INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
17–0611/2018/20–5–991–1004
DOI: 10.17957/IJAB/15.0591
http://www.fspublishers.org

Full Length Article

Transcriptome Survey Reveals Candidate Genes Involved in Lipid Metabolism of *Carya illinoinensis*

Xiaodong Jia[1,2], Mingzhi Li[4], Huiting Luo[2], Min Zhai[2], Zhongren Guo[2], Yongrong Li[3], Yushan Qiao[1*] and Liangju Wang[1*]
[1]*College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China*
[2]*Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China*
[3]*Nanjing Green Universe Pecan Scientific Company, Nanjing 210007, China*
[4]*Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*
[*]For correspondence: qiaoyushan@njau.edu.cn; wlj@njau.edu.cn

Abstract

Pecan [*Carya illinoinensis* (Wangenh.) K. Koch] kernel contains high proportion of fatty acid, which has the benefits of reducing the risk of blood and heart diseases. However, the molecular mechanism of lipid metabolism in pecan kernels has remained unclear. In this study, the changes of fatty acids in developing pecan kernels (cultivar Pawnee) was monitored using gas chromatography (GC) method, and the transcriptome analysis was performed on four critical stages of lipid accumulation. Twenty-five fatty acids were identified and a total of 73,262 unigenes were obtained after assembling, in which 7709 unigenes were annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The top annotated pathway within "Metabolism" catalog in KEGG was "Carbohydrate metabolism". A putative pecan lipid metabolic pathway was reconstructed including 26 homologs consist of 93 unigenes. The stearoyl-ACP desaturase (SAD) and the omega-6 fatty acid desaturase (FAD2) were found play vital roles in the formation of the high oleic acid content and corresponding fatty acid ratio in pecan kernel. The acetyl coenzyme A carboxylase (ACCase), the glycerol-3-phosphate acyltransferase (GPAT), the diacylglycerol acyltransferase (DGAT), and the phospholipid: diacylglycerol acyltransferase (PDAT) were found may play important roles in pecan lipid accumulation too. Transcription factor (TF) families like AP2-EREBP, bZIP, ARF, and C3H might have positive transcriptional regulations to lipid biosynthesis in pecan, while MYB might have negative transcriptional regulations. To our knowledge, this is the first report of global transcriptional profile of pecan kernels. These data will provide a theoretical basis for future genetic modification of fatty acids in pecan and other oilseed. © 2018 Friends Science Publishers

Keywords: *Carya illinoinensis* (Wangenh.) K. Koch; Lipid metabolism; Pecan; Transcriptome

Introduction

Pecan [*Carya illinoinensis* (Wangenh.) K. Koch] is a North American deciduous tree that belongs to the Juglandaceae family. It had been introduced into China in the 1900s and reconsidered as extremely promising tree oil with great health benefits. Pecan nuts contain high proportion of lipid, which are predominantly composed of triacylglycerol (TAGs). The diverse characters of plant oils are primarily determined by the fatty acid composition of TAG. In mature pecan kernels, the main constituents of fatty acids are unsaturated fatty acids (UFA). Among all fatty acids, the quantity of oleic acid was the highest, followed by linoleic acid, palmitic acid, stearic acid, and α-linolenic acid. The unique health benefits of pecan oil lie in its high oleic acid content. It is reported that pecan nut can reduce the risk of coronary heart disease by reducing high blood cholesterol level (Rajaram *et al.*, 2001; Hudthagosol *et al.*, 2011). The extraordinarily high content of oleic acid also makes pecan a good material for the study of the mechanism involved in the metabolism of monounsaturated fatty acids (MUFAs).

Both control of the oil composition and improvement of the total oil yield have been the major goals of breeders and molecular biologists. Our previous research showed that there are considerable variations of fatty acid contents among different pecan varieties introduced into China (not published yet). Understanding the lipid metabolism process and its molecular mechanism is the foundation to achieve the above objectives. Previous pecan researches mainly focused on marker developing or linkage mapping (Vendrame *et al.*, 1999; Conner and Wood, 2001; Cerna-Cortés *et al.*, 2003; Grauke *et al.*, 2003; Beedanagari *et al.*, 2005) and several genes were also cloned and analyzed (Zhang *et al.*, 2015). But genomic resources are still unavailable for further studies related to nutritional accumulation mechanism. Lipid biosynthesis begins with the *de novo* synthesis of fatty acid in the plastid. Acetyl-CoA were converted to malonyl-CoA and catalyzed

To cite this paper: Jia, X., M. Li, H. Luo, M. Zhai, Z. Guo, Y. Li, Y. Qiao and L. Wang, 2018. Transcriptome survey reveals candidate genes involved in lipid metabolism of *Carya illinoinensis*. *Int. J. Agric. Biol.*, 20: 991–1004

by the plastidial ACCase. Then the plastidial fatty acid synthase (FAS) catalyzed the chain extension, two carbons at a time to form 16:0- and 18:0-ACP (Shintani and Ohlrogge, 1995). The latter product can be further modified by SAD (Kachroo *et al.*, 2007). GPAT and DGAT catalyzed the first and last acylation steps of G-3-P (Murata and Tasaka, 1997; Ohlrogge and Browse, 1995). The fatty acid desaturase (*FAD*) gene family controlled the desaturation of fatty acids, among which *FAD2*, *FAD6* catalyzed the generation of linoleic acid; whereas, *FAD3*, *FAD7* and *FAD8* catalyzed the generation of linolenic acid (Ohlrogge and Browse, 1995). TFs like ABI3, LEC1, LEC2, FUS3, WRI1 and MYB were reported to have effects in regulating lipid biosynthesis, previously (Santos-Mendoza *et al.*, 2008; Baud and Lepiniec, 2010).

De novo transcriptome sequencing provides a powerful approach to explore the mechanisms underlying lipid biosynthesis. For example, by using RNA sequencing (RNA-Seq), genes involved in lipid biosynthesis in olive (Muñoz-Mérida *et al.*, 2013), persea Americana (Kilaru *et al.*, 2015), peanut (Yin *et al.*, 2013), soybean (Chen *et al.*, 2012), flaxseed (Venglat *et al.*, 2011), and oil palm (Dussert *et al.*, 2013) were identified and profiled. To set up the foundation for genomic studies of pecan, an RNA-Seq experiment was conducted using samples from different key developmental stages of pecan, and to address three objectives: (1) to obtain basic transcriptome data for pecan; (2) to identify key regulatory genes associated with lipid metabolism; and (3) to determine the transcriptome mechanisms for the formation of the corresponding fatty acid ratio in pecan kernels.

Materials and Methods

Plant Materials

Pecan cultivar Pawnee was used in this study because of its wide cultivation area in China and its high lipid content. The experimental trees were planted in the scientific orchards of the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences under normal management of water and fertilizer. Fifteen healthy pecan trees (10-years-old) were selected as sample trees. The nuts were sampled from the four directions, east, south, west, and north, at 10-days interval from 95 to 165 days after full blossoming in 2013 (S1-S8 equal to 95-165 d). Two healthy nuts were sampled from each direction makes eight nuts each tree. Nuts from every five trees were combined together to form three biological replicates. The samples were immediately placed in an ice box and transported back to the laboratory. The kernels were separated and stored at -70°C for analysis. Before use, samples were powdered and homogeneously mixed.

Preparation of Fatty Acid Methyl Esters (FAMEs)

Lipids were extracted using Soxhlet extraction method with~10 g kernel and 150 mL n-hexane. Solvent were then removed at~50°C using a Rotovap (Büchi Rotovapor R-114, Büchi Labortechnik AG, Switzerland). FAMEs were prepared according to the official method of the ISO (International Standardization Organization) (Standardization 2000). Briefly, pecan oil (about 30 mg) was weighed into a 50 mL flask together with 4 mL sodium hydroxide methanol solution (NaOH/MeOH, 0.5%) and a few defatted zeolites. The flask was refluxed in a water bath at 85°C under nitrogen protection and shaken every 60 s until the oil disappeared. And then 5 mL boron trifluoride catalyst reagent was added to the flask. After keep refluxing for 3 min, isooctane (3 mL) was added, and the reaction was stopped immediately. FAMEs were extracted by adding a saturated solution of NaCl (20 mL) with violent shaken. After separate into layers, the supernatant was collected by syringe and filtrated with 0.45 μm polytetrafluoroethylene (PTFE) filters (Branch Bollion Lung Experimental Equipment Co., Ltd, Tianjing, China) and stored at 4°C for GC analysis. Methyl heptadecanoate was added as internal standard (Nu-Chek-Prep, Elysian, MN, USA).

Fatty Acid Profiling

FAMEs were analyzed on an Agilent 6890N gas chromatograph. A Supelco SP-2340 column (100 m×0.25 mm, 0.20 μm, Pennsylvania, USA) was used. Chromatographic program was as follow: initial temperature of 100°C, last for 2 min, ramp at 5°C/min to 200°C and hold for 1 min, ramp at 10°C/min to 280°C and hold for 10 min. The injector temperature was 250°C, the flame ionization detector (FID) was 200°C. The helium, air, and hydrogen flows were set at 1.6, 300, and 35 mL/min, respectively. Three replicates were measured on each period. The fatty acids identifications were carried out using Supelco FAME mixture of standards (Nu-Chek-Prep, Elysian, MN, USA). The fatty acids contents were calculated using area normalization method and demonstrated as contents in kernels.

RNA Isolation, Library Construction, and RNA Sequencing

Samples of four key stages were chosen according to the lipid (Jia *et al.*, 2016) and fatty acid results to perform transcriptional profiling, which were S3, S5, S6 and S8 (Fig. 1). Total RNA was isolated from the kernel tissue of each of the three biological replicates using plant RNA extraction kit (TIANDZ, Inc Beijing, China, with DNase). The quantity of the isolated total RNA samples were examined by a NanoDrop2000 Spectrophotometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis (1.2%), while the quality were valued by an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Four libraries (three biological replicates of each period were combined into one library) were construction and sequenced using

Illumina HiSeq2500 platform. All data were up loaded to the NCBI database under BioProject ID PRJNA347430.

Assembling, Annotation and Differentially Expressed Genes (DEGs) Analysis

The raw reads were cleaned and *de novo* assembled using Trinity program (Pertea *et al.*, 2003; Haas *et al.*, 2013). Functions of unigenes were annotated using BLAST program (E-value < 1E-5). Sequences were compared with public databases including NCBI, non-redundant protein database (NR), reviewed protein sequence database (Swiss-Prot), Clusters of Orthologous Groups of proteins (COG), KEGG, Gene Ontology (GO) and Pfam database (version 26.0). TFs were labeled using iTAK software (Zheng *et al.*, 2016). The unigenes expressions were calculated by RPKM value. The DEGseq R package (1.12.0) was used to screen DEGs between each two samples (Wang *et al.*, 2010). The q-value (adjusted p-values) ≤ 0.005 and the absolute value of log2 ratio ≥ 1 were set as the threshold for significantly differential expression (Storey and Tibshirani, 2003). Hierarchical clustering analysis was carried out for all DEGs union set between each two samples.

Quantitative RT-PCR Analysis

The primer sequences were designed using Beacon designer 7 (Supplementary Table 1). Total RNA samples were prepared as described previously. The first-strand cDNA was then synthesized using a Prime Script RT Reagent Kit (TaKaRa, Japan). The RT-qPCR reactions were performed on the ABI7500 fast Real-Time PCR system (Applied Biosystems, USA) of three biological replicates and three technical repeats with the SYBR® Premix ExTaq™ Kit (Takara, Dalian, China). The *Actin* gene was selected as the internal control (Zhang *et al.*, 2015). The relative mRNA expression levels were calculated using the 2-ΔΔCt method (Livak and Schmittgen, 2001). Data analyses were conducted using SPSS version 17.0 statistical software.

Results

Fatty Acid Profiling

Fatty acid compositions of all stages were measured (Table 1). Five dominant components were oleic acid ($C18:1\Delta^{9c}$, 68.39% at S7), linoleic acid ($C18:2\Delta^{9c,\ 12c}$, 22.44%), palmitic acid (C16:0, 5.63%), stearic acid (C18:0, 1.93%), and α-linolenic acid ($C18:3\Delta^{9c,\ 12c,\ 15c}$, 0.96%). The combined relative content of these five fatty acids were 41.68% at S1, 70.89% at S2, 97.16% at S3, and over 99% from S4-S8. The percentage of oleic acid increased significantly ($p< 0.05$) from 17.87% at S1 to 68.21% when fruit was completely ripened. The percentage of linoleic acid reached the highest point at S3 and then decreased, whereas, α-linolenic acid decrease significantly ($p < 0.05$) from 4.34% at S1 to 0.88% at full maturity.

Twenty other minor fatty acids were identified at S1 and S2, while this number reduced to 8 at S4-S8. And the percentages of these minor fatty acids were decrease significantly ($p < 0.05$) at mature when compare with the early stages of pecan kernel. For example, the percentages of C15:1, C16:1and C17:1 all exceeded 3% at S1 and then declined to traceable amount as the kernel matured.

Sequencing and Assembly

From the 4 libraries, 61,394,362, 63,832,162, 55,796,924 and 58,437,290 100-bp paired-end clean reads were obtained, respectively. To ensure the reliability of the libraries, we performed quality controls. Each sample received a Q20 ratio greater than 95%, a Q30 ratio greater than 90%, and a GC proportion between 45–49%, demonstrating high and reliable sequencing quality for analysis in the next step (Supplementary Table 2). Using the high-quality reads, 132,720 transcripts were assembled and 73,262 unigenes were identified (Supplementary Table 3, Supplementary Fig. 1).

Functional Annotation and Classification of Unigenes

The analyses showed that 27,935 unigenes (38.13%) had significant matches in the Nr database, 15,120 unigenes (20.63%) in the Nt database, 18,783 unigenes (25.63%) in the Swiss-Prot database, 9222 unigenes (12.58%) in the KOG database, and 19,103 unigenes (26.07%) in the PFAM database (Table 2). A total of 31,415 unigenes (42.88%) were successfully annotated in at least one database, with 2243 unigenes (3.06%) in all seven databases. The annotation rate (42.88%) was not very low in comparison to other plants whose genomes have not been studied. Unigenes without matching sequences in these databases may be new genes.

In total, 22,019 unigenes (30.05%) had a GO annotation, and these unigenes were assigned to 47 functional terms with three ontologies (Biological Process, Cellular Component, and Molecular Function) (Fig. 2). Of these unigenes, “Biological Process” occupied the majority, followed by “Cellular Component” and “Molecular Function”. In the comprehensive range of GO categories, “cellular process” (13,517 unigenes), “metabolic process” (12,563 unigenes), “binding” (12,634 unigenes), “catalytic activity” (10,599 unigenes) and “cell” (8279 unigenes) were the top five dominant.

The functions of 7709 unigenes were also classified using the KEGG database into 31 KEGG pathways (Fig. 3). The top five annotated pathways within “Metabolism” catalog were “Carbohydrate metabolism” (763 unigenes), “Energy metabolism” (553 unigenes), “Amino acid metabolism” (506 unigenes), “Lipid metabolism” (375 unigenes) and “Metabolism of cofactors and vitamins” (234 unigenes).

Table 1: Changes in lipid content (g·100 g^{-1} FW) and fatty acid composition (% of total fatty acids) in developing kernels of Pawnee pecans (Mean ± SD)

Analyte	Sample name							
	S1	S2	S3	S4	S5	S6	S7	S8
Lipid*	4.18 ± 1.07^{g}	3.12 ± 0.89^{h}	4.96 ± 1.01^{f}	14.04 ± 0.87^{e}	32.31 ± 0.85^{d}	56.17 ± 1.11^{a}	49.49 ± 0.98^{b}	41.03 ± 0.86^{c}
Saturated (SFA)								
C8:0	0.08 ± 0.001	0.09 ± 0.001	nd	nd	nd	nd	nd	nd
C10:0	0.14 ± 0.003	0.11 ± 0.001	nd	nd	nd	nd	nd	nd
C12:0	0.32 ± 0.005	0.21 ± 0.002	nd	nd	nd	nd	nd	nd
C13:0	0.34 ± 0.003	0.28 ± 0.002	nd	nd	nd	nd	nd	nd
C14:0	0.20 ± 0.001^{b}	0.40 ± 0.001^{a}	0.05 ± 0.005^{c}	0.02 ± 0.002^{c}	0.03 ± 0.001^{c}	0.03 ± 0.001^{c}	0.03 ± 0.001^{c}	0.04 ± 0.002^{c}
C15:0	0.12 ± 0.001^{a}	0.12 ± 0.001^{a}	0.11 ± 0.001^{a}	0.01 ± 0.001^{a}	0.01 ± 0.001^{a}	0.01 ± 0.001^{a}	0.01 ± 0.002^{a}	0.01 ± 0.001^{a}
C16:0	3.94 ± 0.11d	13.00 ± 0.36^{a}	9.53 ± 0.26^{b}	4.78 ± 0.12^{c}	4.84 ± 0.12^{c}	5.65 ± 0.02^{c}	5.63 ± 0.02^{c}	5.55 ± 0.04^{c}
C17:0	0.11 ± 0.003^{a}	0.11 ± 0.002^{a}	0.19 ± 0.005^{a}	0.04 ± 0.001^{a}	0.04 ± 0.001^{a}	0.06 ± 0.001^{a}	0.05 ± 0.001^{a}	0.05 ± 0.001^{a}
C18:0	2.01 ± 0.04b	4.65 ± 0.13^{a}	1.92 ± 0.01^{b}	1.80 ± 0.05^{b}	1.83 ± 0.05^{b}	2.00 ± 0.005^{b}	1.93 ± 0.02^{b}	2.00 ± 0.04^{b}
C20:0	0.22 ± 0.005^{ab}	0.33 ± 0.006^{a}	0.22 ± 0.006^{ab}	0.08 ± 0.002^{b}	0.08 ± 0.002^{b}	0.09 ± 0.001^{b}	0.08 ± 0.001^{b}	0.09 ± 0.002^{b}
C22:0	0.46 ± 0.01	0.48 ± 0.01	nd	nd	nd	nd	nd	nd
C24:0	0.17 ± 0.001	0.16 ± 0.001	nd	nd	nd	nd	nd	nd
Monounsaturated (MUFA)								
C14:1n-5	0.51 ± 0.01^{a}	0.54 ± 0.01^{a}	0.05 ± 0.001^{b}	nd	nd	nd	nd	nd
C15:1n-5	3.90 ± 0.11^{a}	2.57 ± 0.05^{b}	0.30 ± 0.01^{c}	0.04 ± 0.001^{c}	0.04 ± 0.001^{c}	0.03 ± 0.001^{c}	0.02 ± 0.001^{c}	0.02 ± 0.001^{c}
C16:1n-7	3.47 ± 0.09^{a}	3.14 ± 0.08^{a}	0.34 ± 0.01^{b}	0.07 ± 0.002^{b}	0.07 ± 0.001^{b}	0.09 ± 0.001^{b}	0.08 ± 0.002^{b}	0.02 ± 0.001^{b}
C17:1n-7	6.59 ± 0.12^{a}	3.42 ± 0.08^{b}	0.97 ± 0.02^{bc}	0.12 ± 0.006^{c}	0.14 ± 0.005^{c}	0.15 ± 0.003^{c}	0.14 ± 0.001^{c}	0.05 ± 0.001^{c}
C18:1n-9	17.87 ± 0.11^{g}	29.58 ± 0.87^{f}	58.05 ± 0.32^{e}	77.80 ± 0.61^{a}	74.38 ± 0.81^{b}	63.71 ± 0.06^{d}	68.39 ± 0.29^{c}	68.21 ± 0.37^{c}
C20:1n-9	3.29 ± 0.09^{a}	2.95 ± 0.06^{a}	0.68 ± 0.01^{b}	0.19 ± 0.001^{b}	0.18 ± 0.002^{b}	0.21 ± 0.001^{b}	0.17 ± 0.001^{b}	0.18 ± 0.001^{b}
C22:1n-9	2.60 ± 0.05^{a}	1.38 ± 0.02^{b}	0.41 ± 0.001^{b}	nd	nd	nd	nd	nd
Polyunsaturated (PUFA)								
C18:2n-6	13.52 ± 0.38^{e}	19.41 ± 0.58^{c}	26.28 ± 0.84^{a}	14.43 ± 0.29^{e}	17.19 ± 0.48^{d}	26.80 ± 0.03^{a}	22.44 ± 0.18^{b}	22.88 ± 0.32^{b}
C18:3n-3	4.34 ± 0.09^{a}	4.25 ± 0.11^{a}	1.38 ± 0.03^{b}	0.77 ± 0.01^{b}	0.77 ± 0.01^{b}	1.05 ± 0.004^{b}	0.96 ± 0.005^{b}	0.88 ± 0.01^{b}
C20:2n-6	2.49 ± 0.05^{a}	1.42 ± 0.02^{b}	0.38 ± 0.01^{c}	nd	nd	nd	nd	nd
C20:4n-6	0.80 ± 0.001	0.16 ± 0.001	nd	nd	nd	nd	nd	nd
C20:5n-3	1.13 ± 0.003	0.64 ± 0.001	nd	nd	nd	nd	nd	nd
C22:2n-6	1.12 ± 0.01	1.71 ± 0.01	nd	nd	nd	nd	nd	nd

Each value is a mean ± standard deviation (SD) of triplicate analysis results of different samples. Fatty acid: C8:0, caprylic acid; C10:0, carpic acid; C12:0, dodecanoicacid; C13:0, tridecanoic Acid; C14:0, myristic acid; C15:0, pentadecylic acid; C16:0, palmitic acid; C17:0, margaric acid; C18:0, stearic acid; C20:0, arachidic acid; C22:0, behenic acid; C24:0, lignocericacid; C14:1n-5, tetradecenoic acid; C15:1n-5, pentadecenoic acid; C16:1n-7, hexadecenoic acid; C17:1n-7, heptadecenoic acid; C18:1n-9, oleic acid; C20:1n-9, eicosenoic acid; C22:1n-9, erucic acid; C18:2n-6, linoleic acid; C18:3n-3, α-linolenic acid; C20:2n-6, eicosadienoic acid; C20:4n-6, arachidonic acid; C20:5n-3, eicosapentanenoic acid (EPA); C22:2n-6,docosadienoic acid

Means with same letters in the same line were not significantly different according to the Turkey's test at $p < 0.05$

Nd means not detected

* Data of lipid content was cited from our previous report (Jia *et al.*, 2016)

Supplementary Table 1: Gene-specific primers sequence for detection by RT-qPCR

Unigene No.	Enzyme	Forward primer (5′-3′)	Reverse primer (5′-3′)
-	*Actin*	GCTGAACGGGAAATTGTC	AGAGATGGCTGGAAGAGG
comp 61443_c0	ACC(CTα)	GATTAGGAAGGTGAAGAAG	GCTGTATAGGAGTGAGAT
comp 54534_c0	ACC(BC)	CTTCTGATTCCTCCACTT	CACTTTCTCTACCTTTCTC
comp 56179_c0	MAT	CTATCTATGTCACGAGTC	GAATCCAGTCCTATTACAC
comp 49975_c0	HAD	TACTCAATCACCCTATCC	CCACTCTTTCTCAATCTC
comp 59929_c0	KAS II	CGCTCCTTATCAATCTTAG	GTATTACAACAGGCTTCTC
comp 59424_c1	SAD	ATAGATGAATCCGAGGTAG	CCCTTCCAACTTATCAAAC
comp 60218_c0	FATA	GGTTCCATAGTTGTAAGTG	AGAGGAGAATAGTAGTAGC
comp 48868_c0	DGAT1	CTGCTGAATGAAGAAGTC	GAGAAGACAGGAGACATA
comp 57877_c0	PDAT	GTAACTTGGATTGGTCTG	CTGCTCATCTCATCATAC
comp 51666_c0	FAD2	GAAGTAAGAGGTAGCAATG	CAACAAGAGTGAAGAACAG

Supplementary Table 2: QC summary of clean data

Sample	Raw Reads	Clean Reads	Clean Bases	Error (%)	Q20 (%)	Q30 (%)	GC Content (%)
S3_1	31772999	30697181	3.07G	0.03	97.62	92.59	45.29
S3_2	31772999	30697181	3.07G	0.03	98.11	94.28	45.29
S5_1	33234362	31916081	3.19G	0.03	97.52	92.29	46.63
S5_2	33234362	31916081	3.19G	0.03	97.93	93.78	46.64
S6_1	28788445	27898462	2.79G	0.03	97.37	91.68	48.49
S6_2	28788445	27898462	2.79G	0.03	97.95	93.51	48.51
S8_1	30370513	29218645	2.92G	0.03	97.51	92.37	45.36
S8_2	30370513	29218645	2.92G	0.03	97.98	94	45.37

Table 2: All-in-one list of functional annotation of the pecan unigenes

Database	73,262 all unigenes	
	Number of annotation	Percent of annotation (%)
Nr	27,935	38.13
Nt	15,120	20.63
Swiss-Prot	18,783	25.63
KOG	9222	12.58
PFAM	19,103	26.07
GO	22,019	30.05
KEGG	7709	10.52
Annotated in at least one database	31,415	42.88
Annotated in all databases	2243	3.06
Unknown	41,847	57.12

Supplementary Table 3: Length distribution of assembled transcripts and unigenes

	Total	≥ 500 bp	≥ 1000 bp	N50	N90	Total length	Max length	Min length	Mean length
Transcripts	132,720	75,245	50,395	1907	427	144,773,395	15,805	201	1091
Unigenes	73,262	27,029	14,412	1299	276	52,669,633	15,805	201	719

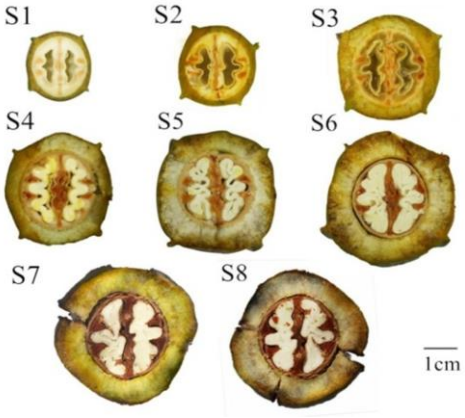
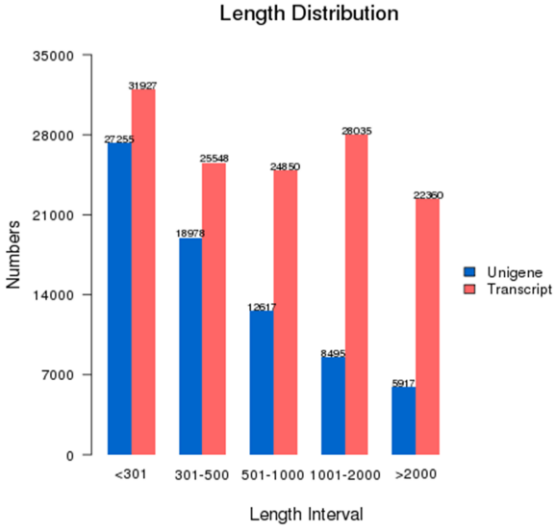


Fig. 1: Different developmental stages of *Carya illinoinensis* Pawnee. S1 and S2, lipid was accumulated in low speed; S3, beginning of the fast lipid accumulation stage; S4 and S5, middle of the fast lipid accumulation stage; S6, end of the fast lipid accumulation stage; S7, kernel mature; S8, later stage of kernel maturity. Scale bars = 1.0 cm



Supplementary Fig. 1: Length distribution of transcripts and unigenes

Analysis of Differentially Expressed Unigenes

To calculate the expression of unigenes, we mapped the clean reads of each sample to the reference unigenes. In the four samples, 51,151,550 (83.32%), 54,565,652 (85.48%), 48,009,544 (86.04%) and 49,706,198 (85.06%) were uniquely mapped to the reference unigenes, respectively. The Venn diagram shows the distribution of unigenes among the four samples, in which 21,902 genes were mutually expressed among all four samples, while 12,265 (S3), 815 (S5), 623 (S6) and 3827 (S8) were expressed only in one sample, respectively (Supplementary Fig. 2). To characterize the expression model of DEGs, a heat map was constructed after H-cluster (Supplementary Fig. 3). Three different expression patterns appeared, which were the highest expression level appeared at S3 (green box), S5 (yellow box), and S6 or S8 (red box). Furthermore, differentially expressed genes between different samples were selected and compared. There were 2510 DEGs in S5, 3086 DEGs in S6 and 3685 DEGs in S8 compared with S3.

To classify DEGs accurately, KEGG functional enrichment analysis was performed. The top 20 represented pathways of both up-regulated and down-regulated were examined between S5 and S3, S6 and S3, S8 and S3. Both "Carbon metabolism" and "Glycolysis/Gluconeogenesis" can be seen among the top 20 up-regulated represented pathways between S5/S6 and S3, so were "Pyruvate metabolism", "Fatty acid biosynthesis" and "Fatty acid metabolism", showing their involvement in the early stages of kernel development. On the contrary, "Biosynthesis of unsaturated fatty acids" can only be seen in top 20 up-regulated represented pathways between S6/S8 and S3,

Supplementary Table 4: KEGG classification of lipid metabolic pathways

Pathway Hierarchy1	Pathway Hierarchy2	KEGG Pathway	Pathway ID	Gene Number
Metabolism	Lipid metabolism	Fatty acid metabolism	ko00071	58
		Fatty acid biosynthesis	ko00061	38
		Fatty acid elongation	ko00062	34
		Biosynthesis of unsaturated fatty acids	ko01040	35
		Linoleic acid metabolism	ko00591	23
		alpha-Linolenic acid metabolism	ko00592	52
		Cutin, suberine and wax biosynthesis	ko00073	13
		Ether lipid metabolism	ko00565	33
		Glycerolipid metabolism	ko00561	65
		Glycerophospholipid metabolism	ko00564	86
		Sphingolipid metabolism	ko00600	26
		Steroid biosynthesis	ko00100	34
		Steroid hormone biosynthesis	ko00140	4
		Synthesis and degradation of ketone bodies	ko00072	6

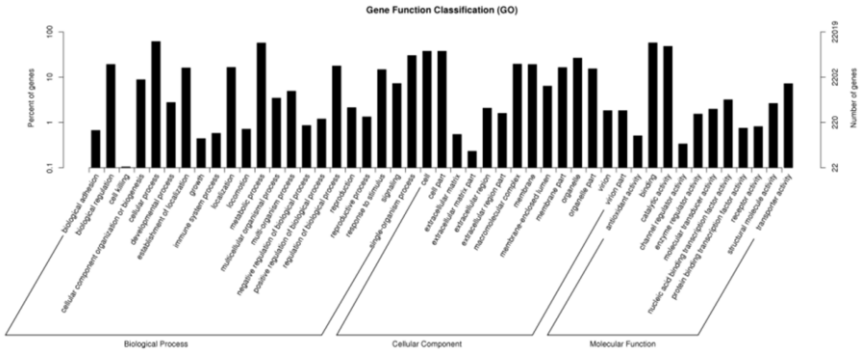


Fig. 2: Functional classification of pecan transcriptome based on gene ontology categorization

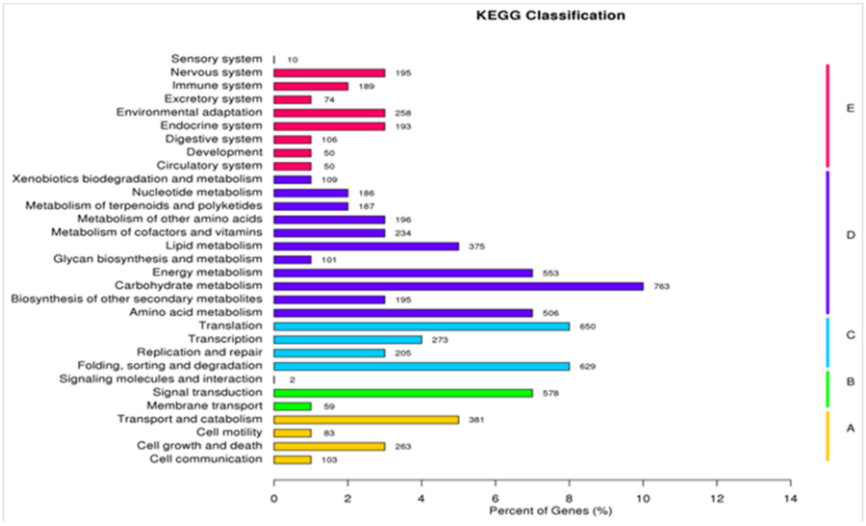


Fig. 3: Distribution of pecan transcriptome genes based on KEGG pathway analysis

while "Fatty acid degradation" can only be found in top 20 up-regulated represented pathways between S8 and S3, consistent with the cascade of lipid metabolic pathway.

Candidate Genes Involved in Lipid Metabolic Pathway

In total, 507 unigenes were annotated implicated in lipid metabolic pathway of pecan (Supplementary Table 4). By compare these unigenes with key enzymes of lipid metabolic pathways, 26 putative homologs consist of 93 unigenes were selected to reconstruct a putative pecan lipid metabolic pathway (Fig. 4). To better understand the expression pattern of these unigenes, their RPKM values in different sample stages were then searched manually, listed in Supplementary Table 5 and marked in Fig. 4.

Supplementary Table 5: Unigenes involved in lipid biosynthesis in pecan

Gene ID	Pathway	EC	Symbol	Definition	RPKM value			
					S3	S5	S6	S8
comp 47818_c0*	HIF-1 signaling pathway	1.2.4.1	PDH(E1α)	pyruvate dehydrogenase E1 component subunit alpha	125.99	243.95	220.93	38
comp 49853_c0	HIF-1 signaling pathway	1.2.4.1	PDH(E1α)	pyruvate dehydrogenase E1 component subunit alpha	78.16	35.25	34.09	44.61
comp 41186_c0	HIF-1 signaling pathway	1.2.4.1	PDH(E1α)	pyruvate dehydrogenase E1 component subunit alpha	64.42	30.85	39.16	43.67
comp 56759_c0	HIF-1 signaling pathway	1.2.4.1	PDH(E1β)	pyruvate dehydrogenase E1 component subunit beta	73.66	45.73	49.11	66.64
comp 15843_c0*	HIF-1 signaling pathway	1.2.4.1	PDH(E1β)	pyruvate dehydrogenase E1 component subunit beta	142.54	274.07	103.82	9.03
comp 42458_c0	HIF-1 signaling pathway	1.2.4.1	PDH(E1β)	pyruvate dehydrogenase E1 component subunit beta	9.07	11.49	10.27	0.42
comp 47921_c0*	Glycolysis / Gluconeogenesis	2.3.1.12	PDH(E2)	pyruvate dehydrogenase E2 component	71.95	57.44	48.78	5.23
comp 58490_c0	Glycolysis / Gluconeogenesis	2.3.1.12	PDH(E2)	pyruvate dehydrogenase E2 component	42.81	35.05	17.49	28.7
comp 42528_c0	Glycolysis / Gluconeogenesis	2.3.1.12	PDH(E2)	pyruvate dehydrogenase E2 component	4.7	4.81	8.69	7.11
comp 51933_c0	Glycolysis / Gluconeogenesis	2.3.1.12	PDH(E2)	pyruvate dehydrogenase E2 component	13.46	5.84	5.18	6.13
comp 61443_c0*	Fatty acid biosynthesis	6.4.1.2	ACC(CTα)	acetyl-CoA carboxylase, carboxyl transferase subunit alpha	81.54	99.46	89.27	11.14
comp 58593_c0*	Fatty acid biosynthesis	6.4.1.2	ACC(CTα)	acetyl-CoA carboxylase carboxyl transferase subunit alpha	2.75	0.11	1.38	17.8
comp 58109_c1	Fatty acid biosynthesis	6.4.1.2	ACC(CTβ)	acetyl-CoA carboxylase carboxyl transferase subunit beta	449.06	613.5	389.24	347.86
comp 182_c0	Fatty acid biosynthesis	6.4.1.2	ACC(BC)	acetyl-CoA carboxylase, biotin carboxylase subunit	0	2.22	0	0.7
comp 485026_c0	Fatty acid biosynthesis	6.4.1.2	ACC(BC)	acetyl-CoA carboxylase, biotin carboxylase subunit	0.53	0	0.19	0.73
comp 463558_c0	Fatty acid biosynthesis	6.4.1.2	ACC(BC)	acetyl-CoA carboxylase, biotin carboxylase subunit	2.29	0	0.78	0
comp 54534_c0*	Fatty acid biosynthesis	6.4.1.2	ACC(BC)	acetyl-CoA carboxylase, biotin carboxylase subunit	105.21	130.93	207.33	23.56
comp 53371_c0	Fatty acid biosynthesis	6.4.1.2	ACC(BCCP1)	acetyl-CoA carboxylase biotin carboxyl carrier protein	32.12	40.37	46.44	42.47
comp 54877_c0*	Fatty acid biosynthesis	6.4.1.2	ACC(BCCP2)	acetyl-CoA carboxylase biotin carboxyl carrier protein	133.87	196.06	127.87	28.81
comp 21897_c0	Fatty acid biosynthesis	2.3.1.39	MAT	ACP S-malonyltransferase	3.11	0	0	0
comp 56179_c0*	Fatty acid biosynthesis	2.3.1.39	MAT	ACP S-malonyltransferase	59.14	49.73	24.06	0.37
comp 58842_c0	Fatty acid biosynthesis	2.3.1.180	KASIII	3-oxoacyl-ACP synthase III	47.34	70.11	64.5	38.44
comp 41805_c0	Fatty acid biosynthesis	2.3.1.180	KASIII	3-oxoacyl-ACP synthase III	6.04	6.77	5.07	2.76
comp 39121_c0*	Fatty acid biosynthesis	4.2.1.59	HAD	3-hydroxyacyl-ACP dehydratase	32.73	12.5	7.87	1.95
comp 49975_c0*	Fatty acid biosynthesis	4.2.1.59	HAD	3-hydroxyacyl-ACP dehydratase	70.73	59	34.74	6.27
comp 72239_c0	Fatty acid biosynthesis	1.1.1.100	KAR	3-oxoacyl-ACP reductase	14.37	3.98	3.35	16.57
comp 326790_c0	Fatty acid biosynthesis	1.1.1.100	KAR	3-oxoacyl-ACP reductase	1.25	0	1.29	0
comp 51677_c0	Fatty acid biosynthesis	1.1.1.100	KAR	3-oxoacyl-ACP reductase	60.84	22.67	19.4	32.02
comp 49781_c0*	Fatty acid biosynthesis	1.3.1.9	EAR	enoyl-ACP reductase I	216.71	118.2	86.4	8.79
comp 48615_c1	Fatty acid biosynthesis	1.3.1.9	EAR	enoyl-ACP reductase I	24.07	19.02	31.65	5.74
comp 56623_c0*	Fatty acid biosynthesis	1.3.1.9	EAR	enoyl-ACP reductase I	131.28	263.19	157.12	13.34
comp 42434_c0	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	6.64	1.35	1.66	0.8
comp 631358_c0	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	1.52	0	0	0
comp 61344_c0	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	13.77	6.45	2.66	4.7
comp 32074_c0	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	10.2	4.09	7.66	0
comp 49718_c0	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	46.36	22.49	15.84	2.66
comp 59929_c0*	Fatty acid biosynthesis	2.3.1.179	KASII	3-oxoacyl-ACP synthase II	106.44	139.67	114.47	13.32
comp 50658_c1	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	3.46	1.65	0.38	0
comp 56480_c0*	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	153.82	718.45	378.25	188.9
comp 59424_c1*	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	415.67	807.05	640.05	42.36
comp 43207_c0	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	20.36	50.17	162.01	44.87
comp 41892_c0*	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	1.7	0	0.86	0.84
comp 50658_c0	Fatty acid biosynthesis	1.14.19.2	SAD	acyl-ACP desaturase	4.43	2.61	0.94	0.05
comp 59516_c0	Fatty acid biosynthesis	3.1.2.14	FATB	fatty acyl-ACP thioesterase B	124.93	72.67	78.95	56.87
comp 49891_c0*	Fatty acid biosynthesis	3.1.2.14	FATB	fatty acyl-ACP thioesterase B	0	0.17	0.76	46.18
comp 60218_c0*	Fatty acid biosynthesis	3.1.2.14	FATA	fatty acyl-ACP thioesterase A	77.68	117.06	122.33	28.3
comp 42537_c1	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	11.95	1.99	6.66	3.18
comp 54479_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	29.88	53.39	44.12	11.14
comp 32684_c0	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	18.54	7.11	4.34	22.78
comp 51471_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	15.38	2.87	1.53	13.33
comp 61675_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	38.67	45.1	50.29	3.87
comp 61249_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	12.91	2.51	1.65	18.76
comp 55999_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	23.19	18.7	20.71	137.97
comp 60061_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	27.55	5.2	5.03	67.55
comp 57223_c0*	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	47.28	12.05	18.07	9.4
comp 69222_c0	Fatty acid metabolism	6.2.1.3	LACS	long-chain acyl-CoA synthetase	22.22	7.05	4.27	7.91
comp 26590_c0	Glycerolipid metabolism	2.3.1.15	GPAT	glycerol-3-phosphate acyltransferase	4.68	0.66	0.67	0.23
comp 57219_c0	Glycerolipid metabolism	2.3.1.15	GPAT	glycerol-3-phosphate acyltransferase	12.6	2.17	1.68	3.69
comp 19365_c0*	Glycerolipid metabolism	2.3.1.15	GPAT	glycerol-3-phosphate acyltransferase	97.48	30.93	22.86	290.68
comp 55994_c0*	Glycerolipid metabolism	2.3.1.15	GPAT	glycerol-3-phosphate acyltransferase	413.42	359.22	230.88	856.23
comp 42663_c0	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	6.19	4.29	2.31	8.31
comp 61557_c1	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	1.35	0.46	0	0.91
comp 40387_c0*	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	21.52	34.96	74.55	35.68
comp 58091_c0	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	6.72	1.2	6.39	3.15
comp 58656_c0	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	1.93	5.02	9.08	8.48
comp 42160_c0*	Glycerophospholipid metabolism	2.3.1.51	LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	65.99	100.63	100.01	18.22
comp 46533_c1	Glycerolipid metabolism	3.1.3.4	PAP	phosphatidate phosphatase LPIN	7.42	5.76	4.59	5.1
comp 62062_c0	Glycerolipid metabolism	3.1.3.4	PAP	phosphatidate phosphatase LPIN	17.6	12.86	8.18	7.03
comp 23132_c0	Glycerolipid metabolism	3.1.3.4	PAP	phosphatidate phosphatase LPIN	0.55	0.75	3.56	0.86
comp 47480_c1	Glycerolipid metabolism	3.1.3.4	PAP	phosphatidate phosphatase LPIN	29.32	13.89	17.25	17.14
comp 58537_c0*	Glycerolipid metabolism	3.1.3.4	PAP	phosphatidate phosphatase LPIN	0.46	7.73	41.75	5.35
comp 48868_c0*	Glycerolipid metabolism	2.3.1.20	DGAT1	diacylglycerol O-acyltransferase 1	4.26	19.24	13.1	11.48
comp 39618_c0	Glycerolipid metabolism	2.3.1.20	DGAT1	diacylglycerol O-acyltransferase 1	2.68	1.43	0	0
comp 278924_c0	Glycerolipid metabolism	2.3.1.20	DGAT1	diacylglycerol O-acyltransferase 1	4.58	0	0	0

Supplementary Table 5: Continued

Supplementary Table 5: Continued

comp 48830_c0	Glycerolipid metabolism	2.3.1.75	DGAT2	diacylglycerol O-acyltransferase 2	20.04	14.6	8.41	33.58
comp 43859_c0	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	6.42	0.99	1.55	0.46
comp 60363_c0	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	11.09	5.84	9.37	6.94
comp 49015_c0*	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	16.45	1.6	1.52	10.34
comp 34196_c0	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	5.2	1.87	1.15	3.02
comp 57877_c0*	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	3.39	38.72	78.76	33.72
comp 194253_c0	Glycerolipid metabolism	2.3.1.158	PDAT	phospholipid:diacylglycerol acyltransferase	1.47	0.33	0.09	1.25
comp 59185_c1*	Glycerophospholipid metabolism	2.7.8.2	CPT	ethanolaminephosphotransferase	31.87	9.93	4.55	9.07
comp 56578_c0	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD6	omega-6 fatty acid desaturase (delta-12 desaturase)	12.16	16.38	6.9	2.65
comp 1090332_c0	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD2	omega-6 fatty acid desaturase (delta-12 desaturase)	1.09	0	0	0
comp 61050_c1*	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD2	omega-6 fatty acid desaturase (delta-12 desaturase)	1.23	0.64	0	1.84
comp 51666_c0*	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD2	omega-6 fatty acid desaturase (delta-12 desaturase)	87.25	501.22	486.89	168.57
comp 47537_c0	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD8	omega-3 fatty acid desaturase (delta-15 desaturase)	3.69	1.79	0.43	0.62
comp 33955_c0	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD8	omega-3 fatty acid desaturase (delta-15 desaturase)	9.96	2.14	0.62	0.12
comp 40895_c0*	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD8	omega-3 fatty acid desaturase (delta-15 desaturase)	26.02	57.47	16.49	17.71
comp 53307_c0*	Biosynthesis of unsaturated fatty acids	1.14.19.-	FAD8	omega-3 fatty acid desaturase (delta-15 desaturase)	23.42	12.47	18.74	2.7
comp 57633_c0	Glycerophospholipid metabolism	2.3.1.23	LPCAT	lysophosphatidylcholine acyltransferase	21.56	5.5	4.46	28.58
comp 60595_c0*	Glycerophospholipid metabolism	2.3.1.23	LPCAT	lysophosphatidylcholine acyltransferase	23.65	4.45	3.41	8.33
comp 52387_c0*	Glycerophospholipid metabolism	2.3.1.23	LPCAT	lysophosphatidylcholine acyltransferase	4.09	0.75	0.36	25.6

Asterisk (*) means DEGs of S5, S6 and S8 compared with S3 involved in lipid metabolic pathway

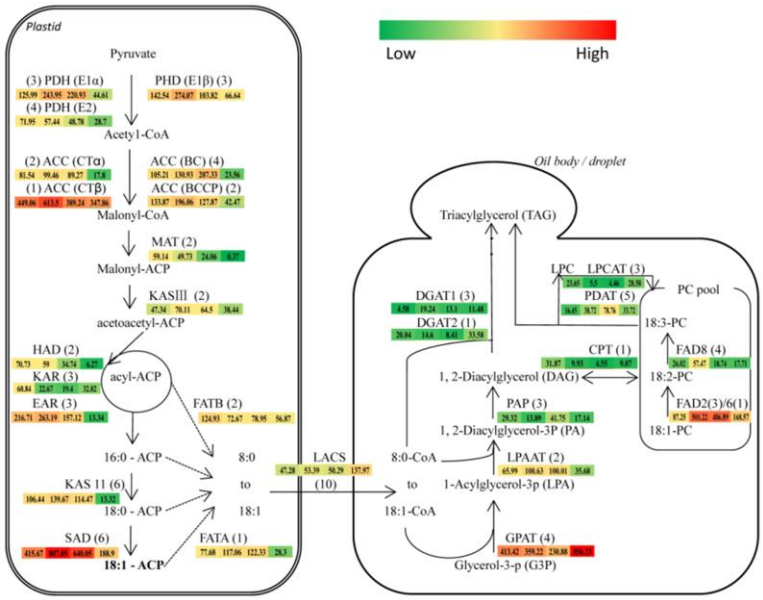


Fig. 4: Reconstruction of the lipid metabolic pathway in pecan. The basic structure of pecan lipid metabolic pathway was constructed using RNA-seq result and referenced from previous report of oil palm (Dussert *et al*., 2013). Numbers in parenthesis represent the number of unigenes. Fatty acids are abbreviated: 8:0, octanoic acid; 16:0, palmitic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3 linolenic acid. Enzyme/protein abbreviations are as follows: PDH (E1α), pyruvate dehydrogenase E1 component subunit alpha; PDH (E1β), pyruvate dehydrogenase E1 component subunit beta; PDH (E2), pyruvate dehydrogenase E2 component (dihydrolipoamideacetyl transferase); ACC (CTα), acetyl-CoA carboxylase carboxyl transferase subunit alpha; ACC (CTβ), acetyl-CoA carboxylase carboxyl transferase subunit beta; ACC (BC), acetyl-CoA carboxylase, biotin carboxylase subunit; ACC (BCCP), acetyl-CoA carboxylase biotin carboxyl carrier protein; MAT, [acyl-carrier-protein] S-malonyltransferase, KAS III, 3-oxoacyl-[acyl-carrier-protein] synthase III; HAD, 3-hydroxyacyl-[acyl-carrier-protein] dehydratase; KAR, 3-oxoacyl-[acyl-carrier protein] reductase; EAR, enoyl-[acyl-carrier protein] reductase I; KAS II, 3-oxoacyl-[acyl-carrier-protein] synthase II; SAD, stearoyl-ACP desaturase (delta-9 desaturase); FATB, fatty acyl-ACP thioesterase B; FATA, fatty acyl-ACP thioesterase A; LACS, long-chain acyl-CoA synthetase; GPAT, glycerol-3-phosphate acyltransferase; LPAAT, 1-acyl-sn-glycerol-3-phosphate acyltransferase; PAP, phosphatidate phosphatase; DGAT1, diacylglycerol O-acyltransferase 1; DGAT2, diacylglycerol O-acyltransferase 2; PDAT, phospholipid:diacylglycerolacyltransferase; CPT, ethanolaminephosphotransferase; FAD2/6, omega-6 fatty acid desaturase (delta-12 desaturase); FAD8, omega-3 fatty acid desaturase (delta-15 desaturase); LPCAT, lysophosphatidylcholineacyltransferase/lyso-PAF acetyltransferase. Colored bars represent the maximum RPKM values of unigene homologs

The expression levels of these unigenes were then analysis combined with the lipid content and fatty acid composition result to find key candidate genes involved in lipid metabolism.

The expressions of unigenes encoding different enzymes were significantly different, which showed the diversity of enzyme activities. Four enzymes were significantly active during kernel development according to their related unigenes showed extremely high expressions over 500 (RPKM value). These four enzymes were the acetyl-CoA carboxylase carboxyl transferase subunit beta (ACC (CTβ)), SAD, FAD2, and GPAT.

Nine unigenes were found to encode the plastidial ACCase, of which five were highly expressed during S3 to S5. Sixteen unigenes were found encoding the FAS, including KASIII (2), HAD (2), KAR (3), EAR (3) and KASII (6), respectively. Five unigenes were found to encode SAD, among which comp59424_c1 and comp56480_c0 were found to express at an extremely high level of 807.05 and 718.45 (RPKM value) at S5, respectively. Unigene comp60218_c0 was annotated as the fatty acyl-ACP thioesterase A (FATA), while unigenes comp59516_c0 and comp49891_c0 were annotated as the fatty acyl-ACP thioesterase B (FATB). The highest expression level of comp60218_c0 appeared at S6, while comp59516_c0 appeared at S3. Different expression patterns of these unigenes were consistent with the fact that FATA is active to 18:0- and 18:1-ACP, while FATB is more active to8:0- to 16:0-ACP. Four unigenes were founded to encoding GPAT. Three unigenes were annotated as the FAD2, one unigene was annotated as FAD6, and four unigenes were annotated as the omega-3 fatty acid desaturase (FAD8). In addition, 6 unigenes were identified to encode PDAT, 3 unigenes encoded the diacylglycerol O-acyltransferase 1 (DGAT1), and 1 unigenes encoded the diacylglycerol O-acyltransferase 2 (DGAT2), catalyzed the final step of lipid synthesis, the acylation of the sn-3 position of diacylglycerol (DAG) to form TAG.

To confirm the transcriptome result, 10 DEGs related to lipid metabolism were selected randomly for RT-qPCR analysis. All these DEGs were first highly and then low expressed, similar to the pattern observed from the sequencing data (Supplementary Fig. 4). The RT-qPCR result confirmed the validity of the transcriptome data.

Table 3: Differentially expressed transcription factors of different stages compared with S3

TF family	Annotated number	S5 up	S6 up	S8 up	S5 down	S6 down	S8 down
MYB	125	1	1	7	14	14	9
AP2-EREBP	110	2	3	1	1	1	0
C2H2	75	3	3	1	2	2	0
NAC	72	0	1	2	4	3	2
C3H	66	0	1	1	0	1	0
bHLH	64	1	1	1	1	1	1
Orphans	58	1	1	4	4	4	5
HB	53	0	1	4	3	2	1
WRKY	53	0	0	6	10	6	6
bZIP	49	1	1	1	0	1	1
GRAS	45	0	0	1	3	3	4
AUX/IAA	38	1	0	0	0	0	0
SNF2	36	0	0	0	1	1	0
CCAAT	32	0	2	1	0	1	0
SET	32	0	1	1	2	1	0
Trihelix	31	2	1	2	5	3	2
PHD	30	0	0	0	2	1	0
G2-like	28	0	0	1	0	0	0
mTERF	28	0	0	1	0	0	0
C2C2-Dof	27	0	0	0	1	1	1
GNAT	26	0	0	1	0	0	1
SBP	25	0	0	0	1	2	1
MADS	23	0	0	1	2	6	5
TCP	22	0	0	0	3	4	1
TRAF	22	0	0	0	0	0	1
HSF	20	1	1	3	1	1	1
LOB	20	1	2	4	1	0	1
ARF	19	1	1	0	0	0	0
OFP	18	0	0	0	1	0	0
SWI/SNF-BAF60b	16	0	0	0	0	0	1
HMG	14	0	0	0	1	1	2
TUB	14	0	0	0	0	0	1
FHA	13	1	1	0	0	0	0
Jumonji	13	0	1	0	0	0	0
PLATZ	12	0	0	2	0	2	1
Tify	12	0	0	0	0	2	3
zf-HD	12	1	1	2	1	1	1
SRS	7	0	0	0	1	1	1
GRF	6	1	1	1	0	0	1
TAZ	5	0	0	0	1	0	0
LIM	4	0	0	0	1	1	1
S1Fa-like	4	0	0	0	0	1	0
Sigma70-like	4	0	0	0	1	0	0
EIL	3	0	0	1	0	0	0
Pseudo ARR-B	3	0	0	0	1	0	0
MBF1	2	0	1	1	0	0	0
DBP	1	0	0	0	0	0	1
Total number	1605	18	26	51	69	68	55

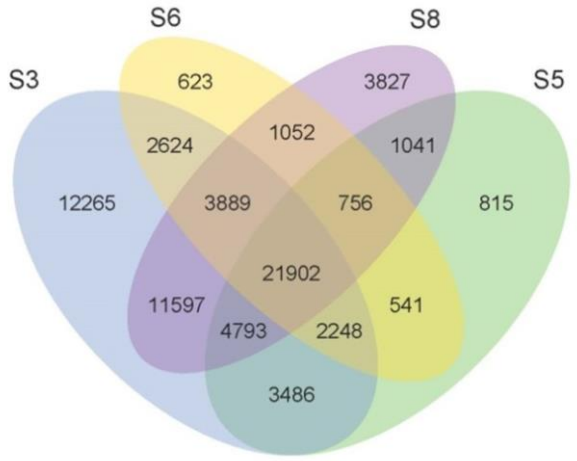
S5 up, S6 up and S8 up, up-regulated transcription factors compared with S3; S5down, S6down and S8 down, down-regulated transcription factors compared with S3

Transcription Factors in Pecan

A total of 1605 TFs from 79 TF families were identified in this study (Supplementary Table 6). Largest number of unigenes (125) was found in MYB family. By compared with S3, differentially expressed transcription factors (DETFs) from 47 TF families were found (Table 3). Among the up-regulated DETF families, 4 unigenes (comp14933_c0, comp14933_c1, comp17377_c0, comp38191_c0) were annotated as AP2-EREP TFs, one unigene (comp34572_c0) was annotated as ARF TF, 3 unigenes (comp41771_c0, comp52398_c1, comp33064_c0) were annotated as bZIP TFs, and one unigene (comp52364_c0) was annotated as C3H TF. Among the down-regulated DETF families, the largest number of unigenes was also found in MYB family.

Supplementary Table 6: Distribution of transcription factors

Category	Total number	Category	Total number	Category	Total number
MYB	125	LOB	20	SWI/SNF-SWI3	4
AP2-REBP	110	HSF	20	Sigma70-like	4
C2H2	75	C2C2-GATA	20	S1Fa-like	4
NAC	72	ARF	19	RWP-RK	4
C3H	66	OFP	18	LIM	4
bHLH	64	SWI/SNF-BAF60b	16	GeBP	4
Orphans	58	TUB	14	CPP	4
WRKY	53	HMG	14	CAMTA	4
HB	53	Jumonji	13	Pseudo ARR-B	3
FAR1	53	FHA	13	PBF-2-like	3
bZIP	49	zf-HD	12	EIL	3
GRAS	45	Tify	12	C2C2-CO-like	3
AUX/IAA	38	PLATZ	12	VOZ	2
SNF2	36	E2F-DP	10	SOH1	2
SET	32	ARID	10	Rcd1-like	2
CCAAT	32	BSD	8	MBF1	2
Trihelix	31	SRS	7	CSD	2
PHD	30	IWS1	7	Coactivator p15	2
mTERF	28	DDT	7	ULT	1
G2-like	28	BES1	7	SAP	1
C2C2-Dof	27	GRF	6	RB	1
GNAT	26	BBR/BPC	6	MED7	1
SBP	25	ARR-B	6	MED6	1
MADS	23	TAZ	5	HRT	1
TRAF	22	LUG	5	DBP	1
TCP	22	C2C2-YABBY	5	Total number	1605
ABI3VP1	22	Alfin-like	5		

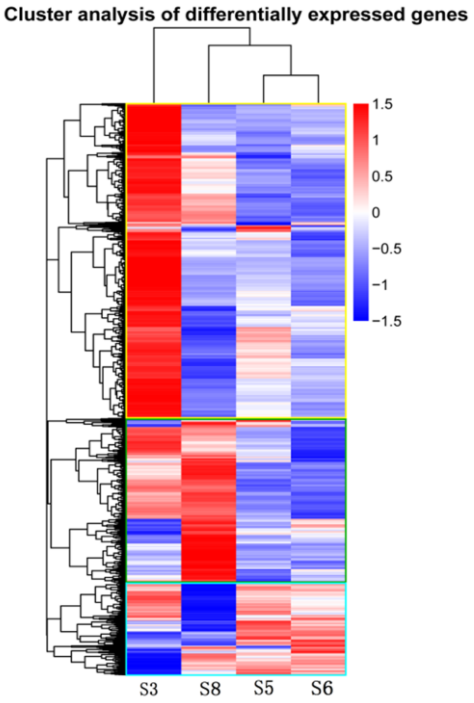


Supplementary Fig. 2: Venn diagrams of DEGs among four samples

Discussion

Fatty Acid Accumulation

The fatty acid types in the early stages of pecan kernel were clearly abundant than the mature stage. This may be explained by the fact that these abundant medium-chain fatty acids were synthesis precursors of long-chain fatty acids such as oleic acid, so they were further assembled into

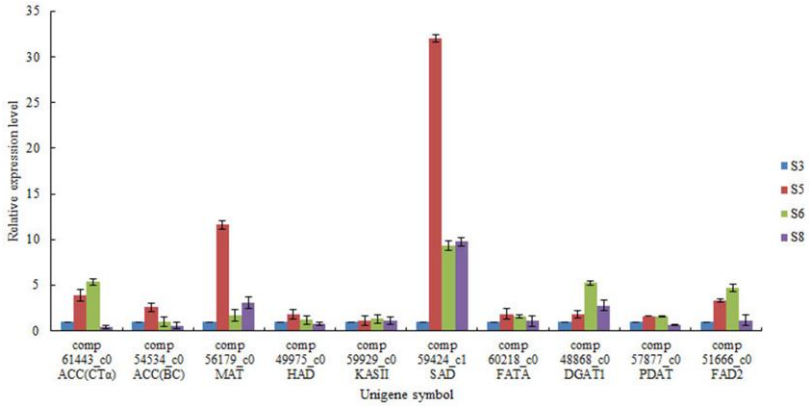


Supplementary Fig. 3: Hierarchical cluster analysis of DEGs. Colors of the bars represented expression levels, red represented a high expression level and blue represented a low expression level. Colors from red to blue indicated the value of log10 (RPKM+1) from low to high. Colors of the boxes represented different expression patterns, green represented the highest expression level appeared at S3, yellow at S5, and red at S6 or S8

downstream products along with the ripen of pecan kernel. The fatty acid compositions of our research were consistent with previous report (Bouali *et al.*, 2013); the difference was our sampling time was 20 d earlier than in the previous report, which is at the beginning of water stage of the pecan kernel.

Carbohydrate Metabolism during Kernel Development

The precursor of lipid biosynthesis is acetyl-CoA, which is transformed from pyruvate. It is reported that pyruvate is formed from carbohydrate which is exported from photosynthetic organs. The carbohydrate is then glycolysised (Baud and Lepiniec, 2009). In pecan kernel, soluble sugar content was high at first and showed an obvious decrease along with the kernel development, while the lipid content was low at the beginning and rise quickly then (Jia *et al.*, 2016). In this study, Carbohydrate metabolism was the top annotated pathway within "Metabolism" catalog in KEGG database. Carbon metabolism and Glycolysis/Gluconeogenesis were the top up-regulated represented KEGG pathways between S5 and S3. These results showed that carbohydrate



Supplementary Fig. 4: RT-qPCR analyses of DEGs involved in lipid synthesis of pecan. Relative expression values normalized to actin are shown as $2^{-\Delta\Delta Ct}$ relative to S3. Error bars represent the SD of three biological replicates with three technical replicates each

metabolism was very active during the kernel development and was highly involved in the lipid metabolism, especially at the beginning of the fatty acid synthesis.

Activity of SAD Enzyme during Synthesis of Oleic Acid

Two unigenes encoding SAD expressed at high levels showing the important role of SAD in fatty acid biosynthesis in pecan (Fig. 4). SAD catalyzes the transform of C18:0 to C18:1. The activity of SAD decided the ratios of saturated fatty acids (SFAs) to MUFAs. Studies on the *Arabidopsis thaliana* genome showed that there are seven isoforms encoding SAD like enzyme, including *ssi2/fab2*, *DES1* to *DES6* (Kachroo *et al.*, 2007). The SAD encoding genes had been characterized from several different species (Shanklin and Somerville, 1991; Cahoon *et al.*, 1998; Whittle *et al.*, 2005). In this study, 6 unigenes were annotated as *DES1*, in which three showed high expression levels while the rest showed low expression levels. We could deduce that temporal specificities and functional differences existed among these isoforms. The expression levels of genes for SAD were consistent with the high oleic acid content in pecan kernel. *SAD* was highly expressed in pecan kernel, indicating its vital role involved in oleic acid biosynthesis. The highest express level of *SAD* in the RT-qPCR confirmed this hypothesis.

***FAD* Gene Family Involved in Final UFAs Ratio**

Desaturation determines the final proportion of UFAs in plant. Studies on the *Arabidopsis thaliana* mutants showed that there are five genetic loci act at C18 fatty acids, in which *FAD2* and *FAD6* encoded omega-6 fatty acid desaturase, while *FAD3*, *FAD7* and *FAD8* encoded omega-3 fatty acid desaturase (Ohlrogge and Browse, 1995). Suppress the expression of *FAD2* can led to high levels of oleic acid in rapeseed (Auld *et al.*, 1992), peanut (Bruner *et al.*, 2001), soybean (Buhr *et al.*, 2002), and cotton seed (Liu *et al.*, 2002). Overexpress the *A. thaliana FAD3* gene can increase the α-linolenic acid content to more than 50% of the total oil (Yadav *et al.*, 1993). The expression levels of genes for desaturases were consistent with the ratio of UFAs. *FAD2* and *FAD3* were expressed at similar levels in sea buckthorn, which had an n-6: n-3 fatty acid ratio of 1:1 (Yang and Kallio, 2001). The *FAD8* gene was highly expressed in tree peony (Li *et al.*, 2015), perilla and Sacha Inchi seed (Wang *et al.*, 2012), in which the content of linolenic acid exceeded 50%. In our study, unigenes encoding FAD2 expressed much higher than that of FAD8, which was consistent with the high linoleic acid and low linolenic acid content in pecan kernel. The expression patterns of unigenes encoding FAD2 and FAD8 play important roles in the formation of the polyunsaturated fatty acids ratio in pecan. High FAD2 enzymes activity led to the relatively high linoleic acid content in pecan kernel.

Rate-limiting Steps of Lipid Metabolism in Pecan

ACCase is another key enzyme. It is the first rate-limiting step in fatty acid synthesis and is subject to feedback inhibition by oleic acid (Shintani and Ohlrogge, 1995; Kozaki *et al.*, 2001; Andre *et al.*, 2012). Attempts to up-regulate fatty acid synthesis by manipulating signal subunits of ACCase yielded complex results. Increasing biotin carboxylase in tobacco leaves had little effect (Shintani *et al.*, 1997), and overexpression of ACC (BCCP) decreased fatty acid synthesis in *Arabidopsis thaliana* (Thelen and Ohlrogge, 2002). Furthermore, the introduction of ACCase into rapeseed seed increased ACCase activity but had no effect on seed lipids (Roesler *et al.*, 1997). In our experiment, several homolog unigenes encoding this enzyme expressed at relatively high levels showed its important roles in pecan kernels which contained high lipid content (Fig. 4). Unigenes encoding ACC (CTα), ACC (BCCP), and ACC (BC) had different expression levels at different stages. ACC (BCCP) tends to be the main active

subunit at the beginning of fatty acid biosynthesis, and subunit ACC (BC) tends to be the main active subunit at the end of fatty acid synthesis due to the altered expression pattern. This result was in agreement with previous reports and validated the complex regulation pattern of heteromeric ACCase (Shintani *et al.*, 1997; Thelen and Ohlrogge, 2002; Andre *et al*., 2012).

Unigene encoding GPAT had the highest expression level among all searched unigenes showed its activation among the lipid metabolic pathway in pecan (Fig. 4). GPAT is reported as another key enzyme of TAG biosynthesis in oil crops (Murata and Tasaka, 1997). Genes encoding GPAT had been cloned from *Arabidopsis thaliana* and several other plants (Weber *et al*., 1991; Johnson *et al.*, 1992; Nishida *et al.*, 1993). In *Brassica napus*, GPAT4 were founded essential for the normal development and oil accumulation of seed (Chen *et al.*, 2014). In this study, high expression levels were found of GPAT encoding unigenes showing the associated important roles of GPAT in pecan lipid metabolism.

The final step of fatty acid synthesis can be performed through two different routes catalyzed by DGAT (acyl-CoA dependent pathway) and/or PDAT and DGTA (acyl-CoA independent pathway) (Ohlrogge and Browse, 1995). This step is considered to be another rate-limiting step. Over-expression of AtDGAT cDNA enhanced oil deposition in wild-type *A. thaliana* (Jako *et al.*, 2001). Both DGAT and PDAT were found in yeast (Oelkers *et al.*, 2002), castor seeds, *Crepis palaestina* (Dahlqvist *et al.*, 2000), tree peony seed oil (Li *et al.*, 2015), and *Brassica napus* (Chen *et al*., 2015). Plants contain two very distinct families of DGAT genes. In our experiment, unigenes encoding both DGAT1 and DGAT2 were identified. Unigenes encoding PDAT were also found, but DGTA were absent. These results implied that both these two routes actively facilitated the incorporation of fatty acids into TAG in pecan kernel.

Transcription Factors in Pecan

Studies had proved that the lipid metabolism is regulated at the transcription level (Baud and Lepiniec, 2010). It is reported that many TFs were related to this process, such as ABI3, LEC1, LEC2, WRI1, FUS3, bZIP, AGL15, AP2-EREP, ARF, C3H, MYB118 and MYB115 (Santos-Mendoza *et al*., 2008; Baud and Lepiniec, 2009; Baud and Lepiniec, 2010). In this study, unigenes encoding TFs families of AP2-EREP, ARF, C3H, bZIP, MYB were found. The WRI1 of AP2-EREBP family was found regulating the fatty acid metabolic pathway in the *Arabidopsis* (To *et al.*, 2012; Marchive *et al.*, 2014). The bZIP can activate *FAD3* expression in *Arabidopsis* seed (Mendes *et al.*, 2013). These two TFs, together with ARF, and C3H were all have positive regulation effect on fatty acid synthesis in plant. DETFs of these four TF families founded in our experiment might play important roles in lipid synthesis and accumulation in pecan kernel. Homolog genes in MYB TF family were regarded to have negative control effect on the lipid synthesis (Li *et al.*, 2017). In this study, 17 down-regulated unigenes were annotated to the MYB family, which might play important roles in the negative transcriptional regulations to the lipid metabolism in pecan.

Conclusion

Using transcriptomic sequencing, we obtained 73,262 assembled unigenes. A putative pecan lipid metabolic pathway was reconstructed. Several enzymes play vital roles in pecan lipid formation; such as SAD is the key enzyme play important roles in the formation of high levels of oleic acid, relatively high linoleic acid content was contributed to the high expression levels of *FAD2* in pecan. Several enzymes controlled the rate-limiting steps in pecan lipid metabolism, including ACCase, GPAT, DGAT, and PDAT. We obtained a large transcriptome database that can be used for future research of seed development of pecan and related breeding programs.

Acknowledgements

We thank the financial fund of the National Natural Science Foundation of China (31200502). Additional funding of this work was provided by the Three Innovations Project of Forestry of Jiangsu Province (LYSX 201644) and the Jiangsu Agriculture Science and Technology Innovation Fund (CX 16 1015).

References

Andre, C., R.P. Haslam and J. Shanklin, 2012. Feedback regulation of plastidic acetyl-CoA carboxylase by 18: 1-acyl carrier protein in *Brassica napus*. *Proc. Natl. Acad. Sci. USA*, 109: 10107–10112

Auld, D.L., M.K. Heikkinen, D.A. Erickson, J.L. Sernyk and J.E. Romero, 1992. Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. *Crop Sci.*, 32: 657–662

Baud, S. and L. Lepiniec, 2010. Physiological and developmental regulation of seed oil production. *Prog. Lipid Res*., 49: 235–249

Baud, S. and L. Lepiniec, 2009. Regulation of *de novo* fatty acid synthesis in maturing oilseeds of *Arabidopsis*. *Plant Physiol. Biochem.*, 47: 448–455

Beedanagari, S.R., S.K. Dove, B.W. Wood and P.J. Conner, 2005. A first linkage map of pecan cultivars based on RAPD and AFLP markers. *Theor. Appl. Genet.*, 110: 1127–1137

Bouali, I., H. Trabelsi, I.B. Abdallah, A. Albouchi, L. Martine, S. Grégoire, G. Bouzaien, M. Gandour, S. Boukhchina and O. Berdeaux, 2013. Changes in fatty acid, tocopherol and xanthophyll contents during the development of Tunisian-Grown pecan nuts. *J. Amer. Oil Chem. Soc*., 90: 1869–1876

Bruner, A.C., S. Jung, A.G. Abbott and G.L. Powell, 2001. The naturally occurring high oleate oil character in some peanut varieties results from reduced oleoyl-PC desaturase activity from mutation of aspartate 150 to asparagine. *Crop Sci*., 41: 522–526

Buhr, T., S. Sato, F. Ebrahim, A. Xing, Y. Zhou, M. Mathiesen, B. Schweiger, A. Kinney, P. Staswick and T. Clemente, 2002. Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J*., 30: 155–163

Cahoon, E.B., S. Shah, J. Shanklin and J. Browse, 1998. A determinant of substrate specificity predicted from the acyl-acyl carrier protein desaturase of developing cat's claw seed. *Plant Physiol.*, 117: 593–598

Cerna-Cortés, J.F., J. Simpson, O. Martínez and R.A. Martínez-Peniche, 2003. Measurement of genetic diversity of native pecan [*Carya illinoensis* (Wangenh) K. Koch] populations established in central Mexico and correlation with dichogamous flowering using AFLP. *J. Food Agric. Environ.*, 1: 168–173

Chen, H., F.W. Wang, Y.Y. Dong, N. Wang, Y.P. Sun, X.Y. Li, L. Liu, X.D. Fan, H.L. Yin, Y.Y. Jing, X.Y. Zhang, Y.L. Li, G. Chen and H.Y. Li, 2012. Sequence mining and transcript profiling to explore differentially expressed genes associated with lipid biosynthesis during soybean seed development. *BMC Plant Biol.*, 12: 122

Chen, J., R.K. Tan, X.J. Guo, Z.L. Fu, Z. Wang, Z.Y. Zhang and X.L. Tan, 2015. Transcriptome analysis comparison of lipid biosynthesis in the leaves and developing seeds of *Brassica napus*. *PLoS One*, 10: e126250

Chen, X., G. Chen, M. Truksa, C.L. Snyder, S. Shah and R.J. Weselake, 2014. Glycerol-3-phosphate acyltransferase 4 is essential for the normal development of reproductive organs and the embryo in *Brassica napus*. *J. Exp. Bot.*, 65: 4201–4215

Conner, P.J. and B.W. Wood, 2001. Identification of pecan cultivars and their genetic relatedness as determined by randomly amplified polymorphic DNA analysis. *J. Amer. Soc. Hortic. Sci.*, 126: 474–480

Dahlqvist, A., U. Stahl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne and S. Stymne, 2000. Phospholipid: diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. USA*, 97: 6487–6492

Dussert, S., C. Guerin, M. Andersson, T. Joët, T.J. Tranbarger, M. Pizot, G. Sarah, A. Omore, T. Durand-Gasselin and F. Morcillo, 2013. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.*, 162: 1337–1358

Grauke, L.J., M.J. Iqbal and A.S. Reddy, 2003. Developing microsatellite DNA markers in pecan. *J. Amer. Soc. Hortic. Sci.*, 128: 374–380

Haas, B.J., A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. LeDuc, N. Friedman and A. Regev, 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, 8: 1494–1512

Hudthagosol, C., E.H. Haddad, K. McCarthy, P. Wang, K. Oda and J. Sabaté, 2011. Pecans acutely increase plasma postprandial antioxidant capacity and catechins and decrease LDL oxidation in humans. *J. Nutr.*, 141: 56–62

Jako, C., A. Kumar, Y. Wei, J. Zou, D.L. Barton, E.M. Giblin, P.S. Covello and D.C. Taylor, 2001. Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.*, 126: 861–874

Jia, X., H. Luo, M. Zhai, M. Qian, Y. Liu, Y. Li, Z. Guo and Y. Qiao, 2016. Dynamic changes and correlation analysis of nutrients content in 'Pawnee' (*Carya illinoensis*). *J. Fruit Sci.*, 9: 1120–1130

Johnson, T.C., J.C. Schneider and C. Somerville, 1992. Nucleotide sequence of acyl-acyl carrier protein: glycerol-3-phosphate acyltransferase from cucumber. *Plant Physiol.*, 99: 771–772

Kozaki, A., K. Mayumi and Y. Sasaki, 2001. Thiol-disulfide exchange between nuclear-encoded and chloroplast-encoded subunits of pea acetyl-CoA carboxylase. *J. Biol. Chem.*, 276: 39919–39925

Kachroo, A., J. Shanklin, E. Whittle, L. Lapchyk, D. Hildebrand and P. Kachroo, 2007. The *Arabidopsis* stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Mol. Biol.*, 63: 257–271

Kilaru, A., X. Cao, P.B. Dabbs, H.J. Sung, M.M. Rahman, N. Thrower, G. Zynda, R. Podicheti, E. Ibarra-Laclette, L. Herrera-Estrella, K. Mockaitis and J.B. Ohlrogge, 2015. Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea Americana* mesocarp. *BMC Plant Biol.*, 15: 203

Li, D., C. Jin, S. Duan, Y. Zhu, S. Qi, K. Liu, C. Gao, H. Ma, M. Zhang, Y. Liao and M. Chen, 2017. *MYB89* transcription factor represses seed oil accumulation. *Plant Physiol.*, 173: 1211–1225

Li, S.S., L.S. Wang, Q.Y. Shu, J. Wu, L.G. Chen, S. Shao and D.D. Yin, 2015. Fatty acid composition of developing tree peony (*Paeonia section* Moutan DC.) seeds and transcriptome analysis during seed development. *BMC Genom.*, 16: 208

Liu, Q., S.P. Singh and A.G. Green, 2002. High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiol.*, 129: 1732–1743

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

Marchive, C., K. Nikovics, A. To, L. Lepiniec and S. Baud, 2014. Transcriptional regulation of fatty acid production in higher plants: Molecular bases and biotechnological outcomes. *Eur. J. Lipid Sci. Technol.*, 116: 1332–1343

Mendes, A., A.A. Kelly, H. van Erp, E. Shaw, S.J. Powers, S. Kurup and P.J. Eastmond, 2013. bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating fatty acid desaturase3. *Plant Cell*, 25: 3104–3116

Muñoz-Mérida, A., J.J. González-Plaza, A. Cañada, A.M. Blanco, C. García-López Mdel, J.M. Rodríguez, L. Pedrola, M.D. Sicardo, M.L. Hernández, R. De la Rosa, A. Belaj, M. Gil-Borja, F. Luque, J.M. Martinez-Rivas, D.G. Pisano, O. Trelles, V. Valpuesta and C.R. Beuzón, 2013. *De novo* assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA Res.*, 20: 93–108

Murata, N. and Y. Tasaka, 1997. Glycerol-3-phosphate acyltransferase in plants. *Biochem. Biophs. Acta-Lipids Lipd Metabol.*, 1348: 10

Nishida, I., Y. Tasaka, H. Shiraishi and N. Murata, 1993. The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*. *Plant Mol. Biol.*, 21: 267–277

Oelkers, P., D. Cromley, M. Padamsee, J.T. Billheimer and S.L. Sturley, 2002. The DGA1 gene determines a second triglyceride synthetic pathway in yeast. *J. Biol. Chem.*, 277: 8877–8881

Ohlrogge, J. and J. Browse, 1995. Lipid biosynthesis. *Plant Cell*, 7: 957–970

Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai and J. Quackenbush, 2003. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19: 651–652

Rajaram, S., K. Burke, B. Connell, T. Myint and J. Sabaté, 2001. A monounsaturated fatty acid-richpecan-enriched diet favorably alters the serum lipid profile of healthy men and women. *J. Nutr.*, 131: 2275–2279

Roesler, K., D. Shintani, L. Savage, S. Boddupalli and J. Ohlrogge, 1997. Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. *Plant Physiol.*, 113: 75–81

Santos-Mendoza, M., B. Dubreucq, S. Baud, F. Parcy, M. Caboche and L. Lepiniec, 2008. Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.*, 54: 608–620

Shanklin, J. and C. Somerville, 1991. Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. *Proc. Natl. Acad. Sci. USA*, 88: 2510–2514

Shintani, D., K. Roesler, B. Shorrosh, L. Savage and J. Ohlrogge, 1997. Antisense expression and overexpression of biotin carboxylase in tobacco leaves. *Plant Physiol.*, 114: 881–886

Shintani, D.K. and J.B. Ohlrogge, 1995. Feedback inhibition of fatty acid synthesis in tobacco suspension cells. *Plant J.*, 7: 577–587

Storey, J.D. and R. Tibshirani, 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA*, 100: 9440–9445

Thelen, J.J. and J.B. Ohlrogge, 2002. Both antisense and sense expression of biotin carboxyl carrier protein isoform 2 inactivates the plastid acetyl-coenzyme A carboxylase in *Arabidopsis thaliana*. *Plant J.*, 32: 419–431

To, A., J. Joubès, G. Barthole, A. Lécureuil, A. Scagnelli, S. Jasinski, L. Lepiniec and S. Baud, 2012. WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*. *Plant Cell*, 24: 5007–5023

Vendrame, W.A., G. Kochert and H.Y. Wetzstein, 1999. AFLP analysis of variation in pecan somatic embryos. *Plant Cell Rep.*, 18: 853–857

Venglat, P., D. Xiang, S. Qiu, S.L. Stone, C. Tibiche, D. Cram, M. Alting-Mees, J. Nowak, S. Cloutier, M. Deyholos, F. Bekkaoui, A. Sharpe, E. Wang, G. Rowland, G. Selvaraj and R. Datla, 2011. Gene expression analysis of flax seed development. *BMC Plant Biol.*, 11: 74

Wang, L., Z. Feng, X. Wang, X. Wang and X. Zhang, 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26: 136–138

Wang, X., R. Xu, R. Wang and A. Liu, 2012. Transcriptome analysis of Sacha Inchi (*Plukenetia volubilis* L.) seeds at two developmental stages. *BMC Genom.*, 13: 716

Weber, S., F.P. Wolter, F. Buck, M. Frentzen and E. Heinz, 1991. Purification and cDNA sequencing of an oleate-selective acyl-ACP:sn-glycerol-3-phosphate acyltransferase from pea chloroplasts. *Plant Mol. Biol.*, 17: 1067–1076

Whittle, E., E.B. Cahoon, S. Subrahmanyam and J. Shanklin, 2005. A multifunctional acyl-acyl carrier protein desaturase from *Hedera helix* L. (English ivy) can synthesize 16-and 18-carbon monoene and diene products. *J. Biol. Chem.*, 280: 28169–28176

Yadav, N.S., A. Wierzbicki, M. Aegerter, C.S. Caster, L. Pérez-Grau, A.J. Kinney, W.D. Hitz, J.R. Booth Jr., B. Schweiger, K.L. Stecca, S.M. Allen, M. Blackwell, R.S. Reiter, T.J. Carlson, S.H. Russell, K.A. Feldmann, J. Pierce and J. Browse, 1993. Cloning of higher plant omega-3 fatty acid desaturases. *Plant Physiol.*, 103: 467–476

Yang, B. and H.P. Kallio, 2001. Fatty acid composition of lipids in Sea Buckthorn (*Hippophaë rhamnoides* L.) berries of different origins. *J. Agric. Food Chem.*, 49: 1939–1947

Yin, D., Y. Wang, X. Zhang, H. Li, X. Lu, J. Zhang, W. Zhang and S. Chen, 2013. *De novo* assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. *PLoS One*, 8: e73767

Zhang, J., Z. Mo, Y. Li, G. Wang, J. Xuan, X. Jia and Z. Guo, 2015. Cloning and functional analysis of MADS-box *CiMADS9* gene from *Carya illinoinensis*. *Acta Hortic. Sin.*, 42: 1049–1056

Zheng, Y., C. Jiao, H. Sun, H.G. Rosli, M.A. Pombo, P. Zhang, M. Banf, X. Dai, G.B. Martin, J.J. Giovannoni, P.X. Zhao, S.Y. Rhee and Z. Fei, 2016. iTAK: A program for genome-wideprediction and classification of plant transcription factors, transcriptional regulators, and proteinkinases. *Mol. Plant*, 9: 1667–1670

(Received 05 June 2017; Accepted 27 December 2017)