BMC Genomics

RESEARCH ARTICLE

Open Access

Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*

Caihui Chen[1†], Yongjie Zheng[2†], Yongda Zhong[3], Yangfang Wu[2], Zhiting Li[1], Li-An Xu[1] and Meng Xu[1*]

Abstract

Background: *Cinnamomum camphora* has been cultivated as an economically important tree for its medicinal and aromatic properties. Selective breeding has produced *Cinnamomum* plants for special uses, including spice strains with characteristic flavors and aromas and high-potency medicinal cultivars. The molecular biology underlying terpenoid biosynthesis is still unexplored.

Results: Gas chromatography-mass spectrometry was used to analyze the differences in contents and compositions of essential oil terpenoids in linalool- and borneol-type chemotypes of *C. camphora*. The data revealed that the essential oils consist primarily of monoterpenes with only very minor quantities of sesquiterpenes and diterpenes and that the essential oil differs in different chemotypes of *C. camphora*, with higher yields of (−)-borneol from the borneol-type than from the linalool-type. To study the terpenoid biosynthesis of signature compounds of the major monoterpenes, we performed RNA sequencing to profile the leaf transcriptomes of the two chemotypes of *C. camphora*. A total of 23.76 Gb clean data was generated from two chemotypes and assembled into 156,184 unigenes. The total length, average length, N50 and GC content of unigenes were 155,645,929 bp, 997 bp, 1430 bp, and 46.5%, respectively. Among them, 76,421 unigenes were annotated by publicly available databases, of which 67 candidate unigenes were identified to be involved in terpenoid biosynthesis in *C. camphora*. A total of 2863 unigenes were identified to be differentially expression between borneol-type and linalool-type, including 1714 up-regulated and 1149 down-regulated unigenes. Most genes encoding proteins involved in terpenoid precursor MVA and MEP pathways were expressed in similar levels in both chemotypes of *C. camphora*. In addition, 10 and 17 DEGs were significantly enriched in the terpene synthase activity and oxidoreductase activity terms of their directed acyclic graphs (DAG), respectively. Three monoterpene synthase genes, *TPS14-like1*, *TPS14-like2* and *TPS14-like3* were up-regulated in the borneol-type compared to the linalool-type, and their expression levels were further verified using quantitative real-time PCR.

Conclusions: This study provides a global overview of gene expression patterns related to terpenoid biosynthesis in *C. camphora*, and could contribute to a better understanding of the differential accumulation of terpenoids in different *C. camphora* chemotypes.

Keywords: Terpenoids, Terpenoid biosynthesis, *Cinnamomum camphora*, Transcriptome, Different chemotypes

* Correspondence: xum@njfu.edu.cn
†Caihui Chen and Yongjie Zheng contributed equally to this work.
[1]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
Full list of author information is available at the end of the article

Chen et al. BMC Genomics (2018) 19:550

Page 2 of 15

Background

Cinnamomum camphora, a member of the Lauraceae family, is an evergreen broad-leaf tree indigenous to southern China and Japan [1]. The essential oil or crystal distilled from *C. camphora* has considerable economic importance as a source of food preservative and additive, and as raw materials for the cosmetic and pharmaceutical industries. The chemical compositions of different *C. camphora* tissues have been previously investigated, and mono- and sesquiterpenes were found abundant in the leaves and twigs [2, 3]. There is chemical polymorphism in *C. camphora*. According to the signature constituent in leaf, *C. camphora* was classified into at least five different chemical variants (chemotypes), including linalool- (58–92%), borneol- (67–82%), camphor- (54–97%), cineole- (32–52%), and nerolidol-types (16–57%) [4, 5]. In recent years, *C. camphora* has become of increasing importance as a source of essential oils, especially for the production of natural borneol and linalool [6]. Borneol, a bicyclic monoterpene alcohol, has been widely used in food and drug industries typically in folk medicine in China and India. Interestingly, natural borneol has been used to increase the permeabilities of the intestinal mucosa and blood-brain barrier to improve the oral bioavailability of some poorly permeable drugs [7].

Terpenoids (or terpenes) constitute the largest class of structurally diverse metabolites, with more than 55,000 members identified in living organisms [8, 9]. Based on the number of 5-carbon units, terpenoids are classified into monoterpene (C10), sesquiterpenes (C15), diterpenes (C20), triterpenes (C30), tetraterpenes (C40) and polyterpenes which have more than eight 5C units [10]. Borneol and linalool, like all other monoterpenes, are formed from the isomeric 5-carbon building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Two independent pathways participate in the biosynthesis of IPP and DMAPP in plants. 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 [11, 12], whereas in the cytosol and peroxisomes, IPP/DMAPP formation occurs from the condensation of acetyl-CoA in the mevalonate acid (MVA) pathway [13, 14]. The equilibrium of IPP and DMAPP is controlled by IPP Delta-isomerase (IDI), which reversibly converts IPP to DMAPP [15]. The condensation of the C5 precursors leads to the formation of monoterpenes, sesquiterpenes, and diterpenes by the activation of terpene synthases (TPS) [16, 17]. Terpenoid synthesis is often correlated with the induction of TPS gene expression [18–20] and several monoterpene and sesquiterpene synthases have been isolated and characterized from various plant species [21, 22]. Besides the regulation at the level of terpene synthase activity, the induction of precursor biosynthetic genes has also been described [23, 24].

Recently, the development of RNA-sequencing (RNA--Seq) provided an opportunity for detailed transcriptomic studies, even in species without a reference genome [25]. De novo transcriptome assembly has been widely applied to investigate and identify the critical genes involved in the biosynthesis of secondary metabolites in medicinal and aromatic plants, such as *Artemisia annua* [26], *Mentha spicate* [27], *Cinnamomum camphora* [28], *Salvia miltiorrhiza* [29], *Lindera glauca* [30] and *Huperzia serrata* [31].

In light of the chemical polymorphism evident in *C. camphora*, it is of interest to search for potential differences in the terpenoid biosynthesis pathway in different chemotypes. We have used metabolic analysis and transcriptome sequencing in order to discover potential monoterpene synthases present in *C. camphora* chemotypes. In the current study, the leaf contents and compositions of terpenoids in borneol- and linalool-types was analyzed with gas chromatography-mass spectrometry (GC-MS) and their transcriptomes compared. After functional annotation and classification, the genes involved in the MEP/MVA pathways and terpene synthesis in both chemotypes were identified. The results contribute to our understanding of the mechanism underlying differences in terpenoid biosynthesis between *C. camphora* chemotypes.
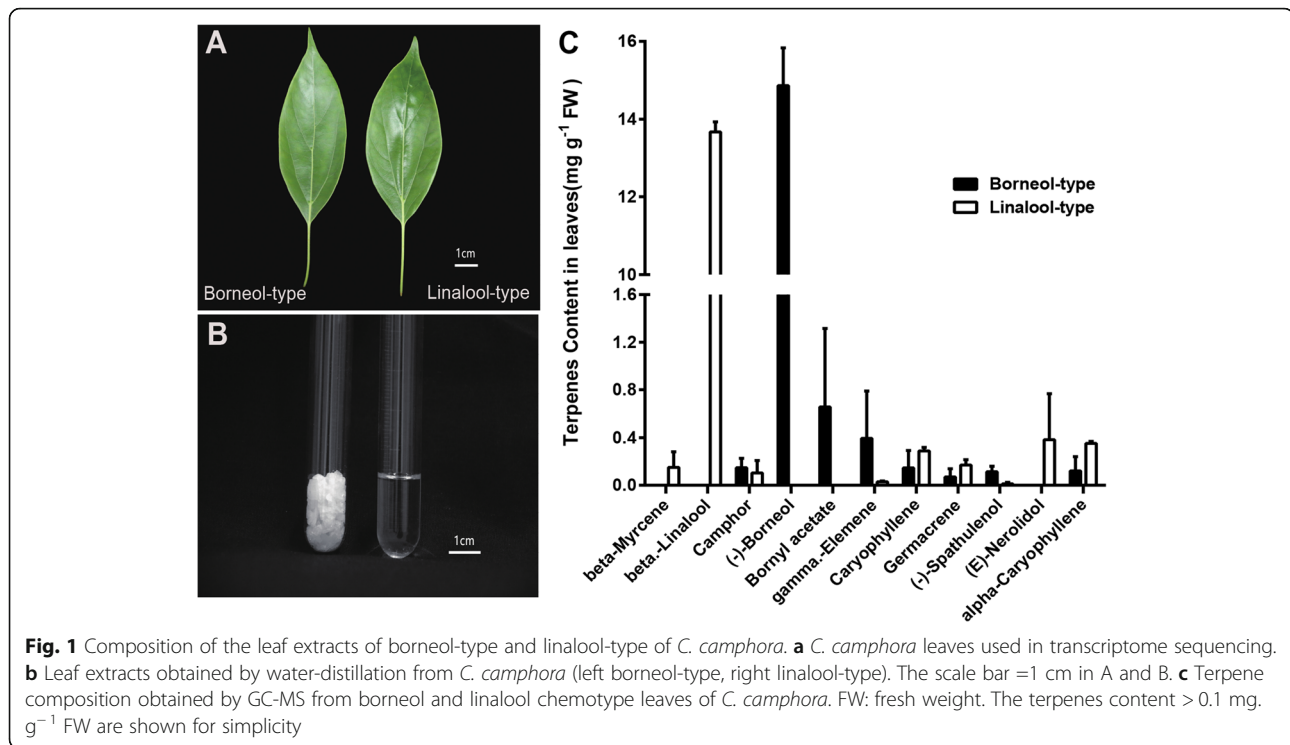
Methods

Plant materials

Leaves of linalool- (No. JF1, JF2) and borneol-type (No. JL1, JL4) *C. camphora* were collected for chemical analysis and RNA sequencing from trees grown at the experimental tree farm of the Jiangxi Academy of Forestry in Nanchang, China (Fig. 1a). Two 8-year-old trees of each chemotype were used as biological replicates. The two replications for the linalool chemotype were termed F_L1 and F_L2, and the borneol chemotype were termed L_L1 and L_L2. Three chemotypes of *C. camphora* (camphor-, cineole- and nerolidol-types) were obtained from the campus of Nanjing Forestry University for further analysis, the compositions of their leaf extracts are listed in Additional file 1. All the samples for RNA extraction were frozen in liquid nitrogen immediately and stored at − 80 °C. At the same time, fresh leaves of different chemotypes were obtained for essential oil isolation.

Measurement of leaf terpenoids

The fresh leaves of each sample were hydrodistilled with a modified Clevenger-type apparatus for 2 h until the extraction was completed. The terpenoid composition analysis of each sample were performed on a SHIMADZU

Chen *et al. BMC Genomics* (2018) 19:550

Page 3 of 15

Fig. 1 Composition of the leaf extracts of borneol-type and linalool-type of *C. camphora*. **a** *C. camphora* leaves used in transcriptome sequencing. **b** Leaf extracts obtained by water-distillation from *C. camphora* (left borneol-type, right linalool-type). The scale bar =1 cm in A and B. **c** Terpene composition obtained by GC-MS from borneol and linalool chemotype leaves of *C. camphora*. FW: fresh weight. The terpenes content > 0.1 mg. g^{-1} FW are shown for simplicity

QP2020 gas chromatograph-mass spectrometer (SHI-MADZU Corporation, Japan), fitted with a DB-5-MS silica capillary column (30 m × 0.32 mm, 0.25 μm film thickness). Helium was used as the carrier gas with a constant flowrate of 0.6 mL/min. The GC temperature program were as follows: 50 °C for 3 min, a gradient of 50–180 °C over 16.25 min followed by holding at 180 °C for 1 min, then a gradient of 180–280 °C over 10 min followed by holding at 280 °C 5 min. Alkanes were used as reference points in the calculation of relative retention indices. The GC inlet was operated at 280 °C in splitless mode with 0.6 μL injection volume. The quadrupole MS operating parameters were set as follows: electron ionization (EI) mode; EI source, 70 eV; transfer line temperature, 250 °C; ion source temperature, 200 °C; emission current, 150 μA; examination voltage, 500 V; 0.2 s for the full scan mode; scan mass range, 29–450 m/z. The peaks were identified by comparing their retention time with that of the known standards, which were determined under the same conditions. A library search was carried out using the Wiley GC-MS Library and the TBAM Library of Essential Oil Constituents.

Total RNA extraction and transcriptome sequencing

Total RNAs from all samples were extracted using the RNeasy Plant Mini Kit (Qiagen, Hilde, Germany), and immediately frozen at − 80 °C until use. RNA degradation and contamination were monitored on 1% agarose gels. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total amount of 1.5 μg of RNA extracted from each leaf sample was used for transcriptome analysis using Illumina's kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Four sequencing libraries were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Boston, Massachusetts, USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. The cDNA library preparations were sequenced on an Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA) and paired-end reads were generated.

De novo assembly and annotation

For the assembly library, raw data (raw reads) in FASTQ format were first processed through CASAVA base-calling. Trimmomatic (version 0.36) [32] were used to trim and crop FASTQ data as well as to remove adapters. The parameters of removing reads containing N or low quality (reads containing more than 50% bases with Q-value ≤20) were default with paired end mode. All the libraries were mixed together to generate an assembly using Trinity (version r20140413pl) [33], and default parameters were used except for the minimum kmer coverage setting to 2. The assembly results were further clustered and filtered to get the non-redundant unigenes by Corset (version 1.07) [34]. Gene function was annotated using BLAST program with an E-value cut-off of 1e^{-5} against the following databases:

Chen *et al. BMC Genomics* (2018) 19:550

Page 4 of 15

NCBI non-redundant protein sequences (Nr, https://blast.ncbi.nlm.nih.gov/), NCBI non-redundant nucleotide sequences (Nt, https://blast.ncbi.nlm.nih.gov/), Protein family (Pfam, http://pfam.sanger.ac.uk/), and EuKaryotic Orthologous Groups (KOG) database (http://www.ncbi.nlm.nih.gov/COG/). Functional annotation by Gene ontology (GO) terms was analyzed by Blast2GO version 2.5 [35] (http://www.geneontology.org). The annotation of gene association with the Kyoto encyclopedia of genes and genomes (KEGG) pathways was performed by KAAS (r140224).

Analysis of the differential expression of unigenes

The number of fragments per kilobase of exon per million fragments mapped (FPKM) was used to estimate the relative expression levels [36] and the FPKM of unigenes were estimated using the software RSEM v1.2.15 [37], with the Trimmed Mean of M-values method. Differential expression analysis of two chemotypes of *C. camphora* samples was performed using read counts with the DESeq R package (v1.10.1) [38]. The resulting *P*-values were adjusted to Q-values to compensate for multiple hypothesis testing [39]. Genes with an adjusted Q-value < 0.05 found by DESeq were assigned as differentially expressed. GO enrichment analysis of differentially expressed genes (DEGs) were performed using the GOseq R package (version 1.10.0) based on the Wallenius non-central hypergeometric distribution, which can adjust for gene length bias in DEGs [40]. KEGG enrichment analysis was performed with KOBAS version 2.0.12 software [41].

Validation of DEGs by qRT-PCR analysis

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was employed using the SYBR Green PCR Master Mix (Takara, Dalian, China) and the ABI ViiA 7 Real-time PCR platform. The primers were aligned to the target gene sequence using Oligo 7.0 (Molecular Biology Insights, Cascade, CO, USA). The qRT-PCR analyses were performed for the following genes: *GPS_like* (Cluster-13,185.62746), *HMGR_like* (Cluster-13,185.61515), *HDS_like1* and *HDS_like2* (Clusters-13,185.59980 and 13,185.80000, respectively), *HDR* (Cluster-13,185.61134), *DXR_like* (Cluster-13,185.63961), *Mg17_like* (Cluster-13,185.50937), *TPS14_like1*, *TPS14_like2* and *TPS14_like3* (Clusters-13,185.98748, 13,185.75128 and 13,185.81558, respectively), *SDR_like1* and *SDR_like2* (short-chain dehydrogenases, Clusters 13,185.106850 and 13,185.105357) and one putative linalool synthase gene were submitted to GenBank (accession no. XJ028228, *C. camphora*). Details of the primers used for the qRT-qPCR assay are listed in Additional file 2, and sequences of these unigenes are listed in Additional file 3. The experiments were performed with three technical replicates that contained 30 ng cDNA, 10 μL of SYBR Green PCR Master Mix, and 600 nM primers and nuclease free water with a final volume of 20 μL per reaction. The Ct values for all genes were normalized to the Ct value of actin (*ACT*, KM086738.1).

Results

The composition of leaf extracts from different chemotypes of *C. camphora*

Specimens of *C. camphora* leaves from two chemotypes (F_L1 and F_L2 were linalool-types, L_L1 and L_L2 were borneol-types) were obtained from the garden of Jiangxi Academy of Forestry and the leaf extracts were isolated by water distillation. There were no morphological differences in between the leaves from the two chemotypes of *C. camphora* (Fig. 1a). Interestingly, the leaf essential oil extracts of the linalool-type remained as oil liquids, whereas that of the borneol-type formed crystals (Fig. 1b). The compositions of leaf essential oil from the fresh leaves were analyzed with GC-MS. A total of 24 terpenoid compounds in leaf distillates were detected, and monoterpenes were the main constituents (Table 1). The major components with yields > 0.1 mg.g^{-1} fresh weight are shown in Fig. 1c. In the extracts of both chemotypes, the signature compound was present in far higher content than all other compounds. In the linalool-type, the most abundant compound, beta-linalool, yielded 13.67 mg.g^{-1} FW, followed by (E)-nerolidol in yields of 0.38 mg g^{-1} FW. The principal constituent of the borneol-type was (–)-borneol (14.86 mg g^{-1} FW), followed by borneol acetate (0.65 mg g^{-1} FW). To investigate the biosynthesis of major compounds, we proceeded with the analysis of the transcriptomes of the two *C. camphora* chemotypes.

RNA sequencing and transcriptomic assembly

To identify genes involved in terpenoid biosynthesis in *C. camphora*, four RNA libraries (F_L1 and F_L2 were linalool-types, L_L1 and L_L2 were borneol-types) were prepared and analyzed on Illumina Hiseq2000 platform with a pair-end length of 150 bp. A total of 249.6 million reads were generated from four libraries. After filtering out low quality sequences, approximately 237.6 million clean reads with 46.5% GC content were obtained. The throughput and quality of the RNA-Seq data are included in Table 2. All the clean reads obtained from two different chemotypes were assembled by the Trinity method, resulting in 156,184 unigenes with an N50 length of 1430 bp. The total number of assembled unigenes might be overestimated due to the absence of a reference genome. Additionally, a total of 179,016,590 reads (75.34% of all clean reads) were perfectly mapped (mismatch = 0) to the reference transcriptome by RSEM, which showed that the quality of these mapped genes was sufficient to conduct the subsequent analysis (Table 3).

Chen *et al. BMC Genomics* (2018) 19:550

Page 5 of 15

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

RI[a]	Component	L_L1 (%)	L_L2 (%)	F_L1 (%)	F_L2 (%)
902	alpha.-Thujene	–	–	–	0.13
964	beta.-Phellandrene	–	–	–	0.11
958	beta-Myrcene	–	–	0.12	1.8
976	beta-Ocimene	–	–	1.17	1.57
1059	Eucalyptol	tr	–	–	0.13
1064	7-Octen-2-ol	–	tr	–	–
1082	beta.-Linalool	–	–	83.44	85.97
1121	Camphor	1.4	0.39	–	1.33
1138	(−)-Borneol	97.73	79.81	–	–
1137	Terpinen-4-ol	–	0.33	tr	0.14
1143	alpha-Terpineol	–	0.88	0.14	0.14
1164	trans-Linalool oxide	–	–	0.18	0.15
1277	Bornyl acetate	–	7.56	–	–
1398	1-methylethenyl	–	0.1	–	tr
1431	gamma.-Elemene	–	4.54	0.13	0.23
1469	Naphthalene	–	–	0.87	0.85
1494	Caryophyllene	–	1.68	1.55	2.04
1507	Caryophyllene oxide	–	0.22	0.15	0.11
1515	Germacrene	–	0.8	0.74	1.38
1522	Cyclohexanemethanol	0.1	–	–	–
1536	(−)-Spathulenol	0.42	0.92	0.15	
1564	(E)-Nerolidol	–	tr	4.71	0.13
1579	alpha-Caryophyllene	–	1.38	2.21	2.14
1593	beta-Eudesmol	0.13	tr	–	–
	total %	99.78	98.61	95.56	98.35
	Oil yield %	1.62	1.74	1.67	1.56

[a]RI, Retention indices calculated against n-alkanes(C-C). tr, trace (< 0.1%). F_L1 and F_L2 were linalool-types, L_L1 and L_L2 were borneol-types

Gene annotation and functional classification

The BLAST alignment was utilized to annotate the 156,184 unigenes of *C. camphora* with an E-value threshold of $1e^{-5}$ in the public databases: Nr, Nt, Pfam, GO and KOG. The database annotation results are summarized in Fig. 2a. In summary, 76,070 (48.70%) unigenes were successfully annotated in at least one database, and 17,306 (11.08%) unigenes shared annotation in all databases. There were 64,524 unigenes with significant matches in the Nr database, accounting for the highest proportion (41.31%), while the lowest proportion (26,778; 17.14%) was obtained from the KOG database.

To identify the active biological pathways in *C. camphora*, the assembled unigenes were mapped to the KEGG annotation system (Fig. 2b). A total of 25,512 unigenes were matched in 19 metabolic pathways and 'translation' (2562), 'carbohydrate metabolism (2,215) and 'folding, sorting and degradation' (1,184) were the three richest pathways. In addition, 490 unigenes were mapped to the 'Metabolism of terpenoids and polyketides', including 'Terpenoid backbone biosynthesis' (ko00900, 147), 'Monoterpenoid biosynthesis' (ko00902, 12), 'Sesquiterpenoid and triterpenoid biosynthesis' (ko00909, 44), 'Diterpenoid biosynthesis' (ko00904, 58), 'Limonene and pinene degradation' (ko00903, 88), 'Carotenoid biosynthesis' (ko00906, 151), 'Brassinosteroid biosynthesis' (ko00905, 23) and 'Zeatin biosynthesis' (ko00908, 98).

Based on sequence homology, 51,768 annotated unigenes were categorized into three ontologies with 56 GO terms (Fig. 2c). Within the category of biological process (BP) category, genes matched to 25 GO terms, the most highly represented of which were 'cellular process' (28,998), 'metabolic process' (27,125) and 'single-organism process' (20,998). For the molecular function (MF) category, 'binding' (27,919) and 'catalytic activity' (22,472) were the two most abundant of 21 GO terms. The largest associated term within the 10 GO terms of the cellular component (CC) category was 'cell' (15,526).

The identification of relative DEGs in *C. camphora* chemotypes and enrichment analysis of transcripts

To fully explore potential differential gene expression between borneol and linalool chemotypes of *C. camphora*, the clean reads were mapped to the unigene database. The normalization of gene expression data was performed using multiple correction methods [42], and the differentially-expressed genes between the borneol-type and linalool-type were characterized by DESeq with Q-value < 0.05 and |log2.Fold_change| > 1, resulting in a total of 2863 unigenes identified, consisting of 1714 up-regulated and 1149 down-regulated genes in the

Table 2 Summary of RNA-seq data from four RNA libraries of linalool and borneol chemotypes of *C. camphora*

Libraries	Raw Reads	Clean Reads	Clean Bases	Q30[a] (%)	GC (%)	Mapped reads
F_L1	66,745,696	63,619,460	9.54G	89.37	45.71	47,532,228(74.71%)
F_L2	61,415,336	58,612,946	8.79G	89.15	45.56	43,734,592(74.62%)
L_L1	61,954,640	59,075,134	8.86G	89.28	45.48	45,017,258(76.20%)
L_L2	59,517,032	56,317,054	8.45G	88.78	45.65	42,732,512(75.88%)
Summary	249,632,704	237,624,594	35.64G			179,016,590(75.34%)

[a]Q30: The percentage of bases with a Phred value > 30

Chen *et al. BMC Genomics* (2018) 19:550

Page 6 of 15

Table 3 Length distribution of transcripts and unigenes

Length distribution	Transcripts		Unigenes	
	Number	Percentage (%)	Number	Percentage (%)
201–500	170,861	61.86	53,633	34.34
501–1000	54,976	19.90	52,225	33.44
1001–2000	32,396	11.74	32,370	20.72
> 2000	17,956	6.50	17,956	11.50
Total	276,189	100.00	156,184	100
Total length (bp)	190,057,842		155,645,929	
Mean length (bp)	688		997	
N50 (bp)	1085		1430	

borneol-type relative to the linalool-type, 738 unigenes of which were borneol-type unique, 524 unigenes of which were linalool-type unique, and 1601 mixed assembly unigenes in the two chemotypes. The distribution of these genes is shown in Fig. 3.

To explore the difference of metabolic pathways between borneol chemotype and linalool chemotype, enriched KEGG pathway analysis has been performed using 2863 differentially expressed genes. The top 20 over-presented KEGG pathways categories were presented as dot plots (Additional file 4). The most significantly enriched pathway for up-regulated DEGs in the borneol chemotype relative to the linalool type was the "monoterpenoid biosynthesis" pathway, with three DEGs annotated. For the up-regulated DEGs in linalool chemotype, the most significantly enriched pathway was "folate biosynthesis".

In the GO enrichment analysis, 124 up-regulated DEGs between the borneol-type and linalool-type profiles were enriched and matched in 7 GO terms that belonged to biological process and molecular function (Fig. 4a). The structure of GO can be described in the form of a directed acyclic graph (DAG) in which each GO term is depicted as a node and the parentages as an arrow. Peroxidase activity and terpene synthase activity were significantly enriched (Fig. 4b).

Candidate genes involved in terpenoid biosynthesis

To explore the regulatory mechanisms for the accumulation patterns of different terpenoids in *C.camphora*, the expression profiles of genes involved in terpenoid biosynthesis were analyzed. A total of 67 expressed unigenes encoding terpenoid biosynthesis enzymes were identified in *C. camphora* (Additional file 3). The expression data of all these unigenes is shown in Fig. 5, as well as the FPKM values in Additional file 5. Most of the genes encoding key enzymes in terpenoid backbone pathway (MEP and MVA pathway, KEGG entry ko00900) exhibited a high transcriptome expression level, were not differentially expressed between the two

chemotypes except for one HMGR and two MCS unigenes. Interestingly, unigenes in the MEP pathway show a higher expression level than those of the MVA pathway (Fig. 5). Both MEP and MVA pathways generate IPP and its isomer DMAPP, which are precursors of the production of terpenoids. It is indicated that active biosynthesis of building blocks make contribution to a large accumulation of different class of terpenoids which is consistent with our composition analysis of leaf extract in *C. camphora* (Fig. 1c).

In addition, we characterized twenty-four unigenes encoding terpenoid synthase, including eight in monoterpenoid biosynthesis, seven in diterpenoid biosynthesis and nine in sesquiterpenoid and triterpenoid biosynthesis(Fig. 5). Among them, three monoterpenoid synthase unigenes, one diterpenoid synthase unigene and two unigenes involved in sesquiterpenoid and triterpenoid synthase were up-regulated in the borneol chemotype, while only one diterpenoid synthase unigenes were up-regulated in the linalool chemotype. Overall, our results of the high expression levels for the enzymes in terpenoid backbone pathway are consistent with a high rate of terpenoid synthesis, and suggest that the different expression level of TPS genes in the two chemotypes may be the reason for their different terpenoid compositions.

qRT-PCR validation of DEGs from the RNA-Seq analysis

To validate the expression patterns of terpenoid biosynthetic genes obtained from RNA-Seq analysis, qRT-PCR was conducted to examine the expression levels of twelve unigenes in the two chemotypes (Fig. 6). The expression levels of these selected genes from qRT-PCR analyses were generally consistent with those deduced from their fragments per kilobase per million mapped (FPKM) data from RNA-Seq (Fig. 6a). The correlation between the qRT-PCR and RNA-Seq measurements was evaluated and the coefficient of determination (or R-squared) was 0.9231(Fig. 6b). The obtained results confirm the reliability of the transcriptomic profiling
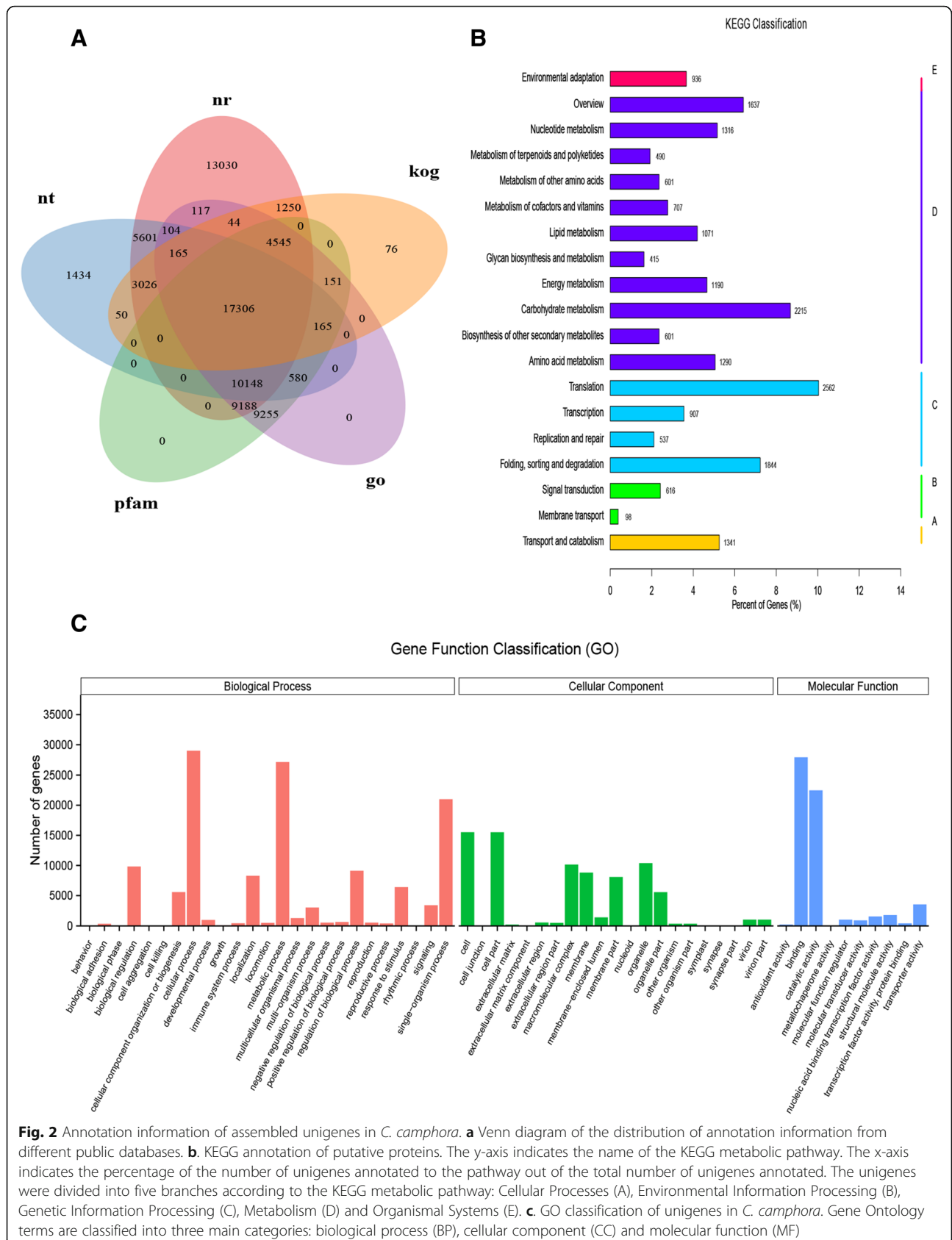
Chen *et al. BMC Genomics* (2018) 19:550

Page 7 of 15

Fig. 2 Annotation information of assembled unigenes in *C. camphora*. **a** Venn diagram of the distribution of annotation information from different public databases. **b**. KEGG annotation of putative proteins. The y-axis indicates the name of the KEGG metabolic pathway. The x-axis indicates the percentage of the number of unigenes annotated to the pathway out of the total number of unigenes annotated. The unigenes were divided into five branches according to the KEGG metabolic pathway: Cellular Processes (A), Environmental Information Processing (B), Genetic Information Processing (C), Metabolism (D) and Organismal Systems (E). **c**. GO classification of unigenes in *C. camphora*. Gene Ontology terms are classified into three main categories: biological process (BP), cellular component (CC) and molecular function (MF)
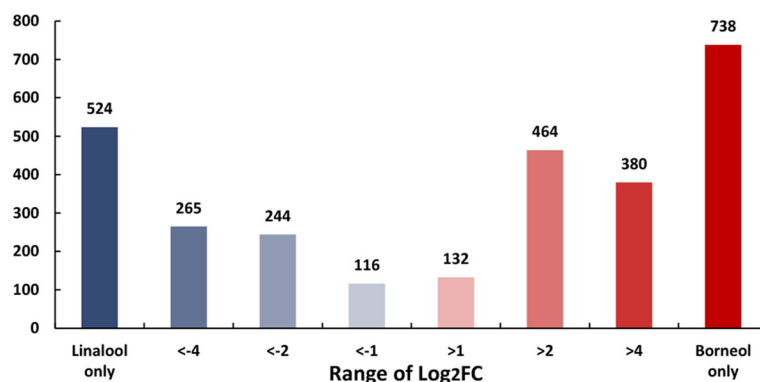
Chen *et al. BMC Genomics* (2018) 19:550

Page 8 of 15

Fig. 3 Number of and differentially expressed genes (DEGs) in Borneol-type and Linalool-type of *C. camphora*. Red: upregulated genes in the borneol-type; Blue: downregulated genes in the borneol-type. Relative expression of DEGs selected at Q-value < 0.05. The darker colors of bars represent larger change in expression. The fold change (FC) was calculated as the ratio between the borneol-type and linalool-type. The *x*- axis represents the range of Log$_2$ FC. The *y*-axis indicates the number of detected DEGs

data estimated from RNA-Seq data. Moreover, the expression of these genes has also been validated in three other chemotypes (camphor-, cineole- and nerolidol-types) by qRT-PCR except for borneol- and linalool-types of *C. camphora* (Fig. 7).
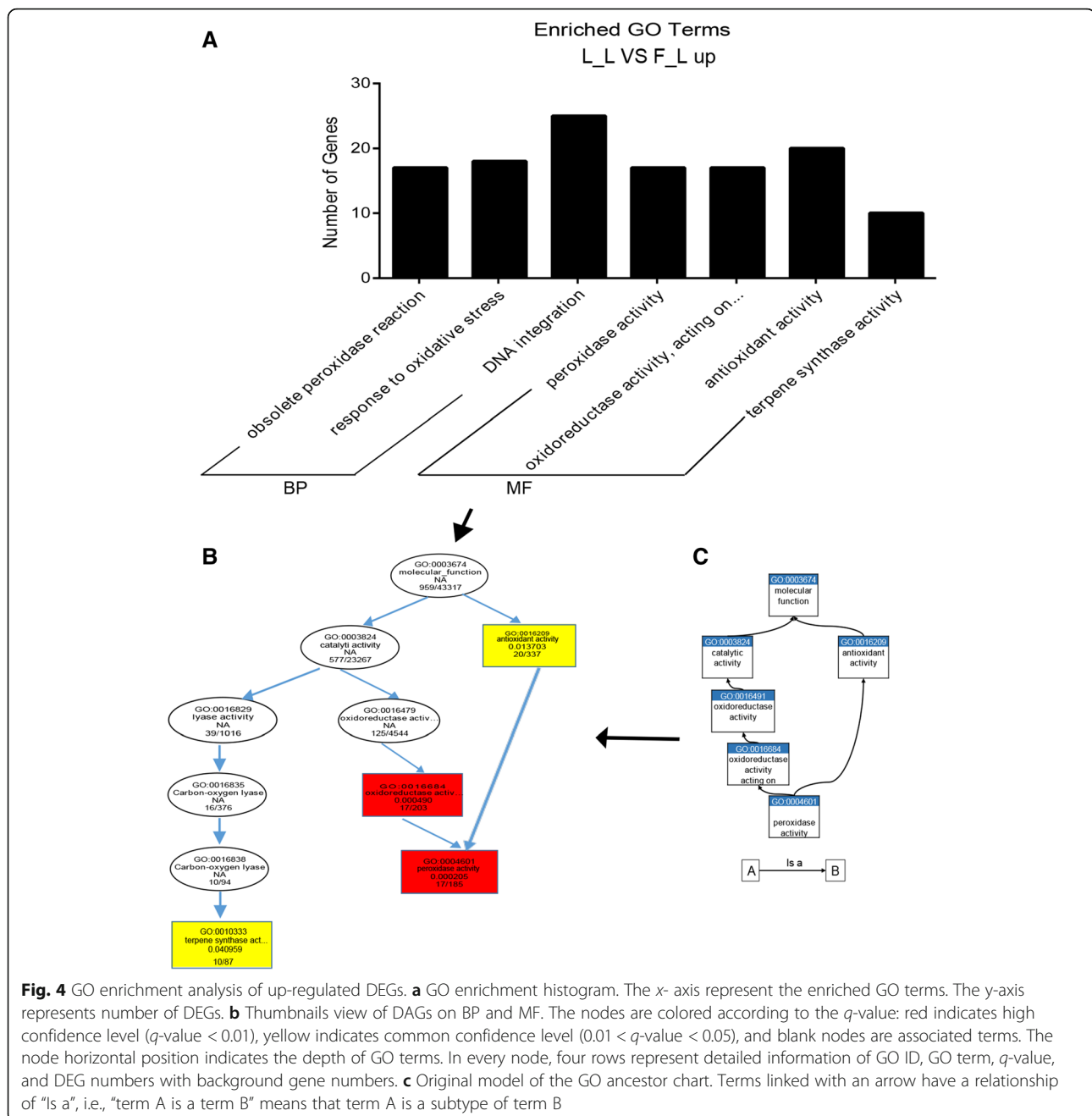
Discussion
Monoterpenes are the main components in leaf extracts in many members of the genus *Cinnamomum*, such as *C. osmophloeum* [43], *C. burmannii* [44] and *C. kanehirae* Hay [45]. In this study, 24 components were identified in borneol and linalool-chemotypes of *C. camphora*. An analysis of the components showed that the essential oil or crystal isolated from *C. camphora* consists primarily of monoterpenes with very minor quantities of sesquiterpenes and diterpenes. However, the most prevalent monoterpene differed between the two chemotypes, with a high content of (−)-borneol in the borneol-type and a higher content of linalool in the linalool-type than in the borneol-type. Therefore, the different major monoterpenes mainly contributed to the differences between the essential oil of the two chemotypes.

In recent years, genes related to terpenoids biosynthesis have been extensively studied in different plants. Terpenoids are rich in the Lauraceae family, while only several genes involved in terpenoid biosynthesis have been successfully identified and functionally characterized so far [21, 43]. In *Litsea cubeba*, three TPS genes encoding monoterpenoid synthase enzymes were isolated and functionally catalyzed the formation of trans-ocimene, α-thujene and (+)-sabinene [46]. In *Laurus nobilis*, TPS enzymes which catalyze the formation of 1,8-cineole, cadinenes and geranyllinalool were characterized [47]. The bioinformatic analysis presented here aimed to discover genes involved in terpenes synthesis and their precursors in the MVA and MEP

pathways. A total of 67 expressed unigenes likely involved in terpenoid biosynthesis were isolated from the transcriptome profile of *C. camphora*. Multiple unigenes were annotated as the same enzyme, and these unigenes may represent various alternatively spliced transcripts or members of a gene family.
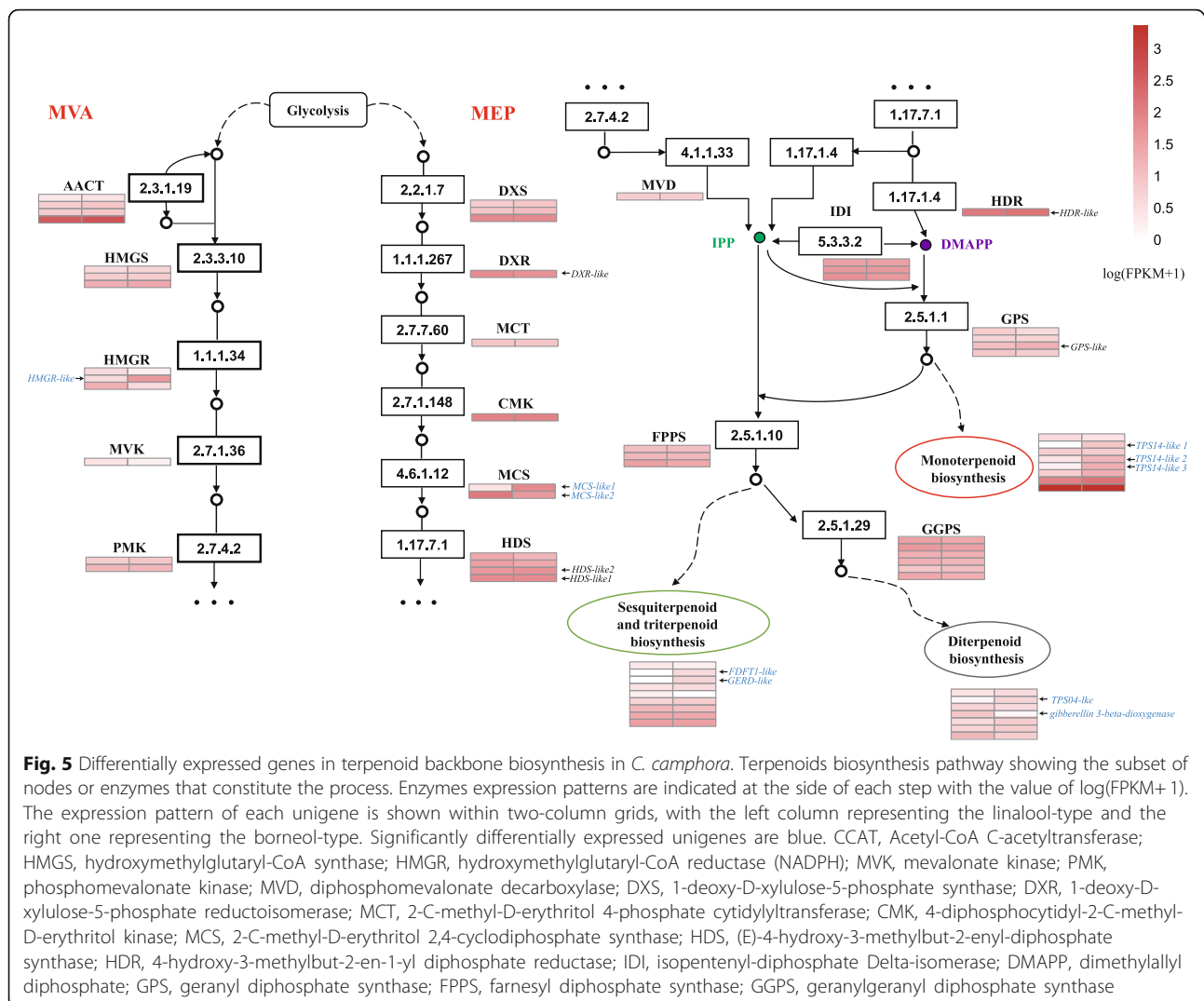
The regulation of plant MVA and MEP pathway gene expression occurs mainly at the transcriptional level [48]. HMGR and DXS are two rate-limiting enzymes in the MVA [49] and MEP pathway, respectively [50]. Thirty expressed unigenes which were identified to be highly homologous with 14 known enzymes have been annotated into the MVA and MEP pathways of terpenoid backbone biosynthesis in *C. camphora*, and shown to exhibit a high transcriptome expression level in both two chemotypes. The MEP pathway provides precursors for the synthesis of monoterpenes and diterpenes in plastids, whereas sesquiterpenes are derived from precursors of the MVA pathway in the cytosol [48]. Cross talk between these two different terpene backbone pathways has been documented, whereas the relative contribution of each pathway to the biosynthesis of the various class of terpenes remain uncertain.

A large number of different TPS and the fact that some TPS produce multiple products are the chief reasons for the variety in terpenes [51]. Linalool synthase isolated from *C. osmophloeum* were able to generate the S-(+)-linalool from geranyl diphosphate (GPP) and (E)-nerolidol from farnesyl diphosphate (FPP) [43]. In addition, the products of TPS can be further modified by oxidation, peroxidation, methylation and acylation etc., such as by cytochrome P450 dependent monooxygenases and short-chain dehydrogenases/reductases (SDRs) [52]. In the GO enrichment analysis, 124 up-regulated DEGs between the borneol-type and linalool-type profiles were enriched, among these 10 and 17 DEGs were

Chen *et al. BMC Genomics* (2018) 19:550

Page 9 of 15

Fig. 4 GO enrichment analysis of up-regulated DEGs. **a** GO enrichment histogram. The *x*- axis represent the enriched GO terms. The y-axis represents number of DEGs. **b** Thumbnails view of DAGs on BP and MF. The nodes are colored according to the *q*-value: red indicates high confidence level (*q*-value < 0.01), yellow indicates common confidence level (0.01 < *q*-value < 0.05), and blank nodes are associated terms. The node horizontal position indicates the depth of GO terms. In every node, four rows represent detailed information of GO ID, GO term, *q*-value, and DEG numbers with background gene numbers. **c** Original model of the GO ancestor chart. Terms linked with an arrow have a relationship of "Is a", i.e., "term A is a term B" means that term A is a subtype of term B

significantly enriched in the terpene synthase activity and oxidoreductase activity term of the DAG, respectively, indicating that terpene synthase activity and oxidoreductase activity were likely to be the terms leading to the differential accumulation of terpenoids in borneol and linalool chemotypes of *C. camphora*. In addition, 37 unigenes related to pathways of prenyl diphosphat, monoterpenoid, sesquiterpenoid and diterpenoid biosynthesis were identified in the current study. Mono-, sesqui–/tri-, and di-TPS have been synthesized by the activities of geranyl diphosphate synthase (GPS), farnesyl diphosphate synthase (FPPS) and geranylgeranyl diphosphate synthase (GGPPS) from GPP, FPP and geranylgeranyl diphosphate (GGPP), respectively [9]. Four GPS, three FPPS and six GGPS have been identified in both borneol and linalool chemotypes but with no difference in expression levels. In addition, a total of seven TPSs including three mono-TPSs, two di-TPSs and sesqui –/tri-TPSs were differentially expressed in the different chemotypes. This suggests that the different expression level of TPS genes in the two chemotypes may be the reason for their accumulation of different terpenoids.

Chen *et al. BMC Genomics* (2018) 19:550

Page 10 of 15

Fig. 5 Differentially expressed genes in terpenoid backbone biosynthesis in *C. camphora*. Terpenoids biosynthesis pathway showing the subset of nodes or enzymes that constitute the process. Enzymes expression patterns are indicated at the side of each step with the value of log(FPKM+ 1). The expression pattern of each unigene is shown within two-column grids, with the left column representing the linalool-type and the right one representing the borneol-type. Significantly differentially expressed unigenes are blue. CCAT, Acetyl-CoA C-acetyltransferase; HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase (NADPH); MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MVD, diphosphomevalonate decarboxylase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; IDI, isopentenyl-diphosphate Delta-isomerase; DMAPP, dimethylallyl diphosphate; GPS, geranyl diphosphate synthase; FPPS, farnesyl diphosphate synthase; GGPS, geranylgeranyl diphosphate synthase

In previous research, (+)-borneol as an intermediate product of camphor biosynthesis has been proved to be derived from the conversion of GPP to bornyl pyrophosphate (BPP) by the action of bornyl pyrophosphate synthetase (BPPS) and BPP was subsequently hydrolyzed to borneol [53]. BPPS, a metal-requiring monoterpene cyclase [54], has to date been cloned only from *Salvia officinalis* [53] and *Lavandula angustifolia* [55]. Linalool, an acyclic monoterpene, is a main product of linalool synthases and bi-functional synthases for both linalool and nerolidol production have been characterized in many plant species [51]. It is interesting that three of the mono-TPSs, *TPS14_like*, *TPS14_like2* and *TPS14_like3*, showed an up-regulated expression in the borneol-type compared to the linalool-type in this study, and their protein were aligned with some linalool synthases and borneol pyrophosphate synthases that have been published already (Additional file 6). However, TPS14 is a

linalool synthase expressed in *Arabidopsis thaliana* flowers [56]. It has been reported that the specific biochemical functions of individual TPS family members cannot be predicted based on sequence similarity alone, as changes in only a few amino acids can lead to drastic changes in the terpenoid profile of a given TPS enzyme [18, 57]. In addition, many TPSs are multi-product enzymes and can often give rise to mixtures of the same compounds in differing proportions [58]. Furthermore, linalool biosynthesis can be altered by alternative transcript splicing of key biosynthetic enzymes in *Camellia sinensis* [59]. These make it difficult to determine transcript abundance of individual *TPS* unigenes using qRT-PCR. Nevertheless, *TPS14_like*, *TPS14_like2* and *TPS14_like3* may play an important role in terpenoid accumulation in the borneol-type of *C. camphora*. The specific function of these three genes remains to be further verified.
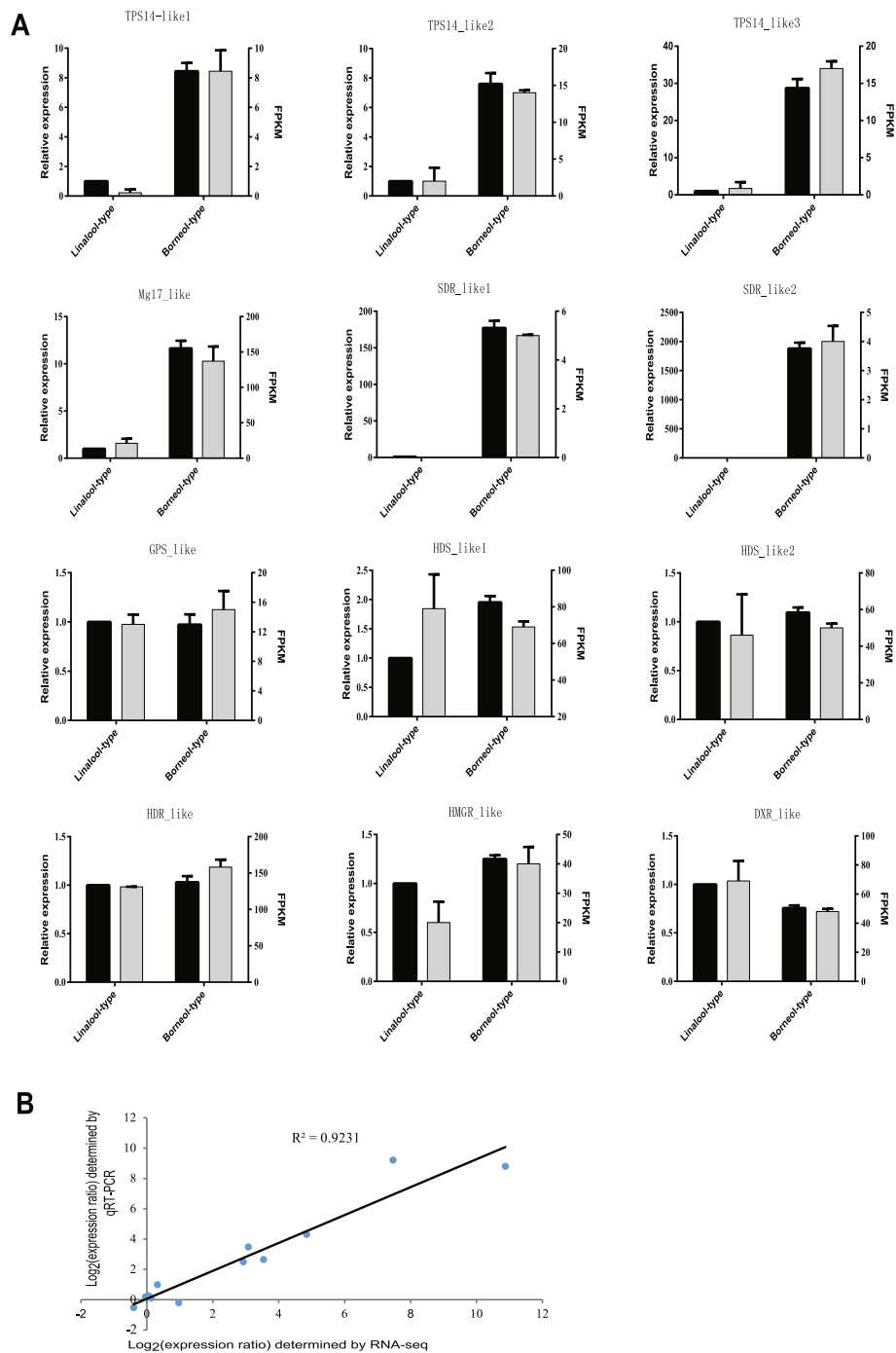
Chen *et al. BMC Genomics* (2018) 19:550

Page 11 of 15

Fig. 6 qRT-PCR validation of selected genes in linalool and borneol chemotypes of *C. camphora*. **a** The gray bars represent the relative expression determined with RT-qPCR (left y-axis) and the black bars represent the level of expression (FPKM) of the transcripts (right y-axis). The relative expression levels were estimated from the threshold of PCR cycle with the delta-delta CT method. The error bars indicate the standard errors from two biological and three technical replicates. **b** Scatter plots show simple linear regression and the R-squared (R^2) between RNA sequencing data and qRT-PCR validation data expressed in terms of log_2FC. The fold change (FC) was calculated as the ratio between the linalool-type and borneol-type of *C. camphora*

Chen *et al. BMC Genomics* (2018) 19:550

Page 12 of 15
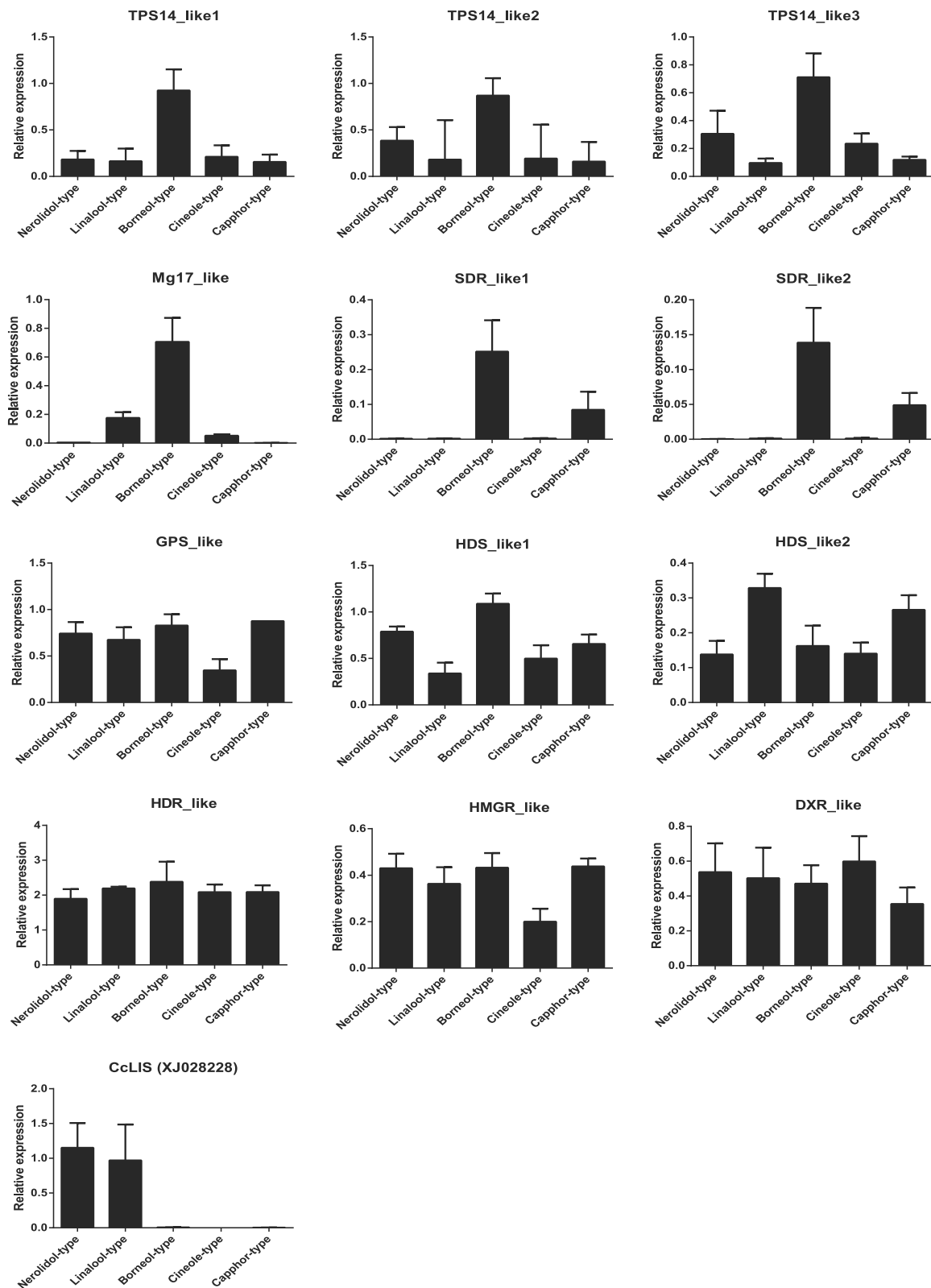
Fig. 7 qRT-PCR validation of selected genes in five chemotypes of *C. camphora*. Five chemotypes include camphor-, cineole-, nerolidol-, borneol- and linalool-types. Relative expression levels were estimated from the threshold of the PCR cycle by the Delta CT method. The values indicate the means of two biological replications

Chen *et al. BMC Genomics* (2018) 19:550

Page 13 of 15

Conclusion

In this study, we conducted leaf transcriptome and metabolic analysis between linalool- and borneol-chemotypes in *C. camphora*. This data provides a comprehensive coverage of terpenoid biosynthesis in *C. camphora*. Monoterpenes were identified as the major components in the leaf of the two chemotypes. Beta-linalool was the most abundant component in the linalool-chemotype while (−)-borneol was the major component in the borneol-chemotype. A comparison of the transcriptomes of these two chemotypes led to the identification of 2863 differentially expressed unigenes. Analysis of these unigenes provides an insight into the gene expression patterns and biological processes active in *C. camphora*. GO and KEGG enrichment analysis revealed that terpene synthase activity and oxidoreductase activity could explain the differential accumulation of terpenoids between borneol and linalool chemotypes in *C. camphora*. A total of 67 candidate unigenes were identified to be involved in terpenoid biosynthesis in *C. camphora*. Most notably, three unigenes involved in monoterpenoid biosynthesis were identified. Further functional studies are needed to elucidate regulatory mechanisms in the formation and accumulation of terpenoids. The transcriptome sequences and gene expression profile provide valuable information for understanding the accumulation of terpenoids in different chemotypes of *C. camphora*. The study also provides worthy resources for bioengineering and synthetic biology study of terpenoids in *C. camphora*.

Additional files

Additional file 1: Composition of the leaf extracts of camphor-, cineole- and nerolidol-types *C. camphora*. (XLSX 11 kb)

Additional file 2: Primer sequences used for quantitative reverse transcription polymerase chain reaction (qRT-PCR). (XLSX 10 kb)

Additional file 3: Sequences of candidate unigenes were identified to be involved in terpenoid biosynthesis. (TXT 108 kb)

Additional file 4: Enriched KEGG pathway analysis of significantly differentially expressed genes between borneol- and linalool-chemotypes. F_L was linalool-type, L_L was borneol-type. The y- axis indicates the pathway name; the x-axis indicates the enrichment factor corresponding to the pathway. The Rich factor means that the ratio of the enriched DEGs number and the number of background genes in corresponding pathway. The greater the Rich factor, the greater the degree of enrichment. The *q*-value is represented by the color of the dot (scale provided to the right of each panel). The number of DEGs is represented by the size of the dots. (PDF 227 kb)

Additional file 5: FPKM values of terpenoid biosynthesis pathway unigenes. (XLSX 15 kb)

Additional file 6: Alignment and phylogenetic tree of the amino acid sequences of linalool and borneol pyrophosphate synthases. (A) Alignment of the amino acid sequences of linalool and borneol pyrophosphate synthases. LaLINs, (−)-(3S)linalool synthase, *Lavandula angustifolia*, ABB73045.1; MaLIS, linalool synthase, *Mentha aquatica*, AAL99381.1; SoSBS, (+)-bornyl diphosphate synthase, *Salvia officinalis*, AAC26017.1; AaQH, (3R)-linalool synthase, *Artemisia annua*, AAF13357.1; AaQH5, (3R)-linalool synthase *Artemisia annua*, AAF13356.1; LeMTS1,

(−)-(3R)-Linalool synthase 1, *Solanum lycopersicum*, AAX69063.1; PaTPS-Lin, (−)-linalool synthase *Picea abies*, AAS47693.1; CcLIS, putative linalool synthase, *Cinnamomum camphora*, XJ028228; OsLIS, linalool synthase, *Oryza sativa*, ACF05530.1; CsNES/LIS, nerolidol/linalool synthase, *Camellia sinensis*, AGX26045.1; FaNES1, (+)-(3S)-Linalool synthase, *Fragaria* x *ananassa*, CAD57081.1; AmNES/LIS-2, nerolidol/linalool synthase, *Antirrhinum majus*, ABR24418.1; AtLIS, linalool synthase, *Arabidopsis thaliana*, AAO85533.1; ObLIS, R-linalool synthase, *Ocimum basilicum*, AAV63789.1; CbLIS, linalool synthase, *Clarkia breweri*, AAD19840.1. (B) Phylogenetic analysis by Maximum likelihood method. Evolutionary analyse were coducted in MEGA7. (PDF 209 kb)

Abbreviations

bp: Base pairs (measuring unit); DAG: Directed acyclic graphs; DEGs: Differentially expressed genes; DMAPP: Dimethylallyl diphosphate; FPKM: Fragments per kilobase of exon per million fragments mapped; FPP: Farnesyl diphosphate; FPPS: Farnesyl diphosphate synthase; FW: Fresh weight; GC-MS: Gas chromatography-mass spectrometry; GGPP: Geranylgeranyl diphosphate; GGPS: Geranylgeranyl diphosphate synthase; GO: Gene ontology; GPP: Geranyl diphosphate; 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 4-phosphate; MVA: Mevalonic acid; Nr: NCBI non-redundant protein sequences; Nt: NCBI non-redundant nucleotide sequences; Pfam: Protein family; qRT-PCR: Quantitative reverse transcription PCR; RNA-Seq: RNA-sequencing; Swiss-Prot: A manually annotated and reviewed protein sequence database; TPS: Terpene synthases

Availability of data and materials

The data supporting the conclusions of this article (raw RNA-Seq reads) are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA): SRP127892, https://www.ncbi.nlm.nih.gov/sra/?term=SRP127892.

Authors' contributions

CC and YJZ (Yongjie Zheng) performed the experiments, analyzed the data and drafted the manuscript. YDZ (Yongda Zhong) analyzed the data. YW collected samples and analyzed the data. LT carried out the experiments and analyzed the data. LX conceived and designed the experiments. MX conceived and designed the experiments, reviewed manuscript. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

The different chemotypes of *Cinnamomum camphora* used in this study was produced by Jiangxi Academy of Forestry in Nanchang and Nanjing Forestry University in Nanjing, China. *C. camphora* was widely cultivated in the south of China. All of voucher specimens were deposited in the Herbarium of Jiangxi Academy of Forestry and Nanjing Forestry University. Experimental research on the different chemotypes of *C. camphora*, including collection of this plant material, complies with institutional, national and international guidelines.

Consent for publication

Not applicable.

Chen *et al. BMC Genomics* (2018) 19:550

Page 14 of 15

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China. [2]Camphor Engineering Technology Research Center for State Forestry Administration, Jiangxi Academy of Forestry, Nanchang 330032, China. [3]Institute of Biological Resources, Jiangxi Academy of Science, Nanchang, Jiangxi, China.

References

1. Babu KN, Sajina A, Minoo D, John CZ, Mini PM, Tushar KV, Rema J, Ravindran PN. Micropropagation of camphor tree (Cinnamomum camphora). Plant Cell Tissue Organ Cult. 2003;74(2):179–83. https://doi.org/10.1023/A:1023988110064.
2. Jiang H, Wang J, Song L, Cao X, Yao X, Tang F, Yue Y. GCxGC-TOFMS analysis of essential oils composition from leaves, twigs and seeds of Cinnamomum camphora L. Presl and their insecticidal and repellent activities. Molecules. 2016;21(4):423–35. https://doi.org/10.3390/molecules21040423.
3. Pragadheesh VS, Saroj A, Yadav A, Chanotiya CS, Alam M, Samad A. Chemical characterization and antifungal activity of Cinnamomum camphora essential oil. Ind Crop Prod. 2013;49(Supplement C):628–33. https://doi.org/10.1016/j.indcrop.2013.06.023.
4. Guo X, Cui M, Deng M, Liu X, Huang X, Zhang X, Luo L. Molecular differentiation of five Cinnamomum camphora chemotypes using desorption atmospheric pressure chemical ionization mass spectrometry of raw leaves. Sci Rep. 2017;7:46579. https://doi.org/10.1038/srep46579.
5. Shi WY, Wen GY, Guo DX, Long GY, Liu YG. Study on chemical constituents of the essential oil and classification of types from Cinnamomum camphora. Acta Bot. 1989;209-214(Sin. 31):5.
6. Su J, Chen J, Liao S, Li L, Zhu L, Chen L. Composition and biological activities of the essential oil extracted from a novel plant of Cinnamomum camphora Chvar. Borneol. J Med Plants Res. 2012; https://doi.org/10.5897/JMPR12.157.
7. Su J, Lai H, Chen J, Lin L, Wong YS, Chen T, Li X. Natural Borneol, a Monoterpenoid compound, potentiates Selenocystine-induced apoptosis in human hepatocellular carcinoma cells by enhancement of cellular uptake and activation of ROS-mediated DNA damage. PLoS One. 2013;8(5):e63502. https://doi.org/10.1371/journal.pone.0063502.
8. Koksal M, Jin Y, Coates RM, Croteau R, Christianson DW. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. Nature. 2011;469(7328):116–20. https://doi.org/10.1038/nature09628.
9. Vranová E, Coman D, Gruissem W. Structure and dynamics of the isoprenoid pathway network. Mol Plant. 2012;5(2):318–33. https://doi.org/10.1093/mp/sss015.
10. Breitmaier E. Terpenes: flavors, fragrances, Pharmaca, pheromones. Wiley-VCH. 2006:1–3. https://doi.org/10.1002/9783527609949.ch1.
11. Cordoba E, Salmi M, León P. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. J Exp Bot. 2009;60(10):2933–43. https://doi.org/10.1093/jxb/erp190.
12. Lichtenthaler HK. The 1-deoxy-d-XYLULOSE-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol. 1999;50(1):47–65. https://doi.org/10.1146/annurev.arplant.50.1.47.
13. Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, Gidoni D, Eyal Y. Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid Mevalonic acid pathway is compartmentalized to peroxisomes. Plant Physiol. 2008;148(3):1219–28. https://doi.org/10.1104/pp.108.127951.
14. Rohmer M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat Prod Rep. 1999;16(5):565–74. https://doi.org/10.1039/a709175c.
15. Okada K, Kasahara H, Yamaguchi S, Kawaide H, Kamiya Y, Nojiri H, Yamane H. Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in Arabidopsis. Plant Cell Physiol. 2008;49(4):604–16. https://doi.org/10.1093/pcp/pcn032.
16. Chen F, Tholl D, Bohlmann J, Pichersky E. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 2011;66(1):212–29. https://doi.org/10.1111/j.1365-313X.2011.04520.x.
17. Tholl D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr Opin Plant Biol. 2006;9(3):297–304. https://doi.org/10.1016/j.pbi.2006.03.014.
18. Zulak KG, Lippert DN, Kuzyk MA, Domanski D, Chou T, Borchers CH, 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(6):1015–30. https://doi.org/10.1111/j.1365-313X.2009.04020.x.
19. Herde M, Gärtner K, Köllner TG, Fode B, Boland W, Gershenzon J, Gatz C, Tholl D. Identification and regulation of TPS04/GES, an Arabidopsis Geranyllinalool synthase catalyzing the first step in the formation of the insect-induced volatile C16 HomoterpeneTMTT. Plant Cell. 2008;20(4):1152–68. https://doi.org/10.1105/tpc.106.049478.
20. Matarese F, Cuzzola A, Scalabrelli G, D'Onofrio C. Expression of terpene synthase genes associated with the formation of volatiles in different organs of Vitis vinifera. Phytochemistry. 2014;105(Supplement C):12–24. https://doi.org/10.1016/j.phytochem.2014.06.007.
21. Yang T, Li J, Wang H-X, Zeng Y. A geraniol-synthase gene from Cinnamomum tenuipilum. Phytochemistry. 2005;66(3):285–93. https://doi.org/10.1016/j.phytochem.2004.12.004.
22. A. Nagegowda D, Dudareva N: Plant biochemistry and biotechnology of flavor compounds and essential oils. In: Medicinal Plant Biotechnology. Wiley-VCH Verlag GmbH; 2008: 469–492. doi: https://doi.org/10.1002/9783527619771.ch20.
23. Ament K, Kant MR, Sabelis MW, Haring MA, Schuurink RC. Jasmonic acid is a key regulator of spider mite-induced volatile Terpenoid and methyl salicylate emission in tomato. Plant Physiol. 2004;135(4):2025–37. https://doi.org/10.1104/pp.104.048694.
24. Richter A, Seidl-Adams I, Köllner TG, Schaff C, Tumlinson JH, Degenhardt J. A small, differentially regulated family of farnesyl diphosphate synthases in maize (Zea mays) provides farnesyl diphosphate for the biosynthesis of herbivore-induced sesquiterpenes. Planta. 2015;241(6):1351–61. https://doi.org/10.1007/s00425-015-2254-z.
25. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2013;14(6):671. https://doi.org/10.1093/bib/bbs046.
26. Soetaert SS, Van Neste CM, Vandewoestyne ML, Head SR, Goossens A, Van Nieuwerburgh FC, Deforce DL. Differential transcriptome analysis of glandular and filamentous trichomes in Artemisia annua. BMC Plant Biol. 2013;13(1):220. https://doi.org/10.1186/1471-2229-13-220.
27. Jin J, Panicker D, Wang Q, Kim MJ, Liu J, Yin J-L, Wong L, Jang I-C, Chua N-H, Sarojam R. Next generation sequencing unravels the biosynthetic ability of spearmint (Mentha spicata) peltate glandular trichomes through comparative transcriptomics. BMC Plant Biol. 2014;14(1):292. https://doi.org/10.1186/s12870-014-0292-5.
28. Jiang X, Wu Y, Xiao F, Xiong Z, Xu H. Transcriptome analysis for leaves of five chemical types in Cinnamomum camphora. Hereditas (Beijing). 2014;36(1):58–68. https://doi.org/10.3724/SP.J.1005.2014.0069.
29. Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. Plant J. 2015;82(6):951–61. https://doi.org/10.1111/tpj.12865.
30. Niu J, Hou X, Fang C, An J, Ha D, Qiu L, Ju Y, Zhao H, Du W, Qi J, et al. Transcriptome analysis of distinct Lindera glauca tissues revealed the differences in the unigenes related to terpenoid biosynthesis. Gene. 2015;559(1):22–30. https://doi.org/10.1016/j.gene.2015.01.002.
31. Yang M, You W, Wu S, Fan Z, Xu B, Zhu M, Li X, Xiao Y. Global transcriptome analysis of Huperzia serrata and identification of critical genes involved in the biosynthesis of huperzine a. BMC Genomics. 2017;18(1):245. https://doi.org/10.1186/s12864-017-3615-8.
32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Chen *et al. BMC Genomics* (2018) 19:550

Page 15 of 15

33. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Xian A, Fan L, Raychowdhury R, Zeng Q. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011; 29(7):644. https://doi.org/10.1038/nbt.1883.

34. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol. 2014;15(7): 410. https://doi.org/10.1186/PREACCEPT-2088857056122054.

35. Götz S, Garcíagómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10): 3420. https://doi.org/10.1093/nar/gkn176.

36. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7): 621. https://doi.org/10.1038/nmeth.1226.

37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323. https://doi.org/10.1186/1471-2105-12-323.

38. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. https://doi.org/10.1186/gb-2010-11-10-r106.

39. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440–5. https://doi.org/10.1073/pnas. 1530509100.

40. Young MD, Wakefield MJ, Smyth GK, Alicia O. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14. https:// doi.org/10.1186/gb-2010-11-2-r14.

41. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics. 2005;21(19):3787–93. https://doi.org/10.1093/ bioinformatics/bti430.

42. Darbani B, Stewart CN, Noeparvar S, Borg S. Correction of gene expression data: performance-dependency on inter-replicate and inter-treatment biases. J Biotechnol. 2014;188(Supplement C):100–9. https://doi.org/10.1016/ j.jbiotec.2014.08.012.

43. 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. 2013;10(1):75–86. https://doi.org/10. 1007/s11295-013-0665-1.

44. Chen L, Su J, Lin Li LB, Li W. A new source of natural D-borneol and its characteristic. J Med Plants Res. 2011;5(15):7.

45. 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 kanehiraeUsing gas chromatography/mass spectrometry, cluster analysis, and principal component analysis. J Wood Chem Technol. 2015;35(3):207–19. https://doi.org/10.1080/02773813.2014.924967.

46. Chang Y-T, Chu F-H. Molecular cloning and characterization of monoterpene synthases from Litsea cubeba (Lour.) Persoon. Tree Genet Genomes. 2011;7(4):835–44. https://doi.org/10.1007/s11295-011-0377-3.

47. Yahyaa M, Matsuba Y, Brandt W, Doron-Faigenboim A, Bar E, McClain A, Davidovich-Rikanati R, Lewinsohn E, Pichersky E, Ibdah M. Identification, Functional Characterization, and Evolution of terpene synthases from a basal dicot. Plant Physiol. 2015;169(3):1683–97. https://doi.org/10.1104/pp.15.00930.

48. Vranova E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol. 2013;64:665–700. https://doi.org/10.1146/annurev-arplant-050312-120116.

49. Friesen JA, Rodwell VW. The 3-hydroxy-3-methylglutaryl coenzyme-a (HMG-CoA) reductases. Genome Biol. 2004;5(11):248. https://doi.org/10.1186/gb-2004-5-11-248.

50. 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(19):4446. https://doi.org/10.1111/j.1742-4658.2006.05446.x.

51. Degenhardt J, Kollner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry. 2009;70(15–16):1621–37. https://doi.org/10.1016/j. phytochem.2009.07.030.

52. Okamoto S, Yu F, Harada H, Okajima T, Hattan J, Misawa N, Utsumi R. A short-chain dehydrogenase involved in terpene metabolism from Zingiber zerumbet. FEBS J. 2011;278(16):2892–900. https://doi.org/10.1111/j.1742-4658.2011.08211.x.

53. Croteau R, Felton M, Karp F, Kjonaas R. Relationship of camphor biosynthesis to leaf development in sage (Salvia officinalis). Plant Physiol. 1981;67(4):820–4. https://doi.org/10.1104/pp.67.4.820.

54. Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci U S A. 2002;99(24): 15375–80. https://doi.org/10.1073/pnas.232591099.

55. Despinasse Y, Fiorucci S, Antonczak S, Moja S, Bony A, Nicole F, Baudino S, Magnard JL, Jullien F. Bornyl-diphosphate synthase from Lavandula angustifolia: a major monoterpene synthase involved in essential oil quality. Phytochemistry. 2017;137:24–33. https://doi.org/10.1016/j. phytochem.2017.01.015.

56. Chen F. Biosynthesis and emission of Terpenoid volatiles from Arabidopsis flowers. Plant Cell Online. 2003;15(2):481–94. https://doi.org/10.1105/tpc.007989.

57. 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 Online. 2003;15(5):1227–41. https://doi.org/10.1105/tpc.011015.

58. Martin DM, 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(4):1908–27. https://doi.org/10.1104/pp.104.042028.

59. Liu GF, Liu JJ, He ZR, Wang FM, Yang H, Yan YF, Gao MJ, Gruber MY, Wan XC, Wei S. Implementation of CsLIS/NES in linalool biosynthesis involves transcript splicing regulation in Camellia sinensis. Plant Cell Environ. 2017; https://doi.org/10.1111/pce.13080.