

Comprehensive identification of terpene synthase genes and organ-dependent accumulation of terpenoid volatiles in a traditional medicinal plant *Angelica archangelica* L.

Miki Suenaga-Hiromori¹, Daisuke Mogi¹, Yohei Kikuchi¹, Jiali Tong¹,
Naotsugu Kurisu¹, Yuichi Aoki², Hiroyuki Amano¹, Masahiro Furutani³,
Takefumi Shimoyama¹, Toshiyuki Waki¹, Toru Nakayama^{1,*}, Seiji Takahashi^{1,**}

¹ Graduate School of Engineering, Tohoku University, Sendai, Miyagi 980-8579, Japan; ² Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai, Miyagi 980-8573, Japan; ³ R&D Center, Sekisui Chemical Co. Ltd., Tsukuba, Ibaraki 300-4247, Japan

*E-mail: toru.nakayama.e5@tohoku.ac.jp Tel & Fax: +81-22-795-7270

**E-mail: takahashis@tohoku.ac.jp Tel: +81-22-795-7272 Fax: +81-22-795-7271

Received June 6, 2022; accepted October 6, 2022 (Edited by T. Koeduka)

Abstract *Angelica archangelica* L. is a traditional medicinal plant of Nordic origin that produces an unusual amount and variety of terpenoids. The unique terpenoid composition of *A. archangelica* likely arises from the involvement of terpene synthases (TPSs) with different specificities, none of which has been identified. As the first step in identifying TPSs responsible for terpenoid chemodiversity in *A. archangelica*, we produced a transcriptome catalogue using the mRNAs extracted from the leaves, tap roots, and dry seeds of the plant; 11 putative TPS genes were identified (*AaTPS1*–*AaTPS11*). Phylogenetic analysis predicted that *AaTPS1*–*AaTPS5*, *AaTPS6*–*AaTPS10*, and *AaTPS11* belong to the monoterpene synthase (monoTPS), sesquiterpene synthase (sesquiTPS), and diterpene synthase clusters, respectively. We then performed in vivo enzyme assays of the *AaTPSs* using recombinant *Escherichia coli* systems to examine their enzymatic activities and specificities. Nine recombinant enzymes (*AaTPS2*–*AaTPS10*) displayed TPS activities with specificities consistent with their phylogenetics; however, *AaTPS5* exhibited a strong sesquiTPS activity along with a weak monoTPS activity. We also analyzed terpenoid volatiles in the flowers, immature and mature seeds, leaves, and tap roots of *A. archangelica* using gas chromatography-mass spectrometry; 14 monoterpenoids and 13 sesquiterpenoids were identified. The mature seeds accumulated the highest levels of monoterpenoids, with β -phellandrene being the most prominent. α -Pinene and β -myrcene were abundant in all organs examined. The in vivo assay results suggest that the *AaTPSs* functionally identified in this study are at least partly involved in the chemodiversity of terpenoid volatiles in *A. archangelica*.

Key words: *Angelica archangelica* L., monoterpene, sesquiterpene, terpene synthase, terpenoid volatiles.

Introduction

Terpenoids are a large and structurally diverse class of plant-specialized metabolites in which more than 50,000 different structures have been identified, including hemiterpenoids (C_5), monoterpenoids (C_{10}), sesquiterpenoids (C_{15}), and diterpenoids (C_{20}). These metabolites are the fragrance and aroma constituents of plants and are involved in responses to biotic and abiotic stresses as well as in the ecological interactions of plants (Tholl 2015). Terpene synthases (TPSs) catalyze the production of terpenoids—either acyclic or cyclic—from prenyl pyrophosphates. Hemiterpenoids,

monoterpenoids, sesquiterpenoids, and diterpenoids are derived from dimethylallyl pyrophosphate (DMAPP, C_5), geranyl pyrophosphate (GPP, C_{10}), farnesyl pyrophosphate (FPP, C_{15}), and geranylgeranyl pyrophosphate (GGPP, C_{20}), respectively (Karunanithi and Zerbe 2019). Previous studies have shown that angiosperm TPS genes consist of multigene families in a plant lineage-specific manner and are derived from an ancestral TPS gene via repeated gene duplication and divergence, resulting in TPS paralogs with different substrate and product specificities (Chen et al. 2011; Karunanithi and Zerbe 2019). In the reactions catalyzed by TPSs categorized into class I TPS, an

Abbreviations: *AaTPS*, *Angelica archangelica* terpene synthase; DMAPP, dimethylallyl pyrophosphate; FPKM, fragments per kilobase of exon per million mapped reads; FPP, farnesyl pyrophosphate; GC, gas chromatography; GC-MS, gas chromatography-mass spectrometry; GGPP, geranylgeranyl pyrophosphate; GPP, geranyl pyrophosphate; IPP, isopentenyl diphosphate; monoTPS, monoterpene synthase; sesquiTPS, sesquiterpene synthase; TPS, terpene synthase.

Published online December 9, 2022



allylic carbocation intermediate is formed by divalent cation-assisted elimination of the diphosphate group from a prenyl diphosphate substrate. The enormous diversity of terpenoid structure, at least in part, results in myriad ways of folding, isomerization, cyclization, and/or rearrangement of the carbocation intermediates generated during TPS catalysis (Christianson 2017; Hare and Tantillo 2016). However, the mechanism and structural determinants governing TPS specificity remain unclear.

Angelica archangelica L. (garden angelica) is a biennial plant belonging to the Apiaceae family with a Nordic origin and was originally a common vegetable and medicinal herb for the Nordic Sami populations (Ojala 1986). Currently, this plant is being commercially cultivated, mainly in Europe, to produce flavor additives, essential oils, and medicinal products. Previous analyses revealed that *A. archangelica* produces an unusual amount and variety of terpenoids, which are responsible for its unique, complex aroma (Holm et al. 2012; Kerrola and Kallio 1994; Kerrola et al. 1994; Lopes et al. 2004; Nivinskiene et al. 2007). For example, essential oils extracted from the seeds of *A. archangelica* are composed of more than 60 different terpenoid structures, including monoterpenoids (β -phellandrene as the most abundant) and other terpenoids ubiquitously found in various plants, such as α -pinene and D-limonene (Holm et al. 2012; Lopes et al. 2004; Nivinskiene et al. 2007). The distinctiveness of terpenoid metabolism in *A. archangelica* is likely due to the involvement of unique TPSs with different specificities and spatial expression patterns. However, to date, none of the TPSs from *A. archangelica* have been identified and characterized.

In this study, transcriptome analysis of *A. archangelica* was performed to comprehensively identify TPSs expressed in this plant. Moreover, these TPS homologs were expressed in *Escherichia coli* cells to determine their TPS activities and specificities. The activity results of the recombinant TPSs were compared with the organ-dependent terpenoid profiles in *A. archangelica* to understand the unique terpenoid chemodiversity observed in this plant.

Materials and methods

Chemicals

β -Phellandrene, α -phellandrene, (+)- δ -cadinene, *trans*- β -farnesene, γ -terpinene, camphene, α -terpinene, β -fmyrcene, β -elemene, and isobutylbenzene were obtained from Tokyo Chemical Industry (Tokyo, Japan). (–)- α -Pinene, (–)- β -pinene, D-limonene, β -ocimene, 3-carene, and α -humulene were obtained from Sigma-Aldrich/Merck Japan (Tokyo, Japan). Sabinene was obtained from BLD Pharmatech (Shanghai, China). α -Copaene, germacrene D, and germacrene B were obtained from Cayman Chemical Company (Ann

Arbor, MI, USA). *t*-Butyl methyl ether was obtained from Wako Pure Chemical Industries (Osaka, Japan). All other chemicals used were of analytical grade.

Plant materials

Seeds of *A. archangelica* were purchased from Mikasa Engei (Tokyo, Japan). The seeds were sown in potting soil (Dio Chemicals, Tokyo, Japan) and vernalized at 4°C for 10 days. For transcriptome analysis, *A. archangelica* plants were grown for 8 months—before flowering—at 22°C under long-day conditions (16-h light and 8-h dark) of white light (approximately 200 $\mu\text{mol m}^{-2} \text{sec}^{-1}$) in a plant growth cabinet (model KCLP-1400ICT, Nippon Medical & Chemical Instruments, Osaka, Japan). The flowers and roots were harvested, immediately frozen in liquid nitrogen (LN₂), and stored at –80°C until total RNA was extracted from the samples. For terpenoid extraction, *A. archangelica* plants were grown for 26 months until they reached stem heights of 30–40 cm. Flowers, immature and mature seeds, leaves, and tap roots (Figure 1) were harvested, immediately frozen in LN₂ after their fresh weights were measured, and stored at –80°C until further use.

Total RNA extraction

Total RNA was extracted from the mature leaves of *A. archangelica* using the phenol/SDS method (Palmiter 1974). Total RNA in the dry seeds and tap roots of *A. archangelica* was extracted through the method developed by Suzuki et al. (Suzuki et al. 2004). The extracted RNA was treated with DNase I (recombinant, RNase-free; Roche, Basel, Switzerland). The total RNA integrity was assessed using the RNA 6000 Nano Kit with the Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA).

Transcriptome analysis and identification of TPS genes

Preparation of cDNA sequencing libraries using the SureSelect Strand-Specific RNA Library Prep System (Agilent Technologies), pair-end sequencing using the HiSeq 2000 sequencing system (Illumina, San Diego, CA, USA) and de novo assembly were outsourced to Medical & Biological Laboratories (Tokyo, Japan). After pair-end sequencing, the raw reads of each RNA sample were assessed using FastQC v0.11.2 and subjected to quality control by removing low-quality

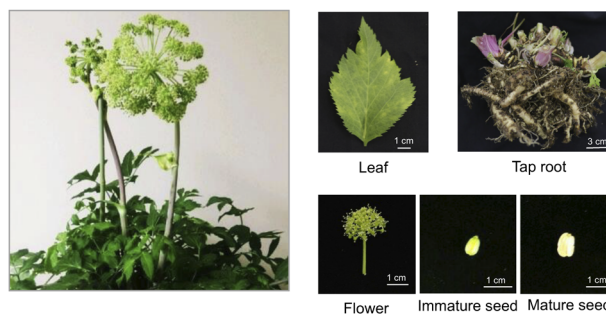


Figure 1. *A. archangelica* plant used in this study.

bases using PRINSEQ 0.20.4 (Schmieder and Edwards 2011); clipping an adapter sequence (5'-AGATCGGAAGAGC-3'); and removing overly short reads (<100 bases) using FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The remaining high-quality trimmed read pairs from three samples (leaves, tap roots, and dry seeds) were collectively assembled using Trinity r2013-02-25 (Grabherr et al. 2011), which yielded 376,899 contigs. After removing contaminated sequences, a non-redundant transcript dataset was generated from the original contigs by mapping each read using RSEM v1.2.30 (Li and Dewey 2011) and removing splicing variants. The resultant unigene dataset (273,968 contigs) was functionally annotated by a BlastX search against a non-redundant National Center for Biotechnology Information protein database, the protein database of *Arabidopsis thaliana* (TAIR10, <https://www.arabidopsis.org>), and the Universal Protein Resource database (<https://www.uniprot.org>). For transcriptome analysis, the trimmed read pairs from each sample were mapped to the 273,968 unigenes using RSEM v1.2.30. The expression level of each transcript was estimated based on the fragments per

kilobase of exon per million mapped reads (FPKM) values.

Unigenes, including a coding sequence highly similar to TPS, were searched by TBLASTN using the TPSs of *A. thaliana* as queries and collectively termed *A. archangelica* TPS (*AaTPS*). Among the 11 unigenes of *AaTPS*, all, except comp128661 (*AaTPS5*), comp123204 (*AaTPS10*), and comp119361 (*AaTPS11*), were considered to include their full-length coding sequences (CDSs) (Table 1). The full-length CDSs of *AaTPS5* and *AaTPS11* were clarified by means of 3'-rapid amplification of cDNA ends (3'-RACE) and 5'-RACE strategies, respectively, using the total RNA prepared from the tap root (see above) and the GeneRacer Kit for full-length, RNA ligase-mediated rapid amplification of 5' and 3' DNA ends (RLM-RACE) (Thermo Fisher Scientific, Waltham, MA, USA). The full-length coding sequence of *AaTPS10* was clarified by means of 5'-RACE using the total RNA prepared from the mature leaves as a template. The total nucleotide sequences of *AaTPS* cDNAs and their deduced amino acid sequences are provided in Supplementary Figures S1 and S2, respectively, and are summarized in Table 1.

Table 1. TPS homologs identified in this study.

Name	Contig ID	CDS (bp)	Protein (aa)	Predicted localization ³⁾	BlastX top hit against uniprot database ⁴⁾	Predicted TPS cluster based on phylogenetics	Major products identified in in vivo enzyme assays
AaTPS1	comp121029	1,854	617	Other	Isoprene synthase from <i>Populus alba</i> (Q50L36, 0.0)	hemiTPS	No product
AaTPS2	comp128785	1,785	594	Chloroplast (42)	(-)- α -Terpineol synthase from <i>Vitis vinifera</i> (Q6PWU2, 0.0)	monoTPS	α -Phellandrene and β -phellandrene
AaTPS3	comp129373	1,770	589	Other	(+)-Limonene synthase from <i>Citrus sinensis</i> (A0A1C9J6A7, 0.0)	monoTPS	α -Pinene
AaTPS4	comp115931	1,758	585	Chloroplast (29)	(+)- α -Terpineol synthase from <i>Vitis vinifera</i> (F6M8I0, 0.0)	monoTPS	α -Pinene
AaTPS5	comp128661 ¹⁾	1,650 ²⁾	549 ²⁾	Other	(R)-Limonene synthase 1 from <i>Vitis vinifera</i> (A0A1C9J6A7, 0.0)	monoTPS	β -Bisabolene
AaTPS6	comp136494	1,713	570	Other	α -Copaene synthase from <i>Eleutherococcus trifoliatus</i> (H9L9E5, 0.0)	sesquiTPS	β -Cedrene, β -barbatene, and β -farnesene
AaTPS7	comp120360	1,692	563	Other	α -Farnesene synthase from <i>Ricinus communis</i> (B9RXW0, 3e-145)	sesquiTPS	α -Bergamotene
AaTPS8	comp129101	1,701	566	Other	Kunzeaol synthase from <i>Thapsia garganica</i> (K4LMW2, 0.0)	sesquiTPS	α -Humulene and germacrene D
AaTPS9	comp127914	1,641	546	Other	Valerianol synthase from <i>Camellia sinensis</i> (A0A167V661, 5e-167)	sesquiTPS	δ -Cadinene
AaTPS10	comp123204 ¹⁾	1,551	516	Other	(3S,6E)-Nerolidol synthase 1 from <i>Fragaria</i> $\times$ <i>ananassa</i> (P0CV94, 0.0)	sesquiTPS	β -Farnesene
AaTPS11	comp119361 ¹⁾	2,325 ²⁾	774 ²⁾	Other	ent-Kaurene synthase from <i>Vitex agnus-castus</i> (A0A2K9RFY0, 0.0)	diTPS	No product

¹⁾Contigs including a partial CDS. ²⁾The sequences for comp119361 and comp128661 were deduced from the full-length cDNA sequences amplified by RT-PCR after determination of the terminal sequences by means of 5'- and 3'-RACE, respectively. ³⁾Predicted by ChloroP. Figures in parenthesis show length (in amino acids) of predicted transit peptide. ⁴⁾Top hit results with accession number and *e*-value in parentheses.

Chemical synthesis of codon-optimized coding sequences of AaTPS

The codon-optimized CDSs of *AaTPS* (except for *AaTPS10*) for heterologous expression in *E. coli*, having *NdeI* and *XhoI* sites on the 5' and 3'-end, respectively, were designed using the GeneArt algorithm (Thermo Fisher Scientific), chemically synthesized by GenScript Biotech Corporation (Tokyo, Japan), and collectively termed synthetic *AaTPS* (*sAaTPS*). It must be mentioned that *N*-terminal 42- and 29-residue sequences of *AaTPS2* and *AaTPS4*, respectively, were predicted to serve as transit peptides for translocation to chloroplast, as analyzed by ChloroP 1.1 Prediction Server (Table 1). Thus, for the expression of these two enzymes, codon-optimized CDSs encoding *N*-terminally truncated forms of *AaTPS2* and *AaTPS4* without the predicted transit peptide were synthesized and termed *sAaTPS2-deltaN* (*sAaTPS2dn*) and *sAaTPS4dn*, respectively. The nucleotide sequences of these codon-optimized CDSs are shown in Supplementary Figure S3.

In vivo enzyme assays of AaTPSs using *E. coli* expression systems

System A

Each of the codon-optimized *sAaTPS* CDSs (except for *AaTPS10*) was co-expressed with *Mycobacterium tuberculosis* GPP synthase (*Rv0989c*) (Mann et al. 2011) in *E. coli* JM109(DE3) cells (Promega, Madison, WI, USA) as follows. The synthesis of a codon-optimized *Rv0989c* gene (*sRv0989c*) with 5'-*BamHI* and 3'-*EcoRI* sites was outsourced to Medical & Biological Laboratories. The *sRv0989c* fragment digested with *BamHI* and *EcoRI* was cloned into multi-cloning site 1 (*BamHI/EcoRI*) of pCOLADuet™-1 (Merck, Darmstadt, Germany) to obtain pCOLADuet-sRv. One of the *sAaTPS* CDSs (except for *AaTPS10*) was digested with *NdeI* and *XhoI* and cloned into multi-cloning site 2 (*NdeI/XhoI*) of pCOLADuet-sRv to obtain a series of plasmids: pCOLADuet-sRv-sAaTPS1 through pCOLADuet-sRv-sAaTPS11 (see Supplementary Figure S4 for general structure). The resulting plasmid was used to transform *E. coli* DH5 α cells (Takara Bio, Shiga, Japan) and purified from the cultured cells using the FastGene® Plasmid Mini Kit (Nippon Genetics, Tokyo, Japan).

Cells of *E. coli* JM109(DE3) were transformed with pCOLADuet-sRv-AaTPS. The transformant cells were grown in LB medium (4 ml) containing 50 mg l⁻¹ ampicillin at 37°C overnight. One hundred microliters of the culture was then inoculated into 40 ml of fresh LB medium and further grown until its optical turbidity at 600 nm reached approximately 0.5. Five milliliter aliquots of the culture were then transferred to sterile 13 ml gas chromatography (GC) glass vials (Agilent Technologies, Santa Clara, CA, USA), and isopropyl β -D-1-thiogalactopyranoside (Nacalai Tesque, Kyoto, Japan) was added to the culture at a final concentration of 1 mM. The vials were sealed with a crimp cap with a polytetrafluoroethylene/silicone septum (Agilent Technologies) and the transformant cells were further grown at 180 rpm for 24 h at 18°C. Terpenoid

volatiles in the headspace of the vials were analyzed by gas chromatography-mass spectrometry (GC-MS) using analytical GC-MS condition 1 (see below).

System B

The *AaTPS10* CDS was amplified by polymerase chain reaction (PCR) using KOD-Plus-Neo DNA polymerase (Toyobo, Osaka, Japan) with a set of primers (123204NdeIFw and 123204XhoIRv; see Supplementary Figure S5). The amplified cDNA was gel-purified using the UltraClean® 15 DNA Purification Kit (Mo Bio Laboratories, Carlsbad, CA, USA), dA-attached using the 10x A-attachment mix (Toyobo), and cloned into the pGEM®-T Easy Vector (Promega). After validating the *AaTPS10* sequence, the resulting plasmid was digested with *NdeI* and *XhoI* to obtain a fragment containing the *AaTPS10* CDS, which was cloned into multi-cloning site 2 (*NdeI/XhoI*) of pCOLADuet™-1 and pCOLADuet-sRv to obtain pCOLADuet-AaTPS10 and pCOLADuet-sRv-AaTPS10, respectively. The resulting plasmids were used to transform *E. coli* Rosetta2(DE3) (Merck) bearing pBR-MScodon, which was introduced to enhance the synthesis of isopentenyl pyrophosphate (IPP) in *E. coli* (see below). Cultivation of transformant cells and GC-MS analysis were carried out as in system A, except that GC-MS analysis was performed with analytical GC-MS condition 1 (see below).

The pBR-MScodon plasmid was constructed as follows. The plasmid pBR322 (Takara Bio) was digested with *SapI* and *BsaAI*, treated with T4 DNA polymerase (Takara Bio) for blunting, and self-ligated to eliminate the *NdeI* site in the plasmid. The resulting plasmid was designated as pBR322N. A DNA fragment containing IPP isomerase of *Saccharomyces cerevisiae* (*ScIDI*) fused with the *rrnB* terminator (*TrrnB*) was amplified from the plasmid pAC-Mev/Scidi, carrying the genes for the mevalonate pathway enzymes of *Streptomyces* sp. strain CL190 and *ScIDI* (Harada et al. 2009), by PCR using primers (*ScidiFw4* and *tacterRv4*) and KOD-Plus-Neo DNA polymerase to attach 5'-*EcoRV* and 3'-*ClaI* sites (see Supplementary Figure S5 for nucleotide sequences of primers and thermal cycling conditions). The amplified fragment was purified, digested with restriction enzymes *EcoRV* and *ClaI*, and ligated with *EcoRV/ClaI*-digested pBR322N using Ligation high Ver. 2 (Toyobo). The resulting plasmid was used to transform *dam*⁻/*dcm*⁻ *E. coli* cells (New England Biolabs Japan, Tokyo, Japan) and purified from the cultured cells to obtain a pBR322N derivative with *ScIDI* and *TrrnB*—pBR-Scidi-TrrnB. The *tac* promoter region (*P_{tac}*) of pAC-Mev/Scidi was also amplified by PCR using primers (*tacproFw4* and *tacproRv3*) and KOD-Plus-Neo DNA polymerase to attach 5'-*BspEI* and 3'-*NdeI-EcoRV* sites (see Supplementary Figure S5). The amplified fragment was purified, digested with *BspEI* and *EcoRV*, and ligated with *BspEI/EcoRV*-digested pBR-Scidi-TrrnB to obtain pBR-Ptac-Scidi-TrrnB. pAC-Mev/Scidi was digested with *NdeI* and *HindIII* to obtain a DNA fragment coding for the mevalonate pathway enzymes of *Streptomyces* sp. strain CL190 and *ScIDI*. That DNA fragment was subsequently ligated with *NdeI/*

HindIII-digested pBR-Ptac-Scidi-TrnB to produce pBR-MS. To replace *ScIDI* in pBR-MS with a codon-optimized *ScIDI*, *SD-ScIDIopt*, which is a DNA fragment with 5'-EcoRV and 3'-SpeI-HindIII sites that encodes a codon-optimized *ScIDI* gene attached to a Shine–Dalgarno sequence (AGGAGG), was designed using the GeneOptimizer algorithm (Thermo Fisher Scientific) and chemically synthesized by Integrated DNA Technologies (Tokyo, Japan). Finally, the EcoRV/HindIII-digested *SD-ScIDIopt* and EcoRV/HindIII-digested pBR-MS fragments were ligated to obtain the pBR-MScodon.

Analysis of terpenoids in *A. archangelica*

Each harvested organ and tissue (i.e., flowers, immature and mature seeds, leaves, and tap roots) of *A. archangelica* (Figure 1) was homogenized in LN₂ using a motor and pestle. Each homogenized plant material (0.3 g of fresh weight) was suspended in 1.5 ml of *t*-butyl methyl ether supplemented with isobutylbenzene (1.2 mM) as an internal standard and shaken overnight at room temperature using a microtube mixer (Model MT-360, TOMY Seiko, Tokyo, Japan) to extract terpenoids. After the mixture was centrifuged, an aliquot of its organic phase (500 μ l) was collected and vigorously mixed with an equal volume of 0.1 M ammonium carbonate. This step was followed by another round of centrifugation. The resulting organic phase (400 μ l) was collected, diluted with 800 μ l of *t*-butyl methyl ether (with 1.2 mM isobutylbenzene), and subjected to GC-MS analysis (condition 2, see below).

GC-MS analysis

GC-MS analysis was performed using a 6890N GC apparatus (Agilent Technologies) and a quadrupole mass selective detector (MSD; Model 5979C, Agilent Technologies) under one of the following conditions.

Condition 1

The vials containing the *E. coli* culture were kept at 80°C and shaken at a low speed using a network headspace sampler G1888 (Agilent Technologies) for 15 min. The headspace sample was injected for 0.15 min in splitless mode, where the temperatures of the loop, transfer line, and injector were maintained at 110, 120, and 250°C, respectively. Separation was achieved using a DB-1701 column (30 m long and 0.25 mm in diameter; Agilent Technologies). The flow rate of the carrier gas (helium) was set to 1.0 ml min⁻¹. The column oven temperature was programmed as follows: the initial oven temperature at 50°C was maintained for 1 min; it was increased to 150°C (at a rate of 5°C min⁻¹) and then to 250°C (at a rate of 100°C min⁻¹), which was sustained for 1 min. The MSD was operated in scan mode, where data acquisition was set between *m/z* 10 and 500 and the solvent delay was set at 3 min. The instrumental parameters were set as follows: the transfer line, source, and quadrupole temperatures were maintained at 250, 230, and 150°C, respectively, and the electron energy was set at 70 eV.

Condition 2

One microliter of the *t*-butyl methyl ether extracts from *A. archangelica* tissues, including isobutylbenzene as an internal standard, was injected into HP-INNOWAX polyethylene glycol (30 m long and 0.25 mm in diameter; Agilent Technologies). The injector was set to a temperature of 250°C for injection at a split ratio of 20:1. The column oven temperature was programmed as follows: the initial temperature at 75°C was maintained for 4 min; it was increased to 250°C (at a rate of 4°C min⁻¹), which was sustained for 10 min. The MSD was operated in scan mode, where data acquisition was set between *m/z* 20 and 300, and the solvent delay was set at 1.9 min. The instrumental parameters were set as follows: the transfer line, source, and quadrupole temperatures were maintained at 250, 230, and 150°C, respectively, and the electron energy was set at 70 eV.

Monoterpenoids and sesquiterpenoids were identified by comparing their retention times and mass spectrums to those for authentic standards or by referencing their mass spectral data to the National Institute of Standards and Technology (NIST) database (<https://webbook.nist.gov/chemistry/name-ser/>) with a threshold coincidence rate of 90% and comparing their retention times to GC retention indexes in the NIST database. The standard curves of isobutylbenzene, β -phellandrene, and β -elemene were obtained by GC-MS analysis of the authentic samples after appropriate dilutions with *t*-butyl methyl ether (condition 2). After normalizing each extract with isobutylbenzene using the standard curve, β -phellandrene and β -elemene levels in each tissue were determined from the peak areas obtained by GC-MS analysis and the corresponding standard curves; the concentrations of other monoterpenoids and sesquiterpenoids were estimated using the β -phellandrene and β -elemene standard curves, respectively.

Results and discussion

Identification of TPS homologs via transcriptome analysis

As the first step in identifying the TPSs responsible for terpenoid chemodiversity in *A. archangelica*, we produced a transcriptome catalogue using the mRNAs collected from its leaves, tap roots, and dry seeds. For the transcriptome analysis, *A. archangelica* dry seeds were commercially obtained for RNA extraction because *A. archangelica* is a biennial plant; therefore, seeds were not immediately harvested. Leaves and tap roots were harvested from *A. archangelica* plants grown in our laboratory. The high-quality read pairs obtained from RNA sequencing of these three samples were collectively assembled to provide 376,899 contigs and 273,968 unigene sequences, and the resultant unigene dataset was functionally annotated as described in the section of Materials and methods. As a result, we identified 11 expressed TPS paralogs: *AaTPS1*–*AaTPS11* (Figure 2,

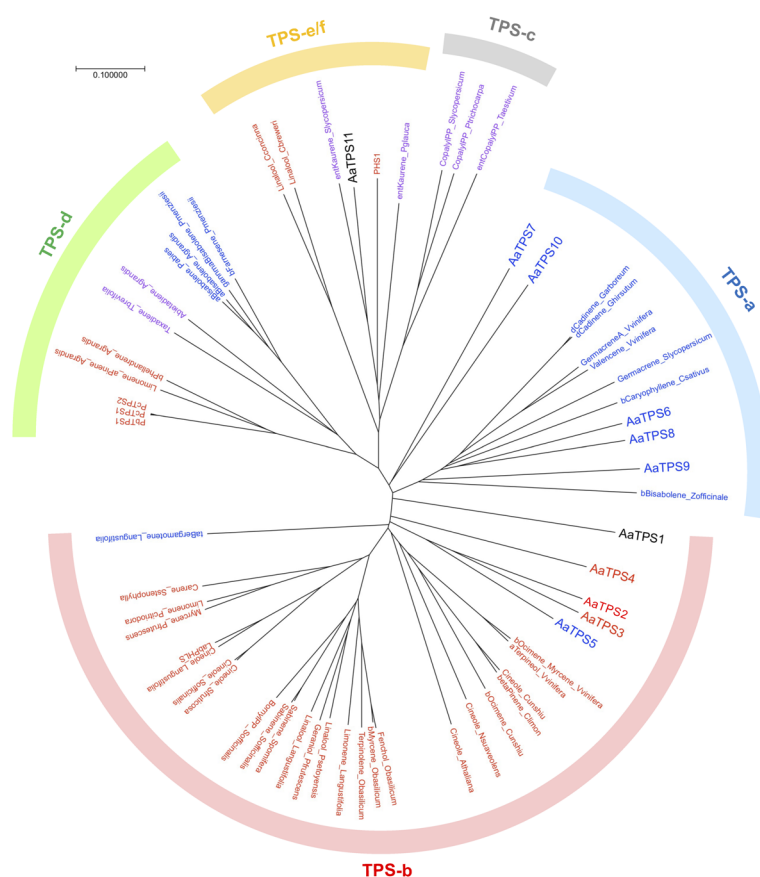


Figure 2. Phylogenetic analysis of TPSs from *A. archangelica*. The tree was constructed by the neighbor-joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The phylogenetic TPS clades (TPS-a, TPS-b, TPS-c, TPS-d, and TPS-e/f) are indicated by blue, red, gray, green arcs, and yellow, respectively. Bar=0.1 amino acid substitution per site. Abbreviated names of MonoTPSs, sesquiTPS, and diTPS in the entries are colored in red, blue, and purple, respectively. Abbreviations and accession numbers of TPS sequences used in this analysis are shown in Supplementary Table S2.

Table 1). Among the 11 TPS homologs, the full-length CDSs for *AaTPS5*, *AaTPS10*, and *AaTPS11* were clarified by cloning the corresponding cDNAs by means of 5'- and 3'-RACE; other TPS paralogs were predicted to encode full-length CDSs (see Supplementary Figures S1 and S2, respectively, for the nucleotide and predicted amino acid sequences). The following characteristics of the 11 *AaTPS*s are summarized in Table 1: the lengths of the CDS (in base pairs) and polypeptide chains (in amino acids), predicted subcellular localizations (based on ChloroP 1.1), BlastP top-hit results, and predicted TPS clusters based on phylogenetics (see below). The FPKM values of these *AaTPS* genes in dry seeds, tap roots, and leaves are shown in Supplementary Figure S6.

The primary structures of these 11 *AaTPS*s shared the structural characteristics conserved among class I plant TPSs, which include the DDXXD and NSE/DTE motifs for divalent cation-mediated holding and elimination of a diphosphate group from a substrate (see Supplementary Figure S7) (Chen et al. 2011); however, *AaTPS1*, having DXDD instead of DDXXD, is an exception. Plant TPS have been classified into

seven phylogenetic clades: TPS-a through TPS-d, TPS-e/f, TPS-g, and TPS-h (Chen et al. 2011; Falara et al. 2011). From the perspective of TPS specificity, most TPSs belonging to the clades TPS-b and TPS-g have been shown to be monoterpene synthases (monoTPSs) and can be grouped into the monoTPS cluster. Clade TPS-a mostly consists of sesquiterpene synthases (sesquiTPSs), forming the sesquiTPS cluster, while TPS-c and TPS-e/f consist mostly of diterpene synthases (diTPSs) and can be put together into the diTPS cluster (Falara et al. 2011). Phylogenetic analysis revealed that each of the 11 *AaTPS*s could be assigned to clades TPS-a (*AaTPS6*–*AaTPS10*), TPS-b (*AaTPS1*–*AaTPS5*), and TPS-e/f (*AaTPS11*), allowing us to predict the specificity of each *AaTPS* (Table 1 and Figure 2). Some *AaTPS*s (*AaTPS2*–*AaTPS6*) also share the RRR₅W motif, which has been proposed to be important in monoterpene ring closure—a process catalyzed by monoTPSs and some sesquiTPSs (see Supplementary Figure S7) (Chen et al. 2011).

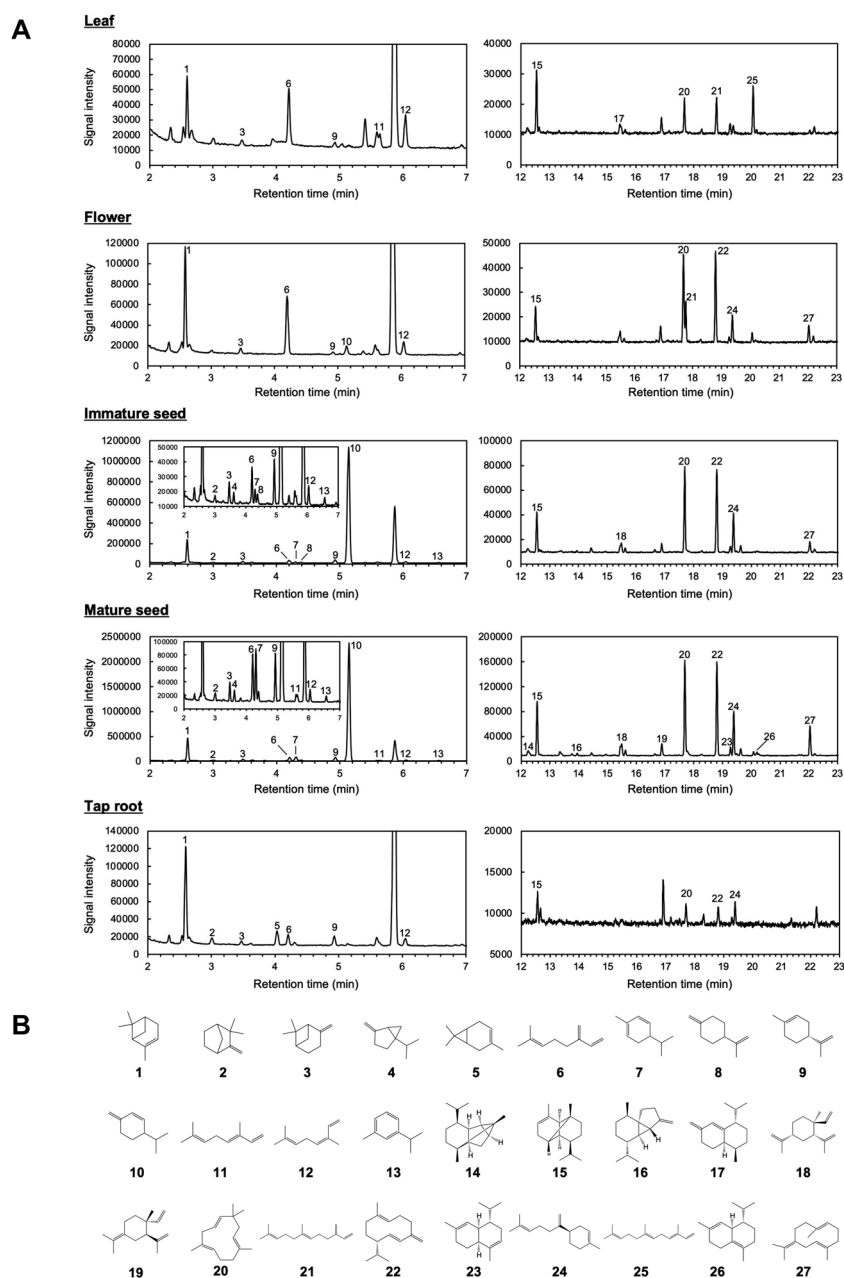


Figure 3. GC chromatograms of terpenoid volatiles extracted from different tissues and organs in *A. archangelica*. (A) From top to bottom panels, leaf, flower, immature seeds, mature seeds, and tap root. Left panels, monoterpenoids and right panels, sesquiterpenoids. The monoterpenoids and sesquiterpenoids were identified by referencing their mass spectral data to the NIST database with a threshold coincidence rate of 90% (see Supplementary Figure S8 for details). Peak at 5.87 min in left panels corresponds to isobutylbenzene added as a spike control. (B) The identified metabolites, numbered (1 through 13 for monoterpenoids and derivatives, and 14 through 27 for sesquiterpenoids), and their structures.

Organ/tissue-dependent compositions of terpenoid volatiles

A. archangelica plants were grown as described in the section of Materials and methods; mono- and sesquiterpenoids were extracted from the flowers, immature and mature seeds, leaves, and tap roots (see Figure 1) and analyzed by GC-MS analysis. The results are presented in Figures 3–5, and Supplementary Figures S8 and S9, and summarized in Supplementary Table S1.

GC-MS analysis identified a total of 13 monoterpenoids and 14 sesquiterpenoids in the *A.*

archangelica plant, with varying compositions in the organs and tissues examined (Figures 3 and 4). The total content of monoterpenoids was higher than that of sesquiterpenoids in all organs and /tissues examined (Figure 5A), concordant with terpenoid compositions of root and seed oils (Holm et al. 2012; Lopes et al. 2004; Nivinskiene et al. 2007) (Figure 5A). Among the organs and tissues examined, the immature and mature seeds accumulated the highest levels of monoterpenoids; β -phellandrene **10** was the most prominent terpenoid [73% and 71% (mol/mol)

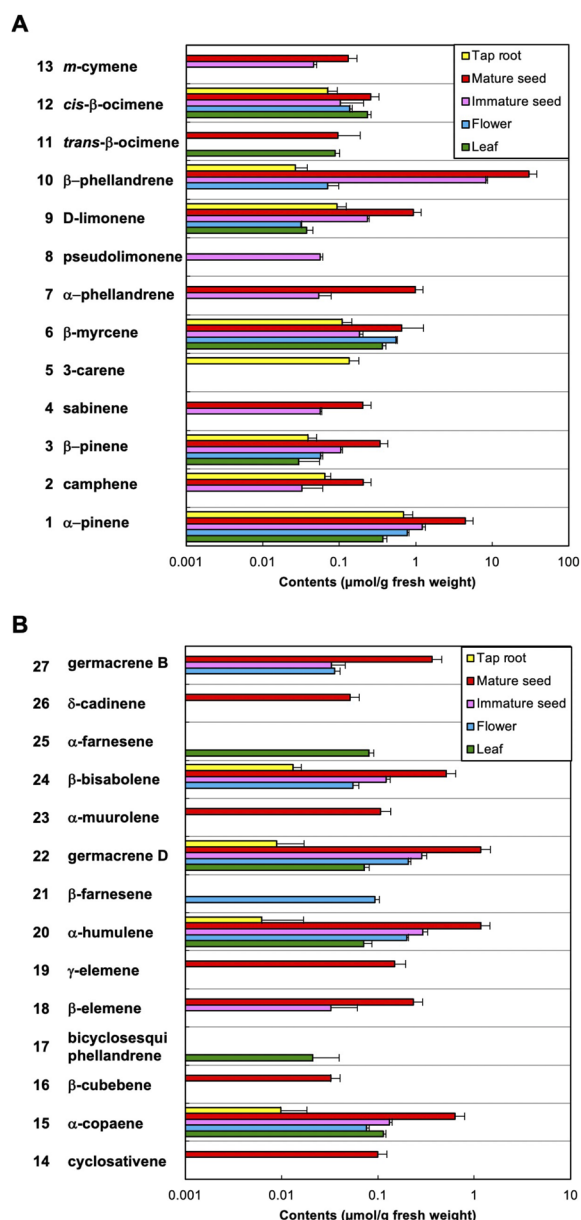


Figure 4. Terpenoid contents in different tissues and organs in *A. archangelica*. (A) Monoterpenoids and derivatives. (B) Sesquiterpenoids. In panels (A) and (B), the contents in tap roots, mature seeds, immature seeds, flower, and leaf (from top to bottom; see also Figure 1) are shown with yellow, red, pink, blue, and green bars, in which data are presented as the average of three independent determinations ($\pm$ standard deviation). Amounts of monoterpene and sesquiterpenoids were expressed as equivalents for β -phellandrene and β -elemene, respectively. Note that the horizontal axis is logarithmic. The identified metabolites are numbered as in Figure 3.

of the total terpenoid volatiles, respectively] (Figures 3, 4A, 5B), as previously reported (Lopes et al. 2004; Nivinskiene et al. 2007). α -Pinene 1 and β -myrcene 6 were relatively abundant in all organs examined in this study (Figures 3, 4A, 5B). More specifically, α -pinene 1 was the most prominent terpenoid in the tap roots and flowers [55% and 35% (mol/mol), respectively] and the second major constituent in the mature and immature

seeds [10% and 11% (mol/mol), respectively] (Figure 5B). β -Myrcene 6 accounted for 25%, 25%, and 8.6% (mol/mol) of the total terpenoid volatiles in the leaves, flowers, and tap roots, respectively, but only 1.5%–1.7% (mol/mol) in the mature and immature seeds (Figure 5B). Sesquiterpenoids were the most abundant in the mature seeds (Figure 5A); among which, germacrene D 22 and α -humulene 20 were the most prominent (Figure 4B). These two sesquiterpenoids also accumulated in all other organs and tissues examined; interestingly, the organ/tissue-dependent accumulation patterns of these sesquiterpenoids were very similar to each other (Figure 4B). It must be mentioned that β -elemene and γ -elemene (Figure 4B, 18 and 19, respectively) are known to be formed in the Cope rearrangement from germacrene A and germacrene B, respectively, under GC conditions (Adio 2009; de Kraker et al. 1998). Thus, the detection of β -elemene and γ -elemene in the GC-MS analysis potentially indicates the occurrence of germacrene A and germacrene B in the harvested organs. Moreover, some terpenoid volatiles identified were not considered as primary enzymatic reaction products. For example, pseudolimonene (Figure 4A, 8) can be formed non-enzymatically from β -pinene (Meehan-Atrash et al. 2021; Sun et al. 2014). *m*-Cymene (Figure 4A, 13) was considered to be formed by the spontaneous dehydration of a monoterpene alcohol formed by a TPS and a subsequent cytochrome P450 enzyme, similar to *p*-cymene formation (Krause et al. 2021).

In vivo assay of AaTPSs

We heterologously expressed AaTPSs in *E. coli* cells to examine their enzymatic activities and specificities. To this end, the codon-optimized genes encoding *N*-terminally truncated forms of AaTPS2 and AaTPS4 (*sAaTPS2dn* and *sAaTPS4dn*, respectively) and the full-length amino acid sequences of other AaTPSs (*sAaTPS*), except for AaTPS10, were chemically synthesized (Supplementary Figure S3 for the nucleotide sequences). Each codon-optimized AaTPS or AaTPS10 cDNA was co-expressed with a codon-optimized *M. tuberculosis* GPP synthase gene (*sRv0989c*) (Mann et al. 2011) in *E. coli* cells to furnish GPP as the substrate for monoTPS using assay systems A or B (see the section of Materials and methods; see also Supplementary Figure S4 for the metabolic pathways). In these systems, FPP, as the substrate of sesquiTPS, was provided by the activity of the endogenous FPP synthase. To facilitate the expression of codon-optimized AaTPS genes, *E. coli* strain JM109(DE3) was utilized as the host for system A; *E. coli* strain Rosetta2(DE3) was used as the host for AaTPS10 expression in system B to enhance the translation efficiency of AaTPS10, which contained codons rarely used in *E. coli*. Moreover, in system B, the host *E. coli* cells also expressed actinomycetous mevalonate

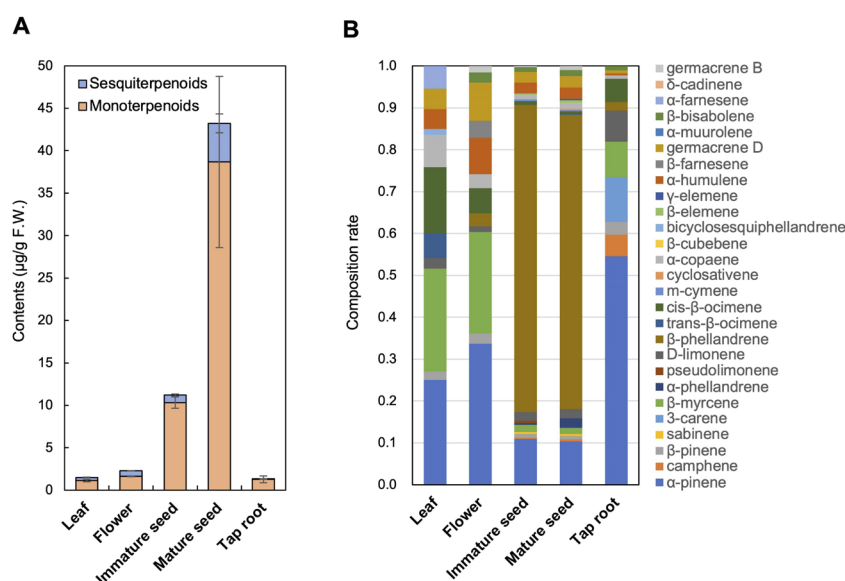


Figure 5. Organ/tissue-dependent distributions of terpenoids in *A. archangelica*. (A) Total mono- and sesquiterpenoid contents. (B) Organ/tissue-dependent terpenoid compositions. Values are calculated using the data presented in Figure 4.

pathway enzymes and yeast *ScIDI* (Harada et al. 2009) to amplify the supply of IPP, a common precursor of terpenoids (see the section of Materials and methods; see also Supplementary Figure S4 for the metabolic pathways designed for system B), resulting in enhanced production of terpenoids by AaTPS10 whose expression level in *E. coli* was lower than those of codon-optimized AaTPSs. In both systems, the *E. coli* transformant cells expressing each AaTPS were grown in a sealed GC vial (see the section of Materials and methods for cultivation conditions), and the terpenoid volatiles in the headspace of the vials were analyzed by GC-MS (see the section of Materials and methods for GC-MS conditions 1). The results are presented in Figure 6, and Supplementary Figures S10–S12. When compared with the volatile metabolites from a control strain that did not bear AaTPS CDS, all transgenic strains expressing each AaTPS, except for AaTPS1 and AaTPS11, accumulated terpenoids derived from the exogenous enzyme.

The results showed that the AaTPSs of clade TPS-b (i.e., AaTPS2, AaTPS3, AaTPS4, and AaTPS5) and those of clade TPS-a (i.e., AaTPS6, AaTPS7, AaTPS8, AaTPS9, and AaTPS10) consistently produced monoterpenoids and sesquiterpenoids (Figure 6 and Table 1), respectively. Interestingly, AaTPS5, a member of clade TPS-b, displayed a strong sesquiTPS activity to produce β -bisabolene 24. Its sesquiTPS activity was much higher than its monoTPS activity (Figure 6), strongly suggesting that AaTPS5 was a sesquiTPS that had evolved from a monoTPS-related ancestor. Detailed analyses of the phylogenetics, specificity, biochemical properties, and subcellular localization of AaTPS5 that would provide important information concerning the evolutionary aspects of TPS specificity and functions will be reported

elsewhere.

In this study, distinct TPS activities for AaTPS1 and AaTPS11 were not detected. Based on amino acid sequence similarity, AaTPS1 was expected to encode an isoprene synthase (IspS) (Table 1). We attempted to detect the formation of isoprene from DMAPP in an *E. coli* strain expressing sAaTPS1 under various GC-MS conditions, but failed (data not shown). The fact that the IspS responsible for isoprene emission is not conserved in all plant species (Sharkey et al. 2008), and that isoprene could not be detected in any of the terpenoid tissues of *A. archangelica* analyzed in this study, even though the transcript levels of AaTPS1 in leaves were comparable to those of other AaTPS paralogs (Supplementary Figure S6), suggest that AaTPS1 is not likely to encode IspS. Although the putative transit peptide was not predicted in the AaTPS1 sequence using the TargetP algorithm, an N-terminal extension of the amino acid sequence of AaTPS1 was found in comparison with those of sesquiTPSs (i.e., AaTPS6, AaTPS7, AaTPS8, AaTPS9, and AaTPS10; Supplementary Figure S7). This suggests that AaTPS1 may have an N-terminal transit peptide. N-terminal extensions, including transit peptides, seldom negatively affect the functional expression of plastidial TPSs in the *E. coli* expression system, causing inclusion body formation (Bohlmann et al. 1998) and interference with the catalytic activity (Lima et al. 2013; Williams et al. 1998) of recombinant TPSs. Therefore, removing the N-terminal extension is a promising approach for the functional expression of recombinant AaTPS1.

AaTPS11 belongs to clade TPS-e/f and is closely related to TPS4 from *Vitex agnus-castus* (VacTPS4; 58% amino acid sequence identities) and *ent*-kaurene

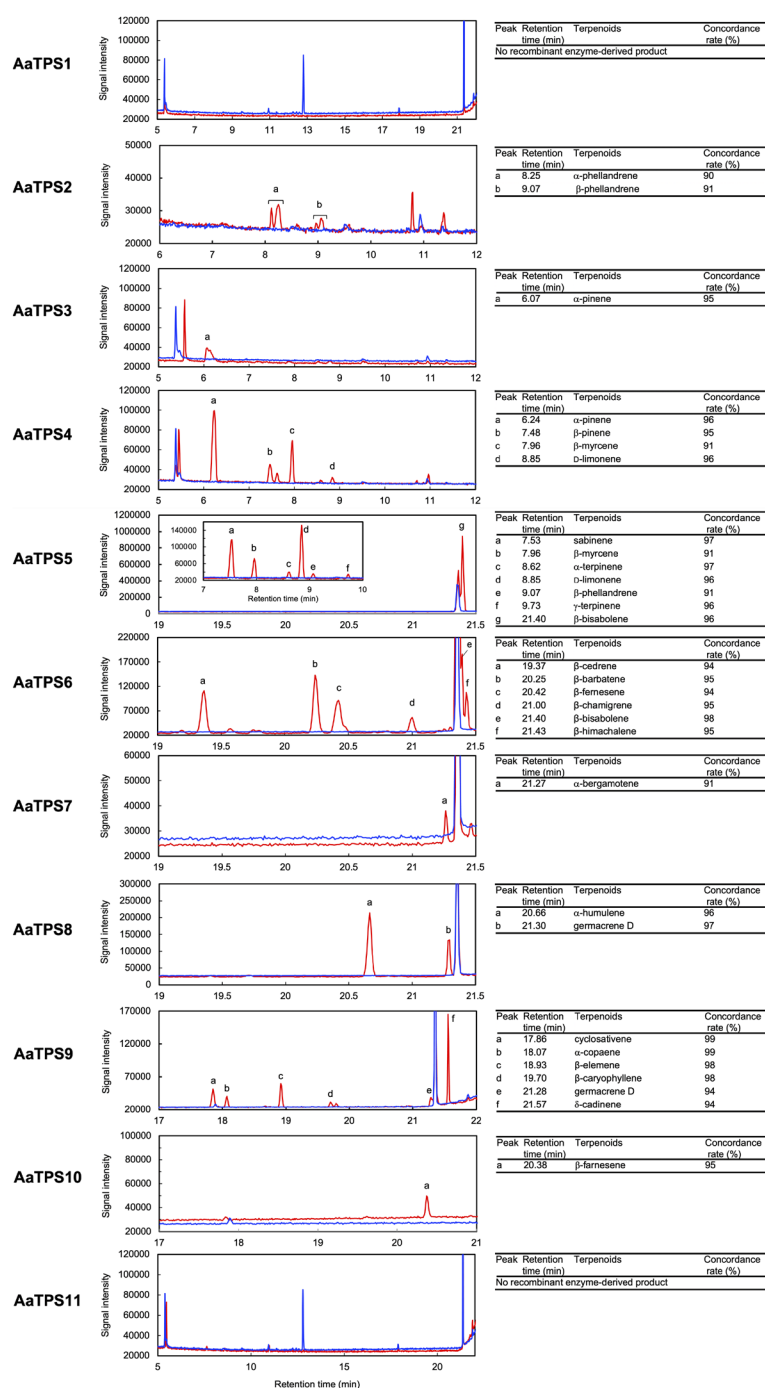


Figure 6. Terpenoids detected in cultures of the transgenic *E. coli* strains expressing AaTPSs. Left panels show GC chromatograms of in vivo enzyme assays of AaTPS1 through AaTPS11 (from top to bottom), with a table right to each GC chromatogram showing peak identification by referencing their mass spectral data to the NIST database with a threshold coincidence rate of 90%. Total ion chromatograms in red and blue corresponding data for the *E. coli* strains expressing each AaTPS and that no exogenous TPS.

synthase (Heskes et al. 2018). In the gibberellin biosynthesis of vascular plants, GGPP is converted to *ent*-kaurene in a sequential reaction catalyzed by two enzymes; GGPP is converted to *ent*-copalyl diphosphate (CPP) by a class II diTPS: *ent*-CPP synthase (TPS-c clade). This step is followed by the conversion of *ent*-CPP to *ent*-kaurene by a class I diTPS: *ent*-kaurene synthase (TPS-e/f clade) (Chen et al. 2011; Karunanithi

and Zerbe 2019). The TBLASTN search against the *A. archangelica* contigs using the *ent*-CPP synthase from *A. thaliana* as a query revealed the existence of three contigs encoding partial protein sequences that showed high similarity to the different regions of *ent*-CPP synthase (data not shown); however, these partial sequences could not be assembled as one unigene covering a full-length CDS. Therefore, *AaTPS11* might encode an *ent*-kaurene

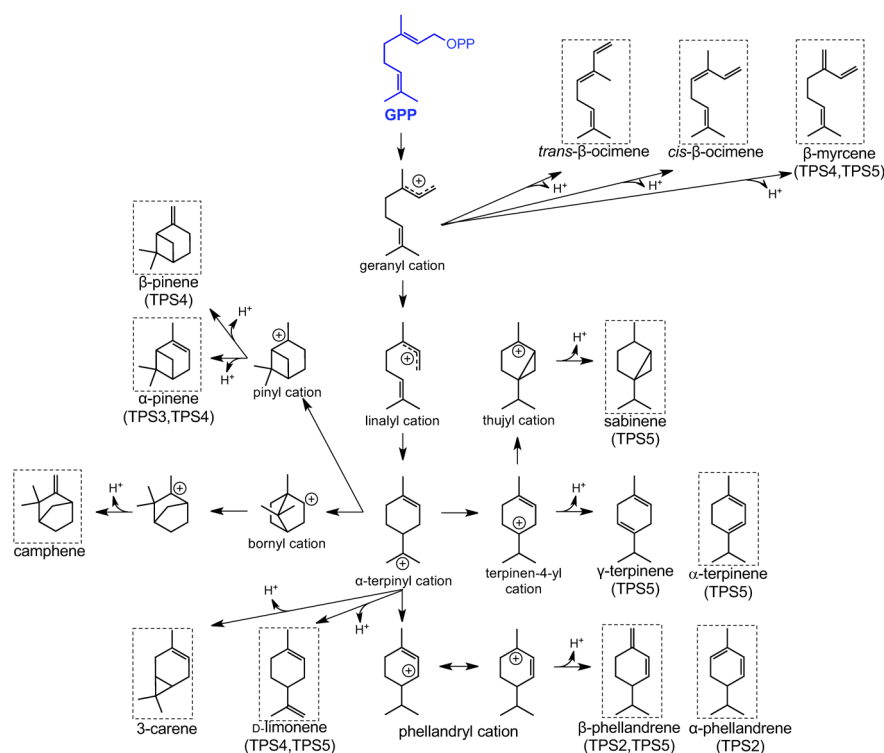


Figure 7. Proposed reaction pathways of TPS-catalyzed formation of monoterpeneoids from GPP. Structures of *A. archangelica* terpeneoid volatiles identified in this study are enclosed by dashed squares. TPS names are shown in parenthesis below the metabolite structures if the enzymes produce the metabolite.

synthase that functions in gibberellin biosynthesis along with an unidentified *ent*-CPP synthase. Another possibility that AaTPS11 might function as a TPS that could use oligoprenyl diphosphate as a substrate was also raised by a report on the identification of β -phellandrene synthase from *Solanum lycopersicum* PHS1 (Schillmiller et al. 2009). PHS1 belongs to the TPS-e/f clade and primarily consists of class I diTPSs (Figure 2); however, its catalysis of a reaction forming some monoterpeneoids, with β -phellandrene as the major product, indicates its unusual substrate preference. PHS1 scarcely accepts GPP but prefers neryl diphosphate (NPP), the *cis*-isomer of GPP. To examine the monoTPS activity of AaTPS11, we heterologously expressed the full-length AaTPS11 as well as a few *N*-terminal truncated forms of AaTPS11—lacking residues expected to include transit peptides—in *E. coli* and conducted *in vitro* TPS assay using the crude protein extracts prepared from *E. coli* cells and various substrates: DMAPP, GPP, FPP, GGPP, and NPP. However, no distinct TPS activity was detected under any reaction conditions (data not shown).

Possible involvement of AaTPSs in terpeneoid chemodiversity in *A. archangelica*

The enzymatic formation of terpeneoid volatiles identified in *A. archangelica* plants could be explained in terms of the generally accepted reaction pathways of TPS catalysis. Specifically, all identified monoterpeneoids were

considered to be produced via the proposed monoTPS reaction pathways involving geranyl and α -terpinyl cations as key intermediates (Figure 7). The formation of sesquiterpeneoids in *A. archangelica* can also be explained by the proposed sesquiTPS reaction pathways, which involve farnesyl cations and their cyclized isomers (e.g., cadinanyl, germacradienyl, humulyl, and bisabolyl cations) as key intermediates (Figure 8). The present *in vivo* assay results strongly suggest that the AaTPSs identified in this study are at least partly involved in the chemodiversity of terpeneoid volatiles in *A. archangelica* plants. Especially for the AaTPSs in TPS-b clade, all of monoterpeneoids derived from heterologously expressed TPSs in *E. coli*, except for γ -terpinene, were identified in *A. archangelica* tissues (Figure 7). On the other hand, some sesquiterpeneoids detected in the *E. coli* culture systems expressing sAaTPS6, sAaTPS7, or sAaTPS9, i.e., β -cedrene, β -himachalene, β -chamigrene, β -barbatene, α -bergamotene, and β -caryophyllene, could not be detected in any *A. archangelica* tissues, suggesting lower catalytic activity than other sesquiTPSs or lower expression levels in the tissues analyzed. In this study, we were unable to identify TPSs that were capable of producing the following terpeneoid volatiles found in *A. archangelica* plants: monoterpeneoids (see Figure 4A), *trans*- and *cis*-ocimenes (11 and 12, respectively), camphene 2, and 3-carene 5; and sesquiterpeneoids (Figure 4B), α -farnesene 25,

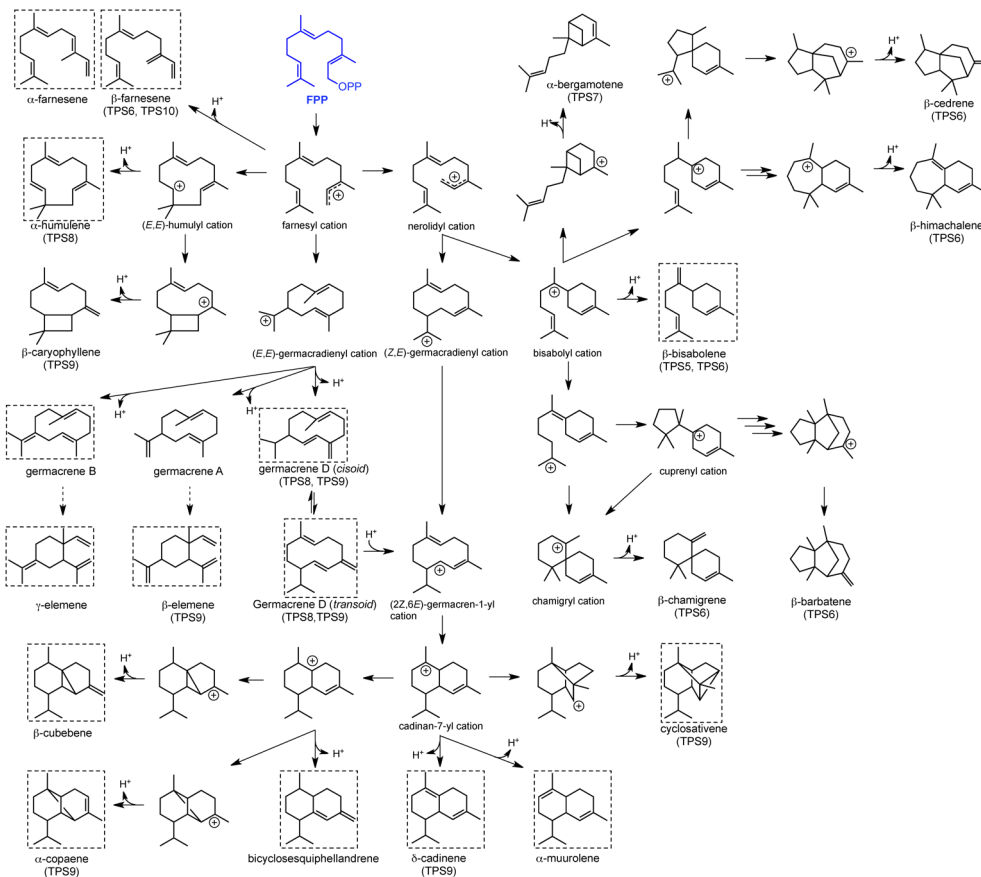


Figure 8. Proposed reaction pathways of TPS-catalyzed formation of sesquiterpenoids from FPP. Structures of *A. archangelica* terpenoid volatiles identified in this study are enclosed by dashed squares. TPS names are shown in parenthesis below the metabolite structures if the enzymes produce the metabolite.

α -muurolene **23**, γ -elemene **20** (germacrene B **27**, see above), bicyclosesquiphellandrene **17**, and β -cubebene **16**. Other unknown TPSs should be responsible for the accumulation of these terpenoid volatiles. This should be addressed in future studies.

Although the precise roles of the respective AaTPSS in organ/tissue-dependent accumulation of terpenoid volatiles in *A. archangelica* remain to be clarified in future studies, it would be useful to discuss the possible involvement of AaTPSS in the accumulation of the major terpenoid volatiles in this plant.

β-Phellandrene

β -Phellandrene **10** is the most prominent terpenoid in the seeds, with up to 71% of total terpenoids (see above). Among the 11 AaTPSs identified in this study, AaTPS2 was the only TPS capable of producing β -phellandrene **10**, as determined by in vivo enzyme assays (Figure 6). However, AaTPS2 was unlikely to be responsible for the accumulation of β -phellandrene **10** in the seeds because it produced α -phellandrene **7** and β -phellandrene **10** at an approximate molar ratio of 2:1 (Figure 6); this product ratio did not account for the exclusive accumulation of β -phellandrene **10** over α -phellandrene

7 in the seeds. Moreover, the FPKM values of *AaTPS2* obtained from transcriptome analysis (Supplementary Figure S6) did not support the organ/tissue-dependent accumulation pattern of β -phellandrene **10**. The FPKM values and TPS assay results also suggest that neither *AaTPS1* nor *AaTPS11* may be the predominant β -phellandrene synthase in seeds. Thus, another unknown TPS that is responsible for the accumulation of a large amount of β -phellandrene **10** remains to be identified in future studies.

α -Pinene and β -myrcene

The in vivo enzyme assay results revealed that AaTPS3 and AaTPS4 were capable of producing α -pinene **1** as their major product; thus, they were considered as the TPSs responsible for the accumulation of α -pinene **1** in this plant (Figure 6). The FPKM values of these two TPSs support their role in α -pinene accumulation in *A. archangelica* leaves and tap roots (Supplementary Figure S6). The in vivo enzyme assay results also showed that AaTPS4 was capable of producing β -myrcene **6** as its major product, suggesting that the enzyme is potentially involved in the accumulation of β -myrcene **6** in this plant. The FPKM values of *AaTPS4* support its role

in β -myrcene accumulation in the plant's leaves and tap roots (Supplementary Figure S6). AaTPS5, which exhibited sesquiTPS activity to produce β -bisabolene **24** as its major product, also produced monoterpenoids, including β -myrcene **6**, as its minor product in the *E. coli* assay system. Nevertheless, AaTPS5 is unlikely to contribute to monoterpene formation in *A. archangelica* because AaTPS5, which does not have an N-terminal extension (Supplementary Figure S7), may not be localized in plastids, where GPP/NPP synthases are generally found. The GPP acceptability of AaTPS5 strengthens the idea that AaTPS5 might be derived from a monoTPS ancestor, as speculated from the phylogenetic analyses (Figure 2).

Germacrene D and α -humulene

AaTPS8 was capable of producing germacrene D and α -humulene, which are sesquiterpenoids, as its major products. This suggests that it might be at least partly responsible for the observed coexistence of germacrene D **22** and α -humulene **20** in this plant. The FPKM values of AaTPS8 support its role in the accumulation of these two sesquiterpenoids in the plant's leaves and tap roots (Supplementary Figure S6). TPSs that catalyze formation of germacrene D as a major product have been identified other Apiaceae, such as AkTPS1 and AkTPS2 from *Angelica keiskei* (Iimura et al. 2020) and DcTPS07, DcTPS11, and DcTPS42 from *Daucus carota* (Muchlinski et al. 2020). AkTPS1 and AkTPS2 show the highest amino acid sequence identity to AaTPS8 (75% and 71%, respectively) among TPSs from *A. archangelica*, suggesting that AaTPS8 might be an orthologue of AkTPS1 and AkTPS2.

In this study, we successfully elucidate the distributions of mono- and sesquiterpenoid volatiles in various tissues of *A. archangelica*. To the best of our knowledge, this is the first report of terpenoid composition analysis of non-stressed physiological plant samples, as previous studies have analyzed terpenoid compositions of *A. archangelica* using processed plant samples, such as essential oils (Lopes et al. 2004; Nivinskiene et al. 2007) and air-dried roots (Kerrola and Kallio 1994; Kerrola et al. 1994). It has been reported that many TPS genes are transcriptionally induced in response to various biotic and abiotic stresses, accompanying with dynamic changes in terpenoid compositions (Pichersky and Raguso 2018). Therefore, the terpenoid composition of the *A. archangelica* tissues, as well as enzymatic properties of AaTPSs, clarified in this study would be useful fundamental information for investigation of physiological roles of terpenoids in response to environmental stresses and identification of other stress-inducible TPS genes. Furthermore, comparable functional analyses of AaTPSs would provide important information for elucidating key residues that

regulate the substrate and product specificities of TPS.

Acknowledgements

We thank the Instrumental Analysis Group of the Technical Division of Tohoku University's School of Engineering for technical assistance with the GC-MS analyses. We are grateful to Dr. Norihiko Misawa (Ishikawa Prefectural University, Japan) for kindly providing the plasmid pAC-Mev/Scidi. This study was supported in part by a JSPS KAKENHI grant (B20H029090).

Author contribution

TS, TW, TN, MF, and ST planned and designed this study; MS-H, DM, YK, JT, NK, YA, TS, and ST annotated the transcriptome data and cloned cDNAs. MS-H, DM, and HA extracted and analyzed the terpenoids in *A. archangelica* plants. MS-H, DM, and HA constructed *E. coli* assay systems and analyzed the terpenoids. TN and ST wrote a draft; All authors reviewed the manuscript.

Conflict of interest

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might affect the objectivity of this review.

Accession numbers

The transcriptome data of this article are available from the DNA Data Bank of Japan at <https://www.ddbj.nig.ac.jp/index-e.html> and can be accessed with the accession numbers DRR378897–DRR378899. The other data underlying this article are available in the article and online supplementary files.

References

- Adio AM (2009) Germacrenes A-E and related compounds: Thermal, photochemical and acid induced transannular cyclizations. *Tetrahedron* 65: 1533–1552
- Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc Natl Acad Sci USA* 95: 4126–4133
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 66: 212–229
- Christianson DW (2017) Structural and chemical biology of terpenoid cyclases. *Chem Rev* 117: 11570–11648
- de Kraker JW, Franssen MCR, de Groot A, König WA, Bouwmeester HJ (1998) (+)-Germacrene A biosynthesis: The committed step in the biosynthesis of bitter sesquiterpene lactones in chicory. *Plant Physiol* 117: 1381–1392
- Falara V, Akhtar TA, Nguyen TT, Spyropoulou EA, Bleeker PM, Schavvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, et al. (2011) The tomato terpene synthase gene family. *Plant Physiol* 157: 770–789
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652
- Harada H, Yu F, Okamoto S, Kuzuyama T, Utsumi R, Misawa N (2009) Efficient synthesis of functional isoprenoids from acetoacetate through metabolic pathway-engineered *Escherichia*

- coli. Appl Microbiol Biotechnol* 81: 915–925
- Hare SR, Tantillo DJ (2016) Dynamic behavior of rearranging carbocations: Implications for terpene biosynthesis. *Beilstein J Org Chem* 12: 377–390
- Heskes AM, Sundram TCM, Boughton BA, Jensen NB, Hansen NL, Crocoll C, Cozzi F, Rasmussen S, Hamberger B, Hamberger B, et al. (2018) Biosynthesis of bioactive diterpenoids in the medicinal plant *Vitex agnus-castus*. *Plant J* 93: 943–958
- Holm Y, Galambosi B, Heikkinen J, Galambosi Z, Hiltunen R (2012) Essential oil composition of *Angelica archangelica* root from northern Finland. *Rev Clin Pharmacol Drug Therapy* 10: 58
- Iimura K, Hattan J, Misawa N, Shindo K (2020) cDNA cloning and functional analyses of ashitaba (*Angelica keiskei*) sesquiterpene synthase genes. *J Oleo Sci* 69: 711–718
- Karunanithi PS, Zerbe P (2019) Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front Plant Sci* 10: 1166
- Kerrola K, Galambosi B, Kallio H (1994) Characterization of volatile composition and odor of angelica (*Angelica archangelica* subsp. *archangelica* L.) root extracts. *J Agric Food Chem* 42: 1979–1988
- Kerrola KM, Kallio HP (1994) Extraction of volatile compounds of angelica (*Angelica archangelica* L.) root by liquid carbon dioxide. *J Agric Food Chem* 42: 2235–2245
- Krause ST, Liao P, Crocoll C, Boachon B, Forster C, Leidecker F, Wiese N, Zhao DY, Wood JC, Buell CR, et al. (2021) The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. *Proc Natl Acad Sci USA* 118: e2110092118
- Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323
- Lima AS, Schimmel J, Lukas B, Novak J, Barroso JG, Figueiredo AC, Pedro LG, Degenhardt J, Trindade H (2013) Genomic characterization, molecular cloning and expression analysis of two terpene synthases from *Thymus caespitius* (Lamiaceae). *Planta* 238: 191–204
- Lopes D, Strobl H, Kolodziejczyk P (2004) 14-Methylpentadecano-15-lactone (Muscolide): A new macrocyclic lactone from the oil of *Angelica archangelica* L. *Chem Biodivers* 1: 1880–1887
- Mann FM, Thomas JA, Peters RJ (2011) Rv0989c encodes a novel (*E*)-geranyl diphosphate synthase facilitating decaprenyl diphosphate biosynthesis in *Mycobacterium tuberculosis*. *FEBS Lett* 585: 549–554
- Meehan-Atrash J, Luo WT, McWhirter KJ, Dennis DG, Sarlah D, Jensen RP, Afreh I, Jiang J, Barsanti KC, Ortiz A, et al. (2021) The influence of terpenes on the release of volatile organic compounds and active ingredients to cannabis vaping aerosols. *RSC Advances* 11: 11714–11723
- Muchlinski A, Ibdah M, Ellison S, Yahyaa M, Nawade B, Laliberte S, Senalik D, Simon P, Whitehead SR, Tholl D (2020) Diversity and function of terpene synthases in the production of carrot aroma and flavor compounds. *Sci Rep* 10: 9989
- Nivinskiene O, Butkiene R, Mockute D (2007) Theseed (fruit) essential oils of *Angelica archangelica* L. growing wild in Lithuania. *J Essent Oil Res* 19: 477–481
- Ojala A (1986) Variation of *Angelica-archangelica* subsp. *archangelica* (Apiaceae) in Northern Fennoscandia. 4. Pattern of geographic-variation. *Ann Bot Fenn* 23: 23–31
- Palmiter RD (1974) Magnesium precipitation of ribonucleoprotein complexes: Expedient techniques for the isolation of undergraded polysomes and messenger ribonucleic acid. *Biochemistry* 13: 3606–3615
- Pichersky E, Raguso RA (2018) Why do plants produce so many terpenoid compounds? *New Phytol* 220: 692–702
- Schillmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, Pichersky E (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc Natl Acad Sci USA* 106: 10865–10870
- Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864
- Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: Why and how. *Ann Bot* 101: 5–18
- Sun H, Ni H, Yang YF, Wu L, Cai HN, Xiao AF, Chen F (2014) Investigation of sunlight-induced deterioration of aroma of pummelo (*Citrus maxima*) essential oil. *J Agric Food Chem* 62: 11818–11830
- Suzuki Y, Kawazu T, Koyama H (2004) RNA isolation from siliques, dry seeds, and other tissues of *Arabidopsis thaliana*. *Biotechniques* 37: 542–544
- Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. *Adv Biochem Eng Biotechnol* 148: 63–106
- Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of limonene synthase preprotein provides a fully active ‘pseudomature’ form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* 37: 12213–12220