Synthetic and Systems Biotechnology 9 (2024) 78–87

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Transcriptome-wide analysis of PIP reductase gene family identified a phenylpropene synthase crucial for the biosynthesis of dibenzocyclooctadiene lignans in *Schisandra chinensis*

Tingyan Qiang [a,b], Yu Chen [a], Bin Li [a], Yuqing Dong [a], Xueping Wei [a], Jiushi Liu [a,**], Bengang Zhang [a,*], Haitao Liu [a,***], Peigen Xiao [a]

[a] State Key Laboratory of Basis and New Drug Development of Natural and Nuclear Drugs, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, 100193, China
[b] College of Traditional Chinese Medicine, Hebei University, Baoding, 071000, China

ARTICLE INFO

Keywords:
Schisandra chinensis (Turcz.) baill.
Dibenzocyclooctadiene lignans
PIP reductase family
Phenylpropenes
Isoeugenol synthase

ABSTRACT

Phenylpropenes, such as isoeugenol and eugenol, are produced as defend compounds, floral attractants, and flavor constituents by phenylpropene synthases belonging to the PIP reductase family. Moreover, isoeugenol is proposed to be involved in the biosynthesis of dibenzocyclooctadiene lignans, the main active compounds of *Schisandra chinensis* (Turcz.) Baill. fruits (SCF). *S. chinensis*, a woody vine plant, is widely used for its medicinal, horticultural, edible, and economic values. In this study, nine *ScPIP* genes were identified and characterized from the transcriptome datasets of SCF. The expression profiles revealed that *ScPIP* genes were differentially expressed during different developmental stages of SCF. Three *ScPIP*s were selected and cloned as candidate genes encoding phenylpropene synthases according to phylogenetic analysis. ScPIP1 was proved to function as isoeugenol synthase (IGS) and designated as ScIGS1 through *in vivo* functional characterization in *Escherichia coli*. Subcellular localization analysis demonstrated that *ScIGS1* was localized in both the cytoplasm and nucleus. The three-dimensional (3D) model of ScIGS1 was obtained using homology modeling. Site-directed mutagenesis experiments revealed that the substitution of residues at positions 110 and 113 impacted the product specificity of ScIGS1 and the mutation of Lys157 to Ala abolishing catalytic function. Moreover, the k_{cat} values of mutants were lower than that of ScIGS1 using a deep learning approach. In conclusion, this study provides a basis for further research on PIP reductases and the biosynthetic pathway of dibenzocyclooctadiene lignans.

1. Introduction

Phenylpropenes, such as isoeugenol, eugenol, chavicol, and isochavicol, are volatile compounds derived from phenylalanine. The structural diversity of phenylpropenes depends on the substituents of the benzene ring and the placement of the double bond in the propenyl side chain. Phenylpropenes contribute to the flavors and scents of several significant herbs, spices, flowers, and fruits [1]. Plants often synthesize and store phenylpropenes to defend against herbivores and microorganisms and to attract insect pollinators [2–5]. Humans have used plant materials containing phenylpropenes since ancient times to preserve and enhance the flavor of food and as medicinal agents [6]. Nowadays, phenylpropenes play a vital role as flavor compounds in various processed food products [7–9]. Moreover, phenylpropenes have multiple effects on human health, such as analgesic, antibacterial, antioxidant, and antiviral [10–13].

Previous research suggested that phenylpropene synthases are NADPH-dependent reductases from the PIP reductase family [14]. The PIP reductase family was named after the three initially identified enzymes, pinoresinol-lariciresinol reductase (PLR) [15], isoflavone reductase (IFR) [16], and phenylcoumaran benzylic ether reductase (PCBER) [17]. The proteins in the PIP reductase family catalyze the

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: liujiushi90@163.com (J. Liu), zhangbengang1958@163.com (B. Zhang), htliu0718@126.com (H. Liu).

https://doi.org/10.1016/j.synbio.2023.11.011
Received 1 August 2023; Received in revised form 25 November 2023; Accepted 30 November 2023
Available online 7 December 2023

reductive reactions by forming the quinone-methide intermediate and contain the NADPH binding domain GXXGXXG in their N-terminus and a highly conserved lysine residue essential for the enzymatic activity [18–20]. In recent years, more members of the PIP reductase family have been identified, such as IGS [6], eugenol synthase (EGS) [6], and leucoanthocyanidin reductase (LAR) [21]. PhIGS1 from *Petunia hybrida* cv. Mitchell and ObEGS1 from *Ocimum basilicum* were first identified to use coniferyl acetate and NADPH as substrates to form isoeugenol and eugenol, respectively [6]. Afterward, only a few IGS and EGS were identified, such as CbIGS1, CbEGS1, and CbEGS2 from *Clarkia breweri* [22], PhEGS1 [22], and ObEGS2-8 [14]. Additionally, IGS was suggested to participate in the biosynthesis of bioactive dibenzocyclooctadiene lignans from *Schisandra chinensis* (Turcz.) Baill. [23,24], the first time IGS was considered to be involved in the biosynthesis of medicinal natural products besides in phenylpropene biosynthesis.

S. chinensis is a deciduous woody vine plant widely used for its medicinal, horticultural, edible, and economic values. With increasing consumption, *S. chinensis* has become an artificially grown crop in China, Korea, and Japan, and has been listed as a key crop for increasing economic development in northeastern China [25,26]. *S. chinensis* fruits (SCF) with characteristic flavors such as spicy and fragrant have been used as traditional Chinese medicine (TCM) for thousands of years [27,28]. Although phenylpropenes have been shown to contribute spicy anise- and clove-like notes to a range of fruits [1,29], the genes encoding phenylpropene synthases have not been identified from *S. chinensis*. Moreover, dibenzocyclooctadiene lignans are the principal bioactive constituents of SCF and they exhibit significant hepatoprotective activity [30–32]. However, the biosynthetic pathway of dibenzocyclooctadiene lignans remains largely unknown, hindering the application of synthetic biology, and the last known enzyme is IGS (Fig. 1), which has not been identified from *S. chinensis* [33]. Although the high-quality genome of *S. chinensis* is not currently available, the transcriptome data of SCF have been sequenced and assembled by many scholars in recent years, which provide valuable resources for further gene family and function analysis [34–36].

In this study, we used bioinformatics methods to identify the PIP reductase family members in *S. chinensis* and analyzed their physicochemical properties, conserved motifs, evolutionary relatedness, and expression patterns. Three candidate genes encoding phenylpropene synthases were selected and cloned. Specifically, a novel phenylpropene synthase gene (named *ScPIP1*) was revealed to encode an enzyme to function as IGS. In addition, subcellular localization, homologous modeling, site-directed mutagenesis, and k_{cat} prediction were performed. These results provide a reference for studying the evolution of enzymatic function in the PIP reductase family and lay the foundation for further elucidating the formation mechanism of the characteristic flavors of SCF and the biosynthetic pathway of dibenzocyclooctadiene lignans.

2. Materials and methods

2.1. Plant materials and chemical reagents

Plant samples were collected in the Beijing medicinal plant garden at the Institute of Medicinal Plant Development (IMPLAD) (40°N and 116°E) in Beijing, and were authenticated as *S. chinensis* by Professor Bengang Zhang of IMPLAD. After entering the fruiting period, we regularly picked SCF with three biological replicates at 20, 50, 80, and 110 days (abbreviation: S1, S2, S3, and S4) after fluorescence (Supplementary Fig. 1). The samples were immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction. Ethyl acetate of UPLC grade was purchased from Merck (Merck KGaA, Darmstadt, Germany). Reference compounds such as eugenol and coniferyl alcohol were purchased from Macklin Biochemical (Shanghai, China). Isoeugenol was obtained from Chengdu DeSiTe Biological Technology (Chengdu, China).

2.2. Identification of the PIP reductase gene family

We previously constructed the full-length transcriptomics of SCF at four different developmental stages under BioProject PRJNA817680 [36], which was used for the identification of PIP reductase family members. The hidden Markov Model (HMM) profile of the conserved NmrA domain (PF05368) was retrieved from the Pfam database [37] using the hmmbuild tool embedded in HMMER3.0 search program of TBtools [38,39]. Subsequently, the HMM profile was used to search the protein database of the full-length transcriptomics of SCF using the hmmsearch tool. Finally, the candidate ScPIPs were submitted to the NCBI Batch Web CD-Search Tool with default parameters to examine the presence of the NmrA domain. The candidate proteins with the conserved sequence GXXGXXG were named randomly from ScPIP1 to ScPIP9.

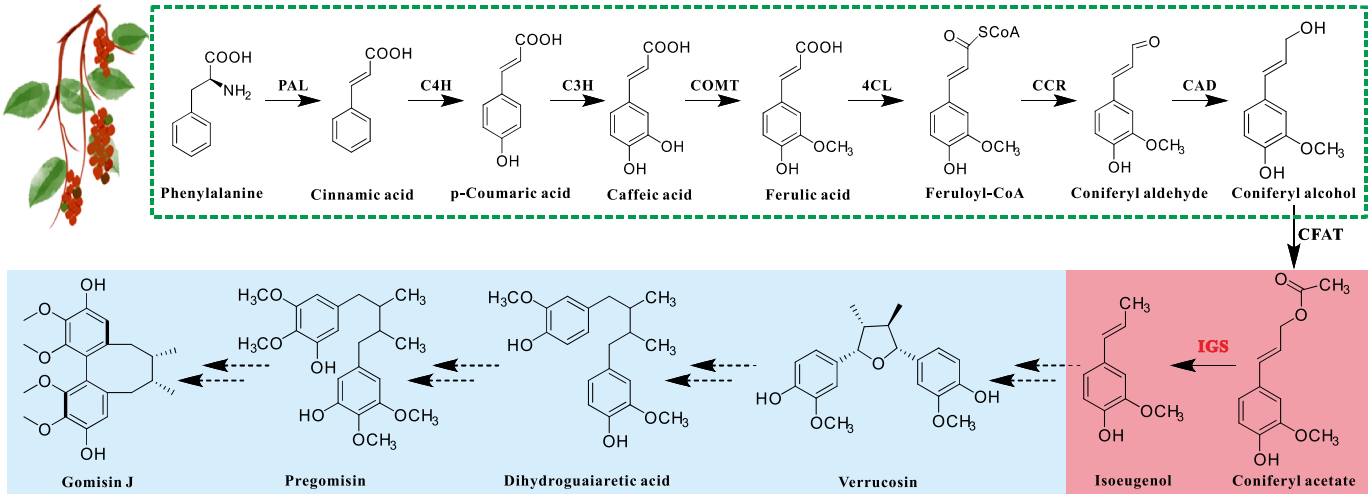

Fig. 1. The proposed biosynthetic pathway of dibenzocyclooctadiene lignans. Abbreviations: PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; C3H, coumarate-3-hydroxylase; COMT, caffeic acid *O*-methyltransferase; 4CL, 4-hydroxycinnamate CoA ligase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; CFAT, coniferyl alcohol acyltransferase and IGS, isoeugenol synthase. Solid arrows represent pathways substantiated by experiments; broken arrows represent pathways assumed based on the comparison of chemical structures.

2.3. Bioinformatic analyses

Firstly, the sequence integrity of ScPIP proteins was analyzed by performing multiple sequence alignments using DNAMAN v9. ExPASy (https://web.expasy.org/protparam/) was used to calculate the length of amino acid, molecular weight (MW), theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY) [40]. Wolf PSORT (https://wolfpsort.hgc.jp/) was employed to predict the subcellular localization of ScPIP proteins [41].

To generate a phylogenetic tree of nine ScPIPs and other PIP members from other plant species, thirty-two biochemically characterized PIP members were downloaded from NCBI (Supplementary Table 1), and nine ScPIP protein sequences were aligned with them using MAFFT 7.4 [42]. The phylogenetic tree was generated using MEGA-X software by the Neighbor-Joining (NJ) method with bootstrap values of 1000 replicates [43]. Evolutionary distances were computed using the Poisson correction method. And the online tool iTOL (https://itol.embl.de/) was used to visually embellish the phylogenetic tree [44].

Furthermore, the conserved motif composition of all PIP proteins was analyzed using the MEME 5.4 online program (http://memesuite.org/tools/meme) [45]. The maximum number of motifs was set to 10, with the width of each motif ranging from 6 to 50 residues. The remaining parameters were set to default.

2.4. Total RNA isolation and cDNA synthesis

The total RNA of SCF was extracted and purified using the EASYspin Plus Complex Plant RNA Kit (Aidlab Biotechnology, Beijing, China) according to the manufacturer's instructions. The quality of the total RNA was verified by measuring at 260 nm and 280 nm optical densities determined by the NanoDrop 2000 ultraviolet spectrophotometer (Thermo Fisher Scientific, San Jose, CA, USA). Additionally, 1 % agarose gel electrophoresis was employed to assess the integrity of the total RNA. The first-strand cDNA was synthesized using the GoScript™ Reverse Transcription Kit (Promega, Beijing, China) according to the manufacturer's protocol and stored at −20 °C.

2.5. Real-time quantitative PCR (qRT-PCR) analysis

To analyze the transcription levels of nine *ScPIP* genes at different development stages of SCF, specific primers were designed by Primer 5 software and are listed in Supplementary Table 2. The qPCR experiments were performed in 25 μL reaction mixture using TB Green Premix Ex Taq (Takara Biotechnology, Dalian, China) on a Bio-Rad CFX96 Real-Time system using the following program: (1) 95 °C for 30 s; (2) 40 cycles of 95 °C for 5 s and 60 °C for 30 s; and (3) 95 °C for 10 s, followed by 65 °C for 5 s, and lastly 95 °C for 5 s. The *ScGAPDH* (glyceraldehyde-3-phosphate dehydrogenase) was used as the internal reference gene. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [46]. Three technological replicates were used for each sample.

2.6. Cloning and sequencing of candidate genes

The specific primers listed in Supplementary Table 2 were designed for gene cloning using CLC Genomics Workbench. The full-length cDNA of *ScPIP1*, *ScPIP2*, and *ScPIP4* was amplified in a 25 μL system consisting of 1 μL cDNA, 0.5 μL forward and reverse primer, 5 μL 5 × PrimeSTA GXL Buffer, 2 μL 2.5 mM dNTP Mixture, 15.5 μL ddH_2O, and 0.5 μL PrimeSTA GXL DNA Polymerase (Takara Biotechnology, Dalian, China). The amplification procedure involved initial denaturation at 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 10 s, annealing at 55 °C for 15 s, extension at 72 °C for 30 s, and a final extension step at 72 °C for 5 min. Subsequently, the amplification products of appropriate length were purified using a Gel DNA Mini Purification Kit (Aidlab Biotechnology, Beijing, China) and ligated into the digested pET-28a vector using a ClonExpress Ultra One Step Cloning Kit (Vazyme, Nanjing, China). The resulting expression constructs were subsequently transformed into *Escherichia coli* DH5α competent cells and sequenced by Tsingke Biotechnology (Beijing, China).

2.7. In vivo enzyme assays in E. coli

Due to the presence of an endogenous enzyme in *E. coli* that catalyzes coniferyl alcohol to coniferyl acetate, we used *E. coli* cells expressing the candidate genes and administered coniferyl alcohol to investigate the function of ScPIP1, ScPIP2, and ScPIP4. In the *E. coli in vivo* feeding assay, *E. coli* BL21 (DE3) harboring *ScPIP1*, *ScPIP2*, and *ScPIP4* expression constructs was cultured in 50 mL Luria Bertani (LB) liquid medium supplemented with 50 μg/mL kanamycin at 37 °C and 180 rpm until the optical density at 600 nm (OD_{600}) reached 0.6–0.8. The expression was induced at 18 °C for 20 h by adding 0.5 mM isopropyl β-D-1-thio-galactopyranoside (IPTG), and coniferyl alcohol dissolved in dimethyl sulfoxide (DMSO) with a final concentration of 50–100 μg/mL was introduced simultaneously.

2.8. Extraction of organic compounds in spent medium

After incubation for 20 h, the cells and spent medium were separated by centrifugation at 12,000 rpm. For the determination of organic compounds in the spent medium, 50 mL of ethyl acetate was added and briefly vortexed with the medium to separate the respective phases. The ethyl acetate layers were collected, and the solvent was recovered *via* vacuum evaporation. Subsequently, the residue was dissolved with methanol and filtered through a 0.22 μm membrane filter into the vials intended for analysis.

2.9. UPLC and GC-MS analyses

The chromatographic analysis of the catalytic production was performed on a Waters ACQUITY UPLC H-Class system equipped with a quaternary pump solvent management system, an autosampler, and a PDA detector. The ACQUITY UPLC BEH C18 (2.1 mm × 100 mm, 1.7 μm) column was chosen to separate the samples. The column and autosampler temperatures were conditioned at 25 °C and 10 °C, respectively. The mobile phase was composed of A (water) and B (acetonitrile) using a gradient elution of 95 % A at 0–3 min, 95-70 % A at 3–7 min, 70-25 % A at 7–10 min, 25-5% A at 10–15 min, and 5 % A at 15–20 min. The flow rate of the mobile phase was set at 0.300 mL/min. The injection volume was 2 μL, and the absorption was monitored at 248 nm. The parameters of GC-MS were set as previously described [33].

2.10. Transient expression and subcellular localization

The pCAMBIA1302 vector driven by the cauliflower mosaic virus (CaMV) 35S promoter was used to determine the subcellular localization of ScIGS1 by utilizing the transient expression system of *Nicotiana benthamiana* leaves. The full-length cDNA of *ScIGS1* without a stop codon was amplified using primers listed in Supplementary Table 2. The amplified sequence was then ligated to the pCAMBIA1302 vector digested by *Spe*I (TaKaRa Biotechnology, Dalian, China). The recombinant plasmid was transferred into *Agrobacterium tumefaciens* strains GV3101 using the conventional freezing-thawing method. The cells were cultured overnight at 28 °C and resuspended in infiltration buffer containing 1 mol L^{-1} Mg_2Cl, 1 mol L^{-1} MES, 0.2 mol L^{-1} acetosyringone, and then were incubated for 2–4 h. The GV3101 suspension harboring pCAMBIA1302-*ScIGS1*-GFP and empty pCAMBIA1302-GFP were transiently injected into the abaxial epidermis of the leaves of five-week-old *N. benthamiana*. After infection for 48 h, the green fluorescence protein (GFP) was observed under the Nikon C2-ER laser scanning confocal microscope.

2.11. Three-dimensional (3D) model construction of ScIGS1 using homology modeling

The BLASTP search for the amino acid sequence of ScIGS1 was performed to determine the appropriate template in the Brookhaven Protein Data Bank (PDB) (https://www.rcsb.org/). The Modeller 10.2 software was used to generate the 3D model of ScIGS1. Five models were produced and selected according to the DOPE (Discrete Optimized Protein Energy) score. The stereochemical property of the model was evaluated by PROCHECK [47] and VERIFY-3D [48]. The visualized 3D model and superimposed diagram of ScIGS1 and CbEGS1 were generated using the PyMOL software.

2.12. Site-directed mutagenesis and deep learning-based kcat prediction

To identify the specific residues determining whether the product is isoeugenol or eugenol, we conducted sequence comparisons of ScIGS1, PhIGS1, CbIGS1, PhEGS1, CbEGS1, CbEGS2, and ObEGS1 using CLC Genomics Workbench. The mutagenic primers (Supplementary Table 2) for each mutation were designed using an online tool (https://crm.vazyme.com/cetool/singlepoint.html). Subsequently, the ScIGS1 mutants were constructed in the pET-28a expression vector using the Mut Express Fast Mutagenesis Kit (Vazyme, Nanjing, China). The mutations were then transformed into *E. coli* DH5α competent cells and confirmed by sequencing in Tsingke Biotechnology. All ScIGS1 mutant proteins were individually heterologously expressed and assayed as described before, except that a multistep gradient from 95 % A (0–3 min), 95-70 % A (3–7 min), 70-25 % A (7–17 min), 25-5% A (17–27 min), and 5 % A (27–32 min). In addition, the enzyme turnover numbers (k_{cat}) of ScIGS1 and its mutants were predicted using a deep learning approach based merely on substrate structure and protein sequences [49].

3. Results and discussion

3.1. Identification and characterization of the PIP reductase family in S. chinensis

As the whole genome of *S. chinensis* is not yet available, systematic transcriptome-wide mining was conducted to explore the *PIP* reductase gene family in *S. chinensis*. Short/long and redundant sequences were removed, and nine *ScPIP* genes were detected and named from *ScPIP1* to *ScPIP9*. As summarized in Table 1, the length of nine ScPIP proteins varies from 307 to 375 amino acids, corresponding to the MW range of 33.45–41.07 kDa. Currently, all identified IGS and EGS-encoded proteins contain no less than 300 amino acids and no more than 330 amino acids, with the longest being PhIGS1 encoding 323 amino acids. Therefore, ScPIP6 and ScPIP7 encoding 375 and 373 amino acids are almost impossible to function as phenylpropene synthases.

The pI of these proteins ranges from 4.70 to 7.67, indicating their acidic to basic nature. The instability index ranges from 29.56 to 42.20, indicating that ScPIPs are stable proteins. The aliphatic index of ScPIPs varies from 92.88 to 101.56, representing their levels of hydrophobicity. In addition, the GRAVY of ScPIPs varies from −0.164 to 0.003, indicating their hydrophilic nature. Moreover, subcellular localization prediction results suggest that the majority of ScPIP proteins (7 out of 9, 77.78 %) localize in the cytoplasm, possibly due to the absence of transit peptides or other sequences that can direct localization in organelles or secretion.

3.2. Sequence alignments, phylogenetic tree, and conserved motifs of PIP reductases

Multiple sequence alignments indicate that all the nine ScPIP proteins had the NADPH binding domain GXXGXXG in their N-terminus, as highlighted in a red rectangle in Fig. 2. To investigate the evolutionary relationship between ScPIPs and other PIP reductases, a phylogenetic tree was constructed using the NJ method with 32 biochemically characterized proteins of the PIP reductase family from other species. As shown in Fig. 3, all the PIP proteins are divided into four clades, which is consistent with the previous research findings [22,50]. Four ScPIPs (ScPIP3, ScPIP5, ScPIP8, and ScPIP9) clustered into clade I with PLRs, suggesting they may function as PLR in *S. chinensis*. ScPIP6 and ScPIP7 clustered into clade II with LARs, indicating that ScPIP6 and ScPIP7 may function as LAR. ScPIP2 and ScPIP4 are clustered into clade III, which includes PCBERs, PTRs, EGSs, and IFRs. In addition, ScPIP1 is clustered into clade IV, which contains IGSs and EGSs.

ScPIP1 has the highest identity with PhIGS1 (59.71 % identity) and CbIGS1 (47.67 % identity) among the ScPIP proteins. Moreover, compared with other ScPIP proteins, ScPIP2 exhibits the highest identity with PhEGS1 (74.03 % identity) and CbEGS2 (77.02 %). ScPIP4 is 52.58 % identical to PhEGS1 and 53.7 % identical to CbEGS2. ObEGS1 and PhIGS1 are 52 % sequence identical. CbIGS1 and CbEGS1 are 58 % and 59 % identical to ObEGS1, respectively. CbEGS2 is only 46 % identical to CbIGS1 and CbEGS1. PhEGS1 is 82 % identical to CbEGS2. In addition, in comparison with ObEGS1, ObEGS2-8 exhibited ~52 %, 43 %, 46 %, 67 %, 46 %, 42 %, and 42 % identity respectively. Therefore, ScPIP1, ScPIP2, and ScPIP4 were selected as candidate genes encoding phenylpropene synthases (IGS and EGS) for subsequent molecular cloning and functional characterization. In addition, conserved motif analysis (Supplementary Fig. 2) showed that the PIP proteins in the same clade had similar motifs. Notably, except for ThPLR, all PIP proteins have motif 4, which contains the GXXGXXG domain.

3.3. Expression profiles of ScPIP genes during fruit development

The expression patterns of nine *ScPIP* genes in SCF at four developmental stages were tested using qRT-PCR to describe their roles during fruit development. As shown in Fig. 4, all the *ScPIP* genes displayed divergence in expression patterns, with most members exhibiting time-specific expressions. Compared to the first stage, *ScPIP1* and *ScPIP2* exhibited higher relative expression levels at the last three stages. Conversely, *ScPIP5* expressed higher at the first three stages than at the last stage. *ScPIP3* had higher expression at the first two stages than at the last two stages. *ScPIP6* and *ScPIP7* expressed higher at the first stage than at the last three stages. *ScPIP8* expressed higher at the middle two stages than at the S1 and S4 stages. However, *ScPIP4* and *ScPIP9*

Table 1
Molecular characteristics of *ScPIP* genes.

Gene name	CDS (bp)	Peptide (aa)	Molecular weight (kDa)	pI	Instability index	Aliphatic index	GRAVY	Subcellular localization
ScPIP1	1,023	340	38.24	7.67	40.94	92.88	−0.142	Cytoplasm
ScPIP2	924	307	33.45	5.85	29.56	98.14	0.003	Cytoplasm
ScPIP3	945	314	35.19	6.02	42.20	101.56	−0.116	Chloroplast
ScPIP4	930	309	34.65	5.94	31.28	97.77	−0.147	Chloroplast
ScPIP5	942	313	35.36	6.20	31.67	94.70	−0.107	Cytoplasm
ScPIP6	1,128	375	41.01	6.10	41.96	96.24	−0.037	Cytoplasm
ScPIP7	1,122	373	41.07	4.70	32.50	94.16	−0.042	Cytoplasm
ScPIP8	1,023	340	38.68	5.96	36.47	93.76	−0.164	Cytoplasm
ScPIP9	942	313	35.37	6.07	37.25	99.04	−0.081	Cytoplasm

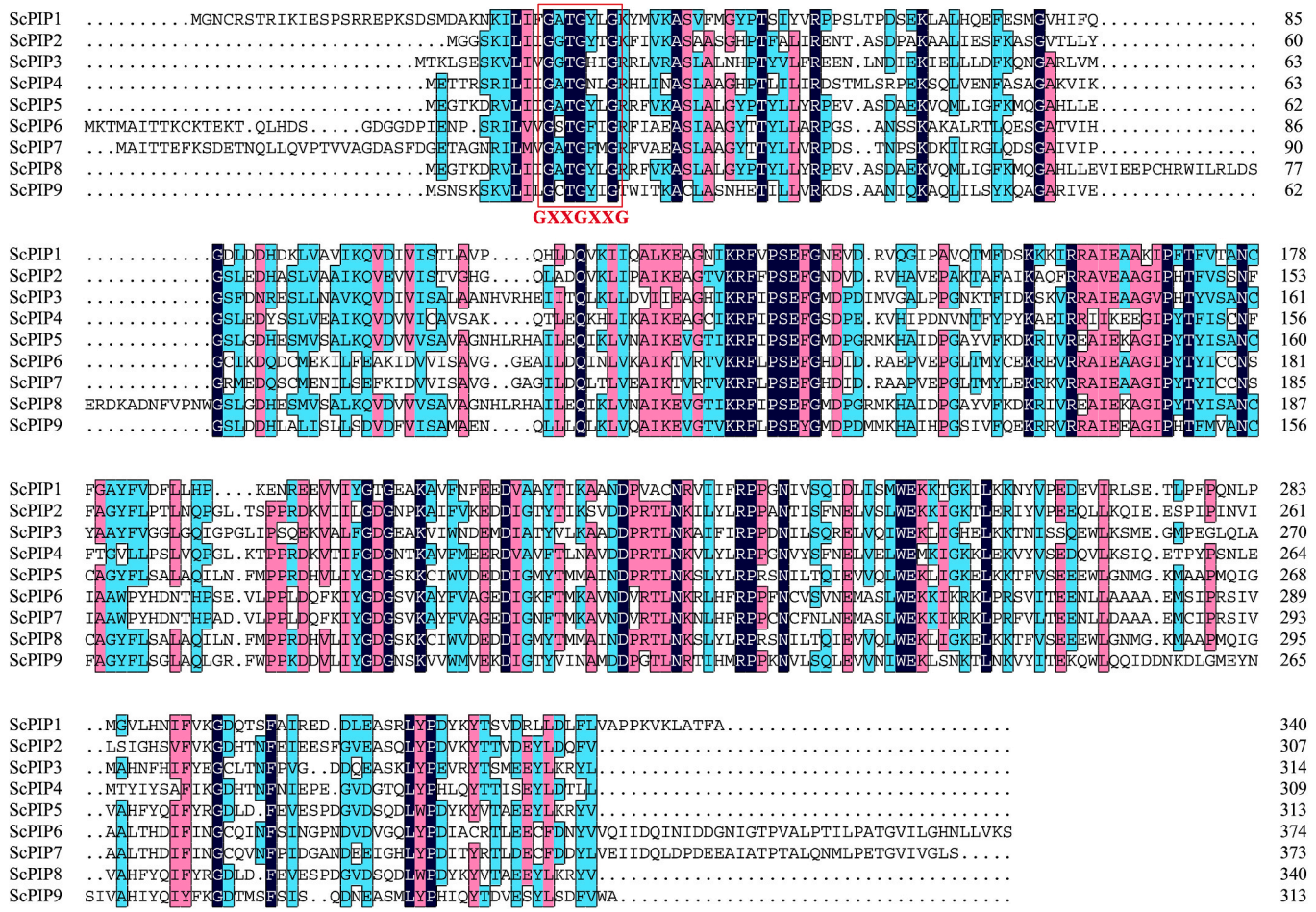

Fig. 2. Multiple sequence alignments of ScPIP proteins. Identical amino acids are shaded, and gaps are indicated by dots. The conserved characteristic PIP reductase domain is marked out in a red rectangle.

exhibited low expression at all four stages. Notably, research showed that *ObEGS1-8* expressed differentially in leaf, leaf stripped of peltate glandular trichomes (PGTs), root, and PGTs [14], which suggests that we should examine the expression profiles of *ScPIPs* in different tissues of *S. chinensis* and associate it with the content profiles of dibenzocyclooctadiene lignans.

As shown in Supplementary Fig. 3A, we also analyzed the SCF samples of four stages using UPLC-Q/TOF-MS (Waters, Milford, MA, USA). The content of several dibenzocyclooctadiene lignans in SCF increased sharply at the S2 stage, and slightly decreased or continued to increase at the S3 and S4 stages (Supplementary Fig. 3B). Therefore, genes involved in the biosynthesis of dibenzocyclooctadiene lignans should exhibit high expression at the last three stages. The expression profiles of *ScPIP1* and *ScPIP2* were consistent with those of *ScCFAT* [33] and the content change pattern of dibenzocyclooctadiene lignans. Therefore, combined with phylogenetic analysis, *ScPIP1* is most likely to function as IGS involved in the biosynthesis of dibenzocyclooctadiene lignans.

3.4. Cloning and functional characterization of ScPIPs

The full-length CDS of *ScPIP1*, *ScPIP2*, and *ScPIP4* were successfully amplified from the cDNA of SCF, as shown in Supplementary Fig. 4. The CDS lengths of *ScPIP1*, *ScPIP2*, and *ScPIP4* were 1023, 924, and 930 bp, respectively, which matched the RNA-seq data. To characterize the enzyme activity of ScPIP1, ScPIP2, and ScPIP4, we transformed *E. coli* with the recombinant plasmids of *ScPIPs* and supplemented the culture medium with coniferyl alcohol, which can be acetylated to coniferyl acetate by the endogenous acyltransferases in *E. coli* [6]. As shown in Fig. 5A, the *E. coli* harboring the recombinant plasmids of *ScPIP1* produced isoeugenol in the medium, similar to the *E. coli* harboring the positive recombinant plasmids of *PhIGS1*. The MS spectrum of isoeugenol is shown in Supplementary Fig. 5. Accordingly, *ScPIP1* was designated as *ScIGS1* and deposited under GenBank Accession Number OP191685. However, the *E. coli* harboring pET-28a, *ScPIP2*-pET-28a, and *ScPIP4*-pET-28a produced neither isoeugenol nor eugenol, suggesting that ScPIP2 (78.9 % identity to PcPCBER) and ScPIP4 (55.66 % identity to ThPCBER) may act as PCBER. In addition, compared to other **phenylpropene synthases,** ScPIP2 and ScPIP4 have mutations (Thr15 and Asn16) in the NADPH binding domain (Supplementary Fig. 6), which might impair their ability to function as **phenylpropene synthases.**

Currently, only one **phenylpropene synthase**, ScIGS1, has been isolated and characterized from SCF. In a previous study, due to a mixture of both eugenol and isoeugenol in roughly similar proportions emitted by *C. breweri* flowers, Koeduka et al. identified CbIGS1, CbEGS1, and CbEGS2 from *C. breweri* flowers [22]. Moreover, due to the presence of eugenol in *O. basilicum* roots, Reddy et al. analyzed the RNA-Seq data of *O. basilicum* roots and identified seven new genes named *ObEGS2-8* apart from *ObEGS1* identified from *O. basilicum* PGTs [14]. Therefore, in the future, we should detect the presence of isoeugenol and eugenol in other parts of *S. chinensis* and screen more PIP family reductase transcripts from the RNA-Seq data.

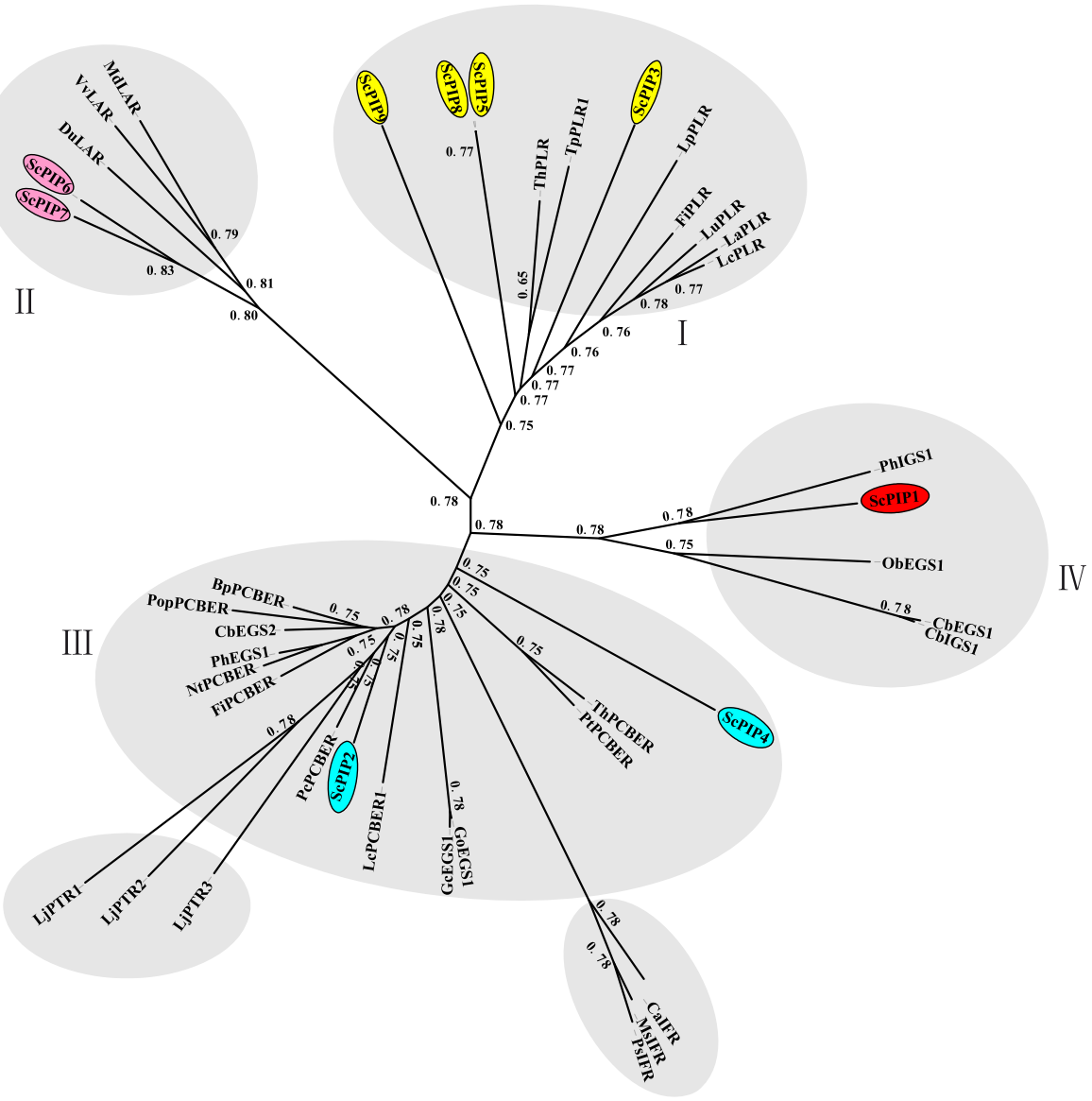

Fig. 3. Phylogenetic tree of PIP reductases from *S. chinensis* and other plant species constructed by MEGA-X. The deduced PIP reductases from *S. chinensis* are shown on a colored background.

3.5. Subcellular localization of ScIGS1

The prediction of protein subcellular localization using the WoLF PSORT Server showed that ScIGS1 was likely to be located in the cytoplasm. To investigate this prediction, the nucleotide sequence without a stop codon and with homologous sequences of restriction enzyme *Spe*I was obtained and ligated into the digested pCAMBIA1302 vector through homologous recombination. The recombinant plasmids were then transformed into *N. benthamiana* leaves through Agrobacterium-mediated transformation. At 48 h after injection, the tobacco leaves were collected to observe the green fluorescence. As shown in Fig. 5B, ScIGS1 showed strong green fluorescent protein signals in both the cytoplasm and nucleus of the tobacco leaf epidermal cells, compared with the control group expressing only green fluorescent protein.

3.6. 3D model of ScIGS1 protein

To clarify the structure of ScIGS1, the 3D model of ScIGS1 was constructed using Modeller 10.2 software based on the X-ray crystal structure of phenylpropene synthase from *C. breweri* (CbEGS1, PDB accession number: 3C1O) as the template, with a sequence identity of 55.66 %. The final unperturbed conformations of the model were acquired by progressively relaxing parts of the initial model. Then the stereochemical quality of the ScIGS1 model was analyzed using PROCHECK. The Ramachandran plot showed that the most favored regions correspond to 89 % of the structure, and only 0.7 % of residues are placed in disallowed regions (Supplementary Fig. 7). An analysis using Verify-3D shows that 85.12 % of the residues have an averaged 3D-1D score greater than or equal to 0.2 (Supplementary Fig. 8). The structure of ScIGS1 comprised an N-terminal domain and a C-terminal domain (Fig. 6A).

In all known reactions catalyzed by PIP reductases, the hydride derived from NADPH is transferred to the substituent positioned para to the 4-hydroxyl group on the phenyl ring of the substrates [51]. Therefore, it is speculated that these two domains of ScIGS1 formed a cleft where the binding sites for the co-factor NADPH and substrate coniferyl acetate are located. As shown in Fig. 6B, the 3D model and Lys157 of ScIGS1 overlapped with the template, which indicates that the 3D model of ScIGS1 is rational. Furthermore, the earlier crystallographic studies of ObEGS1 in complex with an analog of coniferyl acetate, EMDF, clearly

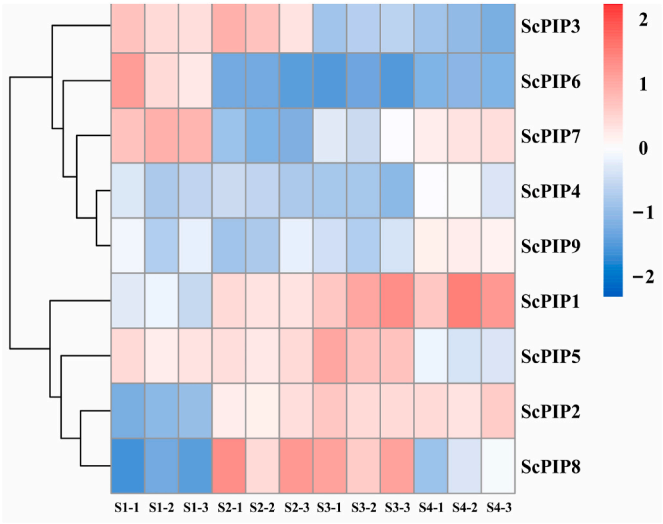

Fig. 4. Expression profiles of *ScPIP* genes at different development stages of SCF.

revealed the substrate-binding mode within the ObEGS1 active site, suggesting there is a water-mediated hydrogen-bonding interaction between the EMDF 4-hydroxy group and the side-chain amino moiety of a conserved lysine residue, Lys132 [18]. Therefore, it's inferred that Lys157 might act as a general base during the catalytic process and be essential for ScIGS1.

3.7. Amino acid residues in ScIGS1 determining product specificity

To determine what led to the specificity of the product, the amino acid residues in ScIGS1 were examined. As shown in Fig. 7A, coniferyl alcohol can be catalyzed by CFAT and IGS or EGS to form isoeugenol and eugenol, respectively. Both IGS and EGS belong to the PIP reductase family, but they can catalyze coniferyl acetate to form isomers with only differences in double bond position, which are unusual reduction reactions. To determine which amino acids affect the product specificity of ScIGS1, multiple sequence alignments of ScIGS1 with other IGS and EGS were performed. As shown in Fig. 7B, these proteins have slightly different numbering of positions due to differences in upstream sequences, but we use the numbers of the ScIGS1 positions in the text to clarify. At position 110, all IGS enzymes have a V (valine), while three EGS enzymes (CbEGS1, CbEGS2, and ObEGS1) have F (phenylalanine) and PhEGS1 has Q (glutamine). At position 113, ScIGS1 has H (histidine) and two other IGS proteins (PhIGS1 and CbIGS1) have Y (tyrosine), while three EGS proteins (CbEGS1, CbEGS2, and ObEGS1) have I (isoleucine), and PhEGS1 has L (leucine). Therefore, the results suggested that the amino acids at these two positions affect the product specificity of ScIGS1.

3.8. Site-directed mutagenesis of ScIGS1 and k_{cat} prediction

To study the amino acid residues that determine the product specificity of ScIGS1, site-directed mutagenesis was performed on the gene. As shown in Fig. 7C and Table 2, changing residue 110 from V to F (mutant 1) gave a product with a 6:94 eugenol/isoeugenol ratio. Changing residue 113 from H to I (mutant 3) gave a protein with the slightly highest ratio of eugenol/isoeugenol production, 7:93, and changing amino acids at positions 110 and 113 (mutant 4) gave a product with a 3:97 eugenol/isoeugenol ratio. All three mutants catalyzed the formation of mainly isoeugenol, with a small amount of eugenol, suggesting that other specific residues of ScIGS1 affect the product specificity. However, mutant 2 (H113Y) did not change the product preference, possibly due to the convergent evolution of secondary metabolite biosynthesis in plants. Mutant 5 (K157A) lost its ability to convert coniferyl acetate to isoeugenol, confirming the role of Lys157 as a catalytic residue for ScIGS1.

Li et al. developed a deep learning approach (DLKcat) for high-throughput k_{cat} prediction and applied this approach to predict genome-scale k_{cat} values for more than 300 yeast species [49]. Moreover, DLKcat can predict the k_{cat} of metabolic enzymes from any organism and capture the effects of single or multiple amino acid substitutions on the k_{cat} values of enzymes. Due to the lack of high-throughput methods for k_{cat} measurement, experimentally measured k_{cat} data are sparse and noisy. Here, we used DLKcat to predict k_{cat} values of ScIGS1 and its mutants and observed that the k_{cat} values of all five mutant enzymes were lower than that of the wild-type enzyme, indicating decreased enzyme activity. Therefore, further research using DLKcat can identify amino acid residues that can enhance the catalytic activity of ScIGS1 and other enzymes.

4. Conclusions

In this study, the identification and characterization of the PIP reductase family in *S. chinensis* were conducted based on the transcriptome datasets of SCF. A total of nine ScPIPs were identified and their physicochemical properties were predicted. Analyses of multiple

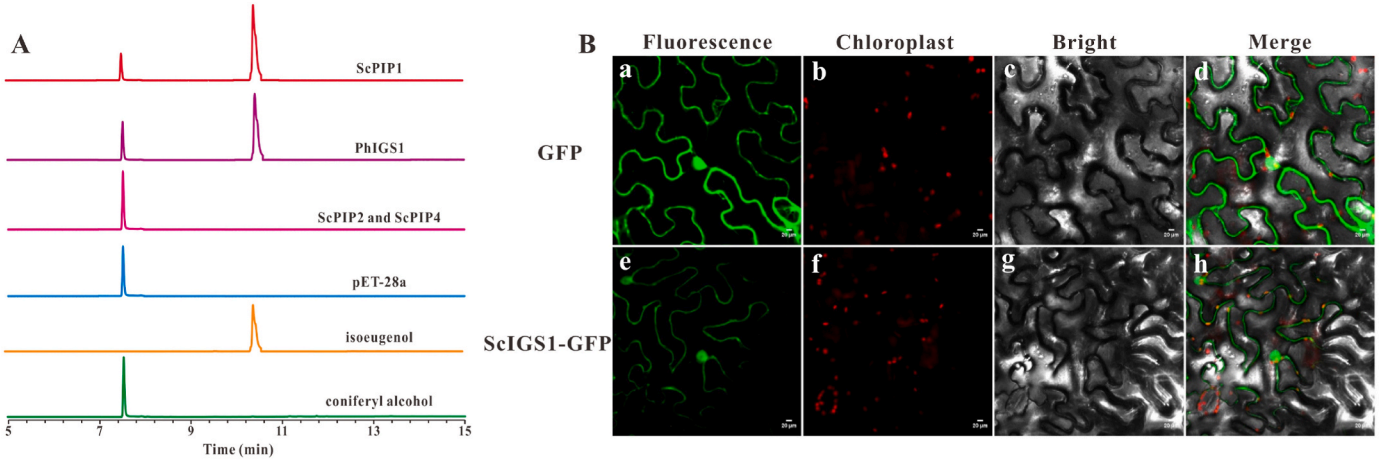

Fig. 5. (A) Analysis of products formed in *E. coli* spent medium catalyzed by ScPIPs and PhIGS1. (B) Subcellular localization of ScIGS1. a, e show the green fluorescence channel; b, f show the chloroplast autofluorescence channel; c, g show the bright field channel; d, h were created from the images shown in the first three panels. Scale bar = 20 μm.

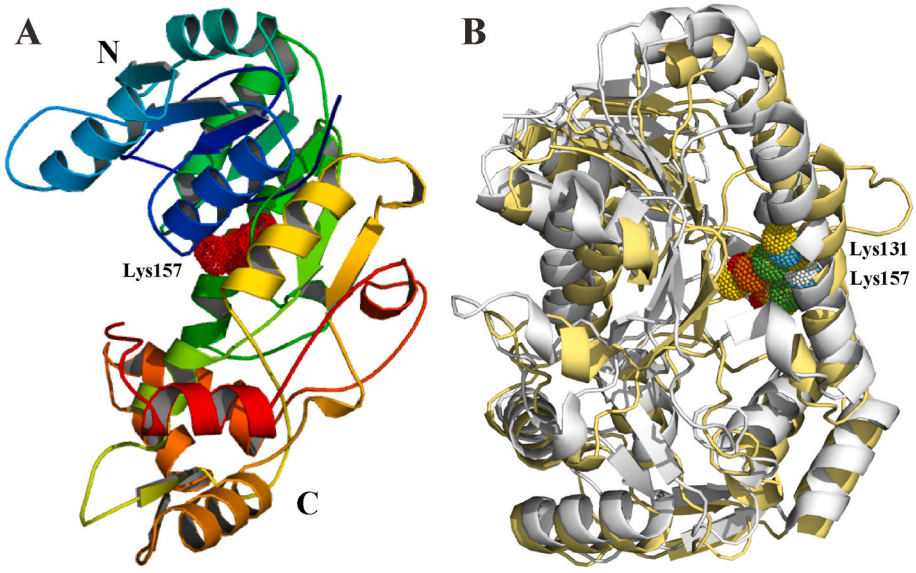

Fig. 6. (A) Homology modeling of 3D model of ScIGS1 constructed by the Modeller 10.2 software. (B) The superimposed diagram of ScIGS1 (yellow) and CbEGS1 (gray). Lys157 and Lys131 are shown in the dot representation.

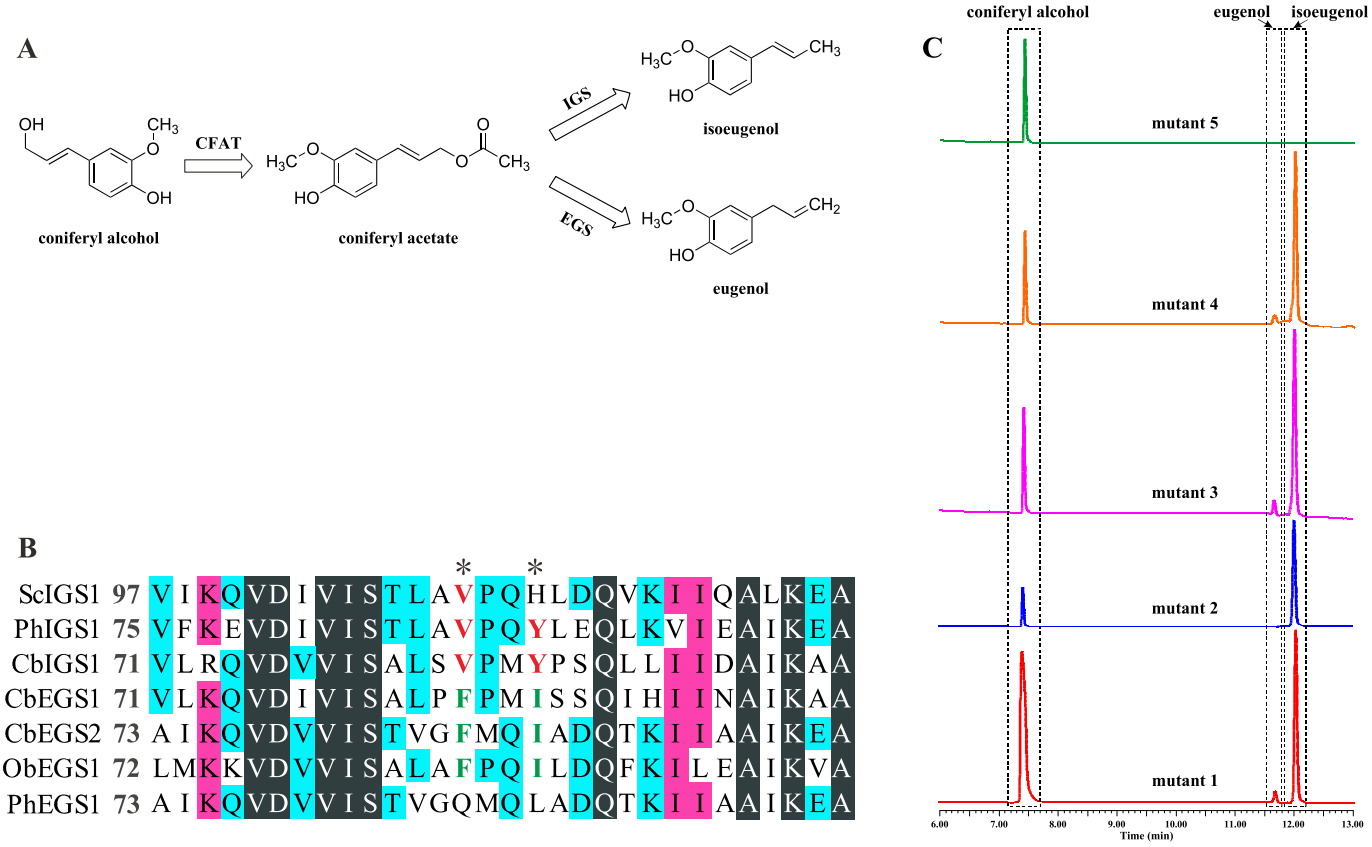

Fig. 7. (A) Biochemical reactions leading to the phenylpropenes isoeugenol and eugenol from coniferyl alcohol. (B) Amino acid sequence alignments of phenylpropene synthases. **The positions of the two amino acids suggested to determine the product specificity are marked by asterisks. The conserved V and Y in IGS enzymes and the conserved F and I in EGS enzymes are shown in red and green, respectively.** (C) Product analyses of reactions catalyzed by the mutations of ScIGS1, using coniferyl alcohol as the substrate.

sequence alignments, evolutionary relatedness, conserved motif composition, and expression profiles were performed. The three candidate *ScPIP* genes encoding **phenylpropene synthases were finally selected and cloned.** The results of *in vivo* functional characterization in *E. coli* demonstrated that ScPIP1 functioned as IGS and was designated as ScIGS1. ScPIP2 and ScPIP4 didn't show the ability to catalyze coniferyl acetate. ScIGS1 was demonstrated to be located in the cytoplasm and nucleus through subcellular localization analysis. Lastly, we also provided insight into the molecular function of ScIGS1 through the 3D model and site-directed mutagenesis experiments.

Table 2
Preferential product formation and the k_{cat} values of ScIGS1 and its mutations.

Mutant no.	Mutations	Product proportions	k_{cat} (S^{-1})
ScIGS1	Wild type	0:100	0.4654
1	V110F	6:94	0.3963
2	H113Y	0:100	0.4352
3	H113I	7:93	0.3991
4	V110F, H113I	3:97	0.3921
5	K157A	0:0	0.4180

SCF has been used as TCM for thousands of years, with characteristic flavors of spicy and fragrant. Dibenzocyclooctadiene lignans, the main active ingredient of SCF, have significant pharmacological activities and broad application prospects. However, current research on SCF has mainly focused on chemical constituents and pharmacological properties, with few reports on the flavor formation mechanism of SCF and the biosynthetic pathway of dibenzocyclooctadiene lignans. Therefore, the identification and characterization of the PIP reductase family and **phenylpropene synthases in this study will** not only assist in clarifying the flavors of SCF but also provide bases for elucidating the biosynthetic pathway of dibenzocyclooctadiene lignans and producing dibenzocyclooctadiene lignans through synthetic biology approaches.

CRediT authorship contribution statement

Tingyan Qiang: Conceptualization, Formal analysis, Data curation, Writing – original draft. **Yu Chen:** Data curation. **Bin Li:** Formal analysis. **Yuqing Dong:** Visualization. **Xueping Wei:** Formal analysis. **Jiushi Liu:** Conceptualization, Investigation. **Bengang Zhang:** Plant identification, Project administration, Writing – review & editing. **Haitao Liu:** Funding acquisition, Project administration, Writing – review & editing. **Peigen Xiao:** Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 82204576) and the CAMS Innovation Fund for Medical Sciences (CIFMS) (No.2022-I2M-JB-009).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2023.11.011.

Supplementary data

Supplementary data to this article can be found online at.

References

[1] Atkinson RG. Phenylpropenes: occurrence, distribution, and biosynthesis in fruit. J Agric Food Chem 2016;66(10):2259–72. https://doi.org/10.1021/acs.jafc.6b04696.
[2] Obeng-Ofori DE, Reichmuth CH. Bioactivity of eugenol, a major component of essential oil of *Ocimum suave* (Wild.) against four species of stored-product Coleoptera. Int J Pest Manag 1997;43(1):89–94. https://doi.org/10.1080/096708797229040.
[3] Chang CL, Cho IK, Li Q. Insecticidal activity of basil oil, *trans*-anethole, estragole, and linalool to adult fruit flies of *Ceratitis capitata, Bactrocera dorsalis*, and *Bactrocera cucurbitae*. J Econ Entomol 2009;102(1):203–9. https://doi.org/10.1603/029.102.0129.
[4] Chang KS, Ahn YJ. Fumigant activity of (*E*)-anethole identified in *Illicium verum* fruit against *Blattella germanica*. Pest Manag Sci 2002;58(2):161–6. https://doi.org/10.1002/ps.435.
[5] Tan KH, Nishida R. Methyl eugenol: its occurrence, distribution, and role in nature, especially in relation to insect behavior and pollination. J Insect Sci 2012;12(56):1–60. https://doi.org/10.1673/031.012.5601.
[6] Koeduka T, Fridman E, Gang DR, Vassão DG, Jackson BL, Kish CM, et al. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. P Natl Acad Sci USA 2006;103(26):10128–33. https://doi.org/10.1073/pnas.0603732103.
[7] Wagner J, Schieberle P, Granvogl M. Characterization of the key aroma compounds in heat-processed licorice (*Succus Liquiritae*) by means of molecular sensory science. J Agric Food Chem 2017;65(1):132–8. https://doi.org/10.1021/acs.jafc.6b04499.
[8] Zeller A, Rychlik M. Character impact odorants of fennel fruits and fennel tea. J Agric Food Chem 2006;54(10):3686–92. https://doi.org/10.1021/jf052944j.
[9] Siano F, Ghizzoni C, Gionfriddo F, Colombo E, Servillo L, Castaldo D. Determination of estragole, safrole and eugenol methyl ether in food products. Food Chem 2003;81(3):469–75. https://doi.org/10.1016/S0308-8146(03)00004-9.
[10] Khalilzadeh E, Hazrati R, Saiah GV. Effects of topical and systemic administration of *Eugenia caryophyllatabuds* essential oil on corneal anesthesia and analgesia. Res Pharm Sci 2016;11(4):293–302. https://doi.org/10.4103/1735-5362.189297.
[11] Chaieb K, Zmantar T, Ksouri R, Hajlaoui H, Mahdouani K, Abdelly C, et al. Antioxidant properties of the essential oil of *Eugenia caryophyllata* and its antifungal activity against a large number of clinical *Candida species*. Mycoses 2007;50(5):403–6. https://doi.org/10.1111/j.1439-0507.2007.01391.x.
[12] Atsumi T, Fujisawa S, Tonosaki K. A comparative study of the antioxidant/prooxidant activities of eugenol and isoeugenol with various concentrations and oxidation conditions. Toxicol Vitro 2005;19(8):1025–33. https://doi.org/10.1016/j.tiv.2005.04.012.
[13] Raja MRC, Srinivasan V, Selvaraj S, Mahapatra SK. Versatile and synergistic potential of eugenol: a review. Pharm Anal Acta 2015;6(5):367–72. https://doi.org/10.4172/21532435.1000367.
[14] Reddy VA, Li CH, Nadimuthu K, Tjhang JG, Jang I. Sweet basil has distinct synthases for eugenol biosynthesis in glandular trichomes and roots with different regulatory mechanisms. Int J Mol Sci 2021;22(2):681. https://doi.org/10.3390/ijms22020681.
[15] Min T, Kasahara H, Bedgar DL, Youn B, Lawrence PK, Gang DR, et al. Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. J Biol Chem 2003;278(50):50714–23. https://doi.org/10.1074/jbc.M308493200.
[16] Wang XQ, He XZ, Lin JQ, Shao H, Chang ZZ, Dixon RA. Crystal structure of isoflavone reductase from Alfalfa (*Medicago sativa* L.). J Mol Biol 2006;358(5):1341–52. https://doi.org/10.1016/j.jmb.2006.03.022.
[17] Gang DR, Kasahara H, Xia ZQ, Mijnsbrugge KV, Bauw G, Boerjan W, et al. Evolution of plant defense mechanisms: relationships of phenylcoumaran benzylic ether reductases to pinoresinollariciresinol and isoflavone reductases. J Biol Chem 1999;274(11):7516–27. https://doi.org/10.1074/jbc.274.11.7516.
[18] Louie GV, Baiga TJ, Bowman ME, Koeduka T, Taylor JH, Spassova SM, et al. Structure and reaction mechanism of basil eugenol synthase. PLoS One 2007;2(10):e993. https://doi.org/10.1371/journal.pone.0000993.
[19] Wang XQ, He XZ, Lin JQ, Shao H, Chang ZZ, Dixon RA. Crystal structure of isoflavone reductase from Alfalfa (*Medicago sativa* L.). J Mol Biol 2006;358(5):1341–52. https://doi.org/10.1016/j.jmb.2006.03.022.
[20] Ü Bayindir, Alfermann AW, Fuss E. Hinokinin biosynthesis in *Linum corymbulosum* reichenb. Plant J 2008;55(5):810–20. https://doi.org/10.1111/j.1365-313X.2008.03558.x.
[21] Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR. Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its own cDNA. J Biol Chem 2003;278(34):31647–56. https://doi.org/10.1074/jbc.M302783200.
[22] Koeduka T, Louie GV, Orlova I, Kish CM, Ibdah M, Wilkerson CG, et al. The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. Plant J 2008;54(3):362–74. https://doi.org/10.1111/j.1365-313X.2008.03412.x.
[23] Umezawa T. Diversity in lignan biosynthesis. Phytochemistry Rev 2003;2(1):371–90. https://doi.org/10.1023/B:PHYT.0000045487.02836.32.
[24] Suzuki S, Umezawa T. Biosynthesis of lignans and norlignans. J Wood Sci 2007;53:273–84. https://doi.org/10.1007/s10086-007-0892-x.
[25] Lee HJ, Cho HI, Lee KE, Kim Y. The compositions of volatiles and aroma-active compounds in dried *omija* fruits (*Schisandra chinensis* Baillon) according to the cultivation areas. J Agric Food Chem 2011;59(15):8338–46. https://doi.org/10.1021/jf200762h.
[26] Choi IY, Lee WH, Lee JJ, Park MJ, Ko JA, Choi JR, et al. Characterization of a *Septobasidium* sp. associated with felt disease of *Schisandra chinensis*. MYCOBIOLOGY 2016;44(1):58–62. https://doi.org/10.5941/MYCO.2016.44.1.58.
[27] Zhai JX, Zhang F, Gao SH, Chen L, Feng G, Yin J, et al. *Schisandra chinensis* extract decreases chloroacetaldehyde production in rats and attenuates cyclophosphamide toxicity in liver, kidney and brain. J Ethnopharmacol 2018;210:223–31. https://doi.org/10.1016/j.jep.2017.08.020.
[28] Lu Y, Chen DF. Analysis of *Schisandra chinensis* and *Schisandra sphenanthera*. J Chromatogr A 2009;1216(11):1980–90. https://doi.org/10.1016/j.chroma.2008.09.070.
[29] Dunkel A, Steinhaus M, Kotthoff M, Nowak B, Krautwurst PDD, Schieberle P, et al. Nature's chemical signatures in human olfaction: a foodborne perspective for

future biotechnology. Angew Chem, Int Ed 2014;53(28):7124–43. https://doi.org/10.1002/anie.201309508.
[30] Zhang F, Zhai JX, Weng N, Gao J, Yin J, Chen WS. A comprehensive review of the main lignan components of *Schisandra chinensis* (North Wu Wei Zi) and *Schisandra sphenanthera* (South Wu Wei Zi) and the lignan-induced drug-drug interactions based on the inhibition of cytochrome P450 and p-glycoprotein activities. Front Pharmacol 2022;13:816036. https://doi.org/10.3389/fphar.2022.816036.
[31] Opletal L, Sovová H, Bártlová M. Dibenzo[*a,c*]cyclooctadiene lignans of the genus *Schisandra*: importance, isolation and determination. J Chromatogr B 2004;812 (1–2):357–71. https://doi.org/10.1016/j.jchromb.2004.07.040.
[32] Szopa A, Ekiert R, Ekiert H. Current knowledge of *Schisandra chinensis* (Turcz.) Baill. (Chinese magnolia vine) as a medicinal plant species: a review on the bioactive components, pharmacological properties, analytical and biotechnological studies. Phytochemistry Rev 2017;16:195–218. https://doi.org/10.1007/s11101-016-9470-4.
[33] Qiang TY, Liu JS, Dong YQ, Mu XL, Chen Y, Luo HM, et al. Identification, molecular cloning, andfunctional characterization of a coniferyl alcohol acyltransferase involved in the biosynthesis of dibenzocyclooctadiene lignans in *Schisandra chinensis*. Front Plant Sci 2022;13:881342. https://doi.org/10.3389/fpls.2022.881342.
[34] Zhang QT, Fan ST, Yang YM, Xu PL, Ai J. De novo transcriptome assembly of *Schisandra chinensis* Turcz. Baill.). Genomics Data 2016;10:153–4. https://doi.org/10.1016/j.gdata.2016.11.008.
[35] Hong CP, Kim CK, Lee DJ, Jeong HJ, Lee Y, Park S, et al. Long-read transcriptome sequencing provides insight into lignan biosynthesis during fruit development in *Schisandra chinensis*. BMC Genom 2022;23(1):1–14. https://doi.org/10.1186/s12864-021-08253-2.
[36] Li HY, Fan CX, Liu JS, Wang B, Li HB. Integration of full-length transcriptomes and anthocyanin metabolite analysis for understanding fruit coloration mechanism in *Schisandra chinensis*. Physiol Mol Biol Plants 2022;28(5):921–33. https://doi.org/10.1007/s12298-022-01179-3.
[37] Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res 2021;49(D1): D412–9. https://doi.org/10.1093/nar/gkaa913.
[38] Eddy SR. What is a hidden Markov model. Nat Biotechnol 2004;22(10):1315–6. https://doi.org/10.1038/nbt1004-1315.
[39] Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 2020;13:1194–202. https://doi.org/10.1016/j.molp.2020.06.009.
[40] Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 2012;40(W1):W597–603. https://doi.org/10.1093/nar/gks400.
[41] Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res 2007;35:W585–7. https://doi.org/10.1093/nar/gkm259.
[42] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30(4):772–80. https://doi.org/10.1093/molbev/mst010.
[43] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35(6):1547–9. https://doi.org/10.1093/molbev/msy096.
[44] Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49(W1):W293–6. https://doi.org/10.1093/nar/gkab301.
[45] Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res 2015;43(W1):W39–49. https://doi.org/10.1093/nar/gkv416.
[46] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 2001;25(4):402–8. https://doi.org/10.1006/meth.2001.1262.
[47] Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–86. https://doi.org/10.1007/BF00228148.
[48] Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356(6364):83–5. https://doi.org/10.1038/356083a0.
[49] Li FR, Yuan L, Lu HZ, Li G, Chen Y, Enggvist MKM, et al. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. Nat Catal 2022;5(8):662–72. https://doi.org/10.1038/s41929-022-00798-z.
[50] Koeduka T. The phenylpropene synthase pathway and its applications in the engineering of volatile phenylpropanoids in plants. Plant Biotechnol 2014;31(5): 401–7. https://doi.org/10.5511/plantbiotechnology.14.0801a.
[51] Kajikawa M, Hirai N, Hashimoto T. A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Mol Biol 2009;69:287–98. https://doi.org/10.1007/s11103-008-9424-3.