



Article

Full-Length Transcriptome Sequencing and Different Chemotype Expression Profile Analysis of Genes Related to Monoterpene Biosynthesis in *Cinnamomum porrectum*

Fengying Qiu ^{1,2}, Xindong Wang ², Yongjie Zheng ², Hongming Wang ³, Xinliang Liu ² and Xiaohua Su ^{1,4,*}

¹ State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China; quifengying1@163.com

² Jiangxi Academy of Forestry, Camphor Engineering Technology Research Center for National Forestry and Grassland Administration, Nanchang 30032, China; x_wangxindong@sina.com (X.W.); zyj920581676@gmail.com (Y.Z.); liuxinliang1988@163.com (X.L.)

³ College of Bioengineering and Biotechnology, Tianshui Normal University, Tianshui 741000, China; hongmw26@163.com

⁴ Key Laboratory of Tree Breeding and Cultivation, National Forestry and Grassland Administration, Beijing 100091, China

* Correspondence: suxh@caf.ac.cn; Tel.: +86-010-6288-9627

Received: 7 November 2019; Accepted: 6 December 2019; Published: 10 December 2019



Abstract: Leaves of *C. porrectum* are rich in essential oils containing monoterpenes, sesquiterpenes and aromatic compounds, but the molecular mechanism of terpenoid biosynthesis in *C. porrectum* is still unclear. In this paper, the differences in the contents and compositions of terpenoids among three chemotypes were analyzed using gas chromatography mass spectrometry (GC/MS). Furthermore, the differential expression of gene transcripts in the leaf tissues of the three *C. porrectum* chemotypes were analyzed through a comparison of full-length transcriptomes and expression profiles. The essential oil of the three *C. porrectum* chemotypes leaves was mainly composed of monoterpenes. In the full-length transcriptome of *C. porrectum*, 104,062 transcripts with 306,337,921 total bp, an average length of 2944 bp, and an N50 length of 5449 bp, were obtained and 94025 transcripts were annotated. In the eucalyptol and linalool chemotype, the camphor and eucalyptol chemotype, and the camphor and linalool chemotype comparison groups, 21, 22 and 18 terpene synthase (TPS) unigenes were identified respectively. Three monoterpene synthase genes, *CpTPS3*, *CpTPS5* and *CpTPS9*, were upregulated in the eucalyptol chemotype compared to the linalool chemotype and camphor chemotype. *CpTPS1* was upregulated in the camphor chemotype compared to the linalool chemotype and the eucalyptol chemotype. *CpTPS4* was upregulated in the linalool chemotype compared to the camphor chemotype and the eucalyptol chemotype. Different unigenes had different expression levels among the three chemotypes, but the unigene expression levels of the 2-C-methyl-D-erythritol 4phosphate (MEP) pathway were generally higher than those of the mevalonate acid (MVA) pathway. Quantitative reverse transcription PCR(qRT-PCR) further validated these expression levels. The present study provides new clues for the functional exploration of the terpenoid synthesis mechanism and key genes in different chemotypes of *C. porrectum*.

Keywords: *Cinnamomum porrectum*; terpenoids; terpenoid biosynthesis; full-length transcriptome; different chemotypes

1. Introduction

Cinnamomum porrectum (Roxb.) Kosterm (*C. porrectum*), an evergreen tree of the family Lauraceae and genus *Cinnamomum*, that grows in evergreen broad-leaved forests or as shrubs, is mainly found in southern China, Bhutan, Cambodia, India, Indonesia, Laos, Malaysia, Myanmar, Nepal, Pakistan, Thailand and Vietnam [1]. The leaves of *C. porrectum* are rich in essential oils, and the whole plant serves as an important source of natural essential oils. *C. porrectum* is widely applied in foods, medicines and the fine chemical industry, including in natural spices and chemical materials [2]. The extracts of the leaves, bark, wood and roots of *C. porrectum* have antioxidant and antibacterial effects [3–5], and *C. porrectum* leaves are traditionally used in carminative tonic, stomachic tonic, and antipyretic preparations [6].

Studies have shown that the essential oil of *C. porrectum* leaves is rich in monoterpenes, sesquiterpenes and aromatic compounds [2,7], and the terpenoids with high contents in this leaf essential oil are mainly linalool, eucalyptol, camphor, citral, and nerolidol. According to the first principal component of its leaf essential oil, *C. porrectum* can be divided into the linalool, camphor, eucalyptol, citral and nerolidol chemotypes [2].

Terpenoids in the essential oil of *C. porrectum* are widely used and have a high development value. Among these terpenoids, linalool is frequently used as an odorant for perfumes and in common chemical products; in addition, linalool has analgesic [8], anti-inflammatory [9], and antitumor [10,11] effects. Eucalyptol, with its antimicrobial and insecticidal effects is widely used in medicine, spices and industry. Citral has a strong lemon odor and is widely used in many perfume compositions; in addition, citral has bacteriostatic effects [12].

In recent years, the biosynthesis and metabolic regulation of plant terpenoids have been studied [13–15], showing that two independent pathways in plants are involved in the biosynthesis of the 5-carbon structural unit isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP) of the universal precursor to isomers of terpenoids. In the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, the biosynthesis of IPP/DMAPP in plastids begins with pyruvate and glyceraldehyde-3-phosphate [16,17], but in the cytoplasm, the mevalonate (MVA) pathway produces IPP [18,19] from acetyl coenzyme A (CoA). The balance between IPP and DMAPP is controlled by IPP delta-isomerase (IDI), which converts IPP to DMAPP reversibly [20]. Terpenoids are generally formed by catalysis of corresponding precursors by terpene synthases (TPSs), while some terpenoids in plants required further modifications, mainly involving methyltransferases, glycosyltransferases, cytochrome P450 monooxygenase, dehydrogenase and reductase. The existence of these modification enzymes greatly enriches the variety of terpenoids and makes the structure and biological activity of terpenoids tend to be diversified [21]. Monoterpene synthase was an important component of leaf essential oil in *C. porrectum*, and the study of monoterpene biosynthesis was of great significance. Monoterpene synthase was the key enzyme in monoterpene biosynthesis, and now many monoterpene synthases have been isolated from plants [22–29]. Linalool synthase (LIS) gene was isolated from different source of *Cinnamomum osmophloeum* [26]. The recombinant LIS protein can produce s-(+)-linalool from GPP and (E) - Neroli from FPP. Bornyl diphosphate synthase (LaBPPS) was cloned from *Lavandula angustifolia* [27], which can catalyze the production of Bornyl or camphor. Through bioinformatics analysis, a monoterpene synthetase was identified from *Laurus nobilis*, and mainly catalyzes the formation of 1,8-cineole [28]. A monoterpene synthetase was cloned from *Cinnamomum tenuipilum*, which was specifically expressed in the geraniol chemotype. After prokaryotic expression and purification, geraniol synthase (*CtGES*) transformed the geraniol diphosphate into a single product geraniol [29]. The genetic transcriptome related to the biosynthesis of terpenoids in *Cinnamomum camphora* was analyzed preliminarily, and some candidate genes involved in the biosynthesis of terpenoids were screened. Three synthetic genes of monoterpene were found to be upregulated in the borneol chemotype [13]. The study of terpenoid synthases and their functions have made great progress. Determining how to use genetic engineering technology to regulate and even transform and improve the composition and yield of the essential oils of plants has become a research hotspot. However, the current research on the terpenoid biosynthetic

pathway of *C. porrectum* is still lacking, and there has been no gene study on the related biosynthesis of major terpenoids in the essential oil of *C. porrectum*, directly restricting the genetic improvement and exploitation and utilization of the quality of *C. porrectum* essential oil.

Transcriptome studies provide a useful perspective for expounding the molecular mechanisms of gene functions, cellular reactions, and different biological processes [30]. High-throughput sequencing (HTS) has recently been used to obtain transcriptome data for nonmodel species and provide valuable genomic information, especially for species without genomic sequences [31]. Second-generation sequencing (e.g., Illumina sequencing) has high-throughput capabilities, provides high-quality readings and enables very accurate research on gene expression and quantification. However, existing transcriptome analyses are based on incomplete genomic sequences and incomplete gene annotations. The fragmentation of genes assembled from second-generation sequencing data (usually 100–150 base pairs [bp] in length), which are limited by read length, results in a lack of information on full-length transcripts. With the advantage of an average read length of 10–15 kb, in combination with multifragment library screening technology, single-molecule real-time sequencing, also known as third-generation sequencing (e.g., PacBio sequencing), achieves transcript analysis without the need for splicing, overcomes the limitations of unigenes assembled from short spliced sequences and incomplete transcript structures in traditional second-generation transcriptomes, and can directly obtain all high-quality transcriptome information of single-RNA molecules from the 5' end to the 3' end. Third-generation sequencing is applicable to species without genomic sequences and has been used in extensive applications [32–36].

In view of this, the present study carried out third-generation de novo transcriptome sequencing of *C. porrectum* leaf tissue for the first time and sequenced the expression profiles of leaf tissues from three different chemotypes (linalool chemotype, eucalyptol chemotype and camphor chemotype) of *C. porrectum* to reveal information on the transcriptome features of *C. porrectum*. Using kyoto encyclopedia of genes and genomes (KEGG) pathway analysis technology, we explored the genes of *C. porrectum* involved in the metabolism pathways of terpenoids and further analyzed the differentially expressed genes (DEGs) of monoterpenoids in the three different chemotypes. We also preliminarily screened candidate genes for the biosynthesis of several important monoterpenoids. qRT-PCR verification of candidate genes was performed, and key genes in the synthesis of monoterpenoids, such as linalool, eucalyptol and camphor, were screened. This study lays a foundation for the cloning of key genes in the synthesis of important monoterpenoids of *C. porrectum*, functional research, and the genetic improvement and study of the special economic values of *C. porrectum*.

2. Results

2.1. The Composition of Leaf Extracts from Different Chemotypes of *C. porrectum*

The leaf essential oil of three chemotypes (the eucalyptol chemotype, Y_L1, Y_L2 and Y_L3; the linalool chemotype, F_L1, F_L2 and F_L3; the camphor chemotype, N_L1, N_L2 and N_L3) of *C. porrectum* were extracted by steam distillation, and the chemical constituents of the leaf essential oil were analyzed by GC/MS. The oil yields of the eucalyptol chemotype, linalool chemotype and camphor chemotype were 11.5–16.5 mg/g, 18.6–21.4 mg/g and 6.1–9.6 mg/g, respectively. A total of 32 compounds were detected in the leaf essential oil, with 20 monoterpenoids, 10 sesquiterpenoids and 2 phenolic compounds. The total content of monoterpenoids accounted for more than 88.04% of the total content of essential oils. In the three chemotypes, the content of the first-principal-component compound was much higher than that of other compounds. The eucalyptol content accounted for more than 34.48% of total content in the eucalyptol chemotype, the linalool content accounted for more than 77.58% of the total content in the linalool chemotype, and the camphor content accounted for more than 92.47% of the total content in the camphor chemotype, that is, the monoterpenoids were almost pure camphor (Table 1).

Table 1. Composition of the leaf extracts of *C. porrectum*.

R1a	R1b	Component	Y_L1 (%)	Y_L2 (%)	Y_L3 (%)	F_L1 (%)	F_L2 (%)	F_L3 (%)	N_L1 (%)	N_L2 (%)	N_L3 (%)
		Monoterpenes	92.98	93.65	90.04	89.46	88.04	88.33	99.01	92.47	98.32
929	928	3-Thujene	0.51	0.51	0.28	-	-	-	-	-	-
939	935	α -Pinene	3.16	3.47	2	-	-	-	-	-	-
977	1006	β -Phellandrene	12.16	9.9	9.53	-	-	-	-	-	-
986	978	β -Pinene	3.47	3.41	2.55	-	-	-	-	-	-
989	988	Myrcene	1.78	1.99	1.76	-	-	-	-	-	-
1011	1002	α -Phellandrene	0.37	0.46	0.16	-	-	-	-	-	-
1021	1014	α -Terpinene	1.5	2.39	1.57	-	-	-	-	-	-
1033	1031	D-Limonene	-	-	-	tr	0.2	0.67	-	-	-
1037	1037	Eucalyptol	35.63	34.67	34.48	0.1	0.23	0.59	tr	tr	tr
1061	1058	γ -Terpinene	4.21	5.16	4.15	tr	-	-	-	-	-
1104	1102	Linalool	1.36	0.81	1.64	87.88	86.46	77.58	tr	tr	tr
1108	1108	cis- β -Terpineol	0.57	0.44	0.54	-	-	-	-	-	-
1148	1139	Pinocarveol	1.93	1.92	2.09	-	-	-	-	-	-
1154	1150	Camphor	-	-	0.1	1.13	tr	7.85	97.78	92.47	97.04
1178	1172	Borneol	0.25	0.71	0.57	tr	0.55	0.37	1.23	-	1.28
1185	1176	4-Terpineol	6.18	8.53	7.44	tr	tr	0.23	-	-	-
1198	1193	α -Terpineol	19.75	19.15	20.96	0.35	0.6	0.89	-	-	-
1213	1207	Piperitol	0.16	0.15	0.16	-	-	-	-	-	-
1226	1223	β -Citronellol	0.32	0.35	0.34	-	-	-	-	-	-
1318	1318	Pinanediol	0.18	0.14	-	-	tr	0.15	-	-	-
		Sesquiterpenes	5.64	4.83	9.27	8.41	10.45	8.44	0.00	3.64	0.00
1430	1418	β -Caryophyllene	2.29	4.17	3.41	4.92	2.33	1.62	-	-	-
1466	1454	α -Caryophyllene	0.35	0.5	0.4	0.69	0.92	0.63	-	-	-
1471	1461	Alloaromadendrene	-	-	0.19	0.13	0.21	-	-	-	-
1491	1484	Germacrene D	2.39	-	1.2	0.44	3.82	1.09	-	-	-
1505	1503	β -Chamigrene	-	0.16	0.25	0.11	2.78	1.06	-	-	-
1562	1563	trans-Nerolidol	0.61	-	3.82	1.58	0.29	2.25	-	-	-
1588	1577	(-)-Spathulenol	-	-	-	tr	tr	0.3	-	-	-
1594	1596	Caryophyllene oxide	-	-	-	0.17	-	-	-	1.91	-
1665	1653	α -Eudesmol	-	-	-	-	0.1	1.24	-	-	-
1786	1762	Aristolone	-	-	-	0.37	-	0.25	-	-	-
		Phenols	0.19	0.00	0.00	0.00	0.35	0.18	0.00	0.00	0.00
1398	1404	Methyleugenol	0.19	-	-	tr	0.2	0.18	-	-	-
1495	1491	Methylisoeugenol	-	-	-	tr	0.15	-	-	-	-
		Total %	98.81	98.48	99.31	97.87	98.84	96.95	99.01	94.38	98.32
		Oil yield %	1.50	1.65	1.15	1.90	2.14	1.86	0.61	0.78	0.96

RT: Retention time; R1a: Retention indices calculated against n-alkanes, R1b: Retention indices reported in the literature [37–40]. tr, trace (<0.1%); ‘-’ means ‘not found’. Y_L1, Y_L2 and Y_L3 were eucalyptol chemotype; F_L1, F_L2 and F_L3 were linalool chemotype; N_L1, N_L2 and N_L3 were camphor chemotype. Quantification was done using the calibration curves from the analyses of following standards. β -pinene was the standard of alkenes; eucalyptol was the standard of alcohols; 2-Dodecanone was the standard of ketones; eugenol was the standard of phenols.

2.2. RNA Sequencing and Transcriptomic Assembly

Full-length transcriptome of *C. porrectum* was generated from normalized full-length cDNAs from pooled poly(A) RNA of one sample (leaf) and subjected to single molecule real-time (SMRT) sequencing via the PacBio Sequel system. In total, two SMRT cells generated 9,624,949 subreads (13.3 Gb); the average read length was 1389.96 bp and an N50 were 2147 bp. A total of 683,767 reads of insert (ROIs) were generated and 104,062 those sequences containing two primers and a poly-A tail were identified as full-length reads of inserts(ROIs), which were subsequently classified into chimeric and nonchimeric reads. Of these ROIs, 289,102 were identified as full-length non-chimeric (FLNC) reads with an average read length 2819.5 bp (Figure 1a). Full-length nonchimeric sequences were clustered, and with the Quiver algorithm, each cluster was corrected and integrated into a consensus sequence. The consensus sequence output by Quiver was divided into low-quality sequences (low QV, LQ) and high-quality sequences (high QV, HQ). A total of 213586 full-length consensus transcripts, including 133,966 HQ (> 99% accuracy) transcripts and 79620 LQ transcripts, were generated. Information on each library of clustered reads is shown in Table S1, and the sequence length and sequence quality of the low-QV and high-QV sequences were compared (Figure S1). Subsequent analyses only used high-QV consensus sequences. Iso-Seq produced more unigenes with lengths > 2000 bp than Illumina sequencing.

These results indicated that PacBio Iso-Seq provides a practical method for producing full-length transcripts without assembly steps and is an important improvement for transcriptome studies of species without reference genomes.

Library clusters and high-quality sequences obtained after error correction were finally merged to eliminate redundancy. After this process, the transcriptome of *C. porrectum* had a total of 104,062 transcripts with a total of 306,337,921 bp; the average sequence length was 2944 bp, and the N50 length was 5449 bp (Table 2). For the length distribution of subreads, see Figure 1b, and the length distribution of nonredundant transcripts is shown in Figure 1c.

Table 2. Summary of the final transcript sequences obtained after the elimination of redundancy.

Sample	Total Isoforms	Total Base (bp)	Mean Length (bp)	N50 (bp)
Total	104,062	306,337,921	2944	5449

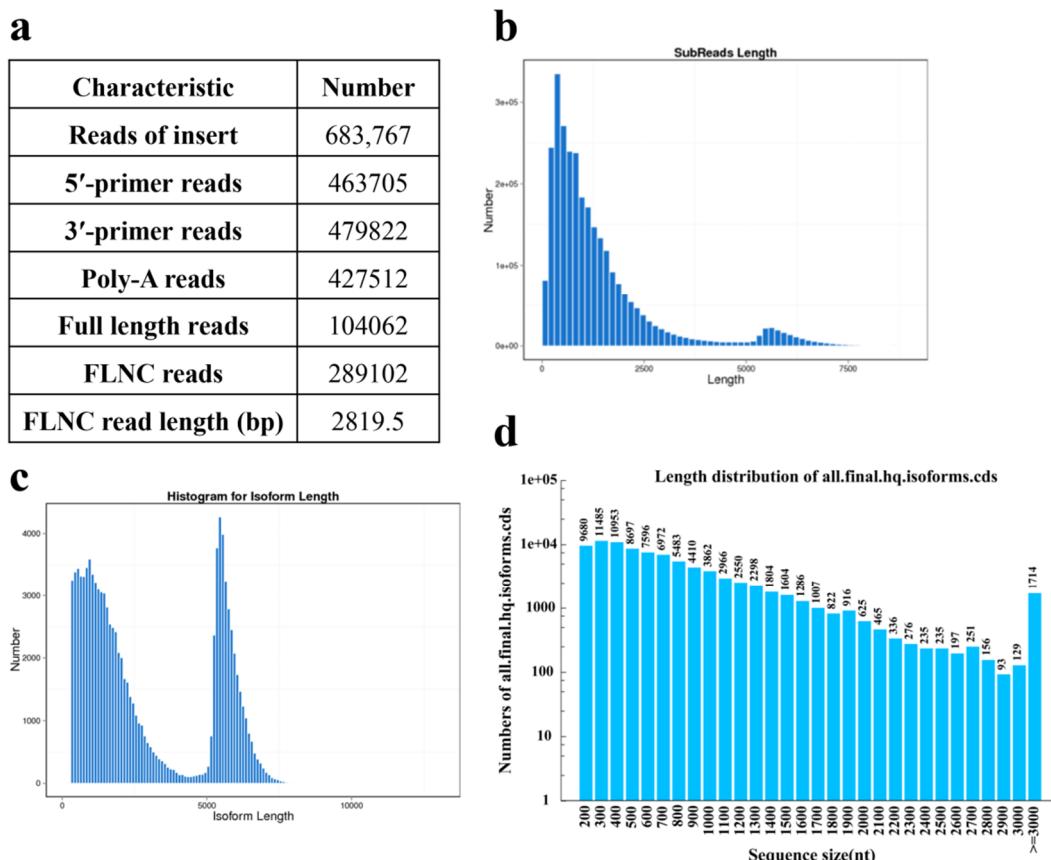


Figure 1. PacBio single-molecule long-read sequencing of *C. porrectum*. (a) Reads classification from PacBio Iso-Seq. (b) distribution of subreads length. (c) Read length distribution. (d) Length distribution of all isoforms coding sequence(CDS).

After clustering and error correction, the TransDecoder software was used to identify the candidate coding regions in each transcript. First, the longest open reading frame was extracted. Then, using BLAST, the SwissProt database was compared with Hmmscan to search for homology to Pfam domains to predict the coding regions. A total of 89,103 complete transcripts were obtained. The details of the prediction results are listed in Table 3, and the coding sequence(CDS) length distribution is shown in Figure 1d.

Table 3. Statistics of CDS prediction results.

Software	Total Number	Total Length	Mean Length	N50	N70	N90	GC (%)
Blast	88763	66645204	750	1041	687	363	44.66
ESTScan	340	205098	603	729	501	336	43.75
Overall	89103	66850302	750	1038	687	363	44.66

2.3. Gene Annotation and Functional Classification

After transcripts were clustered and corrected, functional annotations (NCBI non-redundant protein sequences(NR), NCBI non-redundant nucleotide sequences(NT), Gene ontology(GO), EuKaryotic Orthologous Groups(KOG), KEGG, SwissProt, Interpro) were performed using BLAST, Blast2GO and InterProScan5, and a total of 104,062 transcripts were annotated (Figure 2A). The total number of transcripts annotated by any database was 94,025, accounting for 90.53% of all transcripts. The NR database had the highest annotation rate, accounting for 85.52% of all transcripts and the GO database had the lowest annotation rate, accounting for 34.26% of all transcripts. The total number of transcripts annotated by all seven databases was 21,781, accounting for 20.93% of all transcripts. The overall status of the annotations is shown in Table S2. To better understand the full-length transcripts that were obtained, their functions were investigated using the KEGG pathway database; a total of 104,062 genes were involved in 19 metabolic pathways. The unigenes were divided into six branches according to the KEGG metabolic pathway: organismal systems (A), metabolism (B), human diseases (C), genetic information processing (D), environmental information processing (E), and cellular processes (F). It was found that 1495 unigenes were enriched in the metabolic pathways of terpenoids and polyketides (Figure 2B). The obtained unigenes were GO annotated, and the genes that were annotated successfully were classified according to the 54 biological functional classifications under the three major GO categories: biological process, cellular component and molecular function. In the biological process category, there were relatively high numbers of genes involved in the cellular process, metabolic process, and single-organism process classifications. In the cellular component category, there were more genes involved in the cell, cell part, membrane, and binding processes classifications than in other classifications. In the molecular function category, genes were mainly classified into catalytic activity, structural molecule activity and transporter activity (Figure 2C). The above results reveal a comprehensive functional characterization of the full-length *C. porrectum* transcriptome, and will help further study on gene functions.

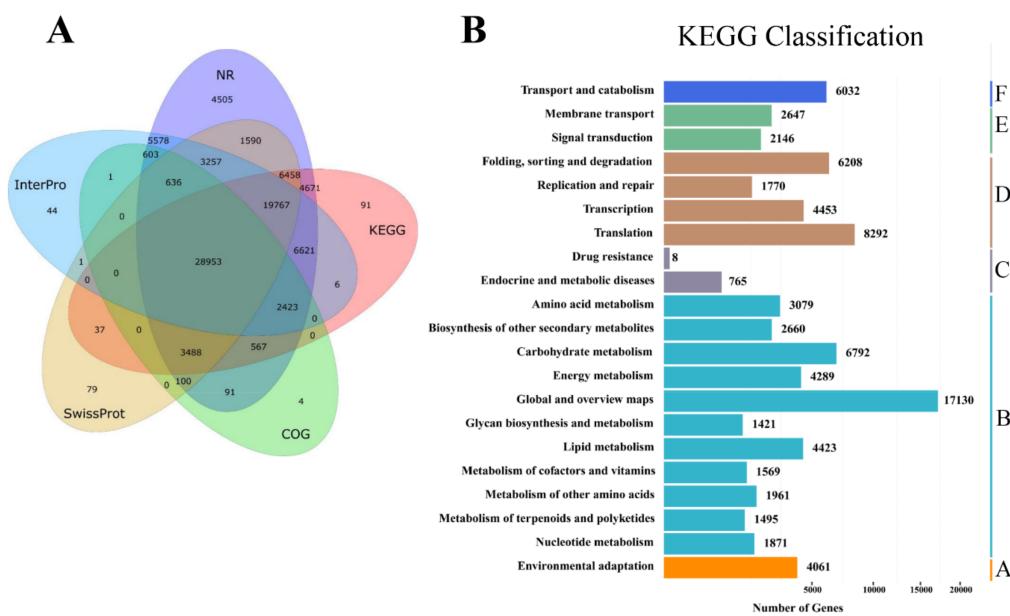


Figure 2. Cont.

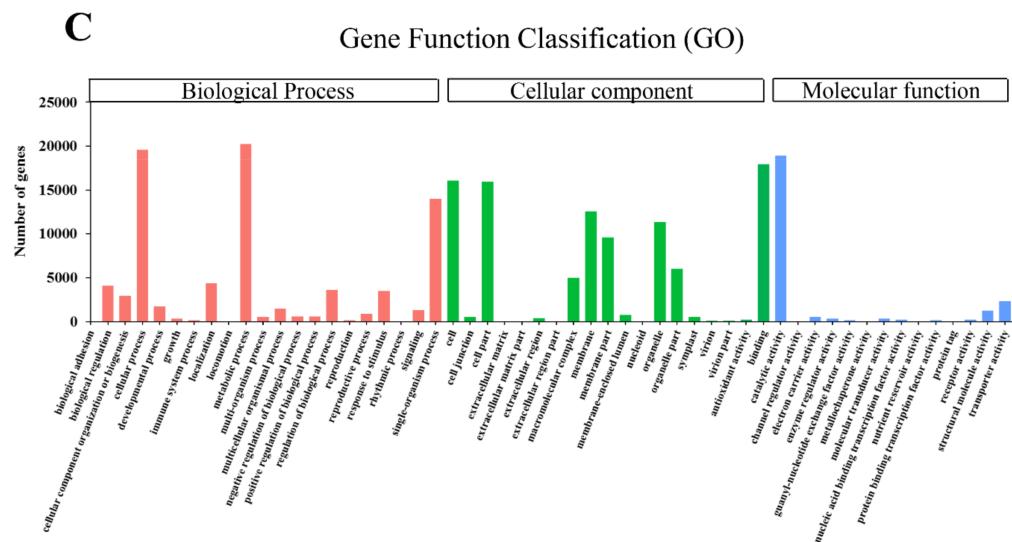


Figure 2. Functional annotation and classification of assembled unigenes in *C. porrectum*. (A) Annotation information of the Venn diagram. (B) KEGG pathway classification of putative proteins. The unigenes were divided into six branches according to the following KEGG metabolic pathways: organismal systems A. metabolism B. human diseases C. genetic information processing D. environmental information processing E. and cellular processes F. (C) GO classification of unigenes in *C. porrectum*. GO terms are classified into three main categories: biological process (BP), cellular component (CC) and molecular function (MF).

2.4. The Identification of Relative DEGs in *C. porrectum* Chemotypes and Enrichment Analysis of Transcripts

The expression profiles of the *C. porrectum* leaf tissues from the eucalyptol, linalool and camphor chemotypes underwent sequencing analysis. The flux and quality of the RNA-Seq data are shown in Table 4. There was a total of 384.31 M clean read data obtained from the three different chemotypes with three biological repetitions, and the total degree of mapping was between 73% and 82%.

Table 4. Summary of RNA-seq data from nine RNA libraries of eucalyptol chemotype, camphor chemotype and linalool chemotype of *C. porrectum*.

Sample	Total Raw Reads (M)	Total Clean Reads (M)	Total Clean Bases (Gb)	Clean Reads Q20 (%)	Clean Reads Q30 (%)	Total Mapping (%)
AYYS1	45.72	42.54	6.38	98.79	96.38	80.41
AYYS2	45.72	42.45	6.37	98.82	96.46	79.94
AYYS3	45.72	42.36	6.35	98.78	96.37	81.93
NZ1	47.35	43.01	6.45	98.68	96.07	78.49
NZ2	47.35	43.24	6.49	98.73	96.2	81.5
NZ3	47.35	42.95	6.44	98.63	95.93	78.7
YXFZC1	47.35	42.51	6.38	98.27	94.78	76.74
YXFZC2	47.35	42.5	6.38	98.29	94.82	73.63
YXFZC3	47.35	42.75	6.41	98.31	94.9	74.33
summary	421.26	384.31	57.65	/	/	/

AYYS1, AYY2 and AYY3 were eucalyptol chemotype; NZ1, NZ2 and NZ3 were camphor chemotype; YXFZC1, YXFZC2 and YXFZC3 were linalool chemotype.

To fully explore the potential differential gene expression between the linalool and eucalyptol chemotypes, between the camphor and linalool chemotypes, and between the camphor and eucalyptol chemotypes, the clean reads were mapped to the unigene database. Gene expression data were standardized using a variety of calibration methods, and the DEGs between different chemical types were characterized by differentially expressed sequence (DESeq) with a *Q*-value < 0.05 and |log2. Fold_change| > 1 [41]. In the eucalyptol chemotype and linalool chemotype comparison group, a total of 31,508

unigenes were identified, including 20,270 upregulated genes and 11,238 downregulated genes. A total of 2975 unigenes were unique to the eucalyptol chemotype, 5391 unigenes were unique to the linalool chemotype, and 23,142 unigenes were shared by the two chemotypes (Figure 3A). In the camphor chemotype and eucalyptol chemotype comparison group, a total of 25,522 unigenes were identified, including 10,834 upregulated genes and 14,688 downregulated genes. Of these unigenes, 4969 were unique to the camphor chemotype, 3624 were unique to the eucalyptol chemotype, and 16,929 were shared by the two chemotypes (Figure 3B). In the camphor chemotype and linalool chemotype comparison group, a total of 29,691 unigenes were identified, including 17,520 upregulated genes and 12,171 downregulated genes. A total of 3011 unigenes were unique to the camphor chemotype, 4160 unigenes were unique to the linalool chemotype, and 22,520 genes were shared by the two chemotypes (Figure 3C). To explore the differences in metabolic pathways between the linalool and eucalyptol chemotypes, between the camphor and linalool chemotypes, and between the camphor and eucalyptol chemotypes, KEGG enrichment pathways were analyzed using 31,508, 25,522 and 29,691 DEGs, respectively. The top 20 KEGG pathway categories are presented in the form of a dot plot (Figure S2A–C). The pathway with the most significant enrichment of DEGs in the camphor and linalool chemotype comparison group was the “monoterpene biosynthesis” pathway.

The “monoterpene biosynthesis” pathway was also enriched significantly in the camphor and eucalyptol chemotype comparison group.

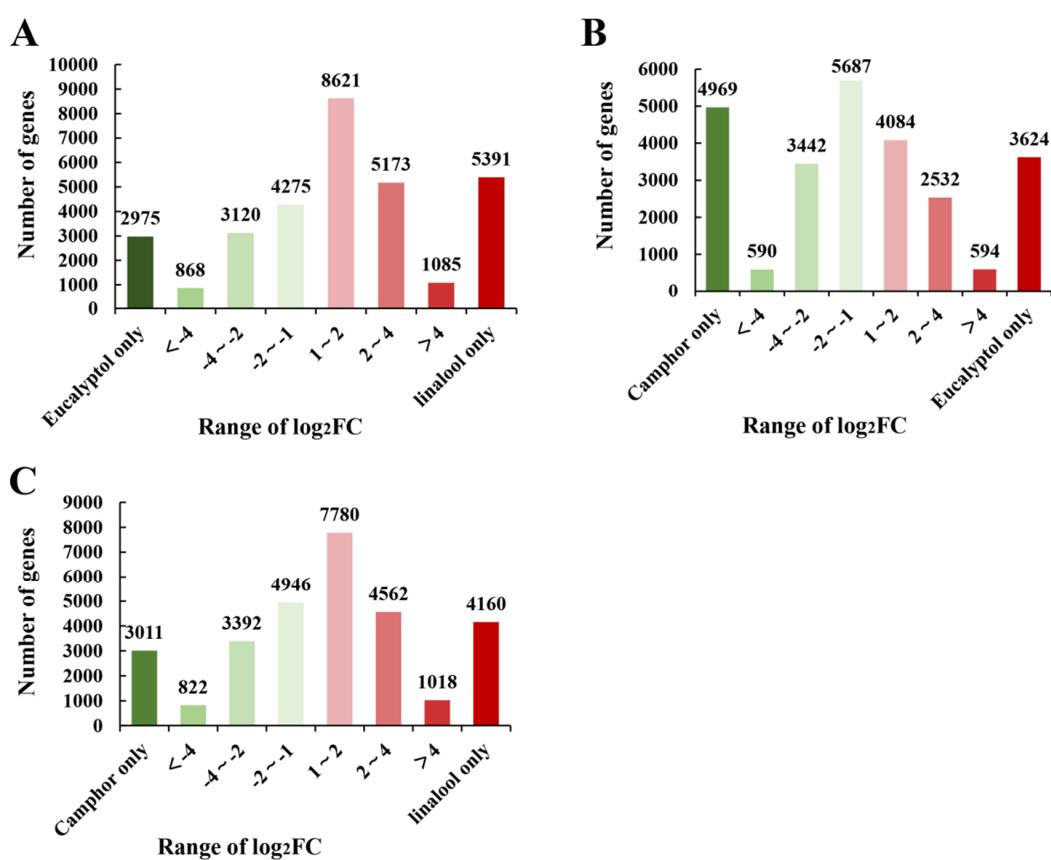


Figure 3. (A) Number of DEGs in the eucalyptol chemotype and linalool chemotype of *C. porrectum*. Red: upregulated genes; green: downregulated genes. Relative expression of DEGs selected at a *Q*-value < 0.05. The fold change (FC) was calculated as the ratio between the linalool chemotype and eucalyptol chemotype. The x-axis represents the range of Log₂ FC. The y-axis indicates the number of detected DEGs. (B) Number of DEGs in camphor chemotype and eucalyptol chemotype of *C. porrectum*. The fold change (FC) was calculated as the ratio between the eucalyptol chemotype and camphor chemotype. (C) Number of DEGs in camphor chemotype and linalool chemotype of *C. porrectum*. The fold change (FC) was calculated as the ratio between the linalool chemotype and camphor chemotype.

2.5. Candidate Genes Involved in Terpenoid Biosynthesis

To explore the regulatory mechanisms for the accumulation patterns of different terpenoids in *C. porrectum*, the expression profiles of genes involved in terpenoid biosynthesis were analyzed. In the eucalyptol chemotype and linalool chemotype, the camphor chemotype and eucalyptol chemotype, and the camphor chemotype and linalool chemotype comparison groups, 52, 49 and 66 DEGs for the biosynthesis of terpenes were identified, respectively (Figure 4, Figure S3 and Figure S4). The expression level of all these unigenes are shown in the Figure 4, Figures S3 and S4, and the fragments per kilobase per million mapped (FPKM) values are shown in Table S3. Most of the genes exhibited a high transcriptome expression level, which encoding key enzymes in the MEP and MVA pathway (KEGG entry ko00900). The expression levels of different unigenes among the three different chemotypes were different, but the unigene expression level of the MEP pathway was higher than that of the MVA pathway as a whole (Figure 4, Figures S3 and S4). Both MEP and MVA pathways produce IPP and its isomer, dmap. The active biosynthesis of building blocks contributes to the accumulation of various terpenoids, which is consistent with our analysis in the components of the extracts from the leaves of *C. porrectum* (Table 1). Moreover, in the eucalyptol chemotype and linalool chemotype comparison group, we screened 21 unigenes encoding TPS, namely, 7 monoterpane synthases, 3 diterpenoid synthases and 11 sesquiterpene and triterpene synthases (Figure 4). The eucalyptol chemotype upregulated 5 unigenes of monoterpane synthases, 1 unigene of diterpene synthase and 4 unigenes involving sesquiterpene and triterpene synthases. However, in the linalool chemotype, 2 unigenes of monoterpane synthases, 2 unigenes of diterpene biosynthesis, and 7 unigenes of sesquiterpene and triterpene synthases were upregulated (Figure 4). In the camphor chemotype and eucalyptol chemotype comparison group, we screened 22 unigenes encoding TPS, namely, 9 monoterpane synthases, 3 diterpenoid synthases and 10 sesquiterpene and triterpene synthases (Figure S3). The camphor chemotype upregulated 3 unigenes of monoterpane synthases, 2 unigenes of diterpene synthases and 6 unigenes of sesquiterpene and triterpene synthases. However, in the eucalyptol chemotype, 6 unigenes of monoterpane synthases, 1 unigene of diterpene biosynthesis, and 4 unigenes of sesquiterpene and triterpene synthases were upregulated (Figure S3). In the camphor chemotype and linalool chemotype comparison group, we screened 18 unigenes encoding TPS, namely, 5 monoterpane synthases, 2 diterpenoid synthases and 11 sesquiterpene and triterpene synthases (Figure S4). The camphor chemotype upregulated 2 unigenes of monoterpane synthases, 2 unigenes of diterpene synthase and 5 unigenes of sesquiterpene and triterpene synthases. However, in the linalool chemotype, 3 unigenes of monoterpane synthases and 6 unigenes of sesquiterpene and triterpene synthases were upregulated, and all the unigenes of diterpene synthases were downregulated (Figure S4). On the whole, the observed high expression levels of enzymes in the terpene skeleton pathway were consistent with the high rate of terpenoid synthesis and suggest that the different terpenoid compositions in the three chemotypes may arise from differences in the expression levels of TPS genes.

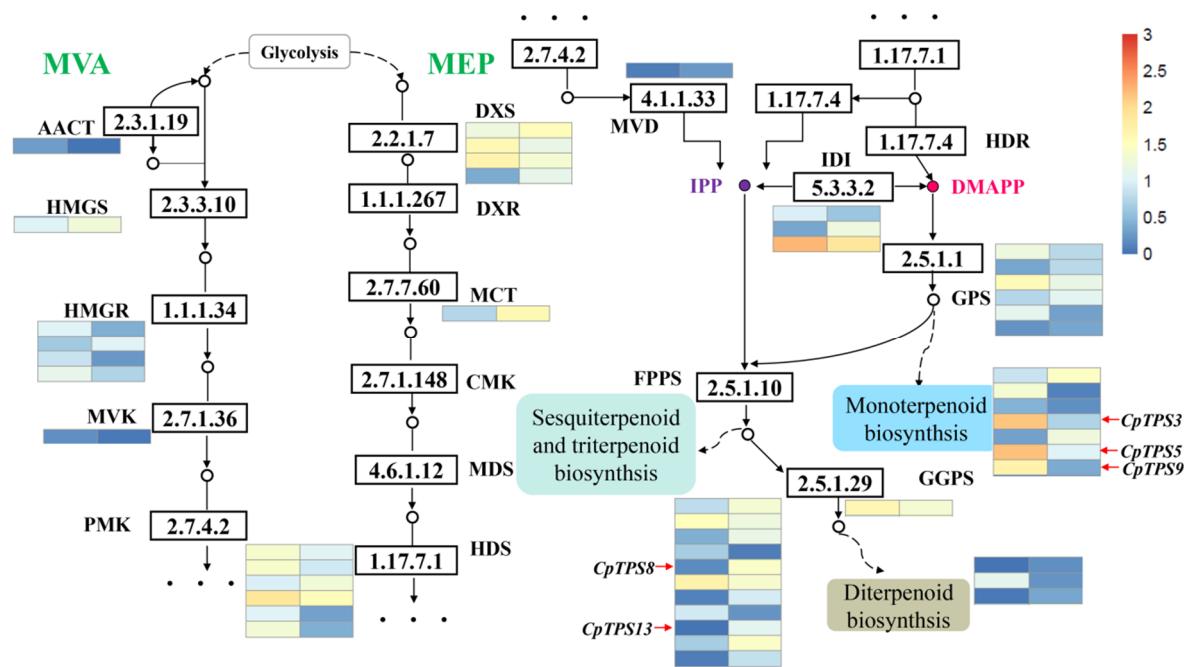


Figure 4. Differentially expressed genes (DEGs) in terpenoid backbone biosynthesis in the eucalyptol chemotype and linalool chemotype comparison group of *C. porrectum*. Enzymes expression patterns are indicated with the value of $\log(\text{FPKM} + 1)$. The expression pattern is shown within two columns, with the left column representing the eucalyptol chemotype and the right representing the linalool chemotype. Acetyl-CoA C-acetyltransferase (CCAT); Hydroxymethylglutaryl-CoA synthase (HMGS); Hydroxymethylglutaryl-CoA reductase(HMGR); Mevalonate kinase (MVK); Phosphomevalonate kinase (PMK); Diphosphomevalonate decarboxylase (MVD); 1-deoxy-D-xylulose-5-phosphate synthase (DXS); 1-deoxy-Dxylulose-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); (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS); 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase(HDR); isopentenyl-diphosphate Delta-isomerase (IDI).

2.6. Construction of Phylogenetic Tree of TPS

Terpene synthase is the main cause of the diversity of terpenoids. The mechanism of terpenoid accumulation in different chemotypes of *C. porrectum* was investigated by constructing a phylogenetic tree of TPS (Figure 5). In combination with terpene synthase genes with currently known functions, we constructed a phylogenetic tree of TPS sequences and divided the terpene genes into six subfamilies. Among these subfamilies, the b subfamily responsible for the synthesis of monoterpenoids had the largest number of genes, with 15 members, while there were 12 genes in a subfamily responsible for the synthesis of sesquiterpenes, 0 genes in the c subfamily, 0 genes in the d subfamily, 4 genes in the e/f subfamily, and 2 genes in the g subfamily.

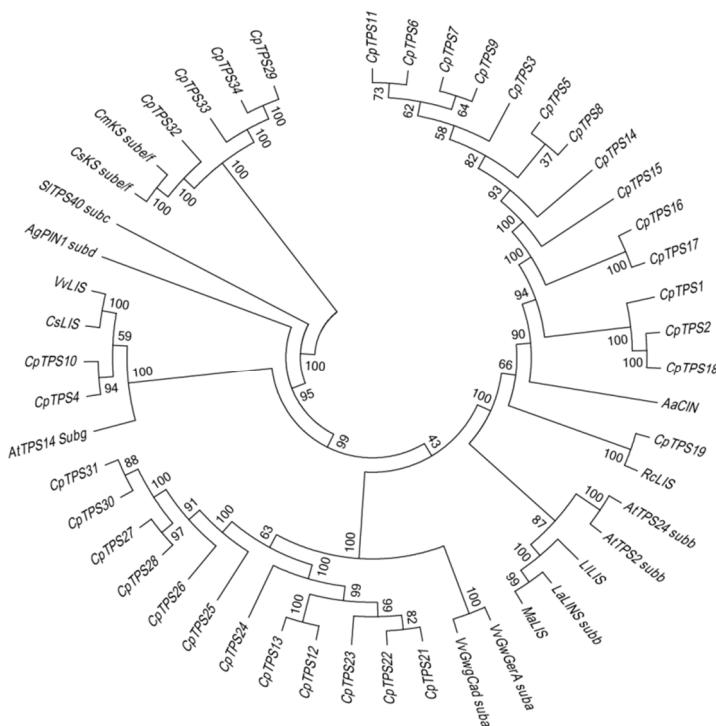


Figure 5. Results of phylogenetic tree of *CpTPS* gene family and representative terpene synthase of the known function in other plants using the neighbor-joining method by MEGA7 software. Bootstrap values are shown as a percentage of 1000 replicates. The sequences of genes in phylogenetic tree of *CpTPS* were shown in Table S4. The accession number for characterized TPS in the phylogenetic tree are: *AtTPS2*(NP_193406.3); *AtTPS14*(NP_001185286.1); *AtTPS24*(NP_189209.2); *SITPS40* (NP_001234008.2); *AgPIN1* (O24475.1); *VvGwgGerA* (ADR66821.1); *VvGwgCad* (ADR74199.1); *LaLIS* (Q2XSC5.1); *CmKS*(XP_022968895.1); *CsKS*(NP_001292675.1); *VvLIS*(ADR74212.1); *LiLIS*(ABD77417.1); *RclIS*(AVR48793.1); *MaLIS*(AAL99381.1); *AaCIN*(PWA51422.1); *CsLIS*(AGX26045.1). The function of known TPS genes in the phylogenetic tree were shown in Table S8.

2.7. qRT-PCR Validation of DEGs from the RNA-Seq Analysis

To verify the expression pattern of the terpene biosynthetic genes obtained from RNA-Seq analysis, the expression of 12 unigenes was examined by qRT-PCR (Figure 6). The qRT-PCR-determined expression levels of these genes were generally consistent with the expression levels deduced from RNA-Seq FPKM data (Figure 6). The results confirm the reliability of the transcriptomic profiling data from RNA-Seq. In addition, we selected 6 of these DEGs and detected their expression in five chemotypes (linalool-, eucalyptol-, camphor-, nerolidol- and citral-type) using qRT-PCR. Notably, *CpTPS3*, *CpTPS5* and *CpTPS9* were specifically expressed in the eucalyptol chemotype; *CpTPS1* was specifically expressed in the camphor chemotype; *CpTPS4* was expressed in all five chemotypes, with the maximum expression in the linalool chemotype. These results, which basically agree with the transcriptome results, also validate the reliability of the transcriptome (Figure 7).

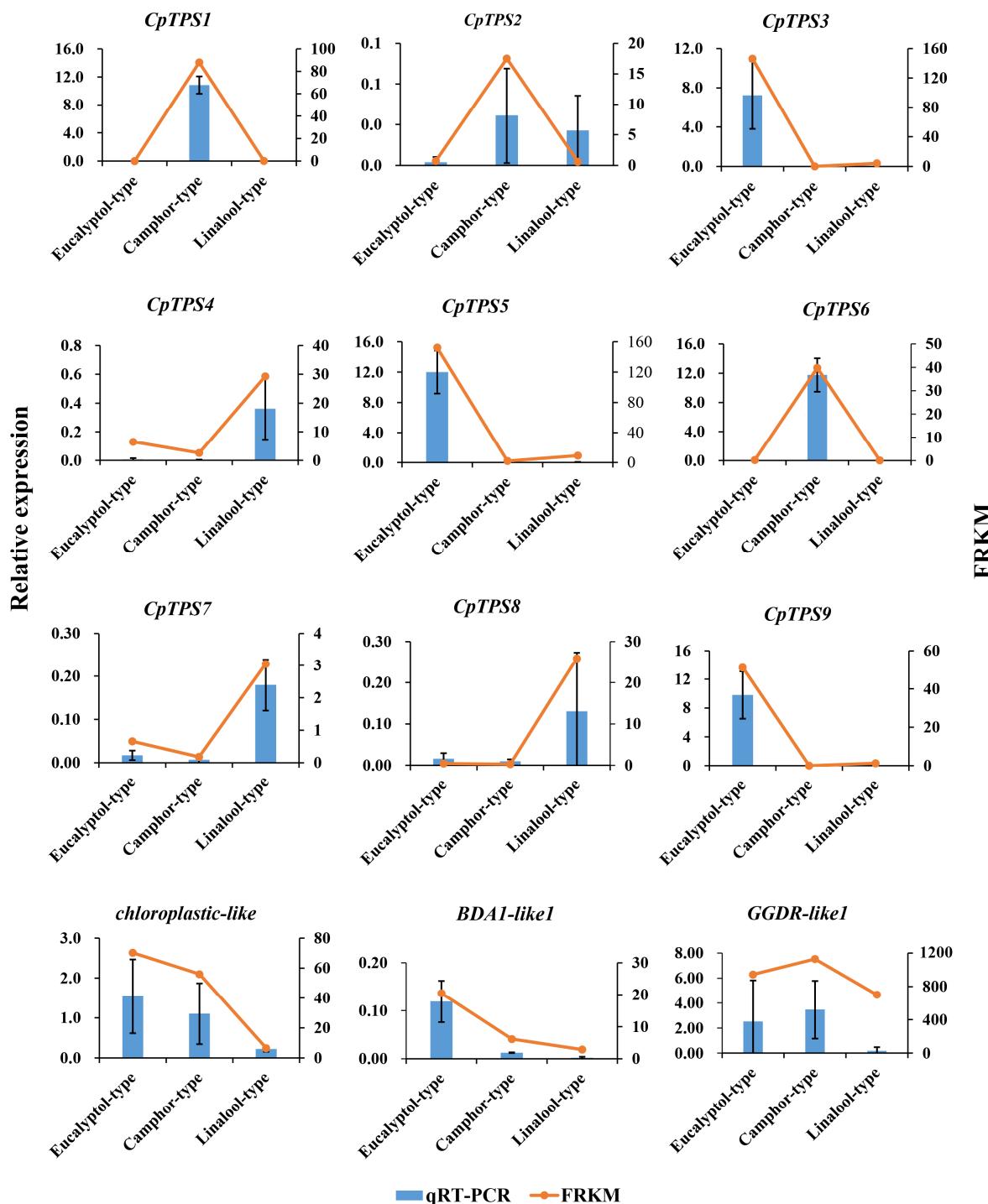


Figure 6. Expression analyses of selected genes (Table S5) using both qRT-PCR and RNA-Seq (FC of FPKM). The relative expression levels were estimated from the threshold of the PCR cycle with the delta CT method. The error bars indicate the standard errors of three biological replicates. Composition of the leaf extracts of the three chemotypes are shown in Table 1.

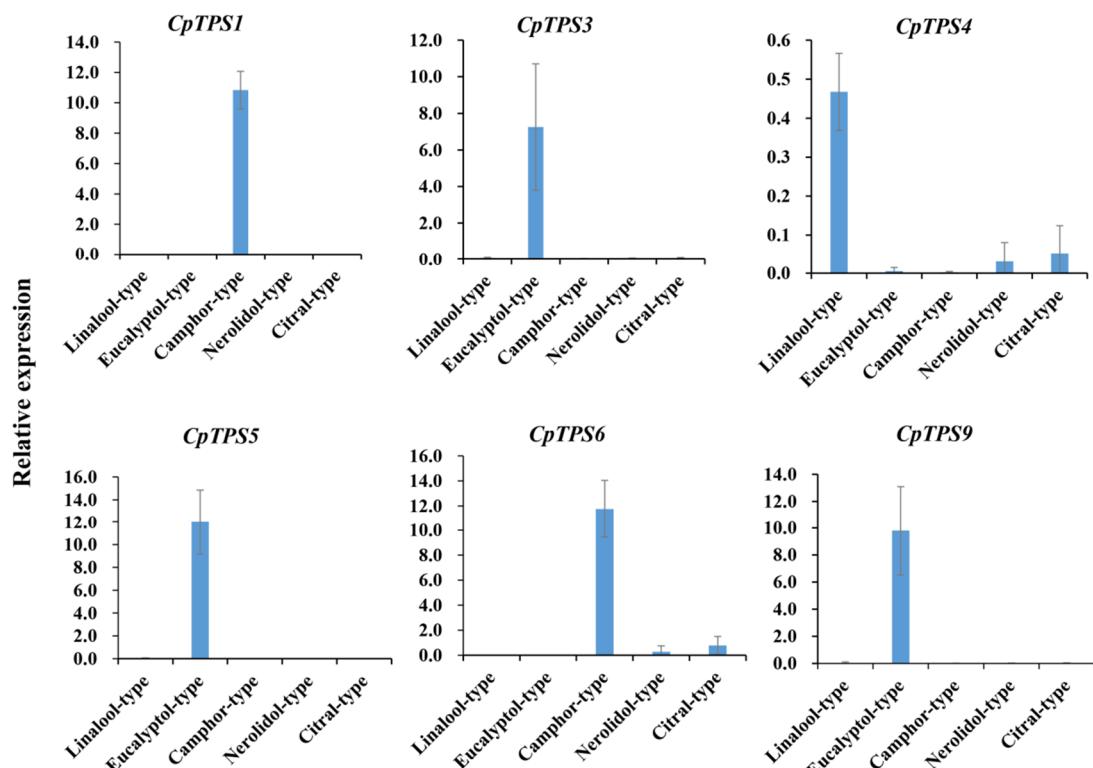


Figure 7. qRT-PCR validation of selected genes in five chemotypes of *C. porrectum*. The five chemotypes are the linalool-, eucalyptol-, camphor-, nerolidol and citral-type. Relative expression levels were estimated from the threshold of the PCR cycle using the Delta CT method. The error bars indicate the standard errors of three biological replicates. Composition of the leaf extracts of the linalool-type, eucalyptol-type and camphor-type are shown in Table 1; Composition of the leaf extracts of the nerolidol-type and citral-type are shown in Table S9.

3. Discussion

At present, research on the *Cinnamomum camphora* transcriptome has mainly been based on cost-effective next-generation RNA sequencing [13]. However, because of the short read length of this method, it is difficult to obtain an accurate full-length sequence using next-generation RNA sequencing without a reference genome [42]. PacBio Iso-Seq, the most popular application of third-generation sequencing, has proven useful for accurately characterizing the diverse landscape of isoforms, as this method can obtain full-length transcripts [43]. Using the PacBio Iso-Seq method, this study provided the first comprehensive set of full-length isoforms in *C. porrectum*. After eliminating redundancy, a total of 104,062 transcripts were obtained, with a total of 306,337,921 bp, an average sequence length of 2944 bp, and an N50 length of 5449 bp. Based on these highly accurate transcripts, a total of 104,062 transcripts were successfully annotated by at least one of seven known functional protein databases (NR, NT, GO, KOG, KEGG, SwissProt, Interpro). Furthermore, 104,062 genes were involved in a metabolic pathway of 19 KEGG pathway databases, and 1495 unigenes were found to be enriched in pathways of terpenoid and polyketide metabolism. These results indicate that PacBio transcriptome sequencing has a high ability to generate full-length transcript sequence information, providing an important reference for genome annotation and gene function analysis.

In recent years, genes associated with terpene biosynthesis have been studied extensively in different plants. Lauraceous plants are rich in terpenoids, and several genes involved in the biosynthesis of terpenoids have been successfully identified, and their functions have been explored [26,29]. Three TPS genes encoding monoterpene synthases were isolated in *Litsea cubeba*: *LcTPS1*, which converts geranyl diphosphate to trans-ocimene, *LcTPS2*, which converts geranyl diphosphate to α -thujene,

and *LcTPS3*, a multifunctional enzyme that converts geranyl diphosphate to α -thujene and (+)-sabinene [44]. In *Laurus nobilis*, the formation of 1,8-cineole, cadinenes and geranylinalool catalyzed by TPS enzymes was characterized [28]. Sixty-seven unigenes were isolated from the transcriptome of *C. camphora*, which might be involved in terpene biosynthesis. The expression of *TPS14-like1*, *TPS14-like2* and *TPS14-like3* were upregulated in the borneol chemotype relative to the linalool chemotype. These results are helpful to better understand the differential accumulation of terpenoids in different chemotypes of *C. camphora* [13]. This study analyzed the transcriptomes of three chemotypes of *C. porrectum*. In the eucalyptol chemotype and linalool chemotype, the camphor chemotype and eucalyptol chemotype, and the camphor chemotype and linalool chemotype comparison groups, 52, 49, and 66 candidate unigenes, respectively, were identified to be related to terpene biosynthesis. Furthermore, the terpenoid synthase genes *CpTPS1*, *CpTPS3*, *CpTPS4*, *CpTPS5* and *CpTPS9* had specific expression or higher expression in one of the chemotypes and may be involved in monoterpene biosynthesis.

It is now generally accepted that MVA and MEP pathway activities are regulated at both the transcript and protein levels to control precursor availability [45]. Gene coexpression networks have revealed that there is no en bloc transcriptional regulation of all genes encoding MVA and MEP pathway enzymes. Isoprenoids synthesized via MVA and MEP pathways are controlled by independent regulatory networks with restricted connectivity [45].

HMGR and DXS are two rate-limiting enzymes in the MVA pathway [46] and the MEP pathway [47]. In the transcriptomes of the linalool- and borneol-chemotypes of *C. camphora*, 30 unigenes, which were highly homologous to 14 known enzymes, were annotated in the MVA and MEP pathways of terpene skeleton biosynthesis in the borneol, and high transcriptome expression levels were shown in both chemotypes [13]. In *C. porrectum*, there were 21 unigenes encoding TPS in the eucalyptol chemotype and linalool chemotype comparative group, 5 monoterpene synthase unigenes upregulated in the eucalyptol chemotype, and 2 monoterpene synthase unigenes upregulated in the linalool chemotype. There were 22 unigenes encoding TPS in the camphor chemotype and eucalyptol chemotype comparative group, 3 upregulated monoterpene synthase unigenes in the camphor chemotype, and 6 upregulated monoterpene synthase unigenes in the eucalyptol chemotype. There were 18 unigenes encoding TPS in the camphor chemotype and linalool chemotype comparative group, 2 upregulated monoterpene synthase unigenes in the chemotype camphor, and 3 upregulated monoterpene synthase unigenes with upregulated expression in linalool chemotype. These results indicate that there may be differences in the gene expression patterns for the regulation and formation of monoterpenoids among different chemotypes. The MEP pathway provides precursors for the synthesis of monoterpenes and diterpenes in plastids, while sesquiterpenes are derived from precursors of the MVA pathway in the cytoplasm [45]. The unigene expression level of the MEP pathway was higher than that of the MVA pathway indicating the abundance of monoterpenes. Crosstalk between these two different terpene skeleton pathways has been documented, and the relative contribution of each pathway to terpene biosynthesis remains uncertain.

TPS genes are responsible for the synthesis of diverse terpenoids molecules from two isomeric isoprene precursors. The phylogenetic tree of *C. porrectum* with the characterized TPS proteins is divided into seven clades, with a majority expansion of TPS-a and TPS-b clades, indicating gene family expansion is correlated with mono- and sesquiterpene diversity.

Monoterpenes are the principal components of many member leaf extracts of *Cinnamomum* species, such as *Cinnamomum osmophloeum* [26], *Cinnamomum kanehirae* [48] and *Cinnamomum camphora* [13]. In this paper, 33 components were identified in the essential oil from leaf tissue of the eucalyptol chemotype, linalool chemotype and camphor chemotype of *C. porrectum*. Analysis of the components showed that the essential oils isolated from the three chemotypes were mainly composed of monoterpenes, including 20 monoterpenoids, basically complying with the study results in *C. camphora* [13]; however, there were significant differences in the varieties and contents of monoterpenoids among the three chemical types. The content of linalool in the linalool chemotype was 77.58%–87.88%, but the linalool content was very low in the other two chemotypes. The content of

eucalyptol in the eucalyptol chemotype was 34.48–35.63%, but the eucalyptol content was extremely low in the other two chemotypes. Notably, the monoterpenoid content of the camphor chemotype was almost pure camphor, but camphor was absent or present in a small amount in the other two chemotypes. The difference in the contents of the three main monoterpenoids (linalool, eucalyptol and camphor) in the three chemotypes of *C. porrectum* greatly coincided with the differential expression of monoterpenene synthase genes, such as *CpTPS1*, *CpTPS3*, *CpTPS4*, *CpTPS5* and *CpTPS9*, among the three different chemotypes. Presumably, the *CpTPS4* gene is highly likely to affect the synthesis of linalool in *C. porrectum* leaf tissue, the *CpTPS3*, *CpTPS5* and *CpTPS9* genes affect the synthesis of eucalyptol, and the *CpTPS1* gene is related to the synthesis of camphor. Compared with the published proteins, *CpTPS4* had the highest similarity with S-+-linalool synthase of *Cinnamomum microratum* (94.44%, Accession RWR97874.1), and *CpTPS1* had the highest similarity with geraniol synthase of *Cinnamomum microratum* (96.55%, Accession RWR98053.1). *CpTPS3*, *CpTPS5* and *CpTPS9* had the highest similarity with alpha thujene synthase/sabinene synthase of *Litsea cubeba* (Accession AEJ91556.1), and the similarities were 90%, 90.21% and 90.69% respectively. Sequence similarity alone can not predict the specific biochemical function of a single TPS family member, because only a few amino acids can cause dramatic changes in the terpenoid structure of a given TPS enzyme [49,50]. In addition, many TPS are prolific enzymes, which can generally produce mixtures of different proportions of the same compounds [51]. In the study of amino acid sequence similarity of limonene synthases in different species, it was also displayed that sequence similarity was more closely related to genetic relationship than function [24,25]. Therefore, it is difficult to judge the specific functions of genes from the similarity of sequences. The specific functions of these five genes need to be further verified. The future functional verification and utilization of these monoterpenoids should focus on these genes.

4. Materials and Methods

4.1. Plant Materials

Three generations of leaf tissue samples for full-length transcriptome sequencing were collected from the Guangdong Baixi Nature Reserve in southern China in June 2017 (23.7129°N, 115.1971°E, Guangdong, China). Five leaf samples of *C. porrectum* were gathered randomly, with an interval of more than 300 m between sample plants, the sampled trees were approximately 20 years old. In the middle and upper part of the tree crown, a whole leaf was taken from the East, West, North and south respectively for RNA extraction. RNA was extracted separately, and the five samples of RNA were mixed in equal amounts for transcriptome sequencing. The leaf samples with linalool, eucalyptol or camphor as the first principal components of essential oil were collected separately and quickly frozen in liquid nitrogen. Leaf samples were placed in a refrigerator at –80 °C for the sequencing of different chemotypes of leaf tissue, and there were three biological replicates for each chemotype. At the same time, sampling was performed separately to determine the content of the essential oil and the chemical composition of the essential oil.

4.2. Determination of Chemical Composition

Isolation of essential oil: Samples of fresh *C. porrectum* leaves (about 200 g) were each weighed immediately after being picked and were kept under hermetic conditions at a low temperature. The volatile essential oil was isolated by steam distillation with a distillation time of 2 h. Isolation of essential oil from each sample was repeated three times. After the essential oil was weighed, it was stored under hermetic conditions at low temperature.

Gas chromatography-mass spectrometry: GC/MS: GC/MS analysis was performed using a Shimadzu QP2020 GC/MS instrument (chromatographic column: SH-RXI-5SILMS, 30 m × 0.25 mm × 0.25 μm). GC/MS procedure: kept at 80 °C for 2 min, elevated to 160 °C at 8 °C/min, and then raised to 250 °C at 8 °C/min, maintained for 2 min. The injection volume was 1.0 μL, and a split ratio of 20:1 was used. The temperature of the injection port was 280 °C, the EI ion source was 230 °C, and the connection line

was 200 °C. The MS scan range (*m/z*) was 50–650. For analysis, essential oils were dissolved in alcohol (30 mg/mL) and directly injected.

Component identification: The components were identified by gas chromatography. The retention index of n-alkanes (c9-c33) was determined under the same operating conditions by comparison with the retention index of literature or the national institute of standards and technology (NIST) 8.0 standard. The retention index of n-alkanes (c9-c33) was further determined by comparison with the retention index of literature or NIST 8.0 standard. Their mass spectra of two columns were compared with those stored in NIST 8.0 libraries or with mass spectra from the available literature [38–40] to further identify the components.

Quantitative analysis: Calibration curve of each of the standard compounds, β-pinene, eucalyptol, 2-Dodecanone and eugenol was generated using six concentration points. The response factor of the respective standard and the percentage of each chemical constituent in the essential oil were determined by following the reported procedure [52,53].

4.3. RNA Extraction and Library Construction

RNA was extracted using the Huayueyang polysaccharide and polyphenol plant RNA extraction kit. The total RNA concentration, RIN value, 28S/18S and fragment size were detected using the Agilent 2100 Bioanalyzer, and RNA purity detection was conducted with NanoDropTM. Total RNA was first synthesized into first-strand cDNA using the Clontech SMARTer PCR cDNA Synthesis Kit, first-strand cDNA was synthesized into second-strand cDNA through PCR amplification, and then double-stranded DNA was used for SMRTbell library construction after secondary PCR amplification. One library > 4K for fragment selection and one nonfragmented library were constructed with BluePippinTM. The two libraries were mixed for sequencing.

4.4. Transcriptome Sequencing and Assembly

Three-generation sequencing of the *C. porrectum* transcriptome library was performed using the PacBio Sequel platform. According to the offline subread data, each inserted fragment was identified and integrated to generate the consensus form of the inserted fragment sequence. The identified inserted fragment sequence was subject to dejointed poly-A sequencing, and then classified as a full-length, non-full-length, chimeric, or nonchimeric sequence. With the ICE algorithm, de novo prediction of isoforms was conducted. The Quiver algorithm was used for isoform correction, and high-QV and low-QV isoforms were output. Redundancy was eliminated from the clustered and error-corrected transcripts using cd-hit to obtain a high-quality full-length unigene library.

4.5. Determination of the Expression Profiles of Three Chemotypes

RNA was sequenced on an Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) with three biological replicates in this study. Sequencing libraries were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Boston, MA, USA) following the manufacturer's recommendations. The number of fragments per kilobase of exon per million fragments mapped (FPKM) of unigenes were calculated using the RSEM (v1.2.15) software to normalize the relative gene expression levels [54,55].

4.6. Transcriptome Annotation and Differential Expression Analysis

Leaf transcriptome functional annotation was performed using BLAST program with an E-value cut-off of 1×10^{-5} against the following databases: NR, InterPro, SwissProt, COG and KEGG. Gene ontology (GO) was analyzed by Blast2GO version 2.5 [56]. Read counts of different chemotypes of *C. porrectum* samples were imported to the DESeq R package (v1.10.1) for differential analysis [57]. The differential expressed transcripts with false discovery rate (FDR) < 0.05 were exported for downstream analysis.

4.7. Construction of Phylogenetic Tree of TPS

Based on the results of the functional annotation, the potential sassafras TPS protein sequence in the transcriptome was obtained and manually corrected for redundancy. First, in the phylogenetic analysis of sassafras TPS, MUSCLE was used to conduct protein sequence alignment to obtain the conserved domain sequence, and then a phylogenetic tree was constructed using MEGA7 with the neighbor-joining (NJ) method and 1000 bootstrap replicates [58].

4.8. qRT-PCR Validation of DEGs

qRT-PCR was employed using the SYBR Green PCR Master Mix (Takara, Beijing, China) on the Roche-LightCycler-480. The primers were designed using Oligo 7.0 (Molecular Biology Insights, Cascade, CO, USA). *CpTPS1*, *CpTPS2*, *CpTPS3*, *CpTPS4*, *CpTPS5*, *CpTPS6*, *CpTPS7*, *CpTPS8*, *CpTPS9*, *chloroplastic-like*, *GGDR-like1* and *BDA1-like1* genes were performed. Details of the primers used for the qRT-qPCR assay are listed in Table S5, and sequences of these unigenes are listed in Table S6. The cDNA was synthesized using the PrimeScriptTM RT Reagent Kit (Perfect Real Time) reverse transcription kit, and the actin (*ACT*) gene was used as the internal reference gene. According to the quantitative reagent SYBR® Premix Ex TaqTM II (Tli RNaseH Plus) instructions, all samples were subjected to three biological replications. The Ct values for all genes were normalized to the Ct value of actin [59].

5. Conclusions

In this study, the transcriptome and metabolism of leaf tissue from the linalool, eucalyptol and camphor chemotypes of *C. porrectum* were analyzed. These data describe the gene expression levels during terpene biosynthesis in *C. porrectum*. Monoterpens were identified as the major component of the three chemotypes. The full-length transcriptome had a total of 104,062 transcripts, including 89,103 complete transcripts, and was annotated, laying the foundation for further functional studies. KEGG enrichment analysis found that 1495 unigenes were enriched in the metabolic pathways of terpenoids and polyketides. In the eucalyptol and linalool chemotype, the camphor and eucalyptol chemotype, and the camphor and linalool chemotype comparison groups, 52, 49 and 66 terpene biosynthesis DEGs were identified. GO and KEGG enrichment analysis indicated that TPS activity and oxidoreductase activity could explain the differential accumulation of terpenoids among the three chemotypes of *C. porrectum*, but further functional studies are required to elucidate the regulatory mechanism of the formation of terpenoids. In conclusion, the full-length transcriptome sequence and the gene expression profiles of this study provide valuable information for understanding the accumulation of terpenoids in different chemotypes of *C. porrectum*, lay a foundation for research on the function of key genes in the synthesis of terpenoids, and offer clues for the study of the terpenoid formation mechanism in *C. porrectum*.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/24/6230/s1>.

Author Contributions: X.S. and F.Q. conceived and designed the experiments; F.Q. and H.W. performed the experiments; X.W., Y.Z. and X.L. analyzed the data; F.Q. wrote the paper.

Funding: This work was supported by the Technology Program of Jiangxi Academy of Forestry (2018512701); The National Key R&D Program (2016YFD0600605); Key R&D Projects of Jiangxi Province (20161ACF60021).

Acknowledgments: We thanks to Tian Jing (Wuhan Botanical Garden, Chinese Academy of Sciences) for plant species identification.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

bp	Base pairs (measuring unit)
CDS	Coding sequence
DEGs	Differentially expressed genes
DESeq	Differentially expressed sequence
DMAPP	Dimethylallyl diphosphate
FDR	False discovery rate
FPKM	Fragments per kilobase of exon per million fragments mapped
FPPS	Farnesyl diphosphate synthase
GC-MS	Gas chromatography-mass spectrometry
GGPS	Geranylgeranyl diphosphate synthase
GO	Gene ontology
GPS	Geranyl diphosphate synthase
IPP	Isopentenyl diphosphate
KEGG	Kyoto encyclopedia of genes and genomes
KO	KEGG ortholog database
KOG	EuKaryotic Orthologous Groups
MEP	2-C-methyl-D-erythritol 4phosphate
MVA	Mevalonate acid
NIST	National institute of standards and technology
Nr	NCBI non-redundant protein sequences
Nt	NCBI non-redundant nucleotide sequences
Pfam	Protein family
qRT-PCR	Quantitative reverse transcription PCR
ROI	Reads of insert
RNA-Seq	RNA sequencing
SMRT	Single molecule real-time
TP	Terpene synthases

References

- Li, S.; Li, X.; Li, J.; Huang, P.; Wei, F.; Cui, H.; Werff, H. *Flora of China*; Science Press and St. Louis: Missouri Botanical Garden Press: Beijing, China, 2008; Volume 7, p. 102.
- Wu, H.; Wang, J.J.; Liu, C.; Zhu, L.F. A study on the chemotypes of *Cinnamomum parthenoxylon*(Jack) Nees. *J. Plant Resour. Environ.* **1992**, *1*, 45–49. (In Chinese)
- Kawamura, F.; Ramle, S.F.M.; Sulaiman, O.; Hashim, R.; Ohara, S. Antioxidant and antifungal activities of extracts from 15 selected hardwood species of Malaysian timber. *Eur. J. Wood Prod.* **2010**, *69*, 207–212. [[CrossRef](#)]
- Buru, A.S.; Pichika, M.R.; Neela, V.; Mohandas, K. In vitro antibacterial effects of *Cinnamomum* extracts on common bacteria found in wound infections with emphasis on methicillin-resistant *Staphylococcus aureus*. *J. Ethnopharmacol.* **2014**, *153*, 587–595. [[CrossRef](#)] [[PubMed](#)]
- Phongpaichit, S.; Kum mee, S.; Nilrat, L.; Itarat, A. Antimicrobial activity of oil from the root of *Cinnamomum porrectum*. *Songklanakarin J. Sci. Technol.* **2006**, *29*, 11–16.
- Palanuvej, C.; Werawatganone, P.; Lipipun, V.; Ruangrungsi, N. Chemical composition and antimicrobial activity against *Candida albicans* of essential oil from leaves of *Cinnamomum porrectum*. *J. Health Res.* **2006**, *20*, 69–76.
- Subki, S.Y.M.; Jamal, J.A.; Husain, K.; Manshoor, N. Characterisation of leaf essential oils of three *Cinnamomum* species from Malaysia by gas chromatography and multivariate data analysis. *Pharmacogn. J.* **2013**, *5*, 22–29. [[CrossRef](#)]
- Batista, P.A.; Werner, M.F.; Oliveira, E.C.; Burgos, L.; Pereira, P.; Brum, L.F.; Santos, A.R. Evidence for the involvement of ionotropic glutamatergic receptors on the antinociceptive effect of (-)-linalool in mice. *Neurosci. Lett.* **2008**, *440*, 299–303. [[CrossRef](#)]

9. Huo, M.X.; Cui, X.R.; Xue, J.D.; Chi, G.F.; Gao, R.J.; Deng, X.M.; Guan, S.; Wei, J.Y.; Soromou, L.W.; Feng, H.H.; et al. Anti-inflammatory effects of linalool in RAW 264.7 macrophages and lipopolysaccharide-induced lung injury model. *J. Surg. Res.* **2013**, *180*, e47–e54. [[CrossRef](#)]
10. Miyashita, M.; Sadzuka, Y. Effect of linalool as a component of *Humulus lupulus* on doxorubicin-induced antitumor activity. *Food Chem. Toxicol.* **2013**, *53*, 174–179. [[CrossRef](#)]
11. Kladniew, R.B.; Polo, M.; Villegas, S.M.; Galle, M.; Crespo, R.; Garcia de Bravo, M. Synergistic antiproliferative and anticholesterogenic effects of linalool, 1,8-cineole, and simvastatin on human cell lines. *Chem. Biol. Interact.* **2014**, *214*, 57–68. [[CrossRef](#)]
12. Shi, C.; Sun, Y.; Liu, Z.Y.; Guo, D.; Sun, H.H.; Sun, Z.; Chen, S.; Zhang, W.T.; Wen, Q.W.; Peng, X.L.; et al. Inhibition of *cronobacter sakazakii* virulence factors by citral. *Sci. Rep.* **2017**, *7*, 43243. [[CrossRef](#)] [[PubMed](#)]
13. Chen, C.H.; Zheng, Y.J.; Zhong, Y.D.; Wu, Y.F.; Li, Z.T.; Xu, L.A.; Xu, M. Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*. *BMC Genom.* **2018**, *19*, 550. [[CrossRef](#)] [[PubMed](#)]
14. Tholl, D. Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* **2015**, *148*, 63. [[PubMed](#)]
15. Xu, Y.; Zhu, C.; Xu, C.; Sun, J.; Grierson, D. Integration of metabolite profiling and transcriptome analysis reveals genes related to volatile terpenoid metabolism in finger citron (*C. medica* var. *sarcodactylis*). *Molecules* **2019**, *24*, 2564. [[CrossRef](#)] [[PubMed](#)]
16. Cordoba, E.; Salmi, M.; Leon, P. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* **2009**, *60*, 2933–2943. [[CrossRef](#)] [[PubMed](#)]
17. Lichtenthaler; Hartmut, K. The 1-Deoxy-D-Xylulose-5-Phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Biol.* **1999**, *50*, 47–65. [[CrossRef](#)] [[PubMed](#)]
18. Sapir-Mir, M.; Mett, A.; Belausov, E.; Tal-Meshulam, S.; Frydman, A.; Gidoni, D.; Eyal, Y. Peroxisomal localization of *arabidopsis* isopentenyl diphosphate isomerase suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. *Plant Physiol.* **2008**, *148*, 1219–1228. [[CrossRef](#)]
19. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **1999**, *16*, 565–574. [[CrossRef](#)]
20. Kazunori, O.; Hiroyuki, K.; Shinjiro, Y.; Hiroshi, K.; Yuji, K.; Hideaki, N.; Hisakazu, Y. Genetic evidence for the role of isopentenyl diphosphate isomerase in the mevalonate pathway and plant development in *Arabidopsis*. *Plant Cell Physiol.* **2008**, *49*, 604.
21. Moses, T.; Pollier, J.; Thevelein, J.M.; Goossens, A. Bioengineering of plant (tri)terpenoids: From metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. *New Phytol.* **2013**, *200*, 27–43. [[CrossRef](#)]
22. Wise, M.L.; Savage, T.J.; Katahira, E.; Croteau, R. Monoterpene Synthases from Common Sage (*Salvia officinalis*). *J. Biol. Chem.* **1998**, *273*, 14891–14899. [[CrossRef](#)] [[PubMed](#)]
23. Bohlmann, J.; Phillips, M.; Ramachandiran, V.; Katoh, S.; Croteau, R. cDNA Cloning, Characterization, and Functional Expression of Four New Monoterpene Synthase Members of the Tpsd Gene Family from Grand Fir (*Abies grandis*). *Arch. Biochem. Biophys.* **1999**, *368*, 232–243. [[CrossRef](#)] [[PubMed](#)]
24. Colby, S.M.; Alonso, W.R.; Katahira, E.J.; McGarvey, D.J.; Croteau, R. 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. *J. Biol. Chem.* **1993**, *268*, 23016–23024. [[PubMed](#)]
25. Williams, D.C.; McGarvey, D.J.; Katahira, E.J.; Croteau, R. Truncation of limonene synthase preprotein provides a fully active ‘Pseudomature’ form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* **1998**, *37*, 12213–12220. [[CrossRef](#)]
26. Lin, Y.L.; Lee, Y.R.; Huang, W.K.; Chang, S.T.; Chu, F.H. Characterization of S-(+)-linalool synthase from several provenances of *Cinnamomum osmophloeum*. *Tree Genet. Genomes* **2014**, *10*, 75–86. [[CrossRef](#)]
27. Despinasse, Y.; Fiorucci, S.; Antonczak, S.; Moja, S.; Bony, A.; Nicole, F.; Baudino, S.; Maqnard, J.L.; Jullien, F. Bornyl-diphosphate synthase from *Lavandula angustifolia*: A major monoterpene synthase involved in essential oil quality. *Phytochemistry* **2017**, *137*, 24–33. [[CrossRef](#)]
28. Yahyaa, M.; Matsuba, Y.; Brandt, W.; Doron-Faigenboim, A.; Bar, E.; McClain, A.; Davidovich-Rikanati, R.; Lewinsohn, E.; Pichersky, E.; Ibdah, M.; et al. Identification, Functional Characterization, and Evolution of Terpene Synthases from a Basal Dicot. *Plant Physiol.* **2015**, *169*, 1683–1697. [[CrossRef](#)]
29. Yang, T.; Li, J.; Wang, H.X.; Zeng, Y.A. Geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* **2005**, *66*, 285–293. [[CrossRef](#)]

30. Hittinger, C.T.; Johnston, M.; Tossberg, J.T.; Rokas, A. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1476–1481. [[CrossRef](#)]
31. Morozova, O.; Hirst, M.; Marra, M.A. Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 135–151. [[CrossRef](#)]
32. Rhoads, A.; Au, K.F. PacBio sequencing and its applications. *Genom. Proteom. Bioinf.* **2015**, *13*, 278–289. [[CrossRef](#)]
33. Abdel-Ghany, S.E.; Hamilton, M.; Jacobi, J.L.; Ngam, P.; Devitt, N.; Schilkey, F.; Ben-Hur, A.; Reddy, A.S. A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **2016**, *7*, 11706. [[CrossRef](#)]
34. Wang, B.; Tseng, E.; Regulski, M.; Clark, T.A.; Hon, T.; Jiao, Y.; Lu, Z.; Olson, A.; Stein, J.C.; Ware, D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **2016**, *24*, 11708. [[CrossRef](#)]
35. Wang, M.; Wang, P.; Liang, F.; Ye, Z.; Li, J.; Shen, C.; Pei, L.; Wang, F.; Hu, J.; Tu, L.; et al. A global survey of alternative splicing in allopolyploid cotton: Landscape, complexity and regulation. *New Phytol.* **2018**, *217*, 163–178. [[CrossRef](#)]
36. Wang, X.; Chen, S.; Shi, X.; Liu, D.; Zhao, P.; Lu, Y.; Cheng, Y.; Liu, Z.; Nie, X.; Song, W. Hybrid sequencing reveals insight into heat sensing and signaling of bread wheat. *Plant J.* **2019**, *98*, 10151032. [[CrossRef](#)]
37. Adams, R.P. *Identification of Essential Oil Components by Gas Chromatography/Mass Spectroscopy*, 4th ed.; Allured Publishing Corporation: Carol Stream, IL, USA, 2007; pp. 1–804.
38. Bordiga, M.; Rinaldi, M.; Locatelli, M.; Piana, G.; Travaglia, F.; Coïsson, J.D.; Arlorio, M. Characterization of Muscat wines aroma evolution using comprehensive gas chromatography followed by a post-analytic approach to 2D contour plots comparison. *Food Chem.* **2013**, *140*, 57–67. [[CrossRef](#)]
39. Ribeiro, V.L.S.; Rolim, V.; Bordignon, S.; Henriques, A.T.; Dorneles, G.G.; Limberger, R.P.; Poster, G.V. Chemical composition and larvicidal properties of the essential oils from *Drimys brasiliensis* Miers (Winteraceae) on the cattle tick *Rhipicephalus* (*Boophilus*) microplus and the brown dog tick *Rhipicephalus sanguineus*. *Parasitol. Res.* **2008**, *102*, 531–535. [[CrossRef](#)]
40. Andrade, E.H.A.; Zoghbi, M.D.G.B.; Lima, M.D.P.; Botânica, C.D.B.; Goeldi, M.P.E. Chemical composition of the essential oils of *Cymbopogon citratus* (DC.) Stapf cultivated in north of Brazil. *J. Essent. Oil Bear. Plants* **2009**, *12*, 41–45. [[CrossRef](#)]
41. Darbani, B.; Stewart, C.N., Jr.; Noeparvar, S.; Borg, S. Correction of gene expression data: Performance-dependency on inter-replicate and inter-treatment biases. *J. Biotechnol.* **2014**, *188*, 100–109. [[CrossRef](#)]
42. He, Y.; Deng, C.; Xiong, L.; Qin, S.S.; Peng, C. Transcriptome sequencing provides insights into the metabolic pathways of patchouli alcohol and pogostone in *Pogostemon cablin* (Blanco) Benth. *Genes Genom.* **2016**, *38*, 1031–1039. [[CrossRef](#)]
43. Tilgner, H.; Jahanbani, F.; Blauwkamp, T.; Moshrefi, A.; Jaeger, E.; Chen, F.; Harel, I.; Bustamante, C.D.; Rasmussen, M.; Snyder, M.P. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* **2015**, *33*, 736–742. [[CrossRef](#)]
44. Chang, Y.T.; Chu, F.H. Molecular cloning and characterization of monoterpane synthases from *Litsea cubeba* (Lour.) Persoon. *Tree Genet. Genomes* **2011**, *7*, 835–844. [[CrossRef](#)]
45. Vranová, E.; Coman, D.; Gruissem, W. Network Analysis of the MVA and MEP Pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* **2013**, *64*, 665–700. [[CrossRef](#)] [[PubMed](#)]
46. Friesen, J.A.; Rodwell, V.W. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol.* **2004**, *5*, 248. [[CrossRef](#)] [[PubMed](#)]
47. Rohdich, F.; Lauw, S.; Kaiser, J.; Feicht, R.P.; Bacher, A.; Eisenreich, W. Isoprenoid biosynthesis in plants-2 C-methyl-d-erythritol-4-phosphate synthase (IspC protein) of *Arabidopsis thaliana*. *FEBS J.* **2006**, *273*, 4446. [[CrossRef](#)] [[PubMed](#)]
48. Cheng, S.S.; Lin, C.Y.; Yang, C.K.; Chen, Y.J.; Chung, M.J.; Chang, S.T. Chemical polymorphism and composition of leaf essential oils of *Cinnamomum kanehirae* using Gas Chromatography/Mass Spectrometry, cluster analysis, and principal component analysis. *J. Wood Chem. Technol.* **2015**, *35*, 207–219. [[CrossRef](#)]
49. Zulak, K.G.; Lippert, D.N.; Kuzyk, M.A.; Domanski, D.; Chou, T.; Borchers, C.H.; Bohlmann, J. Targeted proteomics using selected reaction monitoring reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate-treated Norway spruce (*Picea abies*). *Plant J.* **2009**, *60*, 1015–1030. [[CrossRef](#)]

50. Dudareva, N. (E)-beta-Ocimene and Myrcene synthase genes of floral scent biosynthesis in snapdragon: Function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell.* **2003**, *15*, 1227–1241. [[CrossRef](#)]
51. Martin, D.M.; Faldt, J.; Bohlmann, J. Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.* **2004**, *135*, 1908–1927. [[CrossRef](#)]
52. Naik, D.G.; Dandge, C.N.; Rupanar, S.V. Chemical examination and evaluation of antioxidant and antimicrobial activities of essential oil from *Gymnema sylvestre* R. Br. Leaves. *J. Essent. Oil Res.* **2011**, *23*, 12–19. [[CrossRef](#)]
53. Zheljazkov, V.D.; Cantrell, C.L.; Tekwani, B.; Khan, S.I. Content, composition, and bioactivity of the essential oils of three basil genotypes as a function of harvesting. *J. Agric. Food Chem.* **2008**, *56*, 380–385. [[CrossRef](#)] [[PubMed](#)]
54. Mortazavi, A.; Williams, B.A.; Mccue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621. [[CrossRef](#)] [[PubMed](#)]
55. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
56. Götz, S.; Garcíagómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, SH.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420. [[CrossRef](#)] [[PubMed](#)]
57. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)]
58. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
59. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, e45. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).