Plant Science 344 (2024) 112080

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Functional identification of specialized diterpene synthases from *Chamaecyparis obtusa* and *C. obtusa* var. *formosana* to illustrate the putative evolution of diterpene synthases in Cupressaceae

Tsai-Jung Wu [a], Chi-Chun Lin [a], Li-Ting Ma [b], Chih-Kai Yang [c], Chen-Lung Ho [d], Sheng-Yang Wang [e], Fang-Hua Chu [a,*]

[a] School of Forestry and Resource Conservation, National Taiwan University, Taipei, Taiwan
[b] Academy of Circular Economy, National Chung-Hsing University, Taichung, Taiwan
[c] Department of Forestry, National Pingtung University of Science and Technology, Taipei, Taiwan
[d] Taiwan Forestry Research Institute, Taipei, Taiwan
[e] Department of Forestry, National Chung-Hsing University, Taichung, Taiwan

ARTICLE INFO

Keywords:
Chamaecyparis obtusa var. *formosana*
Chamaecyparis obtusa
Cupressaceae
Diterpene synthase
Gene structure
Co-expression

ABSTRACT

Chamaecyparis obtusa and *C. obtusa* var. *formosana* of the Cupressaceae family are well known for their fragrance and excellent physical properties. To investigate the biosynthesis of unique diterpenoid compounds, diterpene synthase genes for specialized metabolite synthesis were cloned from *C. obtusa* and *C. obtusa* var. *formosana*. Using an *Escherichia coli* co-expression system, eight diterpene synthases (diTPSs) were characterized. CoCPS and CovfCPS are class II monofunctional (+)-copalyl diphosphate synthases [(+)-CPSs]. Class I monofunctional CoLS and CovfLS convert (+)-copalyl diphosphate [(+)-CPP] to levopimaradiene, CoBRS, CovfBRS1, and CovfBRS3 convert (+)-CPP to (-)-beyerene, and CovfSDS converts (+)-CPP to (-)-sandaracopimaradiene. These enzymes are all monofunctional diterpene syntheses in Cupressaceae family of gymnosperm, and differ from those in Pinaceae. The discovery of the enzyme responsible for the biosynthesis of tetracyclic diterpene (-)-beyerene was characterized for the first time. Diterpene synthases with different catalytic functions exist in closely related species within the Cupressaceae family, indicating that this group of monofunctional diterpene synthases is particularly prone to the evolution of new functions and development of species-specific specialized diterpenoid constituents.

1. Introduction

Because plants are sessile organisms, they have evolved many defensive chemicals, including terpenoids. For example, terpenoids participate in the indirect defense of plants (Austel et al., 2016). Many plants interact with carnivores to fight against herbivores forming tritrophic interactions (Heil, 2008). For example, oleoresin of conifers is one of the important compounds used to defend against pathogens and herbivores (Zi et al., 2014; Alicandri et al., 2020). In addition to aiding in physical defense, the non-volatile diterpene resin acid can seal wounds and block the invasion of pathogens (Zi et al., 2014; Alicandri et al., 2020). The complex defense system involving essential oleoresin is considered to be the main reason that conifers can evolve and have become the most abundant group of gymnosperms (Alicandri et al., 2020). In recent years, many diterpenes, such as taxol, have also been found to have high medical value for humans. Taxol can inhibit the normal decomposition of microtubules during cell division and is widely used to treat of ovarian, breast, and lung cancer (Lin et al., 2016).

Diterpenoids are C20 compounds composed of four isoprenyl units. They usually use geranylgeranyl diphosphate (GGPP) as a precursor and are catalyzed by diterpene synthases (diTPSs). In nature, there are currently approximately 12,000 known diterpenoids, which are considered to be ancient and very metabolically diverse (Zi et al., 2014). In plants, in addition to being the precursors of primary metabolites, such as plant hormones, growth regulators, and photosynthetic pigments. GGPP can also be catalyzed into specialized metabolites. It is possible that the diterpenoids of secondary metabolites do not directly affect the growth and development of plants but play an important role

* Corresponding author.
E-mail address: fhchu@ntu.edu.tw (F.-H. Chu).

https://doi.org/10.1016/j.plantsci.2024.112080
Received 9 January 2024; Received in revised form 12 March 2024; Accepted 1 April 2024
Available online 4 April 2024

in the interaction of plants with the environment, herbivores, and plants.

DiTPSs can be roughly separated into three categories according to their functions: bifunctional class II/I diTPSs, monofunctional class II diTPSs and class I diTPSs. Class II diTPSs harbor a catalytic DxDD motif in the γβ-domain, and class I diTPSs possess the catalytic motifs DDxxD and NSE/DTE in the α-domain. Most diTPSs catalyze cyclization at the class II active site first, and the resulting intermediate products are subsequently passed to the class I active site for more complex catalytic reactions to form structurally variable products. Therefore, bifunctional class II/I diTPSs need only a single enzyme to complete the reaction, and the monofunctional diTPSs require two enzymes of class II and class I to carry out the reaction (Zerbe and Bohlmann, 2015).

According to the TPS lineages, modular structures and crystal structures of TPSs, it is speculated that the ancestors of specialized diTPSs are bifunctional diTPSs involved in gibberellin biosynthesis in bryophytes (Alicandri et al., 2020; Jia et al., 2022). Upon the appearance of the evolutionary event in plants, specialized diTPSs underwent multiple rounds of divergence. The diTPSs of gymnosperms and angiosperms belong to different subfamilies. Both bifunctional and monofunctional diTPSs occur in gymnosperms, and only monofunctional diTPSs occur in angiosperms. However, all monofunctional diTPSs are found in Cupressaceae species according to recent studies (Ma et al., 2019; Tasnim et al., 2020; Ma et al., 2021). In contrast to those in Pinaceae, most of the specialized diTPSs in this family are bifunctional. To understand the evolutionary pathway of specialized diTPSs in Cupressaceae and the differentiation between Cupressaceae and Pinaceae, this study focused on finding new diTPSs in *Chamaecyparis obtusa* var. *formosana* and examining the genetic relationships from gymnosperm diTPSs. In this study, tri- and tetracyclic diTPSs from *C. obtusa* var. *formosana* were cloned and functionally characterized to elucidate the mechanism and evolution of specialized diterpene metabolism.

There are five species and one variety of *Chamaecyparis* worldwide, distributed in Japan, North America and Taiwan. *C. formosensis* and *C. obtusa* var. *formosana* are found in Taiwan. *C. obtusa* var. *formosana* is a variety of *C. obtusa*. It is supposed to be originated in Japan and spread to Taiwan through the Ryukyu Islands during the ice age. When the glaciers retreated, the trees remained in humid and rainy high-altitude mountainous areas. Due to the closure of the area, the taxon has undergone long-term isolation and evolution, forming an endemic species to Taiwan. *C. obtusa*, commonly called hinoki, is one of the most popular conifers used for timber production in Japan (Chen et al., 2011; Miyamoto et al., 2013). *C. obtusa* var. *formosana* is also an important native tree in Taiwan, and is distributed at an altitude of 1300–2700 m. Therefore, different taxonomic units, i.e., species and varieties, were also used to realize the evolutionary divergence in this study.

2. Material and methods

2.1. Materials

Plant materials from 30-year-old *C. obtusa*, 13-year-old *C. obtusa* var. *formosana*, and 80-year-old *C. formosensis* were collected from the Chi-Tou Tract of the Experimental Forest at the National Taiwan University. Needles, phloem, and xylem were separated and frozen in liquid nitrogen immediately upon sampling, then stored at −80°C.

Samples of *Taiwania cryptomerioides* and *C. formosensis* utilized for genomic DNA extraction were identical to those described previously (Ma et al., 2019, 2021).

2.2. RNA isolation, cDNA library construction, and identification of diTPSs

Total RNA from the selected tissues was isolated using a modified CTAB method, as referenced in Chang et al. (1993). Subsequently, the total RNA would be purified using the Plant Total RNA Miniprep Purification Kit (GeneMark, Taiwan) along with treatment with DNaseI (ZGene Biotech, Taiwan). The cDNA libraries of leaves and twigs from *C. obtusa* (Co_L, Co_T) and twigs (Covf_T) from *C. obtusa* var. *formosana* were established according to Illumina HiSeq4000 System preparation. Raw reads were uploaded to the SRA database at NCBI (BioProject ID: Co_L: SRR18328269; Co_T: SRR18328270; Covf_T: SRR18328271). *De novo* transcriptome assembly was finished by Trinity (Grabherr et al., 2011). Then the assembly results were overall evaluated by TransRate (Smith-Unna et al., 2016) and BSUCO (Manni et al., 2021). The cDNA library of leaves from *C. obtusa* var. *formosana* (Covf_L) was established by raw reads downloaded at BioProject ID PRJNA506084 from the NCBI SRA (Sequence Read Archive) database at NCBI (Huang et al., 2020). The raw sequence data were trimmed and assembled by CLC Genomics Workbench v12.

All assembled transcripts underwent a search against the NCBI protein non-redundant (NR, https://ftp.ncbi.nlm.nih.gov/blast/db/) database using BLASTX to retrieve their function annotations, focusing on selecting the diTPS candidate genes.

2.3. Cloning of diTPS cDNAs

cDNAs were synthesized from 2 μg of total RNA using SuperScript III reverse transcriptase (Invitrogen, now ThermoFisher Scientific, US). The partial gene sequences were amplified with gene-specific primers using Blend Taq® (Toyobo, Japan) DNA polymerase. Full-length genes were obtained by using 5′ and 3′ rapid amplification of cDNA ends (RACE; Invitrogen, now ThermoFisher Scientific, US) PCR. Full-length genes were confirmed by PCR using Phanta Super-Fidelity DNA Polymerase (Vazyme, PRC). All fragments were cloned into pGEM-T Easy vector (Promega, Madison, WI, USA) and sequencing. All cDNA sequences cloned in this study were submitted to GenBank. The accession numbers are OM988166 (CoCPS), OM988165 (CoLS), OR359412 (CoBRS), OM988164 (CovfCPS), OM988165 (CovfLS), OR359413 (CovfBRS1), OR359414 (CovfBRS2), OR359415 (CovfBRS3), and OR359416 (CovfSDS).

For co-expression, genes required the removal of N-terminal sequences, which might encode plastidial transit peptides. Transit peptides were predicted using ChloroP 1.1 Prediction Server (Emanuelsson et al., 1999). Truncated class I and class II diTPS genes were subsequently subcloned into pET-21a(+) Vector (Novagen, Merck Bioscience) and pCOLADuet-1 Vector (Novagen, Merck Bioscience), respectively. All the diTPS sequences in recombined plasmids were identified by sequencing. Their protein expressions in *Escherichia coli* were confirmed using the western blot method. Primer sequences for this analysis are provided in Table S1.

2.4. Co-expression in E. coli

The co-expression experiments in this study were performed referring to those adopted from published studies (Morrone et al., 2010; Ma et al., 2019; Ma et al., 2021). To enhance diterpene precursor formation, three plasmids were transformed into *E. coli* BL21DE3-C41 strain (Lucigen, Middleton, WI, USA). These three plasmids included a pMBI plasmid (Martin et al., 2003), a GGPP synthase from *T. cryptomerioides* (GenBank #MH894294) cloned into multiple cloning site 2 (MCS2), and a truncated class II diTPS subcloned into multiple cloning site 1 (MCS1) of the pCOLADuet™-1 plasmid (Novagen, Merck Bioscience), along with a truncated class I diTPS subcloned into a pET21a (+) (Novagen, Merck Bioscience). The constructs were transformed into *E. coli* and selected using plates containing 5 μg/mL tetracycline, 15 μg/mL kanamycin, and 25 μg/mL carbenicillin. Colonies were inoculated into 5 mL super optimal broth (SOB) medium supplemented with the aforementioned antibiotics and incubated overnight at 37°C shaking. The following day, cultures were subcultured into fresh medium until reaching an OD_{600} of 0.6. Cultures were then cooled for 1 h at 16°C before inducing with 1 mM IPTG (Sigma-Aldrich, St. Louis, MO, USA). Mevalonate (Sigma-Aldrich, St. Louis, MO, USA) was added to a final concentration of

10 mM in three equal doses every 12 h following the initial IPTG induction, and cultures were maintained at 16°C and 200 rpm for 72 h after the first mevalonate addition. Subsequently, diterpene products were extracted with an equal volume of *n*-hexane at 4°C for 24 h. The extraction was concentrated by evaporating under nitrogen gas and then transferred into amber glass vials for storage at −80°C.

For NMR (nuclear magnetic resonance) analysis, the volume of culture was increased to 1 L; mevalonate was added at 12, 18, and 24 hours after IPTG induction; the extract was concentrated by rotary evaporation.

The plasmids used in each condition are listed in Table S2.

2.5. Gas chromatography-mass spectrometry (GC-MS) analysis

GC-MS analysis was conducted using an Agilent 6890 N GC System (Agilent, US) coupled with an Agilent 5973 Network Mass Selective Detector (Agilent, US). One microliter of each sample was injected into a DB-5MS capillary column (30 m, 0.25 mm inner diameter, Agilent, US) for compound separation. The temperature program was set as follows: initial oven temperature was 60°C, 20°C min $^{-1}$ to 250°C, then 10°C min $^{-1}$ to 270°C, held for 3 min. Injector temperature was 250°C, ion source temperature was 230°C, ionization potential was 70 eV, and carrier gas He at 1 mL/min; scan range 50 – 400 amu.

2.6. NMR analysis

Prior to NMR analysis, the extract was purified by using Silia*Prep*™ Solid Phase Extraction (SPE) Cartridges, Silica, 40–63 μm, 60 Å (SPE-R10030B-06S; Silicycle, Canada) with 0, 5, 10, 15, 20, 25, 100% of ethyl acetate/hexane sequentially. All fractions were analyzed by GC/MS to verify the presence of metabolites and to estimate the quantity by comparison to fixed-concentration sclareol. ≥1 mg of the purified product was dissolved in deuterated chloroform ($CDCl_3$, Sigma-Aldrich, US). Structure analysis used ^{1}H, ^{13}C NMR, HSQC, HMBC, COSY and NOESY spectra acquired with a Bruker Avance 400 MHz FT-NMR spectrometer (Bruker, US) using topspin (Bruker, US).

2.7. Optical rotation measurement

Samples were prepared as in NMR analysis and measured using a JASCO P-2000 polarimeter.

2.8. Genomic DNA extraction and calculation of exon size and intron phase

Genomic DNA was extracted from leaves using the Plant Genomic DNA Purification Kit (GeneMark, Taiwan) following the manufacturer's instructions. Sequences containing introns and exons were obtained using the same primers, enzyme, and vector used in cloning complete diTPS form cDNA. Subsequently, these sequences were aligned with the complete cDNA sequences to characterize the splicing sites.

The primers are described in Table S1, and the enzyme used for PCR is Blend Taq® (Toyobo, Japan). All gDNA sequences cloned in this study were submitted to GenBank. The accession numbers are ON060907 (CoCPS), ON060908 (CoLS), OR359417 (CoBRS), ON060905 (CovfCPS), ON060906 (CovfLS), OR359418 (CovfBRS1), OR359419 (CovfBRS2), OR359420 (CovfSDS), OQ737011 (TcCPS1), OQ737012 (TcCPS2), ON060909 (TcCPS4), ON060910 (TcKSL3), OQ737008 (TcKSL4), OQ737009 (TcKSL6), OQ737010 (TcKSL7), OQ737013 (CfCPS1), OQ737014 (CfKSL1), OQ737015 (CfKSL2), OQ737016 (CfKSL3), OQ737017 (CfKSL4).

The method used for calculating exon size and intron phase in this study was adapted from Trapp and Croteau (2001).

2.9. Phylogenetic analysis

Protein sequence alignments were conducted using MUSCLE (MUltiple Sequence Comparison by Log- Expectation) (Edgar, 2004). Phylogenetic analyses were performed using the maximum-likelihood method with 1000 bootstrap replications in MEGA 11 (Molecular Evolutionary Genetics Analysis version 11) (Kumar et al., 2018). The resulting phylogenetic tree was illustrated by using iTOL (Interactive Tree Of Life) (https://itol.embl.de) (Letunic and Bork, 2021). The protein sequences included in the analysis are listed in Table S3.

3. Results

3.1. Transcriptome mining and diTPS gene cloning

To establish cDNA libraries of leaves and twigs from *C. obtusa* (Co_L, Co_T) and twigs (Covf_T) from *C. obtusa* var. *formosana*, cDNA samples were sequenced on the Illumina HiSeq4000 System and a total of 58.1 M (Co_L), 41.0 M (Co_T) and 45.8 M (Covf_T) clean reads were obtained from 58.8 M (Co_L), 41.5 M (Co_T) and 46.4 M (Covf_T) raw reads, respectively. Trinity was used to perform *de novo* assembly of the clean data, and the assembly results were optimized and evaluated. A total of 43,390 (Co_L), 36,578 (Co_T), and 30,743 (Covf_T) unigenes were obtained, and the average length was 964.18 bp (Co_L), 1029.36 bp (Co_T) and 1183.22 bp (Covf_T), N50 length is 1714 bp (Co_L), 1672 bp (Co_T) and 1881 bp (Covf_T), and the GC percentages were 41.26% (Co_L), 43.70% (Co_T) and 41.69% (Covf_T), respectively.

All unigenes were functionally annotated by NCBI_NR alignment, and a total of 55 diTPS candidate genes were screened out; 25 were derived from the Co_L cDNA library, 13 were derived from the Co_T cDNA library, six were derived from the Covf_L cDNA library, and 11 were derived from the Covf_T cDNA library (Table S4). Candidate genes that participate in specialized metabolism and have relatively complete sequences for cloning were subsequently selected. Through rapid amplification of cDNA ends (RACE), nine candidates were cloned and full-length gene sequences were obtained (Table S5). The full length of the nine diTPS genes ranged from 847 to 863 amino acids (Table S5). Three of the genes were isolated from *C. obtusa*, and six were from *C. obtusa* var. *formosana*. Based on their function described in the following, these genes were termed *CoCPS*, *CoLS*, *CoBRS*, *CovfCPS*, *CovfLS*, *CovfBRS1*, *CovfBRS2*, *CovfBRS3*, and *CovfSDS*, where CovfBRS1, CovfBRS2, and CovfBRS3 share at least 92% identity at the amino acid level (Table S6).

To determine whether the cloned diTPSs were class II or class I diTPSs, the amino acid sequences of nine diTPSs were aligned, and key motifs of diTPSs such as DDxxD, RxR, DDxD, and NSE/DTE were searched (Table S7) (Fig. S1). Among the nine diTPSs, only CoCPS and CovfCPS had class II motif (DxDD), and the class I motifs RxR, DDxxD and NSE/DTE were lost. The other seven genes (*CoLS*, *CoBRS*, *CovfLS*, *CovfBRS1*, *CovfBRS2*, *CovfBRS3*, and *CovfSDS*) had class I motifs RxR, DDxxD, and NSE/DTE. Therefore, we assumed that CoCPS and CovfCPS were monofunctional class II diTPSs and that the others were monofunctional class I diTPSs.

3.2. Use of a co-expression system to identify the function of diTPSs

To identify the functions of diTPSs, *CoCPS* and *CovfCPS* were subcloned without predicted transit peptides into pCOLADuet™ (Novagen, Merck Bioscience), and *CoLS*, *CoBRS*, *CovfLS*, *CovfBRS1*, *CovfBRS2*, *CovfBRS3*, and *CovfSDS* genes were subcloned without predicted transit peptides into pET21a(+) (Novagen, Merck Bioscience) and subsequently transformed into *E. coli* C41 for co-expression. The transportation destination of plant diTPSs is the plastid after the transit peptides are cleaved. The amino acid sequences in different segments of the cloned diTPSs were predicted by the ChloroP 1.1 Server website (Emanuelsson et al., 1999). All the diTPSs identified in this study had

predicted transit peptides. The length of the predicted transit peptide ranged between 45 and 66 amino acids (Table S5). The protein expression was confirmed by western blotting (Fig. S2).

Since most of the substrates of class I diTPSs are products (prenyl diphosphate intermediates) catalyzed by class II diTPSs, the products of class II diTPSs (CoCPS and CovfCPS) were analyzed first. From the BLASTX results for CoCPS and CovfCPS, their functions were predicted to involve the convertion of GGPP to (+)-copalyl diphosphate [(+)-CPP]. According to the results of gas chromatography-mass spectrometry (GC-MS) analysis, CoCPS and CovfCPS converted GGPP to copalol (the product of CPP after dephosphorylation), and the retention time and mass spectrum patterns were the same as those of (+)-CPS [TcCPS4, from *T. cryptomerioides* (Ma et al., 2019)] (Fig. 1 peak 2). To identify its stereostructure, TcKSL3 [levopimaradiene synthase from *T. cryptomerioides* (Ma et al., 2019)], which is stereoselective for CPP and only co-expressed with (+)-CPP but not *ent*-CPP for the synthesis of levopimaradiene, was therefore used. According to the GC-MS analysis results, the co-expression products of CoCPS and CovfCPS coupling with TcKSL3 separately were the same as the co-expression product of (+)-CPS (TcCPS4) coupling with TcKSL3 (Fig. 1 peak 1); thus, the functions of CoCPS and CovfCPS were confirmed to be (+)-CPS. The retention time and mass spectrum patterns of the products of CovfCPS coupled with CovfLS and CoCPS coupled with CoLS were the same as those of the products TcCPS4 coupled with TcKSL3, both of which are levopimaradiene (Fig. 1 peak 1). Therefore, the function of CoLS and CovfLS was to produce levopimaradiene.

CoBRS, CovfBRS1, and CovfBRS3 all resulted in a single new product sharing similar mass spectra at the same retention time of 9.995 min when co-expressed with (+)-CPS (TcCPS4) (Fig. 2, peak 1). However, no product was found for CovfBRS2. The co-expression products of CfKSL3, without activity observed previously (Ma et al., 2021), coupling with (+)-CPS (TcCPS4) were the same as those of CoBRS, CovfBRS1, and CovfBRS3 coupling with (+)-CPS (TcCPS4) in this study. To identify this compound, the product was analyzed by NMR. The ^{1}H NMR and ^{13}C NMR spectra are shown in Figures S3 and S4. The NMR spectral data are ^{1}H (400 MHz) NMR (in $CDCl_3$) δ (ppm) 0.72 (3 H, s), 0.80 (3 H, s), 0.84 (3 H, s), 0.97 (3 H, s), 5.42 (1 H, d, $J = 5.3$ Hz), 5.67 (1 H, d, $J = 5.3$ Hz); ^{13}C (100 MHz) NMR (in $CDCl_3$) δ (ppm) 15.1, 18.6, 20.1, 20.2, 22.0, 25.0, 33.19, 33.23, 33.7, 37.3, 37.4, 39.2, 42.1, 43.6, 49.1, 52.9, 56.1, 61.3, 135.4, 136.2. The product was confirmed to be beyerene by comparing published NMR spectra (Coates and Kang, 1987). The degree of specific optical rotation is $[\alpha]_D^{25}$ −22.8° (*c* 0.1, $CHCl_3$). Thus, the compound was identified as (-)-beyerene. This is the first report of diTPSs producing (-)-beyerene.

GC-MS analysis of the co-expression reaction products of CovfSDS coupling with (+)-CPS (TcCPS4) showed a peak at a retention time of 10.127 min (Fig. 2, peak 2), corresponding to (-)-sandaracopimaradiene, as identified by comparison with published NMR spectra (Tungcharoen et al., 2020) and rotation measurements {$[\alpha]_D^{25}$ −15.6° (*c* 0.1, $CHCl_3$)}. The ^{1}H NMR and ^{13}C NMR spectra are shown in Figures S5 and S6, respectively, and the NMR spectral data are ^{1}H (400 MHz) NMR (in $CDCl_3$) δ (ppm) 0.77 (3 H, s), 0.83 (3 H, s), 0.86 (3 H, s), 1.01 (3 H, s), 2.02 (1 H, m), 2.24 (1 H, dq, $J = 14.0$, 2.0 Hz), 4.86 (1 H, dd, $J = 10.8$, 1.6 Hz), 4.89 (1 H, dd, $J = 17.2$, 1.6 Hz), 5.19 (1 H, s), 5.76 (1 H, dd, $J = 17.2$, 10.8 Hz); ^{13}C (100 MHz) NMR (in $CDCl_3$) δ (ppm) 15.0, 18.8, 19.1, 22.1, 22.6, 26.0, 33.3, 33.8, 34.6, 36.0, 37.4, 38.3, 39.4, 42.2, 50.7, 54.8, 109.9, 128.5, 137.4, 149.2.

Class I diTPS can use prenyl diphosphate intermediates with different stereostructures, but only class II (+)-CPS (CoCPS and CovfCPS) was cloned in this study. Therefore, the labda-13-en-8-ol diphosphate (LPP) product of LPS from *T. cryptomerioides* [TcCPS2 (Ma et al., 2019)] that could serve as substrate for the class I diTPSs found in this study also reacted with class I diTPSs found in this study (Fig. 3). The results showed that peak corresponding to the coupling of TcCPS2 with class I diTPS CoLS or CovfLS occurred at a retention time of 10.322 min (Fig. 3 peak 1), and the product of the mass spectrum was the same as that corresponding to the coupling of TcCPS2 with TcKSL3, both of which are manoyl oxide. There were weak peaks corresponding to manoyl oxide in the chemical profiles of CovfBRS1, CovfBRS2, and CovfSDS coupled with TcCPS2. No other products were observed when TcCPS2 reacted with the class I diTPS CoBRS and CovfBRS3; therefore they did react with LPP under these conditions (Fig. 3). A summary diagram of the functional identification of the diTPSs cloned in this study is shown in Fig. 4.

3.3. Phylogenetic analysis illustrates the correlation between diTPS functions and the relationships of diTPSs

Phylogenetic analysis was performed with the ancient bifunctional diTPS PpCPS/KS (from *Physcomitrella patens*) as the root and the *ent*-CPSs and KSs involved in primary metabolism located in TPS-c and TPS-e/f (Fig. 5). The gymnosperm terpene synthases involved in specialized metabolism were located in TPS-d and could be further separated into TPS-d1, TPS-d2, and TPS-d3, and these three subfamilies were found to function as monoterpene synthases, sesquiterpene synthases, and diTPSs, respectively. The *C. obtusa* and *C. obtusa* var. *formosana* diTPS genes cloned in this study were located within TPS-d3.

The diTPSs of TPS-d3 were separated into several groups according to their functions and relationships, and the genes of Pinaceae and Cupressaceae diTPSs were located in different evolutionary clusters. Even the monofunctional class I diTPSs from Pinaceae with PcmISO1, PbmISO1, and PcmdiTPS3 still grouped with the class II/I bifunctional diTPSs from Pinaceae into another clade rather than agglomerating with monofunctional diTPSs from Cupressaceae. However, these Cupressaceae monofunctional diTPSs belong to class II and are more closely related to members of the Pinaceae than to Cupressaceae monofunctional class I diTPSs. In this group, TcCPS4 (from *T. cryptomerioides*, AFE61356.1), TpdiTPS3 (from *T. plicata*, QND75953), CfCPS1 (from *C. formosensis*, MT275974.1), CoCPS and CovfCPS belong to the same cluster and their functions are all as (+)-CPSs. The monofunctional class I diTPSs in Cupressaceae were grouped into another cluster, that was more diverse than the class II diTPSs. The levopimaradiene synthases were grouped together, including CoLS, CovfLS, CfKSL1 (from *C. formosensis*, QWV53992.1), TpdiTPS2 (from *T. plicata*, QND75952.1), and TcKSL3 (from *T. cryptomerioides*, GU575291.1), which form a cluster different from (-)-beyerene synthases and sandaracopimaradiene synthase. There are four (-)-beyerene synthases, namely, CfKSL3, CoBRS, CovfBRS1 and CovfBRS3, and two (-)-sandaracopimaradiene synthases, namely, CovfSDS and TpdiTPS1.

3.4. Comparative analysis of diTPS gene structure

According to Trapp and Croteau (2001), terpene synthases can be separated into three categories on the basis of their gene structures: all diTPSs from *C. obtusa* and *C. obtusa* var. *formosana* have the characteristics of class I with 12–14 introns (Fig. 6) (Table S8, S9), and there are CDIS (conifer diterpene internal sequence) functional domains in exons 4, 5, and 6. A comparison of the gene structure of Cupressaceae diTPSs with that of the representative gymnosperm specialized metabolic diTPSs revealed that the length of the gDNA from the gymnosperm diTPSs involved in specialized metabolism was approximately 3500–5500 bp. There are 12–15 exons and 12–14 introns in the gDNA of diTPSs. The class II active site is at exon 8, and the class I active site is at exon 12 (calculated based on the ancient typical diTPS GbLS from *G. biloba*) (Fig. 6).

DiTPSs with close relationships in phylogenetic analysis have similar gene structures. Class II diTPSs generally lack intron 1. However, most of class I diTPSs still have the original 14 introns, except for TpdiTPS2 from *T. plicata*, which lost intron 12 and has only 13 introns. In *T. cryptomerioides*, TcKSL4 lost introns 1 and 2 and exons 1 and 2. Therefore, TcKSL4 has a shorter amino acid sequence than other diTPSs. CfKSL2 is

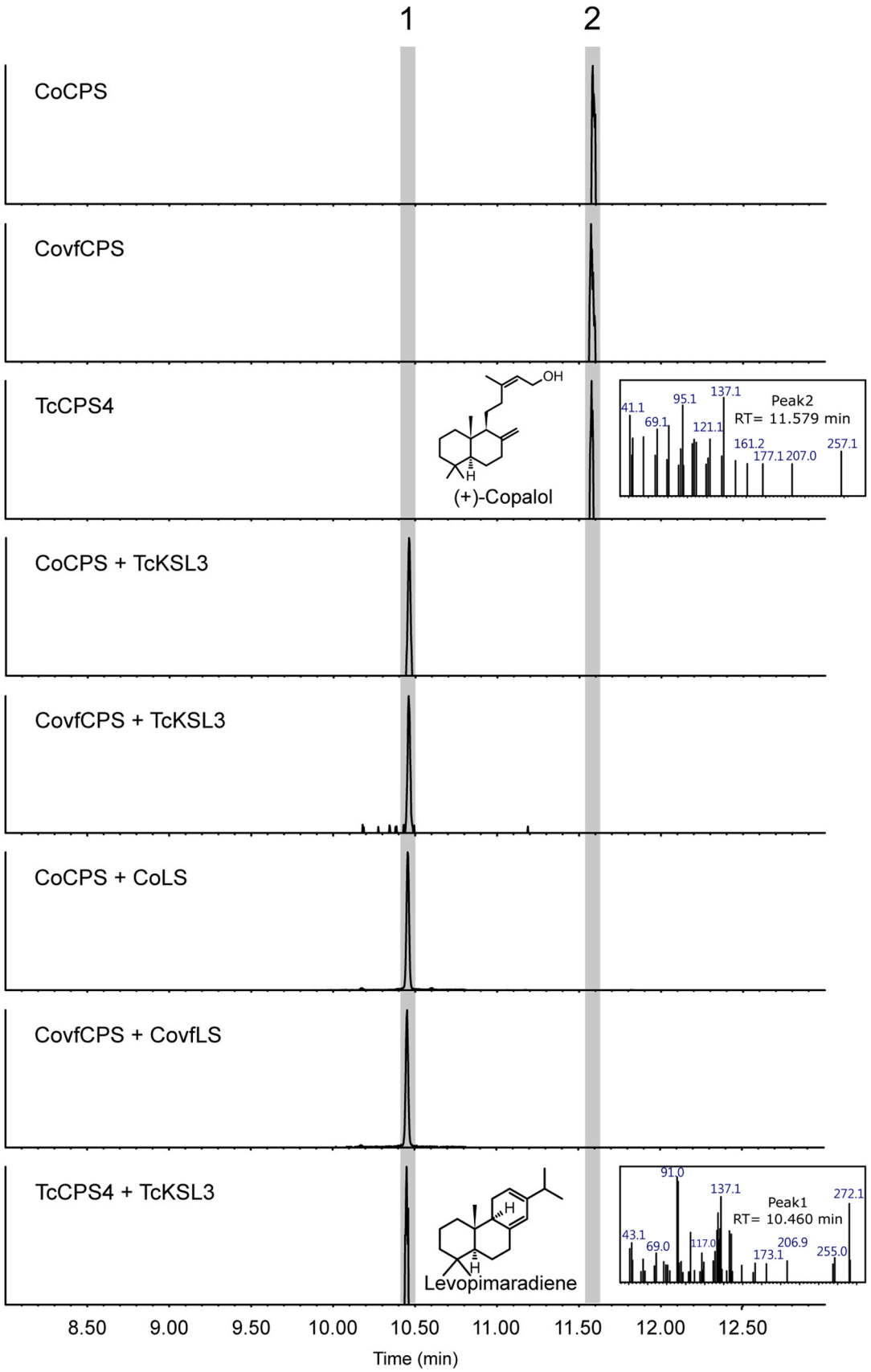

Fig. 1. GC-MS analysis of products extracted from the co-expression assay of the class II diTPSs CoCPS and CovfCPS. The products extracted from the co-expression assay of *T. cryptomerioides* (+)-CPS (TcCPS4, Ma et al., 2019) and coupled with *T. cryptomerioides* levopimaradiene synthase (TcKSL3, Ma et al., 2019) were used as standards. Peak 1: levopimaradiene; Peak 2: (+)-copalol (dephosphorylated product of (+)-CPP). (EIC m/z 91).

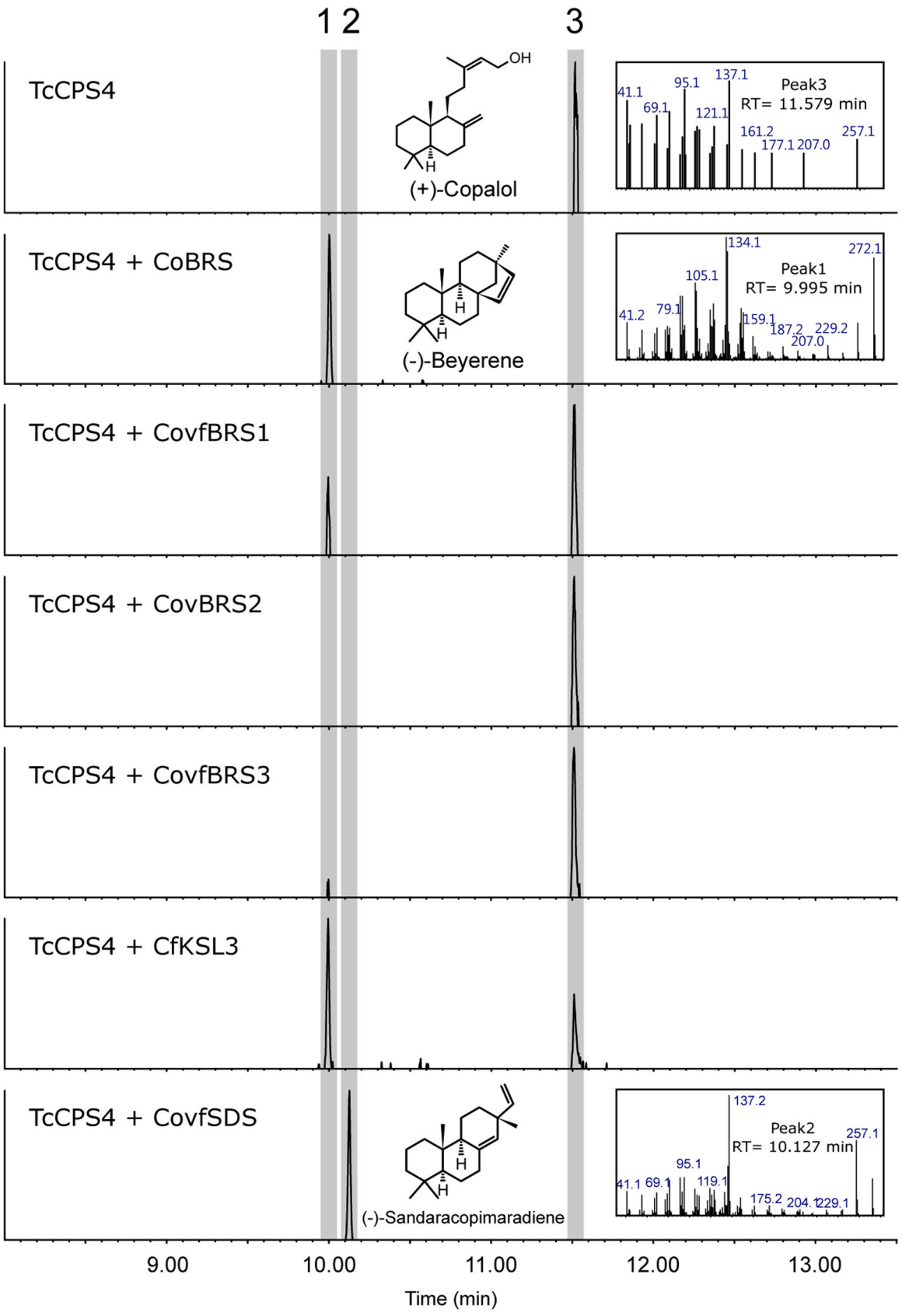

Fig. 2. GC-MS analysis of products extracted from the co-expression assay of *T. cryptomerioides* (+)-CPS (TcCPS4, Ma et al., 2019) coupled with the class I diTPSs CoBRS, CovfBRS1-3, CfKSL3 (Ma et al., 2021) and CovfSDS. Peak 1: (-)-beyerene; Peak 2: (-)-sandaracopimaradiene; Peak 3: (+)-copalol (dephosphorylated product of (+)-CPP). (EIC m/z 257).

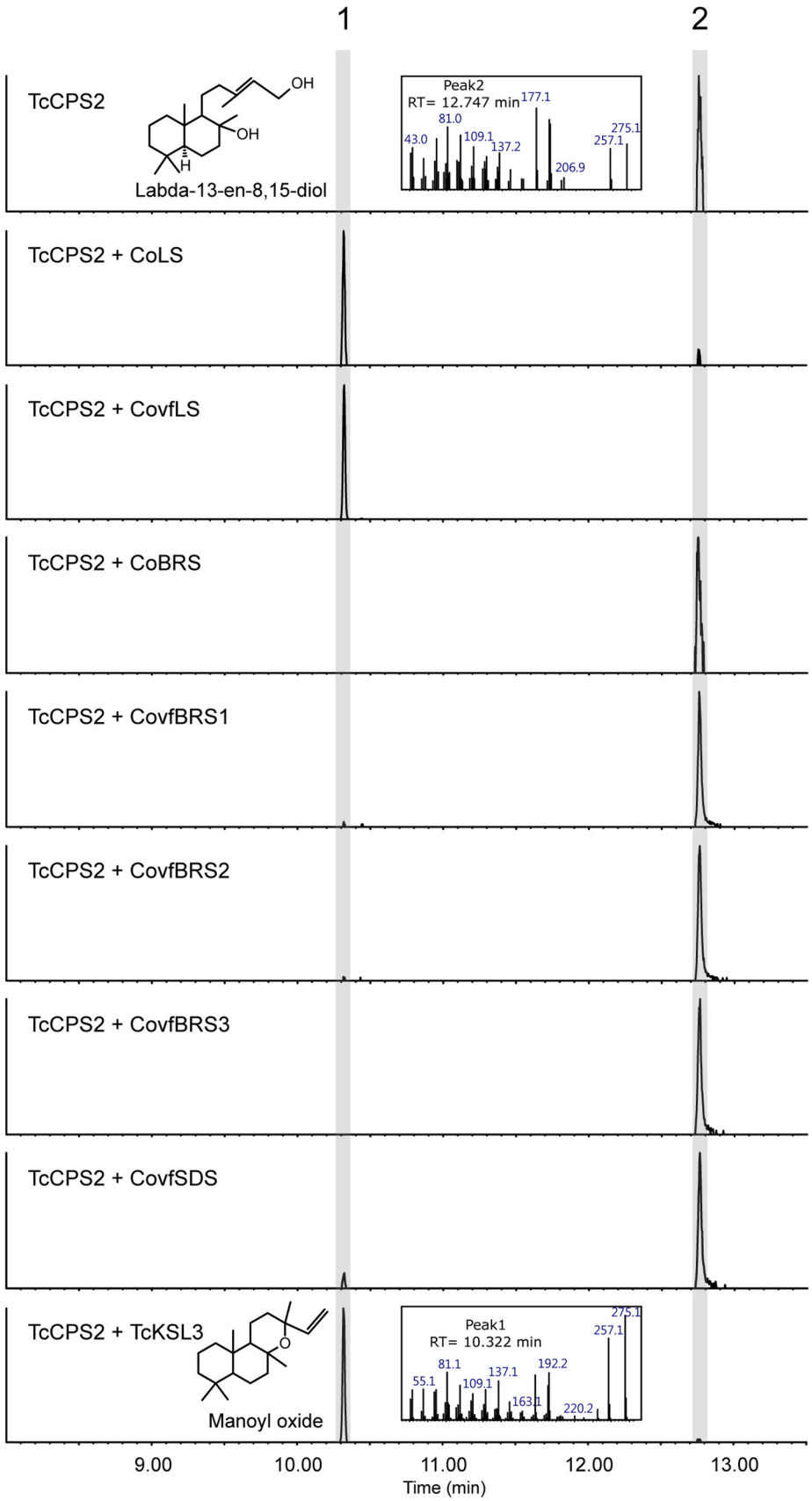

Fig. 3. GC-MS analysis of products extracted from the co-expression of *T. cryptomerioides* LPP synthase (TcCPS2, Ma et al., 2019) with class I diTPSs cloned in this study. The product extracted from the co-expression of TcCPS2 coupled with *T. cryptomerioides* manoyl oxide synthase (TcKSL3, Ma et al., 2019) was used as standard of manoyl oxide. Peak 1: manoyl oxide; Peak 2: labda-13-en-8, 15 diol (dephosphorylated product of LPP). (EIC m/z 257).

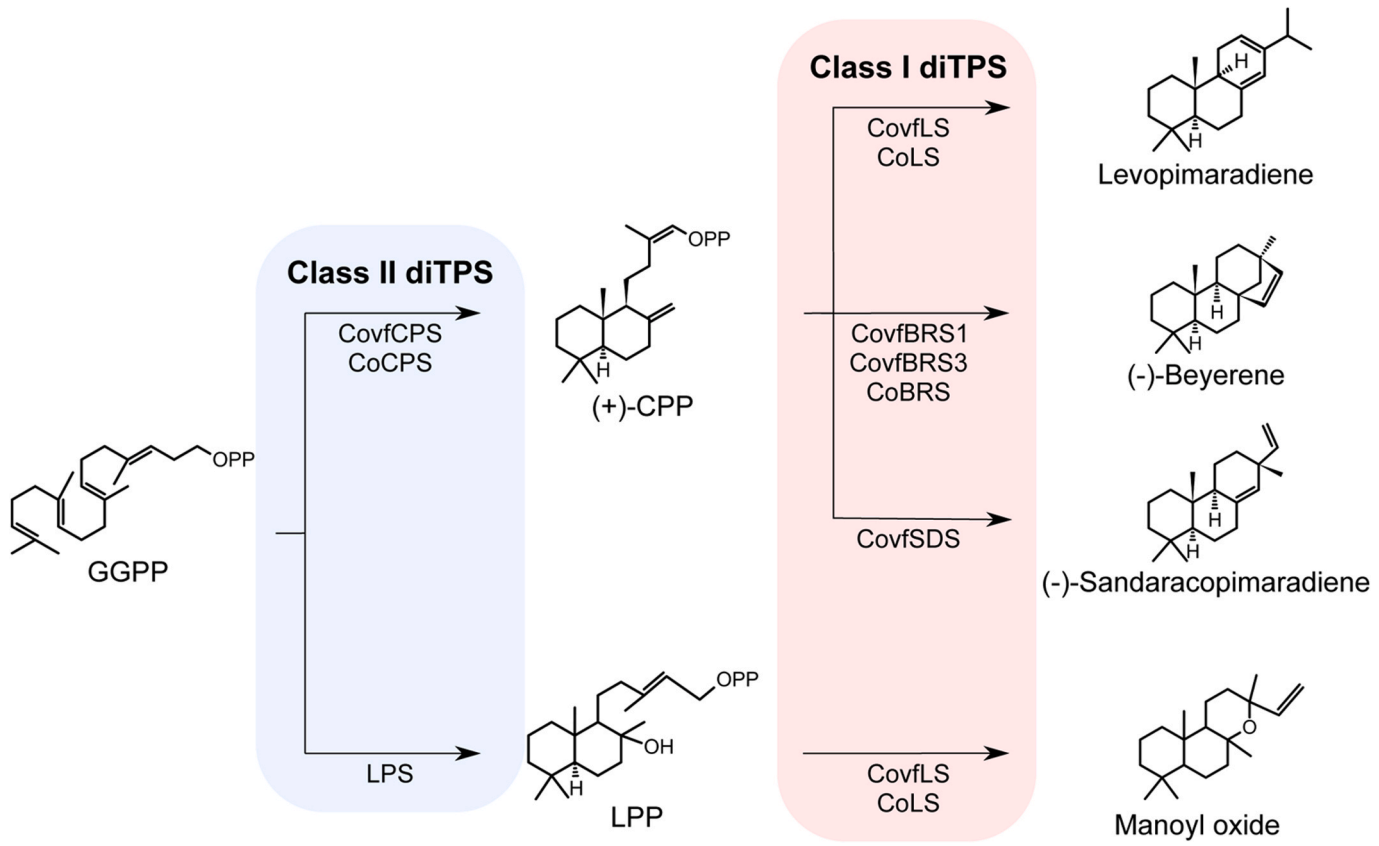

Fig. 4. Summary diagram of the functional identification of the diTPSs identified in this study.

related to TxbreTXS, a taxadiene synthases from *Taxus*, and both of these genes have lost introns 1 and 2.

The detailed sizes of introns and exons in the diTPSs are shown in Table S9. We found that the gene structures of levopimaradiene synthase and (+)-copalyl diphosphate synthase show strong relatedness with exon size in Cupressaceae. For both levopimaradiene synthase and (+)-copalyl diphosphate synthase, the sizes of the exons were almost the same, except for the N-terminal region (Fig. 7). The major difference in intron size was detected. With respect to the distance between taxonomic units, the more closely related the species are the greater the shared identity of their gene structures. In particular, the exact same gene structure of levopimaradiene synthase from *C. obtusa* var. *formosana* and *C. obtusa* may result from *C. obtusa* var. *formosana* being a variety of *C. obtusa*.

4. Discussion

4.1. The evolution of specialized diTPSs in Cupressaceae plants

Early studies suggested that the biosynthesis of labdane-type diterpenes in the primary metabolism of conifers is carried out by two monofunctional diTPSs, while the biosynthesis of specialized labdane-type diterpenes involves primarily bifunctional diTPs (Karunanithi and Zerbe, 2019). However, recent research has shown that in the Cupressaceae family, both primary and specialized labdane-type diterpene metabolism involves only monofunctional diTPSs (Ma et al., 2019, 2021; Tasnim et al., 2020). According to the results of the phylogenetic analysis (Fig. 5), the primary metabolism class II and class I diTPSs responsible for synthesizing *ent*-kuarene in Cupressaceae were classified as TPS-c and TPS-e/f, respectively, similar to the fundings for other plants. The class II genes responsible for specialized metabolism in Cupressaceae belong to TPS-d and form a distinct branch. CovfCPS and CoCPS, which were cloned in this study, are also located on this branch. The neighboring branch consists of bifunctional and class I monofunctional diTPSs from Pinaceae, while the bifunctional diTPSs GbLS from ginkgo is located at the base of these two branches. This finding suggested that the class II specialized diTPSs in Cupressaceae evolved from bifunctional diTPSs by losing class I activity. On the class II specialized metabolism branch, TcCPS4 from *T. cryptomerioides* is more closely related to other species' terpene synthases involved in the synthesis of (+)-CPP, while TcCPS2 for LPP synthesis from the same species is more distant, indicating conservation of the (+)-CPP synthase within the Cupressaceae family.

The class I diTPSs responsible for specialized metabolism in Cupressaceae are distributed on different branches. In *T. cryptomerioides*, TcKSL1 and TcKSL2 belong to TPS-e/f, which are thought to have evolved in parallel from the primary metabolism agent *ent*-kuarene synthase, similar to the class I diTPSs involved in specialized labdane-type diterpene metabolism in angiosperms (Ma et al., 2019; Karunanithi and Zerbe, 2019). There are no specialized class I diTPSs of TPS-e/f found in conifers. Most class I diTPSs involved in specialized metabolism in Cupressaceae form a distinct branch belonging to TPS-d. These terpene synthases are separated from the class I diTPSs in Pinaceae and the class II specialized diTPSs in Cupressaceae, indicating their independent functional evolution. CoLS and CovfLS cluster with other Cupressaceae levopimaradiene synthases, while the beyerene synthases and sandaracopimaradiene synthases cluster separately with their respective enzymes. TcKSL3 shares a sequence identity of 85% with CoLS, both functioning as levopimaradiene synthases. Conversely, TcKSL5, producing phyllocladanol, exhibits a lower sequence identity of 75% with CoBRS, which yields a different product. This suggests that lower sequence similarity may indicate reduced selective pressure, potentially increasing the likelihood of functional changes. Indeed, levopimaradiene is quite important in gymnosperms. Levopimaradiene gives rise to many tricyclic abietane-type diterpene specialized metabolites in gymnosperms, including Pinaceae plants and ginkgo plants.

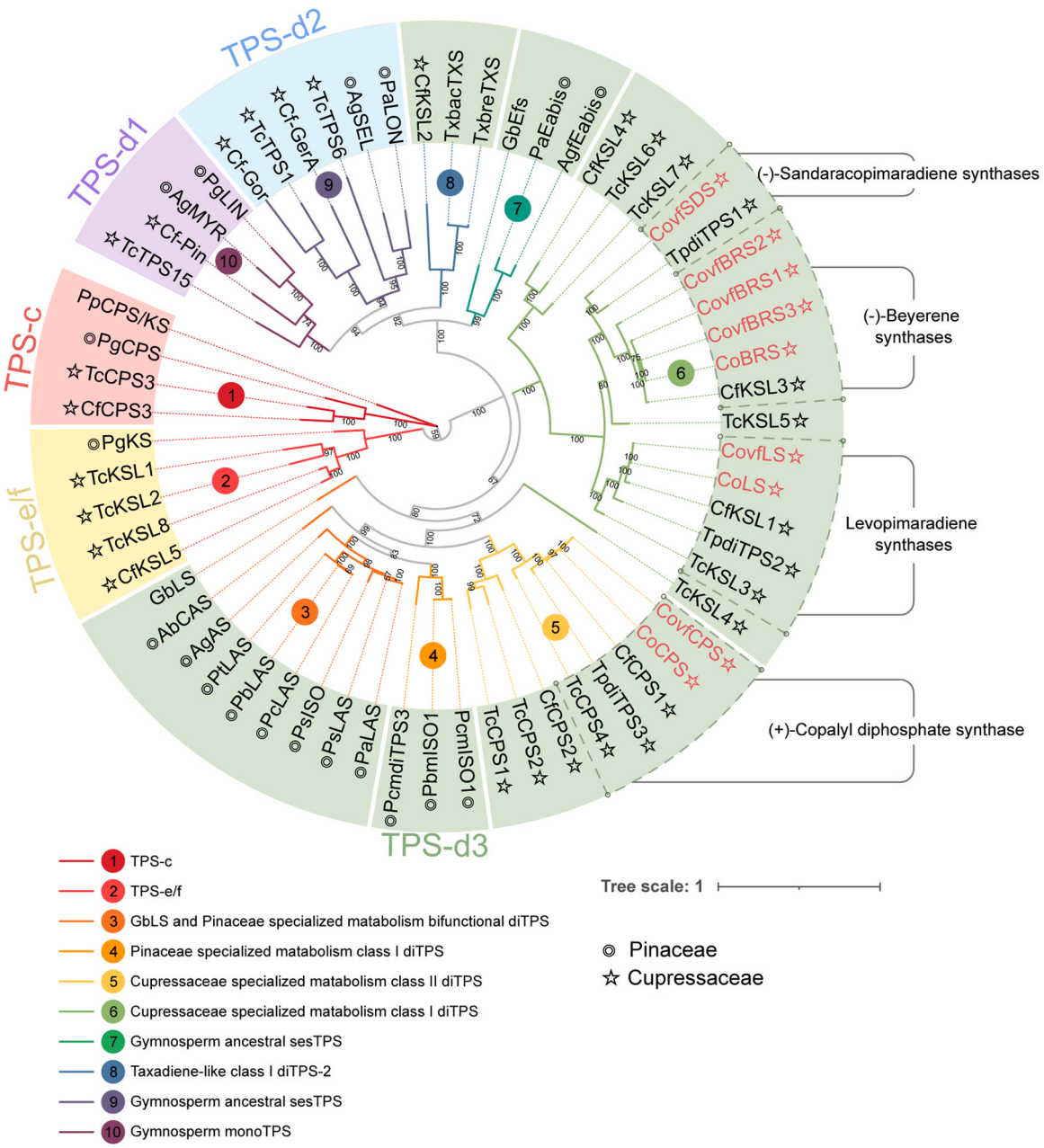

Fig. 5. Phylogenetic analysis of diTPSs. The bifunctional diTPS *ent*-CPS/*ent*-kaurene from *Physcomitrella patens* (PpCPS/KS) was used as the tree root. The genes marked in red were identified in this research. A circle in front of the genes represents Pinaceae members; a star in front of the genes represent Cupressaceae members. Information on amino acid sequences used in Fig. 5 is listed in Table S3. Ab, *Abies balsamea*; Ag, *Abies grandis*; Cf, *Chamaecyparis formosensis*; Co, *Chamaecyparis obtusa*; Covf, *Chamaecyparis obtusa* var. *formosana*; Gb, *Ginkgo biloba*; Pa, *Picea abies*; Pb, *Pinus banksiana*; Pc, *Pinus contorta*; Pg, *Picea glauca*; Pp, *Physcomitrium patens*; Ps, *Picea sitchensis*; Pt, *Pinus taeda*; Tc, *Taiwania cryptomerioides*; Tp, *Thuja plicata*; Txbac, *Taxus baccata*; Txbre, *Taxus brevifolia*.

Despite their low proximity, even though they produce levopimaradiene or abietadiene through bifunctional diTPSs, these genes are highly transcribed (Hall et al., 2013). Alternatively, tetracyclic phyllocladanol has been identified exclusively in Cupressaceae, Podocarpaceae, and Araucariaceae, with no reported occurrences in Pinaceae. The occurrence of tetracyclic beyerene in Cupressaceae is limited to closely related species of the genera *Thujopsis*, *Thuja*, and *Chamaecyparis* (Takahashi et al., 2001; Yang et al., 2012; Küpeli Akkol et al., 2015; Ma et al., 2019; Yasutomi et al., 2021).

According to the current research, bifunctional diTPSs do not exist in Cupressaceae plants. Instead, monofunctional diTPSs are involved in both primary and specialized metabolism, which is different from what has been observed in gymnosperms, where bifunctionaldiTPSs were primarily used for specialized metabolism. Bifunctional diTPSs lack an intermediate channel between the reaction domains of class II and class I. After the product of the class II reaction diffuses out of the class II reaction domain, the synthase enters the class I reaction domain. Therefore, diTPSs do not exhibit strong metabolic channeling, and they do not have significant advantages in terms of reaction efficiency compared to monofunctional enzymes (Zhou et al., 2012a; Hall et al., 2013). In other words, bifunctional diTPSs have the advantage of co-localization, meaning that the class II and class I reaction steps always occur at the same cellular location, ensuring the completion of both reaction steps. However, the advantage of monofunctional enzymes lies in their ability to expand metabolite diversity easily. Class II monofunctional diTPSs need to convert GGPP into a series of bicyclic

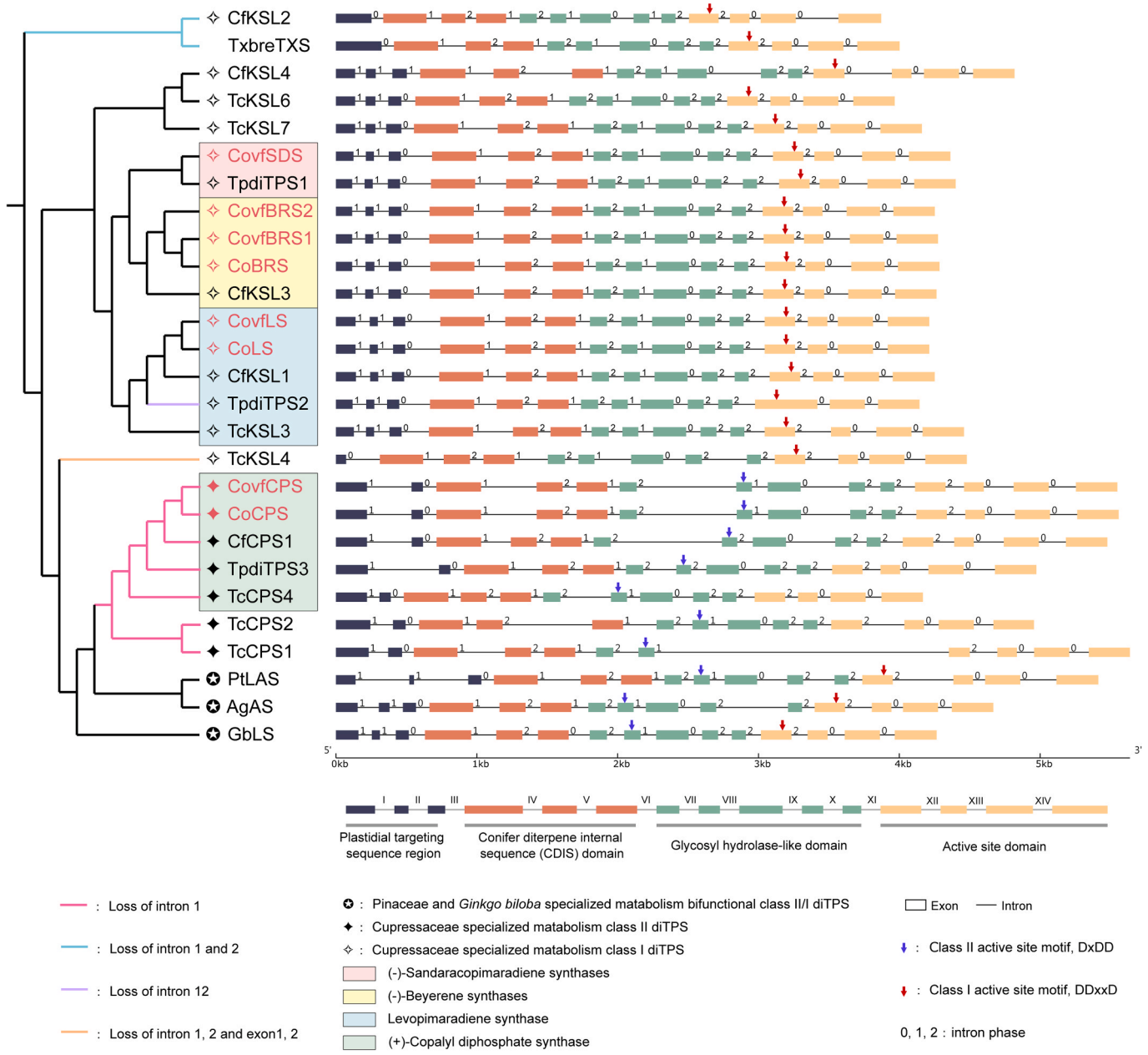

Fig. 6. Gene structures of the gymnosperm diTPSs. The different symbols indicate that the genes are from different species and have different functions. The blue arrow indicates the location of the class II active site motif, and the red arrow indicates the location of the class I active site motif. The number above the line represents the intron phase. Information on the cDNA and gDNA sequences used in Fig. 6 is listed in Table S8.

prenyl diphosphates, and class I monofunctional enzymes may subsequently react with different bicyclic prenyl diphosphates, leading to a broader range of potential products (Karunanithi and Zerbe, 2019). The absence of tetracyclic diterpenes such as phyllocladanol and beyerene in Pinaceae plants compared to Cupressaceae plants may also be attributed to the greater evolutionary potential of class I monofunctional enzymes in acquiring new functions.

Another advantage of monofunctional enzymes is their flexible regulation, which allows for more efficient resource allocation. Enzyme kinetics studies of the bifunctional diTPSs AgAS have shown that the reaction catalyzed by the class I active site is the rate-limiting step (Peters et al., 2000). This finding indicates that the ideal ratio between class II and class I activities is not 1:1, and the use of monofunctional diTPSs allows for modulation of the ratio between class II and class I enzymes. Terpenoid compounds are important components of conifer resin, and Pinaceae, an early diverging lineage within the order Pinales, rely more on resin as a defense element. The resin secretion in Pinaceae is constitutive, while that in other plants, such as Cupressaceae plants, tends to be inducible (Hudgins et al., 2004). In accordance with the differences in resin-secreting tissues, the co-localization of bifunctional diTPSs in Pinaceae ensures the maintenance of a basal level of constitutive resin, while the flexible regulation of monofunctional diTPSs in Cupressaceae allows for better adaptation to inducible resin production.

4.2. Comparison of the amino acid residues in the diTPSs active site and the inferred function

DxDD is a relatively conserved motif in class II diTPSs, especially the third amino acid aspartate (D), which is the most important for the cyclization of class II diTPSs (Köksal et al., 2014). CoCPS, CovfCPS and other (+)-CPSs are conserved in DxDD, and both are DIDD (Table S10). In addition, the His-Asn catalytic dyad is quite conserved in *ent*-CPS

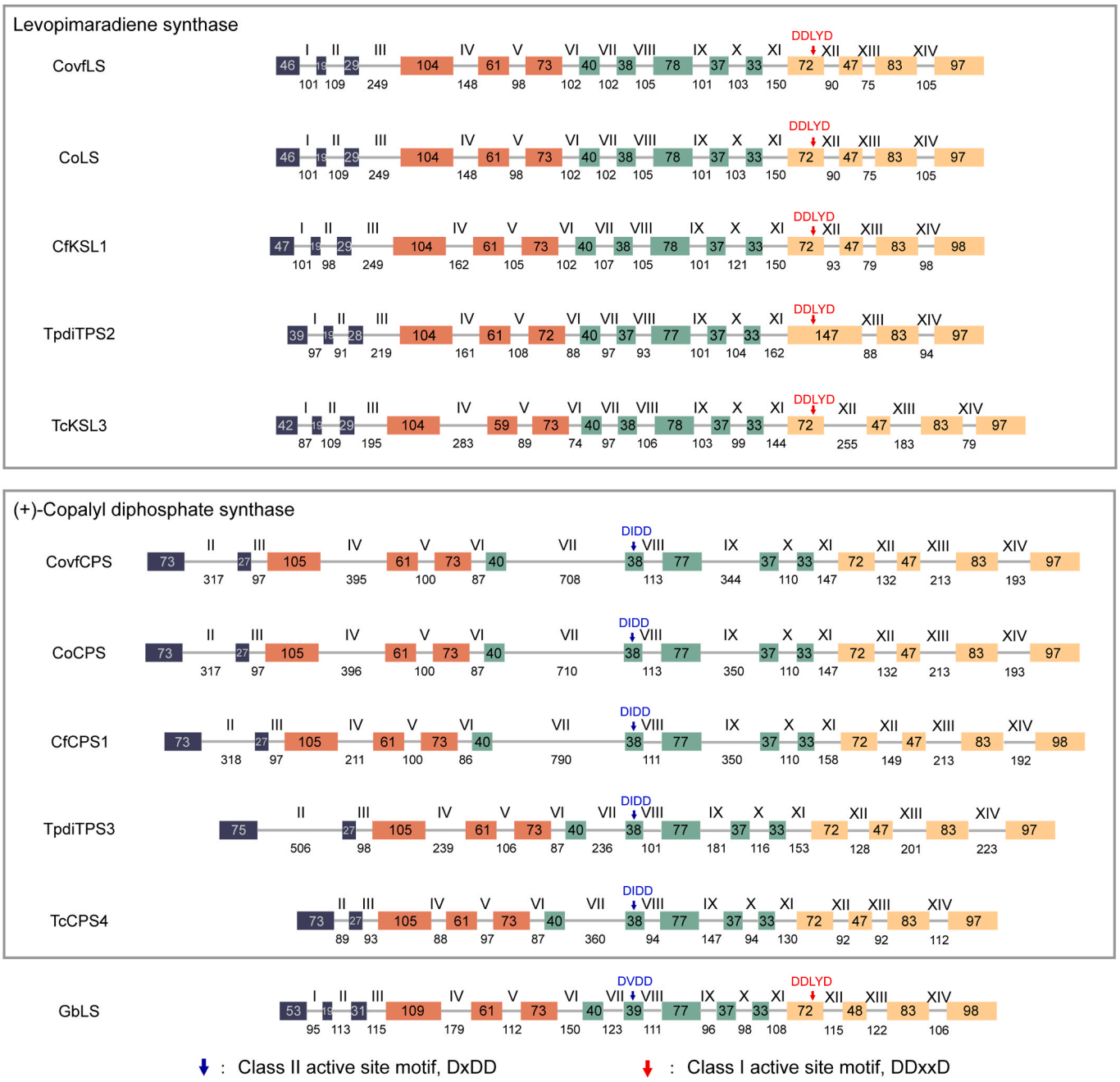

Fig. 7. Gene structures of Cupressaceae levopimaradiene synthase and (+)-copalyl diphosphate synthase and GbLS (from *Ginkgo biloba*). The blue arrow indicates the location of class II active site motif; and a red arrow indicates the location of the class I active site motif.

(Potter et al., 2014; Cui et al., 2015; Pelot et al., 2018; Ma et al., 2019), and both catalytic sites are H259 and N318 [take maize ZmAN2 (AAT70083.1) as an example, Murphy et al., 2018]. In the context of (+)-CPS, it has been observed that in the bifunctional diTPS AgAs in *Abies grandis*, which catalyzes the conversion of GGPP to the intermediate (+)-CPP, and subsequently yields multiple products by mutation analysis. Notably, residues Y287 and H348 play pivotal roles in directing the product towards (+)-CPP. (Criswell et al., 2012; Mafu et al., 2015). These two sites in other (+)-CPSs are also conserved, and this amino acid dyad was not found in diTPSs with other class II diTPS functions. CoCPS and CovfCPS Y-H are in the same position as other (+)-CPSs, so their amino acid sequences correspond to the function of (+)-CPS. SdCPS2 is a CLPS (clerodienyl diphosphate synthase) that catalyzes the transformation of GGPP to clerodienyl diphosphate (CLPP) in *Salvia divinorum* (Pelot et al., 2017). By changing the active site amino acid residue, different products can be produced; for example, if F255 and N313 are substituted with A or W360 is substituted with A, LPP will be generated (Pelot et al., 2017). PvCPS3, which produces (+)-8,13-CPP in switchgrass (*Panicum virgatum*), can also produce prenyl diphosphate intermediates of isomers if mutated at F284 and F331 (Pelot et al., 2018). Therefore, these sites are also important for accessing monofunctional class II diTPSs.

For class I catalytic activity, four residues were found to be critical for the carbocation intermediate rearrangement for bifunctional diTPSs of European spruce (*Picea abies*) (PaLAS: W679, Y686, A713, V717) (Keeling et al., 2008). The point mutation of these residues led to product switching between pimarane- and abietane-type diterpenes. A comparison of several important conserved sequences, revealed that the N-terminal sequence of the Cupressaceae (+)-CPS protein contains the KR(E/D)x_5W conserved sequence previously found in Pinaceae diTPSs

(Keeling et al., 2008). However, the KR(E/D)x_5W conserved sequence was not found in Cupressaceae class I diTPSs or GbLS (from *G. biloba*).

In addition, the conserved sequences of some active sites among Cupressaceae, Pinaceae, and ginkgo levopimaradiene synthase are also different. Cupressaceae (+)-CPS does not conform to the residues that RxR, a class I motif (DDxxD), NSE/DTE should have, or does it conform to the class I catalytic residues that affect the function of levopimaradiene synthase. The RxR and NSE/DTE of Cupressaceae class I levopimaradiene synthase are slightly different from those of Pinaceae levopimaradiene synthase in terms of conserved amino acids. Notably, the class I catalytic residues of class I levopimaradiene synthase in Cupressaceae are replaced by a threonine (T716) at the position of valine (V), but the protein still retains the ability to synthesize levopimaradiene (Table S7).

In this study, the functionally characterized CovfSDS was found to be a (-)-sandaracopimaradiene synthase. Based on the results of previous studies, the carbocation rearrangement mechanism of these terpene synthase reactions was proposed, and the key residues that may influence these reactions were discussed. After (+) CPP reaches the class I active site, the DDxxD motif deprotonates diphosphate, generating the C8-sandaracopimarenyl cation intermediate, which is subsequently deprotonated to become sandaracopimaradiene (Keeling et al., 2008). In *P. abies*, PaISO and PaLAS share a sequence identity of 91%, while PaISO synthesizes isopimaradiene and PaLAS synthesizes four abietane-type diterpenes (Keeling et al., 2008). For PaLAS, a single amino acid mutation from A733 to S733 results in complete conversion of products to isopimaradiene and sandaracopimaradiene. Similar results were obtained from point mutation experiments on the grand fir AgAS. If the four amino acids of PaISO are mutated to be the same as those of PaLAS, 93% of the products are converted to abietane-type diterpenes (Wilderman et al., 2007; Keeling et al., 2008). Compared with these four key residues, the monofunctional levopimaradiene synthases and sandaracopimaradiene synthases in Cupressaceae, three out of the four key amino acids are conserved in these monofunctional terpene synthases, with only one difference (Table S7). The amino acid Y685 in levopimaradiene synthases and F681 in sandaracopimaradiene synthases may contribute to the functional differences observed in the two pathways of these monofunctional diTPSs. In addition, CovfSDS features residues identical to those of (-)-sandaracopimaradiene synthase (TpdiTPS1; *Thuja plicata*) (Tasnim et al., 2020), at the corresponding site, indicating that its potential functions are the same (Table S7).

The enzyme capable of converting (+)-CPP to beyerene was discovered in this study. To date, there have been no studies specifically investigating the enzymatic formation of (-)-beyerene. In this study, CoBRS, CovfBRS1, CovfBRS3, and CfKSL3 were found to be (-)-beyerene synthases using (+)-CPP as a substrate. However, considering that enantiomers have the same energy, the process of converting *ent*-CPP to *ent*-beyerene was investigated. In the biosynthetic mechanism of beyerene, CPP loses diphosphate to form the pimarenyl carbocation through initial cyclization, followed by a second cyclization to form the beyeranyl cation. After deprotonation of the beyeranyl cation, beyerene is formed (Hong and Tantillo, 2010). The class I diTPSs OsKSL5i and OsKSL5j from two subspecies of rice, indica (*Oryza sativa* subsp. indica) and japonica (*O. sativa* subsp. japonica), produce tetracyclic *ent*-isokaurene and tricyclic *ent*-pimaradiene, respectively (Xu et al., 2007; Jia et al., 2017). Point mutation experiments on OsKSL5i and OsKSL5j revealed a single residue that can interchange the products of the two enzymes. When the I residue in OsKSL5i is mutated to T, the product changes to tricyclic *ent*-pimaradiene. Similarly, when the T residue in OsKSL5j is replaced by I, it produces tetracyclic *ent*-isokaurene. Moreover, point mutation experiments on OsKSL4 and Arabidopsis ent-kaurene synthase indicated the importance of this residue in the second cyclization (Xu et al., 2007; Jia et al., 2017). In castor bean (*Ricinus communis*), RcKSL4 is an *ent*-beyerene synthase, the corresponding amino acid at the position is V, and even when replaced by I, it still predominantly produces *ent*-beyerene. However, when replaced by T, the production of *ent*-beyerene decreases (Jia et al., 2017). OsKSL2 primarily produces *ent*-beyerene, and the corresponding amino acid at this position is also V. However, there is no direct experimental evidence supporting the influence of this residue on the reactivity of OsKSL2.

The opposite situation seems to occur for the diTPSs in the Cupressaceae family members that produce tetracyclic diterpenes. TcKSL5 produces the product phyllocladanol, while CoBRS, CovfBRS1, and CovfBRS3 produce (-)-beyerene. The residue at this critical position is T710 (Table S7), which is the same as OsKSL5j, the synthase that produces tricyclic diterpenes. On the other hand, inactive CovfBRS2 has V710, which is the same as the residue found in the *ent*-beyerene. Given that both OsKSL5 and *ent*-beyerene synthase utilize *ent*-CPP as their substrate, it is possible that the crucial reaction residue differs from that of diTPSs in the Cupressaceae family, which use (+)-CPP. Terpene synthases constitute a large gene family believed to have arisen from multiple gene duplications and neofunctionalization (Keeling et al., 2008). Our study revealed significant differences in the crucial reaction residues among these specialized diTPSs in the Cupressaceae family, providing valuable insights into their evolutionary pathways. Notably, our investigation identified three (-)-beyerene synthase genes with amino acid identities ranging from 92% to 97%. Intriguingly, while CovfBRS1 and CovfBRS3 perform the same function, the CovfBRS2 appears to be non-functional.

4.3. The possible evolution of diTPSs for gene structure in Cupressaceae

According to the gene structure analysis (Fig. 6), intron deletions are common events in the evolution of diTPSs. For example, all Cupressaceae monofunctional class II diTPSs lost intron 1; the class I diTPSs TcKSL4 and CfKSL2 lost introns 1 and 2; and *T. plicata* levopimaradiene synthase (TpdiTPS2) lost intron 12. TpdiTPS1, which belongs to the *T. plicata* class I diTPSs, retained 14 introns, and the other Cupressaceae levopimaradiene synthases examined in this study also retained 14 introns (CoLS, CovfLS, CfKSL1, TcKSL3). Therefore, the loss of intron 12 in TpdiTPS2 is more likely to have occurred less recently in the evolution process than that in *T. plicata*; however, more detailed genomic data on *Thuja* or other genera in the Cupressaceae are needed to determine whether only *T. plicata* lost intron 12 or other diTPSs from *Thuja*.

This study analyzed the evolution of Cupressaceae diTPSs through phylogenetic analysis and gene structure analysis. The results of these two methods represent the relationships between the full-length amino acid sequences and the distributions of introns and exons. The evolution of the Cupressaceae specialized diTPSs should be due to the long-term replication and mutation of ancient bifunctional diTPSs and the loss of class I or class II active sites. In the present study, Cupressaceae specialized diTPSs differentiated into several different evolutionary pathways, each of which has similar functions. The differences in the evolutionary pathways of each diTPS group are described in detail below (Fig. 8).

The evolution of Cupressaceae monofunctional class I and class II diTPSs may follow different paths. According to the phylogenetic analysis, Cupressaceae monofunctional class II diTPSs are located in the same cluster. Compared with the Cupressaceae monofunctional class I diTPSs, they are more similar to the Pinaceae bifunctional diTPSs (Fig. 6). The evolution of Cupressaceae class II diTPSs is relatively similar to that of bifunctional Pinaceae diTPSs, which was also shown by phylogenetic analysis and conserved sequences. Both Cupressaceae monofunctional class II diTPSs and Pinaceae bifunctional class II/I diTPSs have KR(E/D)x_5W, but Cupressaceae monofunctional class I diTPSs lack KR(E/D)x_5W.

According to the gene structure analysis, most class I diTPSs still contained 14 introns, but there are still some exceptions. The evolutionary pathway of TcKSL4 is closely related to the evolution of Pinaceae bifunctional diTPSs and Cupressaceae class II diTPSs. The introns 1 and 2 and exons 1 and 2 were lost, and the predicted transport peptides were relatively short. The N-terminus of the amino acid sequence is relatively

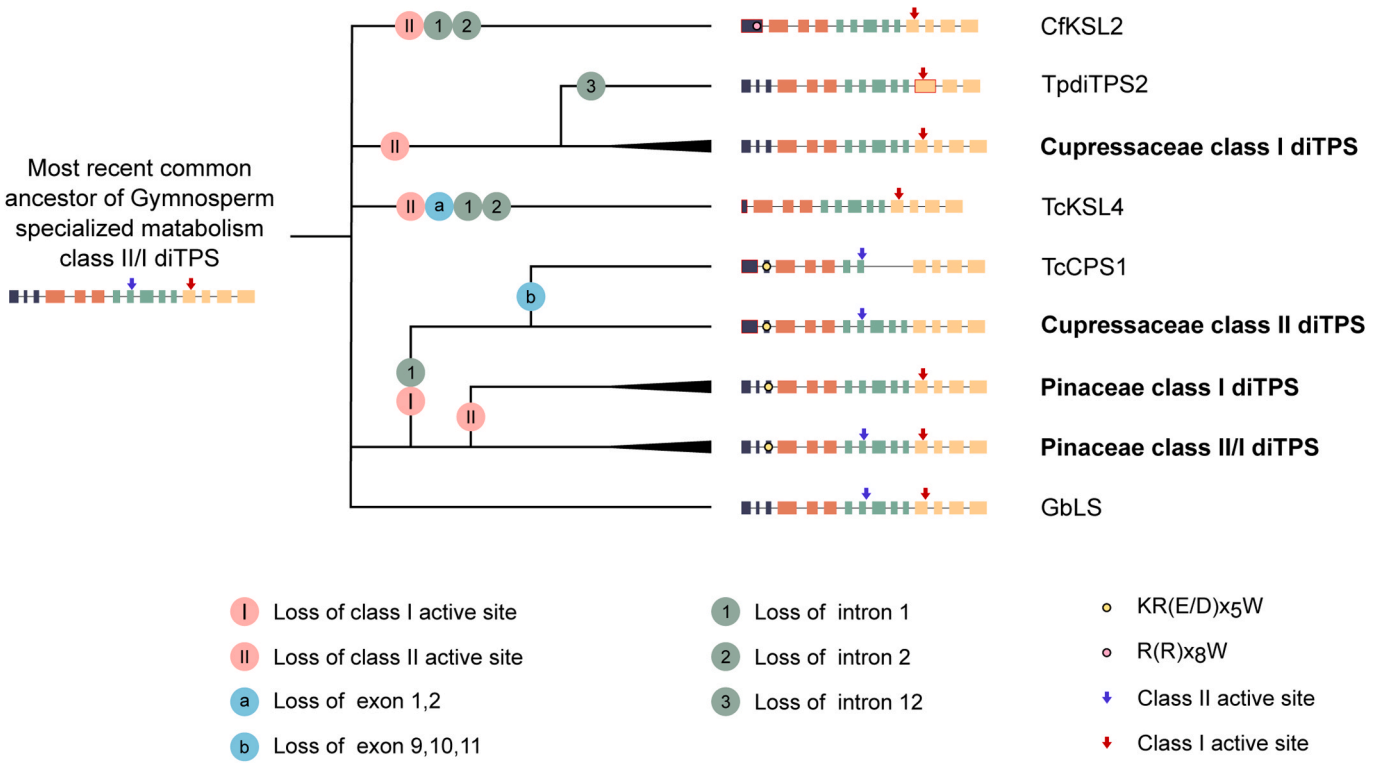

Fig. 8. Putative evolutionary pathways of diTPSs from Cupressaceae. The red box represents the exons with introns or/and exon loss.

short, which may be the cause of the loss of the original function of TcKSL4. In addition, intron 12 of TpdiTPS2 was lost and had one more sequence (S625~R653) than other levopimaradiene synthases (Fig. S7). CfKSL2 lost introns 1 and 2 and was relatively distant from other diTPSs in the phylogenetic analysis. CfKSL2 was inferred to be an excessive gene for the evolution of gymnosperm monoterpene and sesquiterpene synthases (Ma, et al., 2021). In this study, we found that CfKSL2 and the taxadiene synthase from *Taxus* both have R(R)x_8W (Table S9), which is common in monoterpene and sesquiterpene synthase and affects the cyclization of terpenoids, according to previous studies (Chen et al., 2011; Li et al., 2021).

4.4. *The possible evolution of CovfCPS/CoCPS and CovfLS/CoLS*

Levopimaradiene has been proven to be the intermediate product of many specialized diterpenes produced by gymnosperms. For example, levopimaradiene is the precursor of Taiwaniaquinoids which have anti-cytotoxic activity (Chang et al., 2005) and aromatase inhibitory activity (Fillion and Fishlock, 2005) from *T. cryptomerioides*, and the precursor of ginkgolides which has neuroprotective activity (Zhang et al., 2019; Yaro et al., 2019) from *Ginkgo biloba*. The pisiferin from *C. formosensis* also uses levopimaradiene as a precursor and has an unusual 6–7–6 ring skeleton (Ma et al., 2021). In many Pinaceae species, levopimaradiene/abietadiene synthase exists and has a relatively large number of transcripts (Hall et al., 2013). As mentioned above, levopimaradiene synthase might play a crucial role in gymnosperms.

From the perspective of gene structure, intron 1 of all Cupressaceae class II diTPSs was lost, so the evolution of class II diTPSs of Cupressaceae occurred before the differentiation of genera. Furthermore, most Cupressaceae class I diTPSs have more primitive 13 introns; therefore, it can also be speculated that Cupressaceae class I and II diTPSs evolved separately. A comparison of the gene structures and class II active sites of Cupressaceae monofunctional (+)-CPS and Pinaceae bifunctional levopimaradiene synthase showed that the class II active sites that catalyze the production of (+)-CPS from GGPP were all DIDD (Fig. 7). The largest difference in gene structure between Cupressaceae monofunctional (+)-CPS and Pinaceae bifunctional levopimaradiene synthase is the loss of intron 1. The *Chamaecyparis* (+)-CPSs (CovfCPS, CoCPS, and CfCPS1) had longer intron 7 sequences than did the (+)-CPSs of *T. plicata* and *T. cryptomerioides* (Fig. 8).

A comparison the gene structures and class I active sites with Cupressaceae monofunctional levopimaradiene synthases and *G. biloba* levopimaradiene synthase revealed that all of the above class I active sites were DDLYD. Almost all the introns of the Cupressaceae monofunctional levopimaradiene synthase had the same gene structure, with 14 introns, except that TpdiTPS2 had lost intron 12. The gene structures of CovfLS and CoLS were precisely the same. Compared with those of Cupressaceae monofunctional (+)-CPSs, the gene structures of Cupressaceae monofunctional levopimaradiene synthases were less different (Fig. 7).

CovfLS, CoLS, and CovfCPS, CoCPS, two groups of diTPSs with the same function, are almost identical in amino acid sequence and gene structure, but have many differences from other *Chamaecyparis* or Cupressaceae plants. Therefore, from the diTPS gene sequences of *C. obtusa* and *C. obtusa var. formosana*, it can be found that the genetic relationship between the two variants is quite close, which supports the hypothesis that *C. obtusa* and *C. obtusa var. formosana* diverged more recently than Cupressaceae plants did; moreover, the relationship between the two is more similar to that of species.

CRediT authorship contribution statement

Tsai-Jung Wu: Writing – original draft, Investigation, Formal analysis, Data curation. **Chi-Chun Lin:** Writing – original draft, Investigation, Formal analysis, Data curation. **Li-Ting Ma:** Formal analysis. **Chih-Kai Yang:** Data curation. **Chen-Lung Ho:** Formal analysis. **Sheng-Yang Wang:** Methodology. **Fang-Hua Chu:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

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.

Data Availability

No data was used for the research described in the article.

Acknowledgements

We thank Prof. Jay Keasling (University of California, Berkeley, USA) for sharing the pMBI plasmid. Financial assistance was provided by the National Science and Technology Council of Taiwan (MOST 110-2313-B-002–039-MY3).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.plantsci.2024.112080.

References

E. Alicandri, A.R. Paolacci, S. Osadolor, A. Sorgonà, M. Badiani, M. Ciaffi, On the evolution and functional diversity of terpene synthases in the *Pinus* species: a review, J. Mol. Evol. *88* (3) (2020) 253–283, https://doi.org/10.1007/s00239-020-09930-8.

C.-I. Chang, J.-Y. Chang, C.-C. Kuo, W.-Y. Pan, Y.-H. Kuo, Four new 6-nor5(6—>7)abeo-abietane type diterpenes and antitumoral cytotoxic diterpene constituents from the bark of *Taiwania cryptomerioides*, Planta Med. *71* (01) (2005) 72–76, https://doi.org/10.1055/s-2005-837754.

S. Chang, J. Puryear, J. Cairney, A simple and efficient method for isolating RNA from pine trees, Plant Mol. Biol. Report. *11* (2) (1993) 113–116, https://doi.org/10.1007/BF02670468.

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

R.M. Coates, H.Y. Kang, Synthesis and evaluation of cyclobutylcarbinyl derivatives as potential intermediates in diterpene biosynthesis, J. Org. Chem. *52* (10) (1987) 2065–2074, https://doi.org/10.1021/jo00386a031.

J. Criswell, K. Potter, F. Shephard, M.H. Beale, R.J. Peters, A single residue change leads to a hydroxylated product from the classII diterpene cyclization catalyzed by abietadiene synthase, Org. Lett. 14 (23) (2012) 5828–5831, https://doi.org/10.1021/ol3026022.

G. Cui, L. Duan, B. Jin, J. Qian, Z. Xue, G. Shen, X. Qi, Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*, Plant Physiol. *169* (3) (2015) 1607–1618, https://doi.org/10.1104/pp.15.00695.

R.C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. *32* (5) (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.

O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites, Protein Sci.: A Publ. Protein Soc. *8* (5) (1999) 978–984.

E. Fillion, D. Fishlock, Total synthesis of (±)-taiwaniaquinol B via a domino intramolecular Friedel−Crafts acylation/carbonyl α- tert -alkylation reaction, J. Am. Chem. Soc. *127* (38) (2005) 13144–13145, https://doi.org/10.1021/ja054447p.

D.E. Hall, P. Zerbe, S. Jancsik, A.L. Quesada, H. Dullat, L.L. Madilao, J. Bohlmann, Evolution of conifer diterpene synthases: Diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases, Plant Physiol. *161* (2) (2013) 600–616, https://doi.org/10.1104/pp.112.208546.

Y.J. Hong, D.J. Tantillo, Formation of beyerene, kaurene, trachylobane, and atiserene diterpenes by rearrangements that avoid secondary carbocations, J. Am. Chem. Soc. *132* (15) (2010) 5375–5386, https://doi.org/10.1021/ja9084786.

Q. Jia, R. Brown, T.G. Köllner, J. Fu, X. Chen, G.K.-S. Wong, F. Chen, Origin and early evolution of the plant terpene synthase family, Proc. Natl. Acad. Sci. *119* (15) (2022) e2100361119, https://doi.org/10.1073/pnas.2100361119.

M. Jia, K. Zhou, S. Tufts, S. Schulte, R.J. Peters, A pair of residues that interactively affect diterpene synthase product outcome, ACS Chem. Biol. *12* (3) (2017) 862–867, https://doi.org/10.1021/acschembio.6b01075.

P.S. Karunanithi, P. Zerbe, Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity, Front. Plant Sci. *10* (2019). Retrieved from, ⟨https://www.frontiersin.org/articles/10.3389/fpls.2019.01166⟩.

C.I. Keeling, S. Weisshaar, R.P.C. Lin, J. Bohlmann, Functional plasticity of paralogous diterpene synthases involved in conifer defense, Proc. Natl. Acad. Sci. *105* (3) (2008) 1085–1090, https://doi.org/10.1073/pnas.0709466105.

M. Köksal, K. Potter, R.J. Peters, D.W. Christianson, 1.55Å-resolution structure of *ent*-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis, Biochim. Et. Biophys. Acta (BBA) - Gen. Subj. *1840* (1) (2014) 184–190, https://doi.org/10.1016/j.bbagen.2013.09.004.

S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. *35* (6) (2018) 1547–1549, https://doi.org/10.1093/molbev/msy096.

I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation, Nucleic Acids Res. *49* (W1) (2021) W293–W296, https://doi.org/10.1093/nar/gkab301.

L.-G. Lin, C.O.L. Ung, Z.-L. Feng, L. Huang, H. Hu, Naturally occurring diterpenoid dimers: Source, biosynthesis, chemistry and bioactivities, Planta Med. *82* (15) (2016) 1309–1328, https://doi.org/10.1055/s-0042-114573.

R.-S. Li, J.-H. Zhu, D. Guo, H.-L. Li, Y. Wang, X.-P. Ding, S.-Q. Peng, Genome-wide identification and expression analysis of terpene synthase gene family in *Aquilaria sinensis*, Plant Physiol. Biochem.: PPB *164* (2021) 185–194, https://doi.org/10.1016/j.plaphy.2021.04.028.

S. Mafu, K. C. Potter, M. L. Hillwig, S. Schulte, J. Criswell, R. J. Peters, Efficient heterocyclisation by (di)terpene synthases, Chem. Commun. *51* (70) (2015) 13485–13487, https://doi.org/10.1039/C5CC05754J.

V.J.J. Martin, D.J. Pitera, S.T. Withers, J.D. Newman, J.D. Keasling, Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids, Nat. Biotechnol. *21* (7) (2003) 796–802, https://doi.org/10.1038/nbt833.

L.-T. Ma, Y.-R. Lee, N.-W. Tsao, S.-Y. Wang, P. Zerbe, F.-H. Chu, Biochemical characterization of diterpene synthases of *Taiwania cryptomerioides* expands the known functional space of specialized diterpene metabolism in gymnosperms, Plant J. *100* (6) (2019) 1254–1272, https://doi.org/10.1111/tpj.14513.

L.-T. Ma, C.-H. Wang, C.-Y. Hon, Y.-R. Lee, F.-H. Chu, Discovery and characterization of diterpene synthases in *Chamaecyparis formosensis* Matsum. Which participated in an unprecedented diterpenoid biosynthesis route in conifer, Plant Sci. *304* (2021) 110790, https://doi.org/10.1016/j.plantsci.2020.110790.

K. Miyamoto, S. Okuda, Y. Inagaki, M. Noguchi, T. Itou, Within- and between-site variations in leaf longevity in hinoki cypress (*Chamaecyparis obtusa*) plantations in southwestern Japan, J. For. Res. *18* (3) (2013) 256–269, https://doi.org/10.1007/s10310-012-0346-1.

D. Morrone, L. Lowry, M.K. Determan, D.M. Hershey, M. Xu, R.J. Peters, Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: Comparison of MEV and MEP isoprenoid precursor pathway engineering, Appl. Microbiol. Biotechnol. *85* (6) (2010) 1893–1906, https://doi.org/10.1007/s00253-009-2219-x.

K.M. Murphy, L.-T. Ma, Y. Ding, E.A. Schmelz, P. Zerbe, Functional characterization of two class II diterpene synthases indicates additional specialized diterpenoid pathways in maize (*Zea mays*), Front. Plant Sci. *9* (2018) 1542, https://doi.org/10.3389/fpls.2018.01542.

K.A. Pelot, R. Chen, D.M. Hagelthorn, C.A. Young, J.B. Addison, A. Muchlinski, P. Zerbe, Functional diversity of diterpene synthases in the biofuel crop switchgrass, Plant Physiol. *178* (1) (2018) 54–71, https://doi.org/10.1104/pp.18.00590.

K.A. Pelot, R. Mitchell, M. Kwon, L.M. Hagelthorn, J.F. Wardman, A. Chiang, P. Zerbe, Biosynthesis of the psychotropic plant diterpene salvinorin A: Discovery and characterization of the *Salvia divinorum* clerodienyl diphosphate synthase, Plant J. *89* (5) (2017) 885–897, https://doi.org/10.1111/tpj.13427.

K. Potter, J. Crisswell, J. Zi, A. Stubbs, R.J. Peters, Mechanistic analysis of the *ent*-copalyl diphosphate synthases required for plant gibberellin hormone biosynthesis leads to novel product chemistry, *Angew. Chem.* 53 (28) (2014) 7198–7202, https://doi.org/10.1002/anie.201402911.

R. Smith-Unna, C. Boursnell, R. Patro, J.M. Hibberd, S. Kelly, TransRate: Reference-free quality assessment of *de novo* transcriptome assemblies, Genome Res. *26* (8) (2016) 1134–1144, https://doi.org/10.1101/gr.196469.115.

S. Tasnim, R. Gries, J. Mattsson, Identification of three monofunctional diterpene synthases with specific enzyme activities expressed during heartwood formation in western redcedar (*Thuja plicata*) trees, Plants *9* (8) (2020) 1018, https://doi.org/10.3390/plants9081018.

S.C. Trapp, R.B. Croteau, Genomic organization of plant terpene synthases and molecular evolutionary implications, Genetics *158* (2) (2001) 811–832, https://doi.org/10.1093/genetics/158.2.811.

P. Tungcharoen, C. Wattanapiromsakul, P. Tansakul, S. Nakamura, H. Matsuda, S. Tewtrakul, Anti-inflammatory effect of isopimarane diterpenoids from *Kaempferia galanga*, Phytother. Res.: PTR *34* (3) (2020) 612–623, https://doi.org/10.1002/ptr.6549.

M. Xu, P.R. Wilderman, R.J. Peters, Following evolution's lead to a single residue switch for diterpene synthase product outcome, Proc. Natl. Acad. Sci. *104* (18) (2007) 7397–7401, https://doi.org/10.1073/pnas.0611454104.

P. Zerbe, J. Bohlmann, Plant diterpene synthases: Exploring modularity and metabolic diversity for bioengineering, Trends Biotechnol. *33* (7) (2015) 419–428, https://doi.org/10.1016/j.tibtech.2015.04.006.

W. Zhang, Z. Guo, J. Liang, X. Lu, S. Chen, J. Zhu, R. Yu, Functional characterization of the levopimaradiene synthase in *Escherichia coli* and differences in functional expression patterns in distinct cell lines, Nat. Prod. Commun. *14* (6) (2019) 1934578×19850371, https://doi.org/10.1177/1934578×19850371.