

梁文化, 孙旭超, 岳红亮, 等. 水稻超大籽粒形成的重要基因和调控通路的转录组分析[J]. 江苏农业学报, 2020, 36(4): 801-813.
doi: 10.3969/j.issn.1000-4440.2020.04.001

水稻超大籽粒形成的重要基因和调控通路的转录组分析

梁文化, 孙旭超, 岳红亮, 田铮, 陈涛, 赵庆勇, 朱镇, 赵凌, 赵春芳, 姚姝, 路凯, 王才林, 张亚东

(江苏省农业科学院粮食作物研究所/江苏省优质水稻工程技术研究中心/国家水稻改良中心南京分中心, 江苏 南京 210014)

摘要: 以特大粒水稻品系 TD70 和小粒籼稻品种 Kasalath 不同发育时期的幼穗为材料, 进行转录组测序。将获得的干净读长(Clean read)以日本晴(Nipponbare)基因组序列为参考序列进行比对得到唯一读长(即特异比对到参考基因组的 read), 利用 FPKM(Fragments per kilobase of transcript per million fragments mapped)法计算基因的表达水平, 用 GO 和 KEGG 数据库对差异表达基因(Differentially expressed genes, DEGs)进行功能注释、富集和代谢通路分析。结果表明, 在水稻幼穗发育的早、中、后 3 个时期分别获得 3 618 个、4 183 个和 5 254 个差异表达基因, 共有 4 374 个差异表达基因得到基因本体(Gene ontology, GO)功能注释, 主要涉及 DNA 结合、细胞进程、信号转导、细胞增殖、物质转运等方面。KEGG 数据库分析结果显示, 差异表达基因共涉及 119 条代谢通路, 主要包括淀粉和蔗糖代谢通路、内质网中的蛋白质加工通路、激素信号转导通路、细胞周期蛋白通路等。

关键词: 水稻; 粒型; RNA-Seq; 幼穗; 转录组

中图分类号: S511.2⁺101 **文献标识码:** A **文章编号:** 1000-4440(2020)04-0801-13

Transcriptome analysis on critical genes and key pathways in extra-large grain development of rice

LIANG Wen-hua, SUN Xu-chao, YUE Hong-liang, TIAN Zheng, CHEN Tao, ZHAO Qing-yong, ZHU Zhen, ZHAO Ling, ZHAO Chun-fang, YAO Shu, LU Kai, WANG Cai-lin, ZHANG Ya-dong

(*Institute of Food Crops, Jiangsu Academy of Agricultural Sciences/Jiangsu High Quality Rice R&D Center/Nanjing Branch of China National Center for Rice Improvement, Nanjing 210014, China*)

收稿日期: 2019-12-11

基金项目: 国家自然科学基金面上项目(31771761); 国家自然科学基金项目(31901485); 江苏省农业科技自主创新基金项目[CX(18)1001]; 江苏省农业科学院院基金项目(003116111653); 江苏省农业科学院粮食作物研究所基金项目(LZS17-6); 江苏省农业生物学重点实验室开放课题项目(4911707Z201705); 江苏省重点研发计划项目(BE2018357); 国家现代农业产业技术体系项目(CARS-01-62)

作者简介: 梁文化(1984-), 男, 山东临沂人, 博士, 助理研究员, 主要从事水稻遗传育种研究。(E-mail) liangwenhua0228@126.com

通讯作者: 张亚东, (Tel) 025-84390314; (E-mail) zhangyd@jaas.ac.cn

Abstract: Using young panicles of extra-large grain rice TD70 and small grain indica rice Kasalath at different growth stages as materials, transcriptome sequencing was conducted. Unique read was acquired by comparing the obtained clean read with reference genome sequences from Nipponbare. Expression levels of genes were counted by fragments per kilobase of transcript per million fragments mapped (FPKM) method. Functional annotation, enrichment and metabolic pathway of the differentially expressed genes (DEGs) were analysed by using GO and KEGG databases. The results showed that 3 618, 4 183 and 5 254 DEGs were obtained in the early, middle and late stages of

young panicle development, and a total of 4 374 DEGs were annotated with gene ontology (GO), mainly involving DNA binding, cellular processes, signal transduction, cell proliferation and material transfer, etc. The analysis results of KEGG database analysis indicated that DEGs referred to 119 metabolic pathways, mainly including starch and sucrose metabolism, protein processing in endoplasmic reticulum, hormone signal transduction, cyclin and so on.

Key words: rice; grain shape; RNA-Seq; young panicle; transcriptome

水稻粒型包括粒长、粒宽和粒厚等3个基本要素,是影响水稻产量和品质的重要农艺性状^[1-2],因此粒型基因的发掘和利用一直倍受重视。目前人们已经在水稻中定位了400多个与粒型相关的数量性状座位(Quantitative trait locus, QTL),克隆的粒型基因已经超过60个,其中控制粒长的基因有*GS3*、*GL3.1/qGL3*、*GLW7*、*TGW3*、*OsMADS1*、*GS9*等^[3-9],控制粒宽的基因有*GW2*、*GW5/qSW5*、*GS5*、*GW7/GL7*、*GW8*等^[10-16],控制粒厚的基因有*WTGI*^[17-18]。以上研究结果证实,通过调节水稻的籽粒性状可以增加水稻千粒质量,从而提高水稻产量。

*GS3*是水稻中控制粒长、粒质量的重要基因,对粒宽、粒厚也有一定影响,全长的*GS3*基因编码1个跨膜蛋白,是G蛋白的 γ 亚基之一,在水稻幼穗中高表达^[3,19-20]。*qGL3/GL3.1*基因编码含有2个Kelch功能域的蛋白磷酸酶*OsPPL1*,在幼穗中表达量较高,通过调控细胞周期蛋白T1;3来控制颖壳的细胞分裂,最终影响水稻籽粒大小^[4-5]。*GW2*是控制水稻粒宽和粒质量的主效基因,在水稻各组织中都表达,该基因编码1个环型E3泛素连接酶,通过泛素介导的蛋白质降解途径负调控细胞分裂^[10]。*qSW5/GW5*基因编码钙调素结合蛋白,对水稻粒宽、粒质量有显著影响,*qSW5/GW5*基因在水稻各个组织、器官中都表达,在幼穗中表达量最高^[11-12,21]。*GS5*基因是控制水稻粒宽、粒质量和充实度的主效QTL,在幼穗中表达量较高,通过促进外稃/内稃细胞分裂和细胞膨胀而正调控水稻籽粒的大小^[13,22]。*GW7/GL7*基因编码拟南芥LONGIFOLIA蛋白的同源蛋白,在水稻幼穗和分生组织中高表达,通过调控谷粒的细胞分裂而影响粒型^[14-15]。*GW8*是1个SBP(Squamosa promoter binding protein)家族的转录因子,在7cm长的水稻幼穗中表达量最高,该基因的高表达会促进细胞增殖和籽粒灌浆,从而使籽粒变宽,进而提高产量,*GW8*基因与*GW7*基因的启动子结合并抑制*GW7*基因的表达^[16]。

本研究中的水稻材料TD70是来源于天鹅

谷//9520//(72-496/苏御糯)后代粒质量超双亲的超大粒稳定品系,Kasalath为印度的常规籼稻品种,2012年正季在南京种植的TD70千粒质量为65.0g,Kasalath千粒质量为17.4g^[23]。在前期的研究中,笔者利用TD70与小粒籼稻Kasalath构建了重组自交系(RIL)群体,共检测到19个稳定遗传的粒型相关QTL,这些QTL区间包含*GS3*、*qGL3*、*GW2*、*qSW5/GW5*、*GS5*、*GW7*及*GW8*共7个已克隆的粒型基因位点,遗传效应分析结果表明,TD70中这些粒型基因位点对水稻籽粒大小具有正向作用,Kasalath中不含有这些粒型基因的增效位点;进一步研究发现,TD70中不同粒型的基因组合对籽粒大小具有累加效应^[24-26]。由此可见,多个粒型基因的聚合是TD70具有超大籽粒的原因,然而在多个粒型基因聚合后,水稻孕穗期幼穗表达谱中穗发育相关基因的表达和代谢通路的变化并不清晰。

近年来,转录组测序(RNA-Seq)技术已经成为研究表达谱的有力工具,研究者利用RNA-Seq技术在水稻的育性、穗发育、耐盐、耐热等方面展开了相关研究^[27-33]。为了研究水稻粒型相关基因孕穗期在幼穗发育中的调控机制,本研究对超大粒梗稻TD70和小粒籼稻Kasalath孕穗期不同阶段的幼穗进行转录组测序。通过对差异表达基因的基因本体(Gene ontology, GO)富集和代谢通路的分析,从基因表达水平、代谢通路的变化方面对大粒水稻品种形成的分子调控网络进行解析,以期在水稻育种中高效利用粒型基因奠定基础。

1 材料与方法

1.1 材料的种植与取样

大粒梗稻品系TD70为天鹅谷//9520//(72-496/苏御糯)的后代,是本研究团队自主创制的稳定品系;Kasalath是印度的一个常规籼稻品种,由江苏省种质资源保护与利用平台提供。以上材料于2018年种植于江苏省农业科学院试验田中,于2018年5月12日播种,6月12日移栽,每个小区种植4

行,每行 10 株,株距为 13.5 cm,行距为 26.5 cm,每个品种种植 3 个小区,采用常规栽培和管理措施。研究发现,粒型相关基因在水稻幼穗中的表达比较活跃^[5-7]。为了全面分析水稻抽穗前幼穗发育不同阶段表达谱的变化,笔者在水稻孕穗期取 3 个不同发育时期的幼穗[分别记作 P1(幼穗长度为 1~3 cm)、P2(幼穗长度为 4~7 cm)、P3(幼穗长度为 8~10 cm)]进行转录组测序,每个时期的样品取 3 个生物学重复,共 18 个幼穗样本。幼穗样品分离后立即置于液氮中速冻,于 -80 °C 保存至 RNA 提取。

1.2 成熟种子粒型相关性状的测定

分别取 5 株成熟 TD70、Kasalath 单株的种子。每个单株随机挑选 10 粒饱满的种子,使用游标卡尺(精度为 0.01 mm)测量种子粒长、粒宽和粒厚,千粒质量(即单株水稻 1 000 粒风干种子的质量)用电子天平(精度 0.001 g)测定。每个性状以 5 株水稻所测平均值作为最终数值。

1.3 RNA 提取、文库构建及转录组测序

使用 TRIzol[®] RNA Purification Kit (Invitrogen) 对水稻幼穗组织进行总 RNA 的提取,并用 1% 琼脂糖凝胶进行电泳检测。使用 NanoDrop 2000、Qubit2.0、Agilent 2100 分别对 RNA 进行纯度、浓度、完整性检测。以 mRNA 为模板,利用 AMPure XP beads 试剂盒对 RNA 进行纯化、打断、修复、片段选择、PCR 富集等操作,构建 cDNA 文库。对于质量检测合格的文库,由安诺优达基因科技(北京)有限公司使用 Illumina 公司的 NovaSeq 6000 平台进行测序,测序模式为 Paired-end 150 bp。

1.4 测序数据分析

原始数据下机后,首先用 fastQC (Ver 0.11.9) 进行测序数据质量检测与评估;原始测序数据中的接头和低质量序列,采用软件 fastp (Ver 0.20.0) 进行处理,设定参数保证两端序列的最小长度均 ≥ 50 bp,进而获得高质量的干净读长(clean reads)。本研究基于日本晴(Nipponbare)水稻参考基因组(<http://plants.ensembl.org/index.html>, IRGSP-1.0),按照 Trapnell 等已发表的转录组分析的经典流程进行数据分析^[34]。首先,用软件包 Tophat2 (Ver 2.1.1) 将得到的 clean reads 与参考基因组进行比对^[34-35],进一步利用 Samtools (Ver 1.6) 软件获得特异性比对结果^[36]。然后,基于 Tophat 2 比对的结果,使用 Cufflinks (Ver 2.2.1) 软件包进行基因表达量的计算和差异基因的寻找。采用 FPKM (Fragments per

kilobase of transcript per million fragments mapped) 评价每个基因的表达丰度,同时过滤多个样本中 FPKM < 1 的假阳性转录本。进一步利用软件 Cuffdiff (Ver 2.2.1) 对水稻幼穗不同发育阶段的差异表达基因 (Differentially expressed genes, DEGs) 进行鉴定,差异表达基因的筛选条件如下: $\log_2(\text{fold change}) \geq 1$ (fold change 代表某个基因在 2 个材料中的表达量), q 值 < 0.05^[34]。将 DEGs 与美国国家生物信息中心 (National Center for Biotechnology Information, NCBI) 网站上的 nr 蛋白数据库进行比对,预测其基因功能,然后利用 GO、KEGG 数据库对差异表达基因进行功能分类和代谢途径的富集分析。

1.5 荧光定量 PCR (qRT-PCR) 验证

利用 qRT-PCR 定量技术分析部分差异表达基因的表达丰度。使用 Primer Premier (Ver 5.0) 软件进行引物设计,16 个基因的引物序列见表 1。用与 RNA-Seq 建库同批次的 RNA 进行 qRT-PCR 验证。使用 TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech) 进行 cDNA 第一链合成。定量 PCR 在 ABI Step One Plus Real-Time PCR System (Applied Biosystems) 上进行,使用的试剂为 SYBR Premix Ex Taq[™] RT-PCR kit (TaKaRa)。以水稻基因 *OsActin-1* [基因识别号 (ID): Os03g0718100] 作为内标,通过相对定量的方法获得测定值^[37]。

2 结果与分析

2.1 TD70 与 Kasalath 的粒型分析

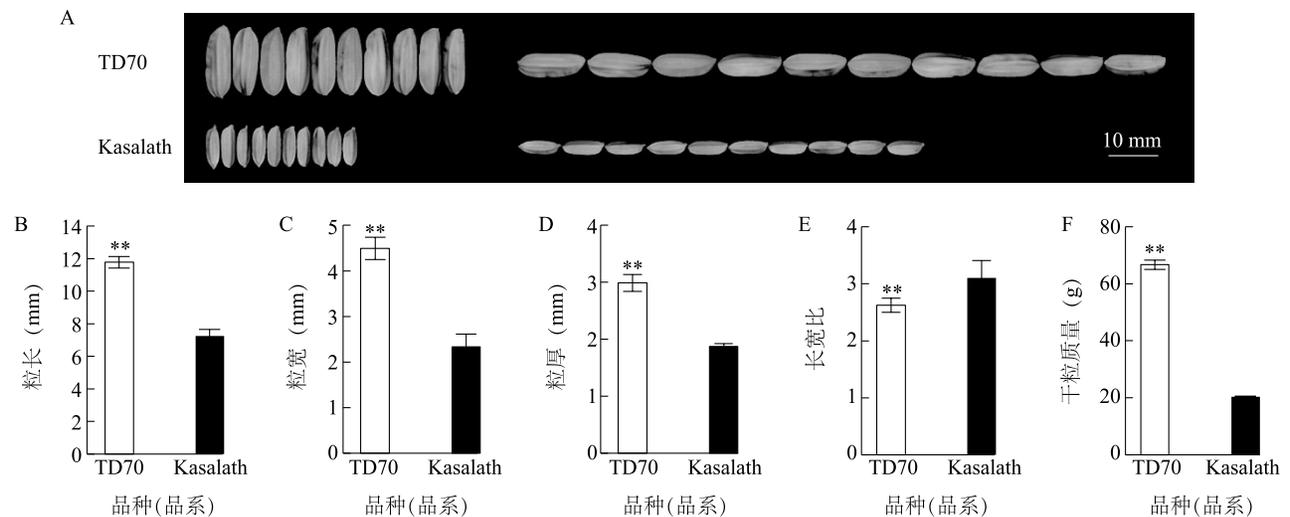
研究发现,特大粒水稻品系 TD70 中聚合了 *GS3*、*qGL3*、*GW2*、*qSW5/GW5*、*GS5*、*GW7* 及 *GW8* 共 7 个对籽粒大小具有正向效应的粒型基因;小粒籼稻 Kasalath 中不含有上述基因^[24-26]。于 2018 年正季测定成熟水稻的籽粒大小,由图 1 可以看出,TD70 平均粒长 11.78 mm,粒宽 4.49 mm,粒厚 2.99 mm,千粒质量 66.72 g,其千粒质量显著高于 Kasalath (17.94 g)。 t 检验结果表明,TD70 的粒长、粒宽、粒厚及千粒质量均高于 Kasalath,2 份材料之间的差异达到极显著水平 ($P < 0.01$);而 TD70 的长宽比则小于 Kasalath,二者之间的差异也达到极显著水平 ($P < 0.01$)。由此可见,多个粒型基因的聚合增加了水稻粒长、粒宽和粒厚,从而产生了超大籽粒,进而提高了千粒质量。

表 1 用于 qRT-PCR 表达分析的引物序列

Table 1 Primers used for qRT-PCR expression analysis

基因 ID	正向引物序列 (5'→3')	反向引物序列 (5'→3')
Os08g0531600	AGGAGTTTGATGAGGCCAAG	GCGTGTAGTATGGGCTCTCC
Os05g0187500	GCTTGAGCGCGTGCCAT	TCCTCTGCAAGAACGCGTGGC
Os02g0244100	TTATGCCTGTAAGTGAACATCTC	CGCAACCCGACAGGAAA
Os07g0603300	GAGCCACAGGGAAGATGCAA	GGAATATTTATATTTCTTCTATAGTTAATCTCA
Os03g0407400	ACGCTGCCTCCAGATGCTGCAG	CACCTTGCTCTGCACAAACAGCG
Os03g0646900	TAACAACATGACGCCTCCACCTG	GTTTCTGCAGCCACTACAACAGC
Os11g0247300	ATGAGGGAGTGCATCTCGAT	CAAGATCGACGAAGACAGCA
Os05g0158500	TTTGGCTCTACAGCGGTGATG	TGTTGAGGGGAACCAGGTGA
Os03g0687700	GCTCGCCATTGCCAAGCT	TTCAGAAAGGGAACAACGAACAC
Os04g0580700	CCACCACCAGATATTGACT	CCTCCTCCATTGCTTCTT
Os09g0309600	TCTGATGGGCATCGACAACG	GGCGCACTCGAAGGAGAGC
Os02g0638650	GATGTCCTTGCCGAATCA	GCTACTGCCGAACGATAT
Os06g0147500	CGCTAGTTTTGCTCATGGTTC	TCGCTCCGTTTTGAGTGCTG
Os12g0114800	GAACCTCAACATGGGCAACG	TGCAGCAAGAAAATAAACAAGCT
Os01g0944200	TGGCAACATCAAAGCAAAACC	TTCACCACCTGCCCTCCC
Os03g0712800	TCTGGACCTGTTGATGAC	GGAGGATGACTTCACTGT
Os03g0718100	CCACTATGTTCCCTGGCATT	GTACTCAGCCTTGCCAATCC

ID 表示识别号。



A: 籽粒形态特征; B: 粒长; C: 粒宽; D: 粒厚; E: 长宽比; F: 千粒质量; 各表型特征图中的 ** 表示经 *t* 检验, 2 个材料间存在极显著差异 ($P < 0.01$)。

图 1 TD70 和 Kasalath 的籽粒表型特征差异

Fig.1 Differences of grains phenotypic characteristics between TD70 and Kasalath

2.2 2 个水稻材料幼穗中转录组测序及序列比对

本研究通过对 2 个水稻材料的 18 个水稻幼穗样本测序文库进行测序, 得到的原始 read 数共为 857 017 092 条, 平均每个测序文库得到 47 612 061 条原始 read。进行去除接头、去污染和去低质量处理后, 平均每个测序文库获得的 clean read 数量为

45 474 692 条。通过 FastQC 软件对 clean 数据的测序质量进行检测, 结果表明, 18 个测序文库碱基的 Q30 (测序碱基准确率为 99.9% 的 read) 比例最低为 91.9%, 最高为 93.0%, 平均为 92.2%, 可见测序质量较高。利用 Tophat 软件包将测序得到的每个 clean 数据回贴到参考基因组中。从表 2 可以看出,

clean 数据比对到参考基因组的最低概率为 83.2%, 最高达到 90.9%。特异性比对 (Unique alignment) 到日本晴基因组中唯一位点的概率为 74.4%~

85.1%, 平均为 79.9%。TD70 比对到参考基因组的概率明显高于 Kasalath, 这可能是籼粳稻之间基因组的差异造成的。

表 2 测序数据统计及与水稻幼穗中基因组比对

Table 2 Summary of sequencing data and comparison of genomes in rice young panicles

样本编号	原始 read 数 (条)	原始碱基数 (bp)	clean read 数 (条)	clean 碱基数 (bp)	clean 数据的 Q30 比例 (%)	比对上参考基因组的 read (%)	特异比对到参考基因组的 read (%)
K-P1-1	50 169 832	7 525 474 800	46 707 476	7 006 121 400	92.1	83.3	74.4
K-P1-2	47 188 462	7 078 269 300	45 000 530	6 750 079 500	92.4	84.5	75.9
K-P1-3	49 602 626	7 440 393 900	46 918 000	7 037 700 000	92.1	83.5	74.8
K-P2-1	49 225 574	7 383 836 100	47 381 448	7 107 217 200	92.1	83.8	74.9
K-P2-2	43 373 592	6 506 038 800	41 170 490	6 175 573 500	92.1	83.4	74.5
K-P2-3	43 643 418	6 546 512 700	40 994 012	6 149 101 800	91.9	83.2	74.4
K-P3-1	47 660 926	7 149 138 900	46 702 074	7 005 311 100	92.5	85.2	77.2
K-P3-2	52 126 028	7 818 904 200	49 520 066	7 428 009 900	93.2	84.6	76.5
K-P3-3	50 338 604	7 550 790 600	47 574 644	7 136 196 600	93.0	83.9	75.9
T-P1-1	48 229 842	7 234 476 300	46 639 190	6 995 878 500	91.9	89.9	83.6
T-P1-2	47 388 156	7 108 223 400	45 872 434	6 880 865 100	92.2	90.2	83.8
T-P1-3	46 000 146	6 900 021 900	44 593 750	6 689 062 500	92.5	90.9	85.1
T-P2-1	46 012 598	6 901 889 700	44 057 950	6 608 692 500	92.0	90.3	84.5
T-P2-2	47 148 816	7 072 322 400	45 549 038	6 832 355 700	92.2	90.4	84.4
T-P2-3	48 715 176	7 307 276 400	46 348 178	6 952 226 700	92.3	90.5	84.8
T-P3-1	47 004 134	7 050 620 100	45 015 864	6 752 379 600	92.1	90.5	84.8
T-P3-2	42 895 246	6 434 286 900	41 018 196	6 152 729 400	91.9	90.6	84.8
T-P3-3	50 293 916	7 544 087 400	47 481 114	7 122 167 100	92.0	89.4	83.3
合计	857 017 092	128 552 563 800	818 544 454	122 781 668 100			
平均	47 612 061	7 141 809 100	45 474 692	6 821 203 783	92.2	87.1	79.9

K 表示 Kasalath; T 表示 TD70; P1、P2 和 P3 分别代表孕穗期 1~3 cm、4~7 cm 和 8~10 cm 长度的幼穗; 1、2、3 分别代表样本的 3 个生物学重复。原始数据、clean 数据均为测序得到的总碱基数; read 表示读长, clean read 表示干净读长。Q30: 测序碱基准确率为 99.9% 的 read。

2.3 2 个水稻材料幼穗中转录组的整体分析

为了分析 TD70 和 Kasalath 在水稻孕穗期幼穗中表达谱的动态差异, 笔者对至少在 1 个样本中表达的基因进行斯皮尔曼相关系数 (Spearman correlation coefficient, SCC) 分析和主成分分析 (Principal component analysis, PCA)。结果显示, 每个样本的 3 个生物学重复间的最低相关系数为 0.947 2, 说明测序样本生物学重复具有较好的重现性。从图 2A 可以看出, 同一水稻材料不同时期的样本间也存在相关性; 2 个水稻材料在相同发育时期的幼穗间存在一定差异; TD70 中后期 (P2 和 P3) 的幼穗具有一定的相似性而与前期的差异较大, Kasalath 也存在类似的情况。由图 2B 可以看出, 2 个水稻材料在早期 (P1) 幼穗中表达谱具有明显差异; 随着水稻幼穗的发育, 在中后期 (P2、P3) 幼穗中表达谱的差异进一步扩大, 这暗示 2 个水稻材料在幼穗发育早期

表达谱已经出现差异, 可能是造成后期水稻籽粒大小差异的原因。

2.4 2 个水稻材料幼穗中差异表达基因分析

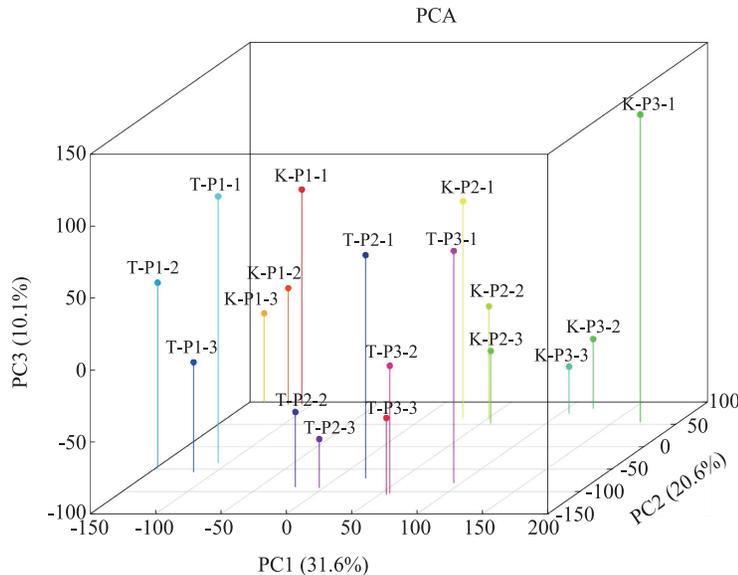
根据 Cuffdiff 的检测结果, 设定差异表达基因的标准为 $q < 0.05$, 差异表达倍数 ≥ 2 , 对 TD70 与 Kasalath 不同大小幼穗的 3 个比较组 (TD70-P1/Kasalath-P1、TD70-P2/Kasalath-P2、TD70-P3/Kasalath-P3) 进行差异表达基因分析。结果显示, P1 期有 3 618 个差异表达基因, 其中 2 436 个基因上调表达, 1 182 个基因下调表达; P2 期共获得 4 183 个差异表达基因, 其中 2 368 个基因上调表达, 1 815 个基因下调表达; P3 期获得的差异表达基因数共 5 254 个, 其中 2 426 个基因上调表达, 2 828 个基因下调表达。由图 3A 可见, 随着水稻幼穗的发育, 2 个材料间差异表达的基因数明显增多, 说明后期有更多基因参与到小穗发育进程中。从差异表达基因文氏图 (图 3B) 看出,

在3个时期都差异表达的基因有1 742个,在 P1、P2、P3 时期特异的差异表达基因分别有1 055个、990 个、2 096个。说明水稻幼穗在不同发育时期的差异表达基因既有共性的,也有不同时期特异性的。

A



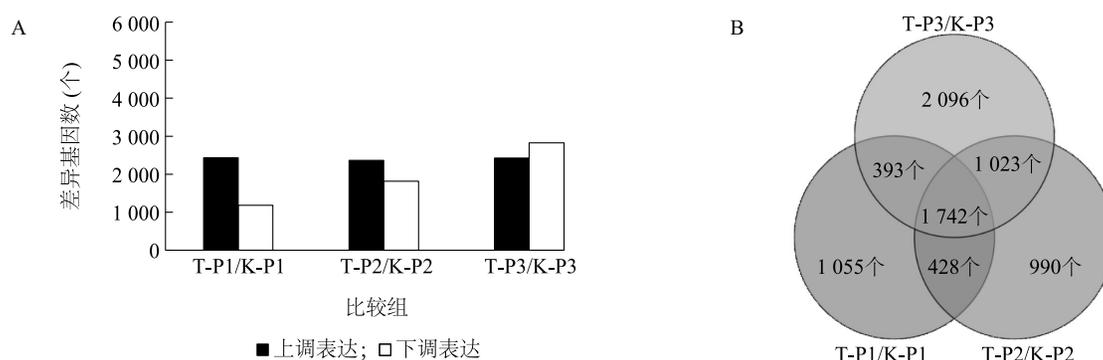
B



A;通过 Spanicleman 相关系数计算 3 个生物学重复之间的相关性,比例尺中数据表示斯皮尔曼相关系数;B:主成分分析显示 TD70 和 Kasalath 不同发育阶段幼穗的转录组聚类关系。图 A 中,a;K-P1-1;b;K-P1-2;c;K-P1-3;d;K-P2-1;e;K-P2-2;f;K-P2-3;g;K-P3-1;h;K-P3-2;i;K-P3-3;j;T-P1-1;k;T-P1-2;l;T-P1-3;m;T-P2-1;n;T-P2-2;o;T-P2-3;p;T-P3-1;q;T-P3-2;r;T-P3-3。K:Kasalath;T:TD70;P1、P2、P3:孕穗期 1~3 cm,4~7 cm,8~10 cm 长度的幼穗;1、2、3:样本的 3 个生物学重复。

图 2 2 个水稻材料幼穗发育不同阶段转录组之间的相关性

Fig.2 Correlation between transcriptomes of two rice materials at different young panicles development stages



A: TD70 和 Kasalath 幼穗发育不同时期差异表达基因统计结果; B: 差异表达基因在水稻幼穗发育 3 个时期的分布情况。K: Kasalath; T: TD70; P1、P2、P3: 孕穗期 1~3 cm、4~7 cm、8~10 cm 长度的幼穗; 1、2、3: 样本的 3 个生物学重复。

图 3 2 个水稻材料幼穗间差异表达基因的统计

Fig.3 Statistics of differentially expressed genes between two rice materials

2.5 2 个水稻材料幼穗中差异表达基因的 qRT-PCR 表达验证

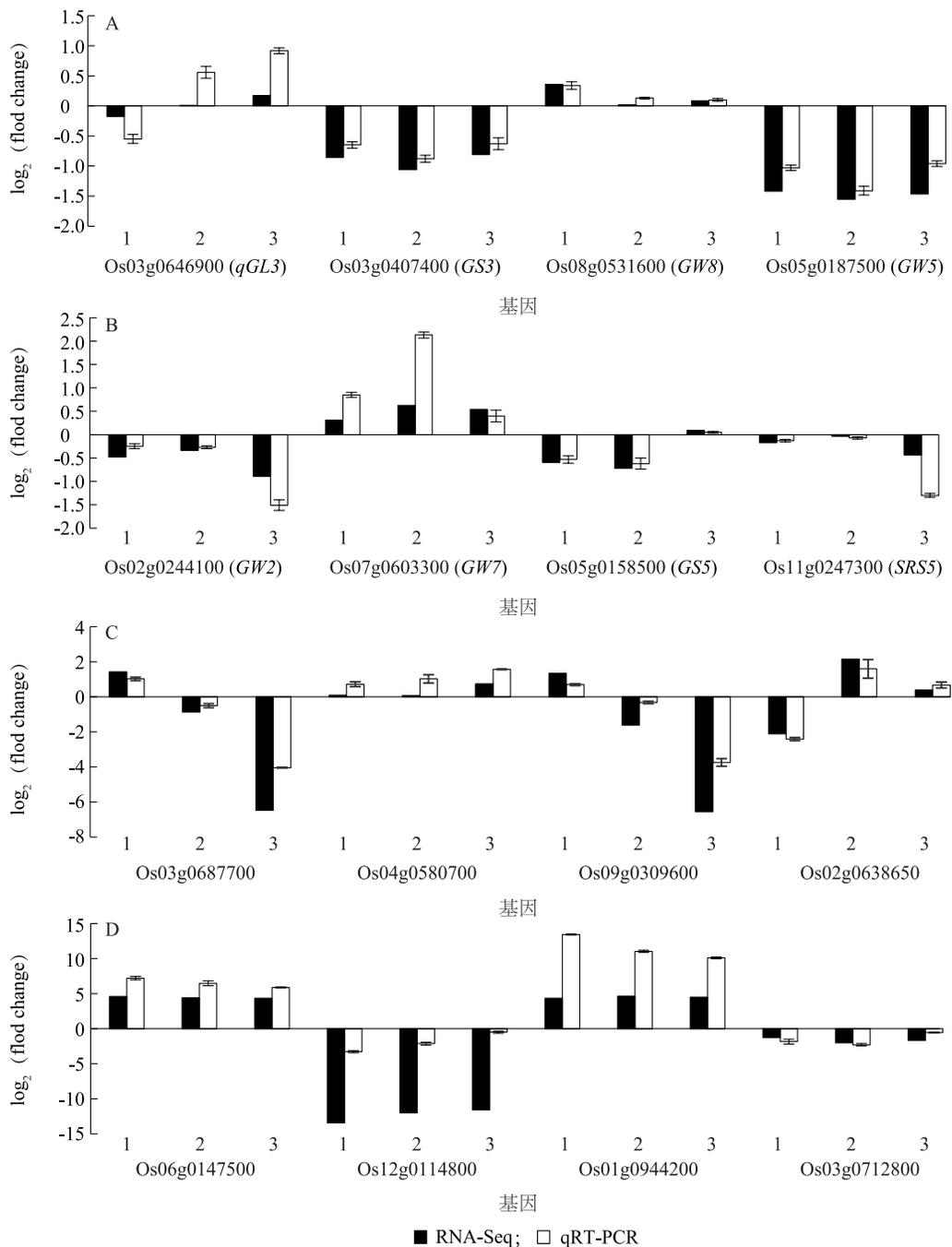
从转录组测序结果中选取 16 个基因, 包括 TD70 中已知的 7 个粒型基因 *GS3*、*qGL3*、*GW2*、*qSW5/GW5*、*GS5*、*GW7* 和 *GW8*。利用 qRT-PCR 分析这些基因在 2 个水稻材料不同发育时期幼穗中的相对表达水平, 对转录组数据进行可靠性验证。图 4 显示, 16 个基因的转录组测序结果与 qRT-PCR 结果一致, 表明转录组测序结果可靠。同时, 对粒型基因在不同发育时期幼穗中的表达模式进行分析, 结果发现, *qGL3* 基因在 TD70 幼穗发育早期下调表达, 随着幼穗的发育, 在中后期表现为上调表达; *GW7*、*GW8* 基因在 TD70 幼穗发育过程中始终上调表达; 其余 4 个粒型基因在 Kasalath 3 个幼穗发育时期的表达水平均高于 TD70。由此可见, 以上 7 个粒型基因在 2 个水稻材料幼穗中具有不同的表达模式, 对水稻籽粒发育调控的作用和时期亦不同。

2.6 2 个水稻材料幼穗中差异表达基因的 GO 富集与代谢通路分析

生物体内生物学功能的行使需要不同基因共同协调完成, 为了探究差异表达基因的生物学功能, 笔者对 2 个水稻材料 3 个幼穗发育时期的差异表达基因(数量分别为 3 618 个、4 183 个、5 254 个)按照 GO 的 3 个类别的生物进程 (Biological processes, BP)、细胞组分 (Cellular components, CC) 和分子功能 (Molecular functions, MF) 进行功能富集分析。结果显示, 共有 4 374 个差异表达基因得到注释, P1、P2 和 P3 这 3 个时期分别注释到 1 825 个 (50.44%)、2 267 个 (54.20%)、2 929 个 (55.75%) 基因。在水

稻幼穗发育的 P1 阶段, 代谢进程 (GO 编号: 0008152)、DNA 结合 (GO 编号: 0005488) 及细胞进程 (GO 编号: 0005623) 分别是生物学进程、分子功能和细胞组分 3 个类别中富集到最多基因的分组。差异表达基因在生物学途径中的生物进程调节 (GO 编号: 0050789)、信号 (GO 编号: 0023052)、生殖进程 (GO 编号: 0022414) 及发育过程 (GO 编号: 0032502) 等二级分类中较为集中。同时, 在生长 (GO 编号: 0040007)、细胞增殖 (GO 编号: 0008283) 等可能与水稻颖壳大小密切相关的代谢途径中也有差异表达基因显著富集。在分子功能方面, 差异表达基因主要富集在催化活性 (GO 编号: 0003824)、转运活性 (GO 编号: 0005215) 及分子功能调节 (GO 编号: 0098772) 等功能类别中。在细胞组分方面, 膜系统 (GO 编号: 0016020)、细胞器 (GO 编号: 0043226) 及蛋白质复合物 (GO 编号: 0032991) 等细胞代谢相关的重要组分的差异表达基因较多。通过以上 GO 富集分析发现, 在水稻幼穗发育过程中, 基因的转录调控、物质的合成与运输以及信号转导等代谢活动较为活跃。此外, 随着水稻幼穗的生长发育, 富集到的差异表达基因数量逐渐增加, 说明在幼穗发育后期, 生物学进程更加活跃、复杂, 因而参与调控的基因也更多 (图 5)。

利用 KEGG 数据库对差异表达基因进行代谢通路分析。结果表明, 在水稻的 3 个幼穗发育时期分别有 217 个、258 个、332 个基因得到注释, 共涉及 119 条 KEGG 代谢通路。同一个 KEGG 中包含 10 个以上差异表达基因的通路主要有内质网中的蛋白质加工通路、淀粉和蔗糖代谢通路、激素信号转导通路、氨基酸生物



A: 识别号 (ID) 为 Os03g0646900、Os03g0407400、Os08g0531600、Os05g0187500 基因分析; B: 识别号 (ID) 为 Os02g0244100、Os07g0603300、Os05g0158500、Os11g0247300 基因分析; C: 识别号 (ID) 为 Os03g0687700、Os04g0580700、Os09g0309600、Os02g0638650 基因分析; D: 识别号 (ID) 为 Os06g0147500、Os12g0114800、Os01g0944200、Os03g0712800 基因分析。fold change 表示某个基因在 2 个水稻材料幼穗中的表达量。x 坐标轴上的 1、2、3 分别表示比较组 T-P1/K-P1、T-P2/K-P2、T-P3/K-P3; T: TD70; K: Kasalath; P1、P2、P3: 孕穗期 1~3 cm、4~7 cm、8~10 cm 长度的幼穗。

图 4 2 个水稻材料幼穗中差异表达基因的 qRT-PCR 分析

Fig.4 The qRT-PCR analysis of differentially expressed genes in two rice cultivars

合成通路、碳代谢通路以及核糖体通路等,其中内质网中的蛋白质加工通路在水稻的 3 个幼穗发育阶段都被显著富集,淀粉和蔗糖代谢通路、碳代谢通路在水稻幼

穗发育后期显著富集。此外,次生代谢产物的生物合成通路也富集到大量基因,表明次生代谢产物在水稻幼穗发育过程中可能也有重要作用(表 3)。

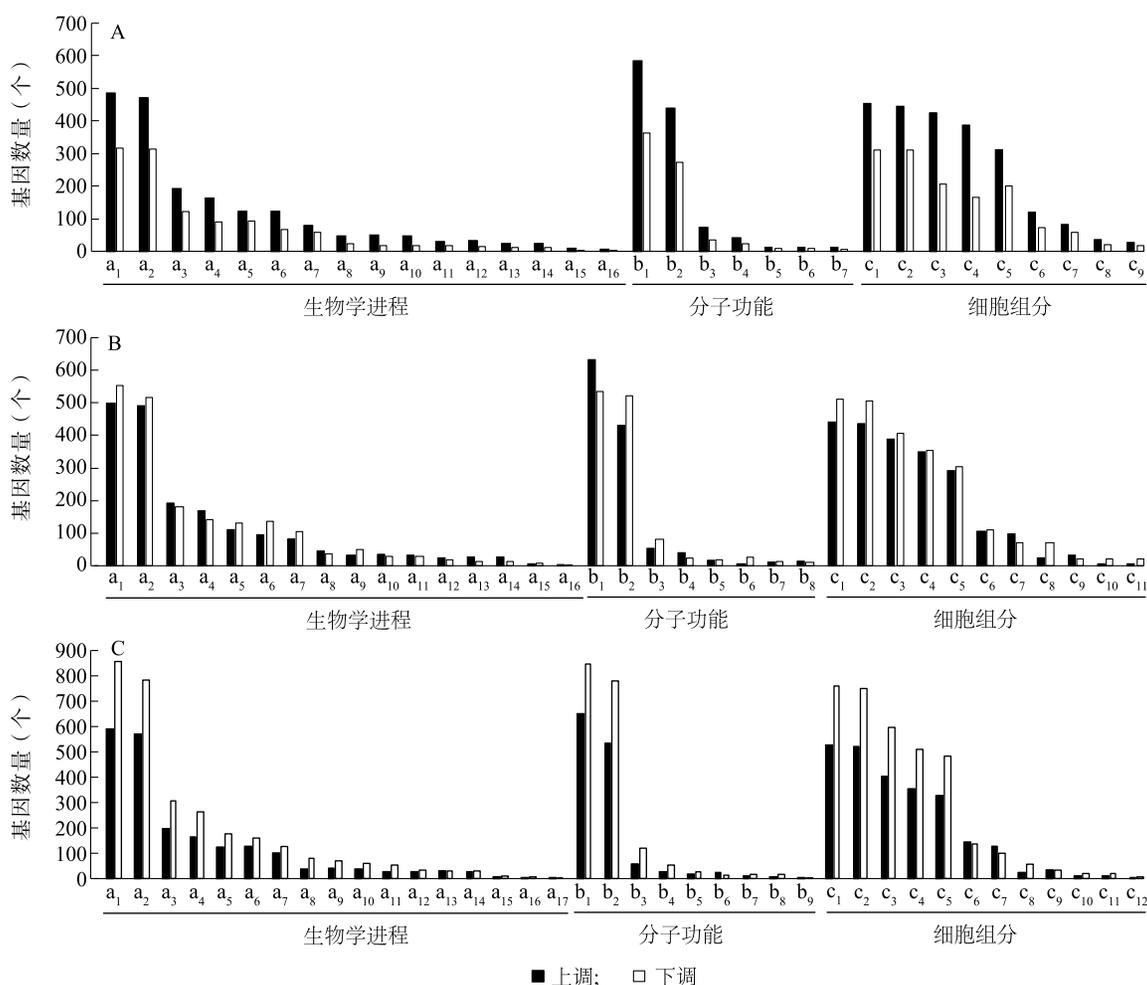


图5 TD70和Kasalath在水稻幼穗不同时期差异表达基因的功能分析

图A中, a₁:代谢进程; a₂:细胞进程; a₃:生物调节; a₄:生物进程调节; a₅:刺激反应; a₆:定位; a₇:细胞组分组织或生物发生; a₈:信号; a₉:多细胞有机体进程; a₁₀:发育进程; a₁₁:生物进程正调节; a₁₂:多种生物进程; a₁₃:生殖进程; a₁₄:生殖; a₁₅:生长; a₁₆:细胞增殖; b₁:DNA结合; b₂:催化活性; b₃:转运活性; b₄:转运调节活性; b₅:结构分子活性; b₆:分子功能调节; b₇:抗氧化活性; c₁:细胞; c₂:细胞部分; c₃:膜; c₄:膜部分; c₅:细胞器; c₆:细胞器部分; c₇:蛋白质复合体; c₈:胞外区; c₉:膜封闭腔。图B中, a₁:代谢进程; a₂:细胞进程; a₃:生物调节; a₄:生物进程调节; a₅:刺激反应; a₆:定位; a₇:细胞组分组织或生物发生; a₈:信号; a₉:发育进程; a₁₀:多细胞有机体进程; a₁₁:生物进程正调节; a₁₂:多种生物进程; a₁₃:生殖进程; a₁₄:生殖; a₁₅:生长; a₁₆:细胞增殖; b₁:DNA结合; b₂:催化活性; b₃:转运活性; b₄:转运调节活性; b₅:分子功能调节; b₆:抗氧化活性; b₇:结构分子活性; b₈:分子转运活性; c₁:细胞; c₂:细胞部分; c₃:膜; c₄:膜部分; c₅:细胞器; c₆:细胞器部分; c₇:蛋白质复合体; c₈:胞外区; c₉:膜封闭腔; c₁₀:细胞连接; c₁₁:共质体。图C中, a₁:代谢进程; a₂:细胞进程; a₃:生物调节; a₄:生物进程调节; a₅:刺激反应; a₆:定位; a₇:细胞组分组织或生物发生; a₈:发育进程; a₉:信号; a₁₀:多细胞有机体进程; a₁₁:生物进程正调节; a₁₂:多种生物进程; a₁₃:生殖; a₁₄:生殖进程; a₁₅:生长; a₁₆:免疫系统过程; a₁₇:细胞增殖; b₁:DNA结合; b₂:催化活性; b₃:转运活性; b₄:转运调节活性; b₅:分子功能调节; b₆:结构分子活性; b₇:分子转运活性; b₈:抗氧化活性; b₉:分子转运活性; c₁:细胞; c₂:细胞部分; c₃:膜; c₄:膜部分; c₅:细胞器; c₆:细胞器部分; c₇:蛋白质复合体; c₈:胞外区; c₉:膜封闭腔; c₁₀:细胞连接; c₁₁:共质体; c₁₂:超分子复合物。

图5 TD70和Kasalath在水稻幼穗不同时期差异表达基因的功能分析

Fig.5 Functional analysis of differentially expressed genes in TD70 and Kasalath at different stages of young panicles

2.7 2个水稻材料幼穗中差异表达的转录因子分析

转录因子通过结合到靶基因启动子区域特定的元件上以调控基因的表达,在植物生长发育中起着

重要作用。已克隆的粒型基因 *GW8* 基因属于 SBP 转录因子家族, *GLW7* 属于 SPL (Squamosa promoter binding protein-like) 转录因子家族, *LG3* 属于 AP2 转

录因子家族, *OsMADS1* 为 MADS-box 家族转录因子^[6,8,15,38]。由此可见, 转录因子在水稻幼穗发育中有重要的调控作用。通过转录因子数据库 (<http://plantfdb.cbi.pku.edu.cn/>) 和 NCBI 的 nr 数据库对差异表达基因进行注释, 共得到来自 43 个家族的 349 个已知或者预测的转录因子, 如图 6 所示, 其中有 3 个 ARF(生长素相应因子) 家族的转录因子, 32 个 bHLH(Basic helix-loop-helix) 家族的转录因子, 34 个 ERF(乙烯相应因子) 家族的转录因子, 10 个 MADS-box 家族的转录因子, 这些差异表达的转录因子可

能对幼穗发育具有重要的调控作用。进一步分析发现, 与 Kasalath 相比, TD70 中粒型基因 *OsLG3* 在 P1 期显著上调表达, 在 P2、P3 期的表达量没有显著差异; *GW8* 基因在幼穗中 3 个时期均上调表达, 但差异不显著, *GLW7* 基因的表达模式与 *GW8* 基因相反; *OsMADS1* 基因表达量在 2 个材料间差异不显著。这可能是由于 TD70 与 Kasalath 的遗传背景差异较大, 同时 TD70 中聚合了多个粒型基因, 而不同基因在调控水平上可能存在互作等。

表 3 KEGG 通路中 10 个以上差异表达基因的代谢通路信息

Table 3 Metabolic pathways information of more than ten differentially expressed genes in KEGG pathway

差异表达基因集	通路名称	差异表达基因数 (个)	差异表达基因数占 总基因数的比例(%)	P 值	KEGG 数据库通路号
T-P1/K-P1	内质网中的蛋白质加工	25	11.52	1.25×10^{-5}	ko04141
	苯丙烷生物合成	16	7.37	0.029 125	ko00500
	淀粉和蔗糖代谢	12	5.53	0.048 287	ko00940
	剪接体	12	5.53	0.080 409	ko04016
T-P2/K-P2	次生代谢产物的生物合成	88	34.11	0.000 207	ko01110
	苯丙烷生物合成	33	12.79	5.61×10^{-8}	ko00940
	碳代谢	23	8.91	0.004 091	ko00500
	淀粉和蔗糖代谢	16	6.20	0.004 647	ko00630
	内质网中的蛋白质加工	16	6.20	0.010 856	ko04146
	植物激素信号转导	14	5.43	0.020 645	ko00710
	氨基酸的生物合成	12	4.65	0.041 443	ko01200
T-P3/K-P3	次生代谢产物的生物合成	110	33.13	0.000 125	ko01110
	苯丙烷生物合成	26	7.83	0.000 159	ko00380
	碳代谢	24	7.23	0.006 078	ko00940
	内质网中的蛋白质加工	19	5.72	0.013 557	ko00480
	糖酵解/糖异生	18	5.42	0.020 041	ko00010
	核糖体	17	5.12	0.026 393	ko04146

K 表示 Kasalath; T 表示 TD70; P1、P2 和 P3 分别代表孕穗期 1~3 cm、4~7 cm 和 8~10 cm 长度的幼穗。

3 讨论

水稻粒型基因的研究一直倍受重视, 目前研究者已经定位并克隆了大量粒型基因, 然而水稻超大籽粒形成的分子机制尚不清晰。本研究对相同发育阶段 TD70、Kasalath 水稻的幼穗进行比较, 在 P1、P2 和 P3 发育阶段分别获得 3 618 个、4 183 个和 5 254 个差异表达基因, 差异表达基因的数量随着幼穗的发育逐渐增加, 可能由于在幼穗发育后期, 调控网络更

加复杂, 参与的基因也更多。GO 富集分析显示, 这些差异表达基因涉及的生物学途径主要集中在信号 (GO 编号: 0023052)、细胞增殖 (GO 编号: 0008283)、催化活性 (GO 编号: 0003824)、转运活性 (GO 编号: 0005215)、膜系统 (GO 编号: 0016020)、蛋白质复合物 (GO 编号: 0032991) 等细胞代谢相关的重要途径。由此可见, 孕穗期幼穗的发育涉及信号转导、细胞分裂、物质转运等重要进程, 这些生物学途径在孕穗期对水稻幼穗的发育起到至关重要的

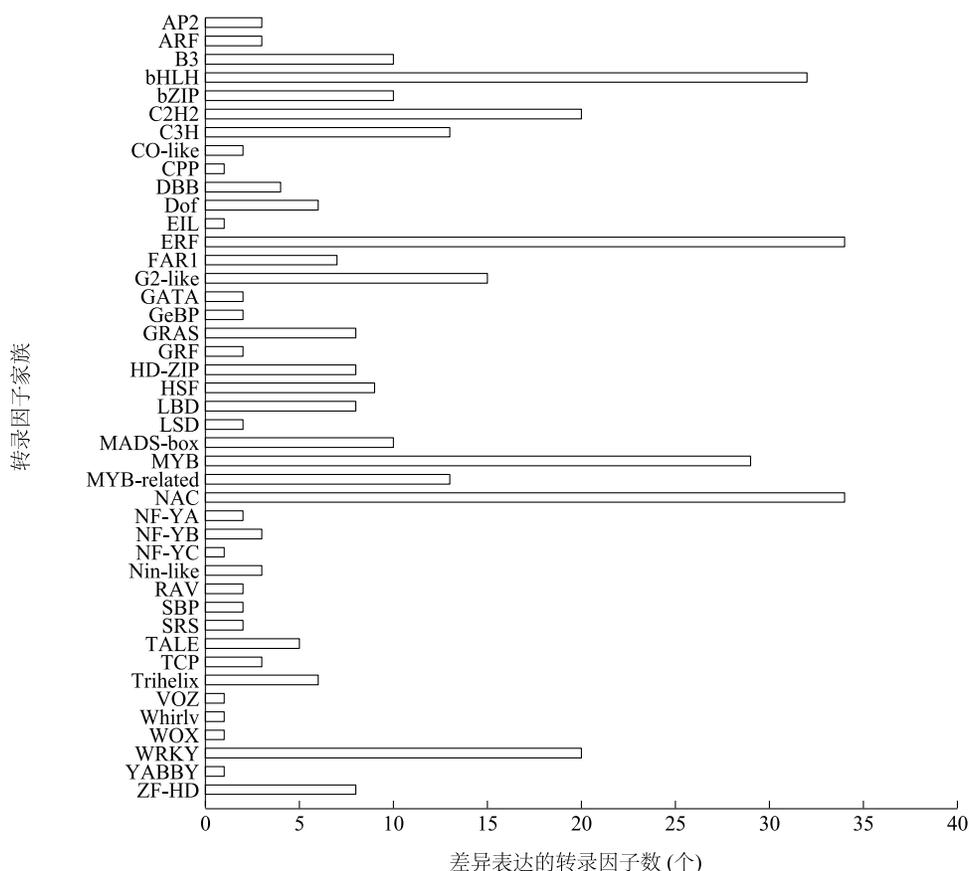


图6 2个水稻材料幼穗中差异表达的转录因子分析

Fig.6 Analysis of differentially expressed transcription factors in two rice materials

作用。以上研究结果与前人对水稻幼穗的转录组研究结果基本一致^[31-33]。研究发现, *GS5*、*GW7/GL7*、*GW8* 等粒型基因通过调控水稻颖壳细胞周期的变化而调控粒型^[14-16,21]。TD70 超大籽粒是多个粒型基因聚合的结果,不同粒型基因涉及膜蛋白、激素以及各类酶类等大分子相关途径,通过调节颖壳细胞分裂和生长而最终影响籽粒大小^[24-26]。通过 KEGG 数据库对幼穗 3 个发育时期都差异表达的基因进行代谢通路的富集分析,结果显示,差异表达基因共涉及 119 条 KEGG 代谢通路;内质网中的蛋白质加工、淀粉和蔗糖代谢、碳代谢、激素信号转导等多个通路代谢旺盛。细胞周期蛋白和细胞分裂素相关基因在幼穗发育的 3 个时期表现出显著差异,推测在幼穗发育过程中 2 个材料颖壳细胞分裂的差异是后期二者籽粒大小差异的重要原因。通过 GO 和 KEGG 分析发现,调控通路的差异涉及生命大分子物质的合成与代谢、激素信号转导及膜系统等,可以为水稻幼穗细胞分裂、生长与分化等生命过程提供重要的物

质基础、能量来源以及重要的分子信号等。

转录因子作为重要的调节蛋白,在水稻穗发育和籽粒形成过程中起着重要作用^[39]。本研究发现在 2 个水稻材料幼穗中 349 个转录因子的表达量存在显著差异,其中包括 AP2、ARF、bHLH、MYB、MADS 等家族的转录因子。在多个物种中的功能鉴定结果表明,植物中 MYB 转录因子家族具有调节植物发育、开花时间等不同的功能^[40]。MADS 转录因子家族在植物生殖器官发育中起到重要的调节作用,*OsMADS1* 基因与水稻 G 蛋白互作对籽粒大小具有调控作用^[8]。综上所述,转录因子对水稻孕穗期幼穗的发育和籽粒形态有重要的调控作用。

本研究通过对不同大小的水稻幼穗进行转录组深度测序,系统地揭示了特大粒水稻 TD70 幼穗发育在不同时期表达谱的变化,为进一步研究水稻籽粒发育过程中基因互作网络以及粒型基因在分子育种中的有效利用提供了理论依据。

参考文献:

- [1] 杨联松,白一松,张培江,等. 谷粒形状与稻米品质相关性研究[J]. 杂交水稻,2001,16(4):48-50,54.
- [2] 徐正进,陈温福,马殿荣,等. 稻谷粒形与稻米主要品质性状的关系[J]. 作物学报,2004,30(9):894-900.
- [3] FAN C C, XING Y Z, MAO H L, et al. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein[J]. Theoretical and Applied Genetics, 2006, 112(6): 1164-1171.
- [4] QI P, LIN Y S, SONG X J, et al. The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3[J]. Cell Research, 2012, 22(12): 1666-1680.
- [5] ZHANG X J, WANG J F, HUANG J, et al. Rare allele of *OsPP-KL1* associated with grain length causes extra-large grain and a significant yield increase in rice[J]. Proceedings of the National Academy of Sciences, 2012, 109(52): 21534-21539.
- [6] SI L Z, CHEN J Y, HUANG X H, et al. *OsSPL13* controls grain size in cultivated rice[J]. Nature Genetics, 2016, 48(4): 447-456.
- [7] YING J Z, MA M, BAI C, et al. *TGW3*, a major QTL that negatively modulates grain length and weight in rice[J]. Molecular Plant, 2018, 11(5): 750-753.
- [8] LIU Q, HAN R X, WU K, et al. G-protein $\beta\gamma$ subunits determine grain size through interaction with MADS-domain transcription factors in rice[J]. Nature Communications, 2018, 9(1): 1-12.
- [9] ZHAO D S, LI Q F, ZHANG C Q, et al. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality[J]. Nature Communications, 2018, 9(1): 1240.
- [10] SONG X J, HUANG W, SHI M, et al. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase[J]. Nature Genetics, 2007, 39(5): 623-630.
- [11] SHOMURA A, IZAWA T, EBANA K, et al. Deletion in a gene associated with grain size increased yields during rice domestication[J]. Nature Genetics, 2008, 40(8): 1023-1028.
- [12] WENG J F, GU S H, WAN X Y, et al. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight[J]. Cell Research, 2008, 18(12): 1199-1209.
- [13] LI Y B, FAN C C, XING Y Z, et al. Natural variation in *GS5* plays an important role in regulating grain size and yield in rice[J]. Nature Genetics, 2011, 43(12): 1266-1269.
- [14] WANG Y X, XIONG G S, HU J, et al. Copy number variation at the *GL7* locus contributes to grain size diversity in rice[J]. Nature Genetics, 2015, 47(8): 944-948.
- [15] WANG S K, LI S, LIU Q, et al. The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality[J]. Nature Genetics, 2015, 47(8): 949-954.
- [16] WANG S K, WU K, YUAN Q B, et al. Control of grain size, shape and quality by *OsSPL16* in rice[J]. Nature Genetics, 2012, 44(8): 950-954.
- [17] WANG S S, WU K, QIAN Q, et al. Non-canonical regulation of SPL transcription factors by a human OTUB1-like deubiquitinase defines a new plant type rice associated with higher grain yield[J]. Cell Research, 2017, 27(9): 1142-1156.
- [18] HUANG K, WANG D K, DUAN P G, et al. Wide and thick *GRAIN 1*, which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice[J]. Plant Journal, 2017, 91(5): 849-860.
- [19] MAO H L, SUN S Y, YAO J L, et al. Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice[J]. Proceedings of the National Academy of Sciences, 2010, 107(45): 19579-19584.
- [20] SUN S Y, WANG L, MAO H L, et al. A G-protein pathway determines grain size in rice[J]. Nature Communications, 2018, 9(1): 851.
- [21] LIU J F, CHEN J, ZHENG X M, et al. *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice[J]. Nature Plants, 2017, 3: 17043.
- [22] XU C J, LIU Y, LI Y B, et al. Differential expression of *GS5* regulates grain size in rice[J]. Journal of Experimental Botany, 2015, 66(9): 2611-2623.
- [23] 丁丹. 水稻5个粒型相关基因的分子标记开发与效应分析[D]. 南京:南京农业大学, 2014.
- [24] 张亚东,张颖慧,董少玲,等. 特大粒水稻材料粒型性状的QTL检测[J]. 中国水稻科学, 2013, 27(2): 122-128.
- [25] ZHANG Y D, ZHENG J, LIANG Z K, et al. Verification and evaluation of grain QTLs using RILs from TD70 \times Kasalath in rice[J]. Genetics & Molecular Research, 2015, 14(4): 14882-14892.
- [26] ZHANG Y D, ZHAO Q Y, ZHAO C F. Distribution of seven grain genes and evaluation of their genetic effects on grain traits[J]. Pakistan Journal of Botany, 2016, 48(3): 1073-1079.
- [27] WANG L, LI P H, BRUTNELL T P. Exploring plant transcriptomes using ultra high-throughput sequencing[J]. Briefings in Functional Genomics, 2010, 9(2): 118-128.
- [28] GUO H B, MENDRIKAHY J N, XIE L, et al. Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis[J]. Scientific Reports, 2017, 7: 40139.
- [29] SHANKAR R, BHATTACHARJEE A, JAIN M. Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses[J]. Scientific Reports, 2016, 6: 23719.
- [30] GONZÁLEZ-SCHAIN N, DRENI L, LAWAS L M F, et al. Genome-wide transcriptome analysis during anthesis reveals new insights into the molecular basis of heat stress responses in tolerant and sensitive rice varieties[J]. Plant & Cell Physiology, 2016, 57(1): 57-68.
- [31] WANG J, ZHANG Q, WANG Y, et al. Analysing the rice young

- panicle transcriptome reveals the gene regulatory network controlled by TRIANGULAR HULL1 [J]. *Rice*, 2019, 12(1): 6.
- [32] ZHANG W H, SUN P Y, HE Q, et al. Transcriptome analysis of near-isogenic line provides novel insights into genes associated with panicle traits regulation in rice [J]. *PLoS One*, 2018, 13(6): e0199077.
- [33] KE S, LIU X J, LUAN X, et al. Genome-wide transcriptome profiling provides insights into panicle development of rice (*Oryza sativa* L.) [J]. *Gene*, 2018, 675: 285-300.
- [34] TRAPNELL C, ROBERTS A, GOFF L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks [J]. *Nature Protocols*, 2012, 7(3): 562-578.
- [35] KAWAHARA, Y BASTIDE M D L, HAMILTON J P, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data [J]. *Rice*, 2013, 6(1): 1-10.
- [36] LI H, HANDSAKER B, WYSOKER A, et al. The sequence alignment/map (SAM) format and SAMtools [J]. *Bioinformatics*, 2009, 25(1/2): 1653-1654.
- [37] LIVAK K J, SCHMITTGEN T D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method [J]. *Methods*, 2001, 25(4): 402-408.
- [38] YU J, XIONG H, ZHU X, et al. *OsLG3* contributing to rice grain length and yield was mined by Ho-LAMap [J]. *BMC Biology*, 2017, 15(1): 28.
- [39] LI N, XU R, LI Y H. Molecular networks of seed size control in plants [J]. *Annual Review of Plant Biology*, 2019, 70(1): 1-30.
- [40] DUBOS C, STRACKE R, GROTEWOLD E, et al. MYB transcription factors in *Arabidopsis* [J]. *Trends in Plant Science*, 2010, 15(10): 573-581.

(责任编辑:徐 艳)