

基于叶球转录组数据比较的甘蓝杂种优势分析

李升娟, 许忠民*, 郭 佳, 张恩慧, 姜 娇, 石汶汶

(西北农林科技大学园艺学院, 陕西杨凌 712100)

摘 要: 对甘蓝两个杂交组合 F_1 的杂种优势进行了田间性状统计分析以及叶球转录组测序分析。所调查的 9 个园艺学性状中, 单球质量和球主叶柄质量在 F_1 代及其亲本之间差异显著, 其中单球质量的中亲优势和超亲优势在两个组合 F_1 中表现突出, 即产量杂种优势较为明显。通过叶球转录组分析, 分别筛选获得了 4 组差异表达基因, 在相同标准下, 两个杂交组合中父本与 F_1 代的差异表达基因明显多于母本与 F_1 代的差异表达基因, 表明母本的表达谱与 F_1 更相似, 即母本在 F_1 叶球杂种优势形成中的贡献较大, 而且上调差异表达基因的差异倍数明显高于下调差异表达基因, 表明上调基因在 F_1 代甘蓝叶球杂种优势建成中有重要作用。进一步将差异表达基因进行 GO (Gene ontology) 分类、COG (Cluster of orthologous groups of proteins) 分类、KEGG (Kyoto encyclopedia of genes and genomes) 通路富集分析以及可变剪接分析, 发现差异表达基因显著富集到生长发育、碳水化合物转运和代谢、信号转导以及氨基酸的合成、转运和代谢等途径。随机选取 7 个差异表达基因进行实时荧光定量 PCR 验证, 结果与 RNA-seq 数据基本一致, 证明转录组数据的可靠性。本研究获得的与甘蓝叶球杂种优势形成相关的差异表达基因, 为后续甘蓝杂种优势分子机制研究提供数据支持。

关键词: 甘蓝; 杂种优势; 转录组分析; 差异表达基因

中图分类号: S 635.1

文献标志码: A

文章编号: 0513-353X (2019) 06-1079-14

Comparative Heterosis Analysis of Cabbage Based on Transcriptome Data of Cabbage Heads Leaves

LI Shengjuan, XU Zhongmin*, GUO Jia, ZHANG Enhui, JIANG Jiao, and SHI Wenwen

(College of Horticulture, Northwest A & F University, Yangling, Shaanxi 712100, China)

Abstract: In two F_1 hybrids, we performed heterosis analysis through field traits survey and transcriptome sequencing of cabbage heads leaves. The result found that among the nine horticultural traits, head weight and main petiole weight showed significant differences between F_1 hybrids and their parents, among which the head weight exhibited a high value of MPH (mid-parent heterosis) and HPH (high-parent heterosis) in the two F_1 hybrids, which means the yield heterosis is obvious. Four groups of differentially expressed genes (DEGs) were compared through transcriptome analysis of cabbage heads leaves. We found that the number of DEGs between F_1 hybrids and paternal lines were significantly higher than that between F_1 hybrids and their maternal lines in the same criteria. It means similar gene expression

收稿日期: 2019-03-25; **修回日期:** 2019-05-27

基金项目: 国家重点研发计划项目 (2016YFD0101702); 国家现代农业产业技术体系建设专项资金项目 (CARS-25); 陕西省重点研发计划项目 (2018NY-059); 西安市科技计划项目 (201806113YF01NC09)

* 通信作者 Author for correspondence (E-mail: xuzhongmin2003@126.com)

profile between maternal lines and their F₁ hybrids or a bigger contribution of maternal lines than paternal lines in the formation process of cabbage heads leaves of F₁ hybrids. Besides, the fold change of up-regulated genes were higher than that of down-regulated genes, which implied an important role of up-regulated genes in the formation process of cabbage heads leaves of F₁ hybrids. The DEGs were further conducted GO (Gene Ontology), COG (Cluster of Orthologous Groups of proteins), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment and alternative splicing analysis. The result found that DEGs are highly involved in growth and development, carbohydrate transport and metabolism, signal transduction and amino acid biosynthesis, transport and metabolism pathways. Seven DEGs were randomly selected to perform qRT-PCR and the result was basically consistent with RNA-seq, demonstrating the reliability of transcriptome data. The DEGs that we screened associated with the formation process of heterosis of cabbage heads leaves, able to provide data support for the subsequent research on the molecular mechanism of cabbage heterosis.

Keywords: *Brassica oleracea* L. var. *capitata*; heterosis; transcriptome analysis; differentially expressed genes (DEGs)

杂种优势是指具有明显遗传差异的双亲杂交产生的 F₁ 代在品质、抗逆以及适应性等方面明显优于双亲的现象。1910 年, 显性假说被最早用于杂种优势现象的解释 (Bruce, 1910), 后来, 超显性假说 (East, 1936)、上位性假说 (Minvielle, 1987; Cheverud & Rautman, 1995) 陆续被学者提出, 但时至今日对杂种优势现象的解释仍未取得一致。随着近年来生物技术的不断发展, 许多学者发现可以从基因差异表达的角度阐述杂种优势。Romagnoli 等 (1990) 最先发现杂种优势与基因差异表达之间的关系, 在研究玉米杂种根尖部位杂种优势时, 利用重组 cDNA 文库杂交筛选获得了与根尖杂种优势相关的差异基因。随后, 表达序列标签技术 (expression sequence tags, EST)、基因芯片技术 (microarray)、SAGE (serial analysis of gene expression tags) 标签技术以及转录组 RNA-seq 技术等都被应用于杂种优势相关基因的研究中, 例如 Song 等 (2010) 利用 SAGE 标签筛选获得了与水稻杂种优势相关的 1 183 个差异表达基因; Ge 等 (2008) 利用表达序列标签技术筛选获得了 191 个水稻 F₁ 代和其亲本之间的差异表达基因等。

转录组 RNA-seq 技术是近几年发展起来的分子生物学技术之一, 是通过高通量测序技术把特定组织或细胞在某一发育阶段或功能状态下转录出来的所有 RNA 序列测出来, 反映出它们的表达水平 (Costa et al., 2010), 转录组能够从整体水平研究基因功能以及基因结构, 揭示特定生物学过程以及疾病发生过程中的分子机理。利用 RNA-seq 技术对蔬菜作物杂种优势的研究较少, 在其他农作物上较多, Zhang 等 (2017) 利用 RNA-seq 技术对大豆两个杂交组合 F₁ 代花器官的杂种优势进行分析, 分别获得了 681 和 899 个 F₁ 与亲本之间的差异表达基因, 这些差异表达基因主要富集在代谢过程、催化活性和次级代谢产物的生物合成等途径, 对大豆花器官发育杂种优势的建成起到重要作用; Zhai 等 (2013) 在分析杂交水稻 ‘Xieyou 9308’ 根部杂种优势时利用 RNA-seq 技术在分蘖期和抽穗期分别获得 829 和 4 186 个差异表达基因。

杂种优势的分子机制研究除了差异基因的筛选与获得, 还包括对某一性状的杂种优势进行遗传分析 (Chen et al., 2018)、挖掘杂种内的基因表达遗传模式 (Bell et al., 2013) 以及对杂种内等位基因的表达情况进行分析 (Springer & Stupar, 2007) 等, 这些方法都是基于杂种与亲本或亲本与亲本间基因表达情况进行的系统分析。此外, 近几年来, 很多研究是从表观遗传学的角度分析杂种优

势的遗传机理, 许多重要的小 RNA、DNA 甲基化和组蛋白修饰过程被证实参与杂种中特定性状的表达以及生物节律杂种优势的调控当中 (Groszmann et al., 2013), 从转录组、代谢组、蛋白质组以及表观基因组等多个遗传学角度进行综合分析评价, 成为未来对植物杂种优势分子机制研究的重要方向之一。

本研究中利用 RNA-seq 技术分析甘蓝 (*Brassica oleracea* L. var. *capitata*) 两个杂交组合 F₁ 代与其各自父母本之间的叶球差异表达基因, 并对差异表达基因进行可变剪接分析、差异表达基因功能注释和功能富集等分析, 最后采用 qRT-PCR 验证获得的差异表达基因, 以期揭示甘蓝叶球杂种优势可能的分子机制。

1 材料与方法

1.1 试材及园艺学性状调查

供试甘蓝由西北农林科技大学园艺学院甘蓝课题组提供。自交系母本 QP13 (A) 和父本 QP03 (B) 杂交获得 F₁ (C); 雄性不育系母本 QP04CMS (D) 和自交系父本 DHP37 (E) 杂交获得 F₁ (F) (图 1)。QP13 是由日本‘珍奇’甘蓝通过 7 年连续自交分离筛选出的性状整齐稳定的甘蓝自交系, 从定植到叶球成熟 65 d 左右, 外叶灰绿色, 蜡粉多, 抗病, 耐裂球, 叶球圆形, 叶球紧实度 0.62。QP03 是由意大利 G507 自交 2 代后进行小孢子培养获得的 DH 系, 从定植到叶球成熟 68 d 左右, 外叶绿色, 蜡粉较少, 抗病, 耐裂球, 叶球正圆形, 外叶直立, 叶球紧实度 0.68。QP04CMS 为细胞质雄性不育系, 通过甘蓝胞质雄性不育材料与抗病、经济性状优良、配合力高的自交系 (保持系) 杂交后, 选择与保持系性状一致的不育株回交, 经 7 年回交转育而成, 其经济性状稳定, 蜜腺发达, 雄蕊全部退化, 无花粉, 雌蕊正常, 植株生长健壮, 从定植到叶球成熟 80 d 左右, 外叶灰绿色, 蜡粉多, 抗病, 耐裂球, 叶球圆形, 叶球紧实度 0.59。DHP37 为甘蓝品种‘8398’经 3 代自交后进行小孢子培养获得的 DH 系, 从定植到叶球成熟 50 d 左右, 外叶绿色, 蜡粉少, 抗病, 耐裂球, 冬性强, 叶球圆形, 叶球紧实度 0.57。



图 1 甘蓝两个组合亲本及其 F₁ 代共计 6 份材料

Fig. 1 Parents and their F₁ generation, total six materials in two groups

6 份材料于 2017 年 7 月 7 日播种, 按照随机区组排列种植于小区内, 每个品种种植 20~30 株, 3 次重复; 由于各试验材料成熟期有一定差异, 因此在取样时参照各自成熟期天数, 以 80% 叶片达到商品成熟为准, 在每小区中间 2 行取样 10 株进行调查和采样 (取株原则为能代表整个小区的平均生长情况, 除去最大株和最小株), 再分别收获称取每个小区的产量。调查性状包括株高、株幅、单球质量、中心柱长、球高、球横径、球主叶柄质量、外叶质量、外叶柄质量等 9 个园艺学性状, 分别计算中亲优势 (MPH) 和超亲优势 (HPH), $MPH(\%) = (F_1 - MP) / MP \times 100$; $MP = (P_1 + P_2) / 2$; $HPH(\%) = (F_1 - HP) / HP \times 100$ 。其中 MP 为两个亲本同一性状的平均值, HP 为两个亲本中较优良亲本的性状值, F_1 为杂交种, P_1 为母本, P_2 为父本。利用 SPSS 16.0 中的 One-way ANOVA, LSD 和 Duncan's 测验进行显著性分析 ($P < 0.05$)。

1.2 RNA 提取和文库构建

甘蓝材料按各自成熟期采样, 剪取 10 cm² 大小的内叶迅速置于液氮中, 每个样品 3 个生物学重复, 为了减小个体间造成的差异, 每重复由 5 个单株混样。使用 TRIzol 试剂盒提取相应材料的 RNA (Tiangen, 北京), 提取后对 RNA 的完整性进行检测。使用寡聚 T 连接的磁珠从总 RNA 中纯化 mRNA, 在 NEBNext 第 1 链合成反应缓冲液 (5×) 中在升高的温度下使用二价阳离子进行破碎, 使用随机六聚体引物和 M-MuLV 逆转录酶合成第一链 cDNA。随后使用 DNA 聚合酶 I 和 RNA 酶 H 进行第 2 链 cDNA 合成, 通过核酸外切酶/聚合酶将剩余的突出端转化为平端, 在 3' 末端 DNA 片段腺苷酸化后, 连接具有发夹环结构的 NEBNext 衔接子以准备杂交。杂交前使用 AMPure XP 系统 (Beckman Coulter, Beverly, USA) 纯化筛选长度为 240 bp 的 cDNA 片段, 然后使用 3 μL 的 USER 酶 (NEB, USA) 与筛选的 cDNA 在 37 °C 下 15 min, 95 °C 下 5 min 进行 PCR 准备, 最后用 Phusion 高保真 DNA 聚合酶、通用 PCR 引物和 InDex (X) 引物进行 PCR, PCR 产物需经 AMPure XP 系统纯化, 最后在 Agilent Bioanalyzer 2100 系统上对文库质量进行评估。

1.3 数据过滤并与参考基因组比对

FASTQ 格式的原始数据 (raw reads) 首先通过内部 perl 脚本去除包含衔接子、poly-N 以及低质量的数据以获取干净数据 (clean reads), 计算干净数据的 Q20、Q30、GC 含量和序列重复水平。使用 Tophat 工具 (v 2.1.1; <http://ccb.jhu.edu/software/tophat/index.shtml>) 将干净的数据映射到参考基因组中 (Brassica oleracea v2.1; <ftp://ftp.ensemblgenomes.org/pub/plants/release-25/fasta/brassicaoleracea/>), 进一步分析读数的理想匹配度和单核苷酸多态性, 并根据参考基因组进行注释。利用 FPKM (每百万个片段的每千碱基转录片段) 来计算基因表达水平。

1.4 差异表达基因及其功能注释和富集分析

在差异表达基因检测过程中, 按照: (1) False Discovery Rate (FDR) ≤ 0.01 ; (2) Fold Change (FC) ≥ 2 标准筛选差异基因。差异倍数 (Fold Change) 表示两样品 (组) 间表达量的比值。错误发现率 (False Discovery Rate, FDR) 通过对差异显著性 P 值 (P -value) 校正获得, 差异分析采用 Benjamini-Hochberg 校正方法对原有假设检验得到的显著性 P 值 (P -value) 进行校正。

基于 GSeq R 包 (v 1.10.1; <http://www.bioconductor.org/packages/release/bioc/html/DESeq.html>) 的 Wallenius 非中心超几何分布 (Young et al., 2010) 进行差异表达基因的 Gene Ontology (GO) 富集分析, 使用 KOBAS 软件 (Mao et al., 2005) 分析 KEGG 途径中差异表达基因的富集情况。基于以下数据库注释基因功能: NCBI 非冗余蛋白质数据库; NCBI 非冗余核苷酸数据库; 大型蛋白质家

族数据库; 蛋白质直系同源数据库; 手动注释和鉴定蛋白质序列数据库; KEGG 同源基因数据库; 基因功能分类数据库。可变剪接分析使用 Cufflinks (<http://cufflinks.cbc.umd.edu/>) 对 Tophat 的比对结果进行拼接, 并使用 Cufflinks 的 Cuffcompare 将 Cufflinks 拼接结果与初始注释结果比较分析, 通过 ASprofile (<http://ccb.jhu.edu/software/ASprofile/>) 软件获取每个样品存在的可变剪接类型及相应表达量。

1.5 qRT-PCR 验证差异表达基因

为了验证 RNA-seq 数据结果的准确性, 随机选取 7 个差异表达基因进行荧光定量 PCR 验证, 内参基因为 *CYP*(GenBank: M55018.1), 引物序列见表 1。第 1 链反转录使用去 gDNA 的 Prime ScriptTM RT 第 1 链合成试剂盒(TaKaRa), 然后使用 abm[®] vaGreen qPCR MasterMix-No dye 试剂盒在 ABI 7500 定量 PCR 仪上进行 qRT-PCR 循环。循环参数为: 95 °C 10 min, 40 ~ 45 个循环: 94 °C 15 s, 60 °C 1 min。所有反应至少重复 3 次, 结果使用 2^{-ΔΔCt} 法 (Adnan et al., 2011) 对表达量数据进行分析。

表 1 实时荧光定量 PCR 基因引物序列
Table 1 Primer sequences for the quantification of transcripts by real-time PCR

基因或编号 Gene ID	引物 (5' – 3') Primer sequence
<i>CYP</i> (M55018.1)	F: AGGAGGAGATTTACCGC; R: TCTCTAACGACATCCATCCC
Bo4g169200	F: TGGTGTCTGCGTCCACGGTAG; R: TCCAGTGAGGCCAGCATACTCTC
Bo8g056920	F: ACTTCCGTGTCTCGCTAGGTACTC; R: TTCGCCTTCTCCTTCTCCTTCTCCTC
Bo01049s040	F: CCAACGGCATGGTGGATGGAC; R: GAGCCACGATCACCGAACCAAC
Bo7g011450	F: AAGTGCATCGAAGCCGTAGTTACC; R: GGTGGAGTAAGACGCATCACAAAGG
Bo7g011440	F: AGAGGAGCTTACGGTGTGGTCTG; R: CTCGTGCCTCACATGCCTAAGAAG
Bo3g127630	F: GCTGCTCTCAGATGCTCGCTATC; R: GCTCTGCCTCGGATTGTTGGATC
<i>Brassica oleracea</i> L_new Gene_1356	F: CGACGATGCACTTCTTCTCCAGAC; R: GCAGTCAACTCCGACGGAATCTAC

2 结果与分析

2.1 杂交组合的杂种优势分析

为了判定两个组合 F₁ 是否存在杂种优势, 对株高、株幅、单球质量、中心柱长、球高、球横径、球主叶柄质量、外叶质量以及外叶叶柄质量进行了调查统计, 并计算中亲优势和超亲优势。结果 (表 2) 表明, 两个杂交组合中亲优势值在 - 0.31 ~ 120.95 之间, 其中仅 F 杂交组合的外叶叶柄质量中亲值为 - 0.31, 表现为负向中亲优势, 其余性状均为正向中亲优势, 单球质量中亲值最高。此外, 两个杂交组合超亲优势值在 - 11.16 ~ 96.61 之间, 其中除 C 杂交组合的株高、株幅和外叶叶柄质量超亲值为负值, F 杂交组合的球横径和外叶叶柄质量超亲值为负值, 表现为负向超亲优势外, 其余性状在两个杂交组合中均表现为正向超亲优势, 单球质量同样在两个杂交组合中超亲值最高, 表明产量杂种优势较为明显。各性状显著性分析发现两个杂交组合中, 单球质量和球主叶柄质量的 F₁ 与其父母本之间具有显著差异, 以上结果表明两个组合 F₁ 杂种优势明显, 尤其表现在单球质量, 即产量上。

表 2 两个杂交组合亲本及其 F₁ 性状统计
Table 2 Statistical analysis of field traits in two groups of F₁ and their parents

亲本及其 F ₁ Parents and F ₁	株高/cm Plant height	株幅/cm Plant width	单球质量/ kg Head weight	中心柱长/cm Center column length	球高/cm Head height	球横径/cm Transverse diameter	球主叶柄 质量/g Main petiole weight	外叶质量/kg Outer leaves weight	外叶叶柄 质量/g Outer leaves stalk weight
A (♀)	19.23±4.35b	42.57±4.61b	0.46±0.01 d	5.23±0.15d	10.43±0.50d	11.13±0.42c	99.23±8.64d	0.26±0.02b	77.43±1.16d
B (♂)	29.57±2.39a	51.73±2.90a	0.59±0.07cd	6.97±0.23a	11.60±0.26d	11.37±0.45c	147.33±3.44c	0.48±0.10a	169.77±7.79a
C (F ₁)	26.27±0.25a	51.13±4.10a	1.16±0.19b	6.07±0.12bc	13.73±1.36c	14.47±0.15a	218.60±3.32a	0.50±0.02a	169.33±7.30a
D (♀)	28.13±2.33a	54.20±3.68a	0.85±0.08c	5.53±0.23cd	15.10±0.62bc	14.87±0.40a	148.63±1.76c	0.47±0.06a	119.20±4.57c
E (♂)	29.50±1.18a	41.50±0.70b	0.68±0.06cd	6.33±0.58b	16.43±0.71ab	13.07±0.71b	140.53±6.62c	0.43±0.02a	130.37±6.96b
F (F ₁)	31.03±3.71a	57.17±1.37a	1.51±0.34a	6.23±0.25b	17.63±1.63a	14.60±1.32a	204.87±11.68b	0.51±0.19a	124.40±4.01bc
C-MPH/%	7.65	8.45	120.95	28.96	24.66	28.59	77.32	34.53	37.00
F-MPH/%	7.69	19.47	97.39	34.74	11.84	4.53	41.69	13.75	- 0.31
C-HPH/%	- 11.16	- 1.16	96.61	12.92	18.39	27.27	48.37	3.45	- 0.26
F-HPH/%	10.31	5.47	78.35	34.34	7.30	- 1.79	37.83	9.29	- 4.58

2.2 转录组数据的获得及差异表达基因的筛选

经 IlluminaHiSeq 平台对两个杂交组合共 6 个样品进行转录组测序后, 去除含有接头和低质量的 Reads, 共获得 138.39 Gb 高质量的 Clean Data, 各样品均达到 6.28 Gb, Q30 碱基百分比在 90.86% 及以上。将各样品的 Clean Reads 与参考基因组进行序列比对发现, 比对到参考基因组的 Reads 数在 67.82%~69.74%之间, 唯一位置的 Reads 数在 64.36%~67.55%之间, 多处位置的 Reads 数在 1.62%~4.00%之间, 正链和负链的 Reads 数基本持平。另外, 文库 mRNA 片段化随机性检验、插入片段长度检验和转录组测序数据饱和度检验评估文库质量发现, mRNA 片段化随机性较高, 磁珠纯化效果较好, 转录组数据充足, 满足后续分析。

进一步将两个杂交组合的 F₁ 代转录组数据与其父母本分别进行比对, 将获得的差异表达基因及注释到 8 个数据库中, 得到如下结果 (表 3)。

表 3 甘蓝杂交组合 F₁ 代与各自父母本之间差异表达基因数量以及注释到 8 个数据库的差异表达基因数量统计
Table 3 Total DEGs statistics between cabbage F₁ generation and their parents, and the count of DEGs annotated into 8 databases

杂交组合 ♀ × ♂ →F ₁	比对 Comparison	差异表达基因数 DEGs number			注释到数据库的差异表达基因 Annotate DEGs								
		总数 Total	上调 Up	下调 Down	总数 Total	COG	GO	KEGG	KOG	NR	Pfam	Swiss-Prot	Egg-NOG
A × B→ C	C vs A	48	25	23	48	13	47	16	24	48	41	42	47
	C vs B	3 191	1 562	1 629	3 062	1 132	2 709	1 090	1 668	3 032	2 398	2 253	2 954
D × E → F	F vs D	2 170	1 454	716	2 093	787	1 857	688	1 080	2 077	1 644	1 551	2 006
	F vs E	3 420	2 098	1 322	3 254	1 221	2 868	1 143	1 765	3 227	2 549	2 353	3 140

两个组合 F₁ 代与其各自父母本获得的差异表达基因数量并不相同, 但表现出了一致的趋势, 即两个 F₁ 代与其父本之间的差异表达基因明显多于 F₁ 与母本之间的差异表达基因。在 A × B → C 的杂交组合中, 其 F₁ 代材料与母本比对 (C vs A) 中, 仅筛选出 48 个差异表达基因, 其中上调基因 25 个, 下调基因 23 个, 而在 F₁ 代与父本比对 (C vs B) 中, 共筛选出 3 191 个差异表达基因, 其中上调基因 1 562 个, 下调基因 1 629 个; 在 D × E → F 的杂交组合中, 其 F₁ 代材料与母本比对 (F vs D) 中, 筛选出 2 170 个差异表达基因, 其中上调基因 1 454 个, 下调基因 716 个, 在 F₁ 代与父本比对 (F vs E) 中, 共筛选出 3 420 个差异表达基因, 其中上调基因 2 098 个, 下调基因 1 322 个。因此, 两个杂交组合 F₁ 与其母本基因表达谱更相似, 推测母本在甘蓝叶球杂种优势中具有重要作用。在 A × B → C 的杂交组合中, 上、下调基因数量相似, 在 D × E → F 组合中, 上调基因数量明显多于

下调基因, 且在 4 对比对中上调基因的差异倍数明显高于下调基因, 证明上调基因在基因差异表达中的主要作用。在 8 个注释到的数据库中, NR 数据库注释到的差异表达基因数量最多, 其次为 EggNOG 和 GO 数据库, 差异表达基因在各数据库中的数量及一致趋势表明筛选差异表达基因的可靠性及正确性。

2.3 差异表达基因 GO 和 COG 分类

为了分析差异表达基因参与的代谢通路及生物学功能, 将差异表达基因进行 GO 分类 (表 4) 和 COG 分类 (表 5)。GO 分类包括 3 大类: 生物学过程、细胞组分和分子功能。表 4 展示了各分类下的 10 个主要词条, 两个杂交组合的 F₁ 代与其父母本之间的差异表达基因数量不同, 但富集到的主要词条大致相同。在生物学过程分类下的 4 对比对中, 差异表达基因富集到的前两个词条为: 细胞过程和单一生物过程, 第 3 位词条 C vs A 比对为胁迫应答, 其余比对为代谢过程。在细胞组分分类下 4 对比对中, 前 3 个富集词条都为细胞部位、细胞和细胞器。在分子功能分类下的 4 对比对

表 4 甘蓝杂交组合 F₁ 代与各自父母本之间差异表达基因在 GO 各分类中的数量
Table 4 The total number of DEGs in GO classification between F₁ hybrid and their parents in cabbage

GO 分类 GO classification	GO 序列号 GO accession	词条 Term	所有差异 基因数量 All DEGs number	A × B → C		D × E → F	
				C vs A	C vs B	F vs D	F vs E
生物学过程 Biological process	GO: 0009987	细胞过程 Cellular process	34 799	42	2 112	1 479	34 799
	GO: 0044699	单一生物过程 Single-organism process	32 069	38	1 981	1 411	32 069
	GO: 0008152	代谢过程 Metabolic process	31 428	32	1 968	1 360	31 428
	GO: 0050896	胁迫应答 Response to stimulus	26 336	36	1 643	1 193	26 336
	GO: 0065007	生物调控 Biological regulation	22 419	30	1 433	1 022	22 419
	GO: 0032502	发育过程 Developmental process	18 604	31	1 196	837	18 604
	GO: 0071840	细胞组分或生物合成 Cellular component organization or biogenesis	16 533	19	1 062	701	16 533
	GO: 0051179	定位 Localization	15 309	24	985	743	15 309
	GO: 0032501	多细胞生物过程 Multicellular organismal process	14 546	21	933	642	14 546
	GO: 0040007	生长 Growth	5 572	8	379	264	374
细胞组分 Cellular component	GO: 0044464	细胞部位 Cell part	42 814	45	2 562	1 738	2 723
	GO: 0005623	细胞 Cell	42 578	45	2 546	1 726	2 710
	GO: 0043226	细胞器 Organelle	37 283	39	2 211	1 490	2 342
	GO: 0016020	细胞膜 Membrane	18 447	21	1 167	849	1 250
	GO: 0044422	细胞器部位 Organelle part	13 187	11	821	553	863
	GO: 0005576	胞外区域 Extracellular region	6 508	7	443	312	442
	GO: 0030054	细胞连接点 Cell junction	6 214	6	431	293	460
	GO: 0044425	细胞膜部位 Membrane part	6 165	6	385	301	401
	GO: 0032991	高分子复合物 Macromolecular complex	5 765	3	344	210	359
	GO: 0031974	膜封闭腔 Membrane-enclosed lumen	1 220	1	59	46	69
分子功能 Molecular function	GO: 0005488	联结 Binding	25 548	30	1 541	1 060	1 632
	GO: 0003824	催化活性 Catalytic activity	21 720	16	1 320	926	1 326
	GO: 0001071	核酸联结转录因子活性 Nucleic acid binding transcription factor activity	3 721	9	234	160	243
	GO: 0005215	转运活动 Transporter activity	3 151	8	200	167	258
	GO: 0004871	信号转导活动 Signal transducer activity	1 353	2	78	45	91
	GO: 0005198	结构分子活动 Structural molecule activity	1 201	1	74	37	83
	GO: 0009055	电子载体活动 Electron carrier activity	920	0	58	33	53
	GO: 0060089	分子传感器活动 Molecular transducer activity	893	1	51	38	62
	GO: 0098772	分子功能调节剂 Molecular function regulator	855	0	40	43	40
	GO: 0016209	抗氧化活性 Antioxidant activity	336	1	20	17	29

中, 前 2 个富集词条为联结和催化活性, 第 3 位词条 C vs A、C vs B 为核酸联结转录因子活性, F vs D、F vs E 为转运活动。这些富集途径对甘蓝叶球杂种优势的建立具有十分重要的作用。

在 4 对比对中, “生长”这一词条下的差异表达基因共有 1 025 个, 其中包括一些与叶球生长发育相关基因, 如 *MADS-box* (*MADS-MIKC*, *MADS-M-TYPE*) 是一类调控植物生长发育与应激反应的超家族, 以调控花器官发育著名, 但同样参与调控芽和子叶的发育 (Ng & Yanofsky, 2001), 此外, 在 4 对比对中都大量检测到与生长发育以及非生物胁迫相关的 NAC 转录因子超家族基因, 其在分生组织发育中常发挥重要作用 (Olsen et al., 2005), 调控植物生长发育、代谢、生物及非生物胁迫的 MYB 超家族基因同样在差异表达基因中显示较高的转录水平 (Dubos et al., 2010), 叶球发育同样离不开生长素的诱导作用, 因此编码生长素响应蛋白的 AUX/IAA 生长素应答基因 (Liscum & Reed, 2002) 在 4 对比对中发现大量转录本, 还有参与植物信号转导及分生组织发育的 GRAS 家族基因 (Bolle, 2004) 以及编码 DNA 结合蛋白, 参与茎尖生长发育的 bZIP (*CPRF3-LIKE*) 基因 (Abe, 2005) 除 C vs A 比对外, 在其余 3 对比对中都检测到大量转录本。以上基因对甘蓝叶球杂种优势的建成必不可少, 是参与甘蓝叶球杂种优势形成的重要基因。

由表 5 可知, COG 分类的 4 对比对富集到的第一位词条相同, 均为 R (一般功能预测), 其中包括多种与细胞壁建成、器官建成以及发育相关基因 (LOC106321495、LOC106325451、LOC106294537、

表 5 甘蓝杂交组合 F₁ 代与各自父母本之间差异表达基因在 COG 各分类中的数量
Table 5 The total number of DEGs in COG classification between F₁ hybrid and their paretns in cabbage

COG 分类 COG classification	分类名称 Classification name	A × B → C		D × E → F	
		C vs A	C vs B	F vs D	F vs E
R	一般功能预测 General function prediction only	5	312	241	345
K	转录 Transcription	1	169	115	187
T	信号转导机制 Signal transduction mechanisms	1	135	117	164
L	复制, 重组与修复 Replication, recombination and repair	1	160	107	161
G	碳水化合物转运与代谢 Carbohydrate transport and metabolism	4	107	96	120
O	翻译后修饰, 蛋白质折叠, 分子伴侣 Posttranslational modification, protein turnover, chaperones	2	96	51	108
E	氨基酸转运与代谢 Amino acid transport and metabolism	3	96	84	95
J	翻译, 核糖体结构与生物合成 Translation, ribosomal structure and biogenesis	0	81	39	93
P	无机离子转运与代谢 Inorganic ion transport and metabolism	1	52	49	84
S	功能未知 Function unknown	0	70	56	84
Q	次生代谢物生物合成, 转运与代谢 Secondary metabolites biosynthesis, transport and catabolism	1	62	47	65
C	能量产生与转换 Energy production and conversion	0	61	38	64
H	辅酶转运与代谢 Coenzyme transport and metabolism	1	28	14	40
I	脂质转运与代谢 Lipid transport and metabolism	0	39	41	40
D	细胞周期控制, 细胞分裂, 染色体分割 Cell cycle control, cell division, chromosome partitioning	1	30	27	39
M	细胞壁/膜/包膜生物合成 Cell wall/membrane/envelope biogenesis	0	48	28	36
U	细胞内运输, 分泌和囊泡运输 Intracellular trafficking, secretion, and vesicular transport	0	20	13	25
Z	细胞骨架 Cytoskeleton	1	17	18	24
V	防御机制 Defense mechanisms	0	18	17	23
F	核苷酸转运与代谢 Nucleotide transport and metabolism	0	16	13	20
A	RNA 加工和修饰 RNA processing and modification	0	17	16	17
B	染色质结构与动态 Chromatin structure and dynamics	0	7	4	8
N	细胞运动 Cell motility	0	2	0	4
W	细胞外结构 Extracellular structures	0	0	0	0
Y	细胞核结构 Nuclear structure	0	0	0	0

LOC106298291、LOC106318601、LOC106300705、LOC106306918、LOC106319440、LOC106302139、LOC106313822、LOC106334001、LOC106321208、LOC106334578、LOC106342758) 等, 第 2 位词条除 C vs A 比对外, 其余比对都为 K (转录), 而 C vs A 比对的第 2 位词条为 G (碳水化合物代谢), 第 3 位词条 F vs D 和 F vs E 比对一致为 T (信号转导机制), 而 C vs A 比对为 E (氨基酸转运和代谢), C vs B 比对为 L (复制, 重组与修复)。以上结果表明 F_1 与亲本之间的差异表达基因在不同功能上相互作用, 协调一致, 共同参与甘蓝叶球杂种优势建成的复杂调控网络。

2.4 差异表达基因 KEGG 通路富集分析

表 6 中显示的是差异表达基因 KEGG 分类中显著性 P -value 较小的前 24 个通路。富集因子, 表示差异基因中注释到某通路的基因比例与所有基因中注释到该通路的基因比例的比值, 富集因子越大, 表示差异表达基因在该通路中的富集水平越显著; P -value 越小, 表示差异表达基因在该通路中的富集显著性越可靠。由于 C vs A 比对差异基因数仅有 48 个, 因而注释到的通路较少, 其中牛磺酸和亚牛磺酸代谢途径相对富集水平较高, 而 C vs B 和 F vs D 比对中差异表达基因都显著富集到了氨基酸生物合成和 α -亚麻酸代谢途径, F vs E 比对中色氨酸代谢、 α -亚麻酸代谢以及植物一病原体相互作用等途径被显著富集。这些显著富集的生物合成或代谢途径都直接或间接地参与到甘蓝叶球生长发育过程当中, 是甘蓝叶球杂种优势形成的重要合成代谢途径。

表 6 甘蓝杂种组合 F_1 代与各自父母本之间差异表达基因在 KEGG 分类中的 P 值和富集因子
Table 6 The P -value and rich factor of DEGs in KEGG classification between F_1 hybrid and their parents in cabbage

KO ID	KEGG 注释 KEGG_annotation	A \times B \rightarrow C				D \times E \rightarrow F			
		C vs A		C vs B		F vs D		F vs E	
		P	富集因子 Rich factor	P	富集因子 Rich factor	P	富集因子 Rich factor	P	富集因子 Rich factor
ko01230	氨基酸生物合成 Biosynthesis of amino acids	—	—	0.0001	1.6815	0.0024	1.6331	0.0994	1.2135
ko00592	α -亚麻酸代谢 Alpha-Linolenic acid metabolism	—	—	0.0026	2.3116	0.0004	3.1537	0.0109	2.0310
ko00514	其他类型的 O-聚糖生物合成 Other types of O-glycan biosynthesis	—	—	0.2886	3.0381	—	—	0.3028	2.8747
ko00660	C5-支链二元酸代谢 C5-branched dibasic acid metabolism	—	—	0.0294	2.8131	0.3080	1.7910	0.5695	1.0647
ko00591	亚油酸代谢 Linoleic acid metabolism	—	—	0.0499	2.4501	0.0082	3.8998	0.0183	2.7820
ko00450	有机含硒化合物代谢 Selenocompound metabolism	—	—	0.1017	1.9987	0.4701	1.2726	0.1212	1.8913
ko00290	缬氨酸, 亮氨酸和异亮氨酸的生物合成 Valine, leucine and isoleucine biosynthesis	—	—	0.0455	1.9922	0.0395	2.3782	0.6216	0.9425
ko00945	二苯乙烯类、二芳基庚类、姜辣素生物合成 Stilbenoid, diarylheptanoid and gingerol biosynthesis	—	—	0.0213	1.9892	0.0585	2.0149	0.2284	1.3689
ko00750	维生素 B6 代谢 Vitamin B6 metabolism	—	—	0.1902	1.9814	0.6219	1.0512	0.2126	1.8748
ko00130	泛醌和其他萜类化合物-醌生物合成 Ubiquinone and other terpenoid-quinone biosynthesis	—	—	0.0452	1.8302	0.1285	1.7478	0.3559	1.2122
ko00030	磷酸戊糖途径 Pentose phosphate pathway	—	—	0.0387	1.7172	0.0485	1.8922	0.2781	1.2499
ko00400	苯丙氨酸, 酪氨酸和色氨酸生物合成 Phenylalanine, tyrosine and tryptophan biosynthesis	—	—	0.0574	1.6571	0.0852	1.7584	0.7863	0.7840
ko00330	精氨酸和脯氨酸代谢 Arginine and proline metabolism	0.1232	7.6778	0.0183	1.6511	0.0093	1.9711	0.2103	1.2499
ko00903	柠檬烯和蒎烯降解 Limonene and pinene degradation	—	—	0.0945	1.6472	0.0555	2.0391	0.1217	1.5586

续表 6

KO ID	KEGG 注释 KEGG_annotation	A × B → C				D × E → F			
		C vs A		C vs B		F vs D		F vs E	
		P	富集因子 Rich factor	P	富集因子 Rich factor	P	富集因子 Rich factor	P	富集因子 Rich factor
ko00250	丙氨酸, 天冬氨酸和谷氨酸代谢 Alanine, aspartate and glutamate metabolism	0.0001	—	0.0515	1.5823	0.0161	2.0149	0.0754	1.4972
ko00630	乙醛酸和二羧酸代谢 Glyoxylate and dicarboxylate metabolism	0.0001	—	0.0340	1.6422	0.0430	1.7970	0.0142	1.7481
ko00710	光合生物中的碳固定 Carbon fixation in photosynthetic organisms	0.1251	7.5546	0.3482	1.1373	0.3692	1.1637	0.0613	1.4604
ko00380	色氨酸代谢 Tryptophan metabolism	—	—	0.4796	1.0850	0.6813	0.8635	0.0018	2.3956
ko04626	植物—病原体相互作用 Plant - pathogen interaction	0.2560	3.4541	0.9399	0.7428	0.4274	1.0641	0.0111	1.4409
ko00430	牛磺酸和亚牛磺酸代谢 Taurine and hypotaurine metabolism	0.0349	28.2543	0.6484	0.9114	0.6187	0.9671	—	—
ko00620	丙酮酸代谢 Pyruvate metabolism	0.1188	7.9814	0.2762	1.2015	0.8621	0.6830	0.1696	1.2993
ko03008	真核生物中的核糖体生物合成 Ribosome biogenesis in eukaryotes	0.1427	6.5708	0.3419	1.1305	0.5339	1.0121	0.3269	1.1365
ko00360	苯丙氨酸代谢 Phenylalanine metabolism	0.1464	6.3924	0.5995	0.9623	0.7002	0.8752	0.9746	0.5853
ko00010	糖酵解/糖异生 Glycolysis/ Gluconeogenesis	0.1495	6.2509	0.2326	1.2099	0.3331	1.1768	0.1584	1.2720
ko00940	苯丙烷类生物合成 Phenylpropanoid biosynthesis	0.2298	3.9025	0.6027	0.9651	0.1700	1.2690	0.9527	0.7147

注: — 表示无差异基因富集到该通路。
Note: — represents no DEGs enriched into the pathway.

2.5 差异表达基因可变剪接分析

基因转录生成的前体 mRNA (pre-mRNA) 有多种剪接方式, 选择不同的外显子, 产生不同的成熟 mRNA, 从而翻译为不同的蛋白质, 构成生物性状的多样性, 这种转录后的 mRNA 加工过程称为可变剪接或选择性剪接。为了进一步了解影响杂种优势基因表达的潜在机制, 对甘蓝两个杂交组合 F₁ 及其双亲的可变剪接事件进行了统计 (表 7)。

表 7 甘蓝杂交组合 F₁ 及其双亲的可变剪接事件数量统计
Table 7 The statistics of the alternative splicing events in F₁ hybrid and their parents in cabbage

可变剪接事件 Alternative splicing event	A × B → C			D × E → F		
	A	B	C	D	E	F
AE: 可变 5'或 3'端剪切 Alternative exon ends (5', 3', or both)	5 500	5 742	5 659	5 893	5 694	6 004
IR: 单内含子滞留 Intron retention (IR_ON, IR_OFF pair)	10 426	9 022	9 870	10 280	8 490	10 028
MIR: 多内含子滞留 Multi-IR (MIR_ON, MIR_OFF pair)	502	414	470	498	392	506
MSKIP: 多外显子跳跃 Multi-exon SKIP (MSKIP_ON, MSKIP_OFF pair)	182	178	176	208	214	194
SKIP: 单外显子跳跃 Skipped exon (SKIP_ON, SKIP_OFF pair)	1 580	1 678	1 598	1 664	1 620	1 824
TSS: 第 1 个外显子可变剪切 Alternative 5' first exon (transcription start site)	49 505	49 278	49 510	49 590	49 279	49 689
TTS: 最后 1 个外显子可变剪切 Alternative 3' last exon (transcription terminal site)	49 049	48 922	48 997	49 102	48 952	49 130
XAE: 可变 5'或 3'端剪切 (模糊边界) Approximate AE	1 058	1 114	1 170	1 175	1 045	1 214
XIR: 单内含子滞留 (模糊边界) Approximate IR (XIR_ON, XIR_OFF pair)	1 750	1 586	1 824	1 820	1 530	1 882
XMIR: 多内含子滞留 (模糊边界) Approximate MIR (XMIR_ON, XMIR_OFF pair)	178	178	198	202	150	228
XMSKIP: 多外显子跳跃 (模糊边界) Approximate MSKIP (XMSKIP_ON, XMSKIP_OFF pair)	200	172	226	184	220	186
XSKIP: 单外显子跳跃 (模糊边界) Approximate SKIP (XSKIP_ON, XSKIP_OFF pair)	850	900	898	908	832	896
总计 Total	120 780	119 184	120 596	121 524	118 418	121 781

可变剪接分为 12 类: AE、IR、MIR、MSKIP、SKIP、TSS、TTS、XAE、XIR、XMIR、XMSKIP、XSKIP。表 7 中甘蓝两个杂交组合 F₁ 及其双亲的可变剪接事件分布具有一致的趋势, TSS (第 1 个外显子可变剪切) 和 TTS (最后 1 个外显子可变剪切) 在杂交组合 A × B 中数量最多, 占总可变剪接事件的比例达 81.6% 以上; 在杂交组合 D × E 中占比达 81.1% 以上。此外, IR 在两个甘蓝杂交组合中的数量仅次于 TSS 和 TTS, 占总可变剪接事件的 7.2% ~ 8.6%, 两个组合 F₁ 的 IR 数量在其双亲数量范围之内, 其他可变剪接事件在各 F₁ 及其双亲中分布相对较少, 因此, TSS、TTS 和 IR 可能在调控甘蓝叶球杂种优势基因表达方面具有重要作用。

2.6 实时荧光定量 PCR 验证

为了确保 RNA-seq 数据的准确性, 随机选取 7 个差异表达基因进行实时荧光定量 PCR 验证, 验证结果如表 8。log₂ratio 为 qPCR 数据经 log₂ 转换后的数值, 实时荧光定量 PCR 和 RNA-seq 的正值均代表上调表达, 负值均代表下调表达。由表 8 可以看到, 在 4 种比对中, 实时荧光定量 PCR 的数据结果与 RNA-seq 测序结果虽然表达倍数略有差异, 但基本呈现出一致的表达趋势, 因此判断 RNA-seq 数据结果具有一定的可靠性。

表 8 实时荧光定量 PCR 与 RNA-seq 比对结果
Table 8 The result of real-time PCR compared with RNA-seq

基因编号 Gene ID	C vs A		C vs B		F vs D		F vs E	
	qPCR log ₂ ratio	RNA-seq	qPCR log ₂ ratio	RNA-seq	qPCR log ₂ ratio	RNA-seq	qPCR log ₂ ratio	RNA-seq
Bo4g169200	0.04	0.05	- 0.08	- 4.50	- 0.19	- 1.25	- 0.42	- 2.27
Bo8g056920	0.14	0.26	3.45	6.08	- 0.40	- 1.18	6.49	无穷大 Infinity
Bo01049s040	- 0.57	- 0.38	3.08	无穷大 Infinity	- 0.14	- 0.61	5.85	无穷大 Infinity
Bo7g011450	- 2.21	- 0.40	2.57	2.54	- 2.01	- 0.43	4.63	6.08
Bo7g011440	- 0.57	- 0.54	0.96	8.35	- 0.04	- 0.16	5.55	11.26
Bo3g127630	- 0.08	- 0.25	3.52	3.56	- 0.91	- 1.98	4.86	5.06
Brassica oleracea L_ new Gene_1356	0.39	0.05	0.29	6.66	- 2.27	- 1.04	1.28	5.42

3 讨论

本研究中选用可育系与不育系作为平行试验是基于两个组合 F₁ 及其双亲具有良好的田间性状表现, 即 F₁ 杂种优势较为明显, 而不育系母本材料: QP04CMS 为 Ogura CMS 细胞质雄性不育类型, 控制这种不育性状的基因已被精确定位于线粒体 DNA 上的 orf138 位点, 在 *atp8* 基因 5' 端的上游, 并与其紧密连锁 (Bonhomme et al., 1991; Grelon et al., 1994)。这种经线粒体基因组重排产生的嵌合基因, 编码的是线粒体非必需的多肽 (ORF138 蛋白), 但相应蛋白的积累会导致 Ogura CMS 的育性变化 (Bonhomme et al., 1992), 即雄蕊败育, 无花粉, 因此, QP04CMS 与保持系配置的杂交组合杂种优势表达并不受育性基因的影响。园艺学性状统计分析结果表明, 两个组合 F₁ 在 9 个园艺学性状中都观察到较高水平的杂种优势, 单球质量杂种优势尤其明显, 表明产量杂种优势为甘蓝杂种优势的重要方面, 此外单球质量和球主叶叶柄质量在双亲及其 F₁ 代之间具有显著差异, 这些结果为进一步研究甘蓝叶球杂种优势的分子机制奠定了一定的生理基础。

近年来, 许多研究开始利用转录组技术分析杂种 F₁ 与亲本之间的差异表达基因, 进而分析植物

某一性状或阶段的杂种优势 (Hu et al., 2016; Zhang et al., 2017; Yang et al., 2018), 而筛选的差异表达基因数量往往表现出母本或父本偏向性, 例如, 在利用 RNA-seq 技术研究玉米小穗和小花分化期杂种优势时, F_1 经筛选获得的与母本之间的差异表达基因明显多于与父本之间的差异表达基因, 表明父本对玉米杂种优势的贡献更大 (Hu et al., 2016)。类似地, 在杂交稻 LYP9 与其亲本的转录组分析过程中, 根据其与其父母本之间的差异表达基因判定 LYP9 的转录组谱在早期发育阶段与 PA64s (母本) 相似, 但在后期更接近 93-11 (父本) (Wei et al., 2009)。本研究中, 在甘蓝成熟期叶球组织中, 两个组合 F_1 与其母本差异表达基因相对与其父本的较少, 表明母本在 F_1 叶球杂种优势形成中的贡献较大。本研究中没有采用正反交试验验证此结果是否与胞质遗传或基因印迹有关, 但甘蓝胞质遗传为母系遗传, 其特点为 F_1 通常只表现母本性状, 而两个组合 F_1 从发芽期到成熟期并未观察到明显的母系遗传偏向。基因印迹或亲源效应 (Parent-of-origin effects) 为一种表观调控机制, 可以使同一基因的父亲或母亲等位基因只有 1 个表达而另 1 个不表达或表达量很少, 目前在植物中确定的印迹基因数量较为有限, 其中模式植物拟南芥中已鉴定出 10 个印迹基因, 而大田作物如玉米中仅鉴定出 6 个 (Berger & Chaudhury, 2009), 大量基因在胚乳中的印迹表达表明基因印迹的组织倾向性以及其对植物早期发育过程的影响 (Stupar et al., 2007), 虽然印迹基因通常只占差异基因的极小部分, 但其作用不可忽视, 因此, 本研究结果与基因印迹的关系有待未来进一步考察。

差异表达基因经 GO 分类、COG 分类以及 KEGG 通路富集分析发现甘蓝叶球杂种优势的形成与生长发育、碳水化合物转运和代谢、信号转导以及氨基酸的合成、转运和代谢等过程密切相关。值得一提的是, 在绝大多数杂种优势分子机制的研究中, 代谢过程被一致认同为与生长或生物量杂种优势相关 (Lisec et al., 2008; Song et al., 2010; Wang et al., 2015), 然而代谢途径在不同植物的不同组织部位和发育阶段中富集水平不尽相同, 其中相当一部分研究认为碳水化合物代谢与杂种优势的建成相关 (Zhang et al., 2008, 2012; Song et al., 2010), 如 Zhang 等 (2012) 研究玉米产量杂种优势相关基因时发现, 能量和碳水化合物代谢是影响玉米产量杂种优势的两个重要代谢途径; 碳水化合物代谢途径显著富集在杂种优势较强而非杂种优势较弱的橡胶树幼苗中, 因而被认为是导致橡胶树生长杂种优势的重要代谢途径 (Yang et al., 2018)。除此之外, Wei 等 (2009) 在水稻 7 个不同组织部位中分别检测 F_1 与亲本、亲本与亲本之间的差异表达基因, 发现相比亲本与亲本之间的差异表达基因, F_1 与亲本之间的差异表达基因显著富集到能量代谢与转运途径, 其他代谢途径如次生代谢途径在杂交稻差异表达基因中被显著富集, 因而被认为与水稻杂种优势建成相关 (Song et al., 2007)。在本研究中, 碳水化合物代谢在 COG 分类中被显著富集, 因此该代谢途径很有可能参与甘蓝叶球杂种优势的建成过程, 是甘蓝叶球杂种优势相关的重要代谢途径。

杂种优势的形成是一个涉及转录组学、蛋白质组学、代谢组学、表观组学等多领域的复杂调控网络, 因此, 作为广泛存在于真核细胞中的一种转录后基因调控模式, 可变剪接 (AS) 涉及到调控植物生长发育 (Tang et al., 2016)、开花 (Chamala et al., 2015)、结果及信号转导 (Tang et al., 2016; Zhang et al., 2016) 等多种重要生命活动过程, 其与杂种优势的相关性不言而喻, 此外, 在植物生物及非生物胁迫条件下, 可变剪接更易大量发生 (Mastrangelo et al., 2012; Laloumet et al., 2017; Liu et al., 2018), 然而, 可变剪接模式的数量与其发挥作用的重要性并无必然关联, 许多研究表明, 植物中一些重要的可变剪接模式只需少量即可对某些基因的表达产生相当大的影响 (James et al., 2012; Yu et al., 2018)。在 Scascitelli 等 (2010) 的研究中, 杂种中一小部分基因存在新的可变剪接模式, 但这些新的可变剪接模式可能是产生新表型, 包括杂种优势特征的因素之一。在本研究中, 双亲及其 F_1 的大部分可变剪接模式为 TSS 和 TTS, 有少量 IR 及其他可变剪接模式存在, 但两个组

合 F₁ 与其亲本的可变剪接模式相比并无较大差异, 因此, 推测与杂种优势相关的可变剪接难以在数量上进行捕捉判断, 需要进一步对可变剪接的功能进行细化, 这也是未来对本文中杂种优势分子机制进一步研究的方向之一。

References

- Abe M. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, 309 (5737): 1052 – 1056.
- Adnan M, Morton G, Hadi S. 2011. Analysis of *rpoS* and *bolA* gene expression under various stress-induced environments in planktonic and biofilm phase using 2^{-ΔΔCT} method. *Mol Cell Biochem*, 357 (1 – 2): 275 – 282.
- Bell G D M, Kane N C, Rieseberg L H, Adams K L. 2013. RNA-Seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biology and Evolution*, 5 (7): 1309 – 1323.
- Berger F, Chaudhury A. 2009. Parental memories shape seeds. *Trends Plant Sci*, 14 (10): 550 – 556.
- Bolle C. 2004. The role of GRAS proteins in plant signal transduction and development. *Planta*, 218 (5): 683 – 692.
- Bonhomme S, Budar F, Féralut M, Pelletier G. 1991. A 2.5 kb *NcoI* fragment of Ogura radish mitochondrial DNA is correlated with cytoplasmic male-sterility in *Brassica* cybrids. *Curr Genet*, 19: 121 – 127.
- Bruce A B. 1910. The mendelian theory of heredity and augmentation of vigor. *Science*, 32: 627 – 628.
- Chamala S, Feng G Q, Chavarro C, Barbazuk W B. 2015. Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front Bioeng Biotechnol*, 3: 33.
- Chen L, Bian J M, Shi S L, Yu J F, Khanzada H, Mustafa Wassan G, Zhu C L, Luo X, Tong S, Yang X R, Peng X S, Yong S, Yu Q Y, He X P, Fu J R, Chen X R, Hu L F, Ouyang L J, He H H. 2018. Genetic analysis for the grain number heterosis of a super-hybrid rice WFYT025 combination using RNA-Seq. *Rice*, 11 (1): 37.
- Cheverud J M, Routman E J. 1995. Epistasis and its contribution to genetic variance components. *Genetics*, 139 (3): 1455 – 1461.
- Costa V, Angelini C, DeFeis I, Ciccociola A. 2010. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*, 2010 (5757): 853916.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci*, 15 (10): 573 – 581.
- East E M. 1936. Heterosis. *Genetics*, 21 (4): 375.
- Ge X M, Chen W H, Song S H, Wang W W, Hu S N, Yu J. 2008. Transcriptomic profiling of mature embryo from an elite super-hybrid rice LYP9 and its parental lines. *BMC Plant Biol*, 8 (1): 114.
- Grelon M, Budar F, Bonhomme S, Pelletier G. 1994. Ogura cytoplasmic male-sterility (CMS) -associated *orf138* is translated into a mitochondrial membrane polypeptide in male-sterile *Brassica* hybrids. *Mol Gen Genet*, 243: 540 – 547.
- Groszmann M, Greaves I K, Fujimoto R, James Peacock W, Dennis E S. 2013. The role of epigenetics in hybrid vigour. *Trends in Genetics*, 29 (12): 684 – 690.
- Hu X, Wang H, Diao X, Liu Z, Li K, Wu Y, Liang Q J, Wang H, Huang C L. 2016. Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. *BMC Genomics*, 17: 959.
- James A B, Syed N H, Bordage S, Marshall J, Gillian A, Nimmo G A, Jenkins G I, Herzyk P, Brown J W S, Nimmo H G. 2012. Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell*, 24: 961 – 981.
- Laloum T, Martin G, Duque P. 2017. Alternative splicing control of abiotic stress responses. *Trends Plant Sci*, 23 (2): 140.
- Liscum E, Reed J W. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol Biol*, 49 (3 – 4): 387 – 400.
- Lisec J, Meyer R C, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjék O, Selbig J, Altmann T, Willmitzer L. 2008. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J*, 53: 960 – 972.
- Liu Z S, Qin J X, Tian X J, Xu S B, Wang Y, Li H X, Wang X M, Peng H R, Yao Y Y, Hu Z R, Ni Z F, Xin M M, Sun Q X. 2018. Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum* L.). *Plant Biotechnol J*, 16 (3): 714 – 726.
- Mao X Z, Cai T, Olyarchuk J G, Wei L P. 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21 (19): 3787 – 3793.

- Mastrangelo A M, Marone D, Laidò G, Leonardis A M E, Vita P D. 2012. Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci*, 185 (4): 40 – 49.
- Minvielle F. 1987. Dominance is not necessary for heterosis: a two-locus model. *Genet Res*, 49: 245 – 247.
- Ng M, Yanofsky M F. 2001. Function and evolution of the plant MADS-box gene family. *Nature Rev Genet*, 2: 186 – 195.
- Olsen A N, Ernst H A, Leggio L L, Skriver K. 2005. NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci*, 10 (2): 79 – 87.
- Romagnoli S, Maddaloni M, Livini C, Motto M. 1990. Relationship between gene expression and hybrid vigor in primary root tips of young maize (*Zea mays* L.) plantlets. *Theor Appl Genet*, 80 (6): 769 – 775.
- Scascitelli M, Cognet M, Adams K L. 2010. An interspecific plant hybrid shows novel changes in parental splice forms of genes for splicing factors. *Genetics*, 184 (4): 975 – 983.
- Song G S, Zhai H L, Peng Y G, Zhang L, Wei G, Chen X Y, Xiao Y G, Wang L, Chen Y J, Wu B, Chen B, Zhang Y, Chen H, Feng X J, Gong W K, Liu Y, Yin Z J, Wang F, Liu G Z, Xu H L, Wei X L, Zhao X L, Ouwerkerk P B F, Hankemeier T, Reijmers T, Heijden R V D, Lu C M, Wang M, Greef J V D, Zhu Z. 2010. Comparative transcriptional profiling and preliminary study on heterosis mechanism of super-hybrid rice. *Mol Plant*, 3 (6): 1012 – 1025.
- Song S H, Qu H Z, Chen C, Hu S N, Yu J. 2007. Differential gene expression in an elite hybrid rice cultivar (*Oryza sativa* L.) and its parental lines based on SAGE data. *BMC Plant Biol*, 7 (1): 49.
- Springer N M, Stupar R M. 2007. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell*, 19 (8): 2391 – 2402.
- Stupar R M, Hermanson P J, Springer N M. 2007. Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. *Plant Physiol*, 145 (2): 411 – 425.
- Tang W, Zheng Y, Dong J, Yu J, Yue J Y, Liu F F, Guo X H, Huang S X, Wisniewski M, Sun J Q, Niu X L, Ding J, Liu J, Fei Z J, Liu Y S. 2016. Comprehensive transcriptome profiling reveals long noncoding RNA expression and alternative splicing regulation during fruit development and ripening in kiwifruit (*Actinidia chinensis*). *Front Plant Sci*, 7: 335.
- Wang L, Greaves I K, Groszmann M, Wu L M, Dennis E S, Peacock W J. 2015. Hybrid mimics and hybrid vigor in *Arabidopsis*. *Proc Natl Acad Sci USA*, 112: 4959 – 4967.
- Wei G, Tao Y, Liu G Z, Chen C, Luo R Y, Xia H A, Gan Q, Zeng H P, Lu Z K, Han Y N, Li X B, Song G S, Zhai H L, Peng Y G, Li D Y, Xu H G, Wei X L, Cao M L, Deng H F, Xin Y Y, Fu X Q, Yuan L P, Yu J, Zhu Z, Zhu L H. 2009. A transcriptomic analysis of super hybrid rice LYP9 and its parents. *Proc Natl Acad Sci USA*, 106 (19): 7695 – 7701.
- Yang H, Wang X, Wei Y, Deng Z, Liu H, Chen J S, Dai L J, Xia Z H, He G M, Li D J. 2018. Transcriptomic analyses reveal molecular mechanisms underlying growth heterosis and weakness of rubber tree seedlings. *BMC Plant Biol*, 18 (1): 10.
- Young M D, Wakefield M J, Smyth G K, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, 11 (2): R14.
- Yu J, Miao J, Zhang Z, Xiong H, Zhu X, Sun X, Pan Y, Liang Y, Zhang Q, Abdul R R M, Li J, Zhang H, Li Z. 2018. Alternative splicing of OsLG3b controls grain length and yield in Japonica rice. *Plant Biotechnol J*, 16: 1667 – 1678.
- Zhai R R, Feng Y, Wang H M, Zhan X D, Shen X H, Wu W M, Zhang Y X, Chen D B, Dai G X, Yang Z L, Cao L Y, Cheng S H. 2013. Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics*, 14 (1): 19.
- Zhang C B, Lin C J, Fu F Y, Zhong X F, Peng B, Yan H, Zhang J Y, Zhang W L, Wang P N, Ding X Y, Zhang W, Zhao L M. 2017. Comparative transcriptome analysis of flower heterosis in two soybean F₁ hybrids by RNA-seq. *PLoS ONE*, 12 (7): e0181061.
- Zhang H Y, He H, Chen L B, Li L, Liang M Z, Wang X F, Liu X G, He G M, Chen R S, Ma L G, Deng X W. 2008. A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol Plant*, 1 (5): 720 – 731.
- Zhang Q S, Zhang X Q, Wang S B, Tan C, Zhou G F, Li C D. 2016. Involvement of alternative splicing in barley seed germination. *PLoS ONE*, 11 (3): e0152824.
- Zhang T F, Li B, Zhang D F, Jia G Q, Li Z Y, Wang S C. 2012. Genome-wide transcriptional analysis of yield and heterosis-associated genes in maize (*Zea mays* L.). *J Integr Agri*, 11 (8): 1245 – 1256.