

Comparative Transcriptome Analysis of Leaf, Stem, and Root Tissues of *Semiliquidambar Cathayensis* Hung T. Chang Reveals Candidate Genes Involved in Terpenoid Biosynthesis

xiaoming tian (✉ tianxiaoming1986@126.com)

hunan botanical garden <https://orcid.org/0000-0001-5591-182X>

Lihong Yan

hunan botanical garden

Liyuan Jiang

hunan botanical garden

Guangfeng Xiang

hunan botanical garden

Gaofei Li

hunan botanical garden

Jia Wu

Central South University

Research Article

Keywords: Semiliquidambar cathayensis Chang, transcriptome, terpenoids, tissue specificity

Posted Date: January 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1249220/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The roots, branches, leaves, bark, and nectar of *Semiliquidambar cathayensis* Hung T. Chang can treat rheumatoid arthritis, lumbar muscle strain, and other diseases. However, limited knowledge of the *S. cathayensis* Chang genome hampers further research on elucidating the pathways underlying active ingredient synthesis and regulation of its accumulation. In this study, we aimed to generate data to elucidate biosynthetic pathway of the active ingredient synthesis and related molecular mechanisms.

Methods

We conducted transcriptome sequencing of *S. cathayensis* Chang leaves, stem epidermis, stem xylem, root epidermis, and root xylem.

Results

We identified 88,582 unigenes, among which, 36,144 unigenes were annotated to the nr protein database, 21,981 to the Gene Ontology database, 11,565 to the Clusters of Orthologous Groups database, 24,209 to the Pfam database, 21,685 to the SWISS-PROT database, and 12,753 to the Kyoto Encyclopedia of Genes and Genomes (KEGG), with 5,072 unigenes common to all six databases. Of those annotated using the KEGG database, 187 unigenes were related to the terpenoid metabolism pathway, and expression analysis of the related genes indicated that the mevalonate and methylerythritol 4-phosphate pathways play different roles in terpenoid biosynthesis in different tissues of *S. cathayensis* Chang.

Conclusions

These findings greatly expand the gene resources of *S. cathayensis* Chang and provide basic data for the study of the biosynthetic pathways and molecular mechanisms of terpenoids.

1. Background

Semiliquidambar cathayensis Hung T. Chang, is a plant species belonging to the family Altingiaceae and the genus *Semiliquidambar* [1], is endemic to China and is scarce and scattered among the mountainous regions of South Jiangxi, North Guangxi, South Guizhou, Guangdong (including Hainan Island), South Hunan, and other areas. With a long history of medicinal use, *S. cathayensis* Chang roots, branches, leaves, bark, and nectar are used to dispel wind, remove dampness, relax muscles, and promote blood circulation. In China, the plant is also used to treat rheumatoid arthritis, lumbar muscle strain, postpartum wind paralysis, hemiplegia, bruises, and sprains [2, 3]. As it is relatively difficult to harvest its branches,

roots, and nectar, the leaves of this plant are typically boiled into medicinal solutions for medicated baths, rubs, or for oral ingestion [4].

Modern pharmacological studies have demonstrated that *S. cathayensis* Chang crude extracts exert analgesic, anti-inflammatory, and anti-hepatitis B effects and pharmacological activities, such as promoting blood circulation and removing blood stasis [3]. Li et al. [5] isolated 35 alkaloids, 12 flavonoids, nine fatty acid types, seven terpenoids, seven cyclic peptide types, five phenylpropanoids, and ten other ingredients from *S. cathayensis* Chang. Meanwhile, using silica gel and Sephadex LH-20 column chromatography, Lu et al. [6] separated ten compounds: 3-acetoxy oleanolic acid methyl ester, β -sitosterol, 3-acetoxyalkene phenolic acid, 2 α ,3 β -dihydroxy-20(29)-ene-lupin-28-acid, (24R)-5 α -stigmaster-3,6-dione, betulinic acid, stearic acid, 3-keto-oleanolic acid, arjunolic acid, and carotin. Zhou et al. [7] reported that the ethyl acetate extract exerts effective anti-inflammatory effects upon croton oil stimulation of mouse ear and identified nine active monomer compounds—oleanolic acid, 3-carbonyl oleanolic acid, 2 α ,3 β -dihydroxy oleanolic acid, 2 α ,3 β ,2 3-trihydroxy oleanolic acid, tannic acid-3,3'-di methyl ether, tannin-3,3',4-trimethyl ether, tannin-3,3'-dimethyl ether-4-O- β -D-xyloside, β -sitosterol, and stearic acid—via silica gel column chromatography, mass spectrometry, and nuclear magnetic separation. The main active ingredients of *S. cathayensis* Chang are secondary metabolites, with terpenoids being important components. However, because genome or transcriptome technology has not yet been applied for identifying the genes encoding active secondary metabolic components, the molecular biosynthesis mechanisms in *S. cathayensis* Chang remain unclear. Studies have been conducted on terpenoid biosynthesis in other traditional Chinese medicinal plants, including *Catharanthus roseus* [8], *Andrographis paniculata* [9], *Aconitum napellus* [10], *Panax ginseng* [11], and *Salvia miltiorrhiza* [12]. Recent research studies have focused on the identification, characterization, expression, and regulation of genes encoding enzymes that play important roles in terpenoid biosynthesis.

High-throughput transcriptome sequencing has played an increasingly important role in the elucidation of medicinal plant functional genes [13], metabolic pathways, molecular regulation mechanisms, and molecular markers [14]. Transcriptome analysis has allowed the exploration of tissue-specific gene expression patterns, clarifying secondary metabolite pathways and networks in medicinal plants [15]. Transcriptome sequencing applied to traditional Chinese medicinal plants, such as *Eucommia ulmoides* [16], *Magnolia officinalis* [17], *S. miltiorrhiza* (cinnabar) root [18], *Aquilaria sinensis* (baimuxiang) [19], *Glycyrrhiza uralensis* (ural) licorice [20], and *Blumea balsamifera* (ainaxiang) [21], has generated abundant genetic data for mining the genes involved in the key secondary metabolite synthesis pathways. In this study, we used Illumina sequencing technology to conduct comparative transcriptomic analysis on tissues from the leaves, stem epidermis, stem xylem, root epidermis, and root xylem of *S. cathayensis* Chang. Through the differential expression patterns of *S. cathayensis* Chang genes, the terpenoids metabolic mechanism could be clarified at the molecular level. Understanding the metabolic mechanisms will provide the foundation for further study on terpenoid biosynthesis, metabolic pathways, regulatory mechanisms, and resource utilization.

2. Results

2.1 High-throughput sequencing and assembly

Transcriptomic sequencing of *S. cathayensis* Chang leaf, stem epidermis, stem xylem, root epidermis, and root xylem tissue samples provided 67.33 Gb of clean data. The clean data obtained for each sample was approximately 7.88 Gb, with Q30 base percentages of 91.61% or higher, and was spliced and assembled using the Trinity software. The matching degree of the mapped reads, those that could be in the transcriptome library, all reached over 72%, indicating effective sequencing data assembly. The majority of the reads were thus assembled; the results are presented in Table 1.

Table 1
Summary of transcriptome sequencing data and assembly results

Sample	Clean Read	Total bases	GC/%	≥Q30/%	Mapped Reads	Mapped Ratio/%
Leaf	35,777,642	10,642,552,078	45.17%	92.14%	25,722,073	71.89%
Stem Epidermis	26,565,020	7,875,115,270	45.89%	91.61%	19,800,817	74.54%
Stem Xylem	31,439,795	9,304,370,786	45.88%	92.06%	23,403,292	74.44%
Root Epidermis	33,885,906	10,059,609,728	45.83%	92.05%	25,671,972	75.76%
Root Xylem	26,761,519	7,936,758,536	46.32%	92.33%	20,166,909	75.36%

The assembly produced 170,504 transcripts with an average length of 1,088 base pairs (bp) and an average N50 value of 1,692 bp, and 88,582 unigenes—the longest transcript of each gene—with an average length of 1,399 bp and an average N50 value of 848 bp (Table 2). These data validated the sequencing and assembly of *S. cathayensis* Chang genes for analysis of the related gene functions.

Table 2
Assembly result statistics table

Length Range	Transcript	Unigene
200-300	29,484(17.29%)	24,451(27.60%)
300-500	30,694(18.00%)	19,981(22.56%)
500-1000	43,716(25.64%)	21,432(24.19%)
1000-2000	41,271(24.21%)	13,969(15.77%)
≥2000	25,339(14.86%)	8,749(9.88%)
Total Number	170,504	88,582
Total Length	185,519,757	75,185,954
N50 Length	1,692	1,399
Mean Length	1088.07	848.77
<p>Note: Length Range: indicates the different length intervals of Unigenes; the number in the table indicates the number of Unigenes in the corresponding interval, the percentage in parentheses indicates the proportion of Unigenes in the corresponding length interval; Total Number: indicates the total number of Unigenes assembled; Total Length: indicates the total length of the assembled Unigenes; N50 Length: indicates the length of Unigenes N50; Mean Length: indicates the average length of Unigenes.</p>		

2.2 Functional annotation of Unigenes

The unigene sequence alignment and function annotation in Table 3 displays the 36,129 (40.79%) unigenes annotated from the National Center for Biotechnology Information (NCBI) nr protein database; 21,977 (24.81%) from the Gene Ontology (GO) database; 16,014 (18.08%) from the NCBI Clusters of Orthologous Groups (COG) database; 24,209 (27.33%) from the Pfam database; 21685 (24.48%) from the SWISS-PROT sequence database; and 12,755 (14.40%) from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Among these annotated sequences, 5,072 unigenes were annotated across all six databases, accounting for 5.72% of the total.

Table 3
The database distribution of unigenes for functional annotation

Databases name	Total number	Percentage of unigenes (%)	300<=length<1000	length>=1000
Nr_Annotation	36144	40.80	12239	16969
GO_Annotation	21981	24.81	7105	10659
COG_Annotation	11565	13.06	3097	5997
Pfam_Annotation	24209	27.33	7076	13613
Swissprot_Annotation	21685	24.48	6999	11758
KEGG_Annotation	12753	14.40	4202	6135
All_Annotated	37604	42.45	12703	17144

Analysis of the species distribution of all unigenes using the nr library (Figure 1) revealed sequences homologous to those of *Vitis vinifera*, *Malus domestica*, *Theobroma cocoa*, *Nelumbo nucifera*, *Nelumbo nucifera*, and *Jatropha curcas*. The reported genes from these closely related species provided a reference point for the unigene annotation.

2.3 Functional Classification of *S. cathayensis* Chang Unigenes in the COG Database

The COG database primarily classifies gene products into orthologs. Comparison of the 88,582 *S. cathayensis* Chang unigenes and COG data yielded 16,014 corresponding sequences, which were divided into 24 categories (Figure 2). *S. cathayensis* Chang unigenes corresponded to a relatively comprehensive inventory of COG functional categories, encompassing most pathways. The majority—2,935 (18.33%)—of the unigenes were predicted for general function only; 1,320 (8.24%) for replication, recombination, and repair; 1,308 (8.07%) for translation, ribosomal structure, and biogenesis; 1,267 (7.91%) for transcription; 1,224 (7.64%) for posttranslational modification, protein turnover, and chaperones; 1,101 (6.88%) for signal transduction mechanisms; 14 (< 0.01%) for cell movement; and 1 for nuclear structure.

2.4 Functional Classification of *S. cathayensis* Chang Unigenes in the GO Database

To understand the distribution of *S. cathayensis* Chang gene functions at a macro level, the unigenes were classified using the GO database, which categorizes genes by genetic molecular function, cellular component, and biological process. The resulting GO annotation classification yielded 21,981 sequences,

24.81% of all the unigenes, which were divided into 17 categories, based on their position in relation to cellular components, with the most abundant being 8,885 cell, 8,885 cell components, 6,534 organelles, and 4,871 cell membranes. The unigenes were also divided by molecular function into 16 categories, the two most abundant being 12,671 catalytic-related proteins and 11,615 binding proteins. Based on the associated biological processes, the *S. cathayensis* Chang unigenes were divided into 20 categories; the most abundant categories included 15,865 unigenes in the metabolic process category, 12,775 in the cellular process category, and 10,634 in the single biological process category, as shown in Figure 3.

2.5 Functional Annotation of *S. cathayensis* Chang Unigenes in the KEGG Database

The KEGG database annotated 12,753 (14.40%) of the *S. cathayensis* Chang unigenes, which were classified by KEGG biochemical pathways into five groups, which were Cellular Processes, Environmental Information Processing, Genetic Information Processing, Metabolism, and Organismal Systems. KEGG enrichment analysis revealed that the annotated unigenes participated in 19 metabolic pathway groups, including 637 related to the other secondary metabolites and 470 to metabolic biosynthesis of terpenoids and polyketides (Figure 4). A comprehensive analysis of the secondary metabolite pathways revealed that 217 unigenes were involved in phenylpropanoid biosynthesis; 46 in flavonoid biosynthesis; seven in anthocyanin biosynthesis; and 187 were terpenoid biosynthesis (Table 4).

Table 4
Biosynthetic pathways of secondary metabolites for *Semiliquidambar cathayensis*
Chang

Metabolic pathway	pathway_id	Gene_number
Ubiquinone and other terpenoid-quinone biosynthesis	ko00130	53
Terpenoid backbone biosynthesis	ko00900	69
Monoterpenoid biosynthesis	ko00902	10
Diterpenoid biosynthesis	ko00904	25
Sesquiterpenoid and triterpenoid biosynthesis	ko00909	30
Phenylpropanoid biosynthesis	ko00940	217
Flavonoid biosynthesis	ko00941	46
Anthocyanin biosynthesis	ko00942	7
Isoflavonoid biosynthesis	ko00943	2
Flavone and flavonol biosynthesis	ko00944	3
Stilbenoid, diarylheptanoid and gingerol biosynthesis	ko00945	26
Isoquinoline alkaloid biosynthesis	ko00950	30

2.6 Differential Expression and Annotation of Genes in Different *S. cathayensis* Chang Tissues

Gene expression analysis for *S. cathayensis* Chang leaf, stem epidermis, stem xylem, root epidermis, and root xylem tissues was conducted based on comparisons of their transcriptome sequencing, assembly, and annotation (Figure 5-6). As shown in Table 5, compared to the gene expression in the stem epidermis (T2), expression levels of more unigenes were downregulated in the stem xylem epidermis (T3), root epidermis tissue (T4), and leaf tissue (T1). Meanwhile, compared to the gene expression in the stem epidermis(T2), expression levels of more genes were upregulated in the stem xylem tissue (T3), root epidermis tissue (T4), and root xylem tissue (T5). Lastly, compared to the gene expression in the root epidermis (T4), expression levels of more genes were upregulated in the root xylem tissues (T5).

Table 5
Number of differentially expressed genes with functional annotation
in *Semiliquidambar cathayensis* Chang

DEG_Set	COG	GO	KEGG	Pfam	Swiss-Prot	Nr
T1vsT2	1119	2226	1003	2770	2557	3563
T1vsT3	1473	2785	1268	3358	3134	4414
T1vsT4	1404	2710	1254	3307	3016	4247
T1vsT5	1341	2661	1189	3242	3045	4241
T2vsT3	805	1791	808	2136	2082	2816
T2vsT4	466	1109	515	1383	1300	1802
T2vsT5	880	1943	884	2299	2227	3009
T3vsT4	668	1513	687	1910	1815	2475
T3vsT5	275	682	312	851	830	1097
T4vsT5	657	1416	615	1779	1672	2313

2.7 Pathway Annotation Analysis of Unigenes in Different *S. cathayensis* Chang Tissues

To investigate the metabolic pathways associated with the identified unigenes, pathway enrichment analysis was performed using the KEGG database. In the comparison of leaf (T1) vs. stem epidermis (T2), photosynthesis; biosynthesis of phenylpropanoids, flavonoids, sesquiterpenoids, and triterpenoids; metabolism of tryptophan and phenylalanine; degradation of aromatic compounds; and other pathways were enriched in the genes. A comparison of the unigene pathway enrichment analysis of different tissues revealed significant differences in the data of the four most differentiated pathways in each combination. As shown in Table 6, the two most differentiated pathways of leaf (T1) vs. stem epidermis (T2), stem xylem (T3), root epidermis (T4), and root xylem (T5) were phenylpropanoid biosynthesis and photosynthesis; the three most differentiated pathways of stem epidermis (T2) vs. stem xylem (T3), root epidermis (T4), root xylem (T5) were photosynthesis, and phenylpropanoid and flavonoid biosynthesis; the three most differentiated pathways of stem xylem (T3) vs. root epidermis (T4) and root xylem (T5), were phenylpropanoid and flavonoid biosynthesis and phenylalanine metabolism; and the most differentiated pathways between root epidermis (T4) and root xylem (T5) were phenylpropanoid, stilbenoid, diarylheptanoid, and gingerol biosynthesis; and starch, sucrose, and phenylalanine metabolism. These results indicate that the secondary metabolites may be distributed differently in various tissues.

Table 6

KEGG enrichment analysis of Unigenes in different tissues of *Semiliquidambar cathayensis* Chang

Type	KEGG Pathway			
	1	2	3	4
T1vsT2	Phenylpropanoid biosynthesis	Photosynthesis	Flavonoid biosynthesis	Sesquiterpenoid and triterpenoid biosynthesis
T1vsT3	Phenylpropanoid biosynthesis	Photosynthesis	Phenylalanine metabolism	Glyoxylate and dicarboxylate metabolism
T1vsT4	Phenylpropanoid biosynthesis	Photosynthesis	Pentose and glucuronate interconversions	Monoterpenoid biosynthesis
T1vsT5	Phenylpropanoid biosynthesis	Photosynthesis	Phenylalanine metabolism	Glyoxylate and dicarboxylate metabolism
T2vsT3	Photosynthesis	Flavonoid biosynthesis	Phenylpropanoid biosynthesis	Phenylalanine metabolism
T2vsT4	Photosynthesis	Phenylpropanoid biosynthesis	Flavonoid biosynthesis	Plant hormone signal transduction
T2vsT5	Photosynthesis	Phenylpropanoid biosynthesis	Phenylalanine metabolism	Plant hormone signal transduction
T3vsT4	Phenylpropanoid biosynthesis	Flavonoid biosynthesis	Phenylalanine metabolism	Stilbenoid, diarylheptanoid and gingerol biosynthesis
T3vsT5	Phenylpropanoid biosynthesis	Phenylalanine metabolism	Flavonoid biosynthesis	Diterpenoid biosynthesis
T4vsT5	Phenylpropanoid biosynthesis	Starch and sucrose metabolism	Phenylalanine metabolism	Stilbenoid, diarylheptanoid and gingerol biosynthesis

Note: The first column represents the sample combination; The first sample is the control group; The latter is the experimental group

2.8 Prediction and analysis of terpenoid synthetic metabolism

To explore terpenoid biosynthesis and metabolism, we identified four pathways involved in terpenoid biosynthesis: terpenoid skeleton biosynthesis (KEGG entry ko00900, 69 unigenes), diterpenoid biosynthesis (ko00904, 25 unigenes), ubiquinone and other terpene quinone biosynthesis pathway (ko00130, 53 unigenes), and sesquiterpene and triterpene biosynthesis (ko00909, 30 unigenes), along with 177 associated sequences with clear functional annotations, most which were involved throughout the biosynthetic pathways. Using *S. cathayensis* Chang KEGG function annotations, 27 genes and enzymes in nine categories were identified as possibly associated with terpenoid synthesis (Table 7). The

terpenoid biosynthesis-related gene expression levels in different tissues were analyzed by comparing reads per kilobase of transcript per million mapped reads (RPKM); the results are presented in Figure 6.

Table 7

Differentially expressed genes related to triterpene and phenolic acids biosynthesis of *Semiliquidambar cathayensis* Chang

Abbreviation	Function enzymes	Unigene ID	Number of Unigene
HMGR	3-hydroxy-3-methyl glutaryl coenzyme A reductase	c107480 c94998	2
IDI	isopentenyl diphosphate isomerase	c80516	1
FPPS	farnesylpyrophosphate synthase	c101256	1
SQS	squalene synthase	c99172 c109863 c132125	3
DXS	1-deoxy-D-xylulose-5-phosphate synthase	c87532 c86485 c94239 c96299 c102448	5
DXR	1-Deoxy-D-xylulose 5-phosphate reductoisomerase	c125318 c99067 c114886	3
HDR	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	c85746 c68258	2
GPPS	Geranyl Diphosphate Synthase	c95462 c88562 c127934 c87983 c93413 c132377 c90089 c50530 c149126	9
CPS	copalyl diphosphate synthase	c97047	1

Two HMGR sequences, c94998 and c107480, were identified in the *S. cathayensis* Chang transcriptome, with the highest expression of the former detected in the leaves, followed by the root xylem, and then the root epidermis, whereas that of the latter was detected in the root xylem. The five DXS sequences in the *S. cathayensis* Chang transcriptome exhibited different expression patterns; higher transcription levels of c94239 were detected in the leaves, whereas those of the remaining four unigenes, c87532, c86485, c96299, and c102448, were highest in the root epidermis, followed by the stem epidermis. The two *S. cathayensis* Chang HDR sequences exhibited the following expression patterns: c85764 expression levels in the leaves were 3.87, 8.49, 3.96, and 5.02 times higher than those in the stem epidermis, stem xylem, root epidermis, and root xylem, respectively, and c68258 transcription levels were low in all tissues.

Among the nine GPPS sequences in the *S. cathayensis* Chang transcriptome, the expression levels of c95462, c87983, c93413, c90089, and c50530 in leaf tissues were three to eight times higher than in other plant tissues. Among the remaining four GPPS sequences, c88562 exhibited higher expression levels in stem epidermis and root epidermis tissues, c127934 and 132377 in root xylem tissue, and c149126 in stem xylem tissue.

2.9 Real-time PCR validation of candidate genes involved in Terpene Biosynthesis

To further verify the high transcription levels of terpene biosynthesis-related genes in the transcriptome database, quantitative reverse transcription-PCR (qRT-PCR) was used to verify the expression of certain *S. cathayensis* Chang genes. The relative expression levels of each candidate gene in different tissues of *S. cathayensis* Chang obtained using qRT-PCR were consistent with the Illumina data (Figure 7), with similar expression fold changes, suggesting that the RNA-seq data were reliable.

3. Discussion

Several terpenoids with different structures and functions play important roles in plant secondary metabolism and are synthesized from the common structural precursors IPP and DMAPP. In plants, IPP and DMAPP are synthesized via two distinct pathways, namely, the mevalonate (MVA) pathway in the cytoplasm and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway in the plastid.

As the first rate-limiting enzyme in MVA pathway, 3-hydroxy-3-methylglutaryl-coenzyme (HMGR) catalyzes the irreversible formation of MVA from HMG-CoA[22], which is also an important regulatory point in the pathway. Two HMGR genes, *hmg1* and *hmg2*, are found in the *A. thaliana* genome, with the former expressed in all tissues, whereas the latter is specifically expressed in the flowers and meristems [23]. Zhang et al. demonstrated that, in *Pogostemon cablin*, roots showed the highest expression levels of *PcHMGR1*. These findings demonstrate the organ specificity of *HMGR* expression, and similar *HMGR* expression patterns were observed in *S. cathayensis* Chang and *P. cablin*. Farnesyl pyrophosphate (FPP) synthase (FPPS) and squalene (SQ) synthase (farnesyl-diphosphate farnesyltransferase, SQS) are two key enzymes downstream of the MVA pathway[24-25]; in *S. cathayensis* Chang, the *FPPS* gene is overexpressed in all tissues, with the highest expression in the leaves. SQS catalyzes the condensation of the two FPP molecules to produce SQ, a common precursor for the biosynthesis of triterpenes and sterols [26]. The SQS gene has been cloned from *Siraitia grosvenorii* (luohan guo) [27], *Salvia divinorum* [28], *Eriobotrya japonica* (loquat) [29], *Helichrysum petiolare* (licorice) [30], *Antrodia cinnamomea* [31], *Panax quinquefolius* (American ginseng) [32], and other medicinal plants. The quantity and functional expression of SQS can vary in different medicinal plants, as demonstrated by the varied expression of the three SQS genes, *PgSQS1*, 2 and 3. In *P. ginseng*, *PgSQS1* is overexpressed in all organs, whereas *PgSQS2* and *PgSQS3* show organ-specific expression [33]; in addition, *Panax notoginseng* [34] and *Artemisia annua* [35] produce only one functional squalene synthase. The three SQS sequences in the *S. cathayensis* Chang transcriptome revealed their highest expression levels in leaves, with c109863 and

c132125, exclusively expressed in leaves. Thus, additional studies investigating the role of SQS sequences in triterpene and sterol synthesis may provide important insights.

The first and second rate-limiting enzymes of the MEP pathway, 1-deoxy-D-xylulose 5-phosphate synthase (DXS) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), respectively, regulate the synthesis of essential substances such as monoterpenes, diterpene aroma compounds, and carotenoids [36]. DXR overexpression in *Mentha aquatica* x *Mentha spicata* (peppermint) [37], and other transgenic plants, promotes an increase in monoterpenoid content; enhanced DXR expression in the hairy roots of *S. miltiorrhiza* effectively regulates or increases the synthesis of terpenoids [38]. The three *S. cathayensis* Chang DXR sequences exhibited the following expression patterns: c125318 and c114886 transcription levels were low in all tissues, whereas expression of c99067 was high, with the expression levels in the stem epidermis and root epidermis being 4.53-fold and 2.31-fold higher, respectively, than those in the leaves, indicating that the stems and roots may be the main monoterpenoids synthesis sites. The terminal active enzyme in the MEP pathway, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase (HDR), provides the precursor terpenoids, and alterations in HDR expression directly affect the production of terpenoid metabolites. HDR overexpression in *Ginkgo biloba* produces a 2.5-fold increase in total ginkgolide content compared with wild type expression [43], and HDR upregulation in *Solanum lycopersicum* (tomato) and *A. thaliana* leads to an increase in carotenoid content [39].

S. cathayensis Chang transcriptome sequence c80516 closely resembles the gene encoding isopentenyl pyrophosphate isomerase (IDI/IPI), an enzyme which restricts the isomerization of terpenoid precursors IPP and DMAPP. Real-time PCR analysis detected the expression of c80516 in all organs of *S. cathayensis* Chang, particularly in the leaves, and its transcription pattern was consistent with that of the *CriDI1* gene in *C. roseus* [40], the *SmIPI1* gene in *S. miltiorrhiza* [41], and the *IDI* gene in *A. napellus* [42].

Geranylpyrophosphate (GPP) synthase (GPPS) catalyzes the formation of GPP from IPP and DMAPP and is present in only a few plant tissues that synthesize monoterpenes. Immune response analysis of GPPS cDNA from *A. thaliana* cloned by Bouvier et al. demonstrated that GPPS is present in chloroplasts and secretory gland cells that do not contain chlorophyll [43]. Li et al. revealed that GPPS may target both the plastid and the cytoplasm. Collectively, these findings indicate that *S. cathayensis* Chang GPPS performs different functions in different organs.

Carbamoyl phosphate synthase (CPS) is a key member of the diterpene synthase family. In *S. miltiorrhiza*, CPS and kaurene synthase-like protein gradually catalyze the reaction mediated by geranylgeranyl diphosphate synthase to produce indanthrone diene [44], similar to *A. paniculate* CPS [45]. CPS in the *A. napellus* transcriptome is preferentially expressed in leaves [10]. Similar to the expression pattern of CPS in *A. paniculata* and *A. napellus*, in *S. cathayensis* Chang, the highest transcriptional activity of CPS was detected in the leaves, with lower activity levels detected in the stem epidermis and root epidermis, and no activity in the stem xylem and root xylem.

4. Conclusions

Although the genes and enzymes involved in the MVA and MEP pathways are different, both the pathways are connected via a metabolic network, allowing metabolites to interchange on the plastid membrane. The contribution of the MVA and MEP pathways to the synthesis of terpenoids remains unclear. This study demonstrated that the expression of MVA and MEP pathway-related genes can be detected in different tissues of *S. cathayensis* Chang, and that different genes exhibit varied expression levels in different tissues and display tissue-specific expression patterns. The expression levels of MVA pathway-related genes, such as *HMGR*, *FPPS*, and *SQS*, were significantly higher in leaves than in other tissues. *SQS* functions at the branching point of the triterpene and sterol synthesis pathway and determines the direction of FPP flow, thus playing an important regulatory role as a key enzyme in the pathway. Certain genes of the MEP pathway, such as those encoding *HDR* and *CPS*, exhibited similar expression patterns, with the highest expression levels in leaves, whereas other genes, such as those encoding *DXR* and *DXS*, exhibited higher expression levels in the stems and roots. Nine *GPPS* unigenes exhibited different expression patterns in different organizations; the relative expression of *FPPS* was significantly higher than *GPPS*. We speculate that the dominant role of the MVA and MEP pathways in terpenoid synthesis varies among different *S. cathayensis* Chang tissues. The outcomes of this study provide a theoretical basis for further understanding of the regulation mechanism of terpenoid biosynthesis.

5. Methods

5.1 Plant Material

A three-year-old *S. cathayensis* Chang plant was collected from the Rare Plant Classification Garden of the Hunan Botanical Garden (Hunan, China), washed, and divided into leaf (T1), stem epidermis (T2), stem xylem (T3), root epidermis (T4), and root xylem (T5) samples. The samples were placed on a filter paper and freeze-dried. The samples were stored in an ultra-low temperature refrigerator at -80°C for transcriptome sequencing.

5.2 RNA Extraction

Total RNA was extracted from 1 g of each sample using Trizol Reagent (Thermo-Fisher, Waltham, MA, USA) according to the manufacturer's instructions. NanoDrop (Thermo-Fisher) and Agilent 2100 (Agilent, Santa Clara, CA, USA) were used to determine RNA purity (OD_{260/280} ratio) and band integrity, respectively.

5.3 Library construction and Transcriptome sequencing

Magnetic beads coupled with an oligo(dT) primer (Thermo-Fisher) were used to enrich the mRNA, and fragmentation buffer was used to randomly interrupt the mRNA. The mRNA was then used as a template with random six-base primers to synthesize the first complementary DNA (cDNA) strand. Buffer, deoxyribonucleotide triphosphates, ribonuclease H, and DNA polymerase I were then added to synthesize the second cDNA strand. Purify cDNA with AMPure XP beads; after the purified double-stranded cDNA

undergoes end-repair, add a tail, and then connect to the sequencing adapter, AMPure XP beads were used for fragment size selection. Finally, PCR enrichment of the cDNA library was performed.

Qubit 2.0 was used for preliminary quantification of the constructed library. The library was diluted to 2 ng μ L⁻¹, and 1.8% agarose gel electrophoresis was used to detect the insert size of the library. Q-PCR quantification was then performed with an effective library concentration of > 2 nmol L⁻¹. Qualified libraries were clustered on an Illumina cbot and sequenced using Illumina HiSeqTM 4000 (Illumina, San Diego, CA, USA).

5.4 Sequence splicing and assembly

The Illumina platform was used to convert the sequenced image data into the corresponding nucleotide sequence data. Quality assessment and credibility analysis on the original sequence files were generated, and low-quality and uncertain sequences were removed during the sequencing process ($Q < 20$). Based on *S. cathayensis* Chang no-reference transcriptome information, the Trinity software was used to mix and assemble sample data, and contigs were assembled by overlapping DNA sequences and aggregated based on sequence paired-end information and contig similarities. Constraint-based local assembly and selection of splice variants (CLASS) was then performed to yield the unigenes.

5.5 Unigene annotation and classification

NCBI Basic Local Alignment Search Tool (BLAST) software was used to compare the unigene sequences with the nr, GO, COG, Pfam, Swiss-Prot, and KEGG six databases to generate annotation data on gene function and classification.

5.6 DEG Analysis

The comparison software Bowtie was used to evaluate the reads obtained from sequencing each sample with the unigene library, and the RPKM value was used to reflect the expression levels of the corresponding unigenes. False discovery rate was used to analyze unigene expression. In the differential gene screening, $FDR \leq 0.01$ and fold change ≥ 2 were used to define differential unigenes. KEGG pathway enrichment analysis was used to determine the most important biochemical metabolic and signal transduction pathways associated with the differentially expressed genes.

5.7 Verification of gene expression via qRT-PCR

qRT-PCR was used to quantify the expression of genes involved in terpenoid biosynthesis in the transcriptome data. Actin was used as an internal reference gene to select candidate genes, call up their nucleotide sequences, and design real-time PCR primers; Supplementary Table 1 provides gene information and shows the primer sequences used. SsoFastTM Eva Green Supermix (Bio-Rad, Hercules, CA, USA) and the Bio-Rad iCycler MyiQ Real-time PCR Systems platform were used to perform real-time PCR experiments, according to the manufacturer's instructions. The results were calculated using the $2^{-\Delta\Delta C_t}$ method, and the experiment was performed in triplicate.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of Data Material

The data presented in this study are available upon request from the corresponding author. The RNA-seq Illumina paired-end reads of the transcriptome for this study have been submitted to NCBI BioProject PRJNA761214.

Competing interests: The authors declare no conflict of interest

Funding: This research was funded by Hunan Forestry Science and Technology Innovation Plan Project, grant number XLK202106-2.

Author Contributions: Conceptualization, XT; methodology, LY and LJ; software, JW.; validation, JW, GX and GL.; data curation, JW; writing-original draft preparation, XT; writing-review and editing, XT and JW acquisition, XT and JW. All authors have read and agreed to the published version of the manuscript

Acknowledgments: The authors would like to thank Chunyan Pei for her assistance with our experiments.

References

1. Tian XM, Yan LH, Xiang GF, et al. Study on grafting and propagation technology of *Semiliquidambar cathayensis* Chang. *Jiangsu Agric Sci*. 2018;46:123–4.
2. Tian XM, Zeng LZ, Yan LH, et al. Study on transcriptome characteristic of *Semiliquidambar cathayensis* Chang. *Hunan For Sci Technol*. 2018;45:40–50.
3. Yang WL, Yao ZS, Luo XQ, et al. Analgesic and anti-inflammatory effects of the ethanol extracts of Jinlu Ban Fenghe (*Semiliquidambar cathayensis*) root. *Jiangxi Sci*. 1999:3–5.
4. Xie SY, Yao KL, Wu XJ, et al. Overview of pharmacological research on *Semiliquidambar cathayensis* H.T. Chang. *J Fujian For Sci Technol*. 2018;45:122–7.
5. Yang L, Liu RH, He JW. Rapid analysis of the chemical compositions in *Semiliquidambar cathayensis* Roots by ultra high-performance liquid chromatography and quadrupole time-of-flight tandem mass spectrometry. *Molecules*. 2019;24. DOI: [10.3390/molecules24224098](https://doi.org/10.3390/molecules24224098).
6. Lu HX, Wu ZL, Liang WJ, Chen ML, Huang BB, Wei QQ. Chemical constituents from *Semiliquidambar cathayensis* Roots. *Zhong Yao Cai*. 2015;38:2543–6.

7. Zhou GX, Yang YC, Shi JG, et al. Studies on chemical constituents from *Semiliquidambar cathayensis*. Chin Trad Herb Drugs. 2002;16–8.
8. Singh SK, Patra B, Paul P, Liu Y, Pattanaik S, Yuan L. BHLH iridoid SYNTHESIS 3 is a member of a bHLH gene cluster regulating terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Plant Direct. 2021;5:e00305. DOI: [10.1002/pld3.305](https://doi.org/10.1002/pld3.305).
9. Cherukupalli N, Divate M, Mittapelli SR, Khareedu VR, Vudem DR. De novo assembly of leaf transcriptome in the medicinal plant *Andrographis paniculata*. Front Plant Sci. 2016;7:1203. DOI: [10.3389/fpls.2016.01203](https://doi.org/10.3389/fpls.2016.01203).
10. Balusamy SRD, Rahimi S, Cho YG, Senthil K, Yang DC. Overexpression of geraniol 10-hydroxylase from panax ginseng conferred enhanced resistance to *Pseudomonas syringae* in Arabidopsis. Plant Growth Regul. 2017;81:305–16.
11. Fang X, Li CY, Yang Y, Cui MY, Chen XY, Yang L. Identification of a novel (-)-5-Epiheremophilene synthase from *Salvia miltiorrhiza* via transcriptome mining. Front Plant Sci. 2017;8:627. DOI: [10.3389/fpls.2017.00627](https://doi.org/10.3389/fpls.2017.00627).
12. Karpaga Raja Sundari B, Budhwar R, Dwarakanath BS, Thyagarajan SP. De novo transcriptome analysis unravels tissue-specific expression of candidate genes involved in major secondary metabolite biosynthetic pathways of *Plumbago zeylanica*: Implication for pharmacological potential. 3 Biotech. 2020;10:271. DOI: [10.1007/s13205-020-02263-9](https://doi.org/10.1007/s13205-020-02263-9).
13. Xu Y, Tian S, Li R, Huang X, Li F, Ge F, et al. Transcriptome characterization and identification of molecular markers (SNP, SSR, and indels) in the medicinal plant *Sarcandra glabra* spp. BioMed Res Int. 2021;2021:9990910. DOI: [10.1155/2021/9990910](https://doi.org/10.1155/2021/9990910).
14. Liu ZQ, Zhu GQ, Chen R, et al. Application of transcriptomics in medicinal plant *Ginkgo*. Acad. J. Shanghai Univ. J Trad Chin Med. 2020;34:88–93.
15. Jin CF, Li ZQ, Li Y, Wang S, Li L, Liu M, et al. Transcriptome analysis of terpenoid biosynthetic genes and simple sequence repeat marker screening in *Eucommia ulmoides*. Mol Biol Rep. 2020;47:1979–90. DOI: [10.1007/s11033-020-05294-w](https://doi.org/10.1007/s11033-020-05294-w).
16. Shi XD, Gu YX, Dai J, et al. Analysis of secondary metabolism pathway genes based on *Magnolia officinalis* transcriptome. Lishizhen Med Mater Med Res. 2018;29:247–50.
17. Yang JRNA. Seq for DEG analysis About triterpenoid saponin synthesis and transcript profiling of *Ardisia crenata Sims*. Master's Degree Sichuan Agric Univ Sichuan Prov China. 2015.
18. Cheng J, Lyu Zh, Wang C, Chen J, Lin T. De novo analysis of *Heortia Vitessoides* (Lepidoptera: Crambidae) transcriptome and identification of putative cytochrome P450 monooxygenase genes. J Entomol Sci. 2019;54:293–315. DOI: [10.18474/JES18-103](https://doi.org/10.18474/JES18-103).
19. Liu YL, Geng YP, Song ML, Zhang P, Hou J, Wang W. Genetic structure and diversity of *Glycyrrhiza* populations based on transcriptome SSR markers. Plant Mol Biol Report. 2019;37:401–12. DOI: [10.1007/s11105-019-01165-2](https://doi.org/10.1007/s11105-019-01165-2).
20. Guan LL, Xia QF, Shi XB, et al. Cloning and analysis of geranyl pyrophosphate synthase (GPPS) sequence of *Blumea balsamifera* L. DC on transcriptome information. Chin J Trop Crops.

- 2016;37:901–9.
21. Singh S, Pal S, Shanker K, Chanotiya CS, Gupta MM, Dwivedi UN, et al. Sterol partitioning by HMGR and DXR for routing intermediates toward with anolide biosynthesis. *Physiol Plant*. 2014;152:617–33. DOI: [10.1111/ppl.12213](https://doi.org/10.1111/ppl.12213).
 22. Manzano D, Fernández-Busquets X, Schaller H, González V, Boronat A, Arró M, et al. The metabolic imbalance underlying lesion formation in *Arabidopsis thaliana* overexpressing farnesyl diphosphate synthase (isoform 1S) leads to oxidative stress and is triggered by the developmental decline of endogenous HMGR activity. *Planta*. 2004;219:982–92. DOI: [10.1007/s00425-004-1301-y](https://doi.org/10.1007/s00425-004-1301-y).
 23. Yang WZ, Cao J, Wu Y, Kong F, Li L. Review on plant terpenoid emissions worldwide and in China. *Sci Total Environ*. 2021;787:147454. DOI: [10.1016/j.scitotenv.2021.147454](https://doi.org/10.1016/j.scitotenv.2021.147454).
 24. Sun ST, McKenna CE. Farnesyl pyrophosphate synthase modulators: A patent review (2006–2010). *Expert Opin Ther Pat*. 2011;21:1433–51. DOI: [10.1517/13543776.2011.593511](https://doi.org/10.1517/13543776.2011.593511).
 25. Tan Y, Tang CH, Feng J, Feng N, Wu YY, Bao DP, et al. Current progress in the study on biosynthesis and regulation of *Ganoderma* triterpenoids. *Acta Edulis Fungi*. 2019;26:125–40.
 26. Su HL. Study on Molecular Structure, biochemical characterization and expression of squalene synthase from *Siraitia grosvenorii*. Doct. Degree Guangxi Med. Univ. Guangxi Prov. *China* 2017.
 27. Rong Q, Jiang D, Chen Y, Shen Y, Yuan Q, Lin H, et al. Molecular cloning and functional analysis of squalene synthase 2(SQS2) in *Salvia miltiorrhiza Bunge*. *Front Plant Sci*. 2016;7:1274. DOI: [10.3389/fpls.2016.01274](https://doi.org/10.3389/fpls.2016.01274).
 28. Li HH, Liu XY, Wang W, et al. Cloning and sequence analysis of squalene synthase gene in *Eriobotrya japonica* L. *Chin J Trop Crops*. 2014;35:1090–4.
 29. Rong Qx, Liu CS, Huang LQ, Zhang N, Nan B, Guo W. Cloning and sequence analysis of squalene synthase gene and cDNA in *Glycyrrhiza uralensis*. *Zhongguo Zhong Yao Za Zhi*. 2011;36:1416–20.
 30. Yuan XL, Zhao N, Hua M, et al. The cloning and expression analysis of squalene synthetase gene (AcSQS) in *Antrodia camphorata*. *J Fujian Norm Univ (Nat Sci Ed)*. 2017;33:50–9.
 31. Jiang SC, Liu WC, Wang Y, et al. Correlation between ginsenoside accumulation and SQS and SQE gene expression in different organs of *Panax quinquefolius*. *Chin Trad Herb Drugs*. 2011;42:579–84.
 32. Kim TD, Han JY, Huh GH, Choi YE. Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. *Plant Cell Physiol*. 2011;52:125–37. DOI: [10.1093/pcp/pcq179](https://doi.org/10.1093/pcp/pcq179).
 33. Li ZD, Zhao RH, Zhang ZC, et al. Advances in biosynthesis and regulation mechanism of notoginseng saponins. *Chin J Exp Trad Med Formulae*. 2018;24:207–13.
 34. Elfahmi Y, Luthfia N, Tati K, Agas C, Sony S. Effect of terbinafine and DMSO on the gene expression level of squalene synthase (Sqs) and Amorpha-4,11-diene synthase (ads) in *Artemisia annua* L. *Procedia Chem*. 2014;13:85–91.
 35. Guirimand G, Guihur A, Phillips MA, Oudin A, Glévarec G, Melin C, et al. A single gene encodes Isopentenyl diphosphate isomerase isoforms targeted to plastids, mitochondria and peroxisomes in

- Catharanthus roseus*. Plant Mol Biol. 2012;79:443–59. DOI: [10.1007/s11103-012-9923-0](https://doi.org/10.1007/s11103-012-9923-0).
36. Tomohisa H, Shinya T, Shunsuke H, et al. Overexpression of 1-deoxy-D-xylulose-5-phosphate reductoisomerase gene in chloroplast contributes to increment of isoprenoid production (PLANT BIOTECHNOLOGY). J Biosci Bioeng. 2008;5:518–26. DOI: [10.1263/jbb.105.518](https://doi.org/10.1263/jbb.105.518).
37. Yan XM, Zhang L, Wang J, Liao P, Zhang Y, Zhang R, et al. Molecular characterization and expression of 1-deoxy-d-xylulose 5-phosphate reductoisomerase (DXR) gene from *Salvia miltiorrhiza*. Acta Physiologiae Plant. 2009;5:1015–22. DOI: [10.1007/s11738-009-0320-5](https://doi.org/10.1007/s11738-009-0320-5).
38. Zhang W, Wang YL, Lin J, et al. Transformation of *Ginkgo biloba* with 1-hydroxy -2- Methy -2-(E)- Buteny -4- diphosphate reductase gene. J Fudan Univ (Nat Sci). 2008;47:598–603. DOI: [10.1007/s11442-008-0073-x](https://doi.org/10.1007/s11442-008-0073-x).
39. Jin BL, Cui GH, Dang BY, et al. Heterologous expression and purification of Recombinant CPS4 from *Salvia miltiorrhiza*. Chin J Exp Trad Med Formulae. 2014;20:94–8.
40. Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J. 2000;24:241–52. DOI: [10.1046/j.1365-3113x.2000.00875.x](https://doi.org/10.1046/j.1365-3113x.2000.00875.x).
41. Zhang XN, Guan HY, Dai ZB, Guo J, Shen Y, Cui G, et al. Functional analysis of the isopentenyl diphosphate isomerase of *Salvia miltiorrhiza* via color complementation and RNA interference. Molecules. 2015;20:20206–18. DOI: [10.3390/molecules201119689](https://doi.org/10.3390/molecules201119689).
42. Zhang DY, Wen H, Wang W, et al. Transcriptional analysis of terpenoid biosynthesis in *Aconitum Carmichaelii*. Chin J Exp Trad Med Formulae. 2017;23:45–50.
43. Kesharwani S, Sundriyal S. Non-hydroxamate Inhibitors of 1-deoxy-d-xylulose 5-phosphate reductoisomerase (DXR): A Critical review and future perspective. Eur J Med Chem. 2021;213:113055. DOI:[10.1016/j.ejmech.2020.113055](https://doi.org/10.1016/j.ejmech.2020.113055).
44. Chen J, Gu W, Duan JL, et al. Study on active components of *Andrographis paniculata* at various growth periods and differential expression of key enzyme gene. Chin Trad Herb Drugs. 2014;45:3149–52. DOI: [10.7501/j.issn.0253-2670.2014.21.020](https://doi.org/10.7501/j.issn.0253-2670.2014.21.020).
45. Tian XM, Yan LH, Jiang L, et al. Chemical comparison of different tissues of *Semiliquidambar cathayensis* Chang by UHPLC-QTOF/MS-based metabonomics. Plant Physiol J. 2021;57:1311–8. DOI: [10.13592/j.cnki.ppj.2020.0558](https://doi.org/10.13592/j.cnki.ppj.2020.0558).

Figures

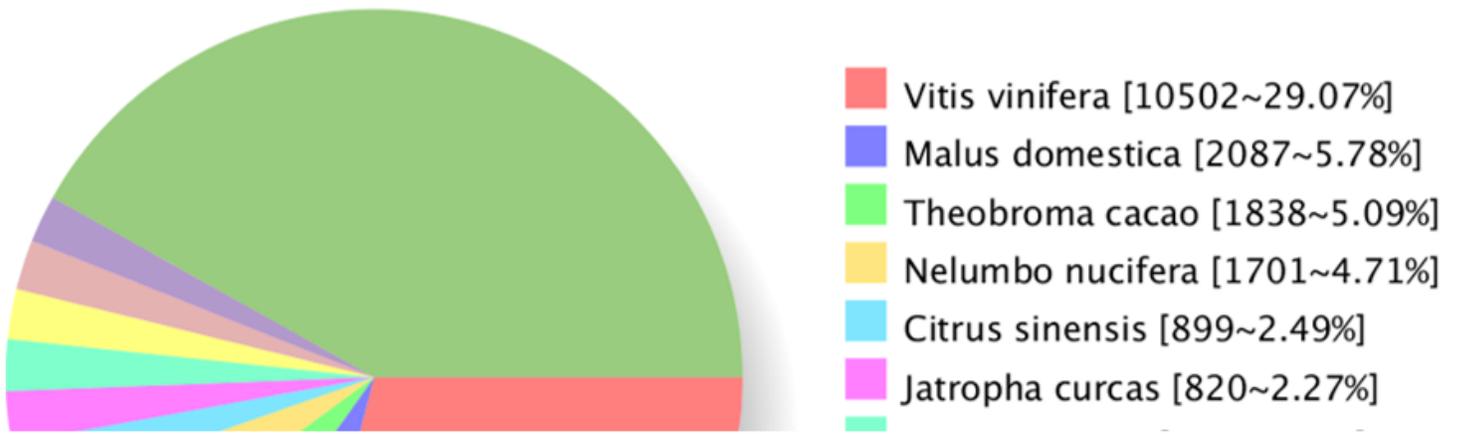


Figure 1

Unigenes for *Semiliquidambar cathayensis* Chang similarity comparison with other species in Nr database

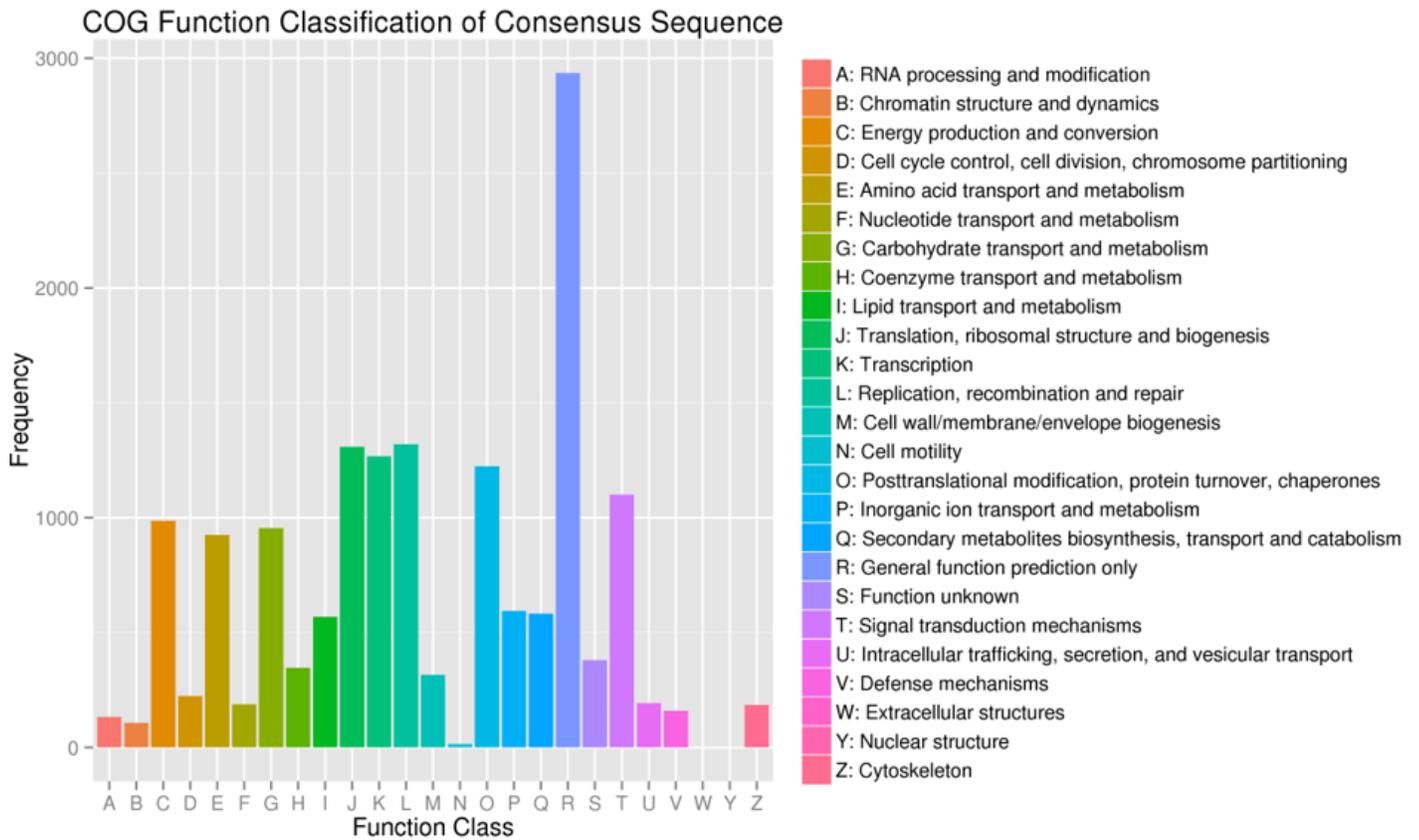


Figure 2

COG classification of unigenes for *Semiliquidambar cathayensis* Chang

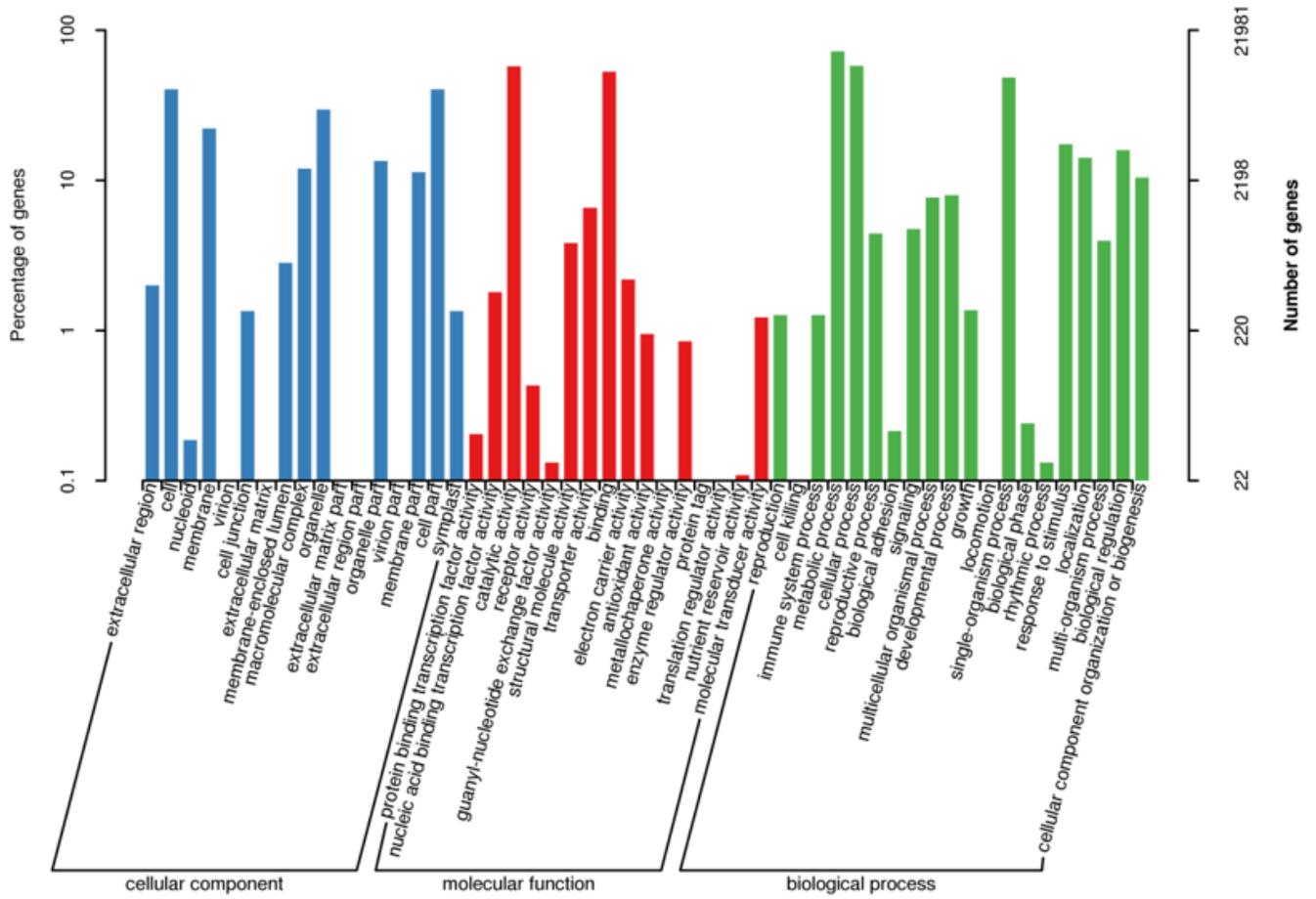


Figure 3

GO classification of unigenes for *Semiliquidambar cathayensis* Chang

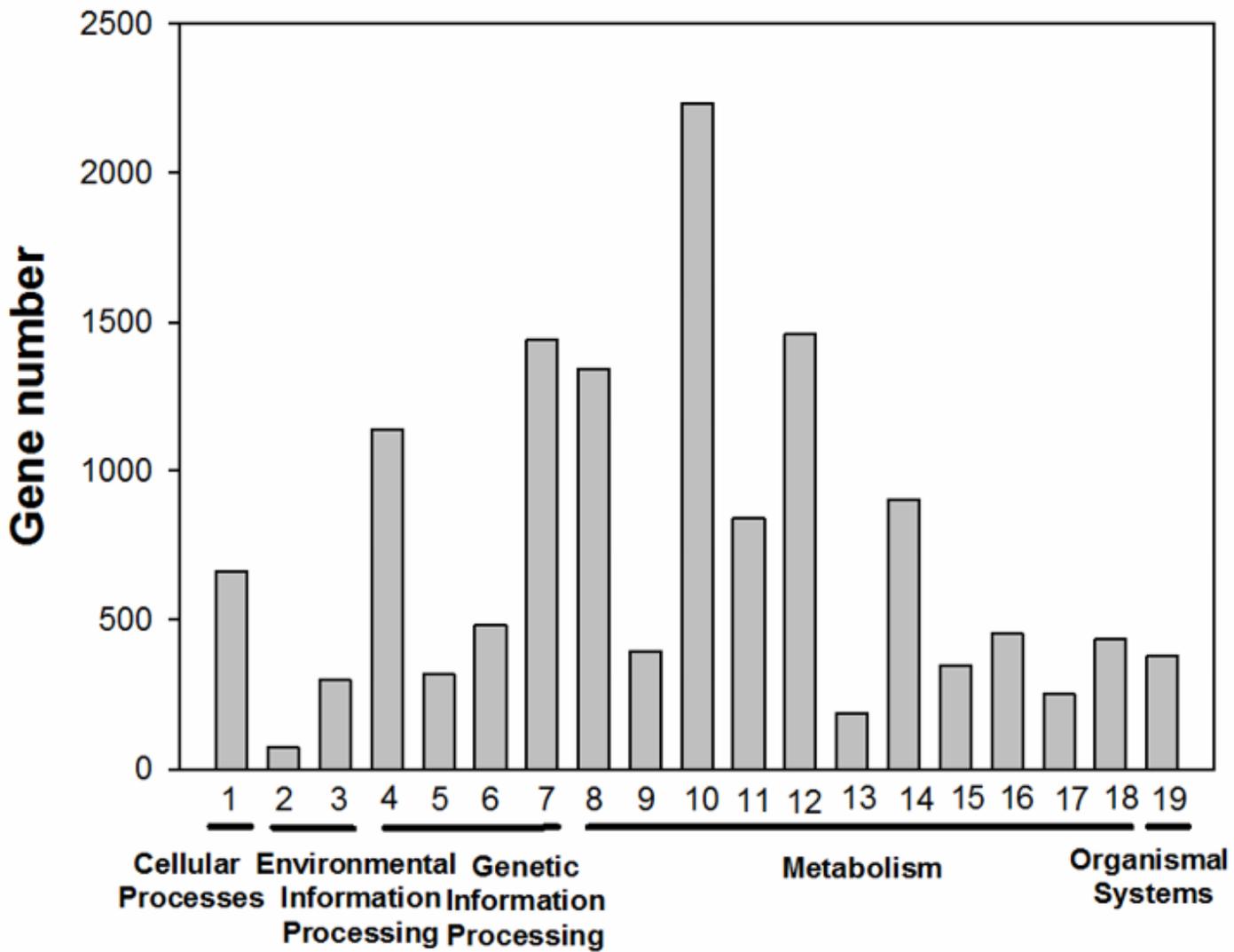


Figure 4

KGGG metabolic pathway for *Semiliquidambar cathayensis* Chang

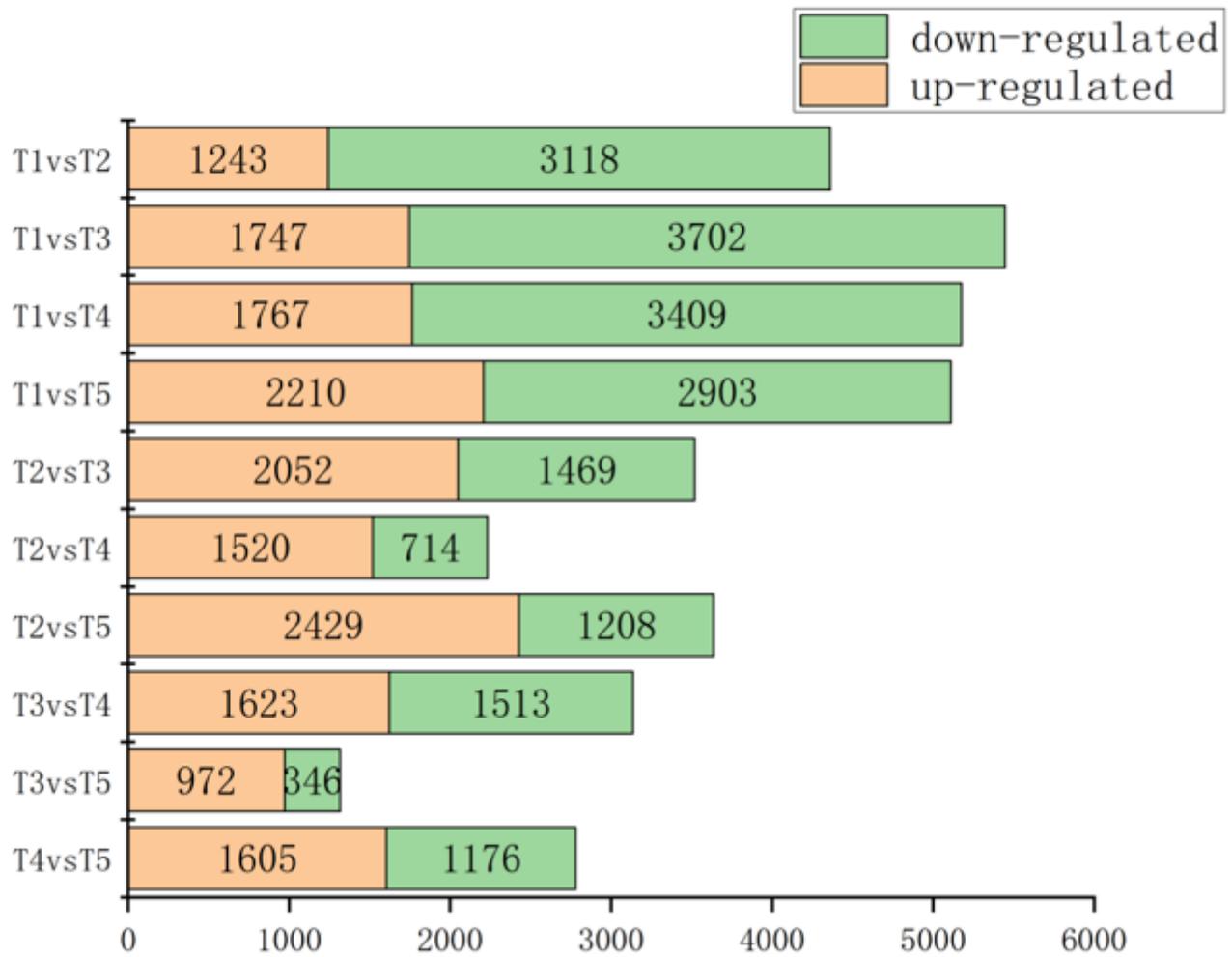


Figure 5

The number of up-regulated and down-regulated genes in different tissues of *Semiliquidambar cathayensis* Chang

Note ☐ Leaf (T1), stem epidermis (T2), stem xylem (T3), root epidermis (T4), root xylem (T5)

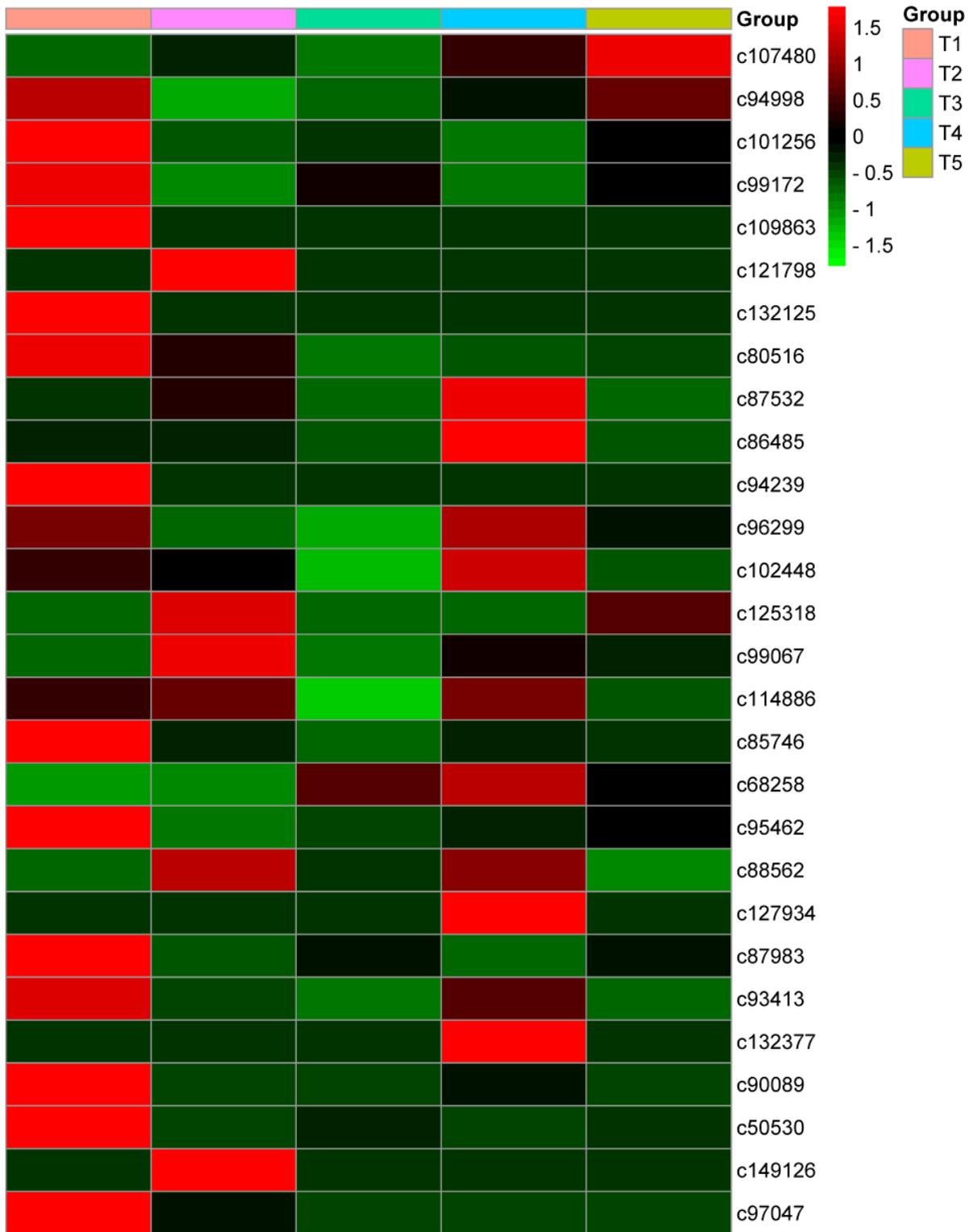


Figure 6

Heat map of gene expression related to terpenoid biosynthesis

Note: Different columns represent different samples, and different rows represent different genes. The color represents the score value after FPKM normalization of the gene expression in the samples

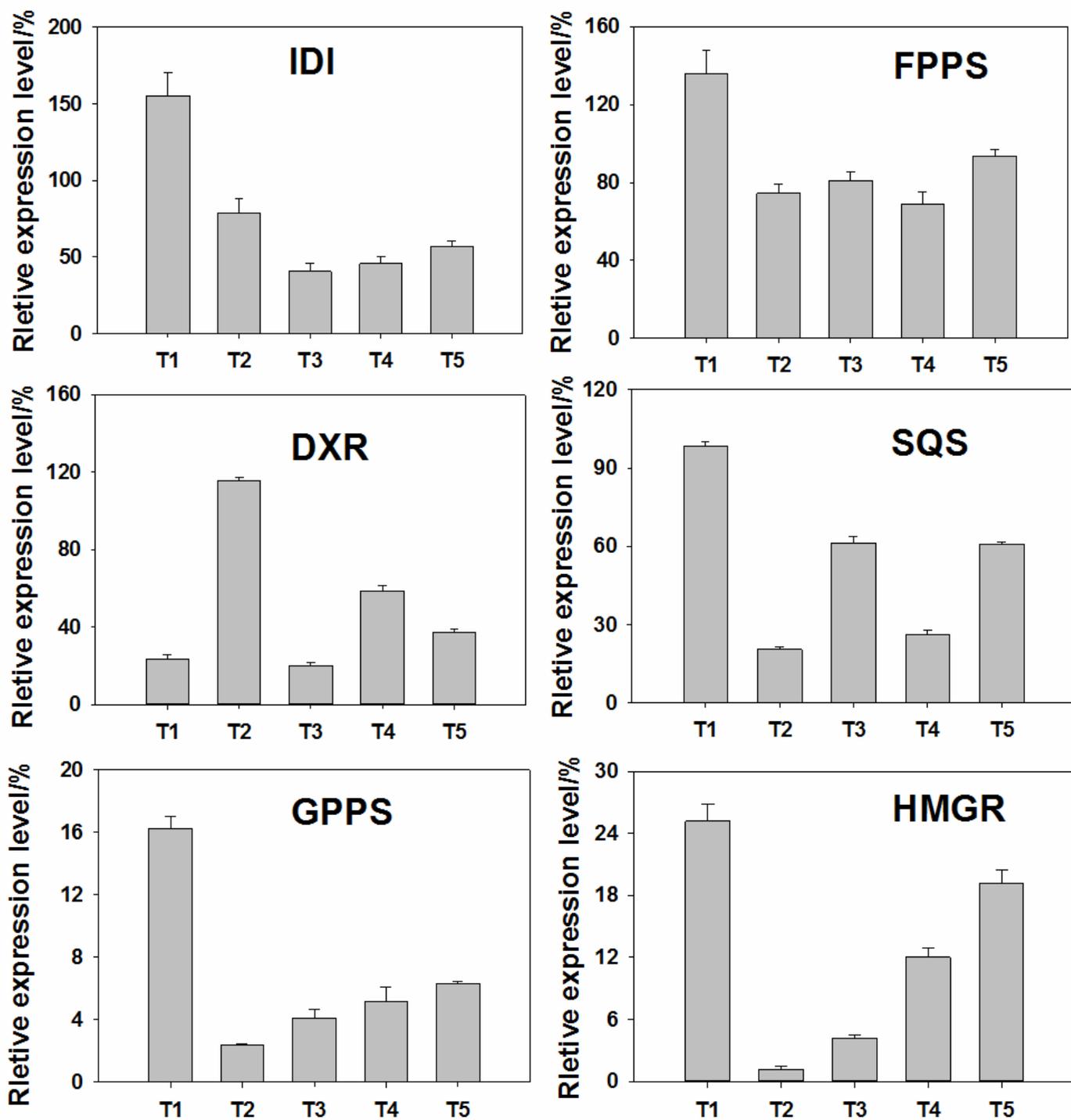


Figure 7

Real-time PCR verification for some important unigenes in terpenoid metabolic pathways

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLEMENTARYFILE.docx](#)