

RESEARCH ARTICLE

Transcriptional and functional characterization of terpene synthase genes of the aromatic plant *Plectranthus hadiensis*

Hinano Mizuno^{*}, Kenro Tokuhito, Koichiro Iwai, Yoshikazu Furuta, Akari Nakasone, Hidenori Tanaka, Hiroki Sugimoto

Toyota Central R&D Labs., Inc., Nagakute, Aichi, Japan

* Hinano.Mizuno.bb@mosk.tytlabs.co.jp



Abstract

Plectranthus hadiensis (Lamiaceae) is recognized for its rich terpene content and potential applications in agriculture, medicine, and aromatherapy. Terpenes are major constituents of *P. hadiensis* essential oil, yet its terpene synthase (TPS) genes remain insufficiently characterized. In this study, we assembled a *de novo* transcriptome from RNA-seq data generated from leaf, stem, and root tissues and identified 26 TPS genes. Phylogenetic analysis classifies these genes into five TPS subfamilies (TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g), broadly associated with sesquiterpene, monoterpene, and diterpene biosynthesis. Expression profiling revealed apparent tissue specificity; notably, *PhTPS1* showed high transcript abundance in the leaf and stem. BLASTP analysis indicated that *PhTPS1* is closely related to Lamiaceae monoterpene synthases, with the top hit being a rosemary (*Salvia rosmarinus*) limonene synthase. Heterologous expression of *PhTPS1* in yeast, followed by headspace gas chromatography-mass spectrometry, detected limonene, confirming *PhTPS1* as a functional limonene synthase. To our knowledge, this is the first limonene synthase gene functionally characterized in *P. hadiensis*. Collectively, these findings provide a curated TPS genes catalog with their tissue-specific expression patterns and identify *PhTPS1* as a promising target for breeding and metabolic engineering to enhance limonene yield.

OPEN ACCESS

Citation: Mizuno H, Tokuhito K, Iwai K, Furuta Y, Nakasone A, Tanaka H, et al. (2026) Transcriptional and functional characterization of terpene synthase genes of the aromatic plant *Plectranthus hadiensis*. PLoS One 21(3): e0345772. <https://doi.org/10.1371/journal.pone.0345772>

Editor: S Ezhil Vandan, Central Food Technological Research Institute CSIR, INDIA

Received: April 14, 2025

Accepted: March 10, 2026

Published: March 23, 2026

Copyright: © 2026 Mizuno et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All transcriptome data generated in this study have been deposited in the DNA Data Bank of Japan (DDBJ) under BioProject accession number PRJDB19996. RNA-seq raw read data are available in the DDBJ Sequence Read Archive

Introduction

Plants synthesize diverse terpenes that contribute to various ecological functions—such as attracting pollinating insects, repelling predators, and communicating with other plants [1]. Terpenes are hydrocarbons composed of isoprene (C5) units [2] and include monoterpenes (C10), sesquiterpenes (C15), and diterpenes (C20), many of which are major constituents of essential oils [1,3,4]. Representative monoterpenes such as limonene, pinene, and camphene contribute to aroma and bioactivity [5],

(DRA) under accession numbers DRR635850–DRR635857. The assembled transcriptome sequences (TSA) have been deposited under DDBJ accession numbers ICX101000001–ICX101060725. TSA accession numbers and CDS coordinates for the 26 TPS genes are provided in Table S11. The coding sequences and corresponding amino acid sequences of the 26 TPS genes are provided as [Supporting Information](#) (S1 and S2 Appendices, respectively).

Funding: This work was supported by Toyota Central R&D Labs., Inc. (Japan). The funder provided support in the form of salaries for authors HM, KT, KI, YF, AN, HT, and HS, and routine research support (materials and laboratory infrastructure), but had no additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

whereas sesquiterpenes and diterpenes include diverse metabolites involved in plant defense and development [6–8].

Monoterpenes, sesquiterpenes, and diterpenes are generally synthesized from geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), respectively (Fig 1) [9]. Terpene synthases (TPS) catalyze the conversion of these prenyl diphosphates into structurally diverse terpene scaffolds. Plant TPS typically contain two conserved Pfam domains (PF01397 and PF03936) [10,11] and are classified into multiple subfamilies (TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f, TPS-g, and TPS-h) that broadly correspond to substrate and product preferences [12,13]. For example, TPS-a is enriched in sesquiterpene synthases, TPS-b contains many monoterpene synthases, and TPS-c and TPS-e/f include diterpene cyclases and synthases, respectively [13]. However, these relationships are not absolute, as TPS enzymes can show substantial functional diversification and multiproduct catalysis [13–17].

Members of Lamiaceae (>250 genera and >7,000 species) synthesize a wide spectrum of terpenes via numerous TPS [18–23]. Research on Lamiaceae TPS has largely used targeted gene- and pathway-level approaches. For example, in *Thymus caespitosus*, cloning and expression analyses identified two monoterpene synthases, a γ -terpinene synthase and an α -terpineol synthase, and defined their exon–intron structures [19, 23]. In *Mentha longifolia*, a genome-wide survey revealed 63 TPS genes across six subfamilies and functional assays verified that at least one TPS catalyzes the conversion of GPP to limonene [20]. In *Origanum vulgare*, seven TPS were isolated; the enzyme expression levels correlated with essential-oil composition, and thymol biosynthesis was found to proceed via γ -terpinene produced by a single monoterpene synthase [22]. Nevertheless, most studies still focus on subsets of enzymes or lineages rather than delivering species-level, tissue-resolved TPS inventories with functional validation [19–23]. *Plectranthus*, a major Lamiaceae genus with >300 species, is a promising source of essential oils and terpenes [24]. Despite extensive ethnobotanical use and demonstrated bioactivity [25], the high intra-generic diversity of *Plectranthus* complicates taxonomic resolution and standardization for product development [26]. This is especially important for medicinal applications, where complex plant extracts require careful evaluation of potential side effects [27]. Moreover, functional TPS studies are limited to *Plectranthus*. In *P. amboinicus*, enzymes related to linalool and nerolidol synthases have been functionally identified, showing only ~60%–70% amino acid identity to TPS reported from other Lamiaceae genera [21]. In a recent study on *P. barbatus*, genome assembly enabled comparative analysis of diterpene-related TPS candidates across Lamiaceae members; however, the study did not focus on comprehensive functional characterization of *Plectranthus* TPS [28]. Collectively, these gaps underscore the need for a systematic, tissue-resolved catalog of the TPS repertoire across the genus *Plectranthus*—extending beyond the two species studied to date—together with functional validation. *Plectranthus hadiensis* essential oils reportedly exhibit repellent, antibacterial, antioxidant, pro-apoptotic, membrane-stabilizing/anti-platelet, and anti-inflammatory activities [29–33]. Despite its potential utility, *P. hadiensis* remains relatively understudied

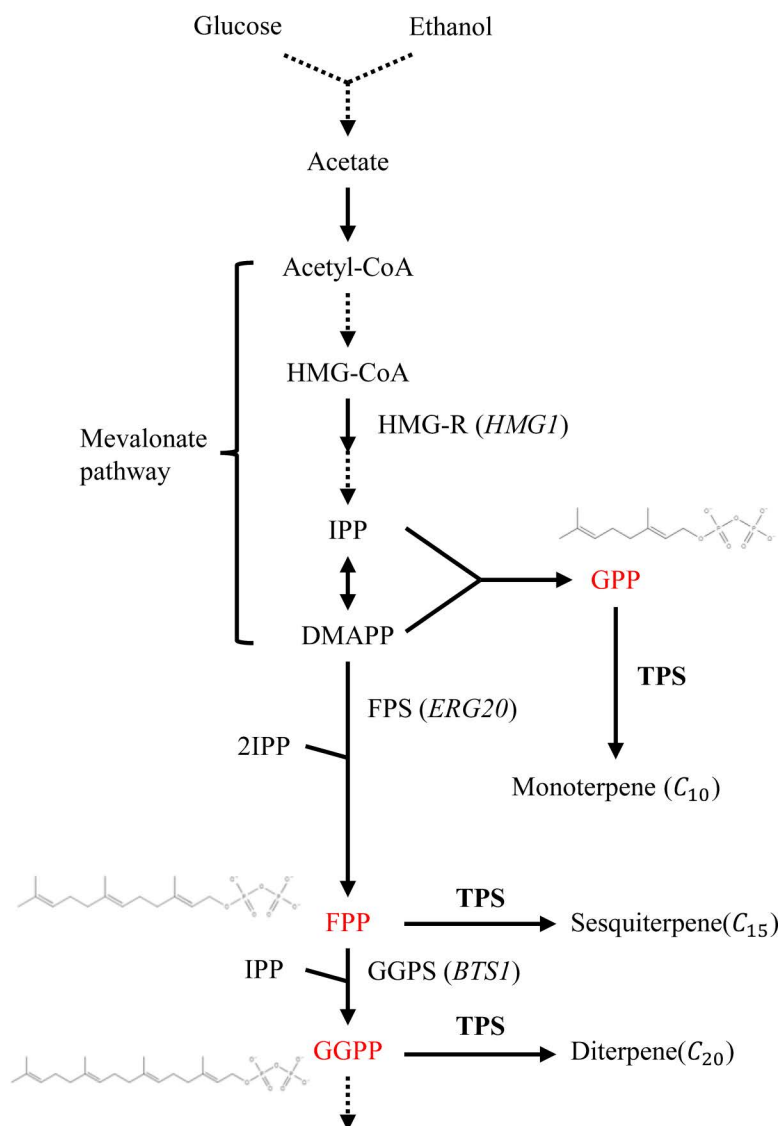


Fig 1. Overview of the terpene biosynthesis pathway (mevalonate pathway). IPP: isopentenyl pyrophosphate (C₅), DMAPP: dimethylallyl pyrophosphate (C₅), HMG-R: HMG-CoA reductase, FPS: FPP synthase, GGPS: GGPP synthase. *HMG1*, *ERG20*, and *BTS1* represent the yeast (*Saccharomyces cerevisiae*) genes encoding the respective synthases.

<https://doi.org/10.1371/journal.pone.0345772.g001>

owing to difficulties in distinguishing it from other congeneric species [26] and is often substituted by *P. barbatus* or *P. amboinicus* [32,34]. Although some genetic studies on *P. hadiensis* have recently been conducted [26,35], comprehensive genetic information, including TPS gene family data, on this species remains limited.

To address the lack of a tissue-resolved TPS catalog and functional evidence in *P. hadiensis*, we set two objectives: (i) to assemble a *de novo*, tissue-resolved transcriptome and delineate the TPS complement by subfamily assignment and expression profiling across the leaf, stem, and root; and (ii) to test whether the most highly expressed TPS candidate encodes a limonene synthase using heterologous yeast expression and headspace gas chromatography-mass spectrometry (GC-MS). An overview of the experimental workflow is shown in S1 Fig. This design enabled, to our knowledge, the first functional validation of a limonene synthase (*PhTPS1*) from *P. hadiensis* and provides a curated TPS resource for this species.

Materials and methods

Sampling and RNA extraction

Plectranthus hadiensis was cultivated for 3 months using a hydroponic system that incorporated hydro balls and was maintained at room temperature (18–27 °C). The plants were grown under long-day conditions (16 h light/8 h dark, approximately 10–15 $\mu\text{mol m}^{-2} \text{s}^{-1}$) to mirror indoor cultivation conditions to achieve uniform growth and obtain high-quality RNA. The characteristic aroma was retained. The leaf (with the leaf blades more than 1 cm in length), stem, and root were sampled (S2 Fig). Total RNA was extracted from each sample using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Sequencing and *de novo* assembly

Complementary DNA (cDNA) library construction and sequencing were outsourced to Nippon Genetics (Tokyo, Japan), and 151 bp paired-end reads were obtained using the NovaSeq6000 system (Illumina, San Diego, CA, USA). Low-quality reads were filtered and adapter sequences were trimmed using fastp v0.23.2 [36] and PRINSEQ v0.20.4 [37]. Quality assessment using fastp indicated that, on average, 93% of bases across all samples had Phred quality scores $\geq$ Q30. Next, randomly selected paired-end reads (1 M) from individual leaf, stem, and root samples were combined and subjected to *de novo* assembly using Trinity v2.15.0 [38,39] to generate the reference cDNA sequences of *P. hadiensis*. Gene prediction was conducted using TransDecoder v5.7.0 [40]. Predicted coding sequences were subjected to functional annotation using InterProScan v5.60-92.0 [41]. Additionally, transcripts per million (TPM) values were calculated using Salmon [42] within the Trinity pipeline, which employs a quasi-mapping approach and normalizes for both transcript length and sequencing depth. We calculated standard assembly statistics (number of contigs, total bases, mean and median contig length, and N50) and assessed completeness with BUSCO v5 [43,44].

To verify the species identity of the plant material, we queried the assembled proteome with the *P. hadiensis* RbcL protein (YP_010593282.1) using BLASTP v2.9.0+ [45,46], retrieved the best RbcL-like hit, and confirmed identity via pairwise alignment. This barcode check was used solely for sample authentication and not for TPS gene identification.

Quantification of gene expression and identification of differentially expressed genes

Reads from a total of eight samples—three leaf samples, two stem samples, and three root samples—were mapped to the reference sequences using HISAT2 v2.2.1 [47]. Gene expression levels for each sample were quantified using StringTie v1.3.4 [48], and TCC-GUI v2021.11.13 [49] was employed to identify differentially expressed genes (DEGs). Genes with absolute M value (the difference in $\log_2$ -based expression between tissues) equal to or higher than 2.5 and with A value (the average of the $\log_2$ -based expression between tissues) higher than 0 were selected as DEGs. Gene Ontology (GO) term enrichment analysis was performed using topGO v2.50.0 [50].

Identification and classification of TPS genes

Genes encoding amino acid sequences that contain both the N-terminal domain (InterPro; IPR001906, Pfam; PF01397) and metal-binding domain (InterPro; IPR005630, Pfam; PF03936) were extracted as TPS genes. A heatmap comparing expression levels among tissues was generated through R v4.2.2 [51] using the pheatmap v1.0.12 [52] package.

Phylogenetic analysis was conducted using amino acid sequences of TPS genes from *P. hadiensis* together with reference TPS sequences from *Arabidopsis thaliana*, representative Lamiaceae species, and additional plant lineages, as needed to cover all major TPS subfamilies defined in previous studies [12,53] (listed in S1 Table). Sequences were aligned with MAFFT v7.453 [54], trimmed using trimAl v1.5.rev1 [55], and a maximum-likelihood tree was inferred with IQ-TREE v3.0.1 [56] with the best-fit substitution model selected by ModelFinder [57]. Branch support was evaluated using the ultrafast bootstrap approximation (UFBoot2) [58] with 1,000 replicates. TPS subfamily assignments

for *P. hadiensis* sequences were determined based on their nearest phylogenetic placement relative to reference TPS sequences in the resulting tree. The final tree was visualized and annotated using Interactive Tree of Life (iTOL) v7 [59] with UFBoot support values.

Molecular cloning of the *P. hadiensis* TPS gene (*PhTPS1*)

One TPS gene that exhibited significantly high expression in the leaf and stem was selected for heterologous expression in yeast to assess its activity. This gene is henceforth referred to as *PhTPS1*. Considering the possibility of allelic variants, RNA sequencing (RNA-Seq) reads were remapped to *PhTPS1* and visualized using IGV v2.3.8 [60] and Jalview v2 [61], thereby allowing for manual inspection of the aligned reads.

A yeast codon-optimized version of the *PhTPS1* sequence, determined after remapping, was artificially synthesized. Similarly, a lemon (*Citrus limon*)-extracted TPS (GenBank: AAM53944.1) involved in limonene synthesis, which is active when expressed in yeast [62], was also synthesized (hereinafter referred to as *CITPS*) as a positive control. Gene synthesis was performed through the GeneArt Gene Synthesis service of Thermo Fisher Scientific (Waltham, MA, USA).

The synthesized TPS sequences were polymerase chain reaction (PCR)-amplified (*PhTPS1*: S2 Table primers 1 and 2, *CITPS*: S2 Table primers 3 and 4) and introduced into a pRS436GAP plasmid [63] (S3 Table No.5) linearized with Sall (Takara Bio, Kusatsu, Japan), using the In-Fusion Dry-Down PCR Cloning Kit w/ Cloning Enhancer (Takara Bio). The plasmids containing *PhTPS1* and *CITPS* were designated pRS436GAP-PhTPS1 (S3 Table No.6) and pRS436GAP-CITPS (S3 Table No.7), respectively.

Sequence confirmation of expression plasmids was performed as follows. For pRS436GAP-PhTPS1 and pRS436GAP-CITPS, the complete coding regions and vector–insert junctions were verified using Sanger sequencing. Sequencing was performed by Eurofins Genomics K.K. (Tokyo, Japan). The verified plasmids were used in all subsequent yeast experiments.

Construction of plant TPS-expressing yeast strains

We used pRS504HMG1/YPH499 [64] (S3 Table No.9) as the host strain. This strain overexpresses *HMG1* encoding the rate-limiting enzyme hydroxymethylglutaryl-CoA reductase in the mevalonate pathway. It was transformed with various plasmids to enhance the production of different terpene precursors—GPP, FPP, and GGPP—consequently generating substrate-producing yeast lines. Specifically, the FPP-producing strain was generated by introducing pRS435GAP-ERG20 [65] (S3 Table No.3), which expresses ERG20 (farnesyl diphosphate synthase)—an enzyme possessing both FPP synthase and GPP synthase activities—under the control of the TDH3 promoter. It has been reported that an A99P mutation in ERG20 reduces FPP synthase activity while retaining GPP synthase activity, leading to GPP accumulation in yeast [64]. Accordingly, the GPP-producing strain was constructed by introducing pRS435GAP-ERG20(A99P) (S3 Table No.4). The GGPP-producing strain was generated by introducing pRS435GAP-GGF [65] (S3 Table No.2) encoding a fusion of Bts1 (geranylgeranyl diphosphate synthase) and Erg20, which enables GGPP production from FPP and IPP. Finally, pRS436GAP-PhTPS1 (S3 Table No.6) or pRS436GAP-CITPS (S3 Table No.7) was transformed into each of the GPP-, FPP-, and GGPP-producing strains using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA) to introduce TPS. The same substrate-producing strains were transformed with a pRS436GAP plasmid lacking TPS genes as a negative control (S3 Table No.5). Yeast strains expressing the pathway and carrying TPS plasmids were selected and maintained on synthetic dropout medium (SD-TRP-LEU-URA) according to plasmid markers and used in GC–MS analyses.

GC–MS analysis

All GC–MS experiments were performed using three independent biological replicates for each yeast strain. The terpene precursor (GPP, FPP, and GGPP) producing yeast carrying each TPS-expression vector and negative control strains—GPP-producing yeast (S3 Table No.19–21), FPP-producing yeast (S3 Table No.16–18), and GGPP-producing yeast (S3

[Table No. 13–15](#)—were pre-cultured in 2 mL of selective medium (SD-TRP-LEU-URA) for 24 h. Yeast pellets were then collected through centrifugation at $500 \times g$ for 1 min at 23 °C. Next, 2 mL of the pellets was transferred into 2 mL of fresh selective medium (SD-TRP-LEU-URA) and cultured for 72 h in a 20-mL screw-neck vial (GERSTEL). The headspace components were subsequently analyzed using GC–MS.

The GC–MS system (8890/5977C; Agilent Technologies, Santa Clara, CA, USA) was equipped with MPS roboticPRO (GERSTEL, Mülheim, Germany) for sample preparation under the following conditions. The vial containing the sample was heated at 60 °C for 15 min to equilibrate the headspace, and volatile compounds from the headspace were extracted for 30 min using a solid-phase microextraction fiber (DVB/CAR/PDMS, 23 gauge, fiber length: 2 cm; Merck, Darmstadt, Germany) before GC–MS injection. Separation was achieved on an HP-5MS UI column (length, 30 m; i.d., 0.25 mm; film thickness, 0.25 μ m) under a constant helium carrier gas flow of 1.2 mL/min. The GC oven temperature was programmed as follows: an initial temperature of 40 °C was maintained for 5 min, increased to 280 °C at 10 °C/min, and maintained at 280 °C for 11 min. Detection was performed using a quadrupole mass selective detector operated at 70 eV in the combined SIM/SCAN mode. Specific target ions (m/z 68, 93, and 136) were monitored in the SIM mode, whereas ions in the range of m/z 30–250 were analyzed in the SCAN mode.

A mixture of C7–C33 n-alkanes (Hayashi Pure Chemical Ind., Ltd., Osaka, Japan) was analyzed under the same GC–MS conditions as those described above to calculate the Kovats retention index (RI). The RI values, calculated from the retention times of the analytes and n-alkane standards, were used alongside mass spectra to identify limonene and other volatiles. Mass spectral data were compared against the NIST/EPA/NIH Mass Spectral Library (NIST 20) for compound identification. Peak areas were compared across samples acquired under identical culture volumes, incubation times, and SPME extraction conditions; therefore, the values represent relative, semi-quantitative estimates rather than absolute concentrations.

Results

Tissue-specific DEG analysis of *P. hadiensis*

To rule out misidentification within the *Plectranthus* genus, we verified the species identity of the plant material using an RbcL barcode. The assembled RbcL protein was identical to the *P. hadiensis* reference (YP_010593282.1) ([S3 Fig](#)) and this step was used only for sample authentication. A sequencing analysis was conducted for RNA isolated from the leaf, stem, and root of *P. hadiensis*. The average read count was approximately 20,000,000 reads per sample ([S4 Table](#)). *De novo* assembly of the reads resulted in a reference cDNA sequence containing 43,489 genes, with an average contig length of 1,016 bases ([S5 Table](#)). BUSCO v5 indicated C = [70.1%] (S: [47.0%], D: [23.1%]), F = [8.3%], M = [21.6%] for the predicted proteins, supporting overall completeness of the reference ([S5 Table](#)). Application of our threshold ($|M \text{ value}| \geq 2.5$ and $A \text{ value} > 0$) revealed 3,759 DEGs in leaf > root, 1,801 DEGs in root > leaf, 1,281 DEGs in leaf > stem, 1,265 DEGs in stem > leaf, 1,986 DEGs in root > stem, and 3,716 DEGs in stem > root. Furthermore, the GO term enrichment analysis revealed that terms associated with photosynthesis were significantly enriched in the leaf compared with those in the root. However, no other notable tendencies were observed ([S4 Fig](#)).

Expression profiles of TPS gene families

We explored coding sequences with two TPS-specific domains (TPS N-terminal and metal-binding domains) in the cDNA sequences. Our analysis revealed 26 TPS genes expressed in *P. hadiensis*. We subsequently classified these identified TPS genes into specific TPS subfamilies, based on phylogenetic placement relative to reference TPS sequences. The molecular phylogenetic tree analysis revealed that *P. hadiensis* possesses TPS genes belonging to five TPS subfamilies ([Fig 2](#)). Specifically, 5/26 (19%) TPS genes were assigned to TPS-a, 6/26 (23%) to TPS-b, 6/26 (23%) to TPS-c, 6/26 (23%) to TPS-e/f, and 3/26 (12%) to TPS-g. No TPS genes belonging to the TPS-d or TPS-h subfamilies were detected.

Tree scale: 1

Colored ranges

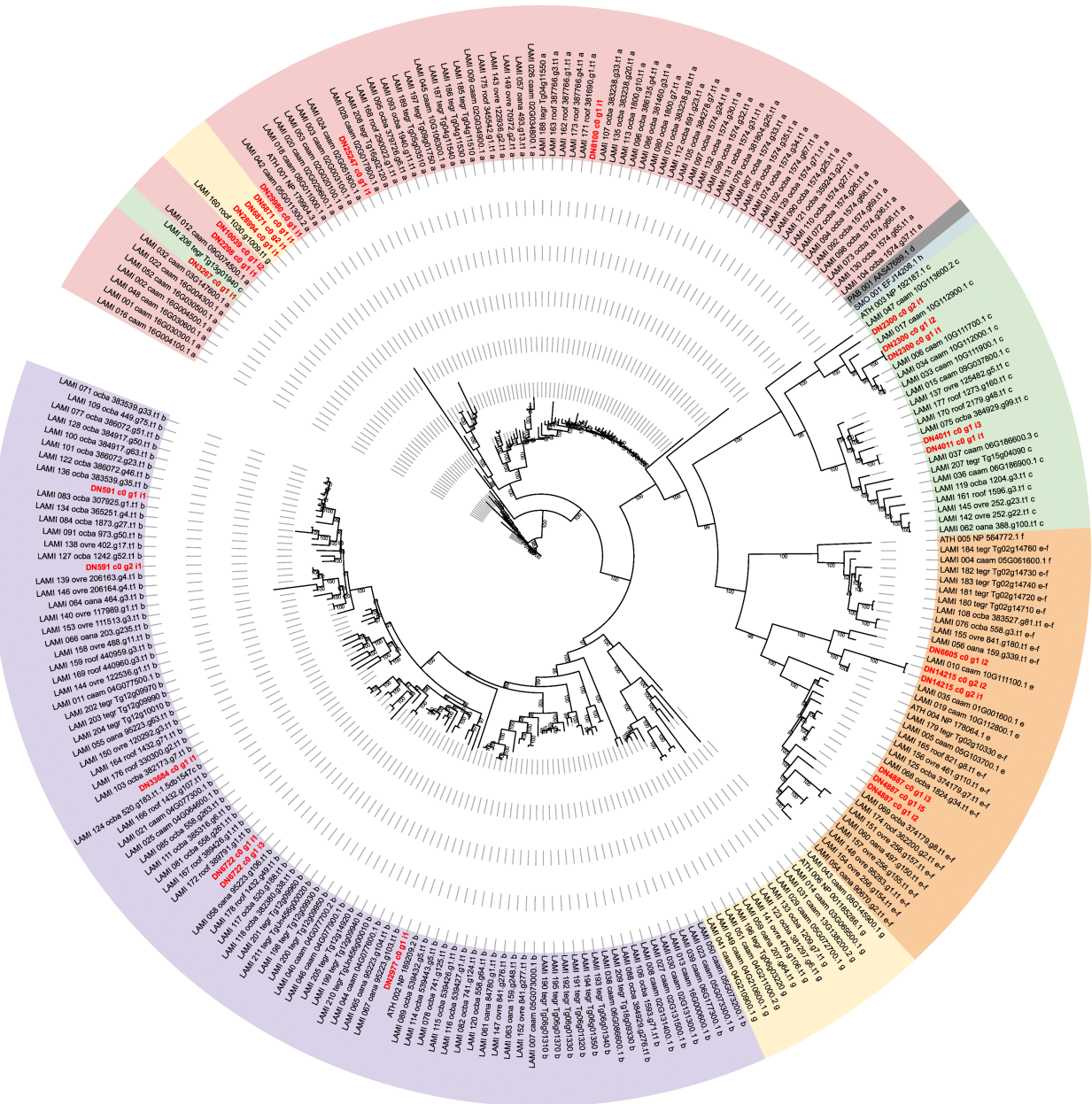
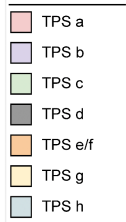


Fig 2. Phylogenetic tree of the 26 TPS genes identified in this study. This maximum-likelihood tree includes amino acid sequences of *P. hadiensis* TPS genes together with reference TPS sequences from *A. thaliana* and representative Lamiaceae species. Branch lengths represent evolutionary distance, and node labels indicate ultrafast bootstrap support values (UFBoot2, 1,000 replicates). TPS subfamilies are indicated by color coding, and *P. hadiensis* TPS genes are shown in red. Subfamily assignments are based on nearest phylogenetic placement relative to reference TPS sequences (see S6 Table).

<https://doi.org/10.1371/journal.pone.0345772.g002>

For a subset of TPS genes, the subfamily assignment was marked as ambiguous due to closely related second-best phylogenetic placements among neighboring TPS clades (S6 Table).

We evaluated the expression distribution of the 26 TPS genes across different tissues. The numbers of DEGs were as follows: 10 in leaf > root, 6 in root > leaf, 1 in leaf > stem, 11 in stem > leaf, 5 in root > stem, and 6 in stem > root. As

shown in Fig 3A, the expression patterns of TPS genes varied depending on tissue type. The normalized heatmap (Fig 3A) further categorizes the TPS genes in *P. hadiensis* into those primarily highly expressed in the leaf and stem (upper 6 genes), in the stem (middle 11 genes), and in the root (lower 9 genes). The TPS genes highly expressed in the root did not include any members of family b, whereas those highly expressed in the stem predominantly belonged to families a, b, and g, as shown by the correspondence between TPS families (left column) and tissue-specific expression patterns in Fig 3A. S7 Table concisely summarizes, for each TPS gene, the putative subfamily, top BLASTP annotation, and tissue-specific expression category; actual expression values (TPM, by tissue/replicate) are provided in S8 Table.

Of the 26 TPS genes identified, 1 (gene ID: DN2977_c0_g1_i1) exhibited a markedly high expression in the leaf and stem (Fig 3B, indicated by an arrow). This gene was designated as *PhTPS1*. In a BLASTP search for the amino acid sequence homology of *PhTPS1*, the top hit was a limonene synthase gene from rosemary (*Salvia rosmarinus*, Accession no. ABD77416.1), a member of the Lamiaceae family (S9 Table). Subsequent top hits predominantly matched monoterpene synthase genes from Lamiaceae species, suggesting a conserved role of the gene in monoterpene biosynthesis within this family. Considering the possibility of allelic variants, we remapped the RNA-Seq reads to *PhTPS1*. Two alleles, *PhTPS1 allele1* and *PhTPS1 allele2*, differed by the

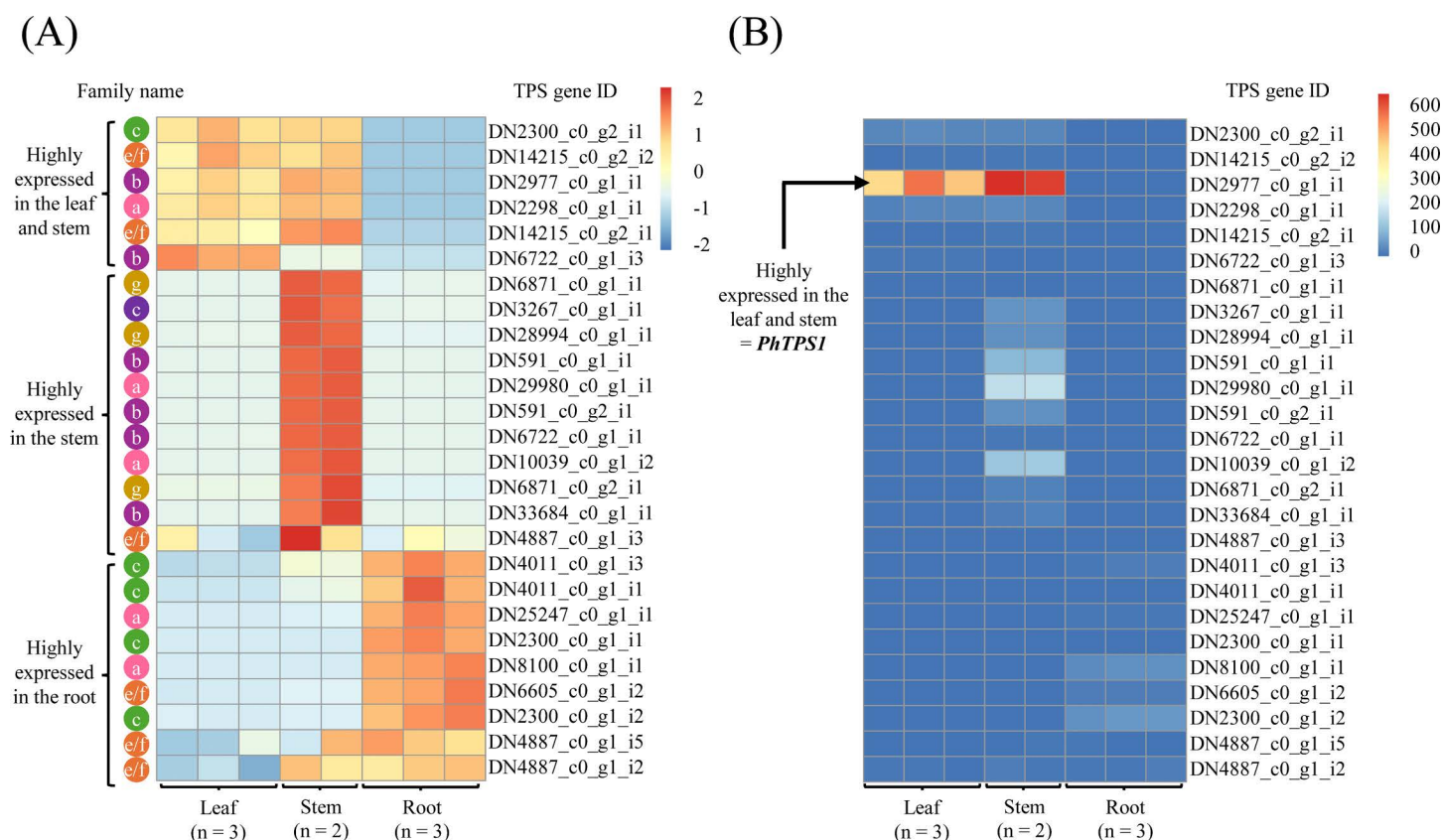


Fig 3. Tissue-specific expression profiles of TPS genes in *P. hadiensis*. (A) Z-score normalized tissue-specific expression profiles of TPS genes in *P. hadiensis*. The heatmap shows the 26 TPS genes with expression levels normalized to a mean of 0 and a variance of 1 (z-score normalization). The putative subfamily for each TPS gene is displayed on the left side of the heatmap. (B) A heatmap showing the expression levels of the 26 TPS genes in the leaf, stem, and root, presented as TPM. The total TPM per sample sums to one million, allowing for comparisons across different samples. (A and B) The color scale (red: higher, blue: lower) indicates relative expression. *PhTPS1*, indicated by an arrow, was highly expressed in the leaf and stem.

<https://doi.org/10.1371/journal.pone.0345772.g003>

presence or absence of a two-amino-acid insertion near the N-terminus (S5 and S6 Figs). Both alleles were expressed at approximately equal proportions in the leaf, stem, and root (S10 Table and S7 Fig). We subsequently cloned and introduced the shorter allele (S5 and S6 Figs, *PhTPS1 allele2*) into yeast to investigate its TPS activity and substrate specificity.

Characterization of the TPS activity of *PhTPS1*

We utilized the GPP-, FPP-, and GGPP-producing yeast lines because it was initially unclear whether *PhTPS1* encodes a mono-, sesqui-, or diterpene synthase (S8 Fig). Overexpression of *HMG1* in the pRS504HMG1/YPH499 background has been reported to activate the mevalonate pathway, thereby increasing the pool of the immediate isoprenoid precursors IPP and DMAPP [64]. Introducing *ERG20*, *ERG20(A99P)*, or *GGF* into this background allowed us to direct the production of GPP, FPP, or GGPP, respectively, in order to evaluate substrate specificity. Notably, the A99P mutation in *ERG20* has been reported to shift the enzyme activity away from FPP synthase while retaining GPP synthase function, thereby enabling GPP accumulation from IPP and DMAPP [65]. Plasmid inserts for *PhTPS1* and *CITPS* were sequence-confirmed using Sanger sequencing prior to transformation (Section 2.5). Yeast transformants were maintained under SD-TRP-LEU-URA selection, and limonene was detected in the *PhTPS1*- or *CITPS*-expressing strains but not in empty-vector controls (Fig 4), supporting functional expression.

We then expressed *PhTPS1* or *CITPS* in each of the GPP-, FPP-, or GGPP-producing strains and analyzed the resulting volatile compounds using GC-MS. A prominent peak corresponding to standard limonene was observed in GPP-producing yeast expressing either *PhTPS1* or *CITPS* (S9 Fig). Mass spectral analysis of this peak closely matched the reference library spectrum of limonene (S10 Fig). Fig 4 shows the GC-MS peak areas of limonene across the three substrate-producing strains. Limonene production was apparent in all three strains in the positive-control yeast expressing *CITPS*, which displayed substantial peak areas. Although the GPP-producing strain expressing *PhTPS1* yielded a smaller limonene peak than that expressing *CITPS*, it produced considerably more limonene than the negative control. These findings confirm that *PhTPS1* encodes a functional limonene synthase.

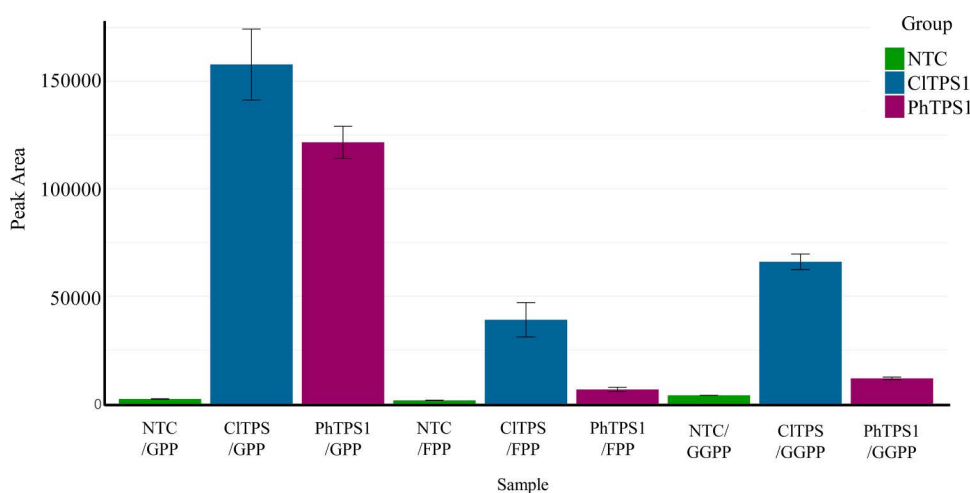


Fig 4. Peak areas of limonene detected in each yeast strain. Error bars indicate the standard deviation across three independent biological replicates.

<https://doi.org/10.1371/journal.pone.0345772.g004>

Discussion

In this study, we performed RNA-Seq across multiple tissues of *P. hadiensis* and identified 26 TPS genes using *de novo* transcriptome assembly. Expression profiling revealed marked tissue specificity among TPS genes, and one gene (*PhTPS1*, gene ID: DN2977_c0_g1_i1) exhibited exceptionally high expression in the leaf and stem. Such highly skewed TPS expression profiles have rarely been reported in transcriptome-wide surveys. A subsequent BLAST analysis revealed that *PhTPS1* has high similarity to limonene synthase genes from other Lamiaceae species, suggesting its potential role in limonene biosynthesis. To validate this speculation, we consequently synthesized *PhTPS1*, cloned and expressed it in yeast, and analyzed the volatile compounds in the culture medium using GC–MS. Limonene production in *PhTPS1*-expressing yeast strains supported its role in limonene biosynthesis.

The identified TPS genes in *P. hadiensis* showed pronounced tissue specificity, indicating that TPS expression in *P. hadiensis* is not restricted to leaf tissue. This pattern is consistent with observations in other species, such as *Rosa chinensis* [66] and *Panicum virgatum* L. [67], where certain TPS genes are also predominantly expressed in the root. Such tissue-specific expression has been interpreted as evidence of functional diversification among TPS genes, with root-expressed genes contributing to belowground chemical defense, interactions with soil microorganisms, or stress induced terpene production, and stem-enriched genes being implicated in constitutive or inducible defense against herbivores and pathogens or in the regulation of volatile emissions along the plant axis [28,68,69]. Collectively, these findings suggest that TPS expression patterns in *P. hadiensis* are likely to reflect diversification of terpene functions across organs. However, *P. hadiensis* is distinguished by the exceptionally high expression of a single gene, *PhTPS1*, in both leaves and stems, implying that *PhTPS1* may play a dominant role in the biosynthesis of a major terpene compound.

Consistent with this hypothesis, limonene was detected among volatile compounds extracted from the culture medium of *PhTPS1*-expressing yeast. Limonene levels were significantly higher in yeast strains engineered to produce GPP than in those producing FPP or GGPP. This difference is likely due to trace GPP synthesis in FPP-producing yeast or GGPP-producing yeast. However, the measurement method employed in this study provides a semi-quantitative estimate based on relative GC–MS peak areas without internal standards or absolute calibration curves. Nevertheless, this approach is sufficient for comparative analyses under identical experimental conditions. Accordingly, the pronounced enrichment of limonene in GPP-producing yeast is consistent with the biochemical role of GPP as the direct substrate for monoterpene biosynthesis, including limonene (Fig 1). In a previous study [29], limonene comprised a significantly high proportion (34.69%) of *P. hadiensis* extracts, and this is consistent with the high expression of *PhTPS1* in *P. hadiensis* observed in our study. Although TPS gene families in plants, particularly in Lamiaceae, often exhibit redundancy and sub-functionalization [13,69], individual TPS genes can still act as major determinants of essential oil composition [69,70]. Taken together, the exceptionally high expression and functional validation suggest *PhTPS1* as a key contributor to limonene biosynthesis in *P. hadiensis*.

We identified five TPS subfamilies in *P. hadiensis* (TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g), whereas TPS-d and TPS-h, largely found in gymnosperms and lycophytes [13], respectively, were not detected. The relative proportions of TPS subfamilies in *P. hadiensis* were 19% (TPS-a), 23% (TPS-b), 23% (TPS-c), 23% (TPS-e/f), and 12% (TPS-g). This composition broadly matches previous surveys of Lamiaceae TPS repertoires, in which TPS-a, -b, -c, and -e/f constitute the major components [53]. However, the proportion of TPS-e/f genes in *P. hadiensis* (23%) is higher than the average reported for Lamiaceae species (approximately 11%) [53]. A similarly elevated representation of TPS-e/f genes has been reported in *Mentha longifolia*, in which approximately 27% of TPS genes belong to this subfamily [20]. These observations suggest that expansion of TPS-e/f genes may have occurred independently in multiple Lamiaceae lineages, including *P. hadiensis*.

Although 26 TPS genes were identified in this study, additional TPS genes may have remained undetected. For instance, *Arabidopsis thaliana* encodes 32 TPS genes [71], whereas *Oryza sativa*, *Sorghum bicolor*, and *Selaginella moellendorffii* harbor 34, 24, and 14 full-length TPS genes, respectively [13]. The number of identified TPS genes can vary according to

detection methods. For example, Yan et al. [66] identified 40, 39, and 13 TPS genes in *O. sativa*, *S. bicolor*, and *S. moellendorffii*, respectively. In Lamiaceae, more comprehensive genome-based studies have reported larger repertoires—for example, 52 TPS genes were identified in *Callicarpa americana* using long-read genome assembly [72]. Similarly, a recent genome-wide analysis of four economically important culinary herbs—sweet basil (*Ocimum basilicum* L.), sweet marjoram (*Origanum majorana* L.), oregano (*Origanum vulgare* L.), and rosemary (*Rosmarinus officinalis* L.)—identified 235 TPS genes in total, ranging from 27 in *O. majorana* to 137 in the tetraploid *O. basilicum* [73]. These studies suggest that *P. hadiensis* may harbor a larger TPS repertoire than indicated by our data. This limitation likely reflects the design of our reference resource: a *de novo* transcriptome assembled from three tissues grown under indoor conditions. Consequently, the BUSCO score reflects expressed gene space rather than the full genomic repertoire. In addition, transcriptome assembly was performed using subsampled RNA-seq data (1 million reads per sample) to reduce computational requirements, which may have further reduced recovery of lowly expressed transcripts. Accordingly, BUSCO completeness of our assembly (70.1%) was lower than that reported for some other non-model Lamiaceae transcriptomes (e.g., *S. miltiorrhiza* drought transcriptome, C: 88.2% [74]; *S. hispanica* tissue transcriptomes, C: 92.2%–92.8% [75]). Despite these constraints, our transcriptome assembly was sufficient to identify 26 TPS candidates and to enable functional validation of *PhTPS1*. Future genome-wide assembly could reveal additional TPS genes and provide a stronger foundation for functional and evolutionary analyses in *P. hadiensis*. In addition, integration of this transcriptomic resource with future genome assemblies, comprehensive volatilome profiling across chemotypes, and targeted in planta functional assays (e.g., CRISPR/Cas-based approaches) will further expand its utility beyond the characterization of limonene biosynthesis.

Conclusion

In this study, we confirmed that *PhTPS1* encodes an active enzyme involved in limonene biosynthesis and is functional when expressed in yeast. The identification of an active TPS gene in *P. hadiensis* is particularly important because not all plant TPS genes are functional [13]. Thus, this study not only contributes to breeding efforts targeting *PhTPS1* for enhanced limonene production but also aids in the evolutionary understanding of TPS genes within the Lamiaceae family.

Supporting information

S1 Table. TPS genes used for classification.
(XLSX)

S2 Table. Primer list.
(XLSX)

S3 Table. Plasmids, vectors, and yeast strains used in this study. All the plasmids and vectors contained the *TDH3* promoter and *CYC1* terminator from *S. cerevisiae* for gene expression.
(XLSX)

S4 Table. Sequencing results for each sample.
(XLSX)

S5 Table. Results of *de novo* assembly under various conditions.
(XLSX)

S6 Table. TPS genes used for phylogenetic classification and subfamily assignment in *P. hadiensis*.
(XLSX)

S7 Table. Summary of TPS genes with phylogeny-based subfamily assignment and tissue-resolved expression (H/M/L/ND; High/Medium/Low/Not detected). Expression categories were assigned per tissue using quartiles

of the mean across replicates: $ND = 0$; $L \leq Q25$; $Q25 < M \leq Q75$; $H > Q75$. Thresholds: leaf $Q25 = 0.15$, $Q75 = 5.56$; stem $Q25 = 2.62$, $Q75 = 42.78$; root $Q25 = 0.39$, $Q75 = 13.37$ (normalized units).

(XLSX)

S8 Table. Raw TPM values for TPS genes by tissue and replicate.

(XLSX)

S9 Table. Blast homology search results for the amino acid sequence of PhTPS1. The amino acid sequence of PhTPS1 was used as the query sequence. The non-redundant protein sequences (nr) database was used for the homology search, and the BLASTP algorithm was used for analysis.

(XLSX)

S10 Table. Number and proportion of reads corresponding to *allele1* and *allele2* in the remapping to *PhTPS1*.

(XLSX)

S11 Table. Transcript IDs, TSA accession numbers, and CDS coordinates of analyzed TPS genes.

(XLSX)

S1 Fig. *P. hadiensis* tissues used in this study (from left to right: whole plant, leaf and stem, and root).

(PDF)

S2 Fig. Workflow for TPS identification, expression profiling, and yeast validation in *P. hadiensis*.

(PDF)

S3 Fig. Amino acid sequence alignments of the putative RbcL protein (TRINITY_DN2794_c0_g1_i1) and the known *P. hadiensis* RbcL protein (YP_010593282.1).

(PDF)

S4 Fig. GO (biological process) terms enriched in the leaf relative to those in the root.

(PDF)

S5 Fig. Nucleotide sequence alignment of *PhTPS1* *Allele1* and *Allele2*.

(PDF)

S6 Fig. Protein sequence alignment of *PhTPS1* *Allele1* and *Allele2*.

(PDF)

S7 Fig. IGV visualization of reads remapped to *PhTPS1*, highlighting the magnified region where two *PhTPS1* alleles are present. The gray vertical bars at the top indicate the read depth at each nucleotide position. The depth from position 66 to position 71 is approximately half of that at the surrounding positions, suggesting the presence of a six-nucleotide deletion in one allele (*allele2*) and retention of these six nucleotides in the other allele (*allele1*).

(PDF)

S8 Fig. Overview of the workflow for yeast strain generation in this study. The numbers in square brackets indicate the identifiers (numbers) of the plasmids, vectors, or yeast strains listed in [S3 Table](#).

(PDF)

S9 Fig. Total ion chromatograms of headspace samples from vials containing cultured GPP-producing yeast, as analyzed using GC–MS. The top panel shows an extracted ion chromatogram focusing on m/z 93 at retention times around 13.1 min.

(PDF)

S10 Fig. Comparison of the mass spectra. The upper spectrum represents the mass spectrum of the volatile compound in the headspace of vials containing cultured PhTPS1/GPP. The lower spectrum shows the reference library spectrum of limonene from the NIST 20 Mass Spectral Library. Major fragment ions at m/z 68, 93, 121, and 136 indicate a high similarity between the sample and library spectra.

(PDF)

S1 Appendix. Nucleotide sequences (coding sequences) of the 26 TPS genes identified in *P. hadiensis*.

(CDS)

S2 Appendix. Amino acid sequences of the 26 TPS proteins identified in *P. hadiensis*.

(PEP)

Author contributions

Conceptualization: Hinano Mizuno, Yoshikazu Furuta, Hiroki Sugimoto.

Data curation: Hinano Mizuno.

Formal analysis: Hinano Mizuno, Koichiro Iwai.

Investigation: Hinano Mizuno, Kenro Tokuhito, Koichiro Iwai, Akari Nakasone, Hidenori Tanaka, Hiroki Sugimoto.

Methodology: Hinano Mizuno, Koichiro Iwai, Hiroki Sugimoto.

Project administration: Hiroki Sugimoto.

Resources: Kenro Tokuhito, Akari Nakasone, Hidenori Tanaka, Hiroki Sugimoto.

Visualization: Hinano Mizuno.

Writing – original draft: Hinano Mizuno.

Writing – review & editing: Hinano Mizuno, Kenro Tokuhito, Koichiro Iwai, Yoshikazu Furuta, Akari Nakasone, Hidenori Tanaka, Hiroki Sugimoto.

References

1. Dhifi W, Bellili S, Jazi S, Bahloul N, Mnif W. Essential Oils' Chemical Characterization and Investigation of Some Biological Activities: A Critical Review. *Medicines* (Basel). 2016;3(4):25. <https://doi.org/10.3390/medicines3040025> PMID: 28930135
2. Moss GP, Smith PAS, Tavernier D. Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995). *Pure and Applied Chemistry*. 1995;67(8–9):1307–75. <https://doi.org/10.1351/pac199567081307>
3. Chen R, Wang M, Keasling JD, Hu T, Yin X. Expanding the structural diversity of terpenes by synthetic biology approaches. *Trends Biotechnol*. 2024;42(6):699–713. <https://doi.org/10.1016/j.tibtech.2023.12.006> PMID: 38233232
4. Christianson DW. Structural and Chemical Biology of Terpenoid Cyclases. *Chem Rev*. 2017;117(17):11570–648. <https://doi.org/10.1021/acs.chemrev.7b00287> PMID: 28841019
5. Geron C, Rasmussen R, Arnts RR, Guenther A. A review and synthesis of monoterpene speciation from forests in the United States. *Atmos Environ*. 2000;34:1761–81. [https://doi.org/10.1016/S1352-2310\(99\)00364-7](https://doi.org/10.1016/S1352-2310(99)00364-7)
6. Seigler DS. Sesquiterpenes. In: Seigler DS, editor. *Plant secondary metabolism*. Boston: Springer. 1998:277–303.
7. Bose SK, Yadav RK, Mishra S, Sangwan RS, Singh AK, Mishra B, et al. Effect of gibberellic acid and calliterpenone on plant growth attributes, trichomes, essential oil biosynthesis and pathway gene expression in differential manner in *Mentha arvensis* L. *Plant Physiol Biochem*. 2013;66:150–8. <https://doi.org/10.1016/j.plaphy.2013.02.011> PMID: 23514759
8. Lanzotti V. *Diterpenes for Therapeutic Use*. Natural Products. Springer Berlin Heidelberg. 2013:3173–91. https://doi.org/10.1007/978-3-642-22144-6_192
9. Helfrich EJN, Lin G-M, Voigt CA, Clardy J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J Org Chem*. 2019;15:2889–906. <https://doi.org/10.3762/bjoc.15.283> PMID: 31839835
10. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716

11. Starks CM, Back K, Chappell J, Noel JP. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science*. 1997;277(5333):1815–20. <https://doi.org/10.1126/science.277.5333.1815> PMID: 9295271
12. Jiang S-Y, Jin J, Sarojam R, Ramachandran S. A Comprehensive Survey on the Terpene Synthase Gene Family Provides New Insight into Its Evolutionary Patterns. *Genome Biol Evol*. 2019;11(8):2078–98. <https://doi.org/10.1093/gbe/evz142> PMID: 31304957
13. 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(1):212–29. <https://doi.org/10.1111/j.1365-313X.2011.04520.x> PMID: 21443633
14. Zhang A, Xiong Y, Fang J, Jiang X, Wang T, Liu K, et al. Diversity and Functional Evolution of Terpene Synthases in Rosaceae. *Plants (Basel)*. 2022;11(6):736. <https://doi.org/10.3390/plants11060736> PMID: 35336617
15. Bao T, Kimani S, Li Y, Li H, Yang S, Zhang J, et al. Allelic variation of terpene synthases drives terpene diversity in the wild species of the *Freesia* genus. *Plant Physiol*. 2023;192(3):2419–35. <https://doi.org/10.1093/plphys/kiad172> PMID: 36932696
16. Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*. 2009;70(15–16):1621–37. <https://doi.org/10.1016/j.phytochem.2009.07.030> PMID: 19793600
17. Zhou F, Pichersky E. More is better: the diversity of terpene metabolism in plants. *Curr Opin Plant Biol*. 2020;55:1–10. <https://doi.org/10.1016/j.pbi.2020.01.005> PMID: 32088555
18. Chrysargyris A. It Runs in the Family: The Importance of the Lamiaceae Family Species. *Agronomy*. 2024;14(6):1274. <https://doi.org/10.3390/agronomy14061274>
19. Lima AS, Schimmel J, Lukas B, Novak J, Barroso JG, Figueiredo AC, et al. Genomic characterization, molecular cloning and expression analysis of two terpene synthases from *Thymus caespitius* (Lamiaceae). *Planta*. 2013;238(1):191–204. <https://doi.org/10.1007/s00425-013-1884-2> PMID: 23624978
20. Chen Z, Vining KJ, Qi X, Yu X, Zheng Y, Liu Z, et al. Genome-Wide Analysis of Terpene Synthase Gene Family in *Mentha longifolia* and Catalytic Activity Analysis of a Single Terpene Synthase. *Genes (Basel)*. 2021;12(4):518. <https://doi.org/10.3390/genes12040518> PMID: 33918244
21. Ashaari NS, Ab Rahim MH, Sabri S, Lai KS, Song AA-L, Abdul Rahim R, et al. Functional characterization of a new terpene synthase from *Plectranthus amboinicus*. *PLoS One*. 2020;15(7):e0235416. <https://doi.org/10.1371/journal.pone.0235416> PMID: 32614884
22. Crocoll C, Asbach J, Novak J, Gershenzon J, Degenhardt J. Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Mol Biol*. 2010;73(6):587–603. <https://doi.org/10.1007/s11103-010-9636-1> PMID: 20419468
23. Mendes MD, Barroso JG, Oliveira MM, Trindade H. Identification and characterization of a second isogene encoding γ -terpinene synthase in *Thymus caespitius*. *J Plant Physiol*. 2014;171(12):1017–27. <https://doi.org/10.1016/j.jplph.2014.04.001> PMID: 24974328
24. Barbosa M de O, Wilairatana P, Leite GM de L, Delmondes G de A, Silva LYS da, Júnior SCA, et al. *Plectranthus* Species with Anti-Inflammatory and Analgesic Potential: A Systematic Review on Ethnobotanical and Pharmacological Findings. *Molecules*. 2023;28(15):5653. <https://doi.org/10.3390/molecules28155653> PMID: 37570622
25. Ahamed AN, Yaser SM, Idhris SM, Padusha MSA, Sherif NA. Phytochemical and pharmacological potential of the genus *Plectranthus*—a review. *S Afr J Bot*. 2023;154:159–89. <https://doi.org/10.1016/j.sajb.2023.01.026>
26. Lukhoba CW, Simmonds MSJ, Paton AJ. *Plectranthus*: a review of ethnobotanical uses. *J Ethnopharmacol*. 2006;103(1):1–24. <https://doi.org/10.1016/j.jep.2005.09.011> PMID: 16289602
27. Hossain CM, Gera M, Ali KA. Current status and challenges of herbal drug development and regulatory aspect: a global perspective. *Asian J Pharm Clin Res*. 2022;31–41. <https://doi.org/10.22159/ajpcr.2022.v15i12.46134>
28. Bryson AE, Lanier ER, Lau KH, Hamilton JP, Vaillancourt B, Mathieu D, et al. Uncovering a multiradiene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory. *Nat Commun*. 2023;14(1):343. <https://doi.org/10.1038/s41467-023-35845-1> PMID: 36670101
29. Sun Y, Liu T, Sun C, Luo Q. Chemical constituents of *Plectranthus tomentosa* extract and its control effect on *Tetranychus kanzawai*. *J Chem*. 2022;2022:5609391. <https://doi.org/10.1155/2022/5609391>
30. Mothana RAA, Abdo SAA, Hasson S, Althawab FMN, Alaghbari SAZ, Lindequist U. Antimicrobial, antioxidant and cytotoxic activities and phytochemical screening of some yemeni medicinal plants. *Evid Based Complement Alternat Med*. 2010;7(3):323–30. <https://doi.org/10.1093/ecam/nen004> PMID: 18955315
31. Menon DB, Gopalakrishnan VK. Terpenoids Isolated From the Shoot of *Plectranthus hadiensis* Induces Apoptosis in Human Colon Cancer Cells Via the Mitochondria-Dependent Pathway. *Nutr Cancer*. 2015;67(4):697–705. <https://doi.org/10.1080/01635581.2015.1019631> PMID: 25837437
32. Menon D, Sasikumar J, Gopalakrishnan VK. Antioxidant and anti-inflammatory properties of terpenoid fraction isolated from the shoot of *Plectranthus hadiensis*. *Int J Pharma Bio Sci*. 2014;5:B197–205.
33. Amarasinghe P, Siriwardhana A, Samarasekera R, Weerasena J. Development of DNA barcodes and optimization of DNA isolation and PCR protocols for a medicinal plant *Plectranthus hadiensis*. *Int J Sci Res*. 2015;4:2852–6.
34. Menon DB, Sasikumar JM, Latha K. Anti-inflammatory and cytotoxic activity of methanolic extract of *Plectranthus hadiensis* stem. *Pharmacologyonline*. 2011;3:275–82.
35. Hao J, Lu Y, Dang M, Xia R, Xu L, Zhu Z, et al. The complete chloroplast genome sequence of *Plectranthus hadiensis* (Lamiaceae) and phylogenetic analysis. *Mitochondrial DNA B Resour*. 2023;8(10):1049–53. <https://doi.org/10.1080/23802359.2023.2262689> PMID: 37810612

36. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560> PMID: [30423086](#)
37. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863–4. <https://doi.org/10.1093/bioinformatics/btr026> PMID: [21278185](#)
38. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52. <https://doi.org/10.1038/nbt.1883> PMID: [21572440](#)
39. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512. <https://doi.org/10.1038/nprot.2013.084> PMID: [23845962](#)
40. TransDecoder. TransDecoder. GitHub. <https://github.com/TransDecoder/TransDecoder>. 2023. Accessed 2024 December 17.
41. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031> PMID: [24451626](#)
42. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197> PMID: [28263959](#)
43. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351> PMID: [26059717](#)
44. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol*. 2021;38(10):4647–54. <https://doi.org/10.1093/molbev/msab199> PMID: [34320186](#)
45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402. <https://doi.org/10.1093/nar/25.17.3389> PMID: [9254694](#)
46. Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, et al. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res*. 2001;29(14):2994–3005. <https://doi.org/10.1093/nar/29.14.2994> PMID: [11452024](#)
47. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4> PMID: [31375807](#)
48. Perteu M, Perteu GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122> PMID: [25690850](#)
49. Su W, Sun J, Shimizu K, Kadota K. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res Notes*. 2019;12(1):133. <https://doi.org/10.1186/s13104-019-4179-2> PMID: [30867032](#)
50. Alexa A, Rahnenführer J. topGO: Enrichment analysis for gene ontology. Bioconductor packages. <https://www.bioconductor.org/packages/release/bioc/html/topGO.html>. 2023. Accessed 2024 December 17.
51. R Core T. R: a language and environment for statistical computing. R Project. <https://www.R-project.org/>. Accessed 2025 March 26.
52. Kolde R. pheatmap: Pretty heatmaps. 2019.
53. Yan X-M, Zhou S-S, Liu H, Zhao S-W, Tian X-C, Shi T-L, et al. Unraveling the evolutionary dynamics of the TPS gene family in land plants. *Front Plant Sci*. 2023;14:1273648. <https://doi.org/10.3389/fpls.2023.1273648> PMID: [37900760](#)
54. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: [23329690](#)
55. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348> PMID: [19505945](#)
56. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*. 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015> PMID: [32011700](#)
57. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285> PMID: [28481363](#)
58. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*. 2018;35(2):518–22. <https://doi.org/10.1093/molbev/msx281> PMID: [29077904](#)
59. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(W1):W78–82. <https://doi.org/10.1093/nar/gkae268> PMID: [38613393](#)
60. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G. Integrative Genomics Viewer. *Nat Biotechnol*. 2011;29:24–6. <https://doi.org/10.1038/nbt.1754>
61. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91. <https://doi.org/10.1093/bioinformatics/btp033>

62. Lückner J, El Tamer MK, Schwab W, Verstappen FWA, van der Plas LHW, Bouwmeester HJ, et al. Monoterpene biosynthesis in lemon (*Citrus limon*). cDNA isolation and functional analysis of four monoterpene synthases. *Eur J Biochem*. 2002;269(13):3160–71. <https://doi.org/10.1046/j.1432-1033.2002.02985.x> PMID: [12084056](#)
63. Ohto C, Muramatsu M, Obata S, Sakuradani E, Shimizu S. Overexpression of the gene encoding HMG-CoA reductase in *Saccharomyces cerevisiae* for production of prenyl alcohols. *Appl Microbiol Biotechnol*. 2009;82(5):837–45. <https://doi.org/10.1007/s00253-008-1807-5> PMID: [19083230](#)
64. Tokuhito K, Muramatsu M, Ohto C, Kawaguchi T, Obata S, Muramoto N, et al. Overproduction of geranylgeraniol by metabolically engineered *Saccharomyces cerevisiae*. *Appl Environ Microbiol*. 2009;75(17):5536–43. <https://doi.org/10.1128/AEM.00277-09> PMID: [19592534](#)
65. Rubat S, Varas I, Sepúlveda R, Almonacid D, González-Nilo F, Agosin E. Increasing the intracellular isoprenoid pool in *Saccharomyces cerevisiae* by structural fine-tuning of a bifunctional farnesyl diphosphate synthase. *FEMS Yeast Res*. 2017;17(4). <https://doi.org/10.1093/femsyr/fox032> PMID: [28854674](#)
66. Yan Y, Li M, Zhang X, Kong W, Bendahmane M, Bao M, et al. Tissue-Specific Expression of the Terpene Synthase Family Genes in *Rosa chinensis* and Effect of Abiotic Stress Conditions. *Genes (Basel)*. 2022;13(3):547. <https://doi.org/10.3390/genes13030547> PMID: [35328100](#)
67. Muchlinski A, Chen X, Lovell JT, Köllner TG, Pelot KA, Zerbe P, et al. Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (*Panicum virgatum* L.). *Front Plant Sci*. 2019;10:1144. <https://doi.org/10.3389/fpls.2019.01144> PMID: [31608090](#)
68. Tholl D. Biosynthesis and biological functions of terpenoids in plants. *Adv Biochem Eng Biotechnol*. 2015;148:63–106. https://doi.org/10.1007/10_2014_295 PMID: [25583224](#)
69. Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds?. *New Phytol*. 2018;220(3):692–702. <https://doi.org/10.1111/nph.14178> PMID: [27604856](#)
70. Karunanithi PS, Zerbe P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front Plant Sci*. 2019;10:1166. <https://doi.org/10.3389/fpls.2019.01166> PMID: [31632418](#)
71. Aubourg S, Lecharny A, Bohlmann J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol Genet Genomics*. 2002;267(6):730–45. <https://doi.org/10.1007/s00438-002-0709-y> PMID: [12207221](#)
72. Hamilton JP, Godden GT, Lanier E, Bhat WW, Kinser TJ, Vaillancourt B, et al. Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, *Callicarpa americana*. *Gigascience*. 2020;9(9):giaa093. <https://doi.org/10.1093/giga-science/giaa093> PMID: [32893861](#)
73. Bornowski N, Hamilton JP, Liao P, Wood JC, Dudareva N, Buell CR. Corrigendum to: Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. *DNA Res*. 2020;27(4):dsaa025. <https://doi.org/10.1093/dnares/dsaa025> PMID: [33146674](#)
74. Zhang S, Qi X, Zhu R, Ye D, Shou M, Peng L, et al. Transcriptome Analysis of *Salvia miltiorrhiza* under Drought Stress. *Plants (Basel)*. 2024;13(2):161. <https://doi.org/10.3390/plants13020161> PMID: [38256715](#)
75. Wimberley J, Cahill J, Atamian HS. De novo Sequencing and Analysis of *Salvia hispanica* Tissue-Specific Transcriptome and Identification of Genes Involved in Terpenoid Biosynthesis. *Plants (Basel)*. 2020;9(3):405. <https://doi.org/10.3390/plants9030405> PMID: [32213996](#)