biologia plantarum
an international journal for experimental botany

BIOLOGIA PLANTARUM (2022) 66: 146-154 DOI: 10.32615/bp.2021.073

Expression of genes encoding terpenoid biosynthesis enzymes during leaf development of *Eucalyptus camaldulensis*

N. ZHAN[1,2], L. HUANG[3], Z. WANG[4], J. ZHANG[1], Y. XIE[1], X. SHANG[1], G. LIU[1], and Z. WU[1,*]

[1] *China Eucalypt Research Centre, Chinese Academy of Forestry, 524022, Guangdong, P.R. China*
[2] *Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences, Guangdong 510640, P.R. China*
[3] *Research Institute of Tropical Forestry, Chinese Academy of Forestry, 510520, Guangdong, P.R. China*
[4] *Beihai City Protective Forest Farm, 536000, Guangxi, P.R. China*

[*]Corresponding author: E-mail: wzhua2889@163.com

Abstract

To reveal the regulation mechanism of terpenoid biosynthesis in the leaves of *Eucalyptus camaldulensis*, the content of volatiles in eucalyptus leaves and the transcriptome databases of young and mature leaves were analyzed. The results showed that *E. camaldulensis* contains 92 and 89 kinds of volatile substances in the young and mature leaves, respectively. Among them, the content of 1,8-cineole, β-pinene, and other substances was significantly different in young and mature leaves. A total of 99 802 unigenes were obtained from the transcriptome database of young and mature leaves of *E. camaldulensis* and 18 441 genes displayed obviously differential expressions during both developmental stages. There were 6 982 up-regulated unigenes and 11 461 down-regulated unigenes in the young leaf stage compared to the mature leaf stage. The key genes for terpenoid biosynthesis, including *limonene synthase-10, limonene synthase-11, myrcene synthase-1, α- pinene synthase-2,* and *1,8-cineole synthase-2*, were selected for further analysis to explore the mechanism of gene regulation and genetic transformation. The expressions of key genes were validated by RT-qPCR, and their expressions were consistent with RNA-seq data. *WRKY, MYB, NAC*, and *bHLH* transcription factors (TFs) displayed important regulatory effects on the above key genes. Thus, a regulatory network model of terpenoid biosynthesis was constructed using target genes and TFs during leaf development in *E. camaldulensis*. These results provide theoretical evidence for understanding the terpenoid biosynthesis in plants and reference for terpenoids utilization by genetic engineering methods in *E. camaldulensis*.

Keywords: Eucalyptus camaldulensis, regulation mechanism, terpenoid biosynthesis, transcriptome, young and mature leaves.

Introduction

Eucalyptus camaldulensis Brooker & M.W.McDonald belongs to the group *symphyomyrtus*, which is characterized by fast growth, strong adaptability under different conditions, and resistance against drought and alkali, and it is widely used in the cultivation and breeding of *Eucalyptus*. The reports have shown that high content of eucalyptus oil is extracted from fresh leaves of *E. camaldulensis* (Boland *et al*. 1991, Tian *et al*. 2005). The leaves are one of the main organs of the synthesis and release of terpenes, which affect the basic behaviour and function of the plant. Moreover, the synthesis and release of the monoterpenes from leaves are affected by the leaf

Received 29 June 2021, *last revision* 5 November 2021, *accepted* 24 November 2021.
Abbreviations: COG - cluster of orthologous groups of proteins; DEGs - differentially expressed genes; DMAPP - dimethylallyl diphosphate; DXR - 1-deoxy-D-xylulose-5-phosphate reductoisomerase; DXS - 1-deoxy-D-xylulose-5-phosphate synthase; FGs - functional genes; GC-MS - gas chromatography-mass spectrometry; GO - gene ontology; HMGCS - hydroxymethylglutaryl-CoA synthase; IPP - isoprene phosphate; KEGG - Kyoto encyclopedia of genes and genomes; ME-CPP - 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; MEP - methylerythritol 4-phosphate; MVA - mevalonate; RIN - RNA integrity number; RPKM - reads per kb per million reads; TPS - terpene synthase; TFs - transcription factors.
Acknowledgements: This work was funded by the National Natural Science Foundation of China No. 31570615 and the Fundamental Research Funds for the Central Non-profit Research Institution of CAF No. CAFYBB2017ZX001-5.
Conflict of interest: The authors declare that they have no conflict of interest.

age (Perez-Mar *et al.* 2014). Eucalyptus oil is a kind of mixed terpenoid compounds and their oxygen-containing derivatives, which play important roles in the growth and development of plants. Terpenoids in plants could attract or avoid herbivorous insects. Production of volatile terpenoids is considered a substantial and irretrievable investment for plants; however, even if photosynthesis is inhibited, volatile terpenoids production and emission are sustained, suggesting they are beneficial to plants under stress conditions (Greenhagen and Chappell 2001, Dudareva *et al.* 2004, Calvert *et al.* 2018, Velho *et al.* 2020). Terpenoids from eucalyptus have many properties pharmacologically suitable (Hosseini *et al.* 2020). Eucalyptus oil contains main terpenoids, such as 1,8-cineole, α-pinene, β-pinene, myrcene, and other volatile substances. Eucalyptus oil is widely used as an essential oil and food additive and has important value in food production (Gong *et al.* 2009, Yu *et al.* 2009, Yahyaa *et al.* 2018, Santadino *et al.* 2017).

There are two main biosynthesis pathways of terpenoids, one is the mevalonate (MVA) pathway and the second is the methylerythritol 4-phosphate (MEP) pathway (Yu *et al.* 2009). All terpenoids originate from the C_5-isoprene, isoprene phosphate (IPP), and its dimethylallyl diphosphate (DMAPP). DMAPP is only catalyzed by MEP pathway enzymes and 1-deoxy-D-xylulose-5-phosphate synthase (DXS) is considered a key catalyzing enzyme (Cordoba *et al.* 2009, Tholl and Lee 2011, Hofberger *et al.* 2015). Terpenoid synthases are key enzymes in the formation of terpenoids, including monoterpene synthase, sesquiterpene synthase, and diterpene synthase (Schilmiller *et al.* 2009). The sesquiterpene synthase gene in tobacco was firstly discovered in 1992 (Facchini *et al.* 1992), and then a variety of genes encoding terpenoid synthases have been cloned, *e.g.,* from *Taxus chinensis* (Wildung and Croteau 1996), *Abies grandis* (Bohlmann *et al.* 1998a,b, 1999), *Abies fabri* (Steele *et al.* 1998a,b), *Litsea cubeba* (Chang and Chu 2011), and *Populus trichocarpa* (Irmisch *et al.* 2014).

Transcription factors (TFs), known as *trans*-acting factors or *cis*-elements regulating functional genes (FGs) transcription, play significant roles in almost all aspects of plant growth development. Many catalyzing enzymes functioning in the biosynthesis pathways of plant terpenoids are often regulated at the transcription level by TFs, therefore, TFs are supposed to be very important in the biosynthetic pathways of terpenoids (Mahmoud and Croteau 2002).

In this study, to understand the molecular mechanism of terpenoid biosynthesis in *E. camaldulensis* leaves, eucalyptus oil was extracted from young and mature leaves by steam distillation, and its composition and content were analyzed by gas chromatography-mass spectrometry (GC-MS). Moreover, FGs and TFs participating in terpenoid biosynthesis were screened and analyzed on the base of construction and annotation of transcriptome database from two leaf development stages by *HiSeq2000 Illumina* sequencing. So, a regulatory network model of terpenoid biosynthesis in two leaf development stages of *E. camaldulensis* was constructed by correlation analysis between main eucalyptus oil components and key FGs and TFs.

Materials and methods

Plant materials: The leaf samples of *Eucalyptus camaldulensis* Brooker & M.W.McDonald were collected during April 2018 from Southern China Experiment Nursery in Zhanjiang, Guangdong, China (21°15′30.69″N, 110°06′41.95″E). The young leaves (1 week old) and mature leaves (4 weeks old) of nine healthy *E. camaldulensis* plants were collected from the same half-sib families located on the Morehead R. Queensland (15°15' S, 143°34' E). Some collected samples were sealed in the fresh-keeping ice bag after being weighed on the spot and taken back to the laboratory for extracting eucalyptus oil. The other samples were kept in liquid nitrogen for RNA extraction.

For extraction of eucalyptus oil, 60 g of young and mature fresh leaves (leaf stalks removed) were cut into 1 - 2 cm pieces, ground into powder by *JJ-2* organization crusher (*BILON*, Shanghai, China) and circulated by steam extractor for 4 h (Bounatirou *et al.* 2007). The *7890 GC-5975* MS gas chromatograph-mass spectrometer (*Agilent Technologies*, Santa Clara, CA, USA) was used in this research. Each sample was measured 3 times. The GC was equipped with an *HP-5MS* quartz elastic capillary column (30 m × 0.25 mm × 0.25 μm). The solvent was *n*-hexane and the carrier gas was high pure helium (99.999 %). Temperature program: 60 °C for 1 min, up to 170 °C at 3 °C min^{-1} for 2 min, up to 260 °C at 8 °C min^{-1}, and last 260 °C for 5 min. The sample size was 1.0 mm^3, the split ratio was 20:1 and the injector temperature was 260 °C. The MS was operated with ionizing energy of 70 eV, ion source temperature of 230 °C, quadrupole temperature of 150 °C, electron multiplier voltage of 1024 V, and a scan range of 40 - 600 m/z.

The volatile components from young and mature leaves were analyzed by GC-MS and the total ion flow chromatogram and MS peaks were obtained. According to the automatic retrieval of GC-MS data processing system using *NIST08* (National Institute of Standards and Technology) mass spectrum database, and combined with manual map analysis, the components of the extract were qualitatively analyzed. The relative percentage of each component was determined by the normalization of the chromatographic peak area. Its chemical structure was determined by reference to a standard atlas and related literature (Cong and Su 2000). To analyse the absolute content of 1,8-cineole, α-pinene, β-pinene, myrcene, and limonene standard substance was weighed and dissolved in hexane solution for preparation of mother liquor of standard solution and then diluted to the gradient concentration required to make the standard curve. The concentration used were: 1,8-cineole (0.1, 0.2, 0.4, 0.6 and 0.8 mg cm^{-3}), α-pinene (0.04, 0.08, 0.12, 0.16 and 0.20 mg cm^{-3}),

β-pinene (0.03, 0.06, 0.09, 0.12 and 0.15 mg cm^{-3}), myrcene (0.05, 0.10, 0.15, 0.20 and 0.25 mg cm^{-3}) and limonene (0.1, 0.2, 0.4, 0.6 and 0.8 mg cm^{-3}). Each sample was measured 3 times and its average value was calculated. The regression equations between peak area and standard concentrations were as follows:

$X_{1,8\text{-cineole}} = 1.12\times10^{-8}y - 0.9745$ ($R^2 = 0.9984$)
$X_{\alpha\text{-pinene}} = 2.01\times10^{-7}y - 8.7\times10^{-3}$ ($R^2 = 0.9964$)
$X_{\beta\text{-pinene}} = 5.24\times10^{-11}y - 1.34\times10^{-3}$ ($R^2 = 0.9987$)
$X_{myrcene} = 0.59\times10^{-7}y - 0.0092$ ($R^2 = 0.9957$)
$X_{limonene} = 1.80\times10^{-9}y + 0.0554$ ($R^2 = 0.9974$)

where X is the increasing concentration and Y is the peak area. According to the solution concentration, we calculated the absolute content of the sample.

RNA extraction and transcriptome data processing: The total RNA was extracted from the samples according to the instructions of the *EASY Spin Plus Plant RNA* kit (*AidLab*, Beijing, China). RNA concentration was detected by *Nanodrop, 28S/18S* (*Nanodrop Technologies*, Foster, USA) and 28S (sedimentation coefficient) and 18S rRNAs ratio was considered the benchmark for intact RNA. RIN (RNA integrity number) values were detected by *Agilent 2100* bioanalyzer (*Agilent Technologies*) and 1 % (m/v) agarose gel electrophoresis was used to detect the quality and integrity of the extracted RNA. *Illumina HiSeq TM 2000* platform was used for transcriptome sequencing by *Gene Denovo* (Guangzhou, China). *TopHat* (*v. 2.0.10*) was used to compare the filtered reads to the *E. grandis* reference genome (NCBI gcf_000612305.1).

Gene analysis and functional annotation: Using the RPKM (reads per kb per million reads) method, the number of unigenes was calculated by cuffdiff (Trapnell *et al.* 2012). *DEGseq v. 1.14.0* software was used for identifying differentially expressed genes between two samples. The *P*-value could be assigned to each gene and genes with $q \leq 0.05$ and $|\log2| \geq 1$ were used to determine differentially expressed genes (DEGs). The obtained unigenes were compared with the *Nr* (non-redundant protein sequence database), *Swiss-Prot* (protein sequence database), *KEGG* (Kyoto encyclopedia of genes and genomes), and *COG* (Cluster of orthologous groups of proteins) databases. E-value was set to be less than 1e-5, and functional annotation and classification information were obtained according to the similarity of protein/genes. Compared with *Plant Transcription Factor Database v. 5.0* (http://planttfdb.cbi.pku.edu.cn/), 1684 unigene transcription factors were annotated. *Cytoscape* software was used to draw network diagram.

Reverse transcription and qPCR: 5 μg of RNA were taken, cDNA was synthesized using the *M-MLV* reverse transcriptase according to *PrimeScripte* RT reagent kit (*TaKaRa*, Kusatsu, Japan), and diluted 1:10 for subsequent experiments. 12 key genes in the terpene biosynthesis of *E. camaldulensis* were validated by RT-qPCR with thermal cycler apparatus (*BIO-RAD iQ5 7700*, *Applied Biosystems,* California, USA). The total volume of the RT-qPCR reaction system was 20 mm^3 and it contained 10 mm^3 of *2×SYBR Premix Ex Taq™* (*Tli RNaseH Plus*), 0.4 mm^3 of *ROX Dye*, 0.8 mm^3 of forward primer (10 μM), and 0.8 mm^3 of reverse primer (10 μM), 2 mm^3 of cDNA, and 6 mm^3 of RNase free ddH_2O. The thermal profile comprised 95 °C for 3 min, 45 cycles at 95 °C for 10 s for denaturation, 10 s annealing at 58 °C, and 20 s extension at 72 °C (the fluorescence signal was collected at the end of each cycle). Each reaction was repeated three times. Primers were designed using *Primer Express 2.0* software (*PE Applied Biosystems*, Foster, USA) under default parameters. The primer sequences are given in Table 1 Suppl. The *actin* gene was used as the internal reference gene, which is expressed stably in the whole growth process of *E. camaldulensis* leaf. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) and *Excel* software were used to analyze the data.

Results

The volatile substances in young and mature leaves of *E. camaldulensis* were analyzed by GC-MS and the total ion flow chromatogram and mass spectra were obtained. *NIST* standard spectrum library was used to retrieve different peaks by computer and investigated according to the mass spectrometry cracking rule. Its chemical structure was determined by reference to standard atlas and related literature (Cong and Su 2000). The Peak area normalization method was used to calculate the relative content of each component. The results showed that the volatile substances of young and mature leaves of *E. camaldulensis* contained 92 and 89 components, respectively (Tables 2 Suppl. and 3 Suppl.). The young leaves of *E. camaldulensis* contained 45 types of terpenoids, and the mature leaves contained 27 types of terpenoids. There were significant differences in the volatile substances of young and mature leaves of *E. camaldulensis*. The content of 1,8-cineole was much higher in mature leaves of *E. camaldulensis* than that in young leaves. In contrast, the content of β-pinene was higher in young leaves than in mature leaves (Fig. 1). However, there were non-significant differences in the content of α-pinene in young and mature leaves. The limonene was detected in young leaves, while it was not detected in mature leaves. The content of myrcene was not detected either in young or in mature leaves.

With the continuous increase of reads, the total number of unigenes also increased. Finally, the number of unigenes obtained tends to saturation. This indicates that the amount of library data is sufficient (Fig. 1 Suppl.). A total of 54 405 genes were found to be similar to those in the *COG* database and most of them could be associated with multiple genes. 54 405 corresponding relationships were established. According to their functions, the unigenes in the eucalyptus transcriptome could be roughly divided into 25 categories (Fig. 2). The *COG* function of unigenes was relatively comprehensive, involving most life activities, and the expression richness of various genes was different.

The analysis showed that there were 102 391 unigenes that were similar to genes in the *GO* database, and many single unigenes could correspond to multiple genes.

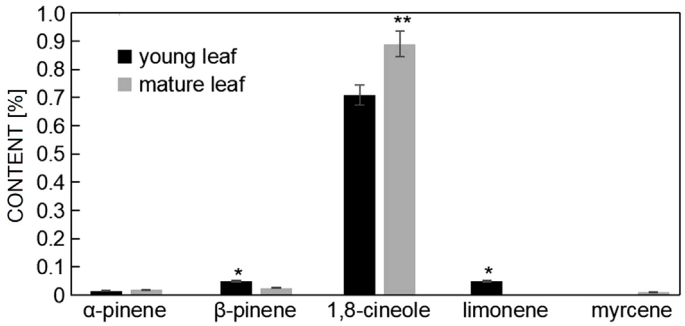

Fig. 1. Comparison of main terpenes in young and mature leaves. Means ± SEs, $n = 9$, * represents the significant difference ($P < 0.05$), and ** represents the highly significant difference ($P < 0.01$).

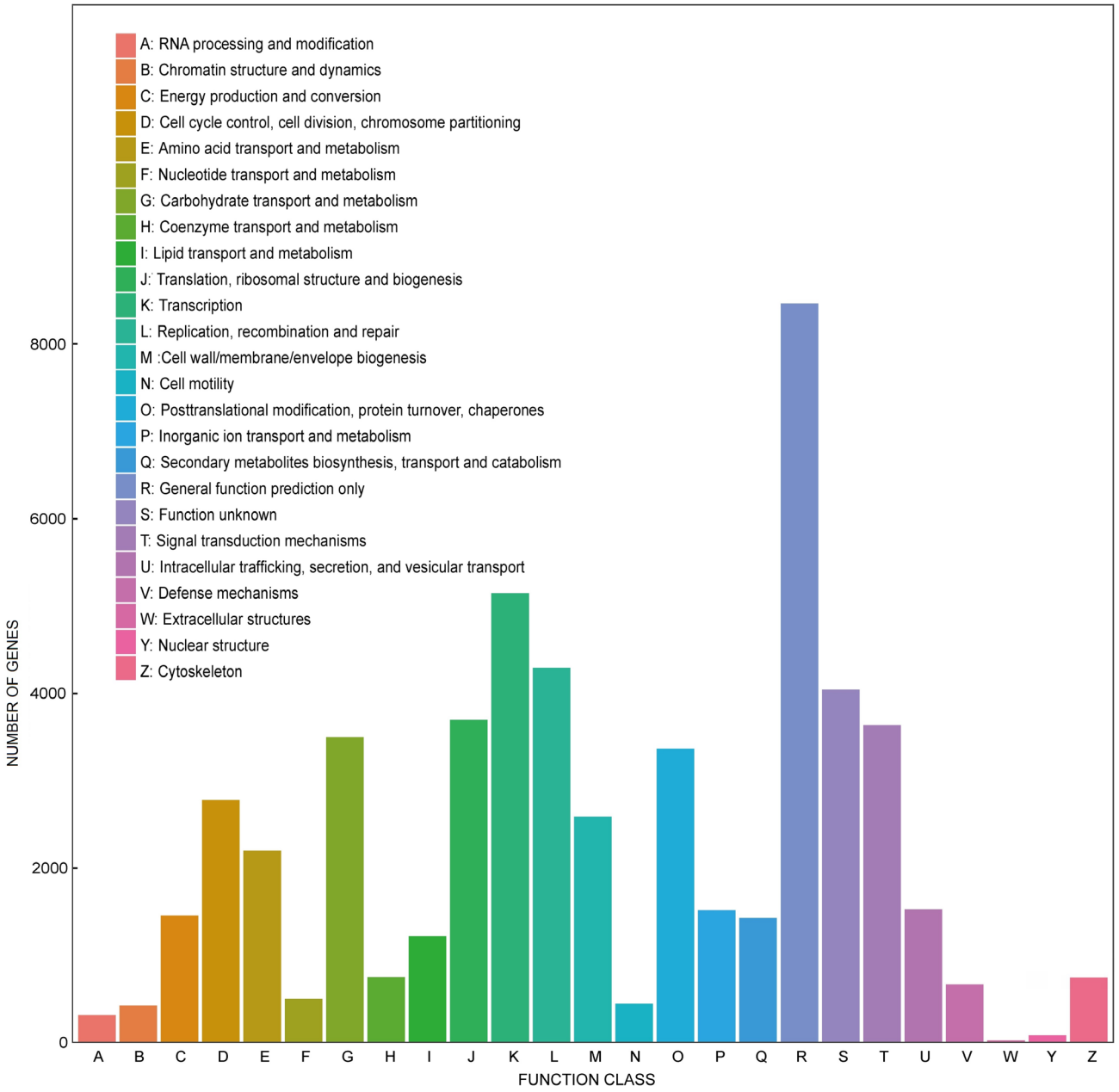

Fig. 2. COG functional classification of genes in *E. camaldulensis*. The capitals represent the unigenes divided into 25 categories with their functions.

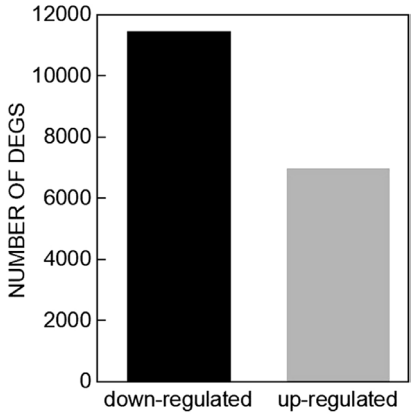

Fig. 3. Differentially expressed genes (DEGs) in mature and young leaves of *E. camaldulensis*.

101 724 corresponding relationships were established and as many comments and classifications as possible were obtained. *GO* enrichment analysis of unigenes could be roughly divided into 42 *GO* terms into 3 categories, including biological processes, cellular components, and molecular functions. A total of 39 969, 37 435, and 24 987 genes were involved in biological processes, cellular components and molecular function categories, respectively. 170 unigenes were involved in the synthesis of terpenoids, which accounted for 1.08 % of the total. A total of 18 443 DEGs were screened from *E. camaldulensis* leaves. Among them, 6 982 genes were up-regulated and 11 461 genes were down-regulated (Fig. 3).

According to the differential expressions and functional analysis of key genes in terpenoid biosynthesis at different stages, the metabolic/biosynthesis pathways for terpenoids were constructed based on the *KEGG* database (Fig. 4). 1-deoxy-D-xylulose-5-phosphate synthase (DXS) is one of the important rate-limiting enzymes in the MEP pathway, and two differentially expressed *DXS* genes were detected during terpenoid biosynthesis in eucalyptus. These two *DXS* genes exhibited down-regulation in mature leaves of *E. camaldulensis* compared to young leaves. 1-deoxy-D-xylulose-5-phosphate reductoisomerase (*DXR*) gene catalyzed reversible reaction and is a branch point of the

Table 1. Differential expression of genes in terpenoid biosynthesis of *E. camaldulensis*.RPKM - Reads per kilobase per million mapped reads**,** FDR - False discovery rate, $\log^2$ ratio - corresponds to log2 FC(fold change).

Unigene number	Gene name	Description	Young leaf RPKM	Mature leaf RPKM	Regulation	FDR	$\log^2$ ratio
0012757	*DXS1*	1-deoxy-D-xylulose-5-phosphate synthase-1	1.710	0.001	down	1.10E-07	-10.739
0012758	*DXS2*	1-deoxy-D-xylulose-5-phosphate synthase-2	2.461	0.146	down	3.80E-08	-4.077
0062059	*DXR*	1-deoxy-D-xylulose-5-phosphate reductoisomerase	9.585	3.284	down	2.66E-58	-1.545
0044938	*ME-CPP*	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	3.035	0.833	down	1.2E-07	-1.866
0027053	*HMGCS*1	hydroxymethylglutaryl-CoA synthase-1	4.005	0.231	down	2.15E-08	-4.114
0039300	*HMGCS2*	hydroxymethylglutaryl-CoA synthase-2	31.518	3.705	down	1.99E-201	-3.089
0047331	*HMGCS3*	hydroxymethylglutaryl-CoA synthase-3	54.552	25.918	down	1.80E-168	-1.073
0062767	*TPS1*	limonene synthase-1	1.657	3.110	up	5.15E-11	2.908
0062773	*TPS2*	limonene synthase-2	9.630	19.246	up	8.83E-06	3.999
0049797	*TPS3*	limonene synthase-3	0.190	1.365	up	3.60E-05	2.844
0062770	*TPS4*	limonene synthase-4	4.662	1.403	down	8.13E-09	-1.732
0095264	*TPS5*	limonene synthase-5	7.661	0.871	down	1.03E-69	-3.137
0024800	*TPS6*	limonene synthase-6	21.361	1.388	down	1.82E-41	-3.944
0033402	*TPS7*	limonene synthase-7	16.787	0.976	down	2.09E-63	-4.105
0038048	*TPS8*	limonene synthase-8	214.941	98.467	down	4.15E-231	-1.126
0046931	*TPS9*	limonene synthase-9	117.009	7.234	down	9.27E-121	-4.016
0038048	*TPS10*	myrcene synthase-1	214.942	98.467	down	4.15E-231	98.467
0024800	*TPS11*	myrcene synthase-2	21.361	1.388	down	1.82E-41	-3.944
0033400	*TPS12*	myrcene synthase-3	9.430	0.174	down	7.57E-60	-5.759
0062771	*TPS13*	α- pinene synthase-1	0.000	0.651	up	10.25E-14	9.346
0062765	*TPS14*	α- pinene synthase-2	3.752	0.267	down	1.12E-18	-3.813
0001449	*TPS15*	α- pinene synthase-3	1.602	0.000	down	2.67E-05	-10.646
0033401	*TPS16*	β-pinene synthase	4.241	0.324	down	1.39E-11	-3.711
0062767	*TPS17*	1,8-cineole synthase-1	1.657	3.110	up	1.03E-69	1.908
0062773	*TPS18*	1,8-cineole synthase-2	9.630	19.246	up	8.83E-06	3.999

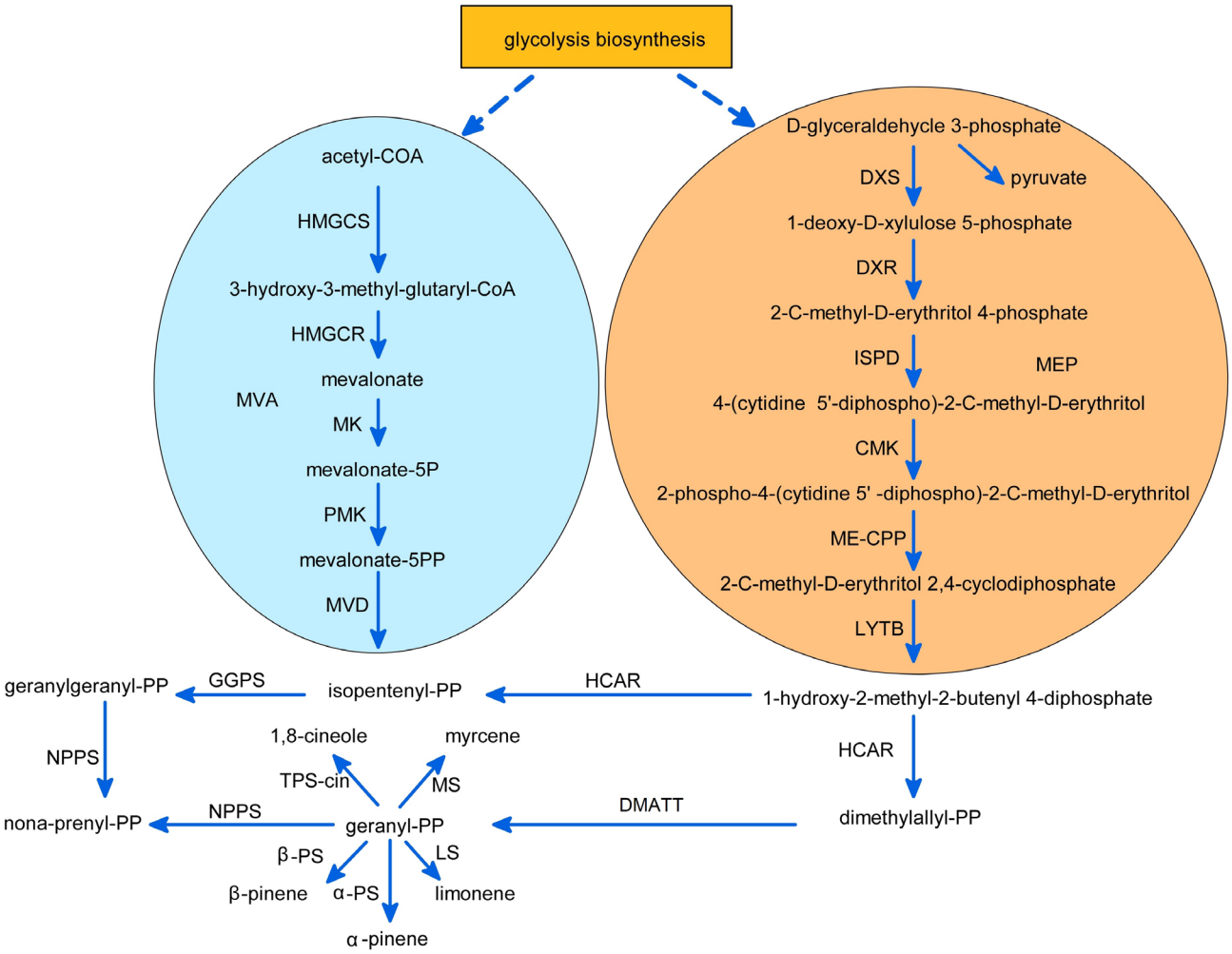

Fig. 4. Biosynthesis pathways for terpenoids in *E. camaldulensis*: MVA - mevalonate, MEP - methylerythritol 4-phosphate, HMGCS - hydroxymethylglutaryl-CoA synthase, HMGCR - hydroxymethylglutaryl-CoA reductase, MK - mevalonate kinase, PMK - phosphomevalonate kinase, MVD - pyrophosphomevalonate decarboxylase, DXS - 1-deoxy-D-xylulose-5-phosphate-synthase, DXR - 1-deoxy-D-xylulose-5-phosphate reductoisomerase, ISPD - isoprenoid synthase, CMK - 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, ME-CPP - 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, LYTB - (E)-4-hydroxy-3-methyl-2-butenyl-pyrophosphate reductase, HCAR - 7-hydroxymethyl chlorophyll *a* reductase, GGPS - deoxyxylose-5-phosphate synthetase, NPPS - nona-prenyl-phosphate synthetase, DMATT - dimethylallyltranstransferase, MS - myrcene synthase, TPS-Cin - 1,8-cineole synthase, α-PS - α-pinene synthase, β-PS - β-pinene synthase, LS - limonene synthase.

"carbon flow" in the MEP pathway and an effective target for regulating the synthesis of terpenoids. 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ME-CPP*) gene expression was down-regulated in mature leaves of *E. camaldulensis* compared to young leaves (Table 1). Hydroxymethylglutaryl-CoA synthase (HMGCS) is the first important rate-limiting enzyme in the MVA pathway and also an important regulatory point of the MVA pathway. During terpenoid biosynthesis of young and mature leaves of *E. camaldulensis*, 3 *HMGCS* genes were down-regulated in mature leaves compared to young leaves (Table 1).

During terpenoid biosynthesis of young and mature leaves of *E. camaldulensis*, 18 differentially expressed terpene synthase (*TPS*) genes were also identified, and six of them were up-regulated and the others were down-regulated in mature leaves compared to young leaves. The expressions of *TPS8* and *TPS9* were down-regulated and significantly different from those in mature leaves compared with young leaves and predicted as related to limonene synthases (TPS8 and TPS9). The limonene content in young leaves of *E. camaldulensis* was detected, while it was not detected in mature leaves (Table 1 and Fig. 1). The expression of the *TPS10* gene in mature leaves was significantly different from that of young leaves and predicted that *TPS10* might be associated with myrcene synthase. The content of myrcene was not detected in young leaves, while it was detected in mature leaves (Table 1 and Fig. 1). The expressions of *TPS13*, *TPS14*, and *TPS15* genes were not significantly different in mature

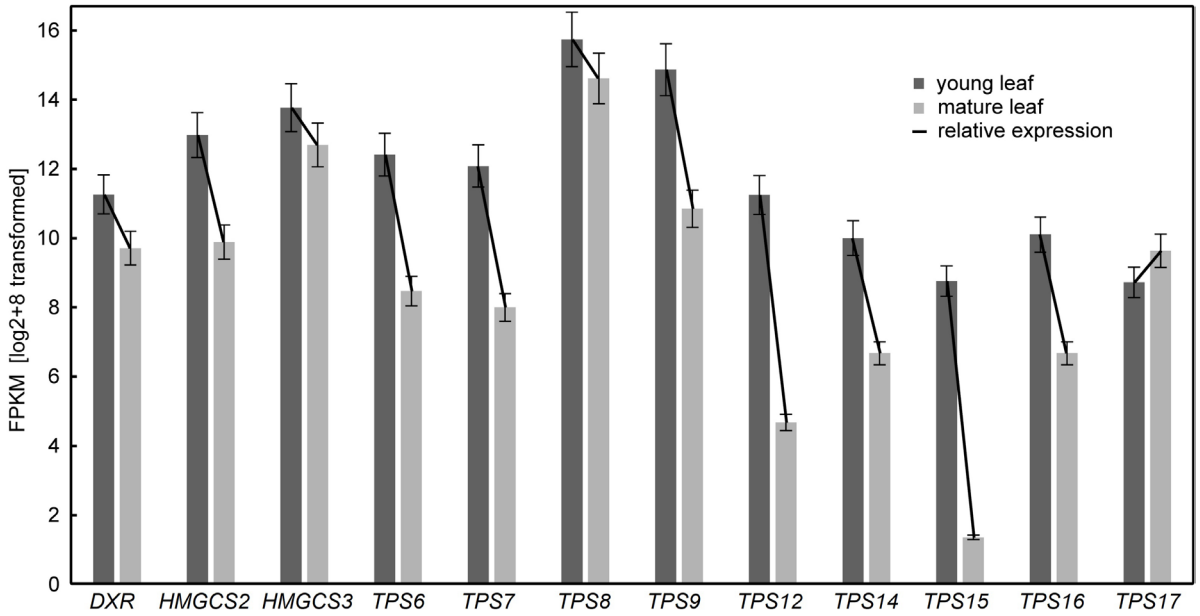

Fig. 5. Expressions of target genes in young and mature leaves of *E. camaldulensis*. Means ± SEs, n = 3; Y-axis represents gene expression; FPKM [log2+8 transformed] - FPMK is a measure of relative gene expression, log2+8 are homogenization data.

leaves compared with young leaves and we predicted that *TPS 13*, *TPS14*, and *TPS15* genes might be connected with α-pinene synthase. There was a non-significantly different content of α-pinene in young and mature leaves of *E. camaldulensis*. The *TPS16* gene was down-regulated in mature leaves of *E. camaldulensis* and predicted that the *TPS16* gene might encode β-pinene synthase. The content of β-pinene in young leaves was higher than that in mature leaves of *E. camaldulensis*. *TPS17* and *TPS18* genes were up-regulated and predicted that they encoded 1,8-cineole synthase. The expression of the *TPS18* gene was up-regulated and significantly different in young leaves from that in mature leaves, and the content of 1,8-cineole in mature leaves of *E. camaldulensis* was higher than that in young leaves (Table 1 and Fig. 1).

TFs regulate plant secondary metabolism by binding *cis*-acting elements to the promoters of terpenoid synthetic genes. It regulates the synthesis of secondary metabolites by activating or inhibiting the expression of downstream genes. In this study, WRKY, MYB, NAC, and bHLH TFs were selected and compared with the target genes, and their regulatory trends were consistent (if TFs were found to be up-regulated then their target genes were also up-regulated, and *vice versa*). WRKY, bHLH, NAC and MYB TFs may regulate the expression of *TPS* genes and promote or inhibit their expression.

Totally 12 target genes were randomly selected for validation of RNA-seq data by RT-qPCR (Fig. 5). The gene-specific primers were designed using *Primer Premier 6*. *DXR*, *HMGCS2*, and *HMGCS3* were more expressed in young leaves compared to mature leaves of *E. camaldulensis*. *TPS7*, *TPS8*, *TPS15,* and *TPS17* also exhibited differential expression patterns in young and mature leaves. Similar expressions were detected by RT-qPCR, which showed the reliability of our RNA sequencing data.

Discussion

With the application of GC-MS, HPLC-MS, GC-EAD, SPME, and other microchemical analysis instruments, and the rapid development of omics technology, molecular and biochemical studies of plant volatiles and metabolic engineering have become easier than before. At present, eucalyptus oil is mostly extracted from eucalyptus leaves by steam distillation and the chemical constituents are separated and identified by GC-MS (Leijs *et al*. 2005). The oil extracted from *E. camaldulensis* leaves is aromatic and volatile, and its composition is formed mainly by terpenes and its oxygen-containing derivatives, which mainly include sesquiterpenes (C15), diterpenes (C20), triterpenes (C30), tetraterpenes (C40), and polyterpenes (>C40) (Yang and Zeng 2005). Terpenes are highly diverse and 25 000 different terpenes have been identified, with thousands being more likely to exist (Gershenzon and Dudareva 2007). To further explore the genetic resources, the transcriptome sequencing of *Eucalyptus cloeziana* (Zhu *et al*. 2018), *E. grandis* × *E. tereticornis* (Xiao *et al.* 2020), *E. urophylla* and *E. dunnii* (Liu *et al*. 2014) has been performed. In this study, some functional genes related to terpenoid biosynthesis were annotated by analyzing the transcriptome of young and mature leaves of *E. camaldulensis*, which laid a foundation for the exploration and functional research of terpenoid genes in *E. camaldulensis*. TFs, by binding with *cis*-acting elements on the promoters of terpenoid synthetic genes, activate or inhibit the expression of downstream genes, thereby regulating the synthesis of secondary metabolites. Studies have shown that bHLH TFs regulate secondary metabolic pathways in plants, and the MYB TF family is involved in plant development, metabolic regulation, and plant hormone signal transduction (Ji *et al*. 2014). AP2 and bHLH TFs affect the isoprene metabolic pathway (Tang

et al. 2007, Zhang *et al*. 2012), and WRKY1 TF regulated the expression of the cadinene synthase gene (CAD1-A) (Xu *et al*. 2004).

Eucalyptus has the largest *TPS* gene family of plants that have been sequenced (Külheim *et al.* 2015, Butler *et al.* 2018), and we also detected many *TPS* genes in this study during terpenoid biosynthesis of young and mature leaves of *E. camaldulensis*. The expressions of *TPS6*, *TPS7*, *TPS8*, and *TPS9* in mature leaves were lower than in young leaves, and these results were verified by RT-qPCR. Combined with the results of limonene content, *TPS6*, *TPS7*, *TPS8*, and *TPS9* genes may play the leading role in limonene synthesis. The expression of the *TPS12* gene in mature leaves was lower than that in young leaves but the content of myrcene in mature leaves of *E. camaldulensis* was higher than in young leaves. WRKY and MYB TFs may inhibit the expression of the *TPS12* gene. The *TPS16* expression was down-regulated in mature leaves compared with young leaves of *E. camaldulensis*, and the content of β-pinene in young leaves was higher than that in mature leaves, which indicated that *TPS16* is a key gene for β-pinene synthesis. *TPS17* and *TPS18* genes exhibited differential expression patterns, and all of them were found to be up-regulated in mature leaves compared with young leaves of *E. camaldulensis*, and the content of 1,8-cineole in mature leaves of *E. camaldulensis* was higher than that in young leaves. These results indicated that *TPS17* and *TPS18* genes are the key genes in 1,8-cineole synthesis.

Cloning the key genes related to the terpenoid synthesis pathway helps to enrich the theory of terpenoid biosynthesis and regulation. Lücker *et al*. (2010) have successfully transferred 3 different lemon monoterpene synthase genes into the wild tobacco by genetic engineering combined with hybridization. Ohara *et al*. (2010) successfully transferred limonene synthase cDNA into *E. camaldulensis* through genetic engineering. The significant genome-wide association studies (GWAS) found in a cluster of *TPS* genes on chromosome 1 led to the successful functional validation of these *TPS* genes, which produced 1,8-cineole and, to a lesser extent, a-pinene (Kainer *et al.* 2019). At present, great progress has been made in the study of terpenoid synthases and their genes in plants and the metabolic network of plants and the regulation of terpenoids have been known by using modern molecular biology. The next step is to clone the key enzyme genes of terpenoid biosynthesis and isolate TFs regulating terpenoid metabolic pathways through genetic engineering.

Conclusions

This study used transcriptome analysis to investigate the regulation of terpenoid biosynthesis genes during leaf development of *E. camaldulensis*. A series of target genes displayed significant differences in the expression between young and mature leaves. The WRKY, MYB, NAC, and bHLH TFs were also selected and compared with the target genes. Our study will contribute to the research of terpenoid biosynthesis engineering, and improve the knowledge of the expression of genes concerning enzymes leading to the synthesis of precursor and final terpenoids, which could be helpful to improve the quality of *E. camaldulensis* plants. Data has been submitted to the SRA (sequence read archive) database of NCBI (https://www.ncbi.nlm.nih.gov). The accession numbers are SAMN11931322 and SAMN11931323.

References

Bohlmann, J., Crock, J., Croteau, J.R.: Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-α-bisabolene synthase from grand fir (*Abies grandis*). - Proc. nat. Acad. Sci. USA **95**: 6756-6761, 1998b.

Bohlmann, J., Croteau, R.: Diversity and variability of terpenoid defences in conifers: molecular genetics, biochemistry and evolution of the terpene synthase gene family in grand fir (*Abies grandis*). - Novartis Found. Symp. **223**: 132-145, 1999.

Bohlmann, J., Meyer-Gauen, G., Croteau, R.: Plant terpenoid synthases: molecular biology and phylogenetic analysis. - Proc. nat. Acad. Sci. USA **95**: 4126-4133, 1998a.

Boland, D.J., Brophy, J.J., House, A.P.N.: Eucalyptus Leaf Oils: Use, Chemistry, Distillation and Marketing. - Inkata Press, Melbourne 1991.

Bounatirou, S., Smiti, S., Miguel, M.G., Faleiro, L., Rejeb, M.N., Neffati, M., Costa, M.M., Figueiredo, A.C., Barroso, J.G., Pedro, L.G.: Chemical composition, antioxidant and antibacterial activities of the essential oils isolated from Tunisian *Thymus capitatus* Hoff. *et* Link. - Food Chem. **105**: 146-155, 2007.

Butler, J.B., Freeman, J.S., Potts, B.M., Vaillancourt, R.E., Grattapaglia, D., Silva-Junior, O.B., Simmons, B.A., Healey, A.L., Schmutz, J., Barry, K.W., Lee, D.J., Henry, R.J., King, G.J., Baten, A., Shepherd, M.: Annotation of the Corymbia terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. - Heredity **121**: 87-104, 2018.

Calvert, J., Baten, A., Butler, J., Barkla, B., Shepherd, M.: Terpene synthase genes in *Melaleuca alternifolia*: comparative analysis of lineage-specific subfamily variation within Myrtaceae. - Plant Syst. Evol. **304**: 111-121, 2018.

Chang, Y.T., Chu, F.H.: Molecular cloning and characterization of monoterpene synthases from *Litsea cubeba* (Lour.) Persoon. - Tree Genet. Genomes **7**: 835-844, 2011.

Cong, P.Z., Su, K.M.: Handbook of Analytical Chemistry: Mass Spectrometry. - Chemical Industry Press, Beijing 2000.

Cordoba, E., Salmi, M., León, P.: Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. - J. exp. Bot. **60**: 2933, 2009.

Dudareva, N., Pichersky, E., Gershenzon, J.: Biochemistry of plant volatiles. - Plant Physiol. **135**: 1893-1902, 2004.

Facchini, P.J., Chappell, J.: Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. - Proc. nat. Acad. Sci. USA **89**: 11088-11092, 1992.

Gershenzon, J., Dudareva, N.: The function of terpene natural products in the natural world. - Nat. chem. Biol. **3**: 408-414, 2007.

Gong, J., Ni, S.F., Luo, R.F., Ren, Y.P., Wu, Y.F., Zhao, G.F.: Progress on ingredient of edible flavor. - China Condiment **34**: 31-33, 2009.

Greenhagen, B., Chappell, J.: Molecular scaffolds for chemical wizardry: learning nature's rules for terpene cyclases. - Proc. nat. Acad. Sci. USA **98**: 13479-13481, 2001.

Hofberger, J.A., Ramirez, A.M., Bergh, E., Zhu, X., Bouwmeester, H.J., Schuurink, R.C., Schranz, M.E.: Large-scale evolutionary analysis of genes and supergene clusters from terpenoid modular pathways provides insights into metabolic diversification in flowering plants. - PLoS ONE **10**: e0128808, 2015.

Hosseini, S.V., Al-Qanbar, M.H., Khazraei, H., Khodaei, S., Iranpour, P.: Evaluation the effects of eucalyptus essential oil and hypertonic saline as scolicidal agents in induction of sclerosing cholangitis in rabbits. - Adv. biomed. Res. **1**: 9, 2020.

Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J., Köllner, T.G.: Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). - BMC Plant Biol. **14**: 270, 2014.

Ji, Y., Xiao, J., Shen, Y., Ma, D.M., Li, Z.Q., Pu, G.B., Li, X., Huang, L.L., Liu, B.Y., Ye, H.C., Wang, H.: Cloning and characterization of AabHLH1, a bHLH transcription factor that positively regulates artemisinin biosynthesis in *Artemisia annua*. - Plant Cell Physiol. **55**: 1592-1604, 2014.

Kainer, D., Padovan, A., Degenhardt, J., Krause, S., Mondal, P., Foley, W.J., Külheim, C.: High marker density GWAS provides novel insights into the genomic architecture of terpene oil yield in Eucalyptus. - New Phytol. **223**: 1489-1504, 2019.

Külheim, C., Padovan, A., Hefer, C., Krause, S.T., Köllner, T.G., Myburg, A.A., Degenhardt, J., Foley, W.J.: The *Eucalyptus* terpene synthase gene family. BMC Genomics, **16**: 1-18, 2015.

Leijs, H., Broekhans, J., Pelt, L.V., Mussinan, C.: Quantitative analysis of the 26 allergens for cosmetic labeling in fragrance raw materials and perfume oils. - J. Agr. Food Chem. **53**: 5487-5491, 2005.

Livak, K.J., Schmittgen, T.D.: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) - Methods **25**: 402-408, 2001.

Liu, Y., Jiang, Y., Lan, J., Zou, Y., Gao, J.: Comparative transcriptomic analysis of the response to cold acclimation in *Eucalyptus dunnii*. - PLoS ONE **9**: e113091, 2014.

Lockhart, D.J., Winzeler, E.A.: Genomics, gene expressand DNA arrays. - Nature **405**: 827-836, 2000.

Lücker, J., Schwab, W., Franssen, M.C.R., Plas, L.H.W.V.D., Bouwmeester, H.J., Verhoeven, H.A.: Metabolic engineering of monoterpene biosynthesis: two-step production of (+)-trans-isopiperitenol by tobacco. - Plant J. **39**: 135-145, 2010.

Mahmoud, S.S., Croteau, R.B.: Strategies for transgenic manipulation of monoterpene biosynthesis in plants. - Trends Plant Sci. **7**: 366-373, 2002.

Ohara, K., Matsunaga, E., Nanto, K., Yamamoto, K., Sasaki, K., Ebinuma, H., Yazaki, K.: Monoterpene engineering in a woody plant *Eucalyptus camaldulensis* using a limonene synthase cDNA. - Plant biotechnol. J. **8**: 28-37, 2010.

Perez-Mar, T. A., Diaz-Espejo, A. : Regulation of photosynthesis and stomatal and mesophyll conductance under water stress and recovery in olive trees: correlation with gene expression of carbonic anhydrase and aquaporins. - J. exp. Bot. **65**: 3143-3156, 2014.

Santadino, M., Lucia, A., Duhour, A., Riquelme, M., Naspi, C., Masuh, H., Liljesthröm, G., Coviella, C.: Feeding preference of *Thaumastocoris peregrinus* on several *Eucalyptus* species and the relationship with the profile of terpenes in their essential oils. - Phytoparasitica **2**: 1-12, 2017.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L., Pichersky, E.: Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. - Proc. nat. Acad. Sci. USA **106**: 10865-10870, 2009.

Steele, C.L., Crock, J., Bohlmann, J., Croteau, R.: Sesquiterpene synthases from grand fir (*Abies grandis*): comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ-selinene synthase and γ-humulene synthase. - J. biol. Chem. **273**: 2078-2089, 1998b.

Steele, C.L., Katoh, S., Bohlmann, J., Croteau, R.: Regulation of oleoresinosis in grand fir (*Abies grandis*): differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. - Plant Physiol. **116**: 1497-1504, 1998a.

Tang, M., Sun, J., Liu, Y., Chen, F., Shen, S.: Isolation and functional characterization of the *JcERF* gene, a putative AP2/EREBP domain-containing transcription factor, in the woody oil plant *Jatropha curcas*. - Plant mol. Biol. **63**: 419-428, 2007.

Tholl, D., Lee, S.: Terpene specialized metabolism in *Arabidopsis thaliana*. - Arabidopsis Book **9**: e0143, 2011.

Tian, Y.H., Liu, X.M., Zhou, Y.H., Qin, R.H.: Composition of essential oils of leaves from *Eucalyptus camadulensis* and *Eucalyptus benthami*. - Fine Chem. **22**: 920-923, 2005.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R.: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. - Nat. Protoc. **7**: 562-578, 2012.

Velho, M.C., Cossetin, L.F., Godoi, S.N.D., Santos, R.C.V., Ourique, A.F.: Nanobiopesticides: development and insecticidal activity of nanoemulsions containing lemongrass or eucalyptus oils. - Nat. Prod. Res. **2**: 1-6, 2020.

Wildung, M.R., Croteau, R.: A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. - J. biol. Chem. **271**: 9201-9204, 1996.

Xiao, Y., Li, J., Zhang, Y., Zhang, X.N., Liu, H.L., Qin, Z.H., Chen, B.W.: Transcriptome analysis identifies genes involved in the somatic embryogenesis of Eucalyptus. - BMC Genomics **21**: 1-13, 2020.

Xu, Y.H., Wang, J.W., Wang, S., Wang, J.Y., Chen, X.Y.: Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-δ-cadinene synthase-A1. - Plant Physiol. **135**: 507-515, 2004.

Yahyaa, M., Ibdah, M., Marzouk, S., Ibdah, M.: Profiling of terpene metabolome in carrot fruits of wild (*Daucus carota* L. ssp. *carota*) accessions and characterization of a geraniol synthase. - J. Agr. Food Chem. **66**: 2378-2386, 2018.

Yang, T., Zeng Y.: Plant terpenoid synthases: molecular biology and genetic engineering. - Acta bot. yunnan. **27**: 1-10, 2005.

Yu, F., Utsumi, R.: Diversity, regulation, and genetic manipulation of plant mono - and sesquiterpenoid biosynthesis. - Cell. Mol. Life Sci. **66**: 3043-3052, 2009.

Zhang, H.B., Bokowiec, M.T., Rushton, P.J., Han, S.C., Timko, M.P.: Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. - Mol. Plant **5**: 73-84, 2012.

Zhu, L.S., Liu, G., He, S.E., Li, H., Liu, X.F., Chen, S.X.: Transcriptome characterization analysis of *Eucalyptus cloeziana* root based on Illumina HiSeq 2000 sequencing technology. - Mol. Plant Breed. **16**: 4245-4254, 2018.