fastqc (Andrews, 2014) was used to obtain and check the clean reads by removing both adaptor and low-quality bases such as length less than 20 bp or N bases more than 10% and also by excluding short reads of less than 25 bp in length. The clean reads of the sequences were aligned to the *A. trifoliata* genome assembly by using HISAT v2.0.438, and the abundance of genes in different samples was computed with StringTie v1.2.3 (Haas et al., 2008). Gene expression levels were estimated in fragments per kilobase of transcript per million fragments mapped (FPKM) values using DESeq2 (Love et al., 2014). Gene expression in 4 stages and 3 tissue samples in *A. trifoliata* was evaluated with the WGCNA (Weight Gene Coexpression Network Analysis) R package v1.51 (Langfelder and Horvath, 2008) to identify modules of coexpressed genes with a soft threshold value (power) = 0.8 and cutHeight = 0.4. A total of 6434 expressed genes were obtained from all 36 samples by filtering out with FPKM < 1 or FPKM variance in the top 50%. Four development stages, three tissues, three AAAs and 15 major metabolic node flavonoids were used as the main traits analysis in the WGCNA. The correlation between the gene expression cluster modules and these objective traits was calculated by default parameters. Hub genes were screened by strongly associated with phenotypes with edge weights ≥ 0.30 and plotted by Cytoscape (Shannon et al., 2003). An integrated AAA and flavonoid biosynthesis pathway model with enzymes encoding genes and metabolites was manually constructed based on the above related KEGG pathways and previous reports (Yoo et al., 2013; Lynch and Dudareva, 2020). Cluster heatmaps of the FPKM values of the expressed genes and metabolite contents were plotted with TBtools (Chen et al., 2020). We predicted AAA and flavonoid biosynthesis- and metabolism-related genes according to KEGG functional gene annotations in the 'phenylalanine, tyrosine and tryptophan biosynthesis' (ko00400), 'phenylalanine metabolism' (ko00360) and 'phenylpropanoid biosynthesis' (ko00940) pathways and previous reports (Lin et al., 2015; El-Azaz et al., 2016; Qian et al., 2019; Barros and Dixon, 2020). WGD paralogs were identified through genome synteny analysis using TBtools v1.098769 with the default parameters (Chen et al., 2020).

Cloning and subcellular location of key genes

Full-length *ASP5* (*EVM0012020*) cDNA was isolated from total RNA extracted from *A. trifoliata* seedlings by reverse transcription PCR using iScript Reverse Transcription Supermix (Bio-Rad) with specific primer pair 1 (Table S1). Then, cDNAs were subcloned into the PC2300-35s-eGFP vector (Biovector Co.,LTD) without a stop codon *via* homologous recombination cloning technology with specific primer pair 2 (Table S1). It was further fused to the N-terminus of the green fluorescent protein (eGFP) gene between the *BamI* and *SalI* sites under the control of the 35S promoter.

The blank control *eGFP* and recombinant *eGFP-ASP5* plasmids were introduced into the *Agrobacterium tumefaciens* GV3101 (pSoup-19) strains, respectively. eGFP-tagged ASP5 proteins were expressed through agroinfiltration (OD600 = 0.5) in *Nicotiana benthamiana* leaves, as previously described (He et al., 2004). eGFP fluorescence was monitored 48 h after agroinfiltration using a laser scanning confocal microscope (Carl Zeiss LSM-800, Oberkochen, Germany). For eGFP detection, the excitation source was an argon ion laser at 488 nm, and emission was observed between 505 and 525 nm. Chloroplast autofluorescence was collected between 680 and 700 nm.

Statistical analysis

The multiple comparison of free amino acid and total flavonoid contents between different fruits were determined with Duncan method using SPSS software version 22.0 (SPSS Inc., Chicago, IL, USA). Then the significant difference at $P < 0.05$ was indicated by lowercase letters. Because high-throughput metabolites of 36 samples were measured by untarget metabolome, the content of each metabolite was normalized into the percentage of relative content using the maximum content as 100%. Principal component analysis (PCA) was used to reflect the relative variation of the metabolite contents between different tissues and development stages. The differential metabolites between different groups of 36 samples were identified by orthogonal projections to latent structures-discriminant analysis (OPLS-DA) according to a fold change (FC) ≥ 2 or FC ≤ 0.5 and a variable importance in project (VIP) value ≥1 (Thévenot et al., 2015). Correlation coefficients and P values between the gene expression cluster modules and traits (tissue, stage, and metabolites) were directly calculated from WGCNA method by default parameters.

Results

Both the component and content of free amino acids in the flesh of *A. trifoliata* fruit

The flesh of mature *A. trifoliata* fruit has a delicious and sweet flavor and a soft juicy texture. We detected all 20 common free amino acids in the ripe flesh of *A. trifoliata* fruit and of four other common fruits: apple, banana, pear and grape (Table 1). In all five fruits, both asparagine (Asn) and glutamine (Gln) were the two richest amino acids, while aspartic acid (Asp) was the least abundant amino acid. Further comparison analysis found that the total content of 20 free amino acids (AAs) was highest (10103 ug/g) in augmelon and lowest (2707 ug/g) in apple. The concentration of the total AAs were significantly higher ($P<0.05$) in augmelon than the other fruits except for pear. In details, five amino acids (Phe, Trp, Tyr, cysteine (Cys), glutamic acid (Glu)) had significantly higher contents in *A. trifoliata* than in the other four fruits (Table 1). In particular, both the total contents (2095.99 ug/g) and the proportion (20.75%) of AAAs, including Phe, Tyr and Trp, were significantly higher in *A. trifoliata* fruit. The essential AAs of *A. trifoliata* fruit were also significantly higher than those of the other tested fruits. These results indicated that there could be a specific biosynthetic mechanism of AAAs in *A. trifoliata*.

Metabolome profile of three *A. trifoliata* fruit tissues at four different stages

To further characterize the changes in metabolite abundance during fruit maturation, the metabolome of three *A. trifoliata* fruit tissues at four different stages was measured and analyzed. A total of 581 metabolites were detected from all tested samples (Figure 1A and Table S2), among which the number of flavonoids mainly derived from Phe was the highest (146). PCA analysis showed that variation of metabolite contents in tissues was larger than that in development stages (Figure 1B). The differential statistical analysis identified that a large number of differential metabolites between the different tissues were also enriched in the metabolic pathway associated with AAAs and flavonoids (Figure S1). Specifically, the relative content of total flavonoids was higher in peel, while the relative content of AAAs was lower than that in both flesh and seed, especially at the mature stage (Figures 1C–E). In contrast, the relative contents of AAAs in flesh had the highest level, while that of total flavonoids had the lowest level at the mature stage (stage 4) (Figure 1C). From the general change in trend, the relative content of AAAs increased at mature stage (stage 4) in flesh and seeds (Figures 1C, E), while the total flavonoids decreased with all development stages in flesh and seeds. In peel, the relative content of total flavonoids was higher than that of AAAs at all development stages, respectively (Figure 1D). In addition, we also measured the absolute content of total flavonoids in *A. trifoliata* fruit. We did not detect flavonoids in mature flesh within the error range. In contrast, the flavonoid contents in the peel and seeds were measured to be 29.11 mg/g and 6.78 mg/g, respectively. In summary, these results suggested that an obvious inverse relationship existed between AAAs and their derivative flavonoids during the developmental process in *A. trifoliata* fruit.

Identification of genes involved in AAA metabolic pathways at the genome level

A total of 104 genes putatively involved in AAA biosynthesis and metabolism pathways were identified in the *A. trifoliata* reference genome according to the KEGG database and previous

TABLE 1 Contents of free amino acids in the flesh of *A. Trifoliate*.

Amino acids (AAs)	*A. trifoliata*	Apple	Banana	Pear	Grape
Alanine (Ala)	333.52 ± 64.07^{a}	414.34 ± 130.08^{a}	550.17 ± 166.44^{a}	519.64 ± 181.93^{a}	894.76 ± 368.45^{a}
Arginine (Arg)	18.75 ± 6.95^{b}	5.75 ± 1.20^{ab}	61.38 ± 5.31^{c}	5.45 ± 1.87^{ab}	4.48 ± 0.93^{a}
Asparagine (Asn)	3016.47 ± 354.83^{b}	485.81 ± 132.22^{a}	1229.54 ± 90.34^{a}	2955.84 ± 192.31^{b}	1142.75 ± 400.46^{a}
Aspartic acid (Asp)	0.06 ± 0.01^{ab}	0.17 ± 0.02^{c}	0.02 ± 0.00^{a}	0.08 ± 0.01^{ab}	0.13 ± 0.05^{bc}
Cysteine (Cys)	36.14 ± 15.49^{b}	9.29 ± 2.86^{a}	12.88 ± 3.10^{a}	4.77 ± 1.90^{a}	10.86 ± 2.83^{a}
Glutamine (Gln)	2888.16 ± 351.32^{b}	1154.79 ± 116.23^{a}	2443.78 ± 172.52^{ab}	2826.74 ± 844.98^{b}	2418.6 ± 307.59^{ab}
Glutamic acid (Glu)	836.17 ± 175.28^{c}	298.71 ± 16.64^{ab}	32.25 ± 1.98^{a}	253.72 ± 44.15^{ab}	348.94 ± 98.46^{b}
Glycine (Gly)	48.17 ± 18.43^{a}	10.99 ± 3.74^{a}	326.19 ± 29.36^{b}	13.44 ± 2.24^{a}	41.17 ± 9.83^{a}
Histidine (His)	137.75 ± 49.66^{a}	4.55 ± 0.14^{a}	1144.15 ± 178.43^{c}	50.79 ± 11.81^{a}	778.62 ± 87.4^{b}
Proline (Pro)	595.5 ± 112.53^{b}	70.4 ± 18.74^{a}	275.18 ± 53.09^{ab}	355 ± 182.16^{ab}	326.3 ± 138.09^{ab}
Serine (Ser)	31.17 ± 1.01^{a}	30.25 ± 2.41^{a}	26.21 ± 4.50^{a}	1536.47 ± 491.78^{b}	30.16 ± 1.92^{a}
Threonine^{d} (Thr)	222.98 ± 10.21^{a}	26.98 ± 4.73^{a}	127.78 ± 8.89^{a}	91.33 ± 30.20^{a}	800.93 ± 133.56^{b}
Valine^{d} (Val)	6.68 ± 3.34^{a}	0.84 ± 0.29^{a}	2.93 ± 0.78^{a}	3.08 ± 0.16^{a}	6.6 ± 1.70^{a}
Methionine^{d} (Met)	58.67 ± 15.81^{a}	10.23 ± 2.87^{a}	372.5 ± 31.81^{c}	199.32 ± 61.61^{b}	55.02 ± 9.97^{a}
Lysine^{d} (Lys)	20.49 ± 3.06^{a}	3.29 ± 1.41^{a}	82.35 ± 22.08^{b}	3.02 ± 0.99^{a}	5.68 ± 1.86^{a}
Isoleucine^{d} (Ile)	5.35 ± 1.63^{a}	0.73 ± 0.03^{a}	30.09 ± 11.19^{b}	17.64 ± 5.90^{ab}	3.05 ± 0.58^{a}
Leucine^{d} (Leu)	21.11 ± 7.18^{a}	0.69 ± 0.33^{a}	281.94 ± 98.47^{b}	7.72 ± 2.76^{a}	17.45 ± 3.42^{a}
Tryptophan^{d} (Trp)	1756.44 ± 266.34^{b}	175.79 ± 17.58^{a}	578.71 ± 114.68^{a}	398.53 ± 97.23^{a}	588.41 ± 211.51^{a}
Phenylalanine^{d} (Phe)	69.44 ± 1.81^{c}	3.32 ± 1.04^{a}	45.88 ± 3.45^{b}	8.1 ± 3.23^{a}	37.91 ± 5.16^{b}
Tyrosine (Tyr)	270.11 ± 41.51^{c}	4.57 ± 1.44^{a}	157.74 ± 21.91^{b}	15.26 ± 4.03^{a}	14.3 ± 6.66^{a}
Total AAs	10103.02 ± 522.33^{c}	2706.92 ± 259.15^{a}	7623.91 ± 255.43^{b}	9250.66 ± 547.46^{c}	7511.84 ± 359.89^{b}
Aromatic AAs	2095.99 ± 245.55^{c} (20.75%)	183.68 ± 18.11^{a} (6.79%)	782.33 ± 136.03^{ab} (10.26%)	421.89 ± 98.88^{b} (4.56%)	640.62 ± 205.26^{ab} (8.53%)
Essential AAs	2161.17 ± 250.06^{c} (21.39%)	221.88 ± 20.86^{a} (8.20%)	1522.17 ± 250.87^{b} (19.97%)	728.74 ± 148.2^{a} (7.88%)	1515.05 ± 99.19^{b} (20.17%)

$^{a, b, c, \ldots}$-Different lowercase letters in the same column show significant differences between different fruits at the $p < 0.05$ level.
The data are presented as the mean ± standard error (ug/g FW), and the percentage number (%) represents the proportion of amino acids among the total amino acids.
^{d}Essential amino acids.

(Figure S2). Among them, 17, 26 and 20 genes were further classified into AAA biosynthesis pathways, including the 'shikimate', 'tryptophan biosynthesis' and 'arogenate and phenylpyruvate' subpathways. Similarly, 14 and 27 genes were classified into AAA downstream metabolism pathways, including the 'phenylpropanoid' and 'flavonoid' biosynthesis subpathways, respectively (Table S3). Further synteny analysis found that 43 (41.3%) out of the 104 genes could be derived from WGD, and this proportion of WGD genes was higher than that 8329 of 24138 (34.5%) in the whole genome background (Table S3). This indicated that AAA metabolic pathways could be reinforced in the evolutionary process.

Correlation analysis between gene expression and related traits of AAAs

Transcriptome analysis of fruit samples was conducted to study the biosynthesis mechanism of AAAs. The correlation between the expressed gene cluster modules and traits including development stages, fruit tissues, AAAs and representative flavonoids was calculated by WGCNA method. A total of five highly correlated gene cluster modules were identified, which exhibited an inverse relationship between AAA and flavonoid content (Figure 2A). We further found that the coexpression correlation values between the 'brown4' and 'skyblue' modules (r= 0.76) and between the 'darkseagreen4' and 'grey60' modules (r= 0.95) were very similar, so we put the 'brown4' and 'skyblue' modules together as an integrated gene cluster 2, the 'darkseagreen4' and 'grey60' modules together as the other integrated gene cluster 3 in the following study (Figure 2B). In addition, the results of WGCNA showed that 15 flavonoids were clustered into two different classes: class 1 (liquiritigenin, butein, cyanidin, pinocembrin, dihydromyricetin and delphinidin) and class 2 (dihydroquercetin, dihydrokaempferol, eriodictyol, naringenin, butin, naringenin chalcone, (+)-gallocatechin, fustin and afzelechin) (Figure 2A). The expression of genes in gene cluster 1 of 'brown' module was positively correlated with AAAs, while negatively correlated with all flavonoids except dihydromyricetin in flesh, especially at stage 4; that in gene cluster 2 was highly positively correlated with Trp, Tyr and class 2 flavonoids, while highly negatively correlated with class 1 flavonoids in seed; and that in gene cluster 3 was highly negatively correlated with AAAs and class 2 flavonoids, while

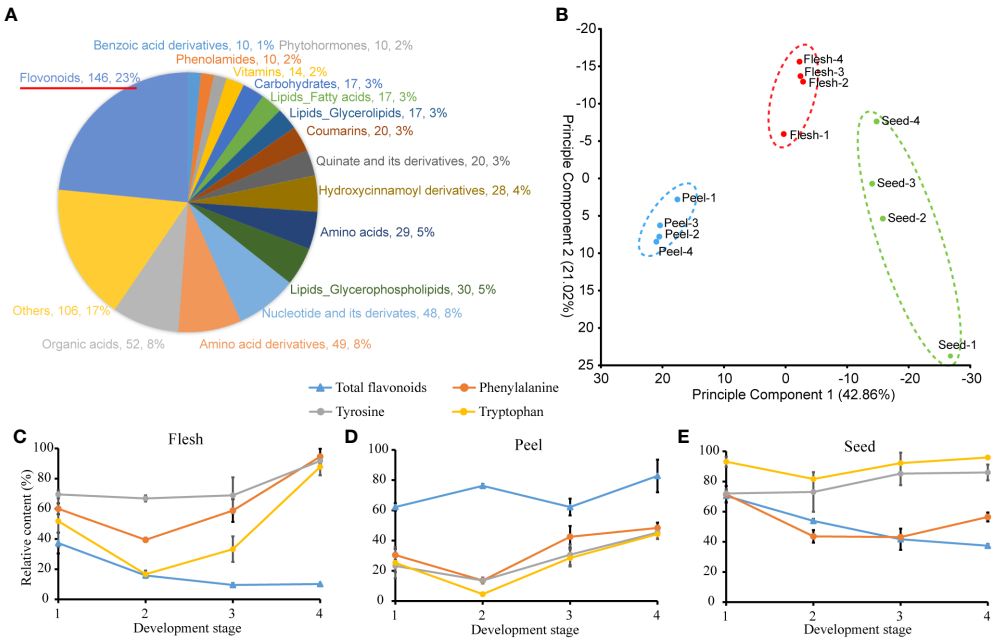


FIGURE 1
Metabolite measurements of flesh, peel and seed tissues at 4 developmental stages. **(A)** Classification of metabolites. **(B)** Plot of principal component analysis of metabolite contents between 3 tissues and 4 development stages. **(C–E)** Relative contents of aromatic amino acids and total flavonoids in different stages and tissues. The maximum content of the given metabolite among 36 samples is as 100%.

highly positively correlated with class 1 flavonoids in peel (Figures 2A, B). Obviously, the AAA metabolic pathway of *A. trifoliata* exhibited obvious tissue or fruit developmental stage specificity (Figure 2B). Then, we focused on the identification of the key genes involved in the metabolic pathway of both AAAs and flavonoids for the five modules.

The evolutionary events mainly experienced candidate key genes involved in the AAA metabolism pathway

Among 104 annotated genes putatively involved in the AAA metabolism pathway, a total of 54 genes, consisting of 29 genes before AAA synthesis and 25 genes after AAA synthesis, were clustered into five highly correlated gene cluster module-trait modules, in which 8, 9, 2, 11 and 24 genes were clustered into three correlative gene clusters, respectively. Based on the results of AAA metabolism at the genome level and transcriptome and metabolome data in this study, we proposed a putative pathway from the tricarboxylic acid cycle to AAAs and then to flavonoids in *A. trifoliata* (Figure S2). Evolutionarily, 30 (56%) of 54 coexpressed genes experienced WGDs (Figure 2C and Table S3), in which 19 and 11 genes were putatively located before and after AAAs in the proposed pathways, respectively (Figures S2, S3). The proportion (66%) of coexpressed genes derived from WGD before AAAs was higher than that (44.0%) after AAAs in the proposed pathways (Figure S2). The results indicated that both the AAA and flavonoid biosynthesis pathways in *A. trifoliata* might simultaneously be reinforced by WGD events, but the degree of AAA biosynthetic processes could be stronger than that of flavonoids.

Identification of key regulatory genes in AAA-related pathways

The reinforced AAA metabolic pathway exhibited obvious tissue or fruit developmental stage specificity (Figure 2). From the 54 coexpression genes on the putative pathway (Figure 2D and Table S3), there were five genes including *PAT* (*EVM0006124*), *ADT* (*EVM0001040*), *GOT2* (*EVM0008507* and *EVM0000841*) and *ASP5* (*EVM0012020*) in gene cluster 1 enriched in the arogenate and phenylpyruvate biological synthetic subpathways of both Phe and Tyr, six genes including three *PAL* (*EVM0013218*, *EVM0015158* and *EVM0002982*) and three *4-CL* (*EVM0023774*, *EVM0011793* and *EVM0005009*) encoding the limited enzymes of phenylpropanoid pathway and five genes *CHI* (*EVM0015959*), *ANS* (*EVM0023082*) and *BZ1* (*EVM0005744*, *EVM0012552* and *EVM0021545*) directly regulating class 1 flavonoids biosynthesis in gene cluster 3, and three genes including *F3H* (*EVM0002323*), *DFR* (*EVM0022942*) and *LAR* (*EVM0020322*) in gene cluster 2

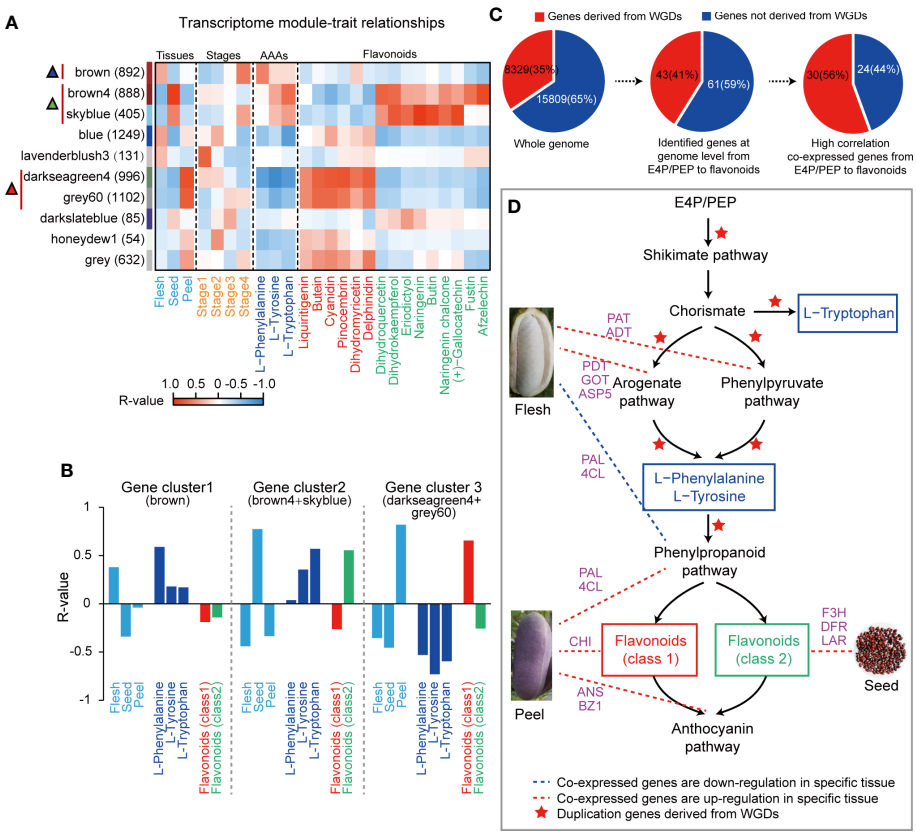


FIGURE 2
Aromatic amino acid biosynthesis and metabolism in *A. trifoliata*. **(A)** Heatmap correlation between the WGCNA gene set module and objective traits. The red vertical line and triangle marker represent the high positive module, and the numbers in brackets are the numbers of genes in each module. Flavonoids were divided into two classes by R-values. **(B)** Association of selected WGCNA modules with different tissues and metabolites. Mean R-values presented in the two integrated modules. **(C)** Percentage of genes derived from WGD events at different levels. **(D)** The postulated pathway for the biosynthesis of aromatic amino acids and flavonoids.

directly regulating class 2 flavonoids biosynthesis, so that they were identified as the candidate genes that plays a key regulatory role in AAAs metabolic pathways (Figure 2D and Figure S4). In addition, among these five key genes, the *ADT* (*EVM0001040*) could encode an enzyme with a prephenate dehydratase (PDT) activity conferring (PAC) domain, the hallmark of PDT enzyme activity (Figure S5). This *ADT* (*EVM0001040*) was derived from WGD events.

Identification and subcellular localization of hub genes

Hub genes have a high degree and importance of connectivity in biological interaction networks. We first screened three hub genes (*PAT*, *GOT2* and *ASP5*) by gene subnetwork analysis, and all three genes belonging to AAA biosynthesis were closely related to coexpressed genes in the gene cluster 3 and connected to several other genes with high connectivity within the network (Figure 3A). Interestingly, all three genes could encode aspartate aminotransferase, a very important enzyme for biosynthesis of Phe and Tyr but have differences in the subcellular location of expression. The *PAT* was probably cytoplasmic, *GOT2* was probably mitochondrial, and *ASP5* was probably located in the chloroplast (Table S3). In the network, both *PAT* and *GOT2* were directly connected with *ASP5* genes, while the direct connection between *PAT* and *GOT2* was obviously absent (Figure 3A), so *ASP5* was treated as the central hub gene in the key node and could play a pivotal role in the regulatory network. Finally, the transient expression of *ASP5* in *N. benthamiana* leaves further confirmed that *ASP5* was localized in the chloroplast (Figure 3B), in which the key node of biosynthesis of both Phe and Tyr was present.

Discussion

Humans survived and profited from the exploitation of plant resources as their final and main food source. In the past,

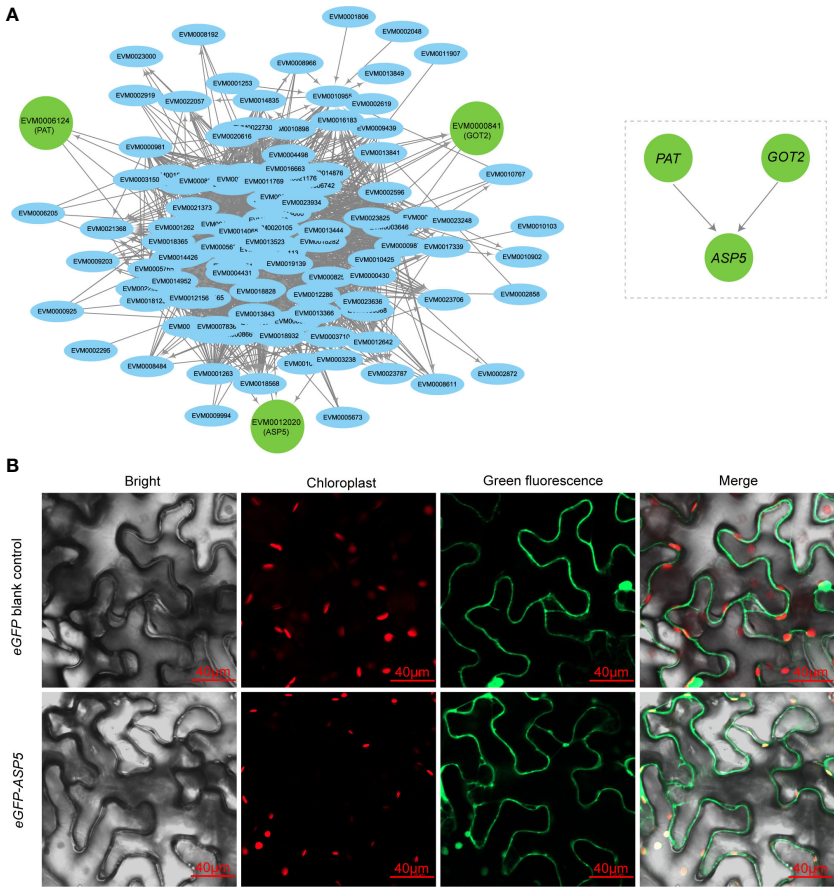


FIGURE 3
Subcellular localization of hub genes. **(A)** Hub genes of the aromatic amino acid biosynthesis pathway in the brown module, as represented by node and edge graphs. **(B)** Subcellular localization of *ASP5*. The *eGFP-ASP5* fusion plasmid was transiently expressed in *Nicotiana benthamiana* leaves. Plasmid only contain *eGFP* was also transiently expressed as control.

especially in remote antiquity, people only considered macronutrients, mainly carbohydrates, proteins and fats, when they found new food sources and largely ignored micronutrients, such minerals and vitamins, which resulted in a number of diseases threatening our health and lives. Today, both the components and contents of natural beneficial materials in food have received increasing attention. Therefore, finding new types of food from various plants, especially multipurpose functional plants, is valuable and significant work.

The fruit of *A. trifoliata*, also called 'augmelon', 'bayuegua' and 'wild bananas', is a new fresh edible fruit, in which the flesh, as the part to directly eat, is rich in vitamins, minerals, free amino acids and proteins (Guan et al., 2022). We found that the total free amino acid content of *A. trifoliata* was the highest among the five tested fruit varieties and further found that the contents of both total AAAs (Trp, Phe and Tyr) were significantly higher than those of the other four popular fruits (Table 1). In addition, in all five fruit varieties, Trp was the most abundant AAA, and Phe was the least abundant AAA, except in grapes. The results indicated that the metabolic pathways of *A. trifoliata* could be different from those of the other four common fruits, and this difference could be explained by evolutionary processes. AAAs are important components of proteins in all organisms. In fact, AAAs are also essential materials for the survival of many animals, including humans, although Tyr can be synthesized from Phe by Phe hydroxylase in animals (Maeda and Dudareva, 2012) because animals derived from fishes have lost the ability to synthesize aromatic amino acids using common substances (Costa et al., 2015). In addition, AAAs, especially Phe, also serve as precursors of various secondary metabolites of plants (Dodd et al., 2017). Hence, the abundant AAA contents in the fruit of *A. trifoliata* provide a great opportunity to further understand the metabolic mechanism of AAAs and to agriculturally exploit new fresh fruit crops from wild plant resources.

We identified a total of 581 metabolites in different fruit tissues and developmental stages (Figure 2A, Table S2), in which flavonoids not only had the largest number of categories

(Figure 2A) but also exhibited tissue specificity (Figure 2). The content of flavonoids was not detectable in ripe flesh, while it was up to 29.11 mg/g in ripe peel, which usually resulted from gene tissue-specific expression rather than the difference at the genome level. The tissue-specific characteristics of metabolites such as flavonoids provide reasonable explanations for the long history as medicinal plants and good environmental adaptation of *A. trifoliata* (Li et al., 2010). Moreover, the obvious inverse relationship between AAA and flavonoid contents also indicated that both AAA synthetic pathways and catabolic pathways could be simultaneously reinforced in evolutionary processes. In addition, we noticed that the derivatives of AAAs were mainly produced by the pathway with the precursor of Phe and Tyr rather than Trp (Table S2), which suggested that the biosynthesis of Trp was reinforced rather than downstream catabolism.

WGD was an important evolutionary force in the plant, and several biological processes, enzymes and transporters in some specific metabolic processes were preferentially reserved after whole-genome triplication of eudicots called 'γ' or independent WGDs around the Cretaceous/Paleogene (K-Pg) boundary (Wu et al., 2020). The shikimate pathway derives AAAs and some eudicot-specific secondary metabolites, such as gallo- and ellagitannins (Okuda et al., 2000; Muir et al., 2011), in which more than 30% of photosynthetically fixed carbon is directed (Tohge et al., 2013), while the phenylpropanoid pathway is one of the most extensively studied specialized metabolic routes, which sits on the boundary between primary and secondary metabolism (Dong and Lin, 2021). In this study, flavonoids were the most abundant, both in number of flavonoids and contents (Table S2). Therefore, we paid much attention to the number and proportion of genes experiencing WGD events in both the shikimate and phenylpropanoid pathways and in the synthetic processes of both AAAs and flavonoids (Figures 2D and S2). We found that the proportion of genes derived from WGD and involved in AAA metabolism pathways was 41.3% (43/104), which was higher than the 34.5% (8329/24138) observed in the whole genome (Table S3). Among 54 coexpressed genes, 30 (56%) experienced WGDs (Figures 2C and S3), and further comparison found that the proportion of genes derived from WGD in the synthetic and catabolic (the main synthesis of flavonoids) pathways of AAAs was 66% (19/29) and 44.0% (11/25), respectively. Moreover, many of these genes exhibited coexpression with 15 major metabolic node flavonoids and both Phe and Tyr. The derivatives of Trp had fewer categories and lower contents, while Trp had the highest content compared with Phe and Tyr, so we would suggest that both the synthetic pathway of all three AAAs and only the catabolic pathway of Phe were reinforced by WGD, which could be an evolutionary genomic basis for the rich free AAA content in the flesh of *A. trifoliata*.

Many studies confirmed that the biosynthetic flux of both Phe and Tyr is mainly through the arogenate pathway located in the chloroplast and is largely regulated by the arogenate dehydratase or prephenate dehydratase (Qian et al., 2019), however, several species, such as *Arabidopsis thaliana* (Bross et al., 2011; Westfall et al., 2015), *Petunia hybrida* (Yoo et al., 2013) and *Pinus pinaster* (El-Azaz et al., 2016), can sufficiently synthesize them through the phenylpyruvate pathway and are mainly regulated by the microbial-like PDT enzyme (El-Azaz et al., 2016). In this study, five genes (two *GOT2*, one *PAT*, one *ADT* and one *ASP5*) in the gene cluster 3 simultaneously took part in both the arogenate and phenylpyruvate pathways of AAA biosynthesis (Figure 2), which indicated that *A. trifoliata* could synthesize Phe and Tyr by both the arogenate and phenylpyruvate pathways. One *ADT* (*EVM0001040*) of the five key regulatory genes in the gene cluster 3 contained a PAC activity domain (Figure S5), which was the hallmark domain of the PDT enzyme, so we would like to accept that *A. trifoliata* could have the other complementary synthetic pathway of Phe and Tyr through the phenylpyruvate pathway (Figure 2D). Interestingly, *EVM0001040* was derived from WGD, affording a chance for the duplicated gene to functionally differentiate. In addition, we further found that all three hub genes (*PAT*, *GOT2* and *ASP5*) in the network putatively encoded aspartate aminotransferase (Figure 3A). Among the three hub genes, *ASP5* could play a more important role in regulating the synthesis of both Phe and Tyr because there was a direct connection between *ASP5* and both *PAT* and *GOT2*, while there was no direct connection between *PAT* and *GOT2*. Moreover, *ASP5* was also localized in chloroplasts (Figure 3B), in which the arogenate pathway is located. The results suggested that the main arogenate pathway of both Phe and Tyr synthesis in *A. trifoliata* had a stronger regulating ability compared with other fruit varieties. Therefore, the possibility that *A. trifoliata* could have the phenylpyruvate pathway of both Phe and Tyr biosynthesis was further increased in our opinion.

Likewise, the catabolism of AAAs, especially Phe, was our last concern because it would be the other important factor affecting the accumulation of AAAs and their derivatives in various tissues of *A. trifoliata* fruit, which was largely regulated by limited enzymes encoded by *PAL* and *4-CL* genes (Barros and Dixon, 2020). In our study, three *PAL* and three *4-CL* coexpressed genes were identified in the gene cluster 2 by WGCNA, and many (66.7%) of them were also duplicated by WGD events (Table S3). The downstream diversifying branch pathways of the phenylpropanoid pathway could be a good explanation for the difference in AAA and derivative contents in various fruit tissues (Figure 2 and Figure S4), which could be regulated by gene tissue-specific expression. For example, the high AAA content in flesh possibly resulted from the blockage of the phenylpropanoid pathway, while the low AAA content in peel and seed could be associated with the enhanced synthetic pathway of class 1 and class 2 flavonoids, respectively.

Conclusion

Comprehensively, we found that the AAA content in the flesh of *A. trifoliata* fruit was significantly higher than that in the flesh of the other four fruits (apple, pear, banana and group), and there was an inverse relationship between the contents of AAAs and their derivative flavonoids in various tissues of *A. trifoliata* fruit. Further multiomics analysis revealed that the evolutionarily reinforced synthetic pathway of all three AAAs, the largely regulating ability conferred by *ASP5* (*EVM0012020*) in the arogenate pathway and the complementary phenylpyruvate pathway endorsed by *ADT* (*EVM0001040*) of both Phe and Tyr biosynthesis afforded reasonable explanations for the high AAA content in the flesh of *A. trifoliata* fruit. Likewise, gene-specific expression could be the main reason for the inverse relationship between AAAs and flavonoids. The results will be helpful to completely understand the metabolic mechanism of AAAs in plants and valuable to commercially accelerate the exploitation of *A. trifoliata* as a fresh fruit crop and medicinal plant by molecular breeding strategies.

Data availability statement

The raw sequencing data presented in the study are deposited in the National Centre for Biotechnology Information (NCBI) with accession numbers from SAMN16551931 to SAMN16551942 under the BioProjectID PRJNA671772. Other original contributions are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

SZ: formal analysis, Software, Writing - Original Draft, Visualization, Methodology. JG: Validation, Investigation, Resources, Writing - Original Draft. CC: Data Curation, Formal Analysis. FT: Investigation, Resources. PL: Conceptualization, Writing - Review and Editing, Project Administration, Funding Acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Science and Technology Department of Sichuan Province, China (2020YFN0091 and 2020YJ0331).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1039550/full#supplementary-material

References

Andrews, S. (2014) *FastQC a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barros, J., and Dixon, R. A. (2020). Plant phenylalanine/tyrosine ammonia-lyases. *Trends Plant Sci.* 25, 66–79. doi: 10.1016/j.tplants.2019.09.011

Barros, J., Serrani-Yarce, J. C., Chen, F., Baxter, D., Venables, B. J., and Dixon, R. A. (2016). Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat. Plants* 2, 16050. doi: 10.1038/nplants.2016.50

Bross, C. D., Corea, O. R. A., Kaldis, A., Menassa, R., Bernards, M. A., and Kohalmi, S. E. (2011). Complementation of the pha2 yeast mutant suggests functional differences for arogenate dehydratases from *Arabidopsis thaliana*. *Plant Physiol. Biochem. Ppb.* 49, 882–890. doi: 10.1016/j.plaphy.2011.02.010

Buck, M. D., Sowell, R. T., Kaech, S. M., and Pearce, E. L. (2017). Metabolic instruction of immunity. *Cell* 169, 570–586. doi: 10.1016/j.cell.2017.04.004

Burton, A. S., Stern, J. C., Elsila, J. E., Glavin, D. P., and Dworkin, J. P. (2012). Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* 41, 5459–5472. doi: 10.1039/c2cs35109a

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

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

Costa, I. R., Thompson, J. D., Ortega, J. M., and Prosdocimi, F. (2015). Metazoan remaining genes for essential amino acid biosynthesis: sequence conservation and evolutionary analyses. *Nutrients* 7, 1–16. doi: 10.3390/nu7010001

Dodd, D., Spitzer, M. H., Van Treuren, W., Merrill, B. D., Hryckowian, A. J., Higginbottom, S. K., et al. (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* 551, 648–652. doi: 10.1038/nature24661

Dong, N. Q., and Lin, H. X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J. Integr. Plant Biol.* 63, 180–209. doi: 10.1111/jipb.13054

El-Azaz, J., de la Torre, F., Ávila, C., and Cánovas, F. M. (2016). Identification of a small protein domain present in all plant lineages that confers high prephenate dehydratase activity. *Plant J.* 87, 215–229. doi: 10.1111/tpj.13195

Fraga, C. G., Clowers, B. H., Moore, R. J., and Zink, E. M. (2010). Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography–mass spectrometry, XCMS, and chemometrics. *Anal Chem.* 82, 4165–4173. doi: 10.1021/ac1003568

Gatellier, P., Kondjoyan, A., Portanguen, S., Grève, E., Yoon, K., and Santé-Lhoutellier, V. (2009). Determination of aromatic amino acid content in cooked meat by derivative spectrophotometry: Implications for nutritional quality of meat. *Food Chem.* 114, 1074–1078. doi: 10.1016/j.foodchem.2008.10.009

Guan, J., Fu, P., Wang, X., Yu, X., Zhong, S., Chen, W., et al. (2022). Assessment of the breeding potential of a set of genotypes selected from a natural population of *Akebia trifoliata* (Three-leaf *Akebia*). *Horticulturae* 8, 116. doi: 10.3390/horticulturae8020116

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

He, X., Anderson, J. C., Pozo, O., Gu, Y.-Q., Tang, X., and Martin, G. B. (2004). Silencing of subfamily I of protein phosphatase 2A catalytic subunits results in activation of plant defense responses and localized cell death. *Plant J.* 38, 563–577. doi: 10.1111/j.1365-313X.2004.02073.x

Kelly, B., and Pearce, E. L. (2020). Amino assets: how amino acids support immunity. *Cell Metab.* 32, 154–175. doi: 10.1016/j.cmet.2020.06.010

Koopman, W. J. H., Nijtmans, L. G. J., Dieteren, C. E. J., Roestenberg, P., Valsecchi, F., Smeitink, J. A. M., et al. (2010). Mammalian mitochondrial complex I: biogenesis, regulation, and reactive oxygen species generation. *Antioxidants Redox Signaling* 12, 1431–1470. doi: 10.1089/ars.2009.2743

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

Laxman, S., Sutter, B. M., Wu, X., Kumar, S., Guo, X., Trudgian, D. C., et al. (2013). Sulfur amino acids regulate translational capacity and metabolic homeostasis through modulation of tRNA thiolation. *Cell* 154, 416–429. doi: 10.1016/j.cell.2013.06.043

Lin, C.-Y., Wang, J. P., Li, Q., Chen, H.-C., Liu, J., Loziuk, P., et al. (2015). 4-coumaroyl and caffeoyl shikimic acids inhibit 4-coumaric acid:coenzyme a ligases and modulate metabolic flux for 3-hydroxylation in monolignol biosynthesis of p*opulus trichocarpa*. *Mol. Plant* 8, 176–187. doi: 10.1016/j.molp.2014.12.003

Li, L., Yao, X., Zhong, C., Chen, X., and Huang, H. (2010). *Akebia*: A potential new fruit crop in China. *HortSci. horts* 45, 4–10. doi: 10.21273/hortsci.45.1.4

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

Lynch, J. H., and Dudareva, N. (2020). Aromatic amino acids: A complex network ripe for future exploration. *Trends Plant Sci.* 25, 670–681. doi: 10.1016/j.tplants.2020.02.005

Maeda, H., and Dudareva, N. (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* 63, 73–105. doi: 10.1146/annurev-arplant-042811-105439

Muir, R. M., Ibáñez, A. M., Uratsu, S. L., Ingham, E. S., Leslie, C. A., McGranahan, G. H., et al. (2011). Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). *Plant Mol. Biol.* 75, 555–565. doi: 10.1007/s11103-011-9739-3

Okuda, T., Yoshida, T., and Hatano, T. (2000). Correlation of oxidative transformations of hydrolyzable tannins and plant evolution. *Phytochemistry* 55, 513–529. doi: 10.1016/S0031-9422(00)00232-6

Qian, Y., Lynch, J. H., Guo, L., Rhodes, D., Morgan, J. A., and Dudareva, N. (2019). Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. *Nat. Commun.* 10, 15–15. doi: 10.1038/s41467-018-07969-2

Schenck, C. A., and Maeda, H. A. (2018). Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry* 149, 82–102. doi: 10.1016/j.phytochem.2018.02.003

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

Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* 14, 3322–3335. doi: 10.1021/acs.jproteome.5b00354

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. (2013). Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00062

Tzin, V., Malitsky, S., Aharoni, A., and Galili, G. (2009). Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in *Arabidopsis*. *Plant J.* 60, 156–167. doi: 10.1111/j.1365-313X.2009.03945.x

Westfall, C. S., Xu, A., and Jez, J. M. (2015). Structural evolution of differential amino acid effector regulation in plant chorismate mutases. *J. Biol. Chem.* 289, 28619–28628. doi: 10.1074/jbc.M114.591123

Widhalm, J. R., and Dudareva, N. (2015). A familiar ring to it: biosynthesis of plant benzoic acids. *Mol. Plant* 8, 83–97. doi: 10.1016/j.molp.2014.12.001

Wu, S., Han, B., and Jiao, Y. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. doi: 10.1016/j.molp.2019.10.012

Yang, H., Chen, W., Fu, P., Zhong, S., Guan, J., and Luo, P. (2021). Developmental stages of *Akebia trifoliata* fruit based on volume. *Hortic. Sci. Technol.* 39, 823–831. doi: 10.7235/HORT.20210072

Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., et al. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants *via* a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* 4, 2833. doi: 10.1038/ncomms3833

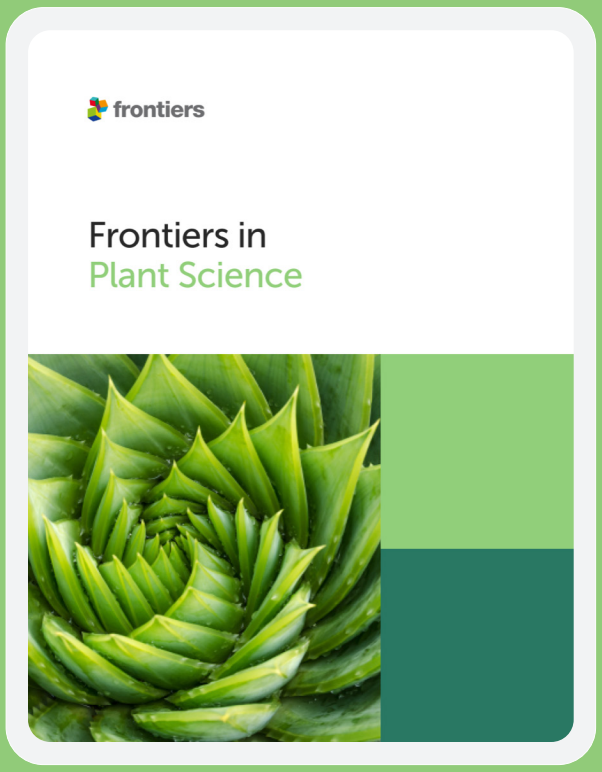
Zhang, X., Wang, X., Yuan, Z., Radford, S. J., Liu, C., Libutti, S. K., et al. (2021). Amino acids-Rab1A-mTORC1 signaling controls whole-body glucose homeostasis. *Cell Rep.* 34, 108830. doi: 10.1016/j.celrep.2021.108830

Frontiers in
Plant Science

Cultivates the science of plant biology and its applications The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact