果 树 学 报 2017,34(9):1084-1094
Journal of Fruit Science

DOI:10.13925/j.cnki.gsxb.20160444

核桃种子油脂转化期转录组分析

杨 丽,陈 虹,潘存德*,尚刘群,方 淼,赵明明

(新疆农业大学林学与园艺学院·新疆教育厅干旱区林业生态与产业技术重点实验室,乌鲁木齐 830052)

摘 要:【目的】从分子水平上探索核桃(*Juglans regia* L.)种子油脂转化时期的脂肪酸合成相关基因的表达模式。【方法】基于Illumina HiSeq™ 4000高通量测序平台,以新疆早实核桃'新新2号'(*J. regia* 'Xinxin2')品种种子的种仁为材料,对核桃种子油脂转化期3个不同阶段(花后60~70 d,简称G1;花后90~100 d,简称G2;花后120~130 d,简称G3)进行转录组测序比较分析。【结果】组装得到174 545条Unigene,其中G2 *vs* G1、G3 *vs* G2和G3 *vs* G1分别有9 408、8 316、6 398条上调表达的差异基因和11 916、10 485、5 218条下调表达的差异基因。代谢通路富集分析发现,3个阶段之间差异基因富集最显著的是脂肪酸生物合成代谢途径。随机挑选8个基因进行荧光定量验证,结果与测序数据一致,表明核桃种子油脂转化期的转录组测序结果可信度高。不同时期基因差异表达富集度最高的为脂肪酸生物合成代谢通路。在脂肪酸生物合成途径上乙酰辅酶A的羧基转移酶α亚基基因*accA*、生物素羧基载体蛋白基因*accB*和生物素羧化酶亚基基因*accC*的表达量在G1到G3持续上升,β-酮酰ACP合酶Ⅱ基因*fabF*、β-酮酰-ACP还原酶基因*fabG*和硬脂酰-ACP去饱和酶基因*FAB2*的表达量在G1到G3也持续上升,并且G2与G1之间上调表达的差异基因最多。【结论】核桃种子的油脂转化在G2阶段脂肪酸合成最活跃,脂肪酸生物合成途径与油脂合成有关的*accA*、*accB*、*accC*、*fabF*、*fabG*、*fabI*和*FAB2*基因均呈上调表达。

关键词:核桃;RNA-seq;转录组;差异基因表达;脂肪酸生物合成

中图分类号:S664.1 文献标志码:A 文章编号:1009-9980(2017)09-1084-11

Transcriptome analysis for walnut seed during oil synthesis period

YANG Li, CHEN Hong, PAN Cunde*, SHANG Liuqun, FANG Miao, ZHAO Mingming

(College of Forestry and Horticulture, Xinjiang Agricultural University·Key Laboratory of Forestry Ecology and Industry Technology in Arid Region, Education Department of Xinjiang, Urumqi 830052, Xinjiang, China)

Abstract:【Objective】Walnut seeds accumulate a lot of fat during the oil synthesis period. It is of great importance for us to understand the genes involved in lipid metabolism for genetic manipulation. Transcriptomic analyses were employed to explore the expression pattern of the genes related to lipid metabolism during the course of lipid transition period in the seeds of walnut.【Methods】In this study, *Juglans regia* 'Xinxin 2' in Xinjiang was used as materials, TIANGEN (DP441) RNA extraction kit was used for extracting mRNA. The seeds were sampled at three different stages of oil synthesis (60 to 70 days after flowering referred to as G1; 90 to 100 days after flowering referred to as G2; 120 to 130 days after flowering referred to as G3). The sequencing of transcriptome was performed by Beijing Novogene Bioinformatics Technology Company. The sequencing platform was Illumina HiSeq™ 4000. The clean reads were obtained by removal of the reads containing the sequence of the joint, the reads in which the N ratio was over 10% and the reads of low quality. There is no reference genome to the walnut, so we used the Trinity software to do the de novo sequencing using the data obtained from each gene as transcriptome, and the longest transcript of each gene was taken as unigene. The sequence features of transcript and the length of the unigene were statistically analyzed respectively for late analysis. Protein Annotation information about

收稿日期:2017-01-06 接受日期:2017-04-18

基金项目:国家自然科学基金(31460210)

作者简介:杨丽,女,在读硕士研究生,研究方向:果树生物技术。E-mail:814662119@qq.com

*通信作者 Author for correspondence. Tel: 0991-8763728,E-mail: pancunde@163.com

the unigene was obtained by sequence alignment in Nr, Nt, KEGG, Swiss-Prot, PFAM, GO and KOG protein databases. The transcripts of Trinity splicing were used as reference sequence (ref), the RSEM software was used (bowtie 2, the parameter is mismatch 0) for mapping the clean reads and ref of each sample, and accounting for the number of readcount for clean reads to each unigene. The data obtained from the gene expression level of readcount were used as input data for differentially expressed genes and the differentially expressed genes were screened by DESeq software($P< 0.05$).【Results】Through the RNA-Seq analysis of the seeds of walnut during the oil transformation period, 174 545 unigenes were obtained, the average length of the unigene was 644 bp and the N50 was 1 050 bp, among them, the length of unigene sequence between 1 000 and 2 000 bp accounted for 8.51%, and the length of sequence over 2 000 bp accounted for 6.52%. Annotation analysis of unigenes indicated that 94 133 transcripts were homologous with those of other species in the public protein database; however, 80 412 sequences were not annotated and might be walnut-specific. Walnut seeds with 69 235 unigenes were annotated with Nr protein database. According to the classification of KOG, 27 333 unigenes were divided into functional categories, and the function classification of GO are 51 769 unigenes. In addition, for KEGG pathway classification, 26 946 unigenes were annotated. There were 9 408, 8 316 and 6 398 differential unigenes in G2 *vs* G1, G3 *vs* G2 and G3 *vs* G1were up-regulated, 11 916, 10 485 and 5 218 differential unigenes in G2 *vs* G1, G3 *vs* G2 and G3 *vs* G1were down-regulated, respectively, which were annotated to the KEGG pathway at each stage of lipid transformation. It was found that the highest concentration of the genes related to the metabolic pathway of the fatty acid bio-synthesis was mostly obvious at the stages of G1, G2 and G3. There were 34 differentially expressed genes of G2 *vs* G1, G3 *vs* G2 and G3 *vs* G1 of fatty acid biosynthetic pathway. In the fatty acid biosynthetic pathway, the expression of carboxylase subunit gene (*accA*), biotin carboxyl carrier protein gene (*accB*) and biotin carboxylase subunit gene (*accC*) of acetyl coenzyme A were continuously up-regulated. The expression of β-ketoacyl ACP synthase gene (*fabF*) and β-ketoacyl ACP reductase gene (*fabG*) at G1 to G3 period continued to increase. The expression of stearyl-ACP desaturase gene (*FAB2*) at G3 phase was higher than that at G2 phase. qRT-PCR was used for the quantitative determination of 8 randomly selected genes, the results were consistent with the sequencing data.【Conclusion】Most of the unigene of the walnut seeds had higher matching degree with the known genes in the existing database. The oil synthesis of walnut seeds was mostly concentrated at the G2 stage. The *accA*, *accB*, *fabF*, *fabG*, *fabI* and *FAB2* genes involved in the lipid biosynthesis were all up-regulated.

Key words: Walnut; RNA-seq; Transcriptome; Differentially expressed genes; Fatty acid biosynthesis

核桃（*Juglans regia* L.）隶属胡桃科核桃属，是世界四大坚果树种之一[1]，其坚果种仁不仅富含维持人体新陈代谢和身体健康所必需的重要矿质元素 Mg 和 P[2-3]，而且还富含食用油的重要成分油脂和不饱和脂肪酸，因此备受食品和保健领域的关注[4]。目前与核桃相关的研究主要集中在种质资源收集与利用[5]、优质丰产栽培[6]、种仁营养保健成分分析[2-3]和化感作用[7]等方面。由于核桃坚果种仁的主要成分是油脂，因此其种子的油脂合成代谢对于改善核桃坚果品质具有重要作用。油料作物油脂合成代谢涉及的关键酶主要有乙酰辅酶 A 羧化酶（acetyl-CoA catboxylase）、脂肪酸合酶（fatty acid synthase）、脂肪酸脱饱和酶（fatty acid desaturase）和甘油三磷酸酰基转移酶（Glycerol-3-Phosphate Acyltransferase）等[8]。在拟南芥（*Arabidopsis thaliana*）[9]、花生（*Arachis hypogaea*）[10]、白菜型油菜（*Brassica carinata*）[11]等 1 a（年）生作物和油桐（*Vernicia fordii*）[12]、油茶（*Camellia oleifera*）[13]等油料树种上的研究已经证实，参与油脂合成代谢的关键酶基因在种子油脂合成过程中具有重要作用[9]。但是，由于核桃基因组、转录组信息相对缺乏，造成核桃油脂合成相关酶的分子标记开发、基因克隆及其调控基因的研究相对滞后，因

此极大地限制了对核桃种子油脂合成代谢涉及的关键酶基因的系统研究。

转录组是指研究对象在某一时段或特定环境条件下产生的所有转录物的集合[14]。转录组测序分析能够快速地获得研究对象在某一状态下的基因转录信息和样品间的表达差异,揭示器官、细胞特定生物学过程的分子机制[14-15]。因此,对于缺乏基因组信息的物种而言,可利用转录组测序分析获得大量转录本信息,从中发掘重要功能基因,进而揭示其优良特性。近年来,转录组测序分析已运用到油桐(*Vernicia fordii*)[12]、油茶(*Camellia oleifera*)[13]和麻风树(*Jatropha curcas*)[16]等关键基因的筛查中,并获得了这些油料树种参与种子油脂合成的ω-6脂肪酸脱饱和酶基因、乙酰辅酶A羧化酶基因等关键基因。为了加深对核桃种子油脂合成代谢途径的认识,笔者以新疆早实核桃'新新2号'(*J. regia* 'Xinxin2')品种种子的种仁为材料,采用Illumina HiSeq™ 4000高通量测序平台,应用转录组测序分析筛查核桃种子油脂转化过程中的差异表达基因,探讨涉及油脂合成代谢的分子机制,其结果不仅为阐释核桃种子成油机制提供理论依据,而且还可借助分子生物学技术手段,为核桃的品种选育工作奠定基础。

1 材料和方法

1.1 材料

'新新2号'(*J. regia* 'Xinxin2')核桃品种种子的种仁。种仁样品采自新疆阿克苏地区乌什县阿克托海乡吉格代力克村核桃生产园,海拔1 394 m。核桃生产园面积3 hm^2,主栽品种为'新新2号',授粉品种为与主栽品种配套的'温185'(*J. regia* 'Wen185'),树龄12 a,南北行向栽植,栽植株行距5 m×6 m,园内栽培管理条件一致,树体健康。采用典型抽样法在选定的核桃生产园内标定'新新2号'种仁样品采集样株10株。依据核桃果实生长发育动态及种子内部油体发育状况[17-18],于2016年'新新2号'花后60~70 d(种仁油脂转化期初始阶段,简称G1)、90~100 d(种仁油脂转化期中间阶段,简称G2)和120~130 d(种仁油脂转化期末尾阶段,简称G3)3个阶段,在标定的种仁样品采集样株上逐日随机采摘10个胚胎发育正常的果实。将采摘的果实剥取种仁,置于液氮中作为转录组测序RNA提取的样品保存备用。

1.2 总RNA提取与测序文库构建

从保存的G1、G2和G3阶段备用样品中选取试验材料样品,将试验材料样品随机分成3份视为重复,分别采用TIANGEN(DP441)多糖多酚植物总RNA提取试剂盒进行试验材料样品RNA的提取,并对提取的RNA进行质量检测。转录组测序委托北京诺禾致源生物信息公司完成,测序平台为Illumina HiSeq™ 4000。

1.3 测序数据分析

1.3.1 *数据处理* 得到测序数据后,首先去除含测序接头的测序片段(reads)、N比例大于10%和低质量的reads,然后利用获得的高质量序列数据(clean reads),统计clean reads的数量、总长度、Q20、GC含量等。

1.3.2 *序列组装* 因核桃无可以参照的基因组,故采用Trinity组装软件对获得的测序数据进行从头组装。取每条基因中最长的转录本作为Unigene(去冗余之后得到的基因序列),对组装得到的转录本和Unigene的长度分别进行统计,并以此进行后续分析。

1.3.3 *Unigene的功能注释* 将获得的核桃Unigene序列在Nr(NCBI官方蛋白序列数据库,e-value=1×10^{-5})、Nt(NCBI官方核酸序列数据库,e-value=1×10^{-5})、KEGG(京都基因与基因组百科全书,e-value=1×10^{-10})、Swiss-Prot(经过注释的蛋白质序列数据库,e-value=1×10^{-5})、PFAM(蛋白质家族结构域数据库,e-value=0.01)、GO(基因本体,e-value=1×10^{-6})和KOG(真核生物蛋白相邻类的聚簇,e-value=1×10^{-3})等核酸、蛋白质数据库中进行序列比对,获得Unigene序列对应的蛋白质注释信息。

1.3.4 *差异基因筛选、KEGG富集分析* 以Trinity拼接得到的转录本为参考序列(ref),采用RSEM软件[20](bowtie 2参数mismatch 0),将每个样品的clean reads与ref比对(mapping),统计每个样品中clean reads比对到每个Unigene上的reads数目(readcount)。以基因表达水平分析中得到的readcount数据作为基因差异表达的输入数据,用DESeq软件[19]进行差异表达基因的筛选($P<0.05$)。

将$P<0.05$的Pathway设定为在差异基因中显著富集的Pathway,使用KOBAS2.0软件进行Pathway富集分析。脂肪酸生物合成途径差异基因维恩图用Venny 2.1.0绘制。利用Cluster 3.0聚类分析软件进

行差异基因聚类分析。

1.4　差异基因表达验证

随机选取8个差异表达基因，以1.2中提取的RNA样品为模板、使用AMV反转录酶（TaKaRa公司生产）进行反转录；采用荧光（SYBRGreen）染料法进行定量PCR（qRT-PCR）验证基因的表达量。使用Oligo 6.0结合Primier 5.0引物设计软件进行引物设计，委托上海生工生物工程股份有限公司合成引物，选用*β-actin*为内参。待验证基因和*β-actin*内参基因及其特异性引物序列见表1。采用$2^{-\Delta\Delta Ct}$计算基因的差异倍数。每个样品设置3次技术重复。

表1　实时荧光定量PCR所用基因及其引物序列

Table 1　Unigene and primers for quantitative RT-PCR analysis

基因 Unigene	注释 Description	引物序列 Primer sequence（5'-3'）
Cluster-6295.114135	glycerol-3-phosphate dehydrogenase [NAD(+)]-like isoform X1 [*Citrus sinensis*]	F:TCTCTCGCTCCATGTCTC R:GATTTGTGAGCAACCCCAT
Cluster-6295.84133	glycerol-3-phosphate acyltransferase 8 [*Vernicia fordii*]	F:ACCCCATCGTAATTGCCAT R:CAGACCACGAGATCACCCTT
Cluster-6295.87458	omega-6 fatty acid desaturase [*Rhus chinensis*]	F:ATGCCTTCAGTGACTACCAA R:ACAAACACTTCATCTCGCTCT
Cluster-6295.99764	oleosin [*Juglans regia*]	F:CAGCACGGTCATCGCCTT R:CGCCGCTACTCCGAATCCAC
Cluster-6295.91376	biotin carboxyl carrier protein of acetyl-CoA carboxylase 2, chloroplastic [*Prunus mume*]	F:TGGCTGGAACATTCTATCGT R:CTAACTGGTTTCGCATCCTC
Cluster-6295.66098	phosphoesterase family protein [*Populus trichocarpa*]	F:ATGACGCCGGAATATCCTT R:GCTCCAACACGACGTAGCC
Cluster-6295.104173	microsomal omega-3 fatty acid desaturase [*Betula pendula*]	F:CATACAGCAACTTGTTCGC R:AAATCAAGTGAGGAACACCG
Cluster-6295.94772	GDSL esterase/lipase 7-like [*Glycine max*]	F:ATGTTCTTCCTGCGGCTA R:AATTGTTTTGCCATTGGTGA
J. regia β-actin		F:CTCTTCCAGCCATCCATGATCG R:CCACTGAGGACAATATTGCCAT

2　结果与分析

2.1　测序数据质量及转录本组装情况分析

G1、G2和G3 RNA测序中得到的Raw reads经去除测序接头、重复冗余、低质量等处理后，获得clean reads，其Q20高质量序列均在95.28%以上，GC含量均在45.41%以上，表明核桃转录组测序数据量和质量较好，能够满足后续分析研究的要求。

进一步对经Trinity拼接得到的转录本及Unigene分别进行长度统计（表2）。有转录本279 675条，平均长度为1 038 bp，N50为2 017 bp，其中：200~500 bp的序列占49.12%，≥500~1 000 bp的序列占18.08%，≥1 000~2 000 bp的序列占16.63%，≥2 000 bp的序列占16.17%。得到Unigene 174 545条，平均

表2　测序数据组装和统计

Table 2　Assembly and statistics of sequencing data

转录本 Transcript			Unigene		
转录本长度 Transcript length/bp	总数 Total number	百分比 Percentage/%	Unigene长度 Unigene length/bp	总数 Total number	百分比 Percentage/%
200~500	137 389	49.12	200~500	119 076	68.22
≥500~1 000	50 563	18.08	≥500~1 000	29 238	16.75
≥1 000~2 000	46 511	16.63	≥1 000~2 000	14 857	8.51
≥2 000	45 212	16.17	≥2 000	11 374	6.52
总数 Total number	279 675		总数 Total number	174 545	
平均长度 Mean length	1 038		平均长度 Mean length	644	
N50长度 N50 length	2 017		N50长度 N50 length	1 050	

长度为644 bp，N50为1 050 bp，其中：200~500 bp的序列占68.22%，≥500~1 000 bp的序列占16.75%，≥1 000~2 000 bp的序列占8.51%，≥2 000 bp的序列占6.52%。

2.2 转录组Unigene功能注释

拼接得到的174 545个Unigene序列分别在Nr、Nt、KO、SwissProt、PFAM、GO和KOG等蛋白质数据库中进行序列比对（图1），共有94 133个转录本被注释，占所有Unigene的53.93%，其中：有69 235个转录本被注释到Nr数据库，占所有Unigene的39.66%；有26 946个转录本被注释到KO数据库，占所有Unigene的15.43%；有55 687个转录本被注释到SwissProt数据库，占所有Unigene的31.90%；有51 769个转录本被注释到GO数据库，占所有Unigene的29.65%；有27 333个转录本被注释到KOG数据库，占所有Unigene的15.65%。

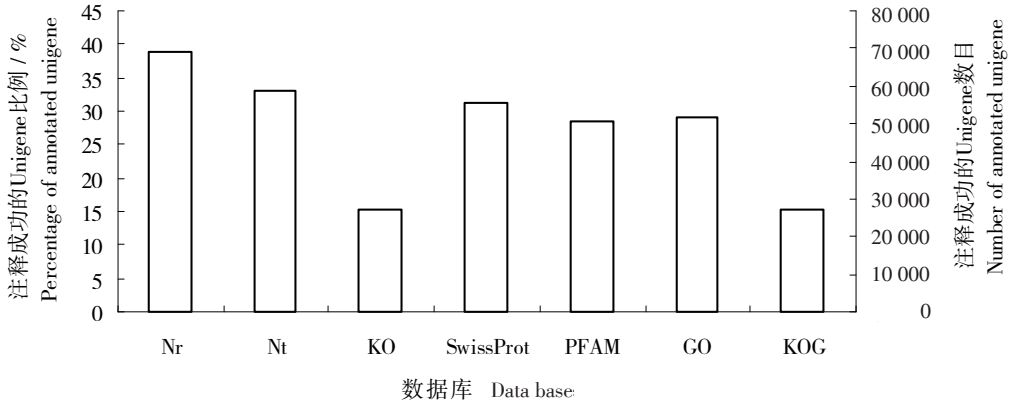

图1 Unigene注释

Fig. 1 Annotation results of Unigene

2.3 核桃种子油脂转化期不同阶段差异基因

G1、G2和G3的表达谱进行两两比对，筛选出差异基因，并统计差异基因的差异表达量（图2）。通过基因表达量计算，在G2 *vs* G1、G3 *vs* G1和G3 *vs* G2中分别有9 408、8 316和6 398条上调表达的差异基因；有11 916、10 485和5 218条下调表达的差异基因。

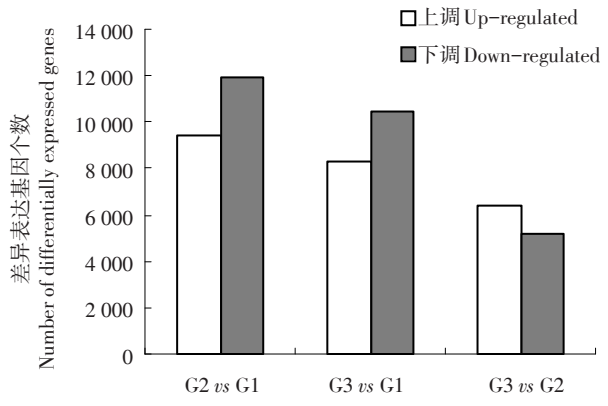

图2 核桃种子油脂转化期不同阶段差异表达基因

Fig. 2 Differentially expressed genes in different stages of walnut seed oil transformation

2.3.1 *差异基因KEGG代谢通路注释与分类* 对筛查出的差异表达基因进行KEGG代谢通路注释，从中选取差异表达基因富集度达到显著水平（$P<0.05$）且显著性水平排前5位的代谢通路。发现核桃种子油脂转化期3个不同阶段（G1、G2和G3）之间差异表达基因富集度最高的代谢通路均为脂肪酸生物合成（fatty acid biosynthesis），在G2与G1之间有118个差异表达基因富集，其中上调表达87个，下调表达31个（表3）；在G3与G2之间有68个差异表达基因富集，其中上调表达29个，下调表达39个（表4）；在G3与G1有123个差异表达基因富集，其中上调表达80个，下调表达43个（表5）。

2.3.2 *差异基因表达量验证* 从差异基因中随机选择的8个差异基因，经q-PCR验证（图3），Cluster-6295.114135、Cluster-6295.87458、Cluster-6295.99764、Cluster-6295.91376、Cluster-6295.66098、Cluster-6295.104173这6个基因的q-PCR相对表达水平与RNA-seq相对表达水平的趋势一致，均表现为从G1→G2→G3表达水平逐渐上升。Cluster-6295.84133和Cluster-6295.94772的q-PCR相对表达水平与RNA-seq相对表达水平的趋势一致，均表现为从G1→G2→G3表达水平逐渐降低。

表 3　G2 与 G1 差异表达基因 KEGG 富集($P<0.05$)

Table 3 KEGG enrichment of differentially expressed genes between G2 and G1 ($P<0.05$)

通路ID Pathway ID	通路 Pathway	富集基因 Enriched gene	上调 Up-regulation	下调 Down-regulation	P值 P-value
ko00061	Fatty acid biosynthesis	118	87	31	0.001 346
ko04075	Plant hormone signal transduction	330	174	156	0.007 785
ko00520	Amino sugar and nucleotide sugar metabolism	226	72	154	0.028 508
ko00071	Fatty acid degradation	120	74	46	0.030 239
ko00620	Pyruvate metabolism	198	122	76	0.030 690

表 4　G3 与 G2 之间差异表达基因 KEGG 富集($P<0.05$)

Table 4 KEGG enrichment of differentially expressed genes between G3 and G2 ($P<0.05$)

通路ID Pathway ID	通路 Pathway	富集基因 Enriched gene	上调 Up-regulation	下调 Down-regulation	P值 P-value
ko00061	Fatty acid biosynthesis	68	29	39	5.04×10^{-6}
ko04075	Plant hormone signal transduction	172	83	89	1.52×10^{-5}
ko00270	Cysteine and methionine metabolism	113	77	36	1.85×10^{-5}
ko00071	Fatty acid degradation	67	25	42	0.000 407
ko00450	Selenocompound metabolism	31	18	13	0.001 358

表 5　G3 与 G1 之间差异表达基因 KEGG 富集($P<0.05$)

Table 5 KEGG enrichment of up-regulation expressed genes between G3 and G1 ($P<0.05$)

通路ID Pathway ID	通路 Pathway	富集基因 Enriched gene	上调 Up-regulation	下调 Down-regulation	P值 P-value
ko00061	Fatty acid biosynthesis	123	80	43	0.001 694
ko04075	Plant hormone signal transduction	354	182	172	0.003 104
ko00950	Isoquinoline alkaloid biosynthesis	60	37	23	0.004 883
ko04141	Protein processing in endoplasmic reticulum	394	175	219	0.015 630
ko00430	Taurine and hypotaurine metabolism	34	14	20	0.016 038

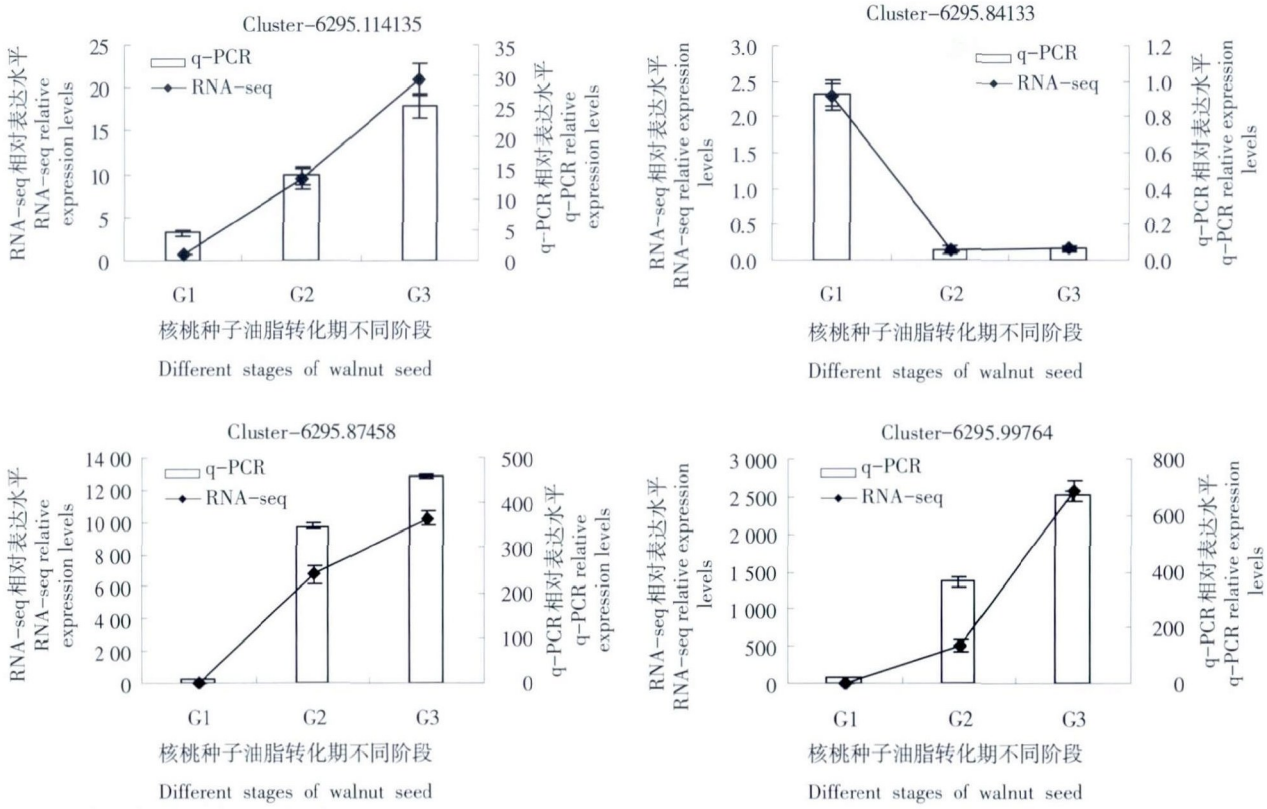

图 3　差异基因相对表达水平

Fig. 3 Levels of differential gene expression

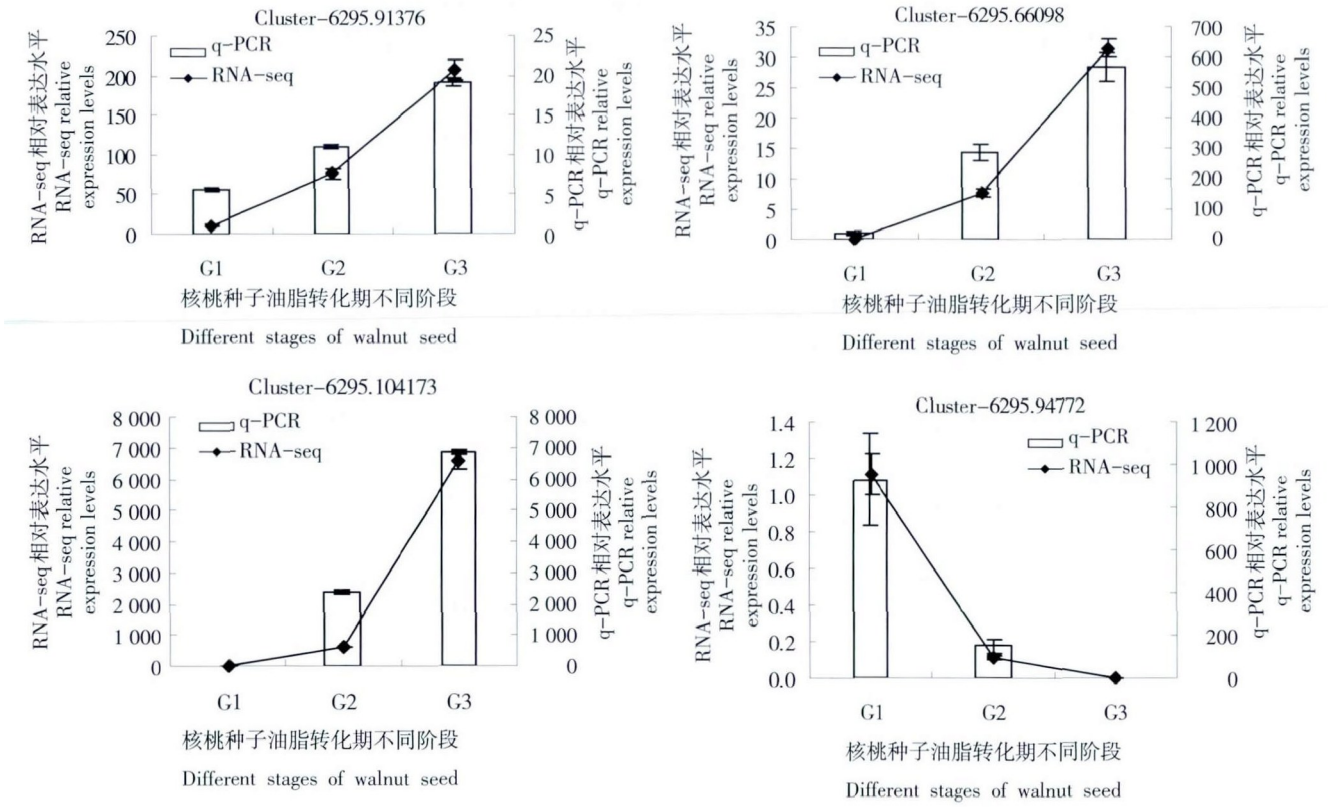

图3 （续）

Fig. 3 （Continued）

2.4 核桃种子脂肪酸生物合成途径关键酶基因表达模式

根据代谢通路富集分析，以差异基因显著富集的脂肪酸生物合成（fatty acid biosynthesis）途径为筛查对象，发现在核桃种子油脂转化期间有153个差异基因富集，其中34个差异基因为G2 *vs* G1、G3 *vs* G2和G3 *vs* G1所共有（图4）。

将G2 *vs* G1、G3 *vs* G2和G3 *vs* G1之间在脂肪酸生物合成途径上共有的34个差异基因，根据表达量进行Cluster聚类，这34个差异基因被分成3类，A1：G1到G3表达量逐渐下降，有10个差异基因。A2：G1到G3表达量持续上升，有17个差异基因。A3：G1到G2表达量逐渐上升，G2到G3表达量逐渐下降，有7个差异基因（图5）。

在脂肪酸生物合成途径中，有17条持续上调的差异基因（图5，表6）。Cluster-6295.63289被注释到生物素羧化酶（biotin carboxylase）基因*ACAC*。Cluster-6295.63289被注释到α-羧基转移酶（α-carboxyltransferase）的基因*accA*。Cluster-6295.85582和Cluster-6295.91376被注释到生物素羧基载体蛋白（biotin carboxyl carrier protein）基因*accB*。Cluster-6295.93304被注释到生物素羧化酶亚基（biotin carboxylase subunit）基因*accC*。Cluster-6295.74924被注释到β-酮酰ACP合酶Ⅱ（β-ketoacyl-ACP synthase Ⅱ）基因*fabF*。Cluster-6295.64644被注释到β-酮酰-ACP还原酶（ketoacyl-ACP reductase）基因

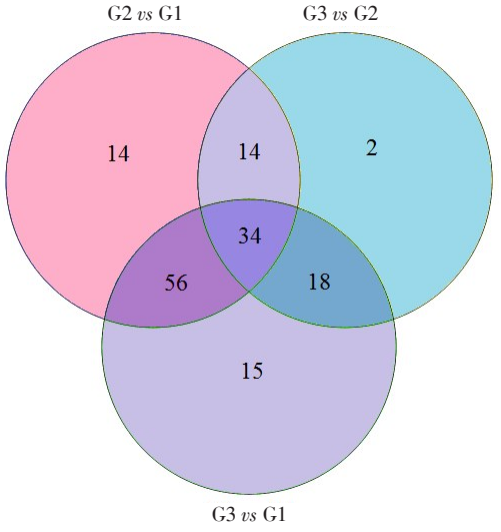

图4 脂肪酸生物合成途径差异基因维恩图

Fig. 4 Venn diagram of differential genes in the fatty acid biosynthesis

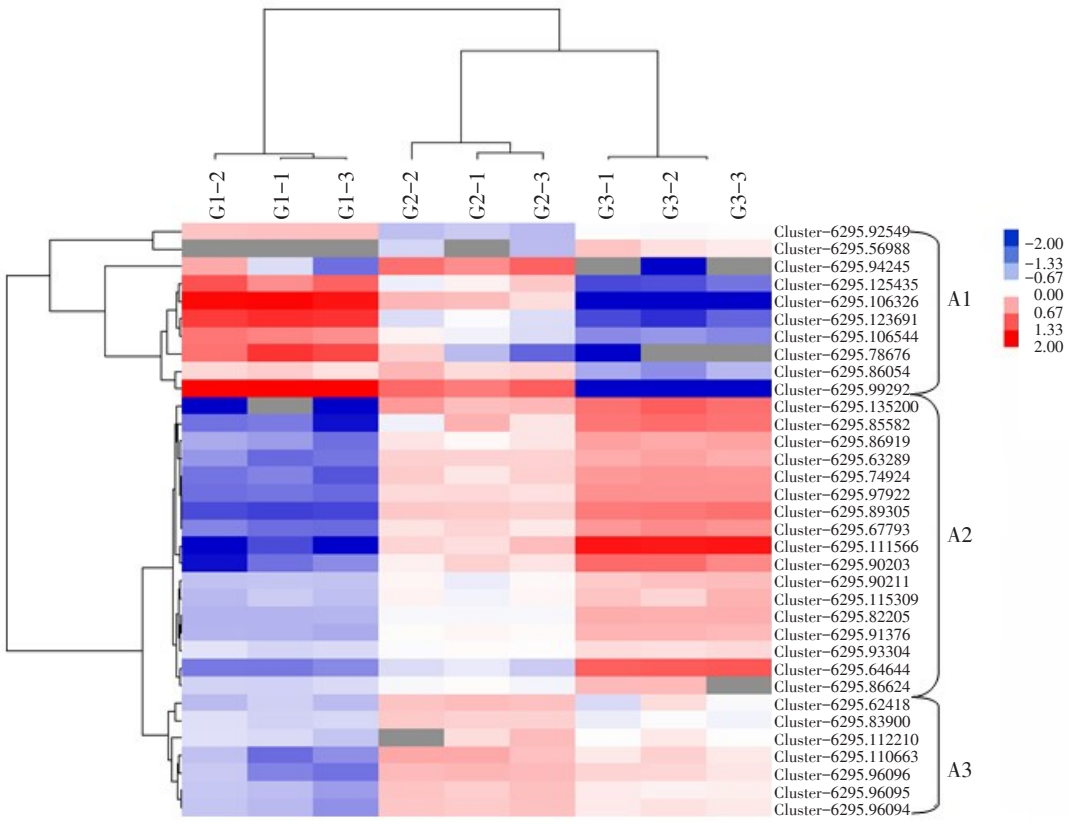

图 5　脂肪酸生物合成途径差异基因聚类分析

Fig. 5　Cluster analysis of differential genes in the fatty acid biosynthesis pathway

fabG。Cluster-6295.90211、Cluster-6295.86919、Cluster-6295.82205、Cluster-6295.97922、Cluster-6295.89305、Cluster-6295.86624和Cluster-6295.90203被注释到硬脂酰-ACP去饱和酶(stearoyl-ACP desaturase)基因*FAB2*。Cluster-6295.111566、Cluster-6295.115309和Cluster-6295.67793被注释到长链酰基辅酶A合成酶(long-chain acyl-CoA synthetase)基因*ACSL*。

表 6　脂肪酸生物合成途径上调基因

Table 6　The up-regulated genes of fatty acid biosynthesis pathway

基因ID Gene ID	KEGG编号 KEGG No.	EC编号 EC No.	基因名称 Gene name	蛋白名称 Protein name
Cluster-6295.74924	K09458	2.3.1.179	*fabF*	β-ketoacyl-ACP synthase Ⅱ
Cluster-6295.90211	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.63289	K01962	6.4.1.2	*accA*	Acetyl-CoA carboxylase carboxyl transferase subunit alpha
Cluster-6295.135200	K11262	6.3.4.14	*ACAC*	Acetyl-CoA carboxylase / biotin carboxylase 1
Cluster-6295.111566	K01897	6.2.1.3	*ACSL*	Long-chain acyl-CoA synthetase
Cluster-6295.115309	K01897	6.2.1.3	*ACSL*	Long-chain acyl-CoA synthetase
Cluster-6295.85582	K02160	Not enzyme	*accB*	Biotin carboxyl carrier protein
Cluster-6295.86919	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.64644	K00059	1.1.1.100	*fabG*	Ketoacyl-ACP reductase
Cluster-6295.82205	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.97922	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.89305	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.67793	K01897	6.2.1.3	*ACSL*	Long-chain acyl-CoA synthetase
Cluster-6295.93304	K01961	6.4.1.2	*accC*	Acetyl-CoA carboxylase, biotin carboxylase subunit
Cluster-6295.86624	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.90203	K03921	1.14.19.2	*FAB2*	Stearoyl-ACP desaturase
Cluster-6295.91376	K02160	Not enzyme	*accB*	Biotin carboxyl carrier protein

3 讨 论

核桃作为重要的油料作物，其分子遗传方面的研究主要集中在种质资源收集和相关分子标记方面[7,22]，种子生长发育过程中基因的表达研究相对较少，而高通量测序对于未开展基因组测序的物种而言，是获得重要基因表达的可取途径，并在芝麻（*Sesamum indicum*）[23]、油桐[12]、油茶[13]等油料作物的研究中得到成功运用。笔者通过RNA-Seq测序技术的应用，获得转录本279 675条，Unigene 174 545条。Unigene平均长度为644 bp，N50为1 050 bp，表明序列组装的质量和长度可以满足转录组分析的基本要求。Unigene的注释表明，有94 133个转录本在公共蛋白质数据库中具有同源物种，但有80 412个转录本没有被注释，这可能与序列片段过短、注释信息的暂时缺乏、核桃物种或组织转录组的特异有关，在很多木本植物转录组分析中也出现过类似的情况，如油桐[12]、茶树（*Camellia sinensis*）[24]等。对核桃种子油脂转化期3个不同阶段的表达谱进行两两比对，发现G2与G1之间的差异基因最多，而G2是核桃种子脂肪酸合成速率最快的阶段[17]，说明转录组分析结果与已有的生理研究结果相一致。核桃种子油脂转化期3个不同阶段之间差异表达基因富集度最高的代谢通路均为脂肪酸生物合成途径，其中在G2与G1之间有上调表达的差异基因最多，为87个，进一步表明G2阶段脂肪酸合成最活跃。进一步进行差异基因qRT-PCR的验证，q-PCR相对表达水平与RNA-seq相对表达水平的趋势一致，表明本研究关于核桃种子油脂转化期的转录组测序结果可信度高。

油脂合成代谢对于改善油料作物品质具有极其重要的作用，其脂肪酸生物合成途径是油脂合成代谢的前提[8]，也是目前了解最为清楚的途径。对核桃脂肪酸生物合成途径的差异基因分析表明，与乙酰辅酶A羧化酶（ACCase）组成相关的基因中，羧基转移酶α亚基基因*accA*、生物素羧基载体蛋白基因*accB*和生物素羧化酶亚基基因*accC*的表达量在G1到G3持续上升，但羧基转移酶β亚基（β-carboxyltransferase，β-CT）基因*accD*并没有持续上调，这与陈昊等[12]在油桐（*Vernicia fordii*）种子转录组的油脂合成途径研究中得到的结果类似，说明ACCase的亚基在核桃油脂转化过程中也是非等量存在，据此推测异质型ACCase的酶活性会随着亚基之间比例的变化而变化，从而影响脂肪酸的合成。随后在脂肪酸合酶复合体的作用下，1个乙酰CoA分子与多个丙二酰CoA分子经过缩合、羰基还原、脱水、再次还原，每经过1个循环将β-酮酯酰基ACP变成比上一个循环多2个碳原子的酰基ACP，生成碳链长度为12~18的脂肪酸[25]。这4个反应分别由β-酮酰ACP合酶、β-酮酰-ACP还原酶、β-羟酰-ACP脱水酶和烯酰-ACP还原酶4类酶催化完成[8]，其中β-酮酰ACP合酶Ⅱ（KASⅡ）基因和β-酮酰-ACP还原酶基因的表达量从G1到G3持续上升，KASⅡ催化棕榈酰基ACP（palmitoyl-ACP，C16-ACP）与丙二酰ACP聚合，生成硬脂酰基ACP（stearoyl-ACP，C18-ACP）[26]，KASⅡ基因缺失的拟南芥（*Arabidopsis thaliana*）*fab1-2*突变体与野生型相比其C18脂肪酸降低，而C16脂肪酸含量增高[27]。在硬脂酰基ACP在硬脂酰去饱和酶（SAD）的作用下，进一步生成含有一个顺式不饱和双键的油酰基ACP[28]。这与核桃种子脂肪酸以C18脂肪酸为主相对应[17]。油料作物中的SAD基因表达量一般均较高，因此在成熟种子中C16∶0和C18∶0等饱和脂肪酸含量较低[29]，羽扇豆（*Lupinus micranthus*）的SAD基因在烟草（*Nicotiana tabacum*）中超表达，获得的烟草转基因植株的油酸（C18∶1Δ9）含量提高了11%[30]。核桃种子中的SAD基因*FAB2*的表达量在核桃种子油脂转化期持续上升。可见，在核桃脂肪酸合成代谢途径和信号通路中有众多基因的参与，因此在进行核桃品种遗传改良时，需综合考虑各基因的功能及表达模式。

4 结 论

核桃种子油脂转化期的转录组测序结果较好，Unigene的功能分类，初步明确了各个基因编码的蛋白质功能，但仍有部分基因无法确定其编码的蛋白质功能，核桃基因组的构建有待进一步完善。核桃种子的油脂转化主要集中在G2和G3阶段，这2个阶段的脂肪酸生物合成途径上，与油脂合成有关的基因*accA*、*accB*、*accC*、*fabF*、*fabG*、*FAB2*等均呈上调表达。

参考文献 **References**：

[1] 郗荣庭，刘孟军．中国干果[M]．北京：中国林业出版社，2005：78-79.
XI Rongting，LIU Mengjun. Chinese dried fruit[M]. Beijing: China Forestry Publishing House，2005: 78-79.

[2] LAVEDRINE F, RAVEL A, VILLET A, DUCROS V, ALARY J. Mineral composition of two walnut cultivars originating in France and California[J]. Food Chemistry, 2000, 68(3): 347-351.

[3] 周鸿升，王希群，徐芝生，郭保香 . 美国核桃特殊营养成分及其临床功效研究进展[J]. 河北林果研究，2010，25(1): 51-56.
ZHOU Hongsheng, WANG Xiqun, XU Zhisheng, GUO Baoxiang. Progress on clinical research explore benefits of walnuts´ unique nutritional profile of U.S.A[J]. Hebei Journal of Forestry & Orchard Research, 2010, 25(1): 51-56.

[4] 王帅，戴涟漪，库雪晶，荣瑞芬 . 核桃营养组成与保健功能研究进展[J]. 中国酿造，2016，35(6): 30-34.
WANG Shuai, DAI Lianyi, KU Xuejing, RONG Ruifen. Research progress on walnut nutritional composition and health function[J]. China Brewing, 2016, 35(6): 30-34.

[5] 吴国良，刘群龙，郑先波，宋宇琴，简在海，彭功波 . 核桃种质资源研究进展[J]. 果树学报，2009，26(4): 539-545.
WU Guoliang, LIU Qunlong, ZHENG Xianbo, SONG Yuqin, JIAN Zaihai, PENG Gongbo. Advances in research on the worldwide walnut germplasm[J]. Journal of Fruit Science, 2009, 26(4): 539-545.

[6] 冯连芬，吕芳德，张亚萍，和红晓 . 我国核桃育种及其栽培技术研究进展[J]. 经济林研究，2006，24(2): 69-73.
FENG Lianfen, LÜ Fangde, ZHANG Yaping, HE Hongxiao. Literature review of researches on breeding and cultivation techniques of chinese walnut[J]. Nonwood Forest Research, 2006, 24(2): 69-73.

[7] 张如义，胡红玲，胡庭兴，杨丽妹，舒兰，阮若玉 . 核桃凋落叶分解对 3 种作物生长、光合及抗性生理特性的影响[J]. 生态与农村环境学报，2016，32(4): 595-602.
ZHANG Ruyi, HU Hongling, HU Tingxing, YANG Limei, SHU Lan, RUAN Ruoyu. Effects of decomposing walnut (*Juglans regia*) leaf litter on growth, photosynthesis and resistance physiology of three recipient plants[J]. Journal of Ecology and Rural Environment, 2016, 32(4): 595-602.

[8] WHITE S W, ZHENG J, ZHANG Y M. The structural biology of type II fatty acid biosynthesis[J]. Annual Review of Biochemistry, 2005, 74(1): 791-831.

[9] BAUD S, GUYON V, KRONENBERGER J, WUILLEME S, MIQUEL M, CABOCHE M, LEPINIEC L, ROCHAT C. Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in *Arabidopsis*[J]. Plant Journal, 2003, 33(1): 75-86.

[10] PATEL M, JUNG S, MOORE K, POWELL G, AINSWOETH C, ABBOTT A. High-oleate peanut mutants result from a MITE insertion into the FAD2 gene[J]. Theoretical and Applied Genetics, 2004, 108(8): 1492-502.

[11] JADHAV A, KATAVIC V, MARILLIA E F, MICHAEL G E, BARTON D L, KUMAR A, SONNTAG C, BABIC V, KELLER W A, TAYLOR D C. Increased levels of erucic acid in *Brassica carinata* by co-suppression and antisense repression of the endogenous FAD2 gene[J]. Metabolic Engineering, 2005, 7(3): 215-220.

[12] 陈昊，蒋桂雄，龙洪旭，谭晓风 . 基于油桐种子 3 个不同发育时期转录组的油脂合成代谢途径分析[J]. 遗传，2013，35(12): 1403-1413.
CHEN Hao, JIANG Guixiong, LONG Hongxu, TANG Xiaofeng. Analysis of oil synthesis metabolism pathways based on transcriptome changes in tung oil tree's seeds during three different development stages[J]. Hereditas, 2013, 35(12): 1403-1413.

[13] 江南，谭晓风，张琳，曾艳玲 . 基于 RNA-Seq 的油茶种子 α-亚麻酸代谢途径及相关基因分析[J]. 林业科学，2014，50(8): 68-75.
JIANG Nan, TAN Xiaofeng, ZHANG Lin, ZENG Yanling. Gene analysis of α- Linolenic acid metabolism of *Camellia oleifera* seeds based on RNA-Seq[J]. Scientia Silvae Sinicae, 2014, 50(8): 68-75.

[14] 祁云霞，刘永斌，荣威恒 . 转录组研究新技术：RNA-Seq 及其应用[J]. 遗传，2011，33(11): 1191-1202.
QI Yunxia, LIU Yongbin, RONG Weiheng. RNA-Seq and its applications: a new technology for transcriptomics[J]. Hereditas, 2011, 33(11): 1191-1202.

[15] WANG Z, GERSTEIN M, SNYDER M. RNA-Seq: a revolutionary tool for transcriptomics[J]. Nature Reviews Genetics, 2009, 10(1): 57.

[16] NATARAJAN P, PARANI M. De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing[J]. BMC Genomics, 2011, 12(4): 191-202.

[17] 陈虹，潘存德，王蓓，胡渊，肖真真，何苗 . 核桃种子发育主要营养物质积累之间的关系及脂肪酸动态变化[J]. 河北农业大学学报，2016，39(1): 57-62.
CHEN Hong, PAN Cunde, WANG Bei, HU Yuan, XIAO Zhenzhen, HE Miao. The relationship among nutrients' accumulation and dynamic changes of fatty acids in seed development of walnut [J]. Journal of Agricultural University of Hebei, 2016, 39(1): 57-62.

[18] 陈虹，潘存德，王蓓，肖真真，胡渊，胡国俊 . 核桃种子油体发育及不同品种种子子叶油体特征差异分析[J]. 中国农业科学，2015，48(19): 3899-3909.
CHEN Hong, PAN Cunde, WANG Bei, XIAO Zhenzhen, HU Yuan, HU Guojun. Oil body observation in seed development and its analysis in seed of *Juglans regia* 'Wen185' and *J. regia* 'Xinxin2' in period of seed maturity[J]. Scientia Agricultura Sinica, 2015, 48(19): 3899-3909.

[19] ANDERS S, HUBER W. Differential expression analysis for sequence count data[J]. Genome Biology, 2010, 11(10): 1-12.

[20] BUSTIN S A, BENES V, GARSON J A, HELLEMANS J, HUGGETT J, KUBISTA M, MUELLER R, NOLAN T, PFAFFL M W,

SHIPLEY G L, VANDESOMPELE J, WITTWER C T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments[J]. Clinical Chemistry, 2009, 55(4): 611-22.

[21] WEI W L, QI X Q, WANG L H, ZHANG Y X, HUA W, LI D H, LÜ H X, ZHANG X R. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers[J]. BMC Genomics, 2011, 12(1): 451.

[22] 王红霞,赵书岗,高仪,玄立春,张志华. 基于 AFLP 分子标记的核桃核心种质的构建[J]. 中国农业科学, 2013, 46(23): 4985-4995.
WANG Hongxia, ZHAO Shugang, GAO Yi, XUAN Lichun, ZHANG Zhihua. A construction of the core-collection of *Juglans regia* L. based on AFLP molecular markers[J]. Scientia Agricultura Sinica, 2013, 46(23): 4985-4995.

[23] 魏利斌,苗红梅,张海洋. 芝麻发育转录组分析[J]. 中国农业科学, 2012, 45(7): 1246-1256.
WEI Libin, MIAO Hongmei, ZHANG Haiyang. Transcriptomic analysis of sesame development[J]. Scientia Agricultura Sinica, 2012, 45(7): 1246-1256.

[24] SHI C Y, YANG H, WEI C L, YU O, ZHANG Z Z, JIANG C J, SUN J, LI Y Y, CHEN Q, XIA T, WAN X C. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds[J]. BMC Genomics, 2011, 12(1):131.

[25] VON W K, OLSEN J G, MCGUIRE K A, HENRIKSEN A. Fatty acid synthesis. Role of active site histidines and lysine in Cys-His-His-type beta-ketoacyl-acyl carrier protein synthases[J]. The FEBS Journal, 2006, 273(4): 695-710.

[26] CARLSSON A S, LABRIE S T, KINNEY A J, VON W K, BROWSE J. A KAS2 cDNA complements the phenotypes of the *Arabidopsis* fab1 mutant that differs in a single residue bordering the substrate binding pocket[J]. The Plant Journal, 2002, 29(6): 761-770.

[27] PIDKOWICH M S, NGUYEN H T, HEILMANN I, ISCHEBECK T, SHANKLIN J. Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(11): 4742-4747.

[28] FUKUCHI-MIZUTANI M, TASAKA Y, TANAKA Y, ASHIKARI T, KUSUMI T, MURATA N. Characterization of delta 9 acyl-lipid desaturase homologues from *Arabidopsis thaliana*[J]. Plant & Cell Physiology, 1998, 39(2): 247-253.

[29] 李晓丹. 油料作物种子脂肪酸累积模式及相关基因的克隆与序列比较研究[D]. 北京: 中国农业科学院, 2007.
LI Xiaodan. Comparative study on pattern of fatty acid biosynthesis and sequences of related genes cloned in oil seed crops[D]. Beijing: Chinese Academy of Agricultural Sciences, 2007.

[30] ZABOROWSKA Z, STARZYCKI M, FEMIAK I, SWIDERSKI M, LEGOCKI A B. Yellow lupine gene encoding stearoyl-ACP desaturase- organization, expression and potential application[J]. Acta Biochimica Polonica, 2002, 49(1): 29-42.