

王莹, 李玉娟, 李敏, 等. 紫叶紫薇新品系叶色变化转录组分析[J]. 江苏农业学报, 2018, 34(5): 1128-1137.
doi: 10.3969/j.issn.1000-4440.2018.05.023

紫叶紫薇新品系叶色变化转录组分析

王莹, 李玉娟, 李敏, 马祥建, 谈峰, 郭聪, 张健

(江苏沿江地区农业科学研究所, 江苏 南通 226541)

摘要: 对紫叶紫薇新品系的叶片进行 RNA-Seq 测序和分析, 共获取 145 338 条 Contig, 得到组装的 Unigene 有 53 654 条, 平均长度 1 186.25 bp。32 667 条 Unigene 可以匹配到蛋白质数据库获得注释信息, KEGG 和 GO 注释与富集分析显示, 差异 DEG 主要调节激素信号转导、碳水化合物代谢、核糖体代谢等生物学功能。58 个表达差异基因分别对比到毛果杨和拟南芥 2 个模式植物的数据库, 筛出与叶绿素代谢相关候选基因 9 个, 其中 2 个候选基因共同参与了叶绿素类黄酮代谢的调控, 2 个候选基因可能参与类黄酮代谢过程, 它们分别属于 *DUF547*、*Polyketide_cyc2*、*Pkinase*、*Calreticulin*、*NAD_Gly3P_dh_N*、*Glyco_hydro_9* 等 13 个基因家族, 主要经氧化还原反应、物质能量代谢、植物激素信号转导等机制介入紫薇叶色调控。本研究结果为探究紫叶紫薇叶色变异的分子机制提供参考。

关键词: 紫叶紫薇; 叶色变化; 转录组

中图分类号: S687.9

文献标识码: A

文章编号: 1000-4440(2018)05-1128-10

Transcriptome analysis of a new strain of purple-leaf crape myrtle (*Lagerstroemia indica*) during leaves color changes

WANG Ying, LI Yu-juan, LI Min, MA Xiang-jian, TAN Feng, GUO cong, ZHANG Jian

(Institute of Agricultural Sciences of Jiangsu Changjiang River Bank District, Nantong 226541, China)

Abstract: This study was conducted on the transcription and bioinformatics analysis of a new strain of purple leaf crape myrtle (*Lagerstroemia indica*) during leaves color changes. A total of 145 338 contigs were obtained, and assembled de novo into 53 654 unigenes, with average length of 1 186.25 bp. KEGG and GO enrichment analysis results showed that differences in gene were significantly associated with some biological functions, mainly involved in the metabolism of the ribosome, carbohydrate metabolism, plant hormone signal transduction and other metabolic pathways. Total of nine unigenes involved in chlorophyll metabolism, two unigenes related to both chlorophyll metabolism and flavonoid metabolism and two unigenes involved in flavonoid metabolism were identified by comparing the 58 unigenes obtained with the GO and KEGG databases of *Populus trichocarpa* and *Arabidopsis thaliana*, which respectively belonged to 13 gene families including *DUF547*, *Polyketide_cyc2*, *Pkinase*, *Calreticulin*, *NAD_Gly3P_dh_N*, *Glyco_hydro_9* etc. They mainly participated in the regulation of leaf color through the mechanism of oxidation-reduction reaction, energy metabolism, material metabolism, and plant hormone signal transduction. The experimental results provide a reference for the study of molecular mechanism of leaf color variation in purple-leaf crape myrtle.

Key words: purple-leaf *Lagerstroemia indica*; leaves color changes; transcriptome

收稿日期: 2017-12-07

基金项目: 南通市应用基础研究项目 (MS12015075); 中央财政农业技术推广项目 [TG (16) 048]; 南通市前沿与关键技术创新项目 (MS22015039)

作者简介: 王莹 (1986-), 女, 江苏南通人, 硕士, 助理研究员, 主要从事彩叶苗木品种引选、栽培方面的研究。(E-mail) 466625450@qq.com

通讯作者: 张健, (E-mail) yjnky@163.com

紫薇 (*Lagerstroemia indica*) 以其观赏性佳、抗逆性好, 受到广泛关注。虽然紫薇属资源在中国分布非常广泛, 但该属植物的育种研究起步较晚, 主要集中在花色, 如李文芳等^[1] 和刘龙昌等^[2] 分别对紫薇花色

素的化学性质及结构进行了初步研究;花香,如徐婉等^[3]对紫薇杂交后代花香气成分做了定性与定量分析;抗逆性,如谢宪等^[4]对紫薇抗寒性的研究。紫薇属稀缺彩叶品种,近年来育种专家们通过引种与选育,逐步获得了一批集观花、观叶性于一体的彩叶紫薇^[5-7]。然而,对彩叶紫薇的研究目前仅集中于传统的引种与栽培技术研究,基础研究方面过于薄弱,尤其关于叶片呈色机理的研究。目前仅能查阅到宋倩等^[8]和王淑安等^[9]关于彩叶紫薇的叶色变异机制的研究报道,分别阐明了彩叶紫薇叶色变化与叶片中色素含量的关系,得到了大量与类胡萝卜素和叶绿素合成相关的候选基因。

作为非模式植物中的一员,紫薇全基因组测序尚未完成,生物信息学信息十分匮乏。随着测序技术日益成熟,高通量测序技术可以在不涉及基因组信息的情况下,检测到所有物种的转录本,为彩叶紫薇叶片呈色机制的研究提供可能。

RNA-Seq 测序技术可获得大量的遗传信息,能诠释植物生化水平的多样性及植物次生代谢分支调控机理,为研究彩叶植物叶色转化带来新的机遇。曹桦等对铁皮石斛叶色突变体进行转录组测序分析,认为 *doGLK1* 表达水平低和 *doFtsz* 不表达导致了突变体叶绿体合成和分裂受阻^[10]。任杰对红花槭品种艳红叶色转化期转录组解析结果表明,花青素替代代谢途径基因差异表达对叶色变化起重要调控作用^[11]。红掌突变体叶色呈色机制的研究结果表明, *AaGLK* 和 *AaARC5* 分别是调控叶绿体发育和分裂的关键基因, *AaDFR* 在花青素合成过程中起着关键作用^[12]。转录组测序技术已经被广泛应用在非模式彩叶植物的叶片呈色机理研究中。

本项目组 2011 年开始进行紫薇实生选种工作,2012 年开始进行杂交育种工作,通过 2 种方法已选育出一批观赏性状优良的彩叶紫薇新品系。其中在黑钻石系列 Best Red 后代实生苗中发现 1 株植株叶片在春季呈现更加明艳的紫红色,初夏开始返青,上位叶紫红色,中位叶呈紫色与绿色相交的花叶色,下位叶绿色,色叶期比亲本长 15~20 d,通过繁育观察,其特异性、一致性、稳定性均达到新品种的申报要求,拟定名为紫晶 1 号,已申报国家林业局植物新品种权。本研究利用 Illumina HiSeq 高通量测序平台对紫晶 1 号不同叶位叶片进行转录组测序分析,比较不同叶片中基因表达的差异,以期通过 RNA-

Seq 技术获得与叶色性状相关的基因序列,为揭示紫叶紫薇叶色变异的分子机制和培育彩叶紫薇新品种提供参考。

1 材料与方法

1.1 试验材料

以江苏沿江地区农业科学研究所紫薇种质圃内新选育的紫叶紫薇紫晶 1 号为试验材料。选取转色期叶色性状稳定、生长健壮、长势一致的植株,采集上位叶(紫叶)、中位叶(花叶)和下位叶(绿叶),用锡纸包裹,在液氮中迅速冷冻。本项目中试验材料总 RNA 的提取与质量检测以及无参考基因组序列的转录组高通量测序分析均由北京百迈客生物科技有限公司进行。

1.2 试验方法

1.2.1 总 RNA 的提取与质量检测 用总 RNA 提取试剂盒(QIAGEN)对不同叶位叶片进行 RNA 提取,样本的纯度、浓度和完整度分别选用 Nanodrop、Qubit 2.0、Agilent 2100 分析。后续进行合格 RNA 的 cDNA 文库构建和 RNA-Seq 测序。

1.2.2 cDNA 文库构建和转录组测序 检测后合格的样品,以 mRNA 为模板,构建 cDNA 双链,利用 AMPure XP beads 试剂盒依次进行纯化、末端修复和片段大小选择,通过 PCR 富集得到紫薇 cDNA 文库。cDNA 文库的质量用 Agilent 2100 检测,再通过 Illumina Hi-Seq 平台完成 RNA-Seq 测序分析。

1.2.3 De novo 组装及质量控制 获得的原始数据(Raw reads)经过滤得到 Clean reads,将其用 Trinity 拼接,得到 Contigs,利用 De Bruijn 图的方法和测序 Read 信息,在片段集合中读取转录本序列,最长 Contig 即为最终转录本的 Unigene。

1.2.4 Unigene 功能注释 使用 BLAST 软件将 DEGs 序列与 NR、Swiss-Prot、GO、COG、KOG、eggNOG4.5、KEGG 数据库分别比对,获得相关注释信息。

2 结果与分析

2.1 转录组测序与 de novo 组装

利用 Illumina Hi-Seq 对 3 个不同叶位叶片(每个叶位 3 个重复)共计 9 个样品进行转录组测序(表 1),共获得 68.56 Gb(过滤后序列),各样品过滤后的序列量均达到 6.87 Gb 以上;Q30 碱基百分比在 91%以上,G、C 含量在 49%以上,可以用于下

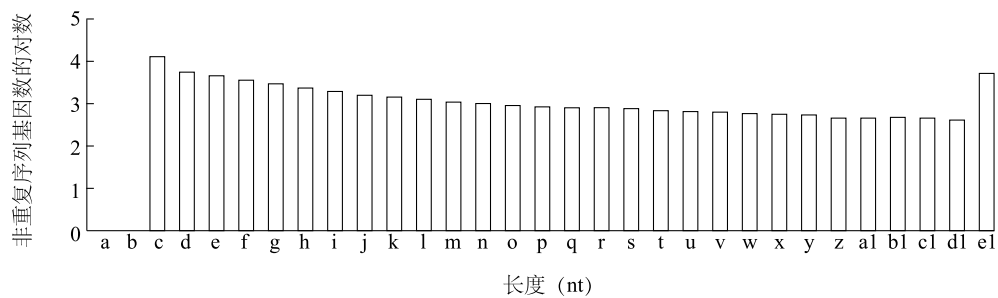
一步的 De novo 组装。利用 Trinity 软件将所有的短序列拼接,获得 145 338 条 Contig,进一步组装得到 53 654 条 Unigene,总长度 63 646 992 bp,平均长度为 1 186.25 bp,其中长度在 200~300 bp 的 Unigene

所占的比例最高,Unigene 的 N50 为 2 229 bp,组装完整性较高(图 1),获得的 Mapped Read 可用于后续分析。

表 1 样品测序数据评估统计

Table 1 The evaluation of sample sequencing data

样品名称	编号	原始序列数	过滤后的序列数	被标记的序列数	G、C 含量 (%)	Q30 (%)
紫叶重复 1	T01	7 449 642 750	24 996 437	18 361 190	50.50	91.87
紫叶重复 2	T02	8 163 098 312	27 336 100	20 721 968	50.86	92.39
紫叶重复 3	T03	7 805 696 038	26 123 290	19 926 535	50.33	92.08
花叶重复 1	T04	8 144 339 834	27 250 805	20 845 995	50.65	92.22
花叶重复 2	T05	7 277 674 194	24 358 779	18 123 943	49.97	91.43
花叶重复 3	T06	7 684 851 184	25 722 535	19 626 492	50.31	91.78
绿叶重复 1	T07	6 874 995 638	23 000 899	16 756 567	50.45	91.57
绿叶重复 2	T08	7 605 846 546	25 473 645	18 722 903	50.15	91.72
绿叶重复 3	T09	7 549 727 182	25 282 499	18 687 976	50.28	91.76



a:0~100;b:101~200;c:201~300;d:301~400;e:401~500;f:501~600;g:601~700;h:701~800;i:801~900;j:901~1 000;k:1 001~1 100;l:1 101~1 200;m:1 201~1 300;n:1 301~1 400;o:1 401~1 500;p:1 501~1 600;q:1 601~1 700;r:1 701~1 800;s:1 801~1 900;t:1 901~2 000;u:2 001~2 100;v:2 101~2 200;w:2 201~2 300;x:2 301~2 400;y:2 401~2 500;z:2 501~2 600;al:2 601~2 700;bl:2 701~2 800;cl:2 801~2 900;dl:2 901~3 000;el:>3 000。

图 1 Unigene 长度分布

Fig.1 Length distribution of unigenes

2.2 差异基因表达量分析

将 FDR 小于 0.01,差异倍数 $FC \geq 4$ 作为筛选标准,对紫晶 1 号 3 个不同叶位叶片中差异表达基因进行比对,共筛选得到 58 条差异表达基因(图 2、表 2);以绿叶为对照,上位叶检测出的差异表达基因数目最多,有 4 792 个转录因子的转录水平发生了改变,其中上调表达的转录因子基因高达 4 017 个,上调倍数在 4.00~1 573.76,最大上调基因 ID 为 c49503.graph_c0;在下调基因中,下位叶与中位叶对比,差异 DEG 最少,其次为上位叶与下位叶对比得到的差异 DEG 数,其中最大下调表达基因 ID 为 c41691.graph_c0,下调倍数达 1 418.35。

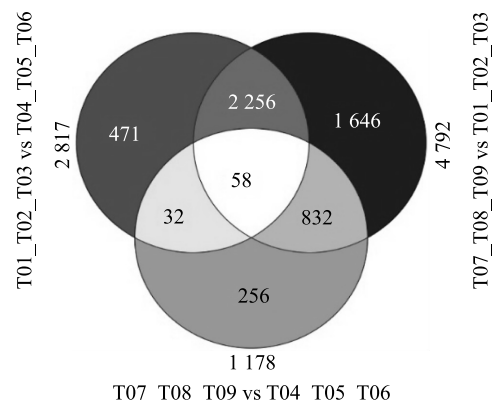


图 2 差异表达基因数目分析

Fig.2 The number of differentially expressed genes (DEGs)

表2 差异表达基因数目统计

Table 2 Statistics on the number of DEGs

差异表达基因集名称	差异表达基因数目	上调基因		下调基因	
		基因数	表达量变化范围	基因数	表达量变化范围
T07_T08_T09_vs_T01_T02_T03	4 792	4 107	4.00~1 573.76	685	4.01~1 418.35
T07_T08_T09_vs_T04_T05_T06	1 178	837	4.00~491.14	341	4.02~430.54
T01_T02_T03_vs_T04_T05_T06	2 817	138	4.01~65.34	2 679	4.00~404.50

2.3 差异表达基因功能注释

通过选择 BLAST 参数 ($E\text{-value} \leq 1e-5$) 和 HM-MER 参数 ($E\text{-value} \leq 1e-10$), 将全部 Unigene 序列与 COG、GO 等 8 个数据库比对 (表 3、表 4), 最后得到 32 667 个有注释信息的基因, 其中长度在 1 000 bp 以上的基因注释效率为 53.05%, 而未得到注释的 Unigene 为 20 987 条 (39.12%)。

2.4 差异表达基因 GO 分类和功能富集

紫晶 1 号差异表达基因在 GO 数据库进行比对, 有 18 718 条 Unigene 在 GO 数据库中得到相关注释, 其中每 2 个样品相比得到的差异基因在 GO 数据库中注释率最多达 56.88% (图 3), 注释的基因参与生物学过程、细胞组分和分子功能调控 (表 5), 功能组涉及到 53 个。生物学过程的注释显示, Unigene 表达差异显著顺序依次为代谢过程, 细胞过程和单组织过程; 在细胞成分分类信息中, 细胞部分和

细胞显著富集; 在分子功能分类中, 富集程度最高的是催化活性, 说明紫晶 1 号转色时期叶片中催化活性受影响较大。

表3 Unigene 注释统计表

Table 3 Statistics on the annotation of unigene

功能数据库名称	注释到该数据库的 Unigene 数	300 ≤ 长度 < 1 000	长度 ≥ 1 000
COG_Annotation	11 099	2 596	6 822
GO_Annotation	18 718	5 324	9 896
KEGG_Annotation	12 585	3 755	6 862
KOG_Annotation	18 989	5 215	10 940
Pfam_Annotation	22 779	5 793	14 413
Swissprot_Annotation	20 454	5 908	12 217
eggNOG_Annotation	30 818	8 812	16 891
nr_Annotation	31 296	9 118	16 871
All_Annotated	32 667	9 636	17 331

表4 注释的差异表达基因数量统计

Table 4 Statistics on the number of annotated DEGs

#DEG_Set	Annotated	COG	GO	KEGG	KOG	Pfam	Swiss-Prot	eggNOG	nr
T07_T08_T09_vs_T01_T02_T03	4 011	1 497	1 856	1 594	2 531	3 339	2 628	3 823	3 638
T07_T08_T09_vs_T04_T05_T06	1 039	291	591	290	433	820	773	992	1 031
T01_T02_T03_vs_T04_T05_T06	2 280	784	592	1 059	1 569	1 944	1 370	2 171	1 924

#DEG_Set: 差异表达基因集名称; Annotated: 得到注释的差异基因数; COG: 蛋白质直系同源数据库; GO: 基因本体论数据库; KEGG: 京都基因与基因组百科全书; KOG: 真核生物蛋白相邻类的聚簇数据库; Pfam: 蛋白质家族数据库; Swiss-Prot: 蛋白序列数据库; eggNOG: 蛋白直系同源组分析数据库; nr: 氨基酸序列数据库。

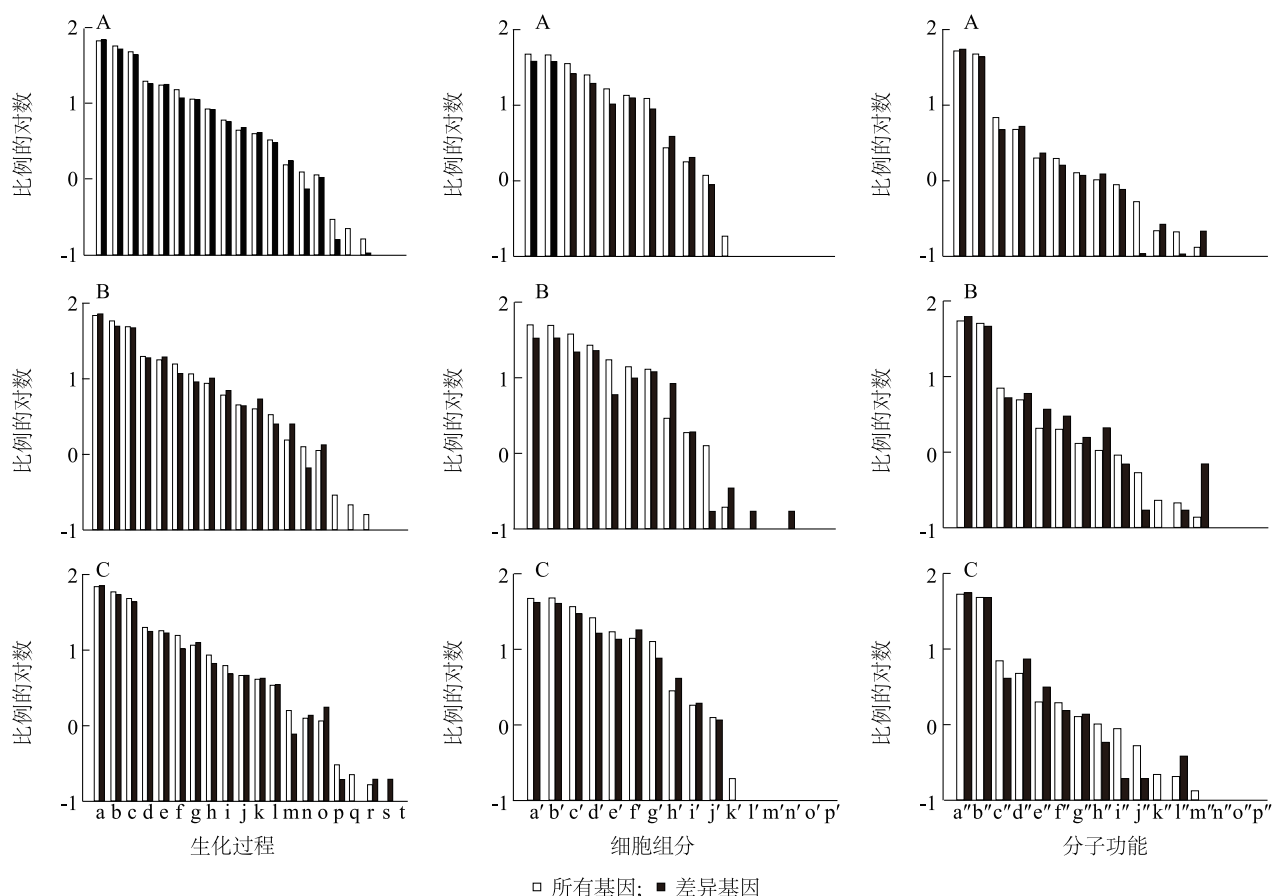
表5 GO 数据库 3 大功能主类所占比例

Table 5 The proportion of major function in GO database

差异表达基因集 (#DEG_Set)	GO 数据库 3 大功能主类所占比例 (%)		
	生物学过程	细胞组分	分子功能
T07_T08_T09_vs_T01_T02_T03	49.83	27.17	23.00
T07_T08_T09_vs_T04_T05_T06	48.70	30.21	21.09
T01_T02_T03_vs_T04_T05_T06	47.97	31.32	20.71

2.5 差异表达基因 KEGG 注释和富集分析

有 7 544 个 (23.10%) Unigene 在 KEGG 数据库得到注释, 涉及了 119 个 KEGG 代谢通路, 其中绿叶与紫叶对比获得注释的 919 个差异表达基因 (DEG) 可以被分配到 118 个 KEGG 通路 (表 6), 绿叶对比花叶获得注释的 187 条 DEG 可以被分配到 88 个 KEGG 通路, 紫叶对比花叶获得注释的 599 条 DEG 可以被分配到 113 个 KEGG 通路, 排名前 10 显著性富集差



A: 绿叶与紫叶对比; B: 绿叶与花叶对比; C: 紫叶与花叶对比。a: 代谢过程; b: 细胞过程; c: 单细胞生物过程; d: 应激响应; e: 生化调节; f: 定位; g: 细胞组织成分或生物发生; h: 发育过程; i: 多细胞生物过程; j: 信号传输; k: 生殖过程; l: 多生物过程; m: 生长; n: 免疫系统; o: 生殖; p: 生物粘附; q: 生物相; r: 节律过程; s: 运动; t: 细胞杀伤。a': 细胞部分; b': 细胞; c': 细胞器; d': 膜; e': 细胞器部分; f': 大分子复合物; g': 生物膜部分; h': 细胞外区域; i': 细胞连接处; j': 膜封闭内腔; k': 类核; l': 细胞外基质; m': 细胞外区域部分; n': 细胞外基质部分; o': 病毒; p': 病毒部分。a'': 催化活性; b'': 结合; c'': 转运体活性; d'': 结构分子活性; e'': 电子载体活性; f'': 核酸结合转录因子活性; g'': 分子转导活性; h'': 酶调节活性; i'': 抗氧化活性; j'': r 受体活性; k'': 鸟苷酸、核苷酸交换因子活性; l'': 蛋白结合转录因子活性; m'': 营养储藏活动; n'': 金属伴侣活性; o'': 蛋白质标签; p'': 翻译调节活性。

图 3 差异表达基因 GO 二级节点注释统计图

Fig.3 Secondary node statistics of DEGs in GO annotation

异基因主要涉及到细胞途径、遗传信号响应、环境信号响应和新陈代谢 4 个方面,其中最大的分类是新陈代谢,包括碳水化合物、氨基酸、多糖生物合成以及能量、次生物质等;环境信号响应类最小。同时,KEGG 的注释结果也显示,差异基因主要参与了次生物质代谢、植物激素信号转导、核糖体代谢和碳水化合物代谢等通路。紫晶 1 号在转色期叶片类黄酮生物合成途径的 KEGG 通路注释结果显示,花青素合成途径中有 9 个基因表达上调,3 个基因下调,肉桂酰辅酶 a 是花色苷合成关键酶,相应的基因 ID 为 c37587.graph_c0、c50617.graph_c0 和 c54586.graph_c0(K05278+K08695+K00660)。

c0(K05278+K08695+K00660)。

2.6 差异基因注释结果分析

叶色转化期间,log₂FC 达 8 倍以上差异表达的基因如表 7 所示,高水平差异表达的注释基因中,以绿叶为对照,紫叶对比花叶的短链脱氢酶还原酶 c27313.graph_c0 基因表达下调水平显著;以紫叶为对照,c52058.graph_c0(表皮原因子基因)也显示出相似的下调趋势。在差异表达 DEG 中,除 c47073.graph_c0、c27117.graph_c0 为假定的脂质转移蛋白基因,其他的基因功能主要涉及到基础代谢、逆境响应因子、性别决定等几个方面。

表 6 排行前 10 位的代谢通路

Table 6 Metabolic pathways ranked in the top 10

差异表达基因集 (#DEG_Set)	通路	代谢通路中的差异表达基因数	百分比 (%)	通路号
T07_T08_T09_vs_T01_T02_T03	核糖体	108	11.75	ko03010
	碳代谢	85	9.25	ko01200
	在内质网蛋白加工	50	5.44	ko04141
	氨基酸生物合成	48	5.22	ko01230
	淀粉、蔗糖代谢	44	4.79	ko00500
	植物激素信号转导	43	4.68	ko04075
	糖酵解/糖异生	43	4.68	ko00010
	RNA 运输	40	4.35	ko03013
	氧化磷酸化	35	3.81	ko00190
	嘌呤代谢	33	3.59	ko00230
T07_T08_T09_vs_T04_T05_T06	核糖体	34	18.18	ko03010
	植物激素信号转导	21	11.23	ko04075
	苯丙素的生物合成	14	7.49	ko00940
	戊糖、葡萄糖醛酸转换	9	4.81	ko00040
	碳代谢	9	4.81	ko01200
	淀粉、蔗糖代谢	9	4.81	ko00500
	角质、亚氨基酸和蜡生物合成	8	4.28	ko00073
	脂肪酸代谢	7	3.74	ko01212
	糖酵解/糖异生	7	3.74	ko00010
	丙酮酸代谢	7	3.74	ko00620
T01_T02_T03_vs_T04_T05_T06	碳代谢	59	9.85	ko01200
	核糖体	59	9.85	ko03010
	在内质网蛋白加工	39	6.51	ko04141
	RNA 运输	33	5.51	ko03013
	氧化磷酸化	31	5.18	ko00190
	糖酵解/糖异生	29	4.84	ko00010
	脂肪酸代谢	28	4.67	ko01212
	氨基酸生物合成	27	4.51	ko01230
	柠檬酸循环	24	4.01	ko00020
	吞噬体	24	4.01	ko04145

表 7 不同叶位叶片转色期 log₂FC 达 8 倍以上的差异表达的基因及注释结果

Table 7 The DEGs with log₂FC value more than eight times and annotation results at different leaf positions in colure-changed period

差异表达基因集 (#DEG_Set)	基因 ID	log ₂ FC	基因长度 (bp)	基因功能注释
T07_T08_T09_vs_T01_T02_T03	c49503.graph_c0	10.620 920	828.88	富含甘氨酸的细胞壁结构蛋白
	c52058.graph_c0	10.610 710	410.00	原表皮因子
	c50421.graph_c0	10.096 670	936.00	花粉 Ole e 1 过敏源和扩展蛋白家族
	c57274.graph_c2	9.863 145	1 965.17	木菠萝素相关凝集素
	c47073.graph_c0	9.629 891	783.00	非特异性脂质转运蛋白
	c58159.graph_c0	9.286 087	1 154.00	铜运输蛋白质 ATX1
	c47991.graph_c0	9.263 041	2 043.50	F-box/kelch-复制蛋白 (At1g67480)
	c52849.graph_c0	9.123 826	2 079.50	富亮氨酸重复类伸展蛋白
	c26990.graph_c0	9.060 468	1 294.00	类甲酸精蛋白
	c27117.graph_c0	8.944 177	836.00	非特异性脂质转运蛋白
	c53710.graph_c0	8.835 734	1 895.00	酪氨酸脱羧酶
	c27281.graph_c0	8.801 827	2 715.00	类枯草菌素蛋白酶
	c53531.graph_c0	8.760 676	1 511.00	GDSL 酯酶/脂肪酶 LTL1
	c51707.graph_c0	8.749 019	1 459.00	Ctr 蛋白
	c45240.graph_c0	8.673 921	2 185.00	漆酶-17
	c31000.graph_c0	8.615 653	451.00	富脯氨酸蛋白
	c53718.graph_c3	8.586 218	488.00	多元醇运输体
	c27013.graph_c0	8.551 246	1 517.00	GDSL 酯酶/脂肪酶 APG
	c29938.graph_c0	8.544 774	989.00	轴向调节物 YABBY
	c37983.graph_c0	8.396 735	956.00	富含甘氨酸的细胞壁结构蛋白
	c53964.graph_c0	8.381 486	1 050.00	GDSL 酯酶/脂肪酶 (At5g45950)
	c44309.graph_c0	8.362 634	1 953.00	雄性不育性蛋白
	c26948.graph_c0	8.074 160	1 800.00	ctr 蛋白
	c27313.graph_c0	-8.017 840	841.00	短链脱氢酶还原酶
	c41691.graph_c0	-10.470 500	1 818.16	抗菌肽
T07_T08_T09_vs_T04_T05_T06	c47073.graph_c0	8.941 690	783.00	非特异性脂质转运蛋白
	c31000.graph_c0	8.205 042	451.00	富脯氨酸蛋白
	c30595.graph_c0	-8.189 620	1 107.00	乙烯响应转录因子 ERF017
	c27313.graph_c0	-8.394 590	841.00	短链脱氢酶还原酶
	c55207.graph_c0	-8.752 090	2 215.00	伯胺氧化酶
T01_T02_T03_vs_T04_T05_T06	c49503.graph_c0	-8.228 590	828.88	富含甘氨酸的细胞壁结构蛋白
	c52058.graph_c0	-8.657 120	410.00	原表皮因子

2.7 叶色相关转录因子分析

在转色期间,3 种不同叶位筛选到差异基因 58 个。与模式植物毛果杨和拟南芥的 GO 和 KEGG 蛋白数据库作比对 (表 8),筛选与叶绿素代谢相关候选基因 9 个,其中 2 个候选基因共同参与了叶绿素

兼类黄酮代谢的调控,2 个候选基因可能参与类黄酮代谢过程,分别属于 *DUF547*、*Polyketide_cyc2*、*Pkinase*、*Calreticulin*、*NAD_Gly3P_dh_N*、*Glyco_hydro_9* 等 13 个基因家族,主要经氧化还原反应、物质能量代谢、植物激素信号转导等机制介入紫薇叶色调控。

在与叶色相关差异表达候选基因中,有 11 条 Uni-gene 与卟啉环代谢和叶绿素合成相关基因表现出同源性;而 c56217.graph_c0 基因在叶绿素降解通路中发挥了一定作用;在花青素苷代谢途径中,尿苷二

磷酸葡萄糖基转移酶、富含亮氨酸重复序列受体蛋白激酶、松脂醇还原酶 I、聚半乳糖醛酸酶 II 等均参与了对花青苷代谢途径的调控。

表 8 与叶色相关差异表达候选基因及 GO 注释

Table 8 The DEGs associated with leaf color and GO annotation

功能	基因 ID	基因长度 (bp)	基因家族	GO 注释
与叶绿素代谢	c42391.graph_c0	2 066.00	<i>DUF547</i>	Chloroplast
	c43468.graph_c0	950.00	<i>Polyketide_cyc2</i>	Chloroplast
	c45897.graph_c0	2 396.20	<i>Pkinase</i>	Chloroplast
	c47022.graph_c0	1 609.00	<i>Calreticulin</i>	Chloroplast
	c49079.graph_c0	2 226.84	<i>NAD_Gly3P_dh_N</i>	Chloroplast
	c50079.graph_c0	1 601.00	<i>Glyco_hydro_9</i>	Chloroplast
	c50421.graph_c0	936.00	<i>Pollen_Ole_e_I</i>	Chloroplast
	c51707.graph_c0	1 459.00	<i>HMA</i>	Chloroplast
	c56217.graph_c0	2 370.00	<i>Inhibitor_I9</i>	Chloroplast chlorophyll catabolic process
叶绿素兼类黄酮代谢	c47870.graph_c0	1 833.00	<i>UFGT</i>	Chloroplast; anthocyanin-containing compound biosynthetic process; anthocyanidin 3-O-glucosyltransferase activity; flavonol biosynthetic process; flavonol 3-O-arabinosyltransferase activity
	c51138.graph_c0	2 405.84	<i>PLN00113</i>	Chloroplast; regulation of anthocyanin biosynthetic process
类黄酮代谢	c29868.graph_c0	1 301.00	<i>SDR</i>	Positive regulation of flavonoid biosynthetic process
	c52763.graph_c0	2 107.52	<i>BURP</i>	Anthocyanin accumulation in tissues in response to UV light

3 讨论

3.1 基于 RNA-Seq 技术的植物叶色基因的发掘

新一代高通量测序技术是基于边合成边测序和可逆终止子的原理,双端测序的使用,不仅可以提高测序的深度,还能提高从头拼接的效率和准确性^[13-15]。一大批彩叶植物包括紫背天葵(*Gynura bicolor*)^[16]、银杏(*Ginkgo biloba*)^[17]、花叶矢竹(*Pseudosasa japonica* f. *Akebonosuji*)^[18]等的转录组已经被测序,为研究彩叶植物色素合成途径及叶色转化带来新的机遇。本项目组对紫晶 1 号 3 种叶色叶片作了转录组测序分析和功能基因预测,获得 145 338 条 Contig,进一步组装得到 53 654 条 Uni-gene, Q30 均大于 91%,被标记的序列数比例(被标记的序列数/过滤后的序列数)均在 72% 以上,Uni-gene 的 N50 为 2 229 bp,测序数据的输出和装配质

量能够满足转录分析的基本要求,为诠释紫晶 1 号叶片色素代谢分支调控机制,为色素合成关键基因的挖掘提供重要参考。

3.2 差异基因表达与功能注释分析

杨海芸等的研究结果表明,同一转录组 Uni-genes 的序列越长,其匹配率越高^[18],本研究中长度在 1 000 bp 以上的基因注释率为 53.05%,而在 300~1 000 bp 的 Unigenes 的注释率只有 29.50%,结果与前者相似。在生物学过程分类中,表达差异最大的是代谢过程,这与吴全金^[19]、郭小鸥^[20]的研究结果相似,表明紫晶 1 号叶片中代谢相关基因的表达存在显著差异,代谢产物的差异可能导致生物学功能与形态的差异。罗成科等在相关研究中指出,DUF 是一大组未表明功能的蛋白家族,一些植物特有的 DUF 蛋白家族在调控植物生长发育、防御病虫害和非生物胁迫应答反应等方面起着重要作

用^[21]。Kodama等的研究结果显示,拟南芥叶绿体运动响应是 *DUF* 家族中 *WEB1* 和 *PMI2* 2个基因互相作用调控的^[22-23],推测 *c42391.graph_c0* 可能与其作用过程相似。韩晓斌等的研究结果表明,阻遏或敲除 *Pkinase* 家族中的 *OsEDR1* 基因,突变体植株均会出现斑叶^[24],推测与 *OsEDR1* 同源的紫薇 *c45897.graph_c0* 转录因子影响了叶绿素的表达,进而使叶色发生变化。在毛果杨、拟南芥、水稻等研究结果中表明, *HMA* 基因家族是植物吸收、转运 Zn、Cu、Cd 等重金属的重要组分,拟南芥 *HMA1* 功能缺失会导致叶绿体中 Cu 含量下降和 Zn 含量上升,进而影响叶绿素的合成^[25-27]。以绿叶为对照,紫叶中属于 *HMA* 家族的 *c51707.graph_c0* 表达量显著上调,进一步推测 *c51707.graph_c0* 是控制紫薇叶色表达的候选基因。花色苷作为类黄酮物质的一类,是植物呈现红、橙、粉、紫、蓝等颜色的主要色素,而葡萄糖醛酸转移酶家族(*UDPGT*)在花色素合成途径起到重要作用。Kotepong等研究马六甲蒲桃果皮发现, *UFGT* 表达强度随着红巴梨果实的成熟而降低^[28],招雪晴等研究指出, *UFGT* 家族的 *PgUFGT* 基因在红宝石和水晶甜石榴品种的发育期内具有不同的表达模式^[29],这与本研究的结果相似,推测 *c47870.graph_c0* 可能参与了紫薇叶片中类黄酮物质的 3-O 位置的糖基化,与之相似的同源基因 *MDP0000543445*、*AtUGT78D2* 等已在红肉苹果^[30]、黄果草莓^[31]等研究中被找到,并已证实其与花青素的代谢活动密切相关。孙海燕等在三色堇花色形成机制的研究中,克隆出属于 *SDR*(短链脱氢/还原酶)超基因蛋白家族的 *VvDFR* 基因,推测该基因可能是花青素合成结构基因^[32]。本研究中 *c29868.graph_c0* 属于 *SDR* 基因家族,在类黄酮的生物合成过程中起到了正向调控作用,在紫叶中表达水平最高,说明 *c29868.graph_c0* 是紫晶 1 号叶片色素合成的关键候选基因。

本研究利用高通量测序技术对紫叶紫薇新品系紫晶 1 号转色期叶片进行了转录组测序,共获得 145 338 条 Contig,进一步组装得到 53 654 条 Unigene。KEGG 和 GO 注释与富集分析显示,差异 DEG 主要调节激素信号转导、碳水化合物代谢、核糖体代谢等生物学功能。将最终获得的 58 条差异表达 Unigene 对比到毛果杨和拟南芥 2 个模式植物的数据库,筛出与叶绿素代谢相关候选基因 9 条,有

2 条候选基因共同参与了叶绿素兼类黄酮代谢的调控,2 条候选基因可能参与类黄酮代谢过程,它们分别属于 *DUF547*、*Polyketide_cyc2*、*Pkinase* 等 13 个基因家族,主要经氧化还原反应、物质能量代谢、植物激素信号转导等机制介入紫薇叶色调控。

参考文献:

- [1] 李文芳,向昌国,王立志,等.紫薇花色化学性质及结构的初步研究[J].广州化工,2016,44(10):67-69.
- [2] 刘龙昌,陶珂,景康康,等.花色多态植物紫薇花部综合特征与繁育系统[J].林业科学,2015,51(1):42-54.
- [3] 徐婉,石俊,蔡明,等.尾叶紫薇与紫薇杂交后代花香气成分分析[J].西北植物学报,2014,34(2):387-394.
- [4] 谢宪,张林,王峰,等.紫薇抗寒性研究进展[J].天津农业科学,2016,22(5):11-14,20.
- [5] 汪志铮.紫薇新品种——“三红”紫薇[J].科学种养,2015(8):56.
- [6] 聂硕,张林,王峰,等.紫薇杂种 F_1 代的性状表现与倍性研究[J].山东农业科学,2017,49(1):17-22.
- [7] 王淑安,王鹏,杨如同,等.紫薇新品种“金幌”[J].园艺学报,2016,43(3):609-610.
- [8] 宋倩,黎榕,刘小平,等.紫薇叶片色素含量变化与叶色转变的关系[J].南方林业科学,2017,45(3):13-15,48.
- [9] 王淑安,王鹏,张振宇,等.紫薇叶片转录组分析及叶绿素和类胡萝卜素等途径基因的鉴定[J].分子植物育种,2016,14(5):1059-1066.
- [10] 曹桦,李涵,苗振,等.铁皮石斛叶色突变体初步研究[J].核农学报,2017,31(3):461-471.
- [11] 任杰.红花槭的良种选育与转色机理研究[D].合肥:安徽农业大学,2015.
- [12] 杨玉霞.花烛品种“Sonate”叶色突变体转录组分析及抗寒性研究[D].南京:南京农业大学,2015.
- [13] 高毅,史建荣.基于高通量测序技术分析麝香草酚处理禾谷镰孢菌后转录组学的变化[J].江苏农业学报,2017,33(6):1257-1265.
- [14] LIANG Y, CHEN S Y, AND LIU G S. Application of next generation sequencing techniques in plant transcriptome[J]. Hereditas, 2011,33(12):1317-1326.
- [15] 欧奇,李鑫,田洋,等.多油辣木转录组高通量测序及分析[J].江苏农业科学,2017,45(20):71-75.
- [16] 张少平,洪建基,邱珊莲,等.紫背天葵高通量转录组测序分析[J].园艺学报,2016,43(5):935-946.
- [17] 王欢利.黄叶银杏呈色基础及适应性差异[D].南京:南京林业大学,2015.
- [18] 杨海芸.花叶矢竹叶色变异机理研究[D].北京:北京林业大学,2015.
- [19] 吴全金.‘白鸡冠’茶树响应光调控的基因差异及理化特征分析[D].福州:福建农林大学,2015.

- [20] 郭小鹏. 薄皮甜瓜果实成熟品质形成与蛋白质组学分析[D]. 沈阳:沈阳农业大学,2017.
- [21] 罗成科,肖国举,李明. 不同未知功能结构域蛋白家族(*DUFs*)基因在植物中的生物学功能[J]. 植物生理学报, 2015,51(2):153-158.
- [22] KODAMA Y, SUETSUGU N, KONG S G, et al. Two interacting coiled-coil proteins, WEB1 and PMI2, maintain the chloroplast photorelocation movement velocity in *Arabidopsis* [J]. *Proc Natl Acad Sci USA*, 2010, 107 (45): 19591-19596.
- [23] KODAMA Y, SUETSUGU N, WADA M. Novel protein-protein interaction family proteins involved in chloroplast movement response [J]. *Plant Signal Behav*, 2011, 6 (4): 483-490.
- [24] 韩晓斌,徐冉,段朋根,等. 水稻斑点叶突变体 *spl101* 和 *spl102* 的筛选及候选基因鉴定[J]. 遗传, 2017, 39(4): 346-353.
- [25] NAGASAWA SN, MORI M, NAKAZAWA N, et al. Mutations in Rice (*Oryza sativa*) heavy metal ATPase2 (*OsHMA2*) restrict the translocation of Zinc and Cadmium[J]. *Plant Cell Physiol*, 2012, 53(1):213-224.
- [26] 王晓桐,李昊阳,徐吉臣. 毛果杨 *HMA* 基因家族的生物信息学分析[J]. 植物生理学报, 2014,50(7): 891-900.
- [27] WONG C K, JARVIS R S, SHERSON S M, et al. Functional analysis of the heavy metal binding domains of the Zn/Cdtransporting ATPase, HMA2, in *Arabidopsis thaliana* [J]. *New Phytol*, 2009, 181 (1): 79-88.
- [28] KOTEPONG P, KETSA S, VAN DOORN W. A white mutant of Malay apple fruit (*Syzygiu malaccense*) lacks transcript expression and activity for the last enzyme of anthocyanin synthesis, and the normal expression of a MYB transcription factor [J]. *Functional Plant Biology*, 2010, 38(1):75-86.
- [29] 招雪晴,李勃,苑兆和. 石榴 *PgUGFT* 基因克隆及表达分析[J]. 西北植物学报, 2017,37(4): 646-653.
- [30] 冀晓昊. 红肉苹果愈伤组织诱导及红色性状发育机理研究[D]. 泰安:山东农业大学,2015.
- [31] 张豫超. 森林草莓黄果表型形成机制研究[D]. 沈阳:沈阳农业大学,2017.
- [32] 孙海燕,安泽伟,李琴,等. 三色堇 *DFR* 基因的克隆及表达分析[J]. 江苏农业科学, 2015,43(11):34-37.

(责任编辑:陈海霞)