



大豆籽粒不同发育时期的转录组分析

陈锦玲, 徐媛, 陈玉梅, 李璐璐, 李惠敏, 秦新民

(广西师范大学 生命科学院, 广西 桂林 541004)

摘要: 为了从分子水平上研究大豆籽粒不同发育时期的油脂合成与累积机理, 对大豆开花 20 d (DD₂₀)、30 d (DD₃₀)、40 d (DD₄₀)、50 d (DD₅₀) 的籽粒进行了转录组测序。通过对转录组测序数据的分析, 共得到原始数据 461 566 988 条, 经过滤获得开花后 20, 30, 40 和 50 d 的 clean reads, 分别为 107 548 920, 111 670 776, 109 339 672 和 108 884 270 条。DD₂₀/DD₃₀、DD₃₀/DD₄₀、DD₄₀/DD₅₀ 比较组中的差异表达基因分别为 4 759, 6 245 和 13 763 个, 其中上调基因分别为 1 801, 2 941 和 5 695 个。差异表达基因的 KEGG pathway 分析中, 分别得到 134, 133 和 136 条代谢通路, 筛选到 8 个与油脂合成相关的差异表达基因, *ACC*、*FATB*、*GPAT*、*DGATI*、*G3PDH*、*KASI*、*SAD* 和 *FAD2*。研究结果对深入研究大豆籽粒脂质合成的调控机理及大豆高油育种具有重要参考价值。

关键词: 大豆; 差异表达基因; 油脂合成; 转录组测序

Sequencing Analysis of Transcriptome During the Different Developmental Stages in Soybean Seed

CHEN Jin-ling, XU Yuan, CHEN Yu-mei, LI Lu-lu, LI Hui-min, QIN Xin-min

(College of Life Science, Guangxi Normal University, Guilin 541004, China)

Abstract: In order to study the mechanism of oil synthesis and accumulation in soybean seeds at different developmental stages at the molecular level, the soybean seeds of 20 d (DD₂₀), 30 d (DD₃₀), 40 d (DD₄₀) and 50 d (DD₅₀) were used as test materials in transcriptome sequencing. Through the analysis of transcriptome sequencing data, 461 566 988 raw reads were obtained. After filtration, the clean reads of DD₂₀, DD₃₀, DD₄₀ and DD₅₀ were obtained respectively, as 107 548 920, 111 670 776, 109 339 672 and 108 884 270. The total differential expression genes of DD₂₀/DD₃₀, DD₃₀/DD₄₀, DD₄₀/DD₅₀ were 4 759, 6 245 and 13 763 respectively, of which up-regulated genes were 1 801, 2 941 and 5 695 respectively. KEGG pathway analysis of differential expression genes obtained 134, 133 and 136 KEGG pathways, screened 8 differential expression genes related to lipid synthesis, as *ACC*, *FATB*, *GPAT*, *DGATI*, *G3PDH*, *KASI*, *SAD* and *FAD2*. The results provide an important reference for further study on the regulation mechanism of soybean seed lipid synthesis, as well as for the breeding of high oil content soybean varieties.

Keywords: Soybean; Differential expressed genes; Lipid synthesis; Transcriptome sequencing

大豆 (*Glycine max*) 属于豆科 (Leguminosae)、大豆属 (*Glycine*) 植物^[1], 既是重要的油料和高蛋白作物, 又是食品、饲料等多种加工工业的优质原材料^[2]。我国大豆种植虽然有着悠久的种植历史和丰富的种质资源, 但在国际市场上却缺乏竞争力, 品质不高是制约我国大豆产业的一个主要因素。因此大豆的种质资源和基因的改良^[3-4], 以及关键基因的调控研究一直受到人们的广泛重视^[5]。

近年来, 农作物的转录组测序取得了许多重大进展。随着拟南芥^[6]、水稻^[7]、棉花^[8]、芝麻^[9]、玉米^[10]、大豆^[11]等植物的全基因组测序的完成, 越来越多的研究者使用 RNA 测序技术研究不同环境条件下或不同生长发育阶段植物的基因表达模式, 为理解基因表达调控机制以及挖掘与特定性状相关的候选基因提供大量重要的信息^[12]。大豆籽粒中

含有丰富的油脂, 其含油量约为 20%, 是世界上主要的食用油植物之一。大豆油脂合成与积累的分子生物学基础是研究大豆基因组学的主要内容, 也是当今世界对大豆品质改良和育种研究中的热门课题。Ke 等^[13]构建了芝麻未成熟种子的均一化 cDNA 文库, 获得了大量参与其种子油脂合成的同源基因, 其中有 117 个 EST 可能参与了相关的生物合成过程。Schmutz 等^[14]将拟南芥中油脂合成相关的基因与大豆基因组做同源比对分析, 发现 1 127 个与油脂合成相关的基因, 与油脂储存相关的基因有 22 个, 这些基因及其相关通路对大豆油脂含量有重要的影响, 通过对某些基因的修饰和调控, 可增大大豆的油脂产量。陆亮等^[15]研究发现, *GmDGK7* 基因的 SNP1 与油脂含量显著关联, *GmTPR* 基因的 SNP2 与油脂含量、硬脂酸、油酸和亚油酸含量均显

收稿日期: 2019-01-21

基金项目: 广西研究生教育创新计划 (XYCSZ2018056)。

第一作者简介: 陈锦玲 (1994 -), 女, 硕士, 主要从事植物分子生物学研究。E-mail: 1078330499@qq.com。

通讯作者: 秦新民 (1956 -), 男, 博士, 教授, 主要从事分子生物学研究。E-mail: 3066958178@qq.com。

著关联。除此之外,一些研究还表明许多转录因子,如 LEC1、LEC2、ABI3 和 WRI1 等在种子发育时油脂的合成与积累过程中发挥重要的调控作用,这些转录因子可能作为分子育种中潜在的靶基因^[16]。近年来,利用基因差异表达分析方法在油脂合成代谢机制方面的研究也有所进展。吕晓杰等^[17]通过基因差异表达分析方法获得 4 个与脂肪酸和甘油三酯代谢相关基因,发现 *ACCase α-CT*、*KAS III*、*LACS2*、*GPAT4* 随着油茶籽的成熟,表达水平呈上升趋势,这些基因有助于油茶籽含油量的增加。Wang 等^[18]对油桐转录组数据进行了差异基因表达分析,鉴定了 *mhylp* 在脂质生物合成中的关键作用,并为探索脂溶性酵母的生理性二态生长提供了丰富的资源。

为了进一步揭示大豆油脂合成代谢调控机制,本研究通过转录组测序技术(RNA-Seq)对大豆不同发育时期籽粒进行测序,分析和筛选了大豆籽粒不同发育时期油脂合成与积累过程相关差异表达基因,为今后大豆的优良品种的培育、以及生产上采取相应措施提高大豆的油脂含量提供理论依据。同时,研究结果也可为其它高油作物的育种提供有益参考。

1 材料与amp;方法

1.1 材料

供试材料为大豆桂夏 7 号,平均生育期为 104 d,其中开花期至收获期约 65 d。由广西壮族自治区农业科学院经济作物研究所选育并提供。

1.2 试验设计

试验于 2018 年 7 月在广西师范大学生物实验楼进行户外盆栽种植,选取籽粒基本一致的桂夏 7 号大豆种子,每盆点播 2~3 株,株距 10 cm。待种子生长发育成种苗后,选取长势均一大豆种苗 60 余株作为试苗。在大豆植株开花当日挂牌,进行时间标记。根据 Fehr 等^[19]提出的大豆生育时期的划分标准,选取 4 个不同发育时期的大豆籽粒:始粒期 R5(植株主茎任意节中,大豆籽粒长达 3 mm);开花后发育 20 d(DD_20)的大豆籽粒;鼓粒期 R6(植株主茎任意节上,至少有一荚果内籽粒达到最大体积);开花后发育 30 d(DD_30)和发育 40 d(DD_40)的大豆籽粒;成熟始期 R7(植株主茎上至少有一荚果籽粒趋向成熟时黄色颜色);开花后发育 50 d(DD_50)的大豆籽粒。样本存放于密封的离心管中,迅速放入液氮中,随后置于-80℃超低温冰箱保存、备用,利用干冰保存法送往华大基因进行转录组测序。对获得的测序数据基本信息进行初步的分析检测,筛选大豆籽粒不同发育过程的 3 个比较

组中与油脂合成相关的差异表达基因,分析相关差异表达基因参与油脂合成的调控作用。挑取 4 个差异表达基因:酰基 ACP 硫酯酶(*FATB*)、硬脂酰-ACP 脱氢酶(*SAD*)、脂肪酸脱氢酶(*FAD2*)、碱性 7S 球蛋白(*BGS*),进行荧光定量 PCR 分析,将 RT-PCR 验证结果与转录测序数据进行比较、分析测序数据的可靠性。

1.3 方法

1.3.1 总 RNA 提取、文库构建和测序 利用 Trizol 试剂盒(Invitrogen, IRI201™ Reagent15596026)提取大豆籽粒总 RNA。具体操作步骤参照试剂盒说明书进行。通过琼脂糖凝胶电泳检测总 RNA 浓度及完整性,并通过 Agilent 2100 检测 RNA 浓度及 OD₂₆₀/OD₂₈₀ 比值,将检测良好的 RNA 用于文库构建和测序。建库和测序委托华大基因科技有限公司完成,通过 BGISEQ-500 平台进行测序。

1.3.2 数据的过滤、拼接 使用过滤软件 SOAP-nuke 对 BGISEQ-500 平台测序获得的原始数据进行筛选,去除包含接头的 reads、未知碱基 N 含量大于 5%的 reads,以及去除低质量的 reads。得到 clean reads 之后,以大豆基因组(<http://www.phytozome.net/soybean>)作为参考基因组,使用 HISAT(<http://www.ccb.jhu.edu/software/hisat>)将 clean reads 比对到参考基因组序列,使用 Cufflinks^[20](<http://cole-trapnell-lab.github.io/cufflinks>)软件将比对结果组装构建转录本,选择最完整的转录本作为 Unigene。

1.3.3 差异基因的筛选、GO 分析和 KEGG pathway 分析 使用 FPKM 法(fragments per kb per million fragments)^[21]计算 Unigene 的表达量,找出在不同样本间差异表达的 Unigene^[22]。利用比对软件将筛选出的差异表达基因进行 GO 分类以及 KEGG pathway 分析,筛选 3 个转录组的比对组合中与油脂合成相关的差异表达基因。

1.3.4 荧光定量 PCR 分析 采用 Aidlab 公司反转录试剂盒(TUREscript 1st Stand cDNA SYNTHESIS Kit)进行总 RNA 反转录,合成 cDNA 第一条链。通过 analytikjena-qTOWER2.2 型荧光定量 PCR 仪、采用 2×SYBR® Green 染料法进行 RT-PCR 分析。以 *18S* 基因为内参基因,利用 2^{-ΔΔCt}法进行结果分析。采用 Beacon Designer 7.9 软件设计基因扩增引物,内参基因 *18S*(F: CAACCATAAACGATGCCGA; R: AGCCTTGCGACCATACTCC);*FATB*(F: TGGATGATGCTGGATTGGAAG; R: TCTGCCTGAACACAAGCC);*SAD*(F: TTGTGCTGCTGCTGGGTGATA; R: GCTCGTGCCGCTGTCATC);*FAD*(F: CCACCTACTTCCAC-

CTCCTTCC; R: GTGACCACACTCG TGAGCAATC)。BGS (F: CATTCAAGGAGTGGCTGGGTTAGG, R: AGGGTAGCGAGAGAGGCAAGT)。

2 结果与分析

2.1 总 RNA 的提取和测序质量分析

检测 RNA 的纯度为 1.73 ~ 1.99, 浓度 > 480 ng·μL⁻¹, 表明 RNA 质量较好, 可以进行后续的文

库构建和转录组测序分析。对 4 个文库测序共得到 461 566 988 条原始数据, 滤后得到 437 443 638 条 Reads, 占原始数据的 94.71% ~ 95.28%。与大豆基因组进行基因序列成功比对到 416 262 745 条序列, 比对率 ≥ 95.00% (表 1)。4 个转录组文库的比对和重新组装后共检查到表达的基因数为 46 913 个, 其中已知基因为 43 586 个, 预测的新基因为 3 325 个, 获得 16 081 个新转录本。

表 1 大豆籽粒不同发育时期测序数据统计

Table 1 Statistics of sequencing data of soybean seeds at different developmental stages				
样本名 Sample name	过滤前的 Reads Reads before filtering/bar	过滤后的 Reads Filtered Reads/bar	过滤后的 Reads 占比 Rate of filtered Reads/%	过滤后的 Reads 比对成功率 Rate of filtered reads alignment gene/%
DD-20	112876726	107548920	95.28	95.00
DD-30	117902640	111670776	94.71	95.15
DD-40	115393748	109339672	94.75	95.25
DD-50	115393874	108884270	94.36	95.23

2.2 差异表达基因的筛选

在 DD_20/DD_30、DD_30/DD_40、DD_40/DD_50 比较组中分别筛选得到 24 767 个差异表达基因, 3 个比较组的差异共表达基因有 1 077 个 (图 1)。在 DD_20 与 DD_30 时期中差异表达基因为 4 759 个, 其中 1 801 个为上调基因, 2 958 个为下调基因; 在

DD_30 与 DD_40 时期中差异表达基因为 6 245 个, 其中 2 941 个为上调基因, 3 304 个为下调基因; 在 DD_40 与 DD_50 时期中差异表达基因为 13 763 个, 其中 5 695 个为上调基因, 8 068 个为下调基因 (图 2)。

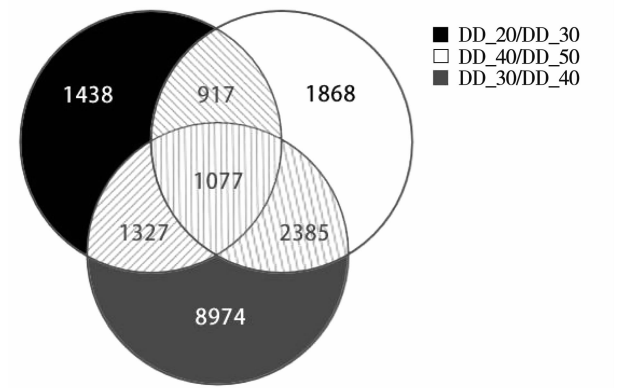


图 1 差异基因 Venn 图

Fig. 1 Venn map of differentially expressed genes

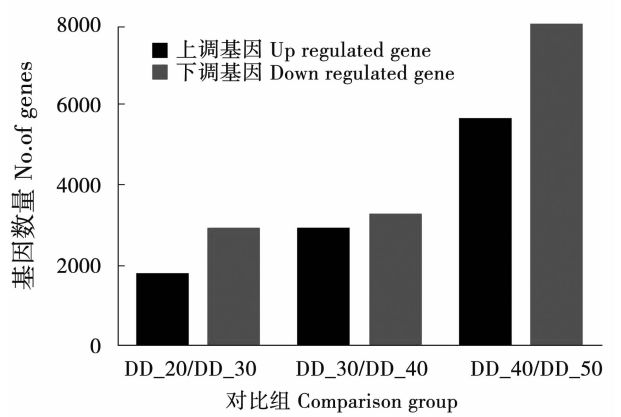


图 2 差异表达基因数量

Fig. 2 Number of differentially expressed genes

2.3 差异表达基因的 GO 功能富集分析

Gene Ontology 分为分子功能 (molecular function)、细胞组分 (cellular component) 和生物过程 (biological process) 三大功能类。差异表达基因的 GO 功能注释分类结果结果显示: DD_20/DD_30 比较组中这三大功能分类分别包含 10, 15 和 18 个层次的子类别; DD_30/DD_40 和 DD_40/DD_50 比较组中这三大功能分类被分为 11, 15 和 19 个层次的子类别。在分子功能分组中, 差异表达基因的功能主要集中在结合 (binding) 和催化活性 (catalytic activity)

类型中, DD_20/DD_30 中分别为 1 691 和 1 660 个, DD_30/DD_40 中分别为 2 161 和 2 199 个, DD_40/DD_50 中分别为 5 118 和 5 176 个; 在细胞组分类别中, 差异表达基因的功能主要集中在细胞 (cell)、膜 (membrane)、膜部分 (membrane part) 和细胞器 (organelle) 4 种类型中, DD_20/DD_30 中分别为 1 373, 1 261, 1 158 和 800 个, DD_30/DD_40 中分别为 1 816, 1 643, 1 504 和 1 072 个, DD_40/DD_50 中分别为 4 347, 3 717, 3 431 和 2 732 个; 在生物过程类别中, 差异表达基因的功能主要集中在细胞过程 (cel-

lular process) 和代谢过程 (metabolic process) 类型中,DD_20/DD_30 中分别为1 410和1 466个, DD_30/DD_40 中分别为1 856和1 953个, DD_40/DD_50 中分别为4 644和4 826个(图3~图5)。结果表明3

个比较组的差异表达基因主要富集于催化活性、细胞和代谢过程生物学通路中,而且3个比较组的功能分类模式相似。

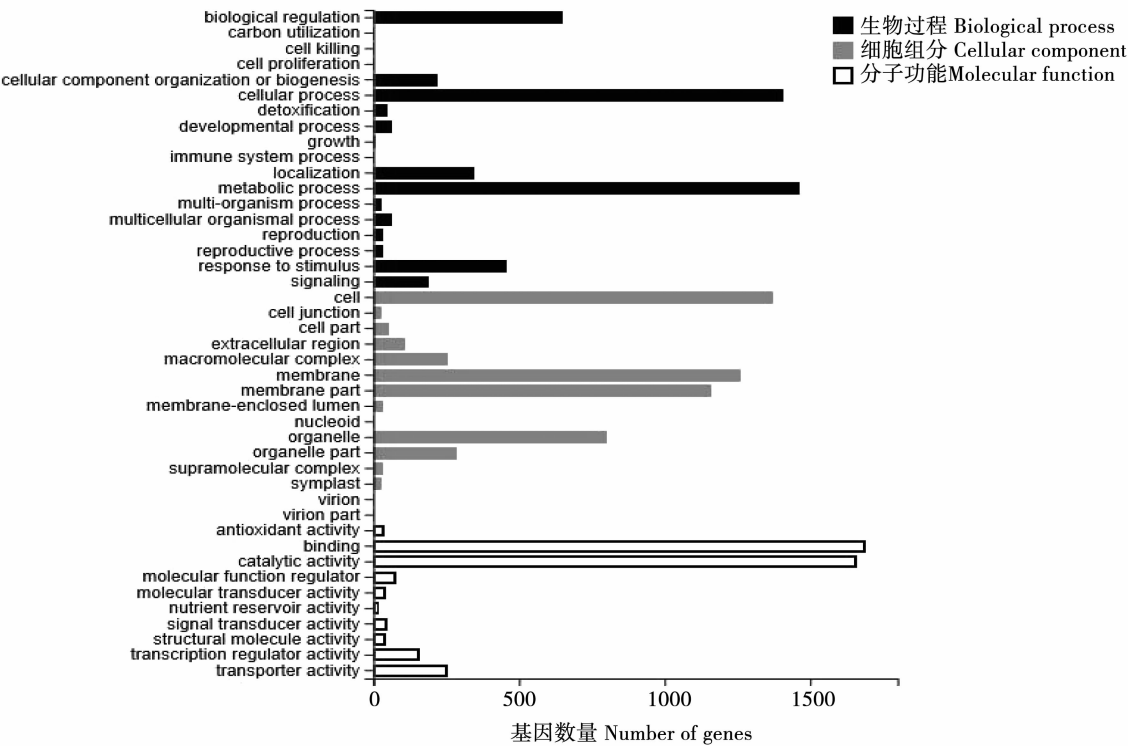


图3 DD_20/DD_30 差异基因 GO 分类
Fig.3 GO classification of DD_20/DD_30 differential genes

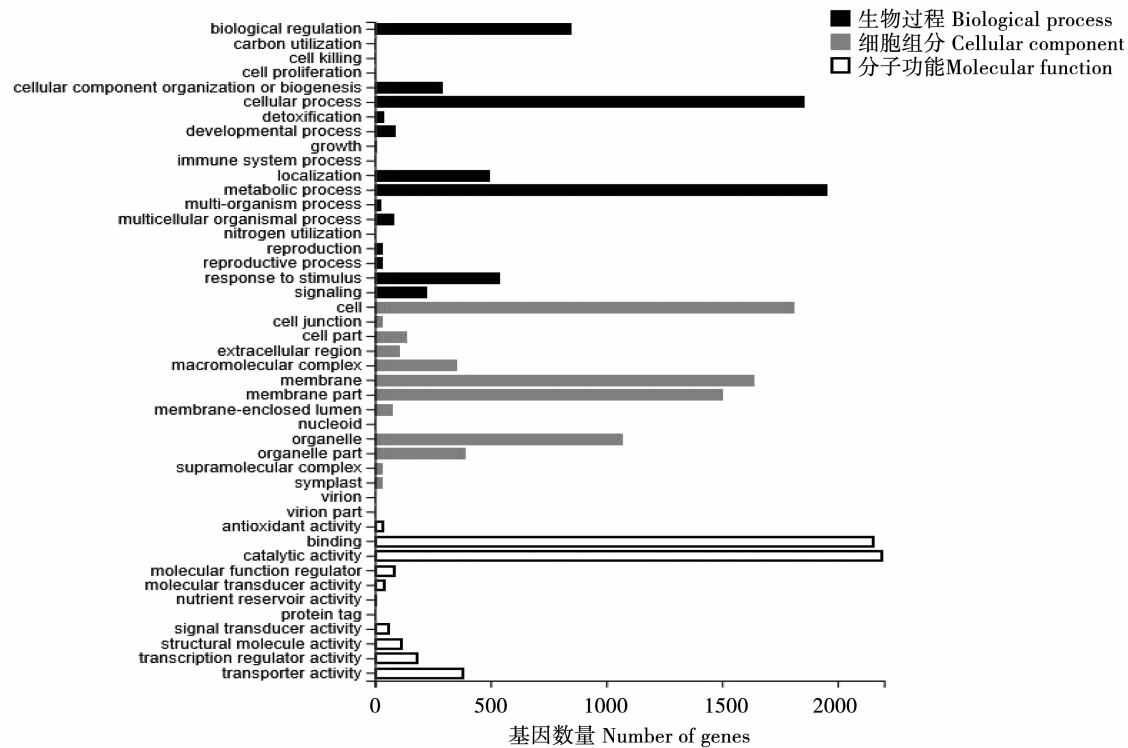


图4 DD_30/DD_40 差异基因 GO 分类
Fig.4 GO classification of DD_30/DD_40 differential genes

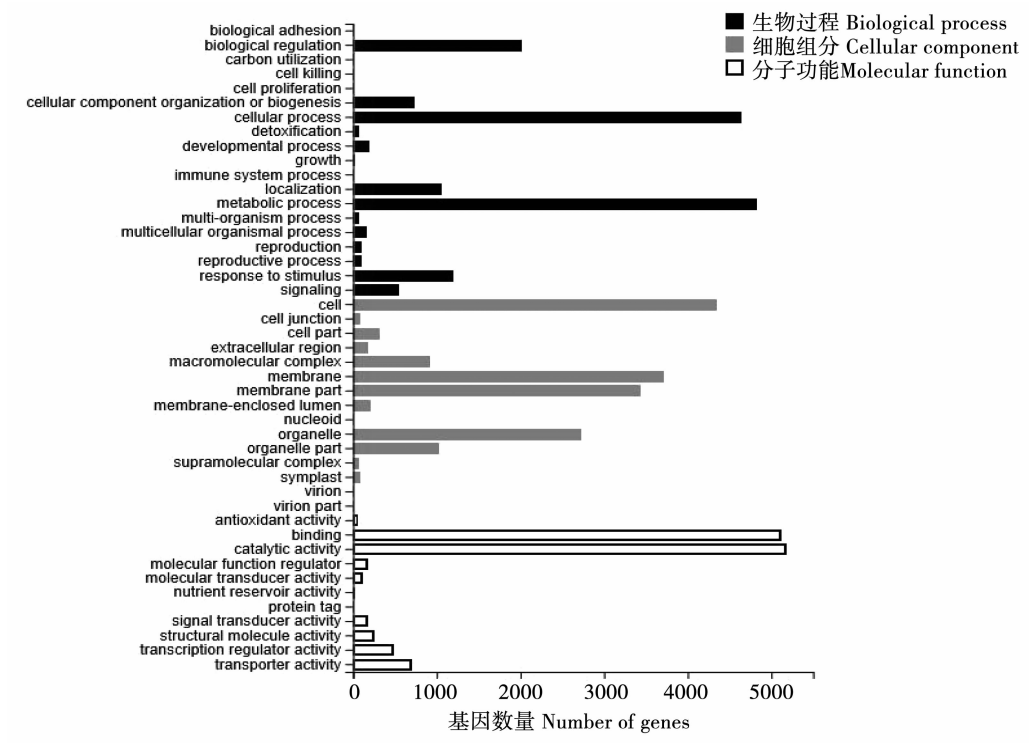


图5 DD_40/DD_50 差异基因 GO 分类

Fig. 5 GO classification of DD_40/DD_50 differential genes

2.4 差异表达基因的 KEGG pathway 分类和富集分析

差异表达基因的 KEGG pathway 代谢通路分为 6 个分支: 代谢 (metabolism)、细胞过程 (cellular processes)、环境信息处理 (environmental information processing)、遗传信息处理 (genetic information processing)、人类疾病 (human disease) (仅限动物)、有机系统 (organismal systems) (表 2)。每一分支下进一步分类统计, 可知这些基因参与 20 种代谢通路, 包括: 氨基酸代谢、脂质代谢、运输和分解代谢、信号转导、转录和翻译、内分泌和代谢疾病、环境适应

等。其中代谢类别分配类目最多, 有 11 种, 细胞过程、人类疾病和有机系统类别分别都只有 1 种, 环境信息处理类别有 2 种, 遗传信息处理类别有 4 种。3 个比较组中, 涉及 Unigenes 分别为 4 352, 5 801 和 10 983 个; 代谢类别分配到 Unigenes 数目最多, 分别占总数 67. 55%, 66. 26%, 64. 32%; 最少为人类疾病类别, 仅占总数 0. 16%, 0. 36%, 0. 39%。通过 KEGG pathway 注释分类, 3 个比较组分别得到 134, 133 和 136 条富集代谢途径, 富集 Unigenes 分别为 2 150, 2 922 和 5 801 个。

表 2 3 个比较组 KEGG pathway 分类

Table 2 KEGG pathway classification of three comparison groups

KEGG 代谢通路 1 KEGG pathway 1	KEGG 代谢通路 2 KEGG pathway 2	差异表达基因数 Differentially expressed genes		
		DD_20/DD_30	DD_30/DD_40	DD_40/DD_50
代谢 Metabolism	氨基酸代谢 Amino acid metabolism	2940	3844	7064
	其它次生代谢产物的生物合成 Biosynthesis of other secondary metabolites			
	碳水化合物代谢 Carbohydrate metabolism			
	能量代谢 Energy metabolism			
	全局和总览图 Global and overview maps			

续表 2

KEGG 代谢通路 1 KEGG pathway 1	KEGG 代谢通路 2 KEGG pathway 2	差异表达基因数 Differentially expressed genes		
		DD_20/DD_30	DD_30/DD_40	DD_40/DD_50
	糖类生物合成与代谢 Glycan biosynthesis and metabolism			
	脂质代谢 Lipid metabolism			
	辅因子与维生素的代谢 Metabolism of cofactors and vitamins			
	其它氨基酸的代谢 Metabolism of other amino acids			
	萜类和聚酮类化合物的代谢 Metabolism of terpenoids and polyketides			
	核苷酸代谢 Nucleotide metabolism			
细胞过程 Cellular processes	运输和分解代谢 Transport and catabolism	178	234	466
环境信息处理 Environmental information processing	信号转导 Signal transduction	316	465	916
	膜运输 Membrane transport			
遗传信息处理 Genetic information processing	复制和修复 Replication and repair	715	963	2021
	转录 Transcription			
	翻译 Translation			
	折叠、分类和降解 Folding, sorting and degradation			
人类疾病 Human disease	内分泌和代谢疾病 Endocrine and metabolic diseases	7	21	43
有机系统 Organismal systems	环境适应 Environmental adaptation	196	274	473
总计 Total	20	4352	5801	10983

2.5 油脂合成相关差异表达基因的筛选

对大豆不同发育期籽粒中油脂合成调控差异表达基因的分析结果表明,4 个样品中均呈现差异表达的油脂合成差异基因为乙酰 - CoA 羧化酶 (*ACC*)、3 - 磷酸甘油醛脱氢酶 (*G3PDH*)、 β - 酮脂酰 - ACP 合酶 (*KASI*)、酰基 ACP 硫酯酶 (*FATB*)、3 - 磷酸甘油脂酰转移酶 (*GPAT*)、二酰甘油脂酰转移酶 (*DGATI*)、硬脂酰 - ACP 脱氢酶 (*SAD*) 和脂肪酸

脱氢酶 (*FAD2*)。

ACC、*GPAT* 和 *DGATI* 的表达模式很相似,它们在前 3 个时期的表达差异不明显,而在 50 d 时其表达量有明显上升;*G3PDH* 在前 3 个时期的表达量变化持续上升,50 d 的表达量少量降低;*FATB* 表达为持续升高,*KASI*、*SAD* 和 *FAD2* 的表达量变化呈现先上升后下降的趋势,且都是在 30 d 的表达量达到最高,随后开始下降(表 3)。

表 3 油脂合成相关差异表达关键酶基因的表达量

Table 3 The expression level of differentially expressed key enzymes related to lipid synthesis

基因编号 Gene ID	酶名称 Enzyme name	缩写 Abbreviation	样本表达量 (FPKM) Sample expression			
			DD_20	DD_30	DD_40	DD_50
547859	乙酰 - CoA 羧化酶 Acetyl-CoA carboxylase carboxyl transferase subunit alpha	ACC	20. 59	18. 87	15. 86	46. 85
100780041	3 - 磷酸甘油醛脱氢酶 Glyceraldehyde-3-phosphate dehydrogenase	G3PDH	5. 52	10. 08	21. 38	16. 25
547575	β - 酮脂酰 - ACP 合酶 Beta-ketoacyl-ACP synthetase I	KASI	49. 41	53. 68	21. 12	14. 87
100797796	酰基 ACP 硫酯酶 Fatty acyl-ACP thioesterase B	FATB	13. 65	14. 88	37. 42	165. 29
100791297	3 - 磷酸甘油脂酰转移酶 Glycerol-3-phosphate acyltransferase	GPAT	5. 08	4. 52	6. 09	17. 4
732606	二酰甘油脂酰转移酶 Diacylglycerolacyltransferase-1	DGAT1	5. 21	4. 51	5. 87	13. 95
100037478	硬脂酰 - ACP 脱氢酶 Stearoyl-ACP desaturase	SAD	542. 65	1262. 78	599. 45	102. 62
547814	脂肪酸脱氢酶 Fatty acid desaturase	FAD2	370. 44	863. 56	626. 95	183. 26

2.6 荧光定量 PCR 验证

为了验证转录组测序的准确性,以 *18S* 为内参基因,挑取 4 个差异表达基因进行 Real-time PCR 分析,其中 *FATB*、*SAD* 和 *FAD2* 与油脂合成相关,*BGS* 与蛋白质合成相关。结果表明 RT-PCR 验证的 3 个与油脂合成相关的基因在大豆籽粒不同发育时期上调或下调变化的倍数与测序结果有些差异,但基因表达量变化趋势与转录组测序中基因的表达量

趋势基本一致(图 6),即 *FATB* 持续上调表达;*SAD* 和 *FAD2* 先上调后下调,且都在大豆籽粒发育 30 d 达到最大值;*BGS* 在大豆籽粒不同发育时期持续上调表达,与转录组测序中基因的表达量变化趋势基本符合,而且符合蛋白质合成积累的规律;这些数据表明转录组测序所得到的数据具有较高的可靠性。

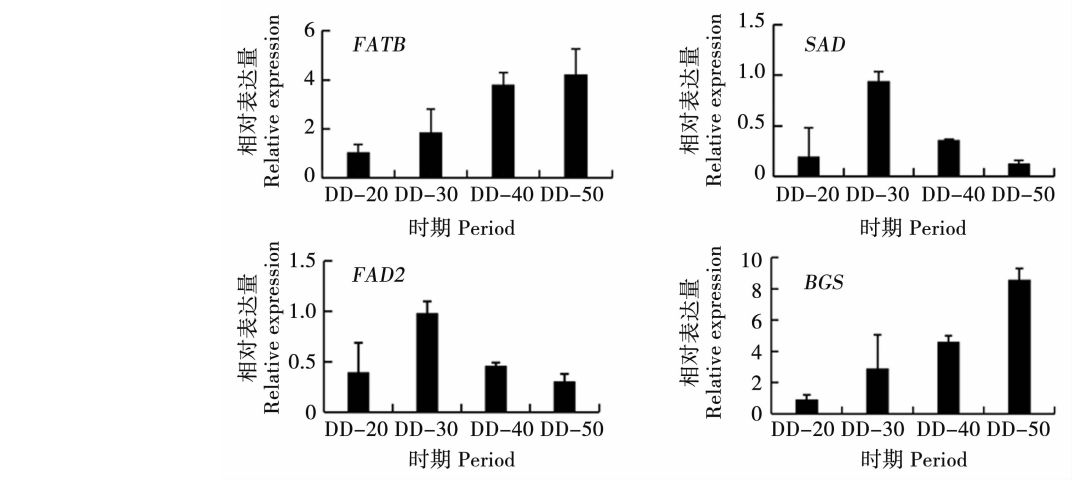


图 6 RT-PCR 验证的基因及其表达水平

Fig. 6 The RT-PCR validation genes and their expression levels

3 讨 论

本试验发现 8 个大豆籽粒不同发育时期油脂合成调控的差异表达基因。其中乙酰 Co A 羧化酶是植物油脂合成的关键酶^[23],该类酶的基因主要在油脂合成活性很高的组织部位表达^[24]。本研究结果显示该酶基因在大豆籽粒发育 50 d 有较高的表达量,而在前 3 个时期其表达量相近。3-磷酸甘油醛脱氢酶是油脂合成前体正相关的关键酶,增强该酶的活性可促进油脂积累^[25]。该酶在籽粒发育 40 d 较高表达而在 20 和 30 d 表达量相对不高。 β -酮脂酰-ACP 合酶 I (*kasI*)是碳链延伸的关键酶,在脂肪酸形成过程具有重要作用,拟南芥 *kas I* 突变型种子含油量显著降低,植株育性也显著降低;互补试验则可以恢复突变体的育性、含油量和脂肪酸组成^[26]。在本研究中,籽粒发育 20 和 30 d 时该酶的表达量相对较高,处于上升趋势,但在后两个时期其表达量下降。酰基 ACP 硫酯酶(*FATB*)主要催化释放棕榈酰 ACP 上的 ACP 基团,从而形成棕榈酸。本研究结果显示 *FATB* 在大豆籽粒发育阶段表达量持续上升,在 50 d 其表达量高达 165.26,与 Drmann 等^[27]发现种子中过量表达的 *FATB* 导致拟南芥种子中的脂肪酸含量大量积累的结果相符。三酰甘油(TAG)是大豆种子存储油的最主要形式,而二酰甘油酯酰转移酶(*DGAT*)和磷酸甘油酯酰转移酶(*GPAT*)是合成 TAG 的关键酶和限速酶^[28],过表达 *GPAT* 基因可以提高转基因植株的油分,Focks 等^[29]在拟南芥中导入外源 *GPAT* 基因能使转基因拟南芥种子的油分积累增加。此外,过量表达 *AtDGAT1* 也能使转基因拟南芥种子的 *DGAT* 活性比野生型高 10%~70%,同时种子的千粒重和含油量也得到提高^[30]。本研究发现的 *GPAT* 和 *DGAT1* 在大豆籽粒发育的 20~50 d 表达量持续上升,与大豆籽粒油脂含量的积累相符。硬脂酰-ACP 脱氢酶(*SAD*)和脂肪酸脱氢酶(*FAD*)是催化饱和脂肪酸形成不饱和脂肪酸的关键酶,调节它们的表达活性会影响脂肪酸组成,而且能改良油分品质。本研究发现 *SAD* 和 *FAD* 的表达模式相似,在籽粒发育 30 d 时表达量最高,整个发育阶段呈现先上升后下降的趋势。

油脂的合成和积累受脂肪酸合成途径中多种酶活性的影响,这些基因的表达还受到转录前、转录和转录后水平,以及外界环境因子的影响^[31]。本研究虽对大豆 4 个不同发育时期的籽粒进行了转录组测序和油脂合成相关差异表达基因的分析,但这些基因以及相关调控因子相互之间的协同调控机理尚不清楚,有待深入研究。

4 结 论

本研究通过 BGISEQ-500 平台对大豆 4 个不同发育时期的籽粒进行转录组测序,分别获得 107 548 920, 111 670 776, 109 339 672 和 108 884 270 条 clean reads。将获得的数据与大豆基因组进行基因序列比对,其比对率达到 95% 以上。对 3 个比较组进行差异表达基因筛选,分别获得 4 759, 6 245 和 13 763 个差异表达基因;GO 功能聚类分析显示 3 个比较组的差异表达基因主要富集于催化活性、细胞和代谢过程的生物学通路中;KEGG pathway 分析中,3 个比较组分别得到 134, 133 和 136 条富集代谢途径,富集 Unigenes 为 2 150, 2 922, 5 801 个,通过筛选获得 8 个与油脂合成相关的差异表达基因,为乙酰-CoA 羧化酶(*ACC*)、3-磷酸甘油醛脱氢酶(*G3PDH*)、 β -酮脂酰-ACP 合酶(*KASI*)、酰基 ACP 硫酯酶(*FATB*)、3-磷酸甘油酯酰转移酶(*GPAT*)、二酰甘油酯酰转移酶(*DGAT1*)、硬脂酰-ACP 脱氢酶(*SAD*)和脂肪酸脱氢酶(*FAD2*)。研究结果为研究大豆籽粒中油脂合成相关基因的调控作用、探讨其积累规律以及开展油脂合成相关基因克隆及功能研究等提供了重要理论依据,同时也为进一步大量挖掘大豆籽粒发育过程中油脂合成相关的重要表达基因奠定了基础。

参考文献

[1] 吕慧颖,王道文,葛毅强,等. 大豆育种行业创新动态[J]. 植物遗传资源学报, 2018, 19(3): 464-467. (Lyu H Y, Wang D W, Ge Y Q, et al. Innovation of soybean breeding industry [J]. Journal of Plant Genetic Resources, 2018, 19(3): 464-467.)

[2] 黄初女,金东淳,董艺兰,等. 浅谈大豆蛋白质品质改良[J]. 吉林农业科学, 2006, 31(1): 37-40. (Huang C N, Jin D C, Dong Y L, et al. Talking about the improvement of soybean protein quality[J]. Journal of Jilin Agricultural Sciences, 2006, 31(1): 37-40.)

[3] 冯丽娟. 不同基因型与环境因素对高油大豆品质性状的影响[D]. 大庆: 黑龙江八一农垦大学, 2007. (Feng L J. The effect of different genotypes and environmental factors on characters of high-oil soybean[D]. Daqing: Heilongjiang Bayi Agricultural University, 2007.)

[4] 王金龙,陈存来. 大豆种子贮藏蛋白组份 11S/7S 研究概况[J]. 山东农业科学, 1998(1): 48-50. (Wang J L, Chen C L. Study on seed storage protein component 11S/7S of soybean[J]. Shandong Agricultural Sciences, 1998(1): 48-50.)

[5] Wang J, Liu L, Guo Y, et al. A dominant locus, qBSC-1, controls β subunit content of seed storage protein in soybean [*Glycine max* (L.) Merri.][J]. Journal of Integrative Agriculture, 2014, 13(9): 1854-1864.

[6] Kaul S, Koo H L, Jenkins J, et al. Analysis of the genome se-

quence of the flowering plant *Arabidopsis thaliana* [J]. Nature, 2000, 408(6814): 796-815.

[7] Feng Q, Zhang Y, Hao P, et al. Sequence and analysis of rice chromosome 4[J]. Nature, 2002, 420(6913): 316-320.

[8] Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii* [J]. Nature Genetics, 2012, 44(10): 1098-1103.

[9] Zhang H, Miao H, Wang L, et al. Genome sequencing of the important oil-seed crop *Sesamum indicum* L.[J]. Genome Biology, 2013, 14(1): 401.

[10] Li P, Ponnala L, Gandotra N, et al. The developmental dynamics of the maize leaf transcriptome[J]. Nature Genetics, 2010, 42(12): 1060-1067.

[11] He W, Xu X, Liu X, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection[J]. Nature Genetics, 2010, 42(12): 1053-1059.

[12] 刘佳伟, 姚丹, 赵东海, 等. 大豆转录组测序研究进展综述[J]. 安徽农学通报, 2015, 21(8): 17-20. (Liu J W, Yao D, Zhao D H, et al. Research status of soybean transcriptome sequencing[J]. Anhui Agricultural Science Bulletin, 2015, 21(8): 17-20.)

[13] Ke T, Dong C, Mao H, et al. Analysis of expression sequence tags from a full-length-enriched cDNA library of developing sesame seeds (*Sesamum indicum*) [J]. BMC Plant Biology, 2011, 11(1): 180-191.

[14] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean[J]. Nature, 2010, 463(7278): 178-183.

[15] 陆亮, 贺建波, 苗龙, 等. 大豆 *GmDGK7* 和 *GmTPR* 基因与油脂相关性状的关联分析[J]. 大豆科学, 2015, 34(6): 938-944. (Lu L, He J B, Miao L, et al. Association of *GmDGK7* and *GmTPR* genes with oil related traits in soybean[J]. Soybean Science, 2015, 34(6): 938-944.)

[16] 李玉兰, 孙勤富, 王幼平. 植物油脂合成的转录调控研究进展[J]. 分子植物育种, 2016, 14(9): 279-288. (Li Y L, Sun Q F, Wang Y P. Research advance in transcriptional regulation of lipid synthesis and accumulation in plant [J]. Molecular Plant Breeding, 2016, 14(9): 279-288.)

[17] 吕晓杰, 潘德灼, 李健, 等. 油茶种仁成熟过程油脂合成代谢的转录组分析[J]. 森林与环境学报, 2018, 38(2): 156-163. (Lyu X J, Pan D Z, Li J, et al. Transcriptomic analysis of lipid synthesis metabolism in *Camellia oleifera* seed kernel during its maturation period[J]. Journal of Forest and Environment, 2018, 38(2): 156-163.)

[18] Wang G, Li D, Miao Z, et al. Comparative transcriptome analysis reveals multiple functions for *Mhy1p* in lipid biosynthesis in the oleaginous yeast *Yarrowia lipolytica* [J]. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 2018, 1863(1): 81-90.

[19] Fehr W R, Caviness C E. Stages of soybean development[R]// Special Report 80. Ames; Cooperative Extension Service, Agriculture and Home Economic Experiment Station, Iowa State University, 1977: 1-11.

[20] Roberts A, Pimentel H, Trapnell C, et al. Identification of novel transcripts in annotated genomes using RNA-Seq[J]. Bioinformatics, 2011, 27(17): 2325-2329.

[21] Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq[J]. Nature Methods, 2008, 5(7): 621-628.

[22] Audic S, Claverie J M. The significance of digital gene expression profiles[J]. Genome Research, 1997, 7(10): 986-995.

[23] Brownsey R W, Zhande R, Boone A N. Isoforms of acetyl-CoA carboxylase: Structures, regulatory properties and metabolic functions[J]. Biochemical Society Transactions, 1997, 25(4): 1232-1238.

[24] Ke J, Wen T N, Nikolau B J, et al. Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme A carboxylase [J]. Plant Physiology, 2000, 122(4): 1057-1071.

[25] Vigeolas H, Waldeck P, Zank T, et al. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter[J]. Plant Biotechnology Journal, 2010, 5(3): 431-441.

[26] 赵彦朋, 梁伟, 王丹, 等. 植物油脂合成调控与遗传改良研究进展[J]. 中国农业科技导报, 2018, 20(1): 14-24. (Zhao Y P, Liang W, Wang D, et al. Regulation of oil biosynthesis and genetic improvement in plant; Advances and prospects[J]. Journal of Agricultural Science and Technology, 2018, 20(1): 14-24.)

[27] Drmann P, Voelker T A, Ohlrogge J B. Accumulation of palmitate in *Arabidopsis* mediated by the acyl-acyl carrier protein thioesterase *FATBI* [J]. Plant Physiology, 2000, 123(2): 637-644.

[28] 赵江哲. 大豆磷脂酶基因 *GmPLD* 和脂合成酶基因 *GmDGAT*, *GmLPAT* 在调控拟南芥生长和种子油含量中的作用[D]. 南京: 南京农业大学, 2013. (Zhao J Z. Phospholipase *GmPLD* and lipid synthetase *GmDGAT* and *GmLPAT* from soybean regulate growth and seed oil in transgenic *Arabidopsis* plants[D]. Nanjing: Nanjing Agricultural University, 2013.)

[29] Focks N, Benning C. Wrinkled1: A novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism [J]. Plant Physiology, 1998, 118(1): 91-101.

[30] 刘正杰, 张园, 王玉美, 等. 陆地棉 *GhDGATI* 基因干涉载体构建与遗传转化[J]. 中国农业大学学报, 2013, 18(5): 1-8. (Liu Z J, Zhang Y, Wang Y M, et al. Construction and transformation of RNAi vector of gene *GhDGATI* in upland cotton [J]. Journal of China Agricultural University, 2013, 18(5): 1-8.)

[31] 周丹, 赵江哲, 柏杨, 等. 植物油脂合成代谢及调控的研究进展[J]. 南京农业大学学报, 2012, 35(5): 77-86. (Zhou D, Zhao J Z, Bai Y, et al. Research advance in triacylglycerol synthesis, metabolism, and regulation in plants[J]. Journal of Nanjing Agricultural University, 2012, 35(5): 77-86.)