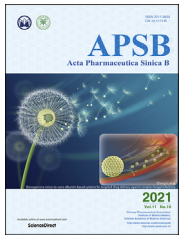
Acta Pharmaceutica Sinica B 2021;11(10):3310–3321

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



ORIGINAL ARTICLE

Functional identification of the terpene synthase family involved in diterpenoid alkaloids biosynthesis in *Aconitum carmichaelii*

Liuying Mao[a,b], Baolong Jin[b], Lingli Chen[b], Mei Tian[b], Rui Ma[b], Biwei Yin[b], Haiyan Zhang[b], Juan Guo[b], Jinfu Tang[b], Tong Chen[b], Changjiangsheng Lai[b], Guanghong Cui[b,*], Luqi Huang[a,b,*]

[a]*College of Pharmacy, Shandong University of Chinese Medicine, Jinan 250355, China*
[b]*State Key Laboratory of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*

Received 18 January 2021; received in revised form 22 March 2021; accepted 2 April 2021

KEY WORDS

Aconitum carmichaelii;
Full-length transcriptome;
Diterpene synthase;
Diterpenoid alkaloids

Abstract *Aconitum carmichaelii* is a high-value medicinal herb widely used across China, Japan, and other Asian countries. Aconitine-type diterpene alkaloids (DAs) are the characteristic compounds in *Aconitum*. Although six transcriptomes, based on short-read next generation sequencing technology, have been reported from the *Aconitum* species, the terpene synthase (TPS) corresponding to DAs biosynthesis remains unidentified. We apply a combination of Pacbio isoform sequencing and RNA sequencing to provide a comprehensive view of the *A. carmichaelii* transcriptome. Nineteen TPSs and five alternative splicing isoforms belonging to TPS-b, TPS-c, and TPS-e/f subfamilies were identified. *In vitro* enzyme reaction analysis functional identified two sesqui-TPSs and twelve diTPSs. Seven of the TPS-c subfamily genes reacted with GGPP to produce the intermediate *ent*-copalyl diphosphate. Five AcKSLs separately reacted with *ent*-CPP to produce *ent*-kaurene, *ent*-atiserene, and *ent*-13-*epi*-sandaracopimaradie: a new diterpene found in *Aconitum*. AcTPSs gene expression in conjunction DAs content analysis in different tissues validated that *ent*-CPP is the sole precursor to all DAs biosynthesis, with AcKSL1, AcKSL2s and AcKSL3-1 responsible for C_{20} atisine and napelline type DAs biosynthesis, respectively. These data clarified the molecular basis for the C_{20}-DAs biosynthetic pathway in *A. carmichaelii* and pave the way for further exploration of C_{19}-DAs biosynthesis in the *Aconitum* species.

*Corresponding authors. Tel./fax: +86 10 64087469 (Guanghong Cui); +86 10 64087892 (Luqi Huang).
E-mail addresses: guanghongcui@163.com (Guanghong Cui), huangluqi01@126.com (Luqi Huang).

Peer review under responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2021.04.008

1. Introduction

Aconitum carmichaelii Debx. is a traditional herb belonging to Ranunculaceae. Its principal and lateral roots after processed, named "wu tou" and "fu zi" respectively, are its most commonly used components in traditional Chinese medicine (Fig. 1), and have been used for the treatment of pain, rheumatics, heart failure, colds, diarrhea, beriberi, and edema for 2000 years[1]. Diterpenoid alkaloids (DAs) are believed to be the predominant bioactive compounds in *A. carmichaelii*, with over 100 isolated[2]. C_{19}-diterpenoid alkaloids, which contain unusual 6,7,5,6 carbon skeletons, are both the dominant bioactive and toxic constituents in *A. carmichaelii*. The most toxic compounds are aconitine, mesaconitine, and hypaconitine, characterized by acetyl and benzoyl esters (Fig. 1C), while alcohol amine-DAs are the primary bioactive compounds and exhibit reduced toxicity[2]. In addition to the C_{19}-type, several C_{18}- and C_{20}-DAs have also been found in *A. carmichaelii*, such as vilmorrianine D, songorine, napelline, atisine[3] (Fig. 1C) and aconicarmissulfonine A: linked to significant analgesic activity[4].

DAs, whether C_{18}, C_{19}, or C_{20}, are all presumed to be derived from the kaurane and atisane diterpenoid families[5–7]. That is, C_{20} atisane and kaurane serve as biosynthetic precursors to C_{18}- and C_{19}-DAs, a mechanism proven by total synthesis[8,9]. Thereby, the biosynthesis of DAs undergoes two principal phases. First, is the formation of the C_{20} diterpene skeleton by terpene synthase (TPS): the universal precursor geranylgeranyl pyrophosphate (GGPP) is cyclized by class II diterpene synthase (copalyl-diphosphate synthase, CPS) to produce *ent*-copalyl diphosphate (*ent*-CPP), then undergoes further cyclization or rearrangement by class I diterpene synthase (kaurene synthase-like, KSL) to form *ent*-kaurene or *ent*-atiserene. Second, is the insertion of the nitrogen into the mature diterpene scaffolds to form C_{20}-DAs, *e.g.*, atisine and napelline (Fig. 1C), which then undergo further development leading to the terminus C_{18}- or C_{19}-DAs. Historically, the specific enzymes and biosynthetic pathways in the plant have garnered substantially less interest than the total synthesis of these compounds: from 1963 to 2018, 24 different DAs have identified *via* total synthesis[10]. One of the few key mechanistic finding has been that L-serine may serve as the nitrogen source of atisine-type DAs[11].

Recently, several RNAsequencing (RNA-seq) approaches using the Illumina platform have analyzed the transcriptome of *A. carmichaelii*[12–17] and *Aconitum heterophyllum*[18]. Many candidate genes, such as terpenoid-related enzymes, monooxygenases, methyltransferases, and BAHD acyltransferases (named according to the first four identified enzymes of the family: BEAT, benzylalcohol *O*-acetyltransferase; AHCT, anthocyanin *O*-hydroxycinnamoyltransferase; HCBT, anthranilate *N*-hydroxycinnamoyl/benzoyltransferase; DAT, deacetylvindoline 4-*O*-acetyltransferase) related to DAs biosynthesis have been identified[13–17,19,20]. However, the limitations of short-read sequencing results in the vast majority of isotigs not representing full-length cDNA sequences, and none of these candidates have yet been functionally identified through *in vivo* or *in vitro* analysis.

In addition to being initiated by class II diterpene synthase, diterpenes can also be initiated by class I diterpene synthase, such as taxa-4(5),11(12)-diene[21], casbene[22,23], and pseudolaratriene[24]. Meanwhile, there are case reports that in addition to the TPS-c (*e.g.*, CPS) and TPS-e/f (*e.g.*, KSL) subfamily genes, the TPS-a and TPS-b subfamily genes have also been linked to diterpene biosynthesis[25,26]. This suggests that C_{19}-DAs backbones could be

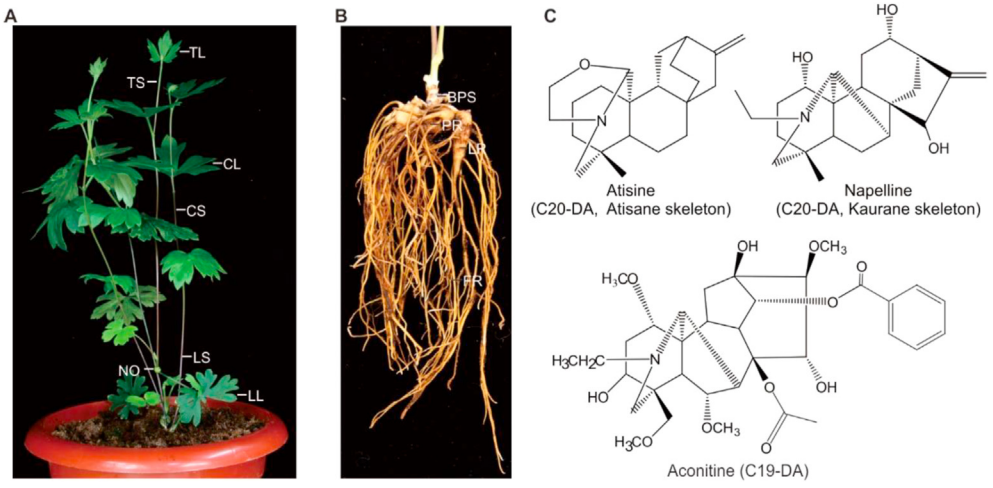


Figure 1 The plant shows tissues used in this study for sequencing and the characteristic DAs in *A. carmichaelii.* The four-month-old plant (A) and roots (B) of *A. carmichaelii* in the greenhouse. C. The characteristic skeleton of DAs in *A. carmichaelii.* PR, principal root; LR, lateral root; FR, fibrous root; BPS, basal part of the stem; TL, Top leaf; CL, central leaf; LF, lower leaf; TS, Top stem; CS, central stem; LS, lower stem, NO, nodular like organs.

synthesized by unexpected enzymes. Accordingly, identifying all the TPS family genes from the whole transcriptome is a critical step in understanding the therapeutic agents of *A. carmichaelii*.

Single-molecule real-time (SMRT) sequencing provides reads > 10 kb and up to 60 kb (PacBio, Menlo Park, CA, USA 2016), enabling more complete and even full-length transcriptome data[27,28]. Herein, we combined full-length isoform sequencing (Iso-seq) and RNA-seq technology to obtain a more reliable *A. carmichaelii* transcriptome. Combined with PCR cloning, we finally identified 19 TPS genes in *A. carmichaelii* and five alternative splicing isoforms. *In vivo* and *in vitro* assay identified 14 functional TPS genes, which establish biosynthetic routes to the diterpene scaffolds *ent*-atiserene, *ent*-kaurene, and *ent*-13-*epi*-sandaracopimaradie, clarifying the molecular basis for *A. carmichaelii* DAs biosynthesis.

2. Materials and methods

2.1. Plant material

A. carmichaelii was acquired from Jiangyou, Sichuan Province, which is the genuine location for "wu tou" and "fu zi" material[29], in June 2018. The plant was grown in a greenhouse until February 2019. Then, 7 lateral roots from the same plant were divided and planted in separate flowerpots. These were then grown in a greenhouse at 25 (±2 °C) under a 16 h-light/8 h-dark cycle provided by a white fluorescent lamp (3000 lux). After 4 months growth, 3 plants with similar conditions were selected for the study. Different tissues were collected and immediately frozen in liquid nitrogen for further study.

2.2. RNA extraction

Total RNA was extracted using an RNA isolation kit (HuaYueYang Biotechnology, Beijing, China) following the manufacturer's instruction. The integrity and concentration of total RNA was monitored using 1.0% agarose gel electrophoresis and NanoDrop ND-100 (NanoDrop Technologies, DE, USA). The electrophoretogram is shown in Supporting Information Fig. S1. The Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was further used to assess the quality of extracted RNA.

2.3. Iso-seq and RNA-seq library construction and sequencing

Due to the relatively high DAs content in the top leaves, principal roots, and lateral roots, to perform Iso-seq we mixed RNA from these three tissues together; concurrently, 13 tissues from three biological replicates were used for RNA-seq. The library construction for Iso-seq was described in Supporting Information Fig. S2. Briefly, mRNA was isolated from total RNA and converted into full-length cDNA using a SMARTer PCR cDNA Synthesis Kit (Clontech, CA, USA). Then, cDNA amplification was performed *via* an advantage 2 PCR kit (Clontech). After purification, size selection using a BluePippin Size selection system was performed (Sage Science, MA, USA) to enrich fragments longer than 4 kb. Then, an equal molar ratio mixed library of enriched fragments with those not processed by size selection was constructed using a SMRTbell Template Prep kit (Pacific Biosciences, CA, USA). The libraries' quality was assessed by Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) and Qubit fuorometer 2.0 (Life Technologies, CA, USA) before sequencing. To prepare the libraries for sequencing, sequencing primer was annealed and polymerase was added to the primer annealed template. Subsequently, the polymerase-bound template was bonded to MagBeads and sequenced on a PacBio RS II instrument (Novogene company, www.novogene.com).

RNA-seq was performed according to the standard protocol. Briefly, mRNA was converted into library templates using a TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA, USA). The cDNA libraries were then sequenced from each end of the cDNA fragments on a HiSeq 2500 (Illumina).

2.4. Full-length transcript data processing and annotation

The SMRT-Analysis software package SMRTlink v7.0 (Pacfic Biosciences) was used for Iso-seq data analysis. Briefly, low-quality reads (<50 bp) were first removed to obtain subreads. The subreads were then classified into circular consensus sequence (CCS) and non-CCS subreads using a ToFu pipeline (Pacific Biosciences). If the sequence contained a 3′ and 5′ primer or a poly (A) tail, the read was then used to obtain the full-length non-chimeric read (FLNC). Redundant FLNCs from same transcripts were discarded to form the consensus reads. The consensus reads were then polished with arrow software, and sequencing errors were further corrected using Illumina reads from the 13 different tissues *via* LoRDEC[30]. Lastly, the redundant transcripts were removed *via* CD-HIT software[31]. The resultant transcripts then underwent the standard annotation process using 7 databases: Nr, Nt, Pfam, KOG/COG, Swiss-prot, KEGG, and GO.

2.5. Cloning, sequence alignment, and phylogenetic analysis of TPS genes

1 μg of total RNA from 5 tissues were reverse transcribed into cDNA using the PrimerScript™ RT reagent kit with gDNA eraser (TaKaRa Corp., Dalian, China) according to the manufacturer's instructions. TPS genes were then amplified with specific primers (Supporting Information Table S1) and cloned into a pET32 plasmid (Merck) with a seamless cloning kit (TransGen Biotech, Beijing, China). TPS alignment was represented using CLC Sequence viewer 7 (Qiagen, Denmark) software. The previously identified 51 diTPSs (Supporting Information Table S2) together with TPS from *Arabidopsis thaliana* were used to construct the maximum likelihood phylogenetic tree using MEGA 6.0[32].

2.6. In vitro assays

The functions of the AcTPSs were tested *via in vitro* assays as has been previously described[33]. AcTPSs were analyzed alone or coupled with others, and an appropriate substrate supply. Briefly, recombinant plasmids were expressed in *Escherichia coli* Transetta (DE3). The crude proteins were then affinity-purified and individually assayed *via* GPP, FPP, GGPP, and CPP. CPP was obtained by incubating 200 μL of the purified ZmCPS2 enzymes with 20−50 μmol/L GGPP (Sigma) for 3 h at 30 °C. AcCPS function was estimated using the GGPP substrate; AcKSL function was assessed *via* the GGPP and CPP substrates; the other TPSs were tested with all possible substrates, respectively. Nine known enzymes, ZmCPS2 (Genbank: NM_001111787), IrCPS4 (KU180502), SmCPS1 (KC814639), SmKSL1 (EF635966), IrKSL4 (KX580633), IrKLS5 (KX580634), EpTPS23 (KP889108), OsKSL10 (DQ823355) and MsTPS1 (MH626616) were used as controls. Different enzymes were mixed with equal volume, typically 300 μL each, in a 2 mL vial. After incubation for 2−4 h at 30 °C, an equal volume of hexane was used to extract

the terpene product. The hexane extracts were dried under a gentle stream of N_2, and the residue suspended in 100 μL hexane. 1 μL was later subjected to GC−MS analysis.

2.7. Metabolic engineering of diterpenes in E. coli

A previously reported modular metabolic engineering system[34] was used in our study. Briefly, the plasmid pIRS overexpressed IDI and DXR[35], which are key genes from endogenous isoprenoid precursor pathway. The plasmid pGG-ZmCPS2 contained a pseudo-mature GGPS from *Abies grandis*[36] and ZmCPS2/An2 from maize[37], which have been designed to efficiently produce GGPP and CPP, respectively. These two plasmids, together with pET32-AcKSLs, were transfected into an *E. coli* C41 overexpression strain (Lucigen). The recombinant cultures were grown in 30 mL TB medium (pH 7.0), with appropriate antibiotics, in 100 mL flasks. These cultures were grown at 37 °C till OD_{600} reached 0.7. The temperature was then dropped to 16 °C for 0.5 h prior to induction with 1 mmol/L isopropylthiogalactoside (IPTG), followed by supplementation with 40 mmol/L pyruvate and 1 mmol/L $MgCl_2$. The induced cultures were grown for additional 48−72 h, then 3 mL of the cultures were used for extraction with an equal volume of hexane. The extraction process was the same in the *in vitro* assays.

2.8. Terpene product analysis by GC−MS chromatography

GC−MS analysis was carried on a Thermo TRACE 1310 gas chromatograph with a TSQ8000 mass detector (Thermo Fisher Scientific) in electron ionization mode. A capillary column TR-5ms (30 mm × 0.25 mm ID; DF = 0.25 μm; Thermo Fisher Scientific) was used with a 1.0 mL/min helium flow rate. The temperature program was as follows: initial column oven temp 50 °C, maintained for 2 min followed by a linear ramp at 40 °C/min to 210 °C; linear ramp at 5 °C/min to 250 °C; linear ramp at 40 °C/min to 300 °C, with a 5 min hold at 300 °C. The ion trap temperature was 280 °C. The terpenes were identified by a comparison of retention time and mass spectra to previously characterized enzymatic products or authentic standards.

2.9. DiTPS gene expression analysis

The expression of diTPS genes were obtained by two means: RNA-seq and qPCR analysis. For the RNA-seq, bowtie2 in RSEM software[38] was used to map RNA-seq reads from the 13 tissues to the SMRT-based reference transcriptome. The read counts were then transformed to FPKM values to determine the gene expression levels of all identified genes. For qPCR analysis, 5 RNA extracts from the fibrous root, principal root, lateral root, leaf, and stem, respectively, were synthesized into cDNA by PrimerScript™ RT reagent kit with a gDNA eraser (TaKaRa Corp.). A SYBR Green kit (TaKaRa Corp.) for quantitative real-time polymerase chain reaction (qPCR) was applied to a ABI7500 real-time PCR detection system according to the manufactory's instructions. The primers sequences are listed in Table S1. Primer specificity was assessed by agarose gel and melting curve analysis. The results were normalized with the housekeeping gene *Actin*. Relative expression levels were calculated as the mean of three technical replicates of three biological replicates.

2.10. GenBank accessions

The full-length transcriptome reported in this paper has been deposited in China National Center for Bioinformation under accession number CRA003781, which is publicly accessible for all researchers at http://bigd.big.ac.cn/gsa. GenBank accession numbers for the functional terpene synthases described in this paper are AcCPS1 (MW478118), AcCPS2-1 (MW478119), AcCPS2-2 (MW478120), AcCPS2-3 (MW478121), AcKSL2-1 (MW478123), AcKSL2-2 (MW478124), AcKSL3-1 (MW478125), AcKSL3-2 (MW478126), AcTPS3 (MW478129) and AcTPS4 (MW478130). The sequences of different isoforms and *E. coli* codon optimized AcKSL1 were listed in Supporting Information isoform.seq.

3. Results

3.1. Full-length transcriptome sequencing using Iso-seq and RNA-seq platforms

Due to the relatively high DAs content in the top leaves, principal root, and lateral root[3], we mixed RNA from these together tissues for Iso-seq. Concurrently, 13 tissues from differing spatial locations (Fig. 1A and B) were sequenced on an Illumina HiSeq 2000 platform to quantify gene/isoform expression levels and correct the Iso-seq reads. Utilizing 40.86 Gb sequenced data, we obtained 37, 182, 436 subreads (Supporting Information Table S3). The Iso-seq subreads' length distribution was shown in Supporting Information Fig. S3. From the subreads we extracted 983,658 circular consensus sequence (CCS) reads. Subsequent trimming, assembly, filtering, classification, and clustering revealed that 52.95% were full-length reads, indicated by the detection of poly (A) or 3′ and 5′ sequences (Supporting Information Table S4). These were then used in the construction of full-length non-chimeric sequences (FLNCs), which ranged from 72 to 12,220 bp, with a 2023 bp average. The cDNA N50 values were determined to be 2612 bp. Of 520,860 full-length cDNAs, 48,240 polished, high-quality, isoform consensus reads were filtered out using arrow software. After isoform-level clustering, next generation short sequencing (Illumina) was performed for error correction, with redundant sequences removed *via* CD-HIT software, and 21,087 unigenes were generated for *A. carmichaelii*: average length 1991 bp, N50 value 2446 bp (Supporting Information Fig. S4). The N50 obtained in this study is 3-fold higher than previously reported (830 bp[13] and 844 bp[14]), providing a novel opportunity to identify all TPS genes involved in DAs biosynthesis.

3.2. Identification of A. carmichaelii terpene synthases

To identify the *A. carmichaelii* TPS gene family, the above obtained unigenes were annotated using seven databases: Nr, Nt, Pfam, KOG/COG, Swiss-prot, KEGG, and GO (Supporting Information Fig. S5). We also mined the data using BLAST analysis against the enzymes like casbene and neocembrene synthases, which are responsible for macrocyclic or unusual skeleton formation[22,23]. In addition, for genes with unpredictable lengths we used BLAST analysis against the reported transcriptome database[13].

A total of 19 TPS unigenes were identified, of which 14 represented full-length sequences (Supporting Information Table S5). The remaining 5 represented alternative splicing events. Transcript13827 (1617 bp), transcript12676 (1443 bp) and transcript12546 (567 bp) had a 515 bp overlap (Supporting Information Fig. S6). The integration of transcript13827 and transcript12676 forms a common full-length CPS gene (average

2400 bp), annotated as *AcCPS1*. It was verified by polymerase chain reaction (PCR) amplification from the cDNA of fibrous root. At the same time, we obtained an isoform *AcCPS1a*, which had 98.70% sequence similarity with *AcCPS1* (Supporting Information Fig. S7). We further used specific primers to clone the three alternative splicing isoforms. We got the expected PCR product of transcript13827 and transcript12676, while didn't obtain any PCR product when using the specific primers of transcript12546. After sequenced the individual clones, we found nine of the transcript12676 clones were the same, which was nearly identical with *AcCPS1* except the 11 nucleotides at the beginning of "ATG". However, the condition was much more complexed for the clones from transcript13827. There were 37 nucleotide variations among the six clones and divided into three types. Four of them had the same sequence with *AcCPS1a*, clone-3-12 was identical with *AcCPS1* except two nucleotide variations. Clone-4-12 was more different from the others. Sixteen nucleotides belonged to *AcCPS1a*, 8 nucleotides belonged to *AcCPS1*, and the other 13 nucleotides were different from both of them (Fig. S7). But all the six clones didn't contain the 34 bp deletion compared with *AcCPS1*, thus made them didn't have the open reading frame except clone-4-12, which occasionally had a nucleotide deletion at the 3′-terminar stop codon (Fig. S7). Thus, we finally obtained two full-length *AcCPS1* isoforms, *AcCPS1* and *AcCPS1a*, together with two alternative splicing isoforms including transcript12676 and clone-4-12 from transcript13827.

Transcript11709 was 2247 bp, which was a little shorter than the common KSL genes (average 2400 bp). When blast against the reported transcriptome data we further got a full-length KSL gene with a 2454 bp open reading frame, which was annotated as *AcKSL1*. Further *AcKSL1* and the transcript11709 were cloned from the cDNA of basal part of the stem. The process of cloning is not smooth for *AcKSL1* (more than five times and from different organs); however, the cloning of transcript11709 was much easier. After sequencing, we got four isoforms of transcript11709. Clone-1-15 was identical with *AcKSL1* besides the early termination. Clone-6-2 and clone-6-5 were extremely identical beside a three nucleotides insertion for clone-6-5 at the positon of 123 bp, and they had about 97% sequence similarity with clone-1-15. Clone-1-8 is much more different from others, with about 94.7% sequence similarity with *AcKSL1*. At the same time, the stop codon of clone-1-8 mutated into "GGA", which resulted in the absence of a complete open reading frame (Supporting Information Fig. S8). Thus, we further got one full-length *AcKSL1* and three alternative splicing isoforms, clone-1-15, clone-6-2 and clone-6-5 of transcript11709. Transcript26960 was identical with the full-length transcript20586, annotated as *AcTPS7*, from 498 to 1743 bp; we thus didn't design primers to clone the transcript26960.

At the process of cloning the other genes from cDNA, we further obtained two more isoforms of *AcCPS2-3*. They have more than 98.6% sequence similarity with *AcCPS2-3* (Supporting Information Fig. S9). In summary, we cloned all the genes from various tissues except transcript19787 (*AcTPS2*). It has 98.09% sequence identity with *AcTPS1*, and the main difference was in the 3′ terminal stop codon (Supporting Information Fig. S10). Finally, we obtained 19 TPS genes and five alternative splicing isoforms from a single *A. carmichaelii* plant (Supporting Information Table S6). We further blast these TPS genes against the previously reported transcriptome[13], it showed that nearly all genes except *AcTPS4* could found corresponding unigenes with similarity higher than 92%

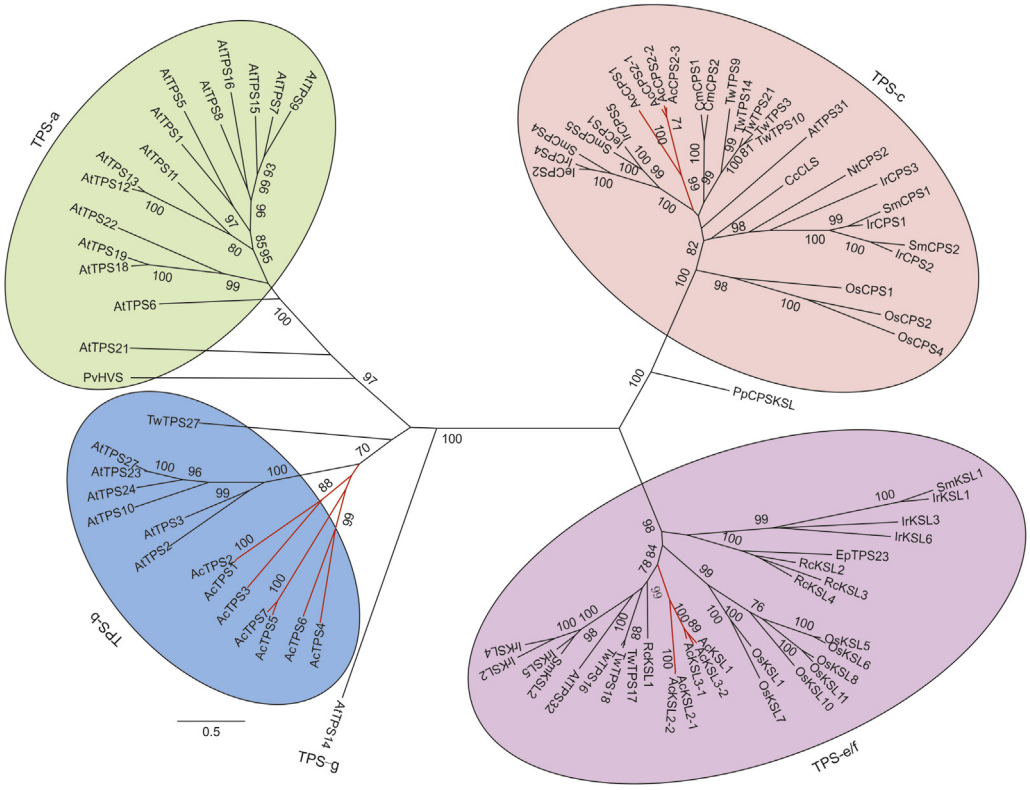


Figure 2 Phylogeny of *A. carmichaelii* terpene synthases. The maximum likelihood tree was reconstructed using MEGA6.0. Numbers on branches indicate the bootstrap percentage values calculated from 1000 bootstrap replicates. TPSs from representative characterized subfamily genes are shown in Table S6. *Physcomitrella patens* copalyl diphosphate synthase/kaurene (PpCPSKS) was used as outgroup.

(Supporting Information Table S7). However, all these unigenes didn't represent the full-length cDNAs except TR36313|c0_g1_i1, which corresponded to *AcTPS2*. Unigene TR48214|c4_g1_i1 had the highest similarity with *AcKSL3-1* (98%), while when using *AcKSL1* and *AcKSL3-2* as the query, the result also to be TR48214|c4_g1_i1 with similarity of 92%. These results demonstrated that the Pacbio full-length platform enabled an in-depth investigation of the TPS gene families in *A. carmichaelii*. In order to make the result clear, the following analysis mainly focused on 16 TPS genes which had identical sequence with the transcriptome data (Table S6), and generally didn't contain various isoforms cloned by PCR, unless otherwise specified.

3.3. Phylogenetic relationship of *A. carmichaelii* terpene synthases

We further categorized the 16 TPS genes by signature sequence motif analysis and phylogenetic comparison with TPS families from *A. thaliana* and 51 previously identified related functional diTPS (Fig. 2). The presence of the catalytic DXDD motif in combination with a close phylogenetic relationship to known TPSs of the TPS-c clade, illustrated that 4 TPS genes were linked to class II diTPS: AcCPS1 and AcCPS2-1 through AcCPS2-3 (Supporting Information Fig. S11). These 4 enzymes formed a single group, which differs from class II diTPS in other species (*i.e.*, *Salvia miltiorrhiza* and *Isodon rubescens*) where class II diTPS usually form two distinct groups[33,39]. AcCPS1 shares ~64% identity with AcCPS2-1 through AcCPS2-3, and the 3 AcCPS2s share >87% sequence identity. Due to the high consistency of AcCPS2s, we speculate that they may come from duplication in a single chromosome or from a similar allele of different chromosomes, given that *A. carmichaelii* has been reported to be a polyploidy plant[40]. The close relationship found between CmCPS1 and CmCPS2, and other CPSs in *S. miltiorrhiza* and *I. rubescens* that have been linked to *ent*-CPP production[33,39,41] indicates that AcCPS1 and AcCPS2s are *ent*-CPP synthases.

The 12 remaining candidate genes were designated as class I TPS, according to the presence of the conserved DDXXD motif (Fig. S11). Of these TPSs, 5 were clustered with diTPSs from the TPS-e/f clade: AcKSL1, AcKSL2-1, AcKSL2-2, AcKSL3-1, and AcKSL3-2 (Fig. 2). Their 5 respective enzymes were also clustered together. AcKSL1 shares ~63% identity with AcKSL2-1 and AcKSL2-2, and ~85% identity with AcKSL3-1 and AcKSL3-2. AcKSL2-1 and AcKSL2-2 share 98% sequence identity, while AcKSL3-1 and AcKSL3-2 share 92% sequence identity, again suggesting they are allelic variants or from tandem duplicates. The other 7 TPSs, AcTPS1 to AcTPS7, were determined to be within the TPS-b subfamily. Most of the characterized TPSs in this subfamily are monoterpene synthases or isoprene synthases[42]. AcTPS1 and AcTPS2 share 98% identity, AcTPS5 and AcTPS7 share 87%, and the others share a substantially lower sequence identity (~40%). This clade included TwTPS27, whose function is corollary to diTPS and can convert copalyl diphosphate to miltiradiene in *Tripterygium wilfordii*[25]. These results indicate that all these TPS candidates have the potential to be involved in *A. carmichaelii* DAs biosynthesis.

3.4. Functional characterization of class II terpene synthases from *A. carmichaelii*

Based on the full-length transcriptome sequencing, we cloned all the *AcCPSs* from *A. carmichaelii*, including seven full-length

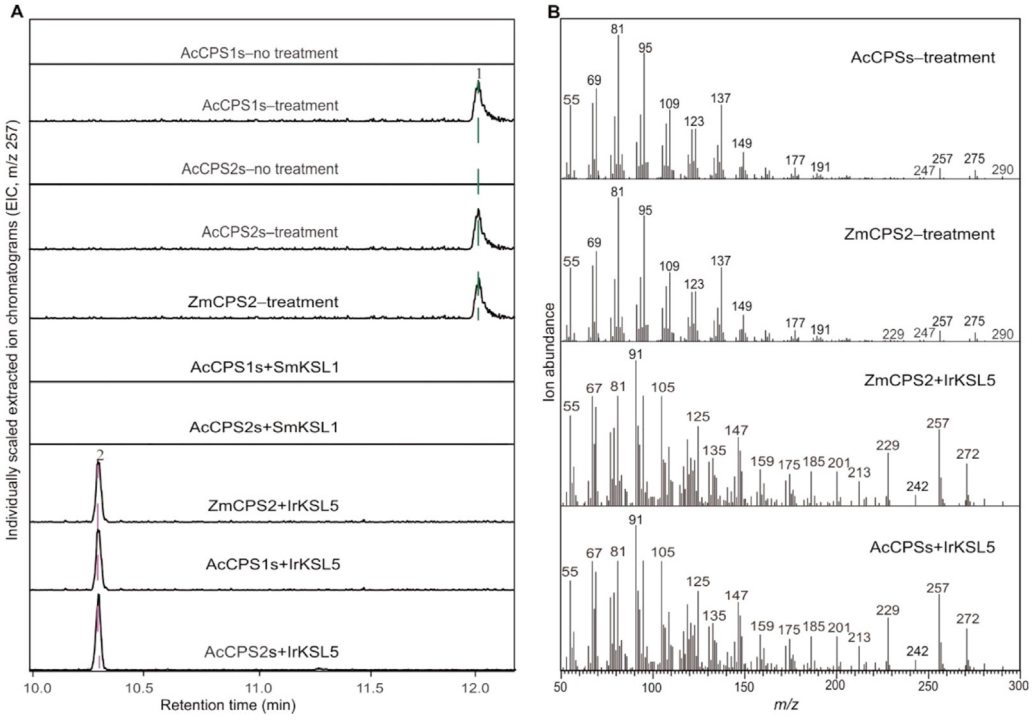


Figure 3 GC−MS analysis of AcCPSs reaction products obtained from *in vitro* assays A. Extracted ion chromatograms (EIC) of *m/z* 257 with AcCPS1s and AcCPS2s alone or in combination with IrKSL5 (specific to *ent*-CPP) or SmKSL1 (specific to *normal*-CPP) with GGPP as the substrate. "Treatment" indicates the dephosphorylation of the product of CPS with Alkaline Phosphatase. B. EI mass spectra of the dephosphorylated product of AcCPSs and ZmCPS2 and of the product of AcCPSs and ZmCPS2 combination with IrKSL5. 1) copalol, 2) *ent*-kaurene.

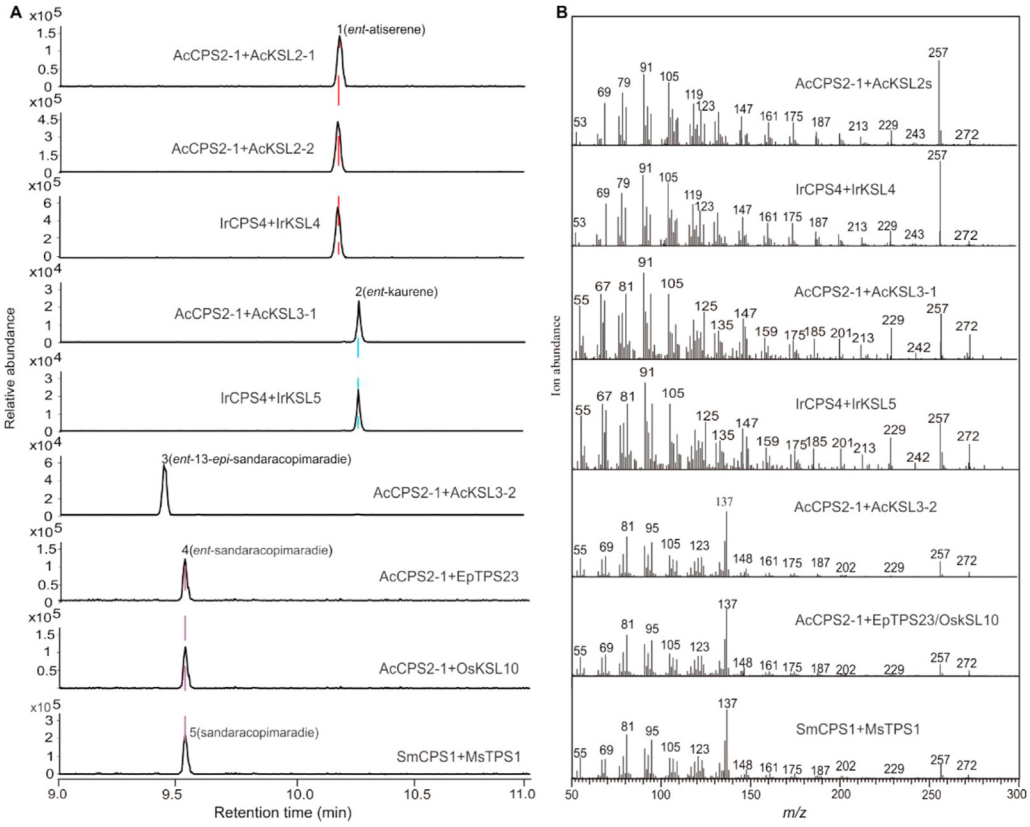


Figure 4 GC−MS analysis of AcKSLs reaction products obtained from *in vitro* assays. (A) EIC of *m/z* 257 of AcCPS2-1 and SmCPS1 in combination with different KSLs. The enzymes SmCPS1, IrCPS4, IrKSL4, IrKSL5, EpTPS23, OsKSL10 and MsTPS1 were used as controls. (B) EI mass spectra of the product of AcCPS2-1, IrCPS4 and SmCPS1 in combination with different KSLs. 1) *ent*-atiserene, 2) *ent*-kaurene, 3) *ent*-13-*epi*-sandaracopimaradie, 4) *ent*-sandaracopimaradie, 5) sandaracopimaradie.

AcCPSs and two alternative splicing isoforms of *AcCPS1* (Table S6). We then functionally characterized these *AcCPSs* through *in vitro* assays with recombinant proteins expressed *via E. coli* (Supporting Information Fig. S12). We first separately incubated the AcCPSs with GGPP to compare with the results from *ent*-CPP synthase ZmCPS2 of *Zea mays*[37]. All enzymes except the two alternative splicing isoforms yielded a product with identical retention time and mass spectra to the product of ZmCPS2 after dephosphorylation (Fig. 3A and B), indicating that the primary product of these AcCPSs is CPP.

To further investigate the stereochemistry of the CPP products, these enzymes were coupled to stereospecific class I diterpene synthases, such as SmKSL1[33,43] (specific to normal-CPP) and IrKLS5[39] (specific to *ent*-CPP). Reactions where AcCPSs were coupled with IrKLS5 resulted in the formation of *ent*-kaurene, while no product was evolved when they were coupled with SmKSL1 (Fig. 3A). That all AcCPSs could produce *ent*-CPP indicates their involvement in *A. carmichaelii* DAs biosynthesis.

3.5. Functional characterization of class I terpene synthases from A. carmichaelii

We cloned 14 class I TPS genes for functional identification except *AcTPS2*, which included 5 *KSLs*: *AcKSL1*, *AcKSL2-1*, *AcKSL2-2, AcKSL3-1* and *AcKSL3-2*, and 6 TPS-b subfamily genes: *AcTPS1*, *AcTPS3* through *AcTPS7*, together with three alternative splicing isoforms of *AcKSL1*. Having identified only the *ent*-CPP activity of the 7 AcCPSs, we first identified the products of the 5 class I diTPS AcKSLs and three alternative splicing isoforms of AcKSL1 in combination with *ent*-CPP synthase (AcCPS2-1). When GGPP was used as the substrate, AcKSL2-1 and AcKSL2-2 turned *ent*-CPP into *ent*-atiserene, a result identical to IrCPS4 in combination with IrKSL4 in *I. rubescens*; AcKSL3-1 produced *ent*-kaurene, identical to the product of IrCPS4 coupled with IrKLS5[39], and AcKSL3-2 coupled with AcCPS2-1 produced a product similar to *ent*-sandaracopimaradie[44] (Fig. 4). We further synthesized two known *ent*-sandaracopimaradie synthases, EpTPS23 from *Euphorbia peplus*[44] and OsKSL10 from *Oryza sativa*[45], along with sandaracopimaradie synthase (MsTPS1) from *Mentha Spicata*[46] to verify the product stereochemistry of AcKSL3-2. When AcKSL3-2, EpTPS23, and OsKSL10 were respectively combined with AcCPS2-1, the product mass spectra of AcKSL3-2 revealed to be identical to the products of EpTPS23 and OsKSL10, though its retention time was ~0.8 min earlier (Fig. 4). While sandaracopimaradie, the product of MsTPS1 combined with normal-CPP synthase (SmCPS1 from *S. miltiorrhiza*[33]), had exactly the same retention time and mass spectra with the product of EpTPS23/OsKSL10. These results confirmed that the product of MsTPS1 with normal-CPP and EpTPS23/OsKSL10 with *ent*-CPP were enantiomers, while the product of AcKSL3-2 was the C-13 epimer of the product of EpTPS23/OsKSL10, due to the fact that they all originated from *ent*-CPP. Thus, the product of AcKSL3-2 was assigned to be *ent*-13-*epi*-sandaracopimaradie.

Due to the poor protein expression of AcKSL1 and its three alternative splicing isoforms in *E. coli*, we could not identify its

function *in vitro*. Therefore, we used a highly efficient diterpene modular metabolic engineering system[35,36] to further verify the function of all 8 of the KSLs. Three vectors, pIRS, pGG-ZmCPS2, and pET32-AcKSLs, were transfected into the C41 over-expressing strain of *E. coli*. After incubation for 72 h the product was analyzed by GC−MS. The results of AcKSL2s and AcKSL3-1 were analogous to the above *in vitro* assays. AcKSL3-2 produced two other products in addition to *ent*-13-*epi*-sandaracopimaradie: one was *ent*-kaurene at 10.30 min based on the same retention time and mass spectra with the product of AcKSL3-1. The other product at 10.75 min showed to be an unidentified diterpene. *Ent*-atiserene was produced by the *E. coli* codon optimized AcKSL1, which had identical retention time and mass spectra with the products of AcKSL2s (Fig. 5). No obvious products were produced by the none codon optimized AcKSL1 and its three alternative splicing isoforms.

To investigate whether class I TPS genes have the capacity to produce diterpenes from GGPP, we first incubated all the recombinant class I TPS with GGPP. No product was formulated in these experiments (Supporting Information Fig. S13). We next incubated the 6 TPS-b enzymes with GGPP and AcCPS2-1, but again no products were formed. We thus conclude that in *A. carmichaelii*, DAs are only produced by the combination of the TPS-c and TPSe/f subfamily genes. In order to identify the function of the TPS-b genes, we incubated them with FPP and GPP, respectively. Our results showed that two genes, AcTPS3 and AcTPS4, generated a product similar to farnesene with FPP, which was identified to be α-farnesene with authentic standard β-farnesene and β-disabolene (Fig. S13), illustrating they are sesquiterpene synthases. The others showed no enzymatic activity *in vitro* assay though AcTPS1 and AcTPS6 had high quality of recombinant proteins (Fig. S12).

3.6. Gene expression patterns of *A. carmichaelii* terpene synthases

In order to investigate the potential physiological roles of the candidate *TPSs*, we performed RNA sequencing across 13 tissues based on spatial location. Six were underground tissues: principal root, fibrous root, cortex, phloem and xylem from lateral root, and the basal part of the stem; and seven were aerial tissues: the top, central, and lower parts of the leaf and stem respectively, and the nodular like organs in the stem (Fig. 1A and B).

Thirteen tissues were able to be differentiated into two groups according to their spatial locations, with the aerial tissues forming one group and the underground tissues the other, illustrating that the expression of TPS genes indeed has a spatial difference (Fig. 6A). The 16 TPS genes formed 3 groups: group I included 3 *AcCPS2*s, *AcKSL1*, and *AcKSL3-2*, all of which had similar moderate expression levels in across all tissues; group II included 4 TPS-b subfamily genes (*AcTPS1*, *AcTPS2*, *AcTPS3*, and *AcTPS6*) and *AcKSL3-1*, which had relatively high expression levels in the aerial leaf and stem; and group III included *AcKSL2-1*, *AcKSL2-2*, *AcCPS1* and 3 TPS-b subfamily genes (*AcTPS4*, *AcTPS5*, and *AcTPS7*), with relatively high expression in the underground tissues, especially for *AcKSL2-1* and *AcKSL2-2*. Notably, *AcCPS1* had specifically high expression levels in the fibrous root samples.

In order evaluate the expression levels from RNA-seq, we further evaluated the mRNA transcription levels of 4 diTPSs in the principal root, lateral root, fibrous root, stem, and leaves by qPCR (Fig. 6B). The results were in agreement with the RNA-seq data: *AcCPS1* was specifically expressed in the fibrous root; *AcCPS2-1* had relatively high expression levels in all tissues; *AcKSL2-1* showed relatively high expression level in principal and lateral

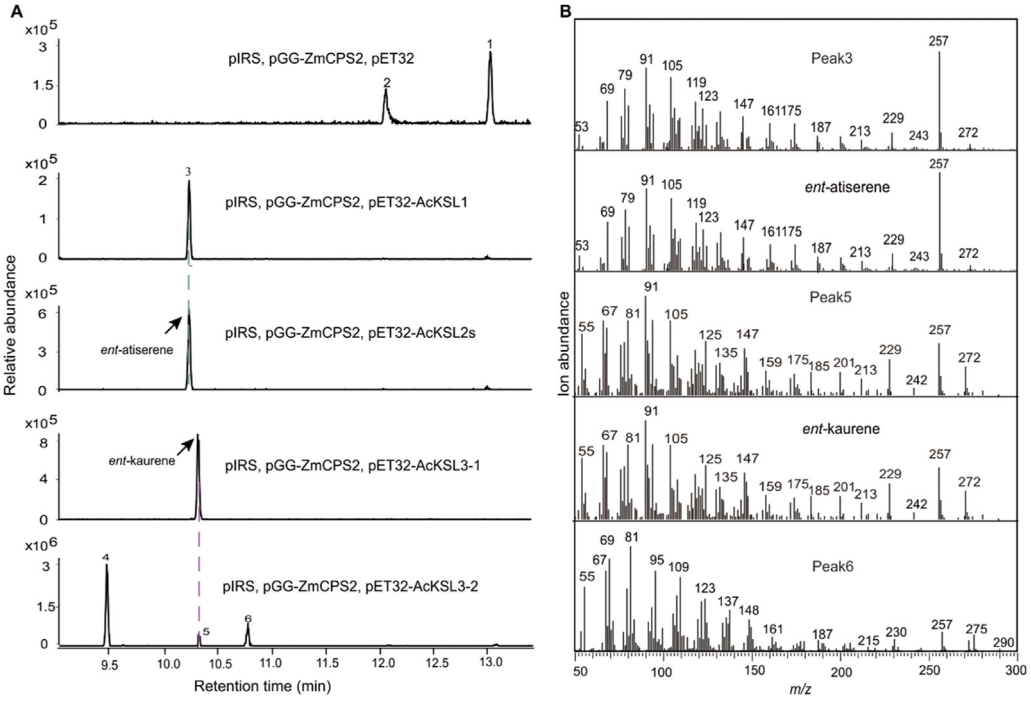


Figure 5 GC−MS analysis of extracts from *E. coli* engineered for diterpene production. (A) EIC of *m/z* 257 with pIRS, pGG-ZmCPS2 in combination with AcKSLs. The empty vector pET32, together with pIRS and pGG-ZmCPS2 was used as the negative control, which produces the hydrolysis products geranylgeraniol and copalol. (B) EI mass spectra of the product of codon optimized AcKSL1 and the two other products of AcKSL3-2. 1) geranylgeraniol, 2) copalol, 3) *ent*-atiserene, 4) *ent*-13-*epi*-sandaracopimaradie, 5) *ent*-kaurene, 6) unidentified diterpene.

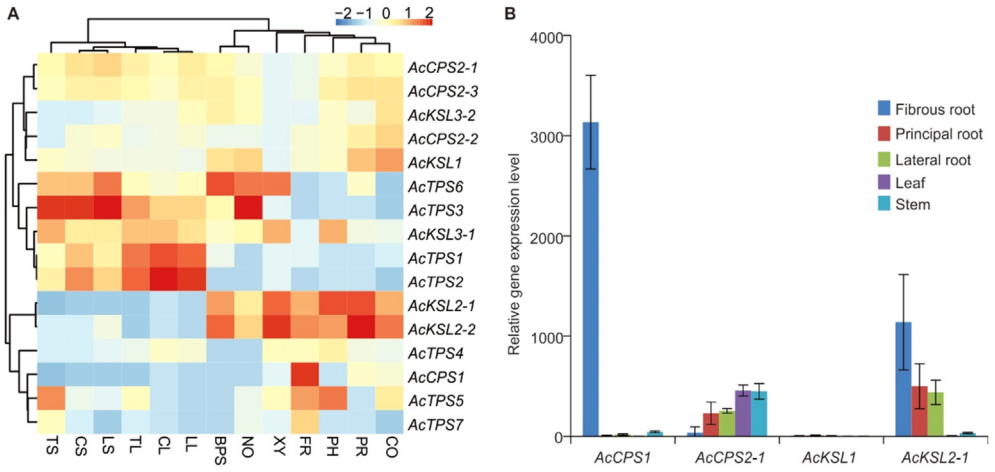


Figure 6 The gene expression patterns of TPS in *A. carmichaelii.* (A) Heat map of 13 spatial tissues obtained by RNA-seq: PR, principal root; FR, fibrous root; CO, cortex; PH, phloem; CY, Xylem; BPS, basal part of the stem; TL, top leaf; CL, central leaf; LL, lower leaf; TS, top stem; CS, central stem; LS, lower stem, NO, nodular like organs. (B) qPCR analysis of 4 diTPSs in 5 tissues. The FKPM values in A and relative expression levels in B were obtained from 3 biological replicates.

root, but lower expression level in stem and leaves; and *AcKSL1* had relative low expression levels in all the tissues.

4. Discussion

Diterpenoid alkaloids (DAs) are the most well-known medicinal compounds of *Aconitum*. Most previous studies have focused on phytochemical composition and pharmaceutical usage of DAs, and significantly less research has gone into the molecular biology of DAs formation in *Aconitum*. To better understand the molecular basis for the complex biosynthesis of DAs; herein, we applied Pacbio full-length transcriptome in conjunction with RNA-seq and PCR cloning to mine for the TPS family and their potential roles, searching specifically for enzymes involved in DAs metabolism. Using this combination approach, we identified 19 TPS genes and five alternative splicing isoforms in *A. carmichaelii*. The diTPS, excepting the alternative splicing isoforms, were functionally identified using *in vitro* and *in vivo* diterpene metabolic engineering system. Using *in vitro* assays—with GPP, FPP, and GGPP as substrates—we eliminated the potential involvement of the TPS-b subfamily in DAs biosynthesis, and established biosynthetic routes to the diterpene scaffolds *ent*-atiserene, *ent*-kaurene, and *ent*-13-*epi*-sandaracopimaradie (Fig. 7) based on 12 diTPS genes, clarifying the molecular basis for *A. carmichaelii* DAs biosynthesis.

The AcTPSs found in the TPS-c and TPS-e/f subfamilies cluster together, analogous to the diTPS found in *T. wilfordii*[25], indicating a recent evolutionary event may responsible to the emergence of different classes of diterpene scaffolds. All of the class II diTPS genes proved to be *ent*-CPP synthase, which is accord with the common view that *ent*-CPP is the only intermediate involved in *A. carmichaelii* DAs biosynthesis[10]. Notably, AcCPS1 is specifically expressed in the fibrous root, as verified by RNA-seq and qPCR analysis, raising interesting questions about its function that necessitate further study. Due to their relatively high expression levels in underground tissues, AcKSL2-1 and AcKSL2-2 are presumed to be the key enzymes involved in DAs biosynthesis. If this is the case, the product of AcKSL2s, *ent*-atiserene, would be the only precursor for atisane-skeleton C_{20}-DAs, such as the atisine-, denudatine-, hetidine-, hetisine- or vakognavine-types, which is agreement with the previous hypothetical biosynthetic pathway based on chemical structure[5]. Besides AcKSL2s, AcKSL1 also identified to be an *ent*-aitiserene synthase in *A. carmichaelii.* The relative low expression level of *AcKSL1* in different tissues (Fig. 6) and the close phylogenic relationship with *ent*-kaurene synthase (AcKSL3-1) (Fig. 2) illustrated AcKSL1 might be the ancestral enzyme for atisane-skeleton C_{20}-DAs biosynthesis, and AcKSL2s undergone a functional divergence from AcKSL1, which shed light on the evolutionary origin of C_{20}-DAs biosynthesis. The product of AcKSL3-1 was *ent*-kaurene, which would be the precursor for the biosynthesis of gibberellin and napelline-type C_{20}-DAs. The product of AcKSL3-2 was found to be *ent*-13-*epi*-sandaracopimaradiene, making it the first enzyme to be identified to synthesize this diterpene. Other *ent*-sandaracopimaradiene-like compound derivatives have been isolated from several species including rice[45], *Kaempferia galanga*[47], and *Guarea rhophalocarpa*[48]; however, it has not reported previously in *Aconitum*. On the other hand, beside *ent*-13-*epi*-sandaracopimaradiene, it also produced *ent*-kaurene and another unknown diterpene in the engineered *E. coli*. Further study into the chemistry and metabolic fate of these diterpenes *in planta* will be needed to fully elucidate its role. Collectively, AcCPS1, AcCPS2s, AcKSL1, AcKSL2s, and AcKSL3-1 appear to be responsible for all the known C_{20}-DAs biosynthesis in *A. carmichaelii*.

Many secondary metabolites are synthesized in highly specific tissues, such as tanshinones and forskolin in the roots. More specifically, tanshinones are synthesized in the root periderm of *S. miltiorrhiza*[27,33] and forskolin is synthesized in the root cork of *Coleus forskohlii*[49]. Accordingly, the diTPS involved in their biosynthesis have specific, high expression levels in these organs[27,33,48]. In our recent study focused on DAs metabolites in different species, we found that despite the principal and lateral root being considered the primary medicinal components of *A. carmichaelii*, the characteristic DAs including aconitine, mesaconitine and hypaconitine have a relatively high content in stem and leaves as well[3]. Indeed, when total DAs content was assessed, the top leaves contained the highest DAs content[3]. It is therefore difficult isolate which enzyme or substrate is responsible for the large array of C_{19}-DAs biosynthetic pathways of *A. carmichaelii*

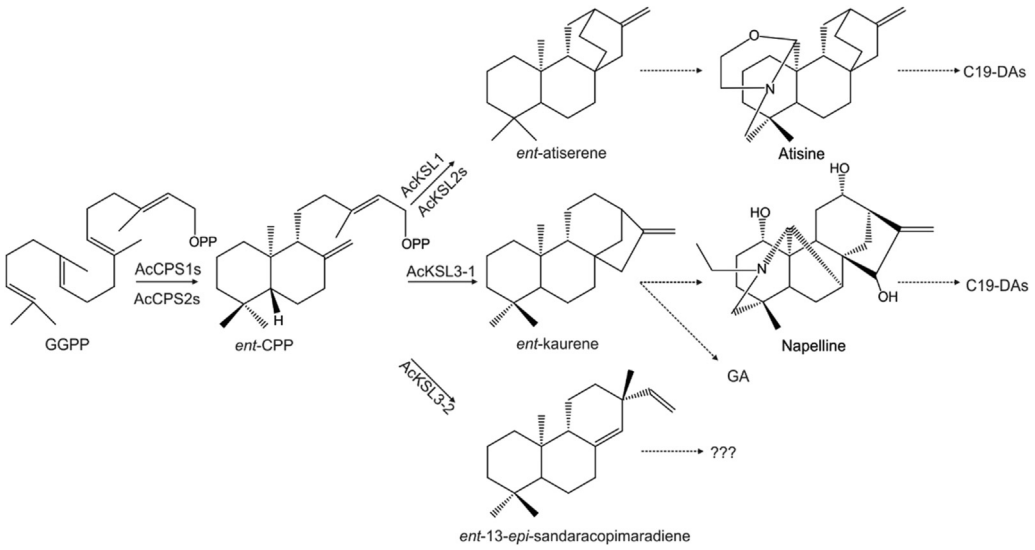


Figure 7 The proposed diterpenoid biosynthesis in *A. carmichaelii*. Twelve identified diterpene synthases, and their corresponding reactions are indicated, along with the possible biosynthetic downstream DAs. Dashed arrows indicate multiple enzymatic reactions.

based on gene expression patterns and metabolite locations alone. Due to the high expression levels of AcKSL2s, we postulate that the atisine type C_{20}-DAs are the main precursor for C_{19}-DAs; however, further proof by RNA interference using microprojectile bombardment (gene gun) or virus-induced gene silencing technology to knock down AcKSLs gene expression *in planta* or through isotope labeling analysis is needed.

The Pacbio isoform sequencing offered information about alternative splicing transcripts[27,50]. In terpene synthase, there were few reports about functional alternative splicing transcripts. In *S. miltiorrhiza*, the kaurene synthase like gene *SmKSL1* and several genes involved in production of isoprenoid precursors were reported to exhibit alternative splicing[27] without functional identification. In *I. rubescens,* the alternative splicing IrKLS3a was identified to alter the product outcome. IrKSL3 produced miltiradiene, while IrKSL3a produced isopimaradiene and miltiradiene[51]. Here, we identified two alternative splicing transcripts of *AcCPS1* and three of *AcKSL1* in *A. carmichaelii*. However, none were identified to be functional genes. Compared with the full-length gene, clone-4-12 from transcript13827 lost α domain of CPS, while transcript12676 lost γ and partial β domain[52], thus it is reasonable for them to lose diTPS functions. However, we are surprised by the loss of function of AcKSL1 and its three alternative splicing transcripts *in vitro*. This was partly due to the low solubility of recombinant protein in prokaryotic expression, after codon optimized of *E. coli* for AcKSL1, we finally identified its function in the highly efficient diterpene modular metabolic engineering system[35,36], indicating we should attempt more platforms such as by transient expression of these genes in *Nicotiana benthamiana*[46] to characterize their functions.

In conclusion, by combining full-length Iso-seq and RNA-seq we obtained a robust *A. carmichaelii* transcriptome and identified all the TPS family genes involved in *A. carmichaelii* DAs biosynthesis. While not conclusive, our results highlight the atisine type C_{20}-DAs as the predominate precursor of the medicinal C_{19}-DAs, clarify the biosynthetic pathway for C_{20}-DAs in *A. carmichaelii*, and pave the way for further exploration of C_{19}-DAs biosynthesis pathway in *Aconitum*.

Acknowledgments

This work was supported by the Major Program of National Natural Science Foundation of China (81891010, 81891013), the National Natural Science Foundation of China (81822046) and National Key R&D Program of China (2018YFA0900600, 2020YFA0908000) and Key project at central government level: the ability to establish sustainable use of valuable Chinese Medicine Resources (2060302, China).

Author contributions

Liuying Mao, Guanghong Cui, Juan Guo and Luqi Huang designed the research. Liuying Mao performed most of the experiment. Changjiangsheng Lai, Biwei Yin prepared the plant material for sequencing. Tong Chen analyzed the transcriptome data. Rui Ma, Haiyan Zhang performed the qPCR analysis. Baolong Jin, Lingli Chen and Mei Tian confirmed the catalytic analysis. Liuying Mao, Baolong Jin and Guanghong Cui wrote the manuscript. Juan Guo, Jinfu Tang revised the manuscript. All authors have read and approved the final manuscript.

Conflict of interest

The authors claim that the researchers in this study have no conflicts of interest.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2021.04.008.

References

1. Singhuber J, Zhu M, Prinz S, Kopp B. Aconitum in traditional Chinese medicine: a valuable drug or an unpredictable risk?. *J Ethnopharmacol* 2009;**126**:18−30.
2. Zhou GH, Tang LY, Zhou XD, Wang T, Kou ZZ, Wang ZJ. A review on phytochemistry and pharmacological activities of the processed

lateral root of *Aconitum carmichaelii* Debeaux. *J Ethnopharmacol* 2015;**160**:173−93.
3. Chen LL, Lai CJS, Mao LY, Yin BW, Tian M, Jin BL, et al. Chemical constituents in different parts of seven species of Aconitum based on UHPLC–Q-TOF/MS. *J Pharmaceut Biomed Anal* 2020;**193**: 113713.
4. Guo QL, Xia H, Shi G, Zhang TT, Shi JG. Aconicarmisulfonine A, a sulfonated C20-diterpenoid alkaloid from the lateral roots of *Aconitum carmichaelii*. *Org Lett* 2018;**20**:816−9.
5. Wang FP, Chen QH. The C19-diterpenoid alkaloids. *Cell Chem Biol* 2010;**69**:482−548.
6. Cherney ME, Baran PS. Terpenoid-alkaloids: their biosynthetic twist of fate and total synthesis. *Isr J Chem* 2011;**51**:391−405.
7. Devkota KP, Sewald N. Terpenoid alkaloids derived by amination reaction. *Nat Prod* 2013;**28**:923−51.
8. Nishiyama Y, Yokoshima S, Fukuyama T. Total synthesis of (−)-cardiopetaline. *Org Lett* 2016;**18**:2359.
9. Nishiyama Y, Yokoshima S, Fukuyama T. Synthesis of cardiopetaline *via* a Wagner-Meerwein rearrangement without preactivation of the pivotal hydroxy group. *Org Lett* 2017;**19**:5833−5.
10. Dank C, Sanichar R, Choo KL, Olsen M, Lautens M. Recent advances towards syntheses of diterpenoid alkaloids. *Synthesis* 2019;**51**:A−AF.
11. Zhao PJ, Gao S, Fan LM, Nie JL, He HP, Zeng Y, et al. Approach to the biosynthesis of Atisine-type diterpenoid alkaloids. *J Nat Prod* 2009;**72**:645−9.
12. Zhang DY, Wen H, Wang W, Peng C, Gao JH. Transcriptional analysis of terpenoid biosynthesis in *Aconitum carmichaelii*. *Zhongguo Shiyan Fangjixue Zazhi* 2017;**23**:45−50.
13. Rai M, Rai A, Kawano N, Yoshimatsu K, Takahashi H, Suzuki H, et al. *De Novo* RNA sequencing and expression analysis of *Aconitum carmichaelii* to analyze key genes involved in the biosynthesis of diterpene alkaloids. *Molecules* 2017;**22**:2155.
14. Zhao D, Shen Y, Shi Y, Shi X, Qiao Q, Zi S, et al. Probing the transcriptome of *Aconitum carmichaelii* reveals the candidate genes associated with the biosynthesis of the toxic aconitine-type C19-diterpenoid alkaloids. *Phytochemistry* 2018;**152**:113−24.
15. Gao JH, Zhang DY, Hou FX, Wen H, Wang QH, Song MW, et al. Transcriptome analysis of *Aconitum carmichaelii* identifies genes involved in terpenoid, alkaloid and phenylpropanoid biosynthesis. *Int J Agric Biol* 2019;**22**:710−20.
16. Yang Y, Hu P, Zhou X, Wu P, Si X, Lu B, et al. Transcriptome analysis of *Aconitum carmichaelii* and exploration of the salsolinol biosynthetic pathway. *Fitoterapia* 2020;**140**:104412.
17. Zhong S, Yin YP, He YN, Li MJ, Zhang M, Li SN, et al. Analysis of transcriptome differences in two leaf-type cultivars of *Aconitum carmichaelii*. *Zhongguo Zhongyao Zazhi* 2020;**45**:1633−40.
18. Pal T, Malhotra N, Chanumolu SK, Chauhan RS. Next-generation sequencing (NGS) transcriptomes reveal association of multiple genes and pathways contributing to secondary metabolites accumulation in tuberous roots of *Aconitum heterophyllum* Wall. *Planta* 2015;**242**: 239−58.
19. Malhotra N, Kumar V, Sood H, Singh TR, Chauhan RS. Multiple genes of mevalonate and non-mevalonate pathways contribute to high aconites content in an endangered medicinal herb. *Aconitum heterophyllum* Wall. *Phytochemistry* 2014;**108**:26−34.
20. Kumar V, Malhotra N, Pal T, Chauhan RS. Molecular dissection of pathway components unravel atisine biosynthesis in a non-toxic Aconitum species, *A. heterophyllum* Wall. *3 Biotech* 2016;**6**:106.
21. Wildung MR, Croteau R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. *J Biol Chem* 1996;**271**:9201−4.
22. Kirby J, Nishimoto M, Park JG, Withers ST, Nowroozi F, Behrendt D, et al. Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces* cerevisiae. *Phytochemistry* 2010;**71**:1466−73.
23. King AJ, Brown GD, Gilday AD, Larson TR, Graham IA. Production of bioactive diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 2014;**26**:3286−98.
24. Mafu S, Karunanithi PS, Palazzo TA, Harrod BL, Rodriguez SM, Mollhoff IN, et al. Biosynthesis of the microtubule-destabilizing diterpene pseudolaric acid B from golden larch involves an unusual diterpene synthase. *Proc Natl Acad Sci U S A* 2017;**114**:974−9.
25. Hansen NL, Heskes AM, Hamberger B, Olsen CE, Hallström BM, Andersen-Ranberg J, et al. The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. *Plant J* 2017;**89**: 429−41.
26. Johnson SR, Bhat WW, Sadre R, Miller GP, Garcia AS, Hamberger B. Promiscuous terpene synthases from *Prunella vulgaris* highlight the importance of substrate and compartment switching in terpene synthase evolution. *New Phytol* 2019;**223**:323−35.
27. Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J* 2015;**82**: 951−61.
28. Xu QS, Zhu JY, Zhao SQ, Hou Y, Li FD, Tai YL, et al. Transcriptome profiling using single-molecule direct RNA sequencing approach for in depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front Plant Sci* 2017;**8**:1205.
29. Miao LL, Zhou QM, Peng C, Meng CW, Wang XY, Xiong L. Discrimination of the geographical origin of the lateral roots of *Aconitum carmichaelii* using the fingerprint, multicomponent quantification, and chemometric methods. *Molecules* 2019;**24**:4124.
30. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 2014;**30**:3506−14.
31. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012; **28**:3150−2.
32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;**30**:2725−9.
33. Cui GH, Duan LX, Jin BL, Qian J, Xue ZY, Shen GA, et al. Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*. *Plant Physiol* 2015;**169**:1607−18.
34. Jia M, Potter KC, Peters RJ. Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. *Metab Eng* 2016;**37**:24−34.
35. Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ. Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl Microbiol Biotechnol* 2010;**85**:1893−906.
36. Cyr A, Wilderman PR, Determan M, Peters RJ. A modular approach for facile biosynthesis of labdane-related diterpenes. *J Am Chem Soc* 2007;**129**:6684−5.
37. Harris LJ, Saparno A, Johnston A, Prisic S, Xu M, Allard S, et al. The maize An2 gene is induced by Fusarium attack and encodes an *ent*-copalyl diphosphate synthase. *Plant Mol Biol* 2005;**59**:881−94.
38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf* 2011; **12**:323.
39. Jin BL, Cui GH, Guo J, Tang JF, Duan LX, Lin HX, et al. Functional diversification of kaurene synthase-like genes in *Isodon rubescens*. *Plant Physiol* 2017;**174**:943−55.
40. Park S, An B, Park S. Recurrent gene duplication in the angiosperm tribe Delphinieae (Ranunculaceae) inferred from intracellular gene transfer events and heteroplasmic mutations in the plastid *matK* gene. *Sci Rep* 2020;**10**:2720.
41. Smith MW, Yamaguchi S, Ait-Ali T, Kamiya Y. The first step of gibberellin biosynthesis in pumpkin is catalyzed by at least two copalyl diphosphate synthases encoded by differentially regulated genes. *Plant physiol* 1998;**118**:1411−9.
42. Chen F, Tholl D, Bohlmann J, Pichersky E. 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* 2011;**66**:212−29.

43. Gao W, Hillwig ML, Huang LQ, Cui GH, Wang XY, Kong JQ, et al. A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org Lett* 2009;**11**:5170−3.
44. Andersen-Ranberg J, Kongstad KT, Nielsen MT, Jensen NB, Pateraki I, Bach SS, et al. Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. *Angew Chem Int Ed Engl* 2016;**55**:2142−6.
45. Xu MM, Wilderman PR, Morrone D, Xu J, Roy A, Margis-Pinheiro M, et al. Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry* 2007;**68**:312−26.
46. Johnson SR, Bhat WW, Bibik J, Turmo A, Hamberger B. Evolutionary mint genomics consortium, et al. A database-driven approach identifies additional diterpene synthase activities in the mint family (Lamiaceae). *J Biol Chem* 2019;**294**:1349−62.
47. Tungcharoen P, Wattanapiromsakul C, Tansakul P, Nakamura S, Matsuda H, Tewtrakul S. Anti-inflammatory effect of isopimarane diterpenoids from *Kaempferia galanga*. *Phytother Res* 2020;**34**:612−23.
48. Camachod RM, Phillipson JD, Croft SL, Kirby GC, Warhurst DC, Solis PN. Terpenoids from *Guarea rhophalocarpa*. *Phytochemistry* 2001;**56**:203−10.
49. Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, Zerbe P, et al. Manoyl oxide (13*R*), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in *Coleus forskohlii*. *Plant Physiol* 2014;**164**:1222−36.
50. Zhang G, Sun M, Wang J, Lei M, Li C, Zhao D, et al. PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *Plant J* 2019;**97**:296−305.
51. Jin B, Guo J, Tang J, Tong Y, Ma Y, Chen T, et al. An alternative splicing alters the product outcome of a class I terpene synthase in *Isodon rubescens*. *Biochem Biophys Res Commun* 2019;**512**:310−3.
52. Köksal M, Hu H, Coates RM, Peters RJ, Christianson DW. Structure and mechanism of the diterpene cyclase *ent*-copalyl diphosphate synthase. *Nat Chem Biol* 2011;**7**:431−3.