

SMRT sequencing of full-length transcriptome and gene expression analysis in two chemical types of *Pogostemon cablin* (Blanco) Benth.

Hongyi Zhang, Wenjing Deng, Changhua Lu, Mengling He and Hanjing Yan

College of Traditional Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou, Guangdong, China

ABSTRACT

Background: *Pogostemon cablin* (Blanco) Benth. also called patchouli, is a traditional medicinal and aromatic plant that grows mainly in Southeast Asia and China. In China, *P. cablin* is divided into two chemical types: the patchouliol-type and the pogostone-type. Patchouliol-type patchouli usually grow taller, with thicker stems and bigger leaves, and produce more aromatic oil.

Methods: To better understand the genetic differences between the two chemical types that contribute to their differences in morphology and biosynthetic capabilities, we constructed *de novo* transcriptomes from both chemical types using the Pacific Biosciences (PacBio) Sequel platform and performed differential expression analysis of multiple tissues using Illumina short reads.

Results: In this study, using single-molecule real-time (SMRT) long-read sequencing, we obtained 22.07 GB of clean data and 134,647 nonredundant transcripts from two chemical types. Additionally, we identified 126,576 open reading frames (ORFs), 100,638 coding sequences (CDSs), 4,106 long noncoding RNAs (lncRNAs) and 6,829 transcription factors (TFs) from two chemical types of *P. cablin*. We adopted PacBio and Illumina sequencing to identify differentially expressed transcripts (DEGs) in three tissues of the two chemical types. More DEGs were observed in comparisons of different tissues collected from the same chemical type relative to comparisons of the same tissue collected from different chemical types. Furthermore, using KEGG enrichment analysis of DEGs, we found that the most enriched biosynthetic pathways of secondary metabolites of the two chemical types were “terpenoid backbone biosynthesis”, “phenylpropanoid biosynthesis”, “plant hormone signal transduction”, “sesquiterpenoid and triterpenoid biosynthesis”, “ubiquinone and other terpenoid-quinone biosynthesis”, “flavonoid biosynthesis”, and “flavone and flavonol biosynthesis”. However, the main pathways of the patchouliol-type also included “diterpene biosynthesis” and “monoterpene biosynthesis”. Additionally, by comparing the expression levels of the three tissues verified by qRT-PCR, more DEGs in the roots were upregulated in the mevalonate (MVA) pathway in the cytoplasm, but more DEGs in the leaves were upregulated in the methylerythritol phosphate (MEP) pathway in the plastid, both of which are important pathways for terpenoids

Submitted 4 June 2021
Accepted 24 January 2022
Published 22 February 2022

Corresponding author
Hanjing Yan,
yanhanjing1211@163.com

Academic editor
Yuriy Orlov

Additional Information and
Declarations can be found on
page 18

DOI [10.7717/peerj.12940](https://doi.org/10.7717/peerj.12940)

© Copyright
2022 Zhang et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

biosynthesis. These findings promote the study of further genome annotation and transcriptome research in *P. cablin*.

Subjects Bioinformatics, Genomics, Molecular Biology, Plant Science

Keywords *Pogostemon cablin* (Blanco) Benth., SMRT, Transcriptome, Two chemical types, Terpenoid biosynthesis

INTRODUCTION

Pogostemon cablin (Blanco) Benth., called patchouli in English, is an economically important plant in the family Lamiaceae and is widely cultivated in the tropical and subtropical areas of Asia (Wu *et al.*, 2010). In China, *P. cablin* is valued for its well-known medicinal properties and used to treat the common cold, headache, fever, vomiting, indigestion, and diarrhea (Chinese Pharmacopoeia Commission, 2020). A crucial essential oil (called patchouli oil) can be produced from patchouli leaf extract. There are more than 24 different sesquiterpenes in patchouli oil, such as patchouli alcohol, eugenol, α -bulnesene, rosmarinic acid, and others (Deguerry, Pastore & Wu, 2006; Hsu *et al.*, 2006). More than 50% of patchouli oil is patchouli alcohol (Miyazawa *et al.*, 2000). Patchouli volatile oil is an important ingredient for perfumes, incense, soaps, and other cosmetic products (Fan *et al.*, 2011; Chen *et al.*, 2013). In China, there are two chemical types, called patchouliol-type patchouli and pogostone-type patchouli. Patchouliol-type patchouli contains a patchouli alcohol content that is much higher than the pogostone content. If the patchouli alcohol content is equal to or less than the pogostone content in the stems and leaves, it is called pogostone-type patchouli (Luo *et al.*, 2003). Patchouliol-type and pogostone-type patchouli also exhibit differences in chemical composition (Hu *et al.*, 2019), morphology (Li *et al.*, 2002) and genetics (Liu *et al.*, 2021).

Patchouli alcohol, the major component in patchouli oil, is a sesquiterpene. Its biosynthesis includes terpenoid backbone biosynthesis and sesquiterpenoid biosynthesis pathways. The synthesis of plant terpenoids involves two main routes: the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP) pathway, with the former located in the cytoplasm and the latter in plastids. Both pathways can synthesize isopentenyl pyrophosphate (IPP), which is the central precursor of the synthesis of all terpene compounds (Sallaud *et al.*, 2009). The direct precursor of sesquiterpene, farnesyl diphosphate (FPP), is derived from IPP. Does the MVA pathway or MEP pathway play a more important role for sesquiterpene synthesis? The answer could be important for increasing the content of patchouli alcohol by the biometabolic engineering pathway.

In recent years most transcriptomes are generated on the Illumina platform using Next-generation sequencing (NGS) technology (Chen, Qiu & Yang, 2019). The first *de novo* assembled 1.15 Gb draft genome sequence for *P. cablin* were performed using NGS technology in 2016 (He *et al.*, 2016). However, NGS sequencing reads are short, which makes it difficult to perform subsequent bioinformatics analyses, such as alternative splicing, assembly, and annotation (Tilgner *et al.*, 2015; Koren *et al.*, 2012). Thus,

single-molecule real-time (SMRT) sequencing (Pacific Biosciences of California, Inc., Menlo Park, CA, USA) was developed, which overcomes the limitation of short-read sequences by enabling the generation of kilobase-sized sequencing reads (Jia *et al.*, 2018). This makes the full-length transcriptome particularly important for basic and applied research on gene function, gene expression, and regulation. In 2019 the first full-length transcriptome of *P. cablin* were analyzed by using SMRT sequencing and 102 transcripts were annotated as 16 encoding enzymes involved in patchouli alcohol biosynthesis (Chen *et al.*, 2019). With the help of highly accurate NGS short reads correction and circular consensus sequence (CCS) reads self-correction, SMRT sequencing achieves results with >99% accuracy (Qian *et al.*, 2019). Therefore, sequencing approaches that combine SMRT and short-read sequencing for transcriptome assembly and differential expression analysis provides a highly accurate assembly that can be used for downstream evolutionary and gene function analysis.

In this experiment, we used SMRT sequencing and NGS to perform transcriptome sequencing of two chemical types, namely, the patchouliol-type and the pogostone-type. Based on the transcriptome data, we obtained transcript functional annotation, predicted open reading frames (ORFs), and analyzed alternative splicing (AS), transcription factors (TFs) and long noncoding RNAs (lncRNAs). In addition, the transcriptomes of the two chemical types will help to further understand the detailed biosynthesis mechanisms of patchouli alcohol and other terpenoids in *P. cablin*.

MATERIALS AND METHODS

Plant materials and RNA preparation

The samples of *P. cablin* were cultivated in Sihui (N23°29'53", E112°46'57") and Gaoyao (N22°55'92", E112°28'38"), Guangdong Province, China. The pogostone-type germplasm came from Shipai, Guangzhou City about 70 years ago and has been cultivated in Gaoyao. Nowadays the pogostone-type germplasm is very rare in the markets. The patchouliol-type germplasm from Sihui is a mainstream product in the markets and is large-scale cultivated in Guangdong and Hainan Province. Two chemical types were cultivated in Sihui and Gaoyao respectively and got the same field management such as weeding, watering and fertilizing in the period from transplanting to harvesting. Patchouli harvesting day was also the cutting breeding day. The apical stem cuttings of two germplasms were planted in Sihui and Gaoyao respectively on November 18, 2017. On March 15, 2018, the cutting seedlings were transplanted to the medicine garden of Guangdong Pharmaceutical University in Guangzhou on the same day. During the growth period, all the samples were given the same field management including weeding and watering. On November 7, 2018, the mature and healthy leaves and corresponding stems were collected in liquid nitrogen. The roots were cleaned up and also collected as samples. Three biological plant replicates per experimental treatment were harvested. All samples were immediately frozen in liquid nitrogen and stored in a -80 °C freezer until RNA isolation. The samples from Sihui belong to the patchouliol-type, while the samples from Gaoyao belong to the pogostone-type. No specific permits were required and no protected species were involved in this study.

Total RNA samples were isolated from each sample using an RNAPrep Pure Plant kit (TIANGEN Biotech, Beijing, China) and treated with DNase I to remove DNA contaminants (TIANGEN) according to the manufacturer's instructions. The RNA quantity and quality were assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). RNA concentration was measured using a Qubit® RNA Assay Kit in a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

Library preparation and SMRT sequencing

The RNAs of three tissue samples of the two chemical types, including roots, stems and leaves, were pooled at equal ratios, respectively. The SMART sequencing library was prepared according to the Iso-Seq protocol as described by Pacific Biosciences. Full length cDNAs were synthesized using the SMRTer cDNA synthesis kit and fragments of 1–6 kb in size were collected using a BluePippin instrument. Two cells (F01 and F02) were run on the PacBio RSII platform. The resulting library was sequenced using the Iso-Seq function of the PacBio RS II system (Pacific Biosciences, Menlo Park, CA, USA) ([Wang et al., 2019](#)).

Illumina RNA-seq library construction and sequencing

Total RNA from 18 samples of three tissues of the two chemical types (three biological replicates with each tissue) was used to construct RNA-seq libraries. First, mRNA was interrupted to short fragments after fragmentation buffer was added. Then the first strand cDNA was synthesized with the resultant mRNA fragments as templates. The second strand cDNA was synthesized by using RNase H and DNA polymerase I. After end repair and A-tailing, the purified double-stranded cDNA fragments were ligated with paired-end adaptors, and then used AMPure XP beads to select fragments size larger than 100 bp. Then the fragments of appropriate size were amplified by PCR to obtain the final cDNA libraries ([Dong et al., 2015](#)). The quality of the libraries was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Finally, the 18 libraries were sequenced using the Illumina HiSeq™ 2500 system (Illumina, San Diego, CA, USA) ([Deng et al., 2018](#)).

PacBio read error correction

Raw SMRT sequencing reads were processed by removing read fragments whose lengths were less than 50 bp and accuracy was less than 80% to retain only the highest quality reads for downstream analysis. The reads of insert (ROIs) with both the 5' and 3' primer sequences and a poly(A) tail present were considered to be full-length transcripts. Chimeric sequences were identified and removed using UCHIME v4.2.40 ([Edgar et al., 2011](#)). The non-chimeric sequences in the full-length sequence are called full-length non-chimeric (FLNC) sequences. The iterative clustering for error correction (ICE) algorithm was used to obtain consensus sequences for isoforms. Sequences were then error corrected using Quiver v1 ([Chin et al., 2013](#)). The Proovread software v2.13.841 ([Hackl et al., 2014](#)) corrected the low-quality full-length transcript with its corresponding Illumina

RNA seq data. The 5' terminal regions of transcripts may be degraded, causing reads derived from the same isoform to be divided into different clusters, leading to the occurrence of redundant sequences in the dataset. Therefore, CD-HIT (parameters: -c 0.95 -s 0.99) was used to reduce redundancy among the FL transcripts. Finally, BUSCO v3.0.2 analysis was performed using *embryophyta_odb10* dataset (Simão *et al.*, 2015) to evaluate the integrity of the dereplicated transcriptome. The PacBio SMRT raw reads and the Illumina short reads have been submitted to the BioProject database of the National Center for Biotechnology Information have been submitted to the BioProject database of National Center for Biotechnology Information (accession numbers [PRJNA606892](#) and [PRJNA607886](#), respectively).

Identification of ORFs, TFs and lncRNAs

TransDecoder v5.5.0 (Haas *et al.*, 2013) was used to identify candidate protein coding regions with default parameters. ORFs of 200 nt in length or greater and those that had detectable Pfam (protein family) (Finn *et al.*, 2014) domains (database downloaded November 2018) were retained for further analysis.

The TFs and transcript regulators (TRs) were predicted with iTAK software (Zheng *et al.*, 2016) from putative protein sequences.

lncRNAs are an important component of the transcriptome. The most widely used methods for analyzing lncRNAs are CPC (Coding Potential Calculator) (Kong *et al.*, 2007), CNCI (Coding-Non-Coding Index) (Sun *et al.*, 2013), CPAT (Coding Potential Assessment Tool) (Wang *et al.*, 2013) and Pfam protein structure domain analysis. Putative protein-coding RNAs were removed from the dataset. Transcripts of >200 nt in length predicted to contain more than two exons were considered candidate lncRNAs and were further screened using CPC/CNCI/CPAT/Pfam. HMMscan searches were first used to identify transcripts that contained pfam motifs. Transcripts that had coding potential, but did not have detectable pfam motifs, were considered potential lncRNAs and were further analyzed using CPC, CNCI, and CPAT. In this study, the results of all four analyses had to agree in order for a transcript to be annotated as a potential lncRNAs.

Functional annotation of transcripts

Functional annotation of transcripts was performed using BLAST (version 2.2.26) (Altschul *et al.*, 1997) to compare transcripts against the NCBI nonredundant protein sequences (NR) (Deng *et al.*, 2006), Swissprot (Bairoch & Boeckmann, 1991), Gene Ontology (GO) (Ashburner *et al.*, 2000), Clusters of Orthologous Groups of proteins (COG) (Tatusov *et al.*, 2000), euKaryotic Ortholog Groups (KOG) (Koonin *et al.*, 2004), Pfam evolutionary genealogy of genes: Nonsupervised Orthologous Groups (eggNOG) (Jensen *et al.*, 2008), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata *et al.*, 1999) databases with an E-value threshold of 10^{-5} . The alignments against the NR database were used in Blast2GO to obtain GO annotations. The KEGG pathway assignments were performed using KOBAS software (Mao *et al.*, 2005).

Screening of differentially expressed transcripts

The clean reads from Illumina RNA-Seq for 18 samples (three biological replicates of roots, stems, leaves of the pogostone-type and the patchouliol-type respectively) were separately mapped to the non-redundant SMRT reference by using Bowtie2 v2.3.4.2 (Langmead & Salzberg, 2012). Quantifications of the transcript expression values were determined by RSEM software (Li & Dewey, 2011) and then normalized to FPKM (fragments per kilobase of transcript per million mapped reads). The DESeq R package (1.10.1) was used to analyze gene expression differences between groups to identify differentially expressed genes (DEGs) (Anders & Huber, 2010). The resulting p-values were adjusted using Benjamini & Hochberg's method to control the false discovery rate (FDR). DEGs were defined as by parameters of FDR < 0.01 and the absolute value of the log₂ ratio ≥ 1. The Venn diagrams of DEGs in three tissues of two chemical types was illustrated using the online Venn diagram production website jvenn (<http://jvenn.toulouse.inra.fr/app/index.html>).

Quantitative real-time PCR (qRT-PCR)

To validate transcript abundance as reflected by FPKM values for the assembled transcripts, qRT-PCR was performed to estimate relative expression levels with a SYBR® Premix Pro Taq HS qPCR Kit (Accurate Biotechnology Company, Guntur, Andhra Pradesh) on a CFX96 Real-Time PCR system (Bio-Rad, Hercules, CA, USA). Gene-specific primers were designed with NCBI and are described in File S1. The conditions for qRT-PCR amplification were as follows: incubation at 95 °C for 30 s, 40 cycles of 95 °C for 5 s and 60 °C for 30 s. The specificity of the primer amplicons was tested using melting curve analysis, and PCR products were verified by sequencing (Chen et al., 2019). The relative expression levels of the selected genes were normalized to the selected reference gene 18S rRNA and determined by the $2^{-\Delta\Delta Ct}$ method. The amplification contained three technical replicates and three biological replicates. The GraphPad Prism 8 software was used to analyze statistical significance. To analyze variance between the gene expressions in the different tissues of samples, one-way ANOVA (with Tukey *post hoc*) was used. A level of $p < 0.05$ was considered statistically significant.

RESULTS

SMRT sequencing data output

Based on PacBio SMRT sequencing technology, clean reads of 11.31 Gb and 10.76 Gb were obtained from the two cells (F01: pogostone-type patchouli; F02: patchouliol-type patchouli). With a full pass ≥ 0 and quality > 0.80, 574,817 and 420,978 ROIs were obtained, respectively (Table 1). In addition, 334,683 and 276,811 FLNC (full-length nonchimeric) read sequences were identified, respectively (Table 2).

Transcript clustering analysis

A total of 169,641,152,493 consensus isoforms were obtained from F01 and F02, including 95,298 and 92,051 high-quality isoforms and 74,216 and 60,333 low-quality isoforms, respectively. The ICE clustering results were shown in Table 3. Dereplication analysis of

Table 1 Reads of insert (ROI) statistics.

	F01	F02
cDNA size	1–6K	1–6K
Reads of insert	574,817	420,978
Read bases of insert	1,128,659,746	840,369,967
Mean read length of insert	1,963	1,996
Mean read quality of insert	0.93	0.94
Mean number of passes	9	11

Table 2 Full-length sequences statistics.

	F01	F02
Reads of insert	574,817	420,978
Number of five prime reads	393,474	322,245
Number of three prime reads	408,234	328,063
Number of poly-A reads	404,257	324,699
Number of filtered short reads	21,001	13,952
Number of non-full-length reads	215,392	126,856
Number of full-length reads	338,424	280,170
Number of full-length non-chimeric reads	334,683	276,811
Average full-length non-chimeric read length	1,846	1,965

Table 3 Results of iterative clustering for error correction (ICE) clustering analysis.

	F01	F02
Number of consensus isoforms	169,641	152,493
Average consensus isoforms read length	2,028	2,147
Number of polished high-quality isoforms	95,298	92,051
Number of polished low-quality isoforms	74,216	60,333
Percent of polished high-quality isoforms (%)	56.18%	60.36%

high-quality transcripts and corrected transcripts was performed using CD-HIT to obtain 134,647 transcript sequences.

Prediction of coding sequences

A total of 126,576 ORFs were identified by using TransDecoder, including 100,638 complete ORFs. The distribution of the coding sequence length of the complete ORF is shown in Fig. 1.

Functional annotation of transcripts

A total of 52,511 transcripts were annotated in the COG; 75,289 annotations in the GO database; 55,168 in the KEGG database; 80,507 in the KOG database; 104,918 in the Pfam

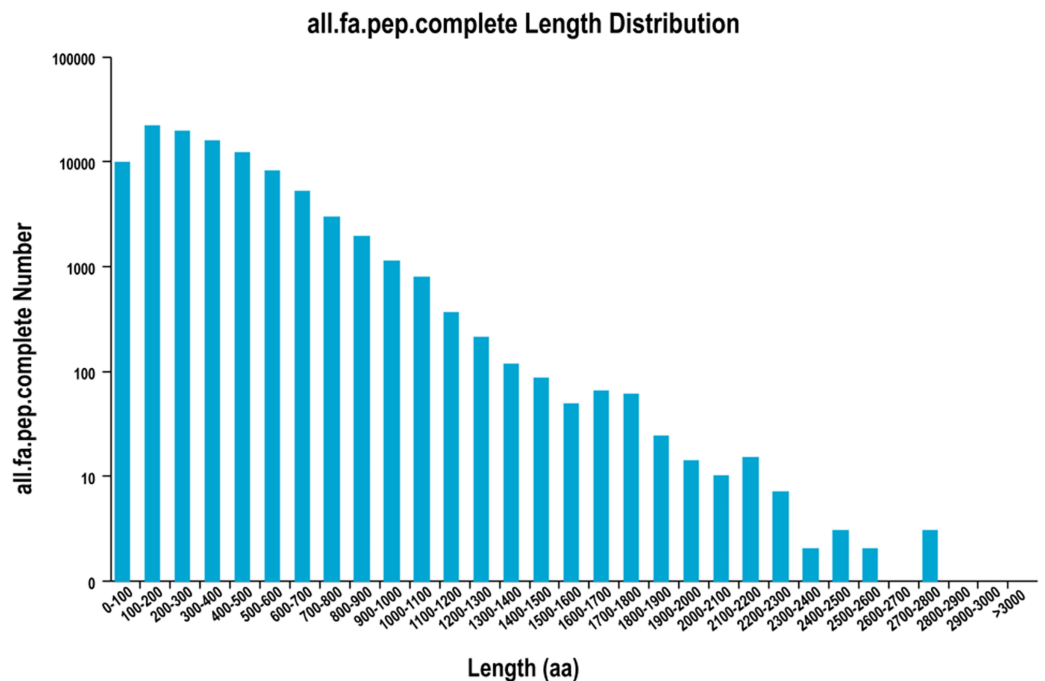


Figure 1 The distribution of the coding sequence lengths of the complete open reading frames.

Full-size  DOI: [10.7717/peerj.12940/fig-1](https://doi.org/10.7717/peerj.12940/fig-1)

Table 4 Annotation of transcripts using public databases.

Annotated databases	Isoform Number	Percentage
COG	52,511	40.94%
GO	75,289	58.70%
KEGG	55,168	43.01%
KOG	80,507	62.77%
Pfam	104,918	81.80%
Swiss-Prot	94,944	74.02%
eggNOG	123,200	96.05%
NR	126,527	98.65%
All	128,264	100%

database; 94,944 in the Swiss-Prot database; 123,200 in the eggNOG database; and 126,527 in the NR database. In addition, a total of 128,264 transcripts were annotated in the eight databases (Table 4 and File S2).

GO annotation

In this study, 75,289 transcripts were assigned GO terms, which were divided into three main categories. “Cell fraction” (GO: 0004464) and “cell” (GO: 0005623) dominated in the cellular component category. A large number of genes belonged to the molecular catalytic category of “catalytic activity” (GO: 0003824) and “binding” (GO: 0005488).

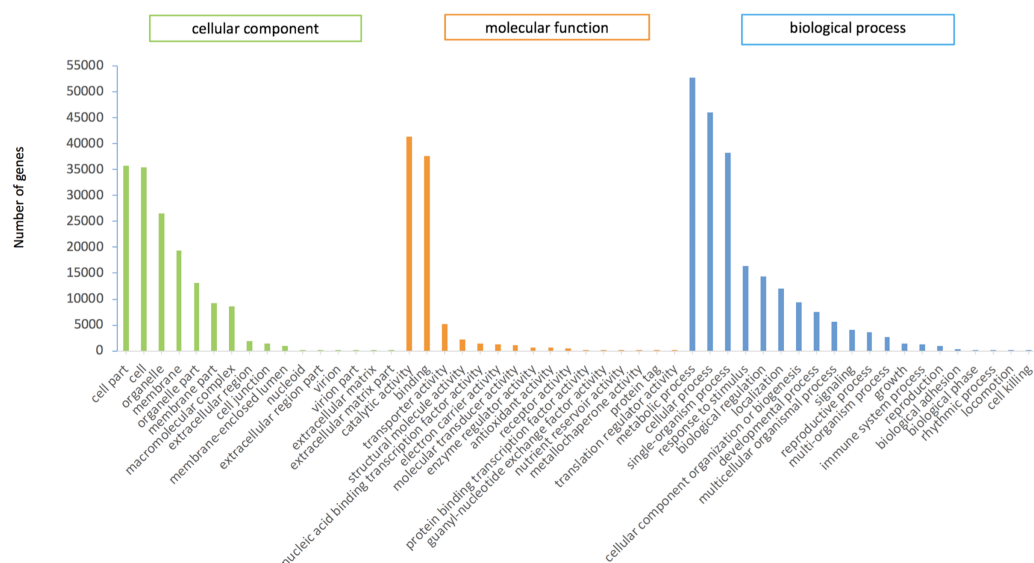


Figure 2 GO functional annotation of *P.cablin* transcripts. The x-axis represents GO categories; the y-axis represents the number of genes. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674_img.jpg\) DOI: 10.7717/peerj.12940/fig-2](https://doi.org/10.7717/peerj.12940/fig-2)

Table 5 The transcripts related to terpenoid biosynthesis.

Secondary metabolites biosynthesis pathways	Transcript numbers	Pathway ID
Terpenoid backbone biosynthesis	566	ko00900
Monoterpenoid biosynthesis	44	ko00902
Diterpenoid biosynthesis	103	ko00904
Sesquiterpenoid and triterpenoid biosynthesis	139	ko00909
Ubiquinone and other terpenoid-quinone biosynthesis	234	ko00130
Porphyrin and chlorophyll metabolism	319	ko00860
Limonene and pinene degradation	98	ko00903
Steroid biosynthesis	192	ko00100
Total	1695	

The main subgroups of biological process category were “metabolic process” (GO: 0008152) and “cellular process” (GO: 0009987) (Fig. 2).

KEGG annotation

KEGG pathway mapping of transcriptome assembly resulted in transcripts assigned to KEGG pathways. We found 1,695 transcripts related to terpenoid biosynthesis (Table 5 and File S3). Among them, 139 were annotated in the sesquiterpene and triterpene biosynthesis pathways, and 566 were annotated in the terpene skeleton biosynthesis pathway.

lncRNA prediction

In this study, 4,106 lncRNA transcripts were predicted by CNCI, CPC, CPAT and Pfam (File S4). The length of lncRNAs varied from 280 bp to 8,941 bp, with 84.46% of lncRNAs having a length $\leq 2,000$ bp. The mean length was 1,084 bp (Fig. 3).

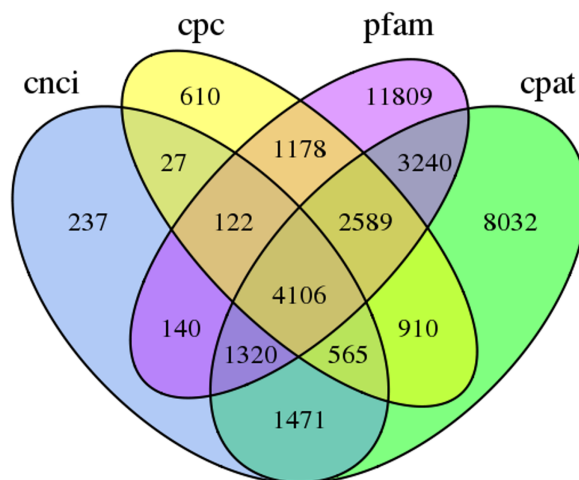


Figure 3 Venn diagram of the number of lncRNAs predicted by CPC, CNCI, CPAT and pfam protein structure domain analysis. Full-size [DOI: 10.7717/peerj.12940/fig-3](https://doi.org/10.7717/peerj.12940/fig-3)

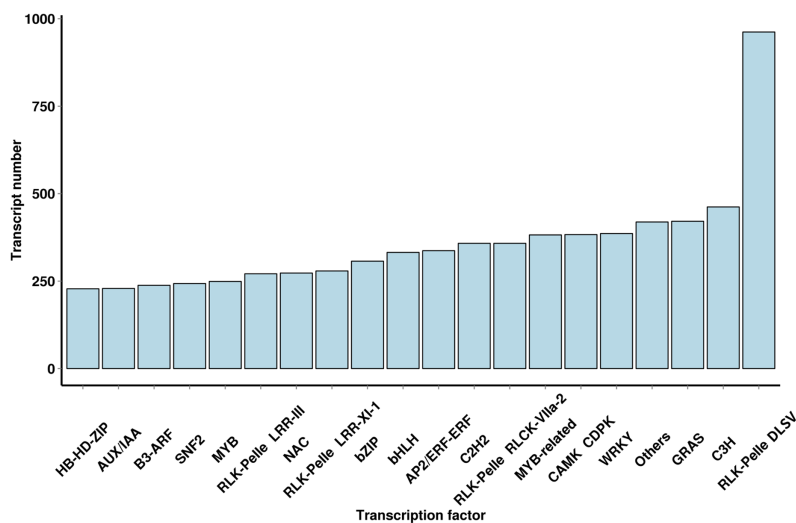


Figure 4 Number and family of top 20 TFs predicted by SMRT.

Full-size [DOI: 10.7717/peerj.12940/fig-4](https://doi.org/10.7717/peerj.12940/fig-4)

Analysis of TFs

Using iTAK software, 6,829 TFs and 2,439 TRs were predicted; they belonged to 91 families of TFs, and a large number of TF families were found (Fig. 4 and File S5). The most abundant TF family was the WRKY (386 transcripts) family, followed by the AP2/ERF-ERF (337 transcripts), bHLH (332 transcripts), NAC (273 transcripts), MYB (249 transcripts) and AUX/IAA (229 transcripts) families. The TF family members provided additional mechanistic insights into the regulation of secondary metabolites.

Analysis of differentially expressed genes (DEGs)

We analyzed DEGs of the roots, stems and leaves in the pogostone-type vs the patchouliol-type patchouli and found 5,842 DEGs (Files S6 and S7). A Venn diagram was

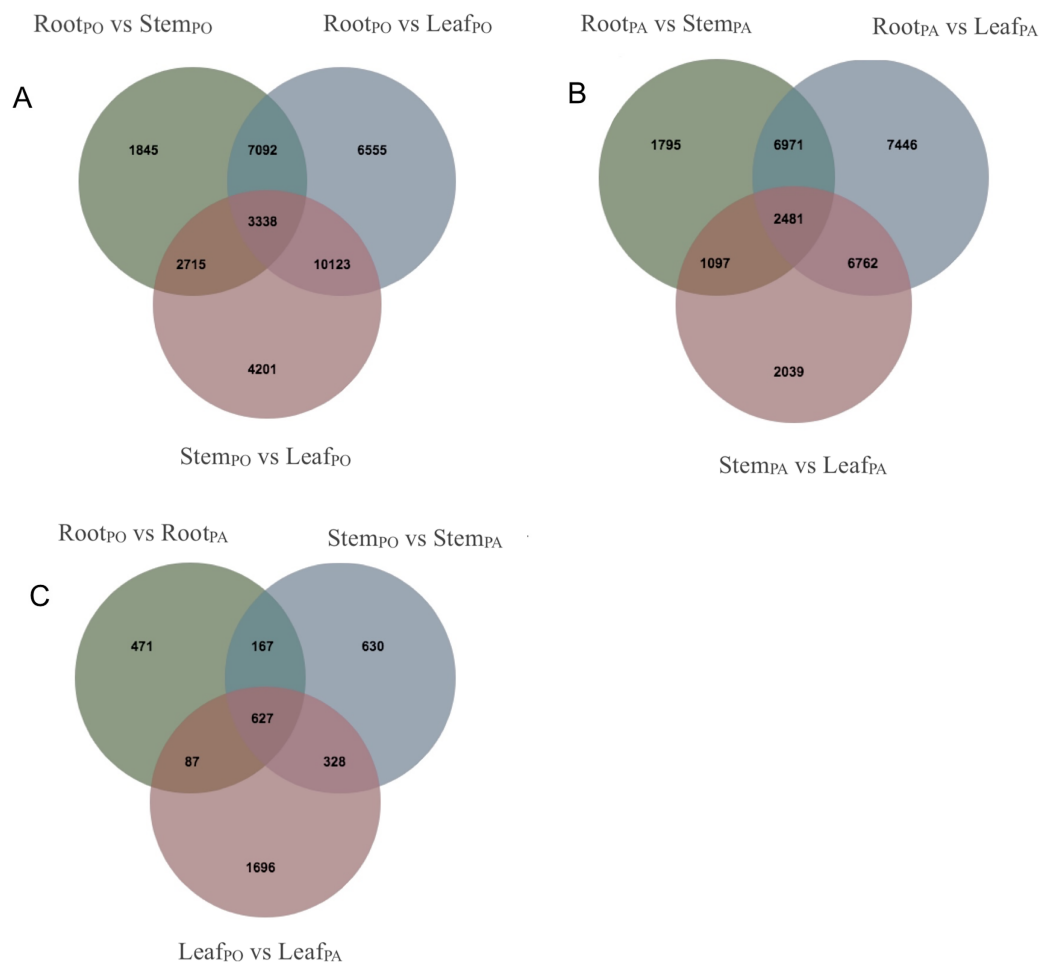


Figure 5 DEGs identified in comparisons among different tissues of *P. cablin*. (A) Venn diagram showing common and unique DEGs among different comparisons about the tissues of the pogostone-type (PO). (B) Venn diagram showing common and unique DEGs among different comparisons about the tissues of the patchouliol-type (PA). (C) Venn diagram showing common and unique DEGs among different comparisons about the tissues of two chemical types.

Full-size DOI: 10.7717/peerj.12940/fig-5

made to visualize the numbers of overlapping and unique DEGs (Figs. 5A–5C). A total of 2,738 DEGs were in the leaves of the two chemical types, of which 1,833 were upregulated and 905 were downregulated. There were 1,752 DEGs in the stems of the two chemical types, of which 1,072 were upregulated and 680 were downregulated. There were 1,352 DEGs in the roots of the two chemical types, with 851 upregulated and 501 downregulated. More DEGs were mainly upregulated in the patchouliol-type than in the pogostone-type. For the same chemical type of patchouli, we compared the DEGs of three tissues. The DEGs in the roots vs leaves (23,660 DEGs in the patchouliol-type and 27,108 DEGs in the pogostone-type respectively) were much more abundant than in the roots vs stems (12,344 DEGs in the patchouliol-type and 14,990 DEGs in the pogostone-type respectively) and in the stems vs leaves (12,379 DEGs in the patchouliol-type and 20,377 DEGs in the pogostone-type respectively) (Figs. 5A and 5B).

KEGG enrichment analysis of DEGs

KEGG enrichment analysis revealed that these DEGs in the roots of the pogostone-type vs the patchouliol-type were significantly enriched in “sesquiterpenoid and triterpenoid biosynthesis” with q value ≤ 0.02815124 (four genes up regulated and two down regulated), “ubiquinone and other terpenoid-quinone biosynthesis” (five genes up regulated), “stilbenoid, diarylheptanoid and gingerol biosynthesis” (three genes up regulated), “flavonoid biosynthesis” (four genes up regulated), “pentose phosphate pathway” (four genes up regulated and five down regulated), “fructose and mannose metabolism” (five genes up regulated), “terpenoid backbone biosynthesis” (six genes up regulated and three down regulated), “plant-pathogen interaction” (13 genes up regulated and two down regulated) and “phenylpropanoid biosynthesis” (seven genes up regulated and two down regulated) (Fig. 6A). In plant-pathogen interaction pathway, DEGs encoding RBOH (respiratory burst oxidase), CML (calcium-binding protein), MEKK1 (mitogen-activated protein kinase kinase kinase 1) were up regulated in the patchouliol-type, which could improve Cell wall reinforcement and plant defense response. The metabolic pathways related to carbohydrate and energy metabolism were “fructose and mannose metabolism” and “pentose phosphate pathway”, which played an important role in plant development. Otherwise, “stilbenoid, diarylheptanoid and gingerol biosynthesis”, “terpenoid backbone biosynthesis”, “ubiquinone and other terpenoid-quinone biosynthesis”, “sesquiterpenoid and triterpenoid biosynthesis”, “flavonoid biosynthesis” and “phenylpropanoid biosynthesis” were important secondary metabolism pathways. Among them “terpenoid backbone biosynthesis” and “sesquiterpenoid” were main ways to synthesize patchouli oil.

The KEGG pathways in the stems of the pogostone-type vs the patchouliol-type were mainly “brassinosteroid biosynthesis” with q value $\leq 2.00E-05$ (five genes up regulated and two down regulated), “sesquiterpenoid and triterpenoid biosynthesis” q value $\leq 4.09E-02$ (five genes up regulated and two down regulated), “flavonoid biosynthesis” (four genes up regulated and one down regulated), “phenylpropanoid biosynthesis” (nine genes up regulated and three down regulated), “ubiquinone and other terpenoid-quinone biosynthesis” (five genes up regulated), “stilbenoid, diarylheptanoid and gingerol biosynthesis” (three genes up regulated and one down regulated) and “plant hormone signal transduction” (13 genes up regulated and 11 down regulated) (Fig. 6B). The top 20 KEGG enrichment pathways of the stems were mainly composed of secondary metabolism and plant hormone signal transduction pathways. Among plant hormone signal transduction pathways, DEGs encoding JAR1 (jasmonic acid-amino synthetase 1), JAZ (Jasmonate ZIM domain-containing protein) and transcription factor MYC2 in Jasmonic acid (JA) biosynthesis and transduction pathway were up regulated in the patchouliol-type. And BR6OX2 (brassinosteroid-6-oxidase 2) in the BR downstream biosynthesis pathway (File S8) and GA2ox (gibberellin 2beta-dioxygenase) DEGs were up regulated in the patchouliol-type (File S9). However DEGs encoding ARF (auxin response factor) and SAUR family protein were down regulated in the auxin signal transduction of the patchouliol-type.

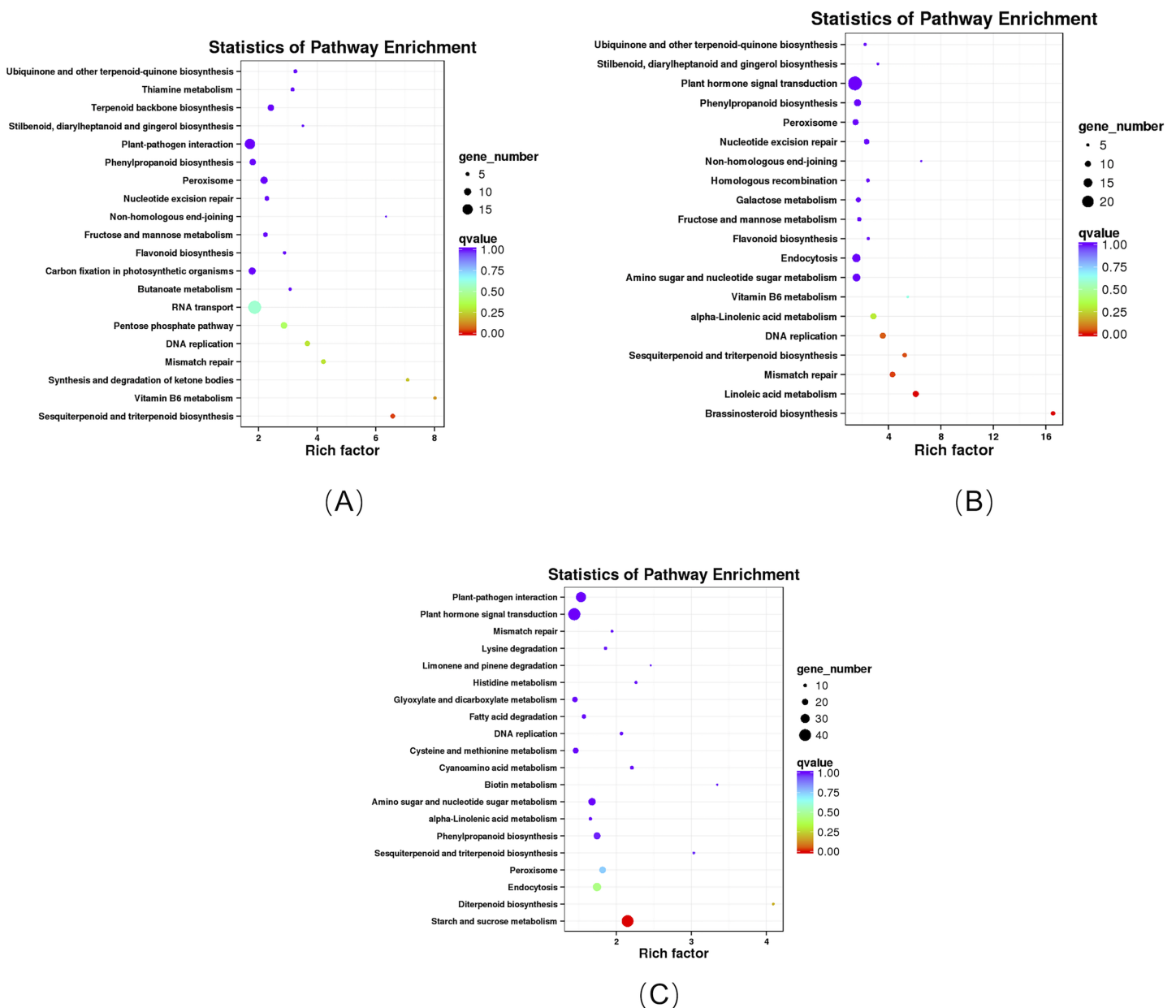


Figure 6 KEGG enrichment analysis of DEGs. (A) KEGG enrichment analysis of DEGs between the root of the pogostone-type vs the root of the patchouliol-type. (B) KEGG enrichment analysis of DEGs between the stem of the pogostone-type vs the stem of the patchouliol-type. (C) KEGG enrichment analysis of DEGs between the leaf of the pogostone-type vs the leaf of the patchouliol-type.

Full-size [DOI: 10.7717/peerj.12940/fig-6](https://doi.org/10.7717/peerj.12940/fig-6)

From KEGG pathway enrichment analysis, we found that the pathways of the leaves in the pogostone-type vs the patchouliol-type were mainly “diterpenoid biosynthesis” with q value ≤ 0.1852 (four genes up regulated and three down regulated), “phenylpropanoid biosynthesis” (17 genes up regulated and five down regulated), “sesquiterpenoid and triterpenoid biosynthesis” (five genes up regulated and two down regulated), “plant hormone signal transduction” (27 genes up regulated and 14 down regulated) and “plant-pathogen interaction” (27 genes up regulated and seven down regulated) (Fig. 6C).

We found that the KEGG enrichment pathways of the leaves were mainly composed of secondary metabolism, plant hormone signal transduction and plant-pathogen interaction pathways. Among plant hormone signal transduction pathways, BSK (BR-signaling kinase), BIN2 (protein brassinosteroid insensitive 2), BZR1/2 (brassinosteroid resistant 1/2) in the brassinosteroid signal transduction pathway were expressed at higher levels in the patchouliol-type than in the pogostone-type (File S10). Furthermore, DEGs encoding JAR1 in the JA signal transduction pathway were up regulated in the patchouliol-type. DEGs encoding GA20ox (gibberellin-44 dioxygenase) and GA2ox in the diterpenoid biosynthesis pathway were up regulated in the patchouliol-type than in the pogostone-type (File S11). In plant-pathogen interaction pathway, DEGs encoding RBOH, CML, calmodulin (CALM), MEKK1 and FLS2 (LRR receptor-like serine/threonine-protein kinase) were up regulated and disease resistance protein RPM1 was down regulated in the patchouliol-type. Otherwise, the DEGs in the pathways of secondary metabolite biosynthesis indicated that many important ingredients, such as sesquiterpenes, triterpenoids, diterpenoids and flavonoids, were mainly synthesized in the leaves.

Transcript profiles involved in terpenoid backbone biosynthesis

To investigate the terpenoid backbone biosynthesis pathway in three tissues of *P. cablin*, we analyzed the transcripts in the terpenoid backbone biosynthesis pathway (Fig. 7 and File S12), and the expression patterns of the terpenoid backbone biosynthesis genes were detected by qRT-PCR (Fig. 8 and File S13).

In the MVA pathway, we found that most of the transcripts, such as hydroxy-3-methylglutaryl-CoA synthase (HMGS), 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), mevalonate kinase (MVK), phosphomevalonate kinase (PMK), and mevalonate diphosphate decarboxylase (MVD), showed higher expression levels in the roots but lower expression levels in the stems and leaves (Fig. 7A). However, in the MEP pathway, most of the transcripts, including 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS), and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR), showed higher expression levels in the leaves, with lower expression levels in the stems and the roots. Then, IPP and dimethylallyl pyrophosphate (DMAPP) are condensed from farnesyl diphosphate synthase (FPPS) to form the sesquiterpene (including patchouli alcohol) intermediate (Fig. 7B). In addition, we found that the expression of patchouli alcohol synthase (PTS) genes in the leaves was much higher than that in the roots and stems (Fig. 8).

DISCUSSION

Evaluation of SMRT sequencing quality

In this study, the full-length transcriptome of *P. cablin* was sequenced using SMRT sequencing and NGS technology. In total, 574,817,420,978 ROIs, including 334,683 and 276,811 FLNCs, were obtained from the two cell lines. Second-generation transcriptome

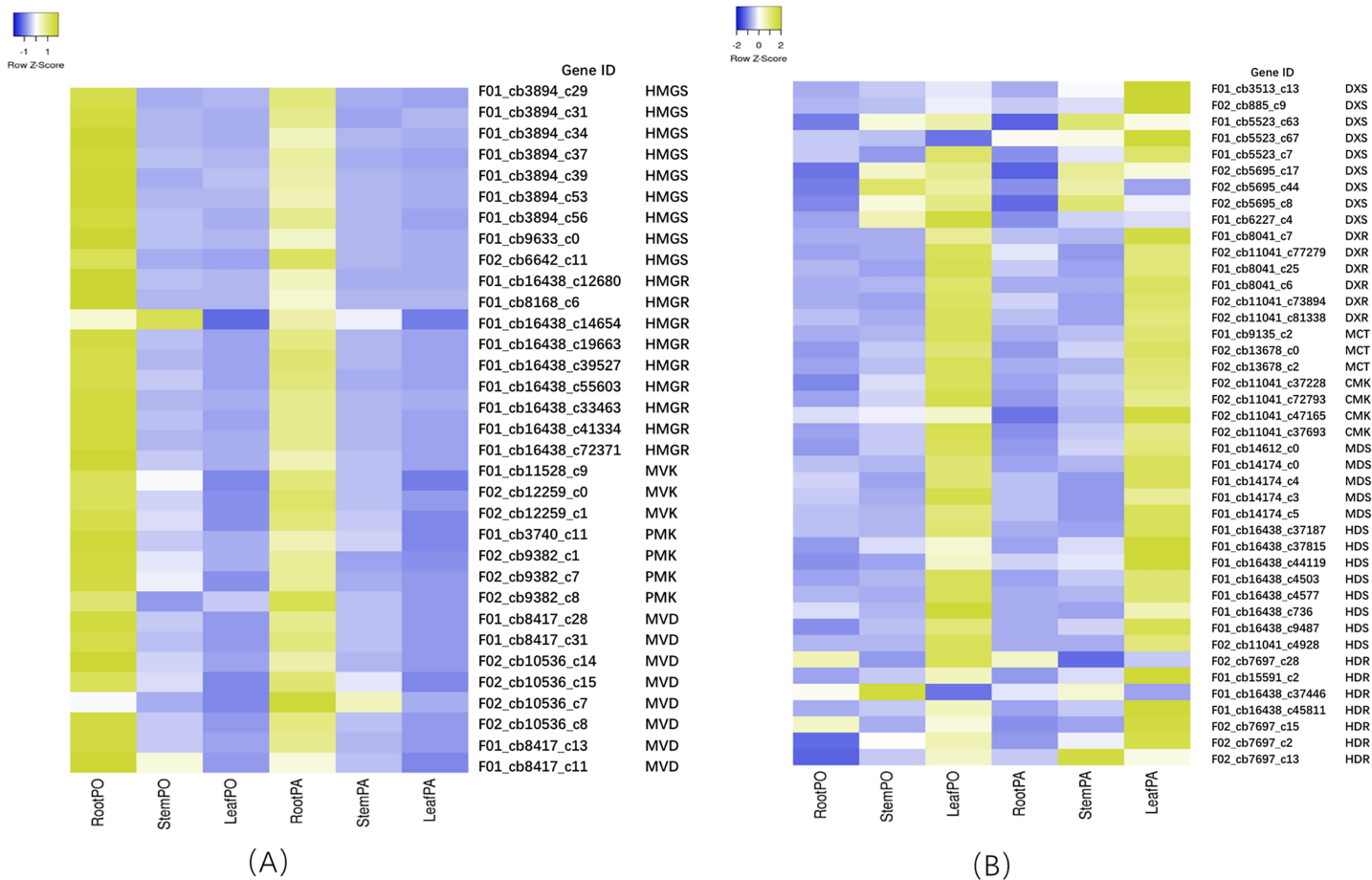


Figure 7 Heatmap of expression levels of the MVA and MEP pathway in root, stem and leaf of two chemical types. (A) DEGs of MVA related genes. (B) DEGs of MEP related genes. PA means the patchouliol-type and PO means the pogostone-type.

Full-size [DOI: 10.7717/peerj.12940/fig-7](https://doi.org/10.7717/peerj.12940/fig-7)

sequencing of 18 samples was completed, and a total of 125.83 G of clean data was obtained, ensuring that Q30 reached 85%. A total of 169,641, 152,493 consensus isoforms were obtained from F01 and F02, including 95,298 and 92,051 high-quality isoforms respectively. 134,647 nonredundant transcript isoforms were obtained using CD-HIT software. The average consensus isoform read lengths in F01 and F02 were 2,028 and 2,147 bp, respectively, which indicated that ROIs were long enough to represent full-length transcripts (Minoche *et al.*, 2015; Tian *et al.*, 2018).

Meanwhile, the second-generation data were used to correct the third-generation low-quality reads. Therefore, the resulting comprehensive transcriptome database for *P. cablin* could provide insight into the structures and functions of genes and the gene expression differences in the two chemical types.

The role of the MEP pathway in terpenoid biosynthesis

There are two pathways for terpenoids biosynthesis in higher plants: the MVA pathway in the cytoplasm and the MEP pathway in plastids such as plant chloroplasts. However, increasing evidence indicates that the MVA and MEP pathways play different roles in

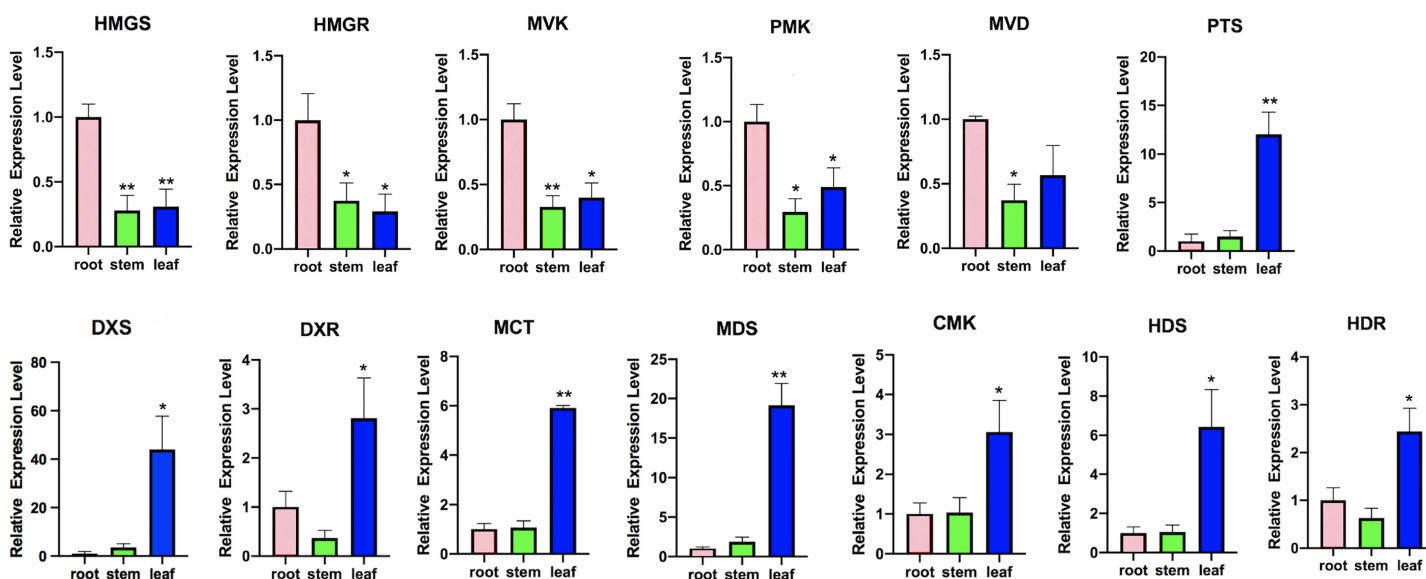


Figure 8 Expression analysis of transcripts involved in terpenoid backbone biosynthesis. Columns represent the value of relative expression levels. Error bars represent the SD from three independent experiments. The significance of the difference was analyzed using a one-way ANOVA (* $P < 0.01$, ** $P < 0.001$).

Full-size [DOI: 10.7717/peerj.12940/fig-8](https://doi.org/10.7717/peerj.12940/fig-8)

terpenoids (Kasahara et al., 2002; Luo et al., 2003; Skorupinska-Tudek et al., 2008).

In transgenic *Salvia miltiorrhiza* lines, DXS showed a much more powerful effect than HMGR in tanshinone biosynthesis, which indicated that genetic manipulation of the MEP pathway resulted in significantly higher yields than MVA pathway enhancement (Kai et al., 2011; Yang et al., 2012). The results suggested that the MEP pathway may play a major role in tanshinone biosynthesis. In our results, the expression levels of related genes in three vegetative tissues were very different in the two pathways. In the MVA pathway, the highest expression level of related biosynthetic genes was found in the roots, with the lower expression in the leaves and stems. However, in the MEP pathway, the highest expression level of related genes was found in the leaves, with the lower expression in the roots and stems. Targeting PTS into *Artemisia annua* plastid produced higher patchouli alcohol compared to its expression in cytoplasm (Fu et al., 2021). Because patchouli alcohol and sesquiterpene mainly synthesized in the leaves of *P. cablin*. Therefore, one of the explanations was that MEP pathway may probably play a much more important role in biosynthesis of patchouli alcohol and sesquiterpene. Another explanation was that because more chloroplasts (one kind of plastids) in leaves than roots or stems and the MEP pathway mainly took place in plastids, the gene expression level of the MEP pathway was higher in the leaves than the roots and the stems.

DEG analysis of the two chemical types

From the results for the DEGs, the number of DEGs between two tissues of the same chemical type was greater than that in the same tissue of two chemical types. Regardless of the patchouliol-type or the pogostone-type patchouli, the number of DEGs between the roots and leaves was the largest, and the number of DEGs between the stems and the

leaves was the smallest. Since each tissue has specific physiological functions and anatomical structures, the difference between the roots and leaves is usually greater than the difference between other plant vegetative tissues, so this pairing demonstrated the largest number of DEGs. The stems and leaves had similar physiological functions, such as photosynthesis, so the number of DEGs between them was minimal.

According to the number of DEGs in the same tissue of the two chemical types, leaves > stems > roots. The number of DEGs between the leaves of the two chemical types was approximately twice that of the roots. The data indicated that in addition to the differences in physiological functions, there may be more DEGs involved in the biosynthesis of secondary metabolites in the leaves, followed by the stems.

Identification of hormone signaling-related genes in growth and development

Patchouliol-type patchouli usually has larger leaves, longer stems, higher stress resistance and more aromatic oil, while pogostone-type patchouli has smaller leaves, shorter stems, poor stress resistance and less aromatic oil.

JA and its derivatives are collectively named jasmonates (JAs). JAs are signaling molecules involved in the regulation of many physiological and developmental processes in plants (Lorenzo *et al.*, 2004). In addition, JAs are important regulators in the plant response to biotic and abiotic stresses, such as ozone exposure, wounding, water deficit and pathogen/pest attack. Currently, JA, MeJA (a volatile methyl ester, methyl jasmonate) and JA-Ile (predominant amino acid conjugate, jasmonoyl-isoleucine) are known to function as signal molecules in plants (Li & Yan, 2014). Our previous study revealed that exogenous foliar spraying with MeJA could effectively increase the content of patchouli alcohol (He *et al.*, 2014). DEG analysis found that JAR1 transcripts were upregulated in the leaves of patchouliol-type patchouli. JAR1 can conjugate Ile to JA-Ile, which binds to the receptor COI1 to activate most JA-elicited responses. The upregulation of the JAR1 DEGs may indicate more JA-elicited responses to induce patchouli alcohol and terpenoid synthesis.

Brassinosteroids (BRs) are a class of plant steroid hormones that play an extremely important role in plant growth and development and responses to several stresses, such as extreme temperatures and drought (Bai *et al.*, 2012; Yang *et al.*, 2011). C6 oxidation is a rate-limiting step in the BR biosynthesis pathway (Wei & Li, 2020), and the downstream enzyme BR6OX2 is a rate-limiting enzyme (Gruszka, Szarejko & Maluszynski, 2011). Loss of either BR6OX2 or BR6OX2 farnesylation could result in reduced brassinolide accumulation (Jamshed *et al.*, 2017). The upregulation of BR6OX2 DEGs possibly resulted in the synthesis of more BRs (Wei *et al.*, 2017).

Gibberellins (GAs) are plant hormones that are essential for many developmental processes in plants, including seed germination, stem elongation, leaf expansion, and trichome development (Achard & Genschik, 2009). The major bioactive GAs, which include GA1, GA3, GA4 and GA7, are derived from a basic diterpenoid carboxylic acid skeleton (Yamaguchi, 2008). According to the KEGG enrichment analysis between two chemical types, the GA20ox and GA2ox transcripts were upregulated in the diterpenoid

biosynthesis pathway of the patchouliol-type. GA20ox and GA2ox play key roles in the entire process of GA metabolism. They are responsible for the last several steps in the synthesis of active GAs, whereas GA2ox members catalyze the inactivation of bioactive GAs (Shang *et al.*, 2017). The higher GA20ox and GA2ox gene expression levels suggest that more GAs may be synthesized in the patchouliol-type than in the pogostone-type. From the expression of related genes in the plant hormone synthesis pathways, we found that JAR1, BR6OX2, GA20ox and GA2ox DEGs were all upregulated in the patchouliol-type. In this case, more JAs, GAs and BRs could be synthesized in the tissues of the patchouliol-type. Therefore, these plant hormones could promote the growth and resistance of patchouliol-type patchouli.

CONCLUSIONS

The two chemical types of patchouli exhibit many differences in morphology and chemical compositions. Our work demonstrated a comprehensive view of the molecular mechanisms of morphology and terpenoid biosynthesis by using a combination of SMRT and next-generation sequencing. Hormone levels may play an important role in the formation of two chemical types. These findings help to understand the formation mechanisms of the two chemical types and terpenoid biosynthesis in *P. cablin*.

ACKNOWLEDGEMENTS

We would like to thank Ms. Yingying Li and Ms. Xiaohua Ou, for their kind help in the project performance, and Ms. Wenyi Chen and Mr. Keke Luo for their technical assistance.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Natural Science Foundation of China (No. 81773829), and the Science and Technology planning project of Guangdong Province (No. 2017030303081 & No. 2017030303082). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: 81773829.

Science and Technology planning project of Guangdong Province: 2017030303081, 2017030303082.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Hongyi Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Wenjing Deng performed the experiments, prepared figures and/or tables, and approved the final draft.
- Changhua Lu performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Mengling He conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Hanjing Yan conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at Genbank: [PRJNA607886](https://www.ncbi.nlm.nih.gov/nuclseq/PRJNA607886), [PRJNA606892](https://www.ncbi.nlm.nih.gov/nuclseq/PRJNA606892).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12940#supplemental-information>.

REFERENCES

- Achard P, Genschik P. 2009. Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. *Journal of Experimental Botany* 60(4):1085–1092 DOI 10.1093/jxb/ern301.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17):3389–3402 DOI 10.1093/nar/25.17.3389.
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11(10):R106 DOI 10.1186/gb-2010-11-10-r106.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* 25(1):25–29 DOI 10.1038/75556.
- Bai MY, Shang JX, Oh E, Fan M, Bai Y, Zentella R, Sun TP, Wang ZY. 2012. Brassinosteroid, gibberellin and phytochrome impinge on a common transcription. module in *Arabidopsis*. *Nature Cell Biology* 14(8):810–817 DOI 10.1038/ncb2546.
- Bairoch A, Boeckmann B. 1991. The SWISS-PROT protein sequence data bank. *Nucleic Acids Research* 19(Suppl):2247–2249 DOI 10.1093/nar/19.suppl.2247.
- Chen X, Li J, Wang X, Zhong L, Tang Y, Zhou X, Liu Y, Zhan R, Zheng H, Chen W, Chen L. 2019. Full-length transcriptome sequencing and methyl jasmonate-induced expression profile analysis of genes related to patchoulol biosynthesis and regulation in *Pogostemon cablin*. *BMC Plant Biology* 19(1):266 DOI 10.1186/s12870-019-1884-x.
- Chen ST, Qiu GL, Yang ML. 2019. SMRT sequencing of full-length transcriptome of seagrasses *Zostera japonica*. *Scientific Reports* 9(1):14537 DOI 10.1038/s41598-019-51176-y.
- Chen MW, Zhang JM, Lai YF, Wang SP, Li P, Xiao J, Fu CM, Hu H, Wang YT. 2013. Analysis of *Pogostemon cablin* from pharmaceutical research to market performances. *Expert Opinion on Investigational Drugs* 22(2):245–257 DOI 10.1517/13543784.2013.754882.
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome

assemblies from long-read SMRT sequencing data. *Nature Methods* **10**(6):563–569
DOI [10.1038/nmeth.2474](https://doi.org/10.1038/nmeth.2474).

- Chinese Pharmacopoeia Commission.** 2020. *Pharmacopoeia of the People's Republic of China I*. Beijing: China Medical Science Press.
- Deguerry F, Pastore L, Wu SQ.** 2006. The diverse sesquiterpene profile of patchouli, *Pogostemon cablin*, is correlated with a limited number of sesquiterpene synthases. *Archives of Biochemistry and Biophysics* **454**(2):123–136 DOI [10.1016/j.abb.2006.08.006](https://doi.org/10.1016/j.abb.2006.08.006).
- Deng YY, Li JQ, Wu SF, Zhu YP, Chen YW, He FC.** 2006. Integrated nr database in protein annotation system and its localization. *Computer Engineering* **32**:71–74
DOI [10.3969/j.issn.1000-3428.2006.05.026](https://doi.org/10.3969/j.issn.1000-3428.2006.05.026).
- Deng Y, Zheng H, Yan ZC, Liao DY, Li CL, Zhou JY, Liao H.** 2018. Full-Length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin biosynthesis, seed formation and development, and stress response. *International Journal of Molecular Sciences* **19**(9):2476 DOI [10.3390/ijms19092476](https://doi.org/10.3390/ijms19092476).
- Dong LL, Liu HF, Zhang JC, Yang SJ, Kong GY, Chu JSC, Chen NS, Wang DW.** 2015. Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics* **16**(1):1039 DOI [10.1186/s12864-015-2257-y](https://doi.org/10.1186/s12864-015-2257-y).
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**(16):2194–2200
DOI [10.1093/bioinformatics/btr381](https://doi.org/10.1093/bioinformatics/btr381).
- Fan L, Jin RH, Liu YH, An M, Chen S.** 2011. Enhanced extraction of patchouli alcohol from *Pogostemon cablin* by microwave radiation accelerated ionic liquid pretreatment. *Journal of Chromatography B* **879**(30):3653–3657 DOI [10.1016/j.jchromb.2011.09.035](https://doi.org/10.1016/j.jchromb.2011.09.035).
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M.** 2014. Pfam: the protein families database. *Nucleic Acids Research* **42**(D1):D222–D230 DOI [10.1093/nar/gkt1223](https://doi.org/10.1093/nar/gkt1223).
- Fu X, Zhang F, Ma Y, Hassani D, Peng B, Pan Q, Zhang Y, Deng Z, Liu W, Zhang J, Han L, Chen D, Zhao J, Li L, Sun X, Tang K.** 2021. High-level patchoulol biosynthesis in *Artemisia annua* L. *Frontiers in Bioengineering and Biotechnology* **8**:1590 DOI [10.3389/fbioe.2020.621127](https://doi.org/10.3389/fbioe.2020.621127).
- Gruszka D, Szarejko I, Maluszynski M.** 2011. Identification of barley *DWARF* gene involved in brassinosteroid synthesis. *Plant Growth Regulation* **65**(2):343–358
DOI [10.1007/s10725-011-9607-9](https://doi.org/10.1007/s10725-011-9607-9).
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A.** 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**(8):1494–1512 DOI [10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084).
- Hackl T, Hedrich R, Schultz J, Förster F.** 2014. Proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**(21):3004–3011
DOI [10.1093/bioinformatics/btu392](https://doi.org/10.1093/bioinformatics/btu392).
- He ML, He F, Xiong Y, Ma JY, Gong FJ, Yan HJ.** 2014. Effect of methyl jasmonate on the content of patchouli alcohol from the leaves of *Pogostemon cablin* (Blanco) Benth. *North Horticulture* **5**:147–150.
- He Y, Xiao H, Deng C, Xiong L, Nie H, Peng C.** 2016. Survey of the genome of *Pogostemon cablin* provides insights into its evolutionary history and sesquiterpenoid biosynthesis. *Scientific Reports* **6**(1):26405 DOI [10.1038/srep26405](https://doi.org/10.1038/srep26405).

- Hsu HC, Yang WC, Tsai WJ, Chen CC, Huang HY, Tsai YC. 2006. α -Bulnesene, a novel PAF receptor antagonist isolated from *Pogostemon cablin*. *Biochemical and Biophysical Research Communications* 345(3):1033–1038 DOI 10.1016/j.bbrc.2006.05.006.
- Hu ZZ, Huang WZ, Lu CH, Luo KK, Zhang HY, He ML, Yan HJ. 2019. Non-targeted metabolomics study of *Pogostemon cablin* from different habitats. *Journal of Chinese Medicinal Materials* 42(2):271–278 DOI 10.13863/j.issn1001-4454.2019.02.007.
- Jamshed M, Liang S, Hickerson NMN, Samuel MA. 2017. Farnesylation-mediated subcellular localization is required for CYP85A2 function. *Plant signaling & behavior* 12(10):e1382795 DOI 10.1080/15592324.2017.1382795.
- Jensen LJ, Julien P, Kuhn M, Mering CV, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research* 36(Database):D250–D254 DOI 10.1093/nar/gkm796.
- Jia D, Wang YX, Liu YH, Hu J, Guo YQ, Gao LL, Ma RY. 2018. SMRT sequencing of full-length transcriptome of flea beetle *Agasicles hygrophila* (Selman and Vogt). *Scientific Reports* 8(1):2197 DOI 10.1038/s41598-018-20181-y.
- Kai GY, Xu H, Zhou CC, Liao P, Xiao JB, Luo XQ, You LJ, Zhang L. 2011. Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metabolic Engineering* 13(3):319–327 DOI 10.1016/j.ymben.2011.02.003.
- Kasahara H, Hanada A, Kuzuyama T, Takagi M, Kamiya Y, Yamaguchi S. 2002. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. *The Journal of Biological Chemistry* 277:45188–45194 DOI 10.1074/jbc.M208659200.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* 35(suppl_2):345–349 DOI 10.1093/nar/gkm391.
- Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf Y, Yin JJ, Natale DA. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* 5(2):R7 DOI 10.1186/gb-2004-5-2-r7.
- Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology* 30(7):693–700 DOI 10.1038/nbt.2280.
- Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357–359 DOI 10.1038/nmeth.1923.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323 DOI 10.1186/1471-2105-12-323.
- Li W, Pian CM, Xu L, Xu HH, Liu XX. 2002. The observation and comparison of *Pogostemon cablin* from different habitats. *Journal of Chinese Medicinal Materials* 25(7):463–465 DOI 10.13863/j.issn1001-4454.2002.07.002.
- Li MS, Yan XF. 2014. Jasmonic acid signaling in plants and its biological functions in relation to environment. *Acta Ecologica Sinica* 34:6779–6788 DOI 10.5846/stxb201303090383.
- Liu XY, Li JH, Yang YT, Yan HJ, Zhang HY, He ML. 2021. Genetic diversity analysis of *Pogostemon cablin* based on AFLP markers. *Chinese Journal of Experimental Traditional Medical Formulae* 27(4):152–158 DOI 10.1016/j.bse.2009.12.006.

- Lorenzo O, Chico JM, Saénchez-Serrano JJ, Solano R. 2004.** Jasmonate-insensitive1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. *The Plant Cell* **16**(7):1938–1950 DOI [10.1105/tpc.022319](https://doi.org/10.1105/tpc.022319).
- Luo JP, Liu YP, Feng YF, Guo XL, Cao H. 2003.** Two chemotypes of *Pogostemon cablin* and influence of region of cultivation and harvesting time on volatile oil composition. *Acta Pharmaceutica Sinica* **38**(4):307–310 DOI [10.16438/j.0513-4870.2003.04.017](https://doi.org/10.16438/j.0513-4870.2003.04.017).
- Mao XZ, Cai T, Olyarchuk JG, Wei LP. 2005.** Automated genome annotation and pathway identification using the KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**(19):3787–3793 DOI [10.1093/bioinformatics/bti430](https://doi.org/10.1093/bioinformatics/bti430).
- Minoche AE, Dohm JC, Schneider J, Holtgräwe D, Viehöver P, Montfort M, Sörensen TR, Weisshaar B, Himmelbaue H. 2015.** Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biology* **16**(1):184 DOI [10.1186/s13059-015-0729-7](https://doi.org/10.1186/s13059-015-0729-7).
- Miyazawa M, Okuno Y, Nakamura SI, Kosaka H. 2000.** Antimutagenic activity of flavonoids from *Pogostemon cablin*. *Journal of Agricultural and Food Chemistry* **48**(3):642–647 DOI [10.1021/jf990160y](https://doi.org/10.1021/jf990160y).
- Ogata H, Goto S, Sato K, Fujibuchi W. 1999.** KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **27**(1):29–34 DOI [10.1093/nar/27.1.29](https://doi.org/10.1093/nar/27.1.29).
- Qian XD, Sun YP, Zhou GF, Yuan YM, Li J, Huang HL, Xu LM, Li LQ. 2019.** Single-molecule real-time transcript sequencing identified flowering regulatory genes in *Crocus sativus*. *BMC Genomics* **20**(1):857 DOI [10.1186/s12864-019-6200-5](https://doi.org/10.1186/s12864-019-6200-5).
- Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M, Tissier A. 2009.** A novel pathway for sesquiterpene biosynthesis from Z, Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell* **21**(1):301–317 DOI [10.1105/tpc.107.057885](https://doi.org/10.1105/tpc.107.057885).
- Shang MY, Wang XT, Zhang J, Qi XH, Ping AM, Hou LP, Xing G, Li G, Li M. 2017.** Genetic regulation of GA metabolism during vernalization, floral bud initiation and development in pak choi (*Brassica rapa* ssp. *chinensis* Makino). *Frontiers in Plant Science* **8**:1533 DOI [10.3389/fpls.2017.01533](https://doi.org/10.3389/fpls.2017.01533).
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015.** BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19):3210–3212 DOI [10.1093/bioinformatics/btv351](https://doi.org/10.1093/bioinformatics/btv351).
- Skorupinska-Tudek K, Poznanski J, Wojcik J, Bienkowski T, Szostkiewicz I, Zelman-Femiak M, Bajda A, Chojnacki T, Olszowska O, Grunler J, Meyer O, Rohmer M, Danikiewicz W, Swiezewska E. 2008.** Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of dolichols in plants. *The Journal of Biological Chemistry* **283**(30):21024–21035 DOI [10.1074/jbc.M706069200](https://doi.org/10.1074/jbc.M706069200).
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. 2013.** Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research* **41**(17):e166 DOI [10.1093/nar/gkt646](https://doi.org/10.1093/nar/gkt646).
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000.** The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**(1):33–36 DOI [10.1093/nar/28.1.33](https://doi.org/10.1093/nar/28.1.33).
- Tian JY, Feng SJ, Liu YL, Zhao LL, Tian L, Hu Y, Yang TX, Wei AZ. 2018.** Single-molecule long-read sequencing of *Zanthoxylum bungeanum* Maxim transcriptome: identification of aroma-related genes. *Forests* **9**(12):765 DOI [10.3390/f9120765](https://doi.org/10.3390/f9120765).
- Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, Chen F, Harel I, Bustamante CD, Rasmussen M, Snyder MP. 2015.** Comprehensive transcriptome analysis using synthetic

long-read sequencing reveals molecular co-association of distant splicing events. *Nature Biotechnology* 33(7):736–742 DOI 10.1038/nbt.3242.

- Wang LJ, Jiang XL, Wang L, Wang W, Fu CL, Yan XC, Geng XX. 2019.** A survey of transcriptome complexity using PacBio single-molecule real-time analysis combined with Illumina RNA sequencing for a better understanding of ricinoleic acid biosynthesis in *Ricinus communis*. *BMC Genomics* 20(1):456 DOI 10.1186/s12864-019-5832-9.
- Wang LG, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W. 2013.** CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research* 41(6):e74 DOI 10.1093/nar/gkt006.
- Wei ZY, Li J. 2020.** Regulation of brassinosteroid homeostasis in higher plants. *Frontiers in Plant Science* 11:583622 DOI 10.3389/fpls.2020.583622.
- Wei ZY, Yuan T, Tarkowská D, Kim J, Nam HG, Novák O, He K, Gou XP, Li J. 2017.** Brassinosteroid biosynthesis is modulated via a transcription factor cascade of COG1, PIF4, and PIF5. *Plant Physiology* 174(2):1260–1273 DOI 10.1104/pp.16.01778.
- Wu YG, Guo QS, He JC, Lin YF, Luo LJ, Liu GD. 2010.** Genetic diversity analysis among and within populations of *Pogostemon cablin* from China with ISSR and SRAP markers. *Biochemical Systematics and Ecology* 38(1):63–72 DOI 10.1016/j.bse.2009.12.006.
- Yamaguchi S. 2008.** Gibberellin metabolism and its regulation. *Annual Review of Plant Biology* 59(1):225–251 DOI 10.1146/annurev.arplant.59.032607.092804.
- Yang DF, Du XH, Liang X, Han RL, Liang ZS, Liu Y, Liu FH, Zhao JJ. 2012.** Different roles of the mevalonate and methylerythritol phosphate pathways in cell growth and tanshinone production of *Salvia miltiorrhiza* hairy roots. *PLOS ONE* 7(11):e46797 DOI 10.1371/journal.pone.0046797.
- Yang CJ, Zhang C, Lu YN, Jin JQ, Wang XL. 2011.** The mechanisms of brassinosteroids' action: from signal transduction to plant development. *Molecular Plant* 4(12):588–600 DOI 10.1016/j.molp.2016.09.014.
- Zheng Y, Jiao C, Sun HH, Rosli HG, Pombo M, Zhang PF, Banf M, Dai XB, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei ZJ. 2016.** iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant* 9(12):1667–1670 DOI 10.1016/j.molp.2016.09.014.