Medicinal Plant Biology

ARTICLE

Open Access

https://doi.org/10.48130/mpb-0024-0017

Medicinal Plant Biology **2024**, in press

Analysis of monoterpene biosynthesis and functional TPSs of *Perilla frutescens* based on transcriptome and metabolome

Shi-Mei Yang[1,4#], Hong-Ye Chu[1#], Yu-Xia Wang[1], Bao-Lin Guo[3], Tian-Yue An[2*] and Qi Shen[1*]

[1] *Institute of Medical Plant Physiology and Ecology, School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China*
[2] *Featured Laboratory for Biosynthesis and Target Discovery of Active Components of Traditional Chinese Medicine, School of Integrated Traditional Chinese and Western Medicine, Binzhou Medical University, Yantai 264003, China*
[3] *The Institute of Medicinal Plant Development, the Chinese Academy of Medical Sciences, The Institute of Medicinal Plant Development (IMPLAD), Beijing 100193, China*
[4] *Guizhou Institute of Subtropical Crop, Guizhou Academy of Agricultural Sciences, Guiyang 550025, China*
[#] Authors contributed equally: Shi-Mei Yang, Hong-Ye Chu
* Corresponding authors, E-mail: antianyue2007@126.com; shenqi@gzucm.edu.cn

Abstract

Perilla is a characteristic medicinal and edible plant. According to the different main monoterpenes in *Perilla* leaves, Perilla can be divided into many chemical types. In this study, four *Perilla* cultivars of different chemical types, included perillaldehyde (PA), perillaketone (PK), perillene (PL) and piperitenone (PT) were selected for metabolome and transcriptome analysis. Totally, 35 terpene compounds were identified and main monoterpenes were consistent with corresponding chemical types, respectively. Combined with transcriptome, a total of 5,920 differentially expressed genes were obtained among them. Totally, 69 genes refered to MVA and MEP pathways and 109 PfTPSs encoded genes were identified, which possessed partial differentially expressed in four chemical types. The core *PfTPSs* in co-expression analysis were functional characterization. Three *PfTPSs* were identified as linalool synthase and one *PfTPSs* were identified as geranyl synthase. This research analyzed the monoterpene biosynthesis and functional characterization of TPSs in *Perilla*, which can give foundation for in-depth research of *Perilla* chemotype metabolic mechanism.

Citation: Yang SM, Chu HY, Wang YX, Guo BL, An TY, et al. 2024. Analysis of monoterpene biosynthesis and functional TPSs of *Perilla frutescens* based on transcriptome and metabolome. *Medicinal Plant Biology* https://doi.org/10.48130/mpb-0024-0017

Introduction

Perilla frutescens (L.) Britton, belonging to Lamiaceae, is a kind of versatile economic crops, and is also a commonly used herbal medicine. It has been widely cultivated in China, Japan, South Korea and many other Asian countries in recent years[1]. The leaves of *Perilla* are used for preparation of cold granules in Traditional Chinese Medicine (TCM), and as vegetables and spices in Asian countries[1,2]. The leaves of *Perilla* possess various chemicals, including terpenes, flavonoids, phenolic acids etc.[3]. The medicinal value of *Perilla* has been officialized in the Chinese pharmacopoeia and the catalogue of affinal drug and diet[1,4]. The essential oils of *Perilla* leaves are the major medicinal flavor components. They are also widely applied in skin care produce and aromatization industry[5,6].

The essential oils of *Perilla* leaves include abundant diversity of chemical types, which are classified into the monoterpene (MT)-type oils and phenylpropene (PP)-type oils[7,8]. Interestingly, there are multiple kinds of monoterpenes in the leaves of Perilla genus, which can be further classified into the following six chemotypes according to their principal constituents: perillaldehyde (PA), perillaketone (PK), perillene (PL), piperitenone (PT), citral (C) and elsholtziaketone (EK) types[9]. Among these monoterpenes, PA are the major aromatic medical ingredient for prescriptions in China and Japan, while PK were thought to be a potent lung toxin[10]. Besides these main monoterpenes, geraniol (GL) is an acyclic monoterpene also commonly found in a wide range of *Perilla* plants[11], and linalool (LL) can be found in all *Perilla* plants and may be a dead-end compound in general monoterpene biosynthetic pathways[12]. Wherein, PA, PL, GL and LL are commercially important for perfumery, food and medicine[13,14]. As multiple chemical types of monoterpenes are enriched in *Perilla*, *Perilla* hence has been thought as a model system for the study of monoterpenes metabolism.

The biosynthesis of monoterpene is specially localized to the glandular trichomes and initiated from the mevalonate (MAV)/methylerythritol 4-phosphate (MEP) pathway in plants[15–17]. Then the terpene synthases (TPS) catalyze prenyl pyrophosphates, the products of MVA and MEP pathways, to the formation of terpene compounds, and cytochrome P450s (CYP450s) further modify the backbones of these terpenes[17]. Recently, some TPSs and CYP450s involved in *Perilla* monoterpene biosynthesis have been reported, including limonene synthase, GL synthase, LL synthase and mono-oxygenase[18–20]. Enzymes participated in PA biosynthesis, such as CYP71D18 and CYP17A7146, have been characterized[18–20] . Eight double-bond reductases (PfDBRs) that catalyze the conversion of isopyrone and soyone to PK were identified by enzymatic method in *vitro*[21]. The recent high-quality and chromosome-scale *Perilla* genome data establishes great foundation for the characterization of *Perilla* monoterpene biosynthesis[22]. Here in this study, we carried out transcriptome analysis of four different *Perilla* chemotypes, and identified some TPSs involved in monoter-

pene biosynthesis, as well as multiple regulatory factors responsible for this biosynthesis pathway. Furthermore, the function of selected TPSs were characterized by heterologous expression system and in *vitro* enzyme assay. These results collectively will help to understand the molecular mechanism of Perilla monoterpene biosynthesis and analyze the biosynthetic pathways of terpenes in *Perilla*.

Methods

Plant materials

The perilla cultivars of four chemotypes, including PA-, PK-, PL- and PT-types, were selected and planted in the greenhouse of Guangzhou University of Chinese Medicine (Guangzhou, Guangdong, China, 113°41' E; 23°07' N). The PA-type cultivar, with purple wrinkled leaves, belongs to *P. frutescens* var. *crispa*. The PK-, PL- and PT-type cultivars, with green non-wrinkled leaves, belong to *P. frutescens* var. *frutescens* (Fig. 1a). The leaves at seedling stages were collected for gas chromatography-mass spectrometry (GC-MS) analysis and RNA extraction. All samples are stored in liquid nitrogen immediately after sampling.

GC-MS analysis

Perilla leaves of four chemotypes were crushed. Then, 0.2 g leaves powder was extracted by petroleum ether and filtered for GC-MS analysis. Analysis conditions include: RXT-5 MS quartz capillary column (30 m × 0.25 μm × 0.25 μm); The front column pressure is 63.9 kPa; The initial temperature of the column was 80 °C and was retained for 1 min. After the heating rate of 15 °C/min, the column was increased to 300 °C and retained for 15 min. MS conditions: ionization mode EI, filament current 0.5mA; Electron energy 70 eV; Multiplier voltage 0.86kV; Ion source 230 °C, solvent delay 1 min; Plasma/nucleus ratio m/z: 40~500. The NIST spectrum library was retrieved by Agilent qualitative software, and the chemical structure analysis was combined to identify the species of components. The relative percentage of each chemical component of volatile oil was calculated by peak area normalization method.

Transcriptome analysis

Total RNA from 12 *Perilla* leaves were extracted using the RNApre Pure plant RNA extraction kit (DP432) (Tiangen, Beijing, China). The mRNA sequencing library was constructed using the NEBNext® Ultra RNA Library Prep Kit (New England Biolabs Inc., Ipswich, MA, USA). Then, the sequencing library was analyzed using the Agilent 2100 Bioanalyzer and was sequenced by an Illumina HiSeq™ 2000 sequencing system (Illumina Inc., San Diego, CA, USA). The original transcriptome data has been uploaded to the NGDC database (National Genomics Data Center) (Number: PRJCA021059).

The *Perilla* genome data came from the National Center for Biotechnology Information (NCBI, Accession No.: GCA_019511825.2)[22]. Trimmomatic software is used for quality control of transcriptome data[23]. STAR(v2.7.10a) software was used to build an index and clen data was compared to the *Perilla* genome[24]. The Python module HTseq is used for *P. frutescens* transcriptome data quantification[25]. Gene expression levels of fragments per kilobase of transcript per million mapped reads (FPKM) were calculated. Then differentially expressed genes (DEGs) were identified using DESeq2[26]. Genes with |log2 (Fold change) |≥1|and $P < 0.01$ were considered DEGs. The online tool eggnog (http://eggnog-mapper.embl.de/) was used to annotate the whole genome protein of *Perilla*[27]. R package ClusterProfiler was used for GO(Gene ontology) and KEGG(Kyoto Encyclopedia of Genes and Genomes) enrichment analysis of differential genes[28].

Identification and characteristic analysis of metabolosynthetic genes

The two hidden Markov models (HMM) of TPS (Terpene_synthase, PF01397 and Terpene_synthase_C, PF03936) were downloaded from Pfam (http://pfam.xfam.org/) and the *Perilla* genome were searched[29]. The identified PfTPS proteins were further determined by online HMMER (www.ebi.ac.uk/Tools/hmmer/) and constructed phylogenetic tree using Neighbor-Joining method in MEGA X (Bootstrap 1000)[30]. Heat maps of candidate genes were drawn by TBtools (v1.112) (https://github.com/CJ-Chen/TBtools)[31]. Intergenomic collinearity analysis using MCScanX[32]. Chromosomal localization and collinearity results were visualized using TBtools. According to the Annotation information of Metabolic pathway synthase in KEGG and eggnog, the encoded gens were identified in MVA/MEP pathway in *Perilla*. The Python script is used to calculate the correlation coefficient between genes expression, using the Pearson correlation coefficient[33].

Real-time PCR analysis and cloning of related genes

The full-length transcripts of *PfTPSs* genes were derived by 5' RACE-PCR and/or 3' RACE-PCR. Then the PCR products were ligated to PLB vector (Tiangen) and sequenced by Sangon Biotech. Primer3Plus (www.primer3plus.com/index.html) was used to design primers for *PfTPS* genes. The primers used for genes cloning were listed in Supplemental Table S1. The fluorescence quantitative reaction system was referred to Wu et al., three replicates were used in each group, and *PfActin* was used as the key gene for analysis[34]. For data analysis, refer to $2^{-\Delta\Delta CT}$ method[35].

Functional characterization in heterologous expression system

All the cloned *PfTPSs* genes were introduced into the heterologous expression vector pETDuet-1. Then the expression plasmids were transformed into C41 *Escherichia coli* (*E. coli*) strain. The positive colony were firstly cultivated in TB medium at 37 °C to an initial OD_{600} of 0.4-0.6, and then the cultures were induced by 1 mM IPTG for another incubation of 72 h at 16 °C. After the cultivation, the cultures were extracted for three times by equal volume of n-hexane, and then the extracts were concentrated by rotary evaporation instrument for gas chromatography-mass spectrometry (GC-MS) detection.

In *vitro* enzymatic reaction

The selected *PfTPSs* genes were introduced into *pET28a* vector and transformed into *E. coli* strain BL21 (DE3). The positive colony were incubated in LB medium at 37 °C to the initial OD_{600} of 0.4–0.6, and then the cultures were induced by 0.5 mM IPTG for another 12 h of cultivation. The cultured cells were harvested and resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 20 mM β-mercaptoethanol, pH 8.0) for 30 min at 4 °C. Then the cells were disrupted by ultrasonication and the lysate was centrifuged at 13,000 g and 4 °C for 30 min. The crude proteins were inside the supernatant. For the soluble PfTPSs proteins, the crude proteins were purified by HIS nickel column.

For in *vitro* enzymatic reaction, crude or purified PfTPSs proteins were added to the reaction buffer containing 200 mM Tris-HCl (pH 7.5), 40 mM $MgCl_2$, 10% glycerol and 1 mM geranyl pyrophosphate (GPP) as the precursor. After incubation at 30 °C for 30 min, the reaction system was extracted equal volume of n-hexane and then detected by GC-MS.

Results

Identification of essential oil components

The volatile essential oils components of leaves from four *Perilla* cultivars were analyzed using GC-MS analysis (Fig. 1a, b). A total of 35 terpenes, including 22 monoterpenes and 13 sesquiterpenes, were detected in these cultivars. The main monoterpenes are consistent with their chemotypes classification. 68.01% of PA was identified as the main compounds in PA-type cultivars, while 71.65% of PL, 88.76 % of PK and 61.20% of PT are the main compounds in PL-type, PK-type, PT-type varieties, respectively (Fig. 1b). The GL (0.07%–0.68%) and LL (0.03%–1.77%) were the ubiquitous and trace metabolites existed in these *Perilla* cultivars. Besides these main chemicals, other monoterpenes and sesquiterpenes were also identified, including limonene, borneol, thujone, verbenol, citral, carvone, trans-shisool, terpine, thymol 2-pinen-4-on, caryophyllene, germacrene, farnesene, trans-nerlidol etc. (Fig. 1c; Supplemental Table S2).

Transcriptome analysis

To explore the molecular mechanism involved in different monoterpenes biosynthesis, the transcriptome analysis was performed for the leaves of four chemotype cultivars. After sequence and data filtration, a total of 579 million clean reads, comprising of 86.90 Gb nucleotide bases with average 46.26% GC were obtained (Supplemental Table S3). The average 98% clean reads were assembled to the *Perilla* genome (GCA_019511825.2) (Supplemental Table S4). Then, the gene annotation and differential expression analysis were carried out among the four cultivars. For the PA type are the main medicinal component of *Perilla* according to Chinese Pharmacopoeia. Hence, more attention focused on the PA-type. Totally 236 specific up-regulated genes compared with other three cultivars (Fig. 2a). In the specific up-regulated genes of PA-type, phylpropanoid and monoterpene biosynthesis were enriched using the KEGG enrichment analysis (Fig. 2b; Supplemental Table S5). Meanwhile, the comparison analysis among the other three different chemotype cultivar, the 79, 92 and 155 specific up-regulated genes were identified in PK -type, PL-type, PT -type, respectively (Fig. 2c, e & g). Genes involved in terpenes, including monoterpene, sesquiterpene, diterpene

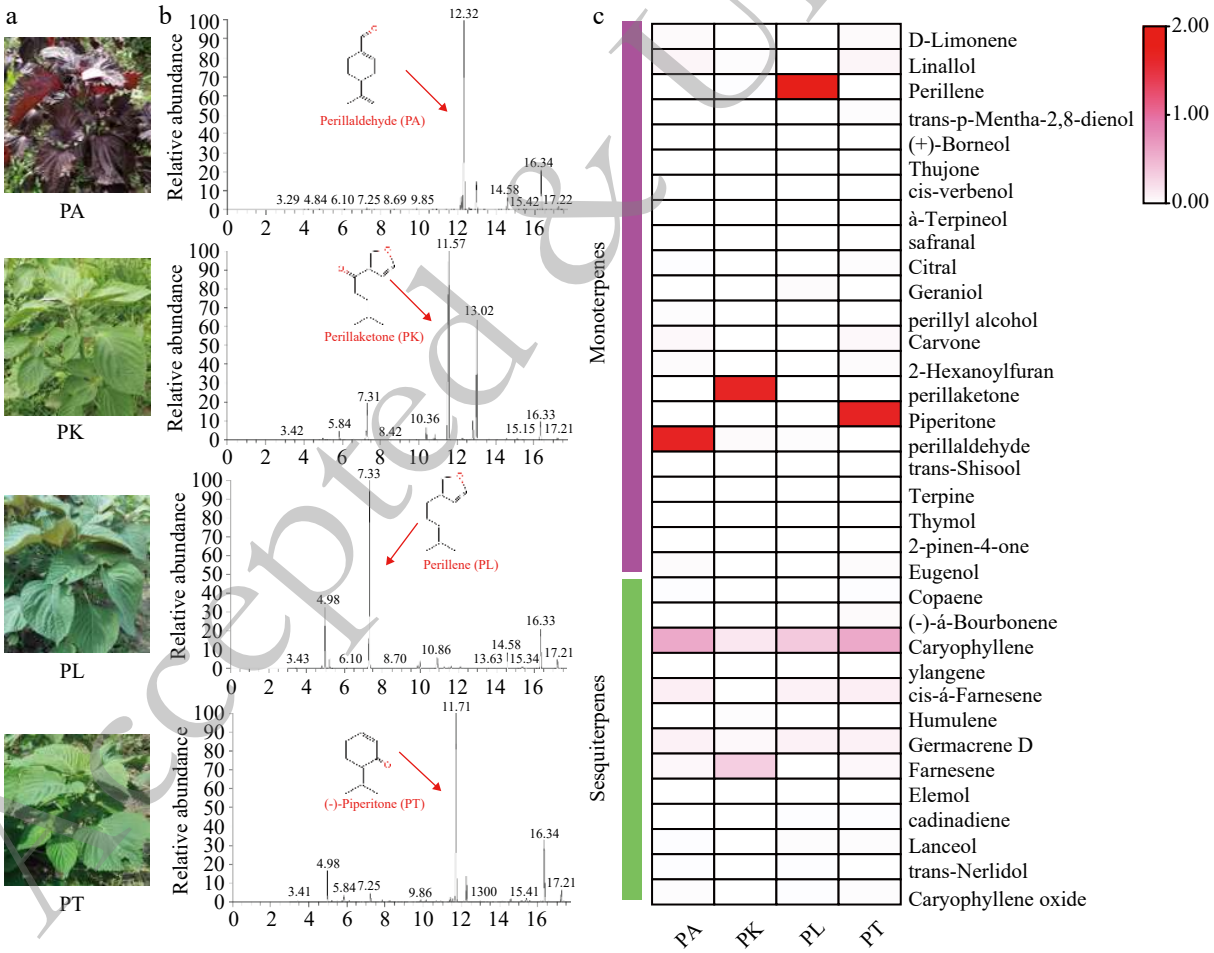

Fig. 1 The analysis of volatile components in four perilla cultivars. (a) the phenotype of four chemical types of *Perilla*; (b) The GC-MS analysis of volatile essential components from four *Perilla* leaves; (c) The heat map of metabolite contents in four *Perilla* leaves.

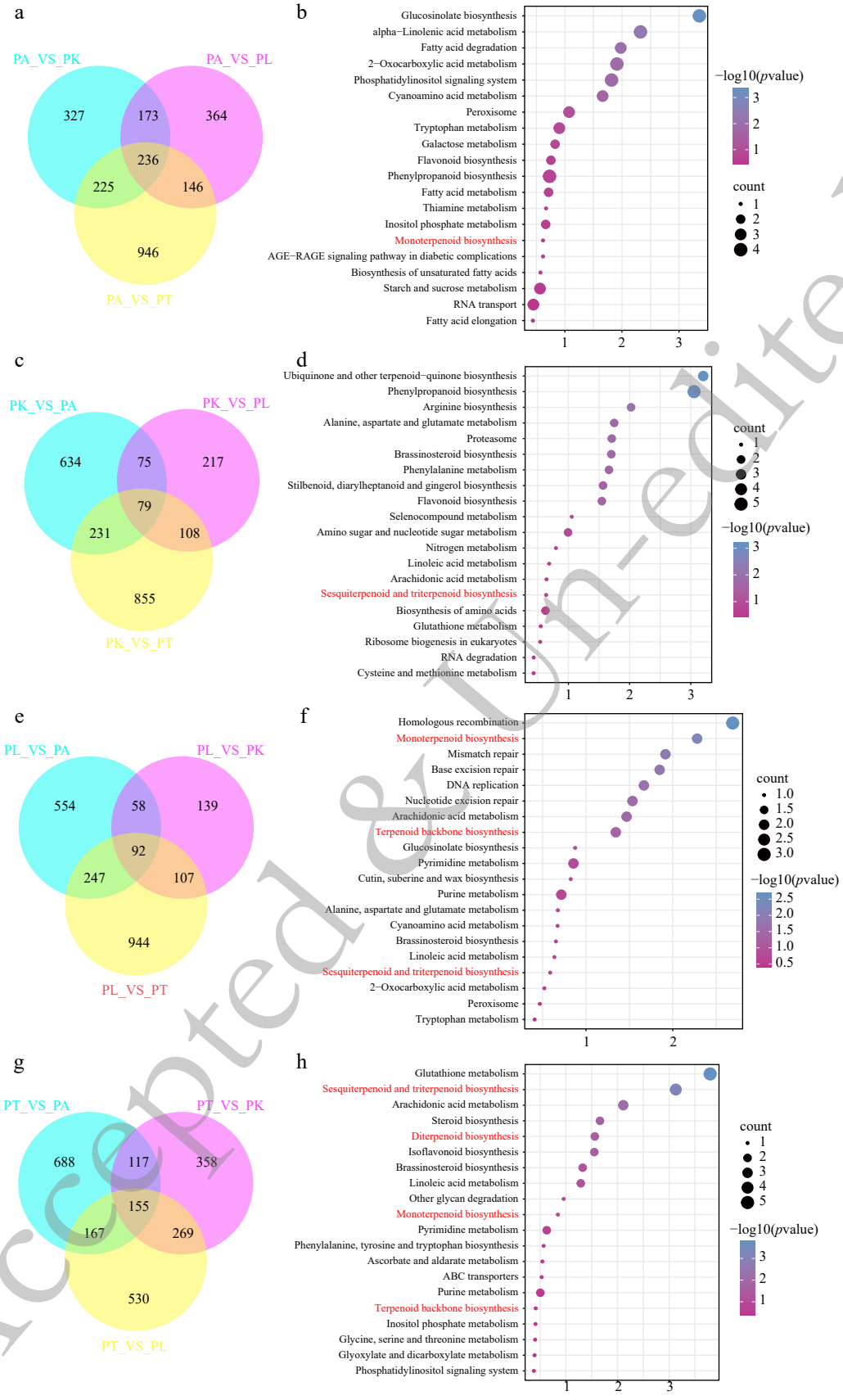

Fig. 2 Different expression genes and KEGG enrichment analysis of four chemotype *Perilla* cultivars. The intersection of PA, PL, PK, PT4 chemotypes with the other 3 chemotypes was indicated in the Venn diagram; (b, d, f) H: KEGG enrichment analysis of special up-expressed genes in PA, PL, PK, PT-type, respectively.

and triterpene, were enriched in corresponding chemical type (Fig. 2d–f; Supplemental Tables S6–S8).

Biosynthetic pathway of the volatile oils in *P. frutescens*

The MEP and MVA pathways are the basic terpene biosynthesis pathway. The 69 genes encoding 17 enzymes in MEP and MVA pathways were identified in the four cultivars (Fig. 3, Supplemental Table S9). The MEP pathway starts with pyruvate, which is catalyzed by DXS to form 1-deoxy-D-xylulose-5-phosphate. Subsequently, it are continuous catalyzed by DXR, MCT, CMK, MDS, HDS, HDR to form MEcPP. There two encoded genes of DXR, MCT, CMK, MDS in *Perilla* and showed upregulated expression in PA-type. In the MVA pathway, Acetyl-CoA are catalyzed by AACT HMGS, HMGR, MVK, PMK, MPDC IPP and DMAPP to IPP. Finally, the equilibrium between IPP and DMAPP are controlled by isopentenyl diphosphate isomerase (IPPI)-encoded genes and the further reaction synthesize by geranyl pyrophosphate synthase (GPPS)-encoded genes to produce the

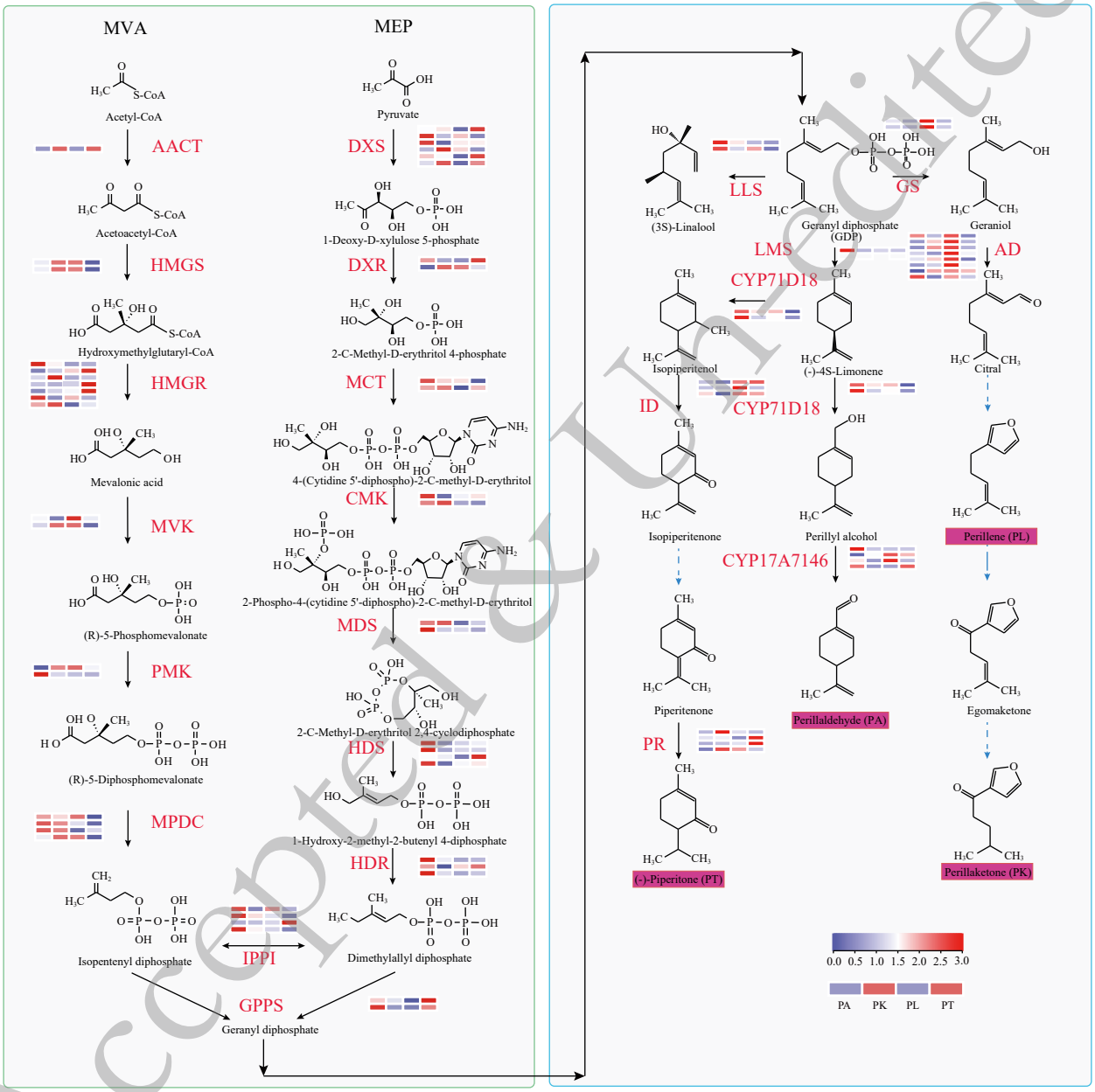

Fig. 3 Synthesis pathway and single thread synthesis pathway of *P. frutescens* isoprene. The MEP pathway:1-deoxy-D-xylulose 5-phosphate synthase (DXS); 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR); 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT); 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK); 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS/MECPS); (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS); 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR); The MVA pathway: acetyl-CoA C-acetyltransferase (AACT); hydroxymethylglutaryl-CoA synthase (HMGS); hydroxymethylglutaryl-CoA reductase (HMGR); mevalonate kinase (MVK); phosphomevalonate kinase (PMK); diphosphomevalonate decarboxylase (MPDC); isopentenyl-diphosphate Delta-isomerase (IPPI); farnesyl diphosphate synthase (FPPS); geranyl diphosphate synthase (GPPS); geranylgeranyl diphosphate synthase (GGPS); limonene synthase (LMS); isopiperitenol dehydrogenase (ID); pulegone reductase (PR); alcohol dehydrogenase (AD).

GPP in plastids (Fig. 3, Supplemental Table S10).

Moreover, the biosynthesis pathway of PA and PT has been reported in *Perilla* and *Mentha*, respectively[18–20,36]. Limonene is the common substrate for the synthesis of PA and PT. Two genes encoded limonene synthase (LMS) were identified in *Perilla*. For the biosynthesis of PA, limonene is catalyzed by CYP71D18 and CYP17A7146. Two genes encoded CYP71D18 were found and up-expressed in PA-type, while four paralogs encoded CYP17A7146 were identified and one of them showed upregulated expression in PA-type (Fig. 3, Supplemental Table S11). The biosynthesis of PT by isopiperitenol dehydrogenase (ID) and pulegone reductase (PR) were identified in *Perilla*, which possess three and four encoded genes and shows different expression in four cultivars, respectively. (Fig. 3, Supplemental Table S11. Moreover, geraniol is catalyzed by GL synthase (GLS). Two encode genes encoded GLS and showed upregulated expression in PL. GL were further catalyzed by alcohol dehydrogenase (AD) to produce citral, which is the precursor of PL and PK. Besides them, LL synthase (LLS) are also the common monoterpene compounds in *Perilla*. Two encode genes encoded LLS and up-expressed in PA-type were identified in *Perilla* (Figure 3).

Identification and characterization analysis of TPS family

The terpene synthases (TPS) use prenyl pyrophosphates as the substrate to synthesize terpenes, which are important for various chemotype formation in *Perilla*. In *Perilla*, totally 109 *TPS* family members were identified using HMM search. The putative PfTPS proteins ranged from 230 to 817 amino acids in length (Supplemental Table S12), with the exon number from 3 to 15 (Fig. 4b). All members contained N-terminal (PF01397) and C-terminal (PF03936) conserved domains of TPS genes (Fig. 4c), and RRX_8W domain existed in the N-terminal, while the typical DDXXD conserved domain, as well as the typical functional domain R_XR, existed in the C-terminal (Fig. 4d). The *PfTPS* genes displayed obviously different expression trend in the four chemotypes (Fig. 4e).

PfTPS family members were divided into five major subfamilies, including *TPS-a* (57 members), *TPS-b* (24 members), *TPS-c* (12 members), *TPS-e/f* (8, members), and *TPS-g* (8 members) (Fig. 4f). The number of *TPS* genes in *Perilla* (109) showed significant expansion compared to that of *Arabidopsis thaliana*[37], *Solanum lycopersicum*[38] and other lamiaceae speices, including *Mentha longifolia*[39], *Salvia miltiorrhiza*[39], *Ocimum tenuiflorum*[39] and *Lavandula angustifolialabiaceae*[40]. Among the *TPS* in *Perilla*, the *TPS-a* and *TPS-b* accounts for the 57.29% and 22.02% proportion as main expanded sub-families in *Perilla*. (Supplemental Table S13). In order to explore the evolutionary relationship of *TPS*, chromosome mapping and collinear block analysis were carried out. The *PfTPS* genes unevenly distributed on the 18 chromosomes. As tetraploid genome of *Perilla*, the distribution of allele genes in pairs is a normal phenomenon. There are 9 *PfTPS* genes were found on chromosome Chr10/11/12/13, 5/8 *PfTPS* in Chr5/12, and 4/6 *PfTPS* in Chr4/6, which showed obvious collinearity in *Perilla* genome (Fig. 4g). Further analysis the collinear relation between *Perilla* and *S. baicalensis*. The collinear block in *Perilla* 11 chromosomes correspond with 7 chromosomes in *S. baicalensis*. Universally, the *TPS* in *Perilla* showed tandem duplication, containing 45 indicating that there is obvious evolutionary relationship between *TPS* of *Perilla* and *S. baicalensis*. However, it is specially that SbChr09 has obvious chromosome blocks correspond with multiple chromosomes of *Perilla* (Fig. 4h).

Co-expression analysis and verification of functional *TPS* genes

To further mine the functional *TPS* genes in various chemotypes, the gene co-expression analysis was performed. Interestingly, *PfTPS18*, *PfTPS46*, *PfTPS47* and *PfTPS49* as significant core genes were identified. In order to present the significant relationship between those *TPSs*, the co-expression network was present the core *TPS* and the terpene biosynthesis genes and TFs, respectively. Firstly, *PfTPS18* as core genes were significant co-expression with 201 the terpene biosynthesis genes and TFs, including *GPPS*, *HMGS*, *HDR*, *AACT* and *ERF*, *MYB*, *NAC* etc (Fig. 5a). Similarly, the *PfTPS46*, *PfTPS47* also were co-expression with *IPPI*, *HDR*, *GPPS*, *FPPS*, *HMGS* and *MPDC* genes, which is important rate-limiting genes in MVA/MEP biosynthesis pathway (Fig. 5b,c). PfTPS49 were co-expression with other five *PfTPSs*, including *PfTPS15*, *PfTPS24*, *PfTPS38*, *PfTPS39* and *PfTPS63*, and associated with The *GPPS*, *HMGS* and other TFs (Fig. 5d).

Among them, 12 *PfTPS* genes and 2 MVA pathway genes were selected for expression verification using qRT-PCR. The significant coincident gene levels were identified in transcriptome sequencing and qRT-PCR (r>0.9). Those *PfTPSs* present general transcription in four cultivars, but showed high expression in one chemotype. Such as *PfTPS18*, *PfTPS*21, *PfTPS76* showed up-regulated expression in PA-types, especially *PfTPS49* present specific high expression level. *PfTPS46*, *PfTPS47*, *PfTPS62* showed up-regulated expression in PL-types. The expression levels of PfTPS87, PfTPS93 and PfTPS108 were similar in the four chemical types (Fig. 5e).

Functional characterization of *PfTPSs* from different chemotypes of *Perilla*

As *PfTPS18*, *PfTPS46*, *PfTPS47* and *PfTPS49* were the significant core genes according to the co-expression analysis, we selected these *PfTPSs* genes for further functional characterization. Due to the expression levels of the above *PfTPSs* in different chemotypes (Fig. 5e), we only successfully cloned genes with predominant expression level in specific chemotypes, such as the highest-expressed *PfTPS18* and specific-expressed *PfTPS49* in PA-type. Thus, the cloned *PfTPSs* genes were named with their chemotypes as follows: *PfTPS46-PL*, *PfTPS46-PK*, *PfTPS18-PA*, *PfTPS47-PA* and *PfTPS49-PA*, respectively. To predict the possible catalytic functions of these PfTPSs, phylogenic analysis was performed.

To identify the catalytic functions of the cloned *PfTPSs*, we ligated the CDSs of different TPS to the expression vectors and transformed them into *E. coli* to characterize the functions of *PfTPSs*. After cultivated for 3 days, the cultures were extracted by n-hexane and then the compounds were detected by GC-MS analysis. Strain with *PfTPS46-PL* produced one peak in GC-MS profile compared to the control group (strain with control vector) (Fig. 6a). The product was determined to be linalool by the comparison of the retention time in total ion chromatograms and the mass spectrum with authentic standard linalool (Fig. 6a and Supplemental Fig. S1a, S1b). As the signal peptide (SP) region in the N terminal of *PfTPS46*-PL might affect the catalytic activity of the enzyme inside *E. coli* cells, we removed this region in the CDS of *PfTPS46*-PL and explore its function using the truncated variant. The strains harboring the

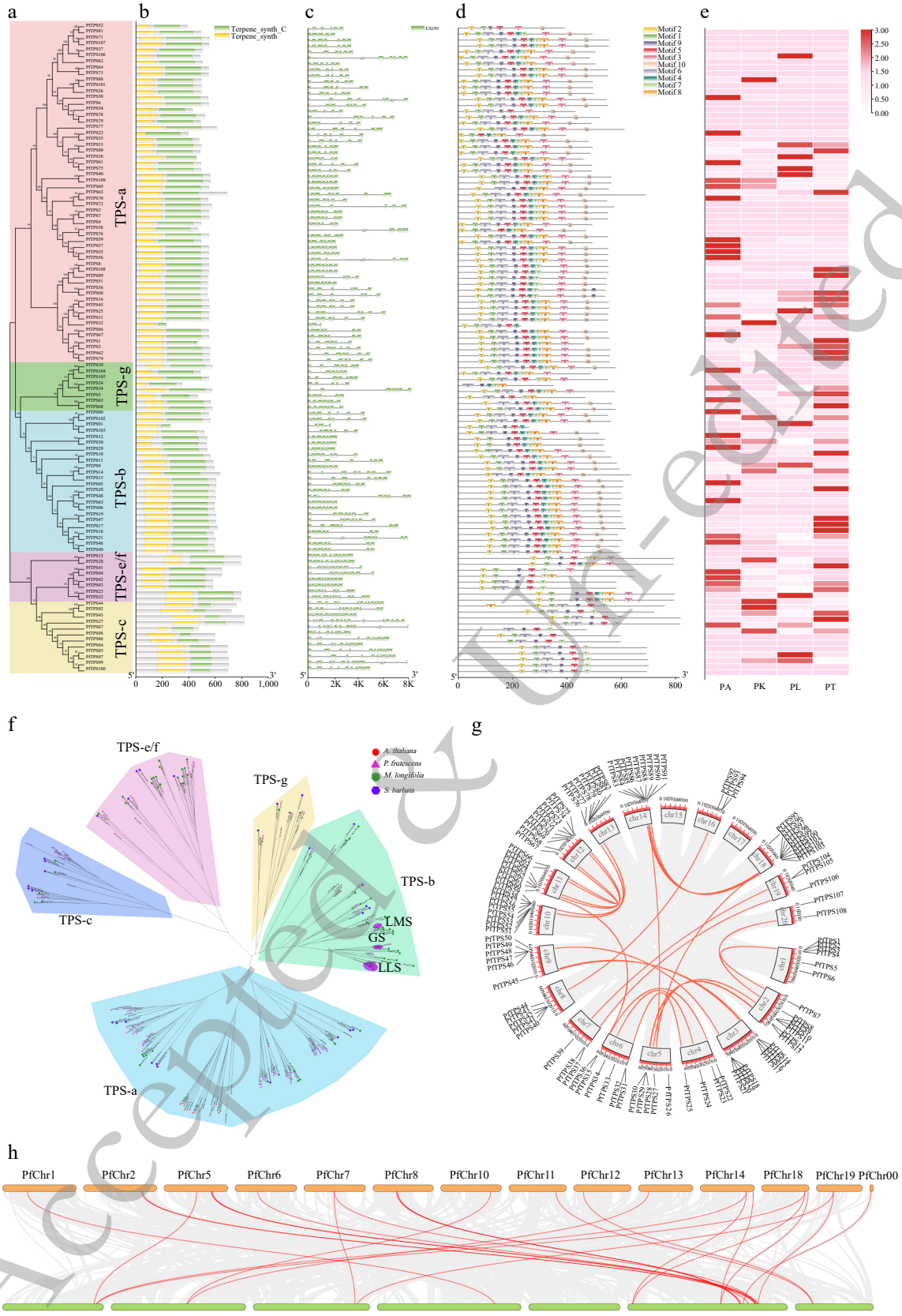

Fig. 4 The *PfTPS* Gene Family Characteristics in *P. frutescens*. A-E: the phylogenetic tree evolutionary tree, conservative domain, gene structure, protein motifs and expression heatmap of *PfTPS* genes; F: The Subfamily classification of PfTPS Family (LLS: linalool synthase; GS: geraniol synthase; LMS: limonene synthase); G: Chromosomal localization and collinearity analysis of *Perilla*; H: Collinearity analysis of *TPS* in *Perilla* and *S. baicalensis*.

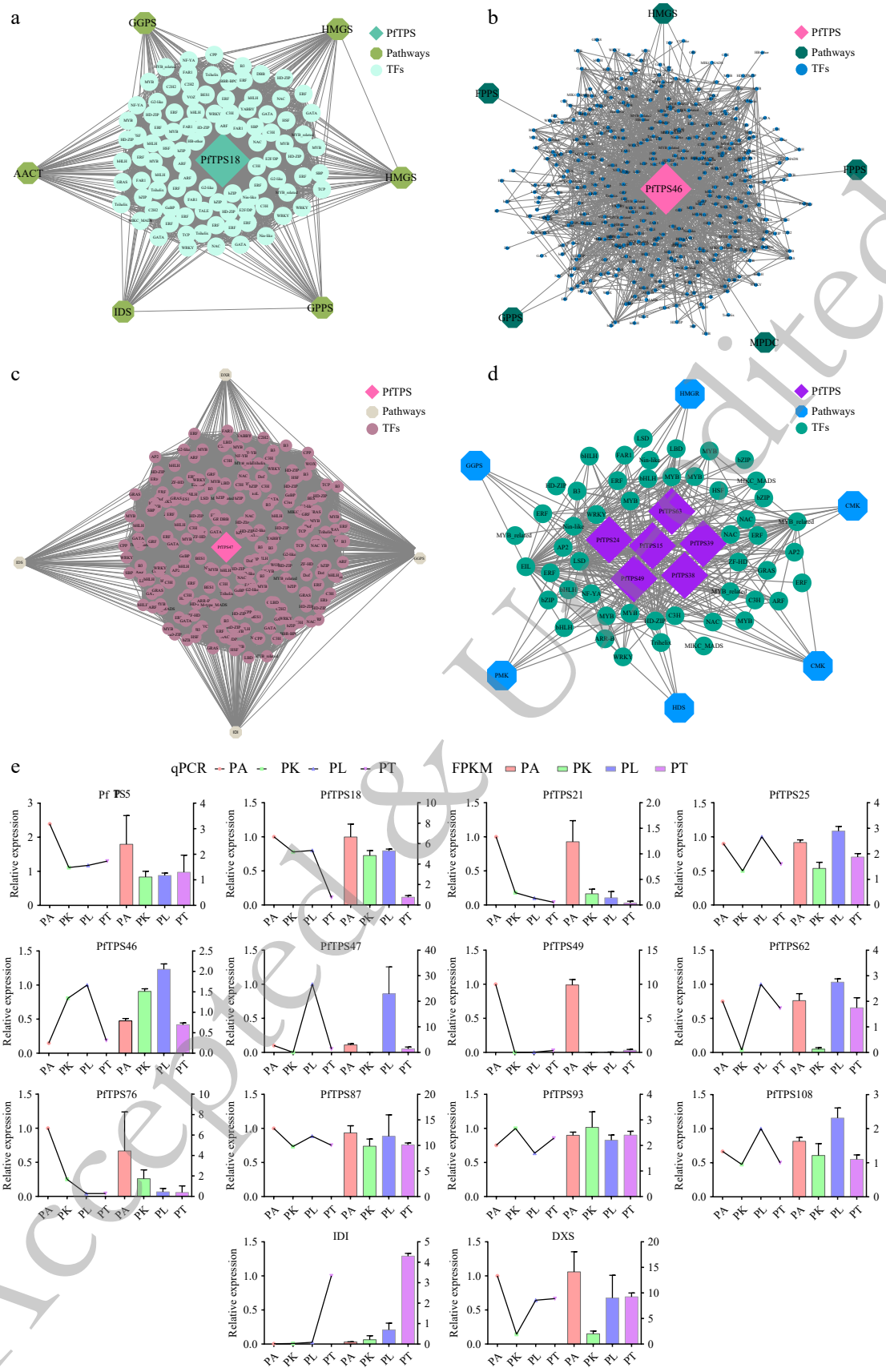

Fig. 5 The co-expression analysis and verification of *PfTPSs*. (a-d) The co-expression analysis with core *PfTPSs*, including *PfTPS18* (a), *PfTPS46* (b), *PfTPS47* (c), *PfTPS49* and other five *PfTPSs* (d). (e) The qRT-PCR verification of *PfTPS* genes and MVA pathway genes (qRT-PCR results (left, line) and transcription results (right, bar)).

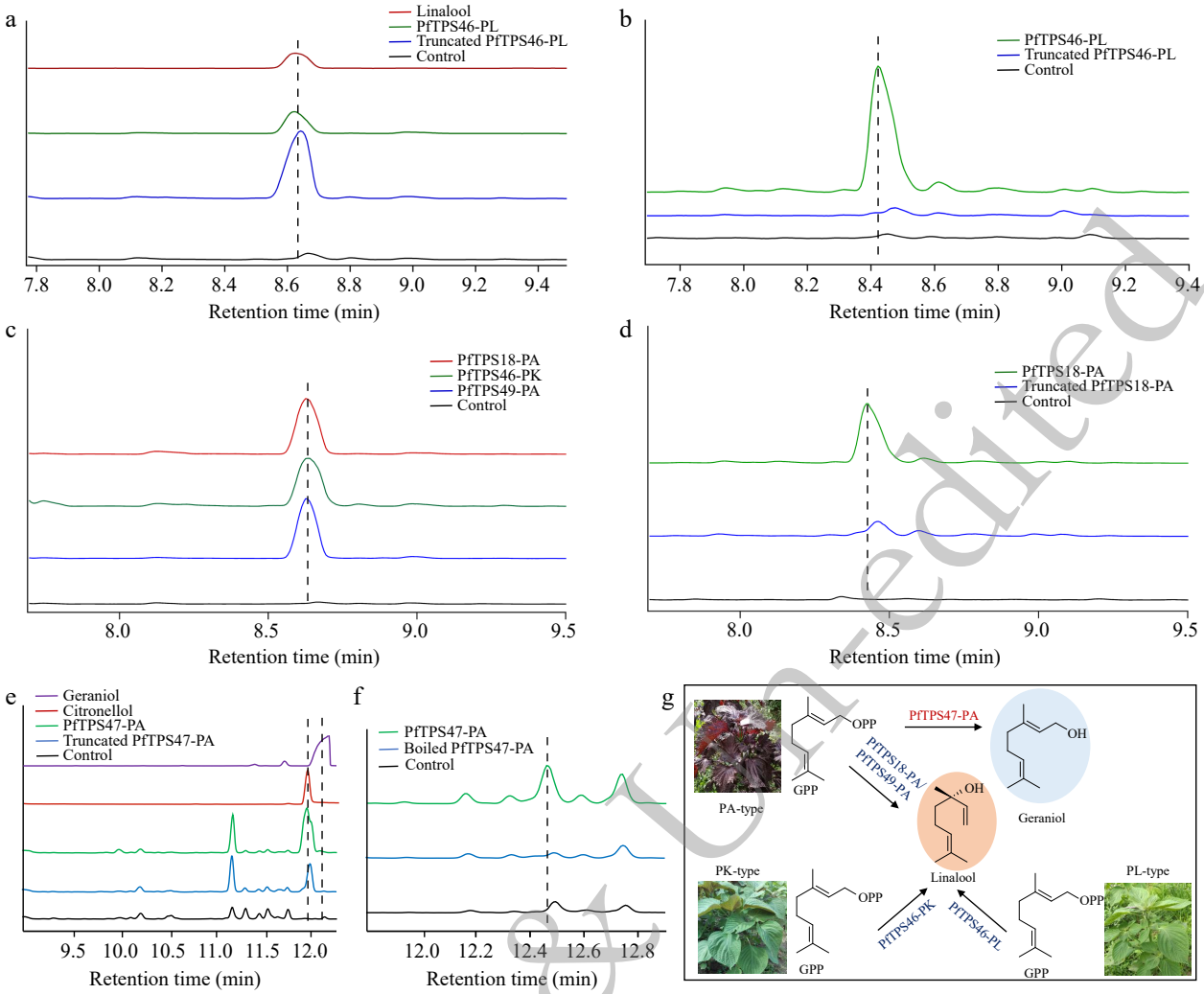

Fig. 6 Functional characterization of four *PfTPSs*. (a) Heterologous functional characterization of *PfTPS46*-PL and (b) the in *vitro* enzymatic reaction of *PfTPS46*-PL; (c) Heterologous functional characterization and (d) the in *vitro* enzymatic reaction of *PfTPS46*-PK, *PfTPS18*-PA and *PfTPS49*-PA; (e) Heterologous functional characterization and (f) the in *vitro* enzymatic reaction of *PfTPS47*-PA. (g) the catalytic model of four *PfTPSs*

truncated *PfTPS46*-PL brought the same linalool product in GC-MS analysis (Fig. 6a and Supplemental Fig. S1c). Next, to further confirm the catalytic function of PfTPS46-PL, we attempted to purify the *PfTPS46*-PL protein and characterized its function using in *vitro* enzymatic reaction. We failed to obtain the proteins of *PfTPS46*-PL due to its insolubility. Thus, the crude proteins of *PfTPS46*-PL were used with geranyl pyrophosphate (GPP) as precursor. Consistent with the result in heterologous expression system, the crude *PfTPS46*-PL protein also produced the sole product linalool (Fig. 6b). According, *PfTPS46*-PL is a linalool synthase.

Next, the functions of other candidate monoterpene synthases were characterized using the same strategy. The products of *PfTPS46*-PK, *PfTPS18*-PA and *PfTPS49*-PA were all found to be linalool in heterologous expression system, as well as the truncated enzymes (Fig. 6c; Supplemental Figs S2, S3). The crude proteins of *PfTPS18*-PA were selected as the representative for the in *vitro* enzymatic analysis. The crude proteins also produced the sole product linalool (Fig. 6d and Supplemental Fig. S4). The results indicated that these *PfTPSs* are linalool synthases.

For the function characterization of *PfTPS47*-PA, two products were detected in *PfTPS47*-PA and truncated *PfTPS47*-PA-harboring strains, with the major product citronellol and the minor product geraniol (Fig. 6e; Supplemental Figs S5, S6). However, the purified *PfTPS47*-PA protein catalyzed GPP to the sole product geraniol (Fig. 6f). Here, we speculated that some certain enzyme inside *E. coli* cells catalyzed geraniol, the product of *PfTPS47*-PA, to citronellol. The results showed that PfTPS47-PA was highly similar to geraniol synthase while other PfTPSs were assigned to the linalool synthase category. Collectively, we identified four linalool synthases and one geraniol synthase in different *Perilla* chemotypes (Fig. 6g).

Discussion

The essential oils of *Perilla* are well-recognized aromatic compound and possess the multiple pharmacological effect. They are also the valuable genetic materials of monoterpene biosynthesis and regulation for the multiple kinds of chemical types. In the study, there were four monoterpene chemotype

cultivars were selected. The important monoterpene biosynthesis pathway and important candidate TPSs was analyzed and verified using the transcriptome sequence and heterologous expression verification.

Terpene biosynthesis initiate from the MVA and MEP pathway in plants[16]. Compared with *Arabidopsis*, the number of encoded genes in the MVA and MEP pathways increased significantly in *Perilla*. The gene amplification could induce gene differentiation and affect the biosynthesis of terpenes[38]. The genes encoded HMGS, PMK and MDPC in MVA pathway, DXR, MCT, CMK and MDS in MEP pathway, and IPPI and FPPS in cross-flow pathway was found obvious expansion in perilla. Interestingly, most MEP pathway genes were up-regulated in the PA-type, which implied the high-efficiency biosynthesis in PA-type of perilla.

The various volatile oil components as chemical type of perilla has been early research. In early stage, the genetic basis for monoterpene chemical type in *Perilla* were verified using the artificial hybrids method. The chemical composition is controlled by a series of multiple alleles (G1, G2, g) and an independent pair of alleles (H, h)[41]. In our research, the109 *TPS* members in *Perilla* genomes were identified. They also showed obviously gene expansion. In especial, the expansion of *TPS-a* and *TPS-b*, reached 57.29% and 22.02% proportion, more than other majority of Lamiaceae plants. The expression and function of TPSs were also significantly differentiated. In past few years, more geraniol synthases, linalool synthases and limonene synthases have been acquired in perilla[18–20]. And the biosynthetic pathway of piperitenon in *Mentha longifolia* was reported. Based on genome-wide identified, we also explore more GS, LLS, LMS, and the PA and PT biosynthesis encoded genes. The expression trend of them are in accordance with volatile oil components. Such as LMS and GS were high expression in PA-types and LLS were high expression in PL-types. Based on co-expression analysis, the four *TPSs* as core genes in various chemical type. The high expression in PA-types and PL-types were selected for function verification.

The heterologous functional characterization and in *vitro* enzymatic reaction are two important methods for the functional characterization of TPSs. PfTPS18, PfTPS46, and PfTPS49 were characterized as linalool synthases and PfTPS47 was characterized as geraniol synthases, respectively. Linalool and geraniol are general compounds. The core genes in co-expression analysis were characterized as linalool and geraniol synthases. More research will be done to identified more chemotype related TPSs. Moreover, the genotype in different cultivar of a certain species is unique and widely used in the recognition of different cultivars in many plants (Fig. S7-S8)[42,43]. For example, the polymorphic variant of one sesquiterpene synthase, *VvTPS24*, in grape conferred the cultivar a different product in the chemotype, which was distinguishable from other grape cultivars (Fig. S9)[44]. However, the identified isozymes in different *Perilla* chemotypes, such as the linalool synthases, including *PfTPS46*-PL and *PfTPS46*-PK, showed no obvious variations in their amino acid sequences. As the constitutions of compounds in different cultivars are determined by the enzymes and their expressions, we guess that the diverse content of linalool in different cultivars is caused by the regulatory element.

Author contributions

Yang SM: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft. Chu HY: Investigation, Resources, Software, Writing – original draft. Wang YX: Data curation, Formal analysis, Investigation, Software, Validation. Lin GB: Formal analysis, Investigation. An TY: Conceptualization, Methodology, Supervision, Writing – review & editing. Shen Q: Funding acquisition, Conceptualization, Methodology, Supervision, Writing – review & editing.

Data availability

The datasets presented in this study are publicly available. The data can be found here: RNA-seq data are available via NGDC with accession (PRJCA021059).

Acknowledgments

This work was funded by the National Natural Science Foundation of China Grant (U22A20446) and the National Natural Science Foundation for Regional Fund (31860391).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (XXXXXX)

Dates

Received 2 February 2024; Accepted 24 June 2024; In press 9 July 2024

References

1. Yu H, Qiu JF, Ma LJ, Hu YJ, Li P, et al. 2017. Phytochemical and phytopharmacological review of *Perilla frutescens* L. (Labiatae), a traditional edible-medicinal herb in China. *Food and Chemical Toxicology* 108:375–91
2. Martinetti L, Ferrante A, Bassoli A, Borgonovo G, Tosca A, et al. 2012. Characterization of some qualitative traits in different perilla cultivars. *Acta Horticulturae* 939:301–308
3. Ahmed HM. 2018. Ethnomedicinal, phytochemical and pharmacological investigations of *Perilla frutescens* (L.) Britt. *Molecules* 24(1):102
4. Wu X, Dong S, Chen H, Guo M, Sun Z, et al. 2023. *Perilla frutescens*: A traditional medicine and food homologous plant. *Chinese Herbal Medicines* 15:369–75
5. Mungmai L, Preedalikit W, Pintha K, Tantipaiboonwong P, Aunsri N. 2020. Collagenase and Melanogenesis Inhibitory Effects of *Perilla Frutescens* Pomace Extract and Its Efficacy in Topical Cosmetic Formulations. *Cosmetics* 7:69
6. Erhunmwunsee F, Pan C, Yang K, Li Y, Liu M, et al. 2022. Recent development in biological activities and safety concerns of perillaldehyde from perilla plants: A review. *Critical Reviews in Food Science and Nutrition* 62:6328–40
7. Ito M, Honda G, Sydara K. 2008. *Perilla frutescens* var. *frutescens* in northern Laos. *Journal of Natural Medicines* 62:251–8
8. Baba M, Yamada K-i, Ito M. 2020. Cloning and Expression of a *Perilla frutescens* Cytochrome P450 Enzyme Catalyzing the Hydroxylation of Phenylpropenes. *Plants* 9:577
9. Michiho I, Mariko T, Gisho H. 1999. Chemical Composition of the Essential oil of *Perilla frutescens*. *Natural medicines* 53:32–6

10. Müller-Waldeck F, Sitzmann J, Schnitzler WH, Graßmann J. 2010. Determination of toxic perilla ketone, secondary plant metabolites and antioxidative capacity in five *Perilla frutescens* L. varieties. *Food and Chemical Toxicology* 48:264−70
11. Sato-Masumoto N, Ito M. 2014. Two types of alcohol dehydrogenase from Perilla can form citral and perillaldehyde. *Phytochemistry* 104:12−20
12. Masumoto N, Korin M, Ito M. 2010. Geraniol and linalool synthases from wild species of perilla. *Phytochemistry* 71:1068−75
13. Chen W, Viljoen AM. 2010. Geraniol — A review of a commercially important fragrance material. *South African Journal of Botany* 76:643−51
14. Dudai N, Segey D, Haykin-Frenkel D, Eshel A. 2006. Genetic Variation of Phenolic Compounds Content, Essential Oil Composition and Anti Oxidative Activity in Israel-Grown *Mentha longifolia* L. *Acta Horticulturae* 69−78
15. Lange BM, Srividya N. 2019. Enzymology of monoterpene functionalization in glandular trichomes. *Journal of Experimental Botany* 70:1095−108
16. Vranová E, Coman D, Gruissem W. 2013. Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. *Annual Review of Plant Biology* 64:665−700
17. Zebec Z, Wilkes J, Jervis AJ, Scrutton NS, Takano E, et al. 2016. Towards synthesis of monoterpenes and derivatives using synthetic biology. *Current Opinion in Chemical Biology* 34:37−43
18. Yuba A, Yazaki K, Tabata M, Honda G, Croteau R. 1996. cDNA Cloning, Characterization, and Functional Expression of 4S-(−)-Limonene Synthase from *Perilla frutescens*. *Archives of Biochemistry and Biophysics* 332:280−7
19. Mau CJD, Karp F, Ito M, Honda G, Croteau RB. 2010. A candidate cDNA clone for (−)-limonene-7-hydroxylase from *Perilla frutescens*. *Phytochemistry* 71:373−9
20. Fujiwara Y, Ito M. 2017. Molecular cloning and characterization of a *Perilla frutescens* cytochrome P450 enzyme that catalyzes the later steps of perillaldehyde biosynthesis. *Phytochemistry* 134:26−37
21. Zhou P, Shao Y, Jiang Z, Dang J, Qu C, et al. 2023. The revealing of a novel double bond reductase related to perilla ketone biosynthesis in *Perilla frutescens*. *BMC Plant Biology* 23:345
22. Zhang Y, Shen Q, Leng L, Zhang D, Chen S, et al. 2021. Incipient diploidization of the medicinal plant Perilla within 10, 000 years. *Nature Communications* 12:5508
23. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114−20
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15−21
25. Srinivasan KA, Virdee SK, McArthur AG. 2020. Strandedness during cDNA synthesis, the stranded parameter in htseq-count and analysis of RNA-Seq data. *Briefings in Functional Genomics* 19:339−42
26. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15:1−21
27. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular biology and evolution* 38:5825−9
28. Wu T, Hu E, Xu S, Chen M, Guo P, et al. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2:100141
29. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, et al. 2019. The Pfam protein families database in 2019. *Nucleic acids research* 47:D427−D32
30. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* 35:1547
31. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant* 13:1194−202
32. Wang Y, Li J, Paterson AH. 2013. MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. *Bioinformatics* 29:1458−60
33. Zhang T, Song C, Song L, Shang Z, Yang S, et al. 2017. RNA sequencing and coexpression analysis reveal key genes involved in α-linolenic acid biosynthesis in *Perilla frutescens* seed. *International journal of molecular sciences* 18:2433
34. Duan W, Shi-Mei Y, Zhi-Wei S, Jing X, De-Gang Z, et al. 2021. Genome-wide analysis of the fatty acid desaturase gene family reveals the key role of *PfFAD3* in α-linolenic acid biosynthesis in Perilla Seeds. *Frontiers in genetics* 12
35. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25:402−8
36. Mahmoud SS, Croteau RB. 2003. Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpene reductase. *Proceedings of the National Academy of Sciences of the United States of America* 100:14481−6
37. Parker MT, Zhong Y, Dai X, Wang S, Zhao P. 2014. Comparative genomic and transcriptomic analysis of terpene synthases in Arabidopsis and Medicago. *IET Systems Biology* 8:146−53
38. Zhou F, Pichersky E. 2020. The complete functional characterisation of the terpene synthase family in tomato. *New Phytologist* 226:1341−60
39. Chen Z, Vining KJ, Qi X, Yu X, Zheng Y, et al. 2021. Genome-wide analysis of terpene synthase gene family in *Mentha longifolia* and catalytic activity analysis of a single terpene synthase. *Genes* 12:518
40. Li J, Wang Y, Dong Y, Zhang W, Wang D, et al. 2021. The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpene biosynthesis. *Horticulture Research* 8:53
41. Tabata M. 2000. Genetics of Monoterpene Biosynthesis in Perilla Plants. *Plant Biotechnology* 17:273−80
42. Singh N, Singh B, Rai V, Sidhu S, Singh AK, et al. 2017. Evolutionary Insights Based on SNP Haplotypes of Red Pericarp, Grain Size and Starch Synthase Genes in Wild and Cultivated Rice. *Frontiers in Plant Science* 8
43. Weeden N, Lamb R. 1985. Identification of Apple Cultivars by Isozyme Phenotypes. *Journal of the American Society of Horticultural Science* 110:509−15
44. Drew DP, Andersen TB, Sweetman C, Møller BL, Ford C, et al. 2015. Two key polymorphisms in a newly discovered allele of the Vitis vinifera *TPS24* gene are responsible for the production of the rotundone precursor α-guaiene. *Journal of Experimental Botany* 67:799−808