

# SCIENTIFIC REPORTS

**OPEN**

## Integrated transcriptome sequencing and dynamic analysis reveal carbon source partitioning between terpenoid and oil accumulation in developing *Lindera glauca* fruits

Jun Niu, Yinlei Chen, Jiyong An, Xinyu Hou, Jian Cai, Jia Wang, Zhixiang Zhang & Shanzhi Lin

Received: 21 May 2015

Accepted: 15 September 2015

Published: 08 October 2015

*Lindera glauca* fruits (LGF) with the abundance of terpenoid and oil has emerged as a novel specific material for industrial and medicinal application in China, but the complex regulatory mechanisms of carbon source partitioning into terpenoid biosynthetic pathway (TBP) and oil biosynthetic pathway (OBP) in developing LGF is still unknown. Here we perform the analysis of contents and compositions of terpenoid and oil from 7 stages of developing LGF to characterize a dramatic difference in temporal accumulative patterns. The resulting 3 crucial samples at 50, 125 and 150 days after flowering (DAF) were selected for comparative deep transcriptome analysis. By Illumina sequencing, the obtained approximately 81 million reads are assembled into 69,160 unigenes, among which 174, 71, 81 and 155 unigenes are implicated in glycolysis, pentose phosphate pathway (PPP), TBP and OBP, respectively. Integrated differential expression profiling and qRT-PCR, we specifically characterize the key enzymes and transcription factors (TFs) involved in regulating carbon allocation ratios for terpenoid or oil accumulation in developing LGF. These results contribute to our understanding of the regulatory mechanisms of carbon source partitioning between terpenoid and oil in developing LGF, and to the improvement of resource utilization and molecular breeding for *L. glauca*.

*Lindera glauca*, a member of the family Lauraceae and the genus *Lindera* Thunb, is widely distributed in the forests at low altitudes only in China, Japan and Korea<sup>1-5</sup>. In China, this plant is one of the most economically and ecologically important and intensively studied tree species owing to its very plentiful resource, superior adaptability, ecological benefits and especially abundant terpenoid contents<sup>2-9</sup>. More recently, the abundance of oil (approximately 15% of fresh weight) has been characterized mainly with capric acid (11.83%), palmitic acid (30.13%), oleic acid (36.18%) and linoleic acid (12.92) in LGF<sup>1</sup>. Thus, LGF has recently emerged as a novel specific experimental material to explore the regulatory mechanism of high-quantity accumulation of terpenoid and oil in plants. Importantly, the precursors required for either terpenoid or oil biosynthesis are mostly from glycolysis and pentose phosphate pathway (PPP)<sup>10,11</sup>,

College of Biological Sciences and Biotechnology, College of Nature Conservation, National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, Beijing Forestry University, Beijing 10083, China. Correspondence and requests for materials should be addressed to S.L. (email: szlin@bjfu.edu.cn)

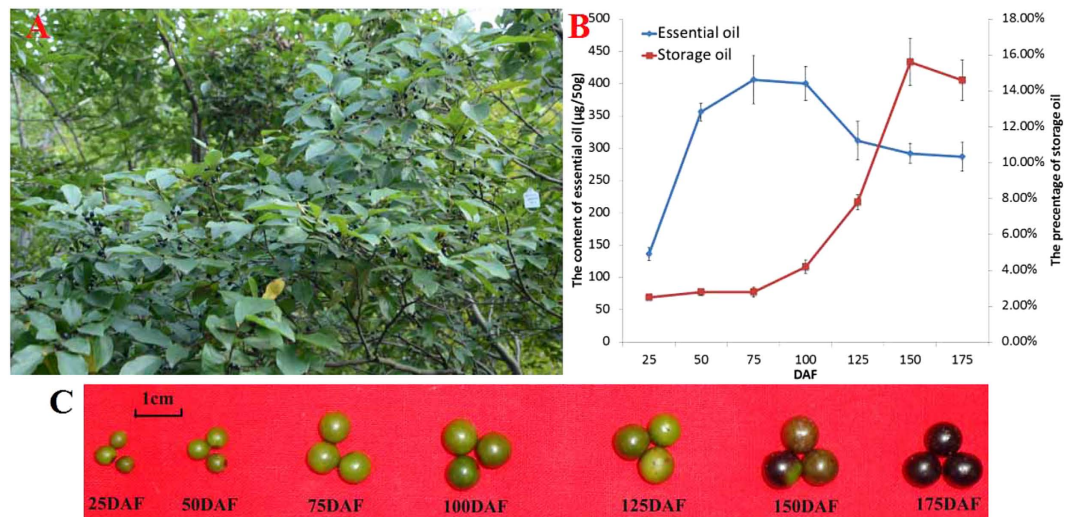
thus the complex and interactional regulatory mechanisms of carbon partitioning between terpenoid and oil in developing LGF.

Carbon is supplied to the heterotrophic organs (fruits, seeds, roots and tubers) mostly as sucrose from photosynthetic tissues, and sucrose synthase (SuSy) or sucrose (SUC) catalyze the cleavage of sucrose to hexoses (glucose and fructose) in cell wall<sup>12</sup>, which are then converted to glyceraldehyde 3-phosphate (G3P) via PPP or pyruvate (PYR) via the glycolysis in both cytosol and plastid, as carbon source for terpenoid and oil biosynthesis<sup>10,11</sup>. In general, the interchange of glycolytic intermediates between cytosol and plastid in plants has been implicated in the highly selective transporters, including glycolipid transporter (GLT), glucose-6-phosphate transporter (GPT), triose phosphate transporter (TPT) and phosphoenolpyruvate transporter (PPT)<sup>13–15</sup>. Additionally, because biological membranes are impermeable to acetyl-CoA as an important precursor for *de novo* fatty acid (FA) biosynthesis in plastids, terpene biosynthesis and FA elongation in cytosol, this molecule must be synthesized within each subcellular compartment by plastidial PYR dehydrogenase complex (PDHC), mitochondrial PDHC or cytosolic ATP-citrate lyase (ACL)<sup>16,17</sup>. Thus, the multiple mechanisms of generating acetyl-CoA for different acetyl-CoA-requiring metabolisms indicate the complexity of carbon flux into TBP and OBP in developing LGF.

In higher plants, the identifications of functional enzymes involved in the biosynthesis of terpenoid ([http://www.genome.jp/kegg-bin/show\\_pathway?map01062](http://www.genome.jp/kegg-bin/show_pathway?map01062)) and oil (<http://aralip.plantbiology.msu.edu>) have been well facilitated by various experimental methods<sup>18–38</sup>. There are two pathways for terpenoid backbone biosynthesis, cytosolic mevalonate (MVA) and plastidial 2-Cmethyl-Derythritol 4-phosphate (MEP) pathway, in which the 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and hydroxymethylglutaryl-CoA reductase (HMGR) are well characterized as the rate-limiting enzymes in MEP and MVA pathway, respectively<sup>18,19</sup>. By geranyl diphosphate synthase (GPPS) and farnesyl diphosphate synthase (FPPS), the produced precursors of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) in MEP and MVA pathway are respectively catalyzed to geranyl diphosphate (GPP) and farnesyl diphosphate (FPP), as the important precursors required for monoterpenoids (C10) biosynthesis by monoterpene synthase (mono-TPS) and sesquiterpenoids (C15) biosynthesis by sesquiterpene synthase (sesqui-TPS), respectively<sup>20–23</sup>. As for oil biosynthesis, carboxylation of acetyl-CoA to malonyl-CoA has been well known as the first rate-limiting step in FA synthesis by a multi-subunit acetyl-CoA carboxylase (ACCase) complex, including carboxyltransferase (CT), biotin carboxylase (BC) and biotin carboxyl carrier protein (BCCP)<sup>13,24–26</sup>. Before entering the synthesis pathway of FAs, malonyl-CoA from acetyl-CoA is firstly transferred to malonyl-ACP by malonyl-CoA: ACP malonyltransferase (MCMT) to provide two-carbon unit at each step of elongation<sup>25,26</sup>. There are four reactions of condensation, reduction, dehydration and reduction occurred in each synthesized cycle, which is catalyzed by a series of regulatory enzymes, including ketoacyl-ACP Synthase III (KASIII), ketoacyl-ACP reductase (KAR), hydroxyacyl-ACP dehydrase (HAD), enoyl-ACP reductase (EAR) and KASI<sup>25,26</sup>. After 7 elongated cycles, the produced saturated C16:0-ACP can either be hydrolyzed by fatty acyl-ACP thioesterase B (FATB) to release free FAs or elongated by KASII to generate 18:0-ACP. Subsequently, the produced 18:0-ACP is desaturated to 18:1-ACP by stearoyl-ACP desaturase (SAD), and then 18:1-ACP be hydrolyzed to free C18:1 by FATA<sup>26</sup>. The resulting free FAs are exported from plastid to endoplasmic reticulum (ER), and converted to fatty acyl-CoAs by long-chain acyl CoA synthetase (LACS), as the substrate for esterification and elongation<sup>26,27</sup>. In ER, fatty acyl-CoAs can be further elongated sequentially by ketoacyl-CoA synthase (KCS), ketoacyl-CoA reductase (KCR1), hydroxyacyl-CoA dehydratase (HCD) and enoyl-CoA reductase (ECR) to produce long chain FAs (>18 carbon)<sup>26,28</sup>. TAG assembly from G3P and acyl-CoAs involves four enzymatic steps: first, two acylations of G3P by glycerol-3-phosphate acyltransferase (GPAT) and lysophosphatidic acid acyltransferase (LPAAT), followed by phosphatidic acid phosphatase (PAP), and a third acylation by diacylglycerol acyltransferase (DGAT) or phospholipid:diacylglycerol acyltransferase (PDAT)<sup>29–35</sup>. During the TAG synthesis, CDP-choline:1,2-sn-diacylglycerol cholinephosphotransferase (CPT) and phosphatidylcholine: diacylglycerol cholinephosphotransferase (PDCT) was characterized that exchanges acyl residues between DAG and phosphatidylcholine (PC) for further desaturated by FA desaturase 2/3 (FAD2/3)<sup>37,38</sup>. All these findings have shown one highly complex process involved in a series of regulatory enzymes for terpenoid and oil biosynthesis in plants. Although considerable progress of *L. glauca* has been made in the contents and compositions of terpenoid and oil<sup>2–9</sup>, literature and molecular resources available for *L. glauca* remain very scarce.

As one of deep sequencing technology, RNA sequencing without recourse to the genomic information has been widely applied for extensive and intensive transcriptional studies in many plant species, such as *Salvia miltiorrhiza*, *Liriodendron chinense*, *Elaeis guineensis*, *Phoenix dactylifera*, *Brassica napus*, *Ricinus communis*, *Euonymus alatus* and *Tropaeolum majus*<sup>39–42</sup>. More recently, our transcriptomic analysis for different tissues of *L. glauca* has been performed by Illumina transcriptome sequencing<sup>43</sup>, but the obtained data are not still suitable for us to deeply explore the comprehensive regulatory mechanisms of carbon source partitioning between the terpenoid and oil accumulation in developing LGF. Thus, the transcriptomic analysis of developing LGF has become an imperative.

In the current study, we used the LGF as a novel specific experimental material, and analyzed the temporal patterns for the contents and compositions of terpenoid and oil at 7 different developing stages (25, 50, 75, 100, 125, 150 and 175 DAF). In result, the representative periods were characterized for comparative deep transcriptomic analysis by using Illumina technology. After functional annotation and



**Figure 1.** The features of *Lindera glauca* fruits. (A) The *Lindera glauca* are fruiting. (B) The dynamical patterns for LGF oil and FAs at different development period. (C) 7 developing stages of LGF.

classification, the genes involved specifically in glycolysis, PPP, TBP and OBP were identified in developing LGF. We then focus on combining differential expression profiling and qRT-PCR to analysis the transcriptional levels for some key enzymes and transcription factors (TFs) associated with the regulation of carbon allocation ratios, highlighting the complexity of alternative carbon flux for terpenoid and oil accumulation in developing LGF for the first time.

## Results

**Different accumulative patterns of terpenoid and oil in developing LGF.** We systematically measured the contents and compositions of terpenoid and oil during the whole developmental stages of LGF from 25 DAF (immature stage) to 175 DAF (fully matured stage) for the first time. The terpenoid contents showed a significant increase from 25 DAF ( $136.05 \pm 9.99 \mu\text{g}/50 \text{g}$ ) to 75 DAF ( $406.09 \pm 36.97 \mu\text{g}/50 \text{g}$ ), and then a decrease from 100 DAF ( $404.19 \pm 26.33 \mu\text{g}/50 \text{g}$ ) to 175 DAF ( $287.06 \pm 22.03 \mu\text{g}/50 \text{g}$ ), while oil contents displayed a gradual increase from 25 DAF ( $2.50 \pm 0.13\%$ ) to 150 DAF ( $15.60 \pm 1.30\%$ ) and then an approximately 1% decline at 175 DAF (Fig. 1B), indicating terpenoid and oil accumulation at the early and middle-late stages during developing LGF, respectively. By GC-MS for the dynamic changes of the terpenoid and oil compositions, we characterized all the terpenoids (mainly including  $\alpha$ -pinene,  $\beta$ -cis-ocimene, eucalyptol, (Z)- $\beta$ -farnesene and  $\alpha$ -selinene) with higher ratios in early development of LGF (Supplementary Table S1). However, for the dynamic patterns of FA components, saturated (27.69% to 48.63%) and monounsaturated (19.80% to 37.72%) FAs showed the similar increasing trend with oil accumulation in developing LGF, but a down-trend for polyunsaturated (52.51% to 13.65%) FAs (Table 1). According to the temporal patterns of the contents and compositions for terpenoid and oil in developing LGF, 3 representative periods (50, 125 and 150 DAF) were characterized from 25–175 DAF for comparative deep transcriptomic analysis to better explore the molecular and metabolic regulatory mechanisms of the alternative carbon flux for terpenoid and oil accumulation in developing LGF.

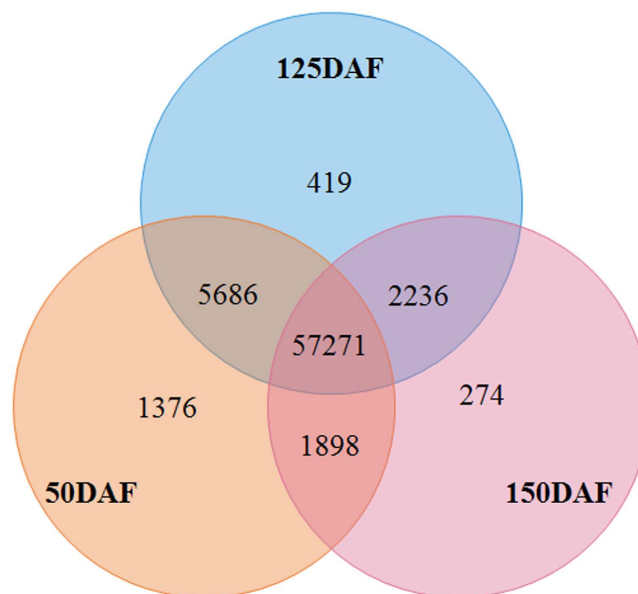
**Transcriptome sequencing and unigene assembly.** To clarify a global overview of the gene expressing profiles in developing LGF, a total of three cDNA libraries were constructed based on the temporal accumulation of terpenoid and oil, and then were respectively sequenced by the Illumina 2000. After filtering the data by discarding low-quality reads and adaptor sequences, 28,021,984, 29,037,610 and 24,298,844 clean reads (average length = 92bp) were obtained from 50, 125 and 150 DAF (Table 2), respectively. The three transcriptome databases can be accessed in the Short Read Archive (SRA) under accession number SRR1952817, SRR1955000 and SRR1956757. Additionally, all the clean reads obtained from LGF of 3 developing stages were together assembled by the Trinity software, resulting in 69,160 unigenes with mean length of 665.79 bp (Supplementary Table S2). From the Venn diagram analysis, 57,272 unigenes were identified to be expressed in the whole developing stage, while 1376, 419 and 274 unigenes were specifically expressed in 50, 125 and 150 DAF, respectively (Fig. 2). The largest number of specific unigenes expressed at 50 DAF indicated more active processes involving in the differentiation and development of LGF in early stage, which was evidenced by the dynamic changes of LGF size (Supplementary Fig. S1). All the obtained unigenes could provide a substantial biology background for the exploration on the molecular regulatory mechanism of carbon source partitioning between terpenoid and oil accumulation in developing LGF.

DAF	C10:0(%)	C12:0(%)	C16:0(%)	C16:1(%)	C18:0(%)	C18:1(%)	C18:2(%)	C18:3(%)	C20:4(%)	Saturated FAs(%)	Monoun-saturated FAs(%)	polyunsaturated FAs(%)
25	—	—	21.22 ± 1.31	—	4.47 ± 0.24	21.80 ± 1.58	29.75 ± 2.88	7.27 ± 0.74	15.49 ± 1.35	25.69	21.8	52.51
50	2.77 ± 0.20	1.59 ± 0.18	22.61 ± 1.85	—	5.12 ± 0.41	21.19 ± 2.31	28.70 ± 4.06	6.44 ± 0.47	11.58 ± 1.55	32.09	21.19	46.72
75	12.49 ± 1.38	5.22 ± 0.56	22.33 ± 1.99	—	3.05 ± 0.16	30.76 ± 4.33	20.17 ± 2.06	0.77 ± 0.09	5.21 ± 0.46	43.09	30.76	26.15
100	16.28 ± 1.43	8.24 ± 1.18	22.06 ± 2.38	1.08 ± 0.06	1.03 ± 0.14	37.14 ± 2.64	11.38 ± 1.40	0.74 ± 0.06	2.05 ± 0.14	47.61	38.22	14.17
125	10.35 ± 1.33	5.83 ± 0.48	30.22 ± 3.05	1.65 ± 0.08	0.95 ± 0.10	28.80 ± 5.47	11.44 ± 1.25	0.76 ± 0.08	—	47.35	42.45	10.2
150	11.08 ± 1.23	5.40 ± 0.36	31.15 ± 3.58	1.57 ± 0.12	0.88 ± 0.06	37.20 ± 3.26	12.07 ± 0.72	0.65 ± 0.09	—	48.51	38.77	12.72
175	11.83 ± 1.00	5.80 ± 0.48	30.13 ± 2.73	1.54 ± 0.09	0.87 ± 0.12	36.18 ± 2.76	12.92 ± 1.23	0.73 ± 0.09	—	48.63	37.72	13.65

**Table 1.** The percentage of FAs in developing LGE.

DAF	Trim reads	Total length(bp)	Average length(bp)	Map reads	Map reads%
50	28,021,984	2,598,396,941	92.73	24,335,597	86.84%
125	29,037,610	2,681,376,343	92.34	25,707,321	88.53%
150	24,298,844	2,233,238,833	91.91	21,287,874	87.61%

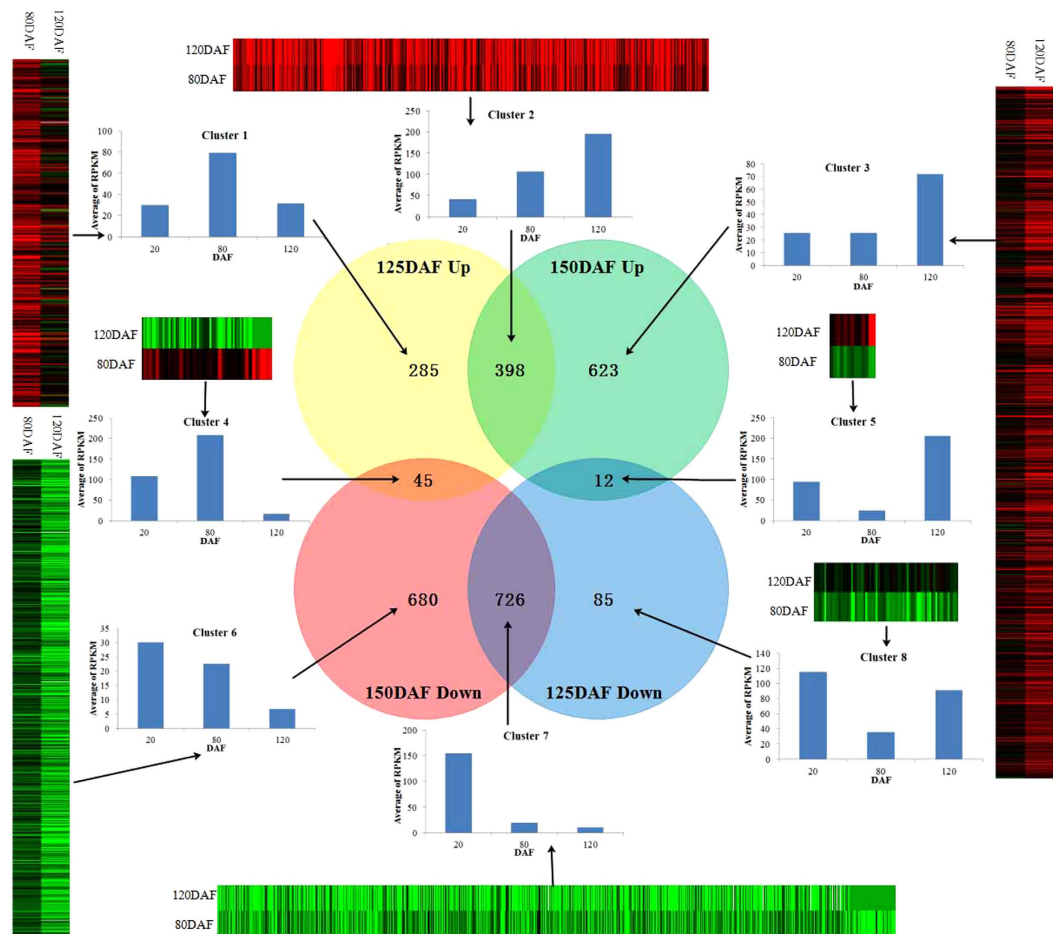
**Table 2.** Reads data statistics.



**Figure 2.** The unigene distribution at different development period.

**Transcript patterns for enzymes involved specially in TBP and OBP.** All the *L. glauca* unigenes from three developmental stages were annotated by using the BLAST algorithm with an E-value  $< 10^{-5}$  and protein identity  $> 30\%$  in public databases. Of the 69,160 unigenes, 14,942 (21.60%), 15,437 (22.32%), 15,529 (22.45%), 13,798 (19.95%), 14,689 (21.24%), 8248 (11.93%) unigenes showed significant similarities to known proteins in NR, SWISS-PROT, TREMBL, CDD, PFAM and COG database, respectively (Supplementary Table S3). Additionally, the species distribution in NR database showed that 6953 (46.53%) unigenes were high similar to the sequences from *Vitis vinifera*, followed by *Populus trichocarpa* (11.67%), *R. communis* (9.56%) and *Glycine max* (5.17%) (Supplementary Fig. 2). To further identify the interactions of all the annotated unigenes, we carried out GO functional enrichment and KEGG pathway analysis. The resulting 20,043 (28.98%) unigenes were assigned to the three main GO categories and 66 subcategories, among which ‘Cellular process’, ‘cell part’ and ‘binding’ were dominant subcategories in ‘biological process’, ‘cellular component’ and ‘molecular function categories’, respectively (Supplementary Fig. S3). Also, a total of 5,820 (8.42%) unigenes were assigned to 298 KEGG pathways and 754 kinds of



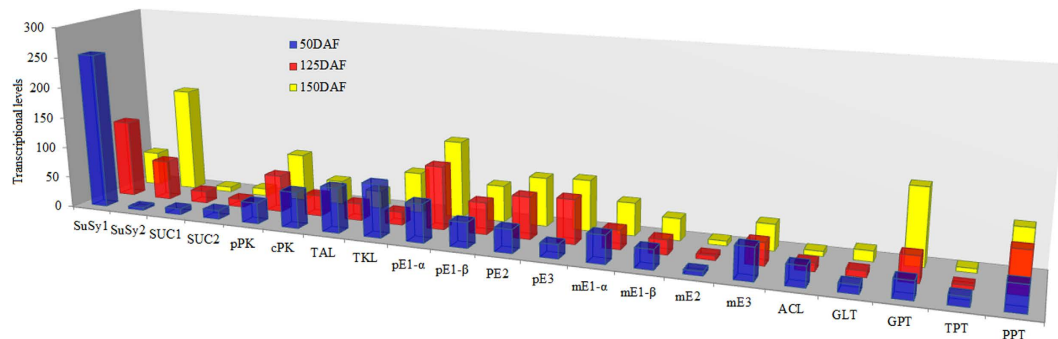


**Figure 3.** The number and distribution of up-regulated or down-regulated unigenes in developing LGF. The detailed sequences were showed in Tables S4.

enzymes (Supplementary Table S3). Notably, a total of 174, 71, 81 and 155 unigenes were implicated in glycolysis, PPP, EBP and OBP, respectively, indicating an intricate regulatory mechanism of carbon flux for oil and terpenoid biosynthesis in developing LGF.

To fully explore the differential unigene expressions in developing LGF, the clean reads from the three developing stages of LGF were mapped to our unigene database, and resulting 24.34 (86.84%), 25.71 (88.53%) and 21.29 (87.61%) million reads were perfect matched (Table 2). The normalization of gene expression data were performed by using multiple correction methods<sup>44</sup> (Supplementary Fig. S4), and the differentially-expressed genes between the two LGF samples from different developmental stages were characterized using DESeq<sup>45</sup> with  $p$ -value cut-off of 0.01 and using the BH method for multiple testing correction<sup>46</sup>, resulting in a total of 2854 unigenes identified with differential expression in developing LGF (Supplementary Table S4). Notably, we characterized 370 and 1303 unigenes to be specific for 50:125 DAF and 50:150 DAF respectively (50 DAF as normalization), but 1096 unigenes shared in the both stages (Supplementary Table S4). According to their expression patterns, all the differentially-expressed unigenes were sorted into 8 clusters, including three up-regulated clusters (cluster 1, 2 and 3), three down-regulated clusters (cluster 6, 7 and 8) and two mixed clusters (cluster 4 and 5) (Fig. 3). Importantly, we characterized that the differentially expressed unigenes involving in glycolysis and PPP distributed in various clusters (cluster 1, 2, 3, 7 and 8), whereas the TBP-specific unigenes were mainly focus on cluster 6 and 7, and OBP-specific unigenes on cluster 1 and 2. Therefore, the differential gene expression for the accumulation of terpenoid and oil may be specifically regulated in response to the circadian clock of developing LGF.

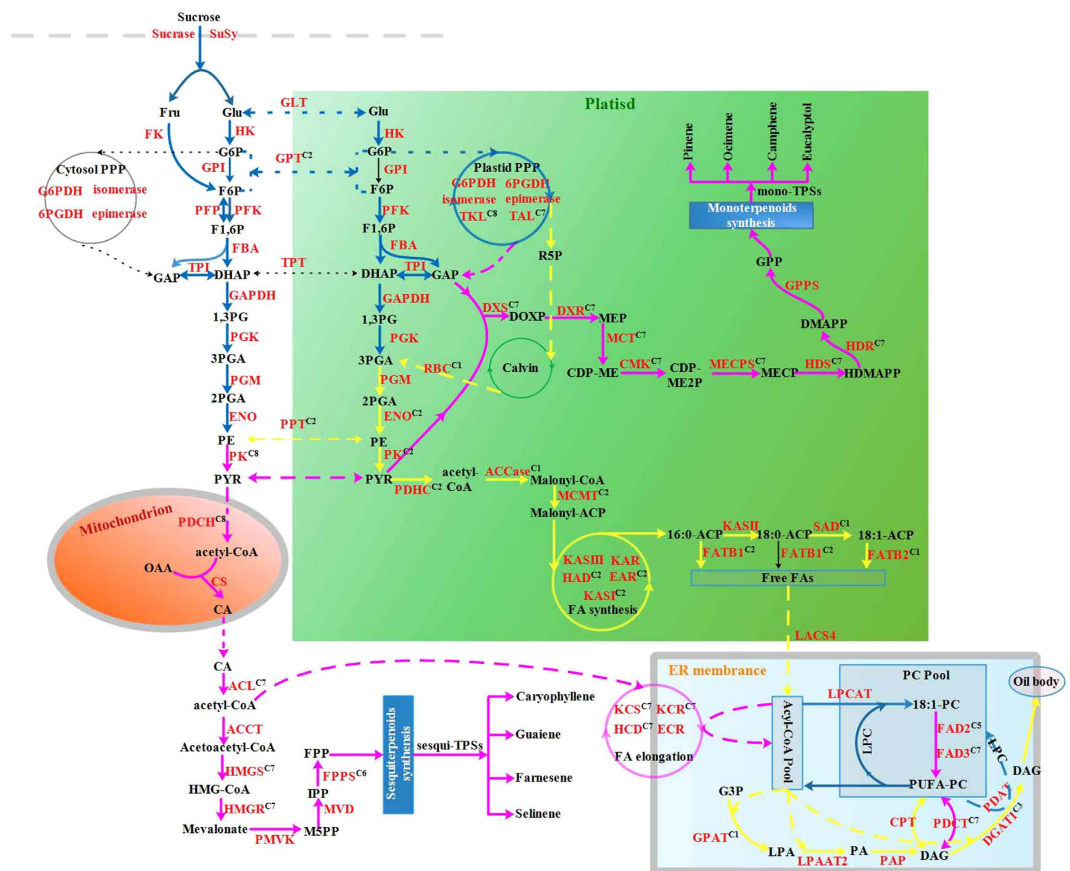
**Temporal allocation of carbon source for TBP and OBP.** Sucrose is known as the most source of carbon required for terpenoid and oil biosynthesis in plants. To deeply understand the allocation of available carbon source for terpenoid and oil biosynthesis in developing LGF, the differential expression profiles of genes for key enzymes, providing the precursors for TBP and OBP, were concretely analyzed (Supplementary Tables S5, S6 and S7). We identified 2 unigenes for two SuSy isozymes with differential profiles (down in *SuSy1* and up in *SuSy2*) in developing LGF, but their expression levels were higher



**Figure 4.** The transcriptional levels for the enzymes involved in the generation of G3P and acetyl-CoA.

than those for two SUC isozymes (SUC1 and 2) with stable expression (Fig. 4), indicating that SuSy, as an important sucrose-cleaving enzyme, play a fundamental role in the supply of hexoses for the terpenoid and oil biosynthetic demand in developing LGF. Indeed, our qRT-PCR analysis attested the fact that *SuSy1* expression with a gradual decrease was observed in 7 developing stages of LGF, while *SuSy2* was gradually up-regulated (Supplementary Fig. S5). This striking difference of transcript level between two SuSy isozymes was entirely consistent with the different accumulative patterns of terpenoid and oil in developing LGF (Fig. 1), indicating that the expressions of *SuSy1* and 2 specifically responded to carbohydrate availability for terpenoid and oil biosynthesis in developing LGF. Additionally, we noticed the differential expression profiles of some crucial enzymes involved in carbon flux from the hexoses into TBP or OBP. Our data showed that the expressions of cytosolic PK and ACL, mitochondrial PDHC, plastidial TKL and TAL were all significantly up-regulated in early development of LGF, but a significant higher expression for plastidial PK and PDHC in middle-late development of LGF (Figs 4 and 5). These results revealed that a main hexose flux via cytosolic glycolysis or plastidial PPP is to provide acetyl-CoA or G3P respectively for terpenoid biosynthesis in early development of LGF, however a greater proportion of hexose to acetyl-CoA flux via plastidial glycolysis is required for oil biosynthesis in middle-late development of LGF. It was interesting to note that the genes encoding ribulose biphosphate carboxylase (RBC), fixation of CO<sub>2</sub> and ribulose 5-phosphate to 3-phospho-D-glycerate (3PGA)<sup>47</sup>, displayed significant higher expression at middle development of LGF (Supplementary Table S5). Combining with 3-fold higher expression of ENO in plastid than in cytosol (Supplementary Table S5), our findings indicated that Calvin pathway is crucial for the supplementation of carbon source for oil biosynthesis in middle development of LGF. Intriguingly, we also found that among the four glycolytic transporters (GLT, GPT, TPT and PPT), only both PPT and GPT showed a significant abundance of transcript in middle-late development of LGF (Fig. 4), as was in accordance with our qRT-PCR results that the expressions of PPT and GPT were up-regulated in 7 developing stages of LGF (Supplementary Fig. S5). Thus, a higher capacity of PPT and GPT provide glycolytic substrates (hexose phosphate, G6P and F6P) and intermediates (triose phosphate, PEP) from the cytosol to plastid for oil biosynthesis in developing LGF. Together, our data highlight that the partitioning of available carbon source (sucrose) is tightly regulated in response of LGF to different developing stages, leading to temporal allocation of sucrose flux to the important precursor (acetyl-CoA or G3P) required for TBP and OBP.

**A higher carbon flux into TBP in early developing LGF.** To explore carbon flux from acetyl-CoA or G3P into TBP for terpenoid accumulation in developing LGF, the differential expression profiles of genes associated with terpenoid biosynthesis were analyzed, and our data revealed that most genes encoding the key enzymes (DXS, HMGCR, GPPS, FPPS, etc.) for terpenoid backbone biosynthesis (MVA and MEP pathway), and the genes for mono-TPSs and sesqui-TPSs, all displayed a higher transcriptional level in early development of LGF (Supplementary Table S6). This indicated that active biosynthesis of building blocks contribute to a rapid accumulation of terpenoid in early development of LGF, which is consistent with our findings on the temporal alternations of sucrose flux and terpenoid content at early phase (Fig. 1). These data reinforce and extend our previous conclusion that LGF are implicated actively in MVA and MEP pathway<sup>43</sup>. Surprisingly, a total of 6 and 4 unigenes showed high homology with the same enzyme of myrcene synthase (MS) and  $\alpha$ -humulene synthase ( $\alpha$ -HS) by which monoterpene (C<sub>10</sub>H<sub>16</sub>) and sesquiterpene (C<sub>15</sub>H<sub>24</sub>) hydrocarbons were synthesized, respectively. The phylogenetic analysis revealed that both MS and  $\alpha$ -HS genes are clustered into 4 classes (Supplementary Fig. S6), showing an involvement of MS or  $\alpha$ -HS in at least 4 different synthesis of monoterpene or sesquiterpene hydrocarbons, which can be confirmed by our detective results of the main four monoterpene hydrocarbons ( $\alpha$ -pinene,  $\beta$ -pinene,  $\beta$ -cis-ocimene and camphene) and five sesquiterpene hydrocarbons ( $\beta$ -farnesene, cedrene-V6,  $\alpha$ -selinene,  $\beta$ -selinene and farnesene) (Supplementary Table S1). Thus, the observed high sequence identity among MS and  $\alpha$ -HS family members indicates the rapid evolution of a species-specific paralogous gene cluster in *L. glauca*. Together, our results of the higher expression levels

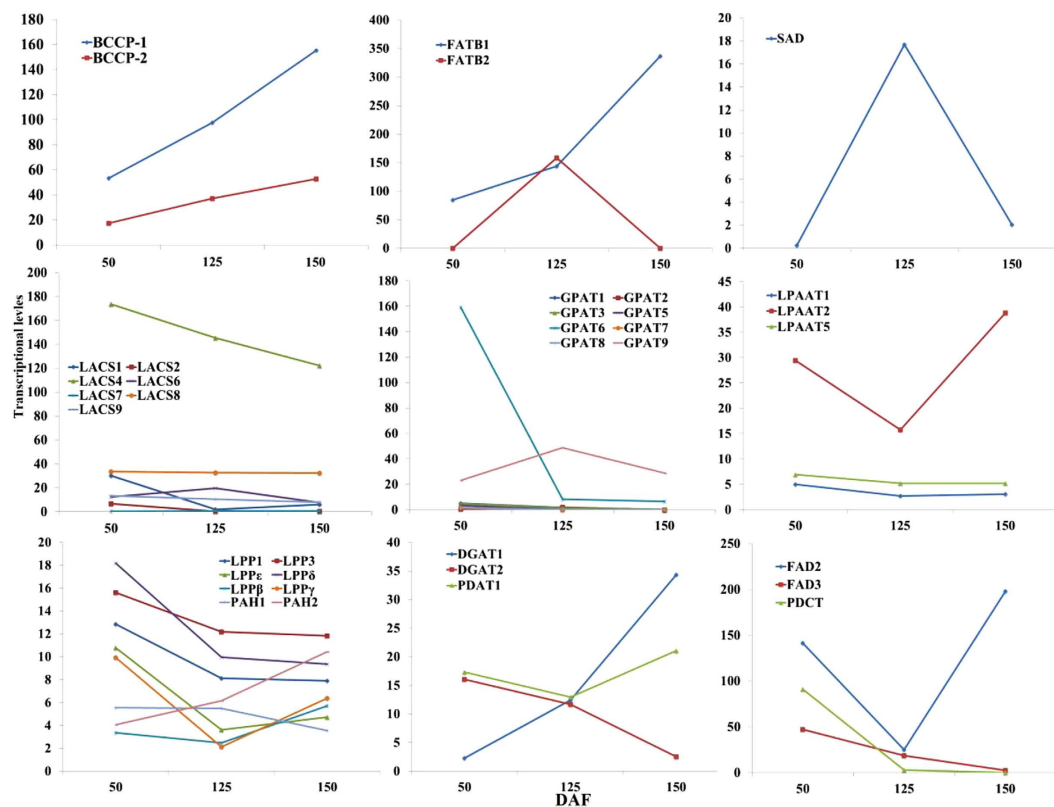


**Figure 5. Compartmentalization of central carbon metabolism in developing LGF.** Pink arrows show the main flux in early development. Yellow arrows show the main flux in middle-late development. Blue arrows show the main flux in the whole development. The superscripts show the results of differential-expressed cluster.

for the enzymes in TBP reveal a higher carbon allocation ratio from sucrose to terpenoid accumulation in early developing LGF.

**A higher carbon flux into OBP in middle-late developing LGF.** In middle-late development of LGF, a greater proportion of hexose to acetyl-CoA flux is used for oil accumulation (Fig. 5). Indeed, the majority of our annotated genes associated with *de novo* FA synthesis were identified with a higher transcriptional level in middle-late development of LGF (Fig. 5 and Supplementary Table S7). It is, therefore, concluded that strongly increased FA synthesis, together with plastid carbon supply, is crucial for the higher oil accumulation in developing LGF. Notably, we characterized 2 genes for two BCCP isoforms (BCCP1 and BCCP2), but the abundance of *BCCP1* expression was 3-fold higher than *BCCP2* (Fig. 6), indicating a more important contribution of BCCP1 subunit to control the carbon flux of acetyl-CoA to malonyl-CoA regulated by ACCase activity for *de novo* FA biosynthesis in developing LGF. The produced FA-ACP (16:0-, 18:0- and 18:1-ACP) must be hydrolyzed by FATB or FATA to release free FAs (saturated and monounsaturated) that are exported from the plastid required for TAG biosynthesis<sup>26</sup>. Unexpectedly, we only identified four unigenes encoding two FATB isozymes (FATB1 and FATB2) but no unigenes similar to FATA (Supplementary Table S7). Moreover, the analysis of transcriptional profiles revealed a gradual increased expression for *FATB1* in developing LGF, and a notably up-regulated expression for *FATB2* only at 125 DAF (comparable to expression pattern of *SAD6*), as was strongly supported by our qRT-PCR results of their expression alterations in 7 developing stages (Fig. 6 and Supplementary Fig. S5). Thus, we concluded that *FATB1* and *FATB2* are implicated respectively in regulation the flux of saturated (C16:0 and C18:0) and monounsaturated (C18:1) FAs from plastid to cytosol for TAG biosynthesis in developing LGF, which was also strengthened by our observation on palmitic (C16:0) and oleic acid (C18:1) as the major FA compositions (Table 1).

Considering the importance of the FA flux from cytosol to ER for FA elongation or esterification in developing LGF, we analyzed the differential expression profiles of the annotated 7 genes of LACS isozymes, and identified 5–10 fold higher expression level for *LACS4* than for the other *LACS* genes (Fig. 6), showing a preferred regulation for the FA flux to ER for oil accumulation by *LACS4* in developing



**Figure 6.** Temporal profile of transcriptional levels for enzymes involved in FA synthesis and TAG assembly.

LGF. As for the biosynthesis of long chain FAs (more than 18 carbon), a series of FA elongases displayed higher transcriptional levels in early development of LGF (Supplementary Table S7), as were consistent with the high ratio of long-chain FA detected in early development of LGF (Table 1). The assembly of TAG from G3P and acyl-CoAs involves four enzymatic steps sequentially by GPAT, LPAT, PAP, and DGAT (or PDAT)<sup>29–35</sup>. As expected, we characterized 8 GPAT isoforms, but only *GPAT9* displayed a significant transcript in middle-late development of LGF (Fig. 6), indicating *GPAT9* as the key enzyme for the initiation of TAG assembly. For the second step of TAG assembly by LPAAT and PAP, indeed, the genes for 3 and 8 isoforms of LPAAT and PAP were respectively identified in developing LGF, where *LPAT2* expression was 3–7 folds higher than *LPAAT1* and *LPAAT5*, whereas all of *PAP* showed no significant expression alteration (Fig. 6). These results indicate that *LPAAT2* is as one major enzyme responsible for generation of PA for oil accumulation, and a coordinated expression of multiple *PAP* isozymes are contributing to DAG synthesis. As for the eventual TAG assembly, we characterized the genes high homology with *DGAT1* (up-regulated), *DGAT2* (down-regulated) and *PDAT1* (stable expression) in developing LGF, which was confirmed by their corresponding changes detected in 7 developing stages by qRT-PCR (Supplementary Fig. S5). Our data revealed an involvement of the overlapping function (*DGAT1* and *PDAT1*) in TAG production for oil accumulation of developing LGF, which was comparable to our previous results in developing seed kernels of Siberian apricot<sup>43</sup>. In addition, we characterized *PDCT* gene with up-regulated only in early development of LGF, and *CPT* with stable expression in developing LGF (Fig. 6). Integrated with the finding that the significant up-regulation of *FAD2/3* was parallel to the high percentage of polyunsaturated FAs in early development of LGF (Fig. 6 and Table 1), we conclude that a high expression of *PDCT* conduce to actively interconvert acyl residues between DAG and PC for further desaturation by *FAD2/3*, resulting in the accumulation of polyunsaturated FAs by PC-derived DAG synthesis in early development of LGF. In summary, our investigations could provide insight into which isoforms of these large gene families are responsible specifically for oil biosynthesis in developing LGF, and their high transcript levels evidence a higher carbon source partitioning from sucrose into OBP in middle-late developing LGF.

**Specifically transcriptional regulation for TBP and OBP.** The TFs related to oil accumulation has been extensively studied, but little is known about TFs involving in regulation of terpenoid accumulation. By the differential expression analysis, we characterized 34 unigenes encoding 31 TFs belonging to 9 families, and the 10, 13 and 8 TFs with high expression were sorted into cluster-1, -2 and -3 in early, middle-late and whole development of LGF, respectively (Supplementary Table S8). Based on



accumulative patterns of terpenoid and oil in developing LGF (Fig. 1), we suspected that the TFs in cluster-1 or -2 respectively regulated the terpenoid or oil biosynthesis in developing LGF, and the TFs of cluster-3 shared in both terpenoid and oil biosynthesis. In cluster-1, except for MYC2, the remaining 10 TFs (TT2, MYB1R1, CP19, NFYA9, NFYA7, MADS6, and bHLH93, 96 and 120) were not previously reported, and therefore they may be novel or specific for terpenoid biosynthesis of developing LGF. In cluster-2, we identified ABI3 and LEC1 as known TFs for oil biosynthesis, but the other 11 TFs (MYB90, NFYC2, bHLH82 and 148, WRKY23 and 65, and ERF5, 18, 26, 76, and 114) were novel or specific for developing LGF. It was noteworthy that the known WR1 of cluster-3, displaying high expression levels in developing LGF, might play a pivotal role in expression regulation for some sugar-responsive genes with the consequence of alternative carbon partitioning from sucrose to terpenoid and oil in developing LGF. Additionally, the other novel 7 TFs in cluster-3 (GLK1, APL, MYB44, TCP2, RAP2-1 and 2–4, and ERF82) displayed the co-regulation for terpenoid and oil accumulation of developing LGF. Together, the above TFs with various expression patterns revealed the complex dynamic regulation mechanisms of carbon partitioning for terpenoid and oil accumulation in developing LGF.

## Discussion

Despite previous phytochemical studies of *L. glauca* has reported on the abundance of terpenoid and oil in fruits<sup>1,2,8</sup>, little is known about the molecular regulatory mechanisms of terpenoid and oil accumulation in developing LGF. Here, we firstly determined the contents and compositions of terpenoid and oil in 7 developing stages of LGF (25–175 DAF), and characterized the accumulation of terpenoid and oil mainly at the early and middle-late development of LGF, respectively. To explore the regulatory mechanisms for the differential accumulation patterns between terpenoid and oil in developing LGF, we selected 3 crucial periods (50, 125 and 150 DAF) for comparative deep transcriptomic analysis for the first time, and identified some key genes implicated specifically in terpenoid and oil accumulation of developing LGF. It is well known that the carbon flux from sucrose to the precursors (G3P and acetyl-CoA) in plants for terpenoid or oil biosynthesis is regulated by some key metabolic enzymes, such as SuSy, HK, PFK, PK, PDHC, ACL, G6PDH, TKL and TAL<sup>10–12,16,17</sup>. By the combining analysis of differential expression profiling and qRT-PCR, we characterized that SuSy as the major enzyme initial the carbon metabolic flow for terpenoid and oil biosynthesis. Impressively, the cytosolic PK and ACL, mitochondrial PDHC, plastidial TKL and TAL, providing the G3P and acetyl-CoA for terpenoid biosynthesis<sup>10,17</sup>, all displayed up-regulated expression in early development of LGF, while the up-regulated expression was showed only for plastidial PK and PDHC in middle-late development, providing the acetyl-CoA for oil biosynthesis<sup>16</sup>. Therefore, combining with the temporal accumulation patterns of terpenoid and oil in 7 developing stages (Fig. 1), these data prompt us to conclude that LGF may allocate more carbon sources from sucrose into TBP in early development, whereas a higher assimilated carbon flux into OBP in middle-late development. Also consistent with this conclusion, all the genes involved in TBP or OBP showed up-regulated expressions in early or middle-late development of LGF, respectively (Fig. 5). Simultaneously, we characterized the involvement of 8 TPSs (4 MSs and 4  $\alpha$ -Hs) specifically in terpenoid biosynthesis, and the more contribution of BCCP2, FATB, LACS4, GPAT9, LPAAT2, DGAT1 and PDAT1 for oil biosynthesis in developing LGF. In addition to the above metabolic enzymes, 31 TFs with the significantly differential expressions were also identified to be highly related to developing LGF, including the known MYC2 for terpenoid biosynthesis<sup>48</sup>, ABI3 and LEC1 for oil biosynthesis<sup>49,50</sup>, and WR1 involved in the activation of a subset of sugar-responsive genes for both terpenoid or oil accumulation<sup>51</sup>. In conclusion, our transcriptome sequencing and dynamic analysis revealed the complex regulatory mechanisms of carbon source partitioning between terpenoid and oil accumulation in developing LGF.

## Methods

**Collection of plant materials.** *L. glauca* is widely distributed in China, so it has not been listed as an endangered or protected species and does not require approval. In this study, the different developmental stages of LGF were collected from Jigong Mountain (E114°06', N32°125') Natural Reserve of Henan Province. Flowers with the same anthesis were marked, and then seeds were harvested at 25 DAF (immature stage), 50 DAF, 75 DAF, 100 DAF, 120 DAF, 150 DAF and 175 DAF (fully matured stage) respectively. The samples were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until use.

**Extraction of terpenoid and oil from developing LGK.** About 50 g of the fresh LGF from each sample (approximate 15 samples per accession) were powdered and subjected to hygro-distillation by using a modified Clevenger-type apparatus for five hours. The essential oil were collected, measured and dried over with anhydrous  $\text{Na}_2\text{O}_4$  and then stored in  $-20^{\circ}\text{C}$  until use. Oil was extracted with petroleum ether using a Soxhlet apparatus at  $45\text{--}50^{\circ}\text{C}$  for 6–8 h until the extraction was completed. The oil content was determined as the difference in weight of the dried fruit sample before and after the extraction. The determination was performed in triplicate.

**GC-MS analysis.** Terpenoid sample analyses were performed on an Agilent 7890Agas chromatograph, equipped with an Agilent 5975C quadrupole mass spectrometer and a HP-INNOWAX capillary column (60 m  $\times$  0.25 mm id, 0.25  $\mu\text{m}$  film thickness J&W Scientific, Folsom, CA). The temperature program was as follows: from  $50^{\circ}\text{C}$  (1 min hold) to  $220^{\circ}\text{C}$  at  $3^{\circ}\text{C}/\text{min}$  and then held at  $220^{\circ}\text{C}$  for 5 min. The carrier

gas was Helium was at 1 mL/min. The GC inlet was set in a splitting mode with split ratio 1:20 and at 230 °C and 1.0 µL of diluted samples (1/10, v/v, in hexane) were injected. The quadrupole MS operating parameters: interface temperature 270 °C; electron impact ionization at 70 eV with scan mass range of 29–450 m/z. As for FA compositions, oil were trans-esterified under standard conditions employing a 5.5:1 molar ratio of methanol to oil using 1 wt% potassium hydroxide as a catalyst at 60–65 °C for 1 h. This was followed by a conventional work-up consisting of separation of phases, washing the resulting methyl esters with water until the water was neutral and drying with magnesium sulfate. The hexane (1 µL) extract was injected into a highly polar HP Innowax capillary column with a length of 30 m (inner diameter 0.32 mm, film thickness 0.5 µm, split 1:20). An Agilent 6890 (CA, USA) equipped with a flame ionization detector was used. The injector and detector temperatures were 250 and 280 °C, respectively. The oven temperature was programmed from 190 °C holding at 3 min to 240 °C at the rate of 15 °C/min for 17 min. The carrier gas was high-purity hydrogen. The peaks were identified by comparing their retention time with that of the known standards, carried out under similar separation conditions. Peak integration was performed by applying HP3398A software. The above experiments were performed in triplicate.

**Total RNA extraction and transcriptome sequencing.** Based on the detective results of LGF oil contents and FA compositions, the experimental LGF from the 3 crucial periods (50, 120 and 150 DAF) were selected as materials for transcriptome sequencing. The equal weight of 5–10 biological samples from every developmental stage were mixed, and then total RNA was extracted from the mixture using RNeasy Plant Mini Kits (Qiagen, Inc., Valencia, CA, USA) according to the manufacturer's protocol. Extracted RNA was qualified and quantified using a Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and all the samples showed a 260/280 nm ratio from 1.9 to 2.1. The equivalent RNA samples for transcriptome analysis were prepared using Illumina's kit following manufacturer's protocol (Illumina, San Diego, CA, USA). The cDNA library was sequenced on the Illumina sequencing platform (HiSeq™ 2000), and two technical repetitions were performed. After removal of the adapter sequences, the low quality sequences (reads with ambiguous bases 'N') and reads with more than 10% Q < 20 bases, all the clean reads were together assembled with the Trinity program, and the Trinities were clustered using TGICL tools into unigenes<sup>52</sup>.

**Sequence annotation.** To understand their functions, the LGF unigenes were annotated using BLASTX alignment with an E-value cut-off of 10<sup>-5</sup> against the following protein databases: NR, SWISS-PROT, TrEMBL, CDD, PFAM and COG database. Functional annotation by GO terms (<http://www.geneontology.org>) was analyzed by Blast2Go software. The KEGG pathways annotation was performed using BLAST all against Kyoto Encyclopedia of Genes and Genomes database.

**Differential expression analysis of unigenes.** According to the previous identification of appropriate reference genes for normalizing transcript expression in developing fruits of *Litsea cubeba*, belonging to the same family (Lauraceae) with our experimental materials<sup>53</sup>, 9 reference genes with the pairwise variation of 0.08 were used for normalizing our transcriptome data (Supplementary Table S9). The normalization of gene expression data were performed by using multiple correction methods<sup>44</sup>, and the distribution of RPKMs before and after correction were shown in Supplementary Fig. S4. Subsequently, the data from the reference gene-based correction on the RPKM were used for differential expression analysis by using DESeq<sup>45</sup> with p-value cut-off of 0.01 and using the BH method for multiple testing correction<sup>46</sup>.

**qRT-PCR assay.** Total RNA was extracted as the description for the cDNA library preparation and was reverse transcribed using the Reverse transcription System (Promega). The amplification primers were designed using PrimerQuest (<http://www.idtdna.com/PrimerQuest/Home/Index>) software with melting temperatures at 62 °C, and the absence of secondary structures was verified by the UNAFold program (<http://eu.idtdna.com/UNAFold>) according to D'haene *et al.*<sup>54</sup>. Also, large subunit ribosomal protein L32e and ubiquitin-conjugating enzyme genes were used as internal controls. The primers list in Supplementary Table S10. The qRT-PCR was performed using the SYBR Premix Ex Taq Kit (TaKaRa) according to the manufacturer's protocol. Negative controls consisting of nuclease-free water instead of template, and reverse transcriptase controls prepared by substituting reverse transcriptase for nuclease-free water in the cDNA synthesis step were included in all analyses for each primer pair. Three technical repetitions were performed for qRT-PCR.

## References

1. Qi, J., Xiong, B., Ju, Y. X., Hao, Q. Z. & Zhang, Z. X. Study on fruit growth regularity and lipid accumulation of *Lindera glauca*. *Chinese Agr Sci Bull* **31**, 29–33 (2014).
2. Wang, Y., Gao, X., Yu, X., Cheng, S. & Kong, L. Study on the resource and its utilizations of *Lindera glauca* in China. *Henan Sci* **12**, 331–334 (1994).
3. Huh, G.-W. *et al.* Flavonoids from *Lindera glauca* blume as low-density lipoprotein oxidation inhibitors. *Nat Prod Res* **28**, 831–834 (2014).
4. Seki, K., Sasaki, T., Haga, K. & Kaneko, R. Two methoxybutanolides from *Lindera glauca*. *Phytochemistry* **36**, 949–951 (1994).

5. Huh, G. W., Shrestha, S., Ahn, E. M. & Baek, N. I. Sterols from *Lindera glauca* Blume stem wood. *J Appl Biol Chem* **54**, 309–312 (2011).
6. Huh, G.-W. *et al.* New diarylpropanoids from *Lindera glauca* Bl. heartwood. *Holzforschung* **66**, 585–590 (2012).
7. Wang, R., Tang, S., Zhai, H. & Duan, H. Studies on anti-tumor metastatic constituents from *Lindera glauca*. *China J Chinese Mater Med* **36**, 1032–1036 (2011).
8. Sun, H. L., Wang, J. X., Gu, X. Z. & Kang, W. Y. Analysis of volatile compounds from leaves and fruits of *Lindera glauca*. *Chinese J Exp Tradit Med Formulae* **7**, 033 (2011).
9. Liu, L., Gu, J. & Chen, J. Studies on the chemical constituents of the leaf of *Lindera glauca* (Sieb Et Zucc) Bl and their uses. *Jiangxi Sci* **1**, 007 (1992).
10. Kruger, N. J. & von Schaewen, A. The oxidative pentose phosphate pathway: structure and organisation. *Curr Opin in Plant Biol* **6**, 236–246 (2003).
11. Plaxton, W. C. The organization and regulation of plant glycolysis. *Annu Rev Plant Biol* **47**, 185–214 (1996).
12. Jin, Y., Ni, D.-A. & Ruan, Y.-L. Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. *Plant Cell* **21**, 2072–2089 (2009).
13. Rawsthorne, S. Carbon flux and fatty acid synthesis in plants. *Prog in Lipid Res* **41**, 182–196 (2002).
14. Weber, A. P., Schwacke, R. & Flügge, U.-I. Solute transporters of the plastid envelope membrane. *Annu Rev Plant Biol* **56**, 133–164 (2005).
15. Knappe, S. *et al.* Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis-AtPPT1 may be involved in the provision of signals for correct mesophyll development. *Plant J* **36**, 411–420 (2003).
16. Ke, J. *et al.* The role of pyruvate dehydrogenase and acetyl-coenzyme A synthetase in fatty acid synthesis in developing Arabidopsis seeds. *Plant Physiol* **123**, 497–508 (2000).
17. Fatland, B. L. *et al.* Molecular characterization of a heteromeric ATP-citrate lyase that generates cytosolic acetyl-coenzyme A in Arabidopsis. *Plant Physiol* **130**, 740–756 (2002).
18. Estévez, J. M., Cantero, A., Reindl, A., Reichler, S. & León, P. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J Biol Chem* **276**, 22901–22909 (2001).
19. Friesen, J. A. & Rodwell, V. W. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol* **5**, 248 (2004).
20. Sawai, S. & Saito, K. Triterpenoid biosynthesis and engineering in plants. *Front Plant Sci* **2**, 25 (2011).
21. Cheng, A. X. *et al.* Plant terpenoids: biosynthesis and ecological functions. *J Integ Plant Biol* **49**, 179–186 (2007).
22. Degenhardt, J., Köllner, T. G. & Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **70**, 1621–1637 (2009).
23. Aubourg, S., Lechary, A. & Bohlmann, J. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol Genet Genomics* **267**, 730–745 (2002).
24. Nikolau, B. J., Ohlrogge, J. B. & Wurtele, E. S. Plant biotin-containing carboxylases. *Arch Biochem Biophys* **414**, 211–222 (2003).
25. Li-Beisson, Y. *et al.* Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists* **11** (2013).
26. Bates, P. D., Stymne, S. & Ohlrogge, J. Biochemical pathways in seed oil synthesis. *Curr Opin Plant Biol* **16**, 358–364 (2013).
27. Koo, A. J., Ohlrogge, J. B. & Pollard, M. On the export of fatty acids from the chloroplast. *J Biol Chem* **279**, 16101–16110 (2004).
28. Baud, S. & Lepiniec, L. Physiological and developmental regulation of seed oil production. *Prog Lipid Res* **49**, 235–249 (2010).
29. Barron, E. J. & Stumpf, P. Fat metabolism in higher plants XIX. The biosynthesis of triglycerides by avocado-mesocarp enzymes. *BBA* **60**, 329–337 (1962).
30. Maisonneuve, S., Bessoule, J. J., Lessire, R., Delseny, M. & Roscoe, T. J. Expression of rapeseed microsomal lysophosphatidic acid acyltransferase isozymes enhances seed oil content in Arabidopsis. *Plant Physiol* **152**, 670–684 (2010).
31. Eastmond, P. J. *et al.* PHOSPHATIDIC ACID PHOSPHOHYDROLASE1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in Arabidopsis. *Plant Cell* **22**, 2796–2811 (2010).
32. Kim, H. U., Li, Y. & Huang, A. H. Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in Arabidopsis. *Plant Cell* **17**, 1073–1089 (2005).
33. Saha, S., Enugutti, B., Rajakumari, S. & Rajasekharan, R. Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol* **141**, 1533–1543 (2006).
34. Shockey, J. M. *et al.* Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell* **18**, 2294–2313 (2006).
35. Zhang, M., Fan, J., Taylor, D. C. & Ohlrogge, J. B. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* **21**, 3885–3901 (2009).
36. Mhaske, V., Beldjilali, K., Ohlrogge, J. & Pollard, M. Isolation and characterization of an *Arabidopsis thaliana* knockout line for phospholipid: diacylglycerol transacylase gene (At5g13640). *Plant Physiol Bioch* **43**, 413–417 (2005).
37. Lu, C., Xin, Z., Ren, Z. & Miquel, M. An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *P Natl Acad Sci USA* **106**, 18837–18842 (2009).
38. Vogel, G. Cholinephosphotransferase and diacylglycerol acyltransferase. *Plant Physiol* **110**, 923–931 (1996).
39. Yang, L. *et al.* Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis. *PLoS One* **8**, e80464 (2013).
40. Yang, Y., Xu, M., Luo, Q., Wang, J. & Li, H. *De novo* transcriptome analysis of *Liriodendron chinense* petals and leaves by Illumina sequencing. *Gene* **534**, 155–162 (2014).
41. Bourgis, F. *et al.* Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *P Natl Acad Sci USA* **108**, 12527–12532 (2011).
42. Troncoso-Ponce, M. A. *et al.* Comparative deep transcriptional profiling of four developing oilseeds. *Plant J* **68**, 1014–1027 (2011).
43. Niu, J. *et al.* Transcriptome analysis of distinct *Lindera glauca* tissues revealed the differences in the unigenes related to terpenoid biosynthesis. *Gene* **559**, 22–30 (2015).
44. Darbani, B., Stewart, C. N., Noeparvar, S. & Borg, S. Correction of gene expression data: performance-dependency on inter-replicate and inter-treatment biases. *J Biotechnol* **188**, 100–109 (2014).
45. Anders, S. & Huber, W. Differential expression of RNA-Seq data at the gene level—the DESeq package. *Heidelberg, Germany: European Molecular Biology Laboratory (EMBL)* (2012).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc. Series B* **57**, 289–300 (1995).
47. Bowes, G., Ogren, W. & Hageman, R. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem Bioph Res Co* **45**, 716–722 (1971).
48. Zhang, H. *et al.* The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J* **67**, 61–71 (2011).
49. Giraudat, J. *et al.* Isolation of the Arabidopsis *ABI3* gene by positional cloning. *Plant Cell* **4**, 1251–1261 (1992).
50. Mu, J. *et al.* LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol* **148**, 1042–1054 (2008).

51. Masaki, T. *et al.* Activator of Spomin: LUC1/WRINKLED1 of *Arabidopsis thaliana* transactivates sugar-inducible promoters. *Plant Cell Physiol* **46**, 547–556 (2005).
52. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652 (2011).
53. Lin, L., Han, X., Chen, Y., Wu, Q. & Wang, Y. Identification of appropriate reference genes for normalizing transcript expression by quantitative real-time PCR in *Litsea cubeba*. *Mol Genet Genomics* **288**, 727–737 (2013).
54. Lovdal, T. & Saha, A. Reference gene selection in *Carnobacterium maltaromaticum*, *Lactobacillus curvatus* and *Listeria innocua* subjected to temperature and salt stress. *Mol Biotechnol* **56**, 210–222 (2014).

## Acknowledgements

This research was supported by the Chinese Key Technology Research and Development Program of Twelfth Five-Year Plan (No. 2013BAD01B06).

## Author Contributions

J.N. and S.L. organized and planned the research, and drafted the manuscript. Y.C., J.A. and X.H. performed the sequence analysis. J.C. and J.W. determined the content and composition of terpenoid and oil. Z.Z. and S.L. provided funding, computational guidance, and were substantially involved in research design, data analysis, and manuscript preparation. All authors have read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Niu, J. *et al.* Integrated transcriptome sequencing and dynamic analysis reveal carbon source partitioning between terpenoid and oil accumulation in developing *Lindera glauca* fruits. *Sci. Rep.* **5**, 15017; doi: 10.1038/srep15017 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>