

练冬梅,赖正锋,姚运法,等. 玫瑰茄转录组测序及花青素合成相关基因表达分析[J]. 江苏农业科学,2020,48(6):41-45.
doi:10.15889/j.issn.1002-1302.2020.06.008

玫瑰茄转录组测序及花青素合成相关基因表达分析

练冬梅,赖正锋,姚运法,张少平,林碧珍,洪建基

(福建省农业科学院亚热带农业研究所,福建漳州 363005)

摘要:以玫瑰茄(*Hibiscus sabdariffa* Linnaeus)花瓣和花萼为研究对象,采用 Illumina Hiseq™ 2500 高通量测序技术,对花瓣和花萼进行转录组测序及花青素合成相关基因差异表达分析。结果表明,花瓣和花萼共获得 13.94 Gb 有效数据(clean data),Q30 碱基百分比均达到 93.0% 以上;共获得 1 399 个差异表达基因(DEGs),包括 65 个上调基因,1 334 个下调基因,且功能注释的基因有 1 176 个;筛选出了与花青素合成相关结构基因 *CHI*、*FLS*、*ANR*、*CHI* 在花萼中表达显著,*FLS*、*ANR* 在花瓣中表达显著。本研究丰富了花青素相关研究,可为阐明玫瑰茄花青素合成机制提供理论依据。

关键词:玫瑰茄;花瓣;花萼;转录组;花青素;差异表达基因

中图分类号: Q786;S681.901 **文献标志码:** A **文章编号:** 1002-1302(2020)06-0041-05

玫瑰茄(*Hibiscus sabdariffa* L.)为锦葵科木槿属一年生草本植物,别称洛神花(roselle、gongura)、芙蓉茄、山茄子、红果梅、苏丹茶、美丽纳等。原产西非至南亚,广泛分布于全球热带和亚热带地区,玫瑰茄在台湾、广西、广东、福建和云南等南部地区均有栽培。玫瑰茄 2004 年被卫生部和卫计委列入新食品原料名单^[1],是一种药食两用的植物。玫瑰茄花重要部位花瓣(亮黄色)和花萼(紫红色)具有药用功能,均是具有重要开发价值的器官,富含花青素、黄酮、多酚酸、有机酸等营养成分,花瓣部位主要用于开发花茶,而花萼部位利用率最高,已广泛用于制作玫瑰茄酱^[2]、饮料^[3]、酸奶^[4]、果酒^[5]等,在食品工业中被用作着色、调味添加剂;花萼具有利尿、抗氧化、抗癌、降血压、降血脂和降血糖等保健功能^[6-9],在玫瑰茄保健品开发方面具有巨大潜力。当前,国内对玫瑰茄研究主要集中在栽培技术、化学成分及功效分析、玫瑰茄花青素及玫瑰茄产品加工技术上,但其理论基础研究还很薄弱,例如玫瑰茄花青素合成相关代谢途径的研究尚属空白。

花青素是构成植物颜色的主要水溶性色素之一,属于类黄酮,主要以糖苷的形式存在于植物液

泡中^[10],迄今从自然界分离和鉴定出的花青素苷多达 600 种,主要从矢车菊素苷元(cyanidin)、飞燕草素苷元(delphinidin)、天竺葵素苷元(pelargonidin)、芍药花素苷元(peonidin)、矮牵牛素苷元(petunidin)、锦葵素苷元(malvidin)等 6 种花青素苷元衍生而来^[11]。花青素苷合成始于苯丙氨酸,途经多步酶促反应,主要由上游基因(*CHS*、*CHI*、*F3H*、*F3'H*和*F3'5'H*等)和下游基因(*DFR*、*ANS*、*GT*、*AT*和*MT*等)参与表达。研究表明,植物花青素兼具营养、药理以及基因工程中改良花色等作用^[12-14]。近年来,已有许多生物通过高通量转录组测序技术完成了全转录功能基因组测序^[15-16]。关于花青素合成相关基因在其他植物中也有相关研究^[17-18]。玫瑰茄分子遗传学与功能基因组学研究较落后,众多玫瑰茄农艺性状的分子调控机制仍不清楚,关联转录组学是一种鉴定调控目标性状候选基因的新方法。本研究利用 Illumina Hiseq™ 2500 高通量测序技术对玫瑰茄花瓣、花萼进行转录组测序,根据转录组数据及其花青素合成代谢途径相关基因分析,探讨玫瑰茄花瓣、花萼花青素合成机制和相关差异表达基因,以期后续玫瑰茄相关研究提供理论参考。

1 材料与方法

1.1 试验材料

玫瑰茄为福建省漳州市漳浦县种植品种,种植地点位于福建省农业科学院亚热带农业研究所试验农场,2017 年 6 月 15 日种植,8 月 25 日开始采集样品。采集标准:分别采集同一株玫瑰茄的花蕾

收稿日期:2018-11-19

基金项目:国家麻类产业技术体系专项(编号:CARS-16S07)。

作者简介:练冬梅(1987—),女,福建永安人,硕士,助理研究员,主要从事特色作物育种及综合利用研究。E-mail:woshildm1987@163.com。

通信作者:洪建基,研究员,主要从事特色作物育种及综合利用研究。E-mail:76227508@qq.com。

(开花前一天 09:00 采集) 10 朵,共 3 株,将花蕾的花瓣和花萼分离后进行混合,立即用液氮速冻,置于 -70 ℃ 超低温冰箱保存备用。

1.2 试验方法

1.2.1 高通量测序及数据组装 采用 PureLink Plant RNAreagent Kit 试剂盒提取玫瑰茄花瓣、花萼的总 RNA。分别采用 Nanodrop、Qubit 2.0、Aglient 2100 技术检测 RNA 样品的纯度、浓度和完整性,构建 cDNA 文库,再使用 Q-PCR 方法对文库的有效浓度进行准确定量,以保证文库质量。库检合格后,用 Illumina Hiseq™ 2500 进行高通量测序,获得原始数据,数据过滤后去除其中的接头序列及低质量读序 (reads),获得高质量的有效数据 (clean data),用 Trinity 软件对 clean data 进行组装,组装成为转录本序列 (transcripts),将不同样品组装结果中多个可变剪接的 transcripts 聚类到一个基因,得到单基因序列 (unigene) 库。

1.2.2 差异基因的功能注释 使用 BLAST 软件将 Unigene 与 NR、Swiss-Prot、GO、COG、KOG、NOG、KEGG、Pfam 等 8 个数据库进行比对 (E 值 $\leq 10^{-5}$),获得 Unigene 的注释信息。

1.2.3 花青素合成相关差异基因分析 利用上述数据库对玫瑰茄花瓣、花萼花青素合成进行检索,分析花青素合成相关基因在 KEGG 数据库功能注释及代谢路径。按照 FPKM (每千个碱基的转录每百万映射读取的片段) 法计算差异基因的表达量,如果倍数变化 (\log_2 Fold change) > 0 ,认为是上调表达,反之,便认为是下调表达。

2 结果与分析

2.1 数据组装及分析

玫瑰茄花瓣和花萼经高通量测序和质量控制,

共获得 13.94 Gb clean data,其中花瓣 clean data 为 7.29 Gb,花萼 clean data 为 6.65 Gb,其 Q30 碱基百分比均达到 93.0% 以上,表明测序结果可靠,可用于后续的分析。对组装结果进行统计 (表 1),总共产生 72 029 条单基因序列和 148 309 条转录本序列,N50 (对一条染色体进行测序,将测序得到的 reads 进行拼接,当拼接的序列长度达到染色体长度的一半时,叫做 N50 长度) 分别为 1 010、1 274 bp,组装完整性较高;其中长度区间位于 200 ~ 300 bp 的 Unigene 数量最多,为 24 697 条,占 34.29%。

表 1 组装结果统计分析

长度区间 (bp)	转录本序列数量 (条)	单基因序列数量 (条)
200 ~ 300	30 782	24 697
301 ~ 500	32 504	19 722
501 ~ 1 000	38 978	14 379
1 001 ~ 2 000	34 636	9 413
$\geq 2 001$	11 409	3 818
总数	148 309	72 029
总长度 (bp)	127 628 445	47 491 188
N50 长度 (bp)	1 274	1 010
平均长度 (bp)	860.56	659.33

2.2 Unigene 功能注释

为获得 Unigene 的功能注释信息,通过 NR、Swiss-Prot、KEGG、COG、KOG、NOG、GO、Pfam 等 8 个数据库进行注释分析 (表 2),在 72 029 条 Unigene 中,共获得 44 887 条 (62.32%) Unigene 的注释。以上 8 个数据库比对分析中,分别获得 44 274 (61.47%)、27 510 (38.19%)、15 908 (22.09%)、10 438 (14.49%)、23 790 (33.03%)、40 386 (56.07%)、20 981 (29.04%)、25 191 条 (34.97%) Unigene 功能注释。

2.3 差异表达基因 (DEGs) 分析

通过对玫瑰茄花瓣和花萼差异基因的表达分

表 2 单基因序列注释统计

生物信息数据库	注释单基因数 (条)	300 bp \leq 单基因长度 \leq 1 000 bp 的单基因数 (条)	单基因长度 $> 1 000$ bp 的单基因数 (条)	百分比 (%)
NR	44 274	21 611	12 708	61.47
Swiss-Prot	27 510	12 889	9 063	38.19
KEGG	15 908	7 512	4 902	22.09
COG	10 438	3 832	4 887	14.49
KOG	23 790	10 925	7 861	33.03
NOG	40 386	19 284	12 460	56.07
GO	20 981	10 075	6 302	29.04
Pfam	25 191	10 666	10 624	34.97
总注释单基因序列数	44 887	21 762	12 714	62.32

析,共获得 1 399 个 DEGs,包括 65 个上调基因,1 334 个下调基因;将 DEGs 单基因序列分别注释到 COG、GO、KEGG、KOG、Pfam、Swiss – Prot、NOG、NR 等 8 个数据库中,共获得 1 176 条基因功能注释,其中 COG 数据库 267 条,GO 数据库 526 条,KEGG 数据库 285 条,KOG 数据库 562 条,Pfam 数据库 853 条,Swiss – Prot 数据库 773 条,NOG 数据库 1 066 条,NR 数据库 1 165 条,其中 NR 数据库注释比最高,达 99.06%。

2.3.1 GO 功能注释 对玫瑰茄花瓣和花萼的 DEGs 进行 GO 功能注释,共分为三大类,分别为生物过程 (biological process)、细胞组分 (cellular component)和分子功能 (molecular function)。GO 分类统计显示,526 条 DEGs 被归到 40 个功能小类(表 3)。生物学过程中 DEGs 在单一生物过程、细胞过程和代谢过程 3 个功能小类中分布数量最多;细胞组分中 DEGs 在细胞、细胞成分和膜结构 3 个功能小类中分布数量最多;分子功能中 DEGs 在催化活性和结合活性 2 个功能小类中分布数量最多。

2.3.2 COG 功能注释 在玫瑰茄花瓣和花萼的转录本中,将注释到 COG 数据库的 267 条 DEGs 进行直系同源分类,并获得 21 个功能分类(图 1),差异基因注释主要集中在 R(一般性功能预测)、T(信号转导机制)、K(转录)和 L(复制、重组和修复)中。

2.3.3 KEGG 功能注释 将 DEGs 通过 KEGG 数据库比对(表 4),有 283 个 DEGs 得到注释,分别富集在 64 条代谢通路,其中涉及较多的是淀粉和蔗糖代谢(38 个)、戊糖和葡萄糖醛酸互相转化(35 个)、植物病原体相互作用(24 个)、甘油磷脂代谢(11 个)、胞吞作用(11 个)和苯丙素生物合成(10 个)等。本研究着重选择与花青素合成密切相关的代谢通路[类黄酮生物合成(5 个)],并找到与花青素代谢相关的差异基因。

2.4 花青素合成相关基因差异性分析

通过对玫瑰茄花瓣和花萼转录组测序结果进行功能注释、功能分类及代谢途径分析发现(表 5),花青素合成途径中,查尔酮 – 黄烷酮异构酶 (chalcone isomaerase,CHI)基因在玫瑰茄花萼中具有显著表达优势,上调表达 4.9 倍,黄酮醇合成酶 (flavonol synthase, FLS)、花青素还原酶 (anthocyanidin reductase, ANR)基因则在玫瑰茄花瓣中具有显著表达优势。这些差异表达基因参与着玫瑰茄花瓣和和花萼的花青素合成代谢途径,影响着玫瑰茄花青素的合成。

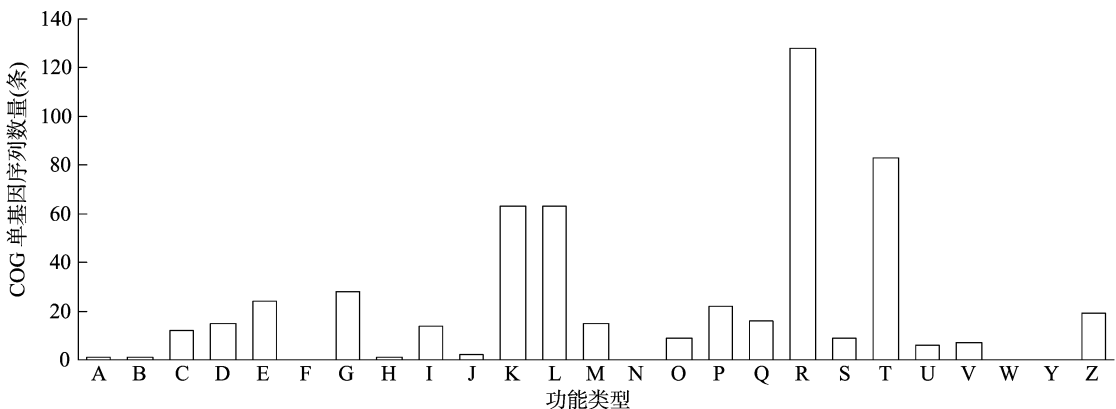
表 3 差异表达基因 GO 功能注释

项目	GO 层级	单基因序列数量 (个)
生物学过程	代谢过程	247
	细胞过程	258
	单一生物过程	261
	刺激应答	93
	生物调节	75
	定位	86
	细胞成分或生物合成	99
	发育过程	95
	多细胞生物过程	92
	信号传导	31
	繁殖过程	76
	多生物过程	62
	生长	57
	免疫系统过程	3
	繁殖	7
细胞组分	运动	2
	细胞	193
	细胞成分	193
	细胞器	119
	膜结构	141
	细胞器部分	33
	膜结构部分	82
	复杂大分子	27
	胞外区	23
	膜封闭腔	4
	细胞连接	6
分子功能	共质体	6
	保外区域部分	1
	催化活性	293
	结合活性	232
	转运活性	42
	结构分子活性	4
	核酸结合转录因子活性	8
	电子载体活性	13
	分子转导活性	7
	酶调节活性	6
	抗氧化活性	3
	蛋白结合转录因子活性	3
	鸟嘌呤核苷酸交换因子活性	2
	营养存储活性	2

3 讨论

本研究通过 Illumina HiSeq™ 2500 高通量测序技术构建玫瑰茄花瓣和花萼转录组数据库,比较分析花瓣和花萼数据,发现花青素合成相关代谢途径中差异表达的基因。

玫瑰茄花瓣、花萼中花青素代谢表现:*p* – 香豆酰 – 辅酶 A (*p* – coumaroyl – CoA) 和 3 分子丙二醛 – 辅酶 A (3 × Malnoyl – CoA), 在查尔酮合成酶



A—RNA 的加工和修饰；B—染色体结构和动态变化；C—能源的产生和转化；D—细胞周期调控，细胞分裂，染色体分配；E—氨基酸转运与代谢；F—核苷酸转运与代谢；G—碳水化合物转运与代谢；H—辅酶转运与代谢；I—脂质转运与代谢；J—翻译、核糖体结构和生物合成；K—转录；L—复制、重组和修复；M—细胞壁/细胞膜的生物发生；N—细胞运动；O—次生代谢产物的合成、转运和代谢；P—无机离子转运与代谢；Q—次生代谢产物的合成、转运和代谢；R—一般性功能预测；S—未知功能；T—信号转导机制；U—胞质运输、分泌和囊泡运输；V—防御机制；W—细胞外结构；Y—核结构；Z—细胞骨架

图1 差异表达基因 COG 注释分类统计

表 4 差异表达基因 KEGG 功能注释

编号	KEGG 功能注释	单基因数列数量 (个)
ko04070	磷脂酰肌醇信号系统	9
ko04075	植物激素信号转导	6
ko00430	牛磺酸和亚牛磺酸代谢	1
ko00906	类胡萝卜素生物合成	1
ko00073	角质、小檗碱和蜡的生物合成	2
ko00010	糖酵解/糖异生	3
ko01220	芳香族化合物的降解	1
ko00941	类黄酮生物合成	5
ko01230	氨基酸的生物合成	4
ko00500	淀粉和蔗糖代谢	38
ko00531	糖胺聚糖降解	2
ko00650	丁酸代谢	1
ko04145	吞噬	3
ko00196	光合作用 - 天线蛋白质	1
ko00564	甘油磷脂代谢	11
ko00330	精氨酸和脯氨酸代谢	3
ko00350	酪氨酸代谢	3
ko00020	柠檬酸循环(TCA 循环)	1
ko03013	RNA 运输	1
ko00908	玉米素生物合成	1
ko00460	氰氨基酸代谢	1
ko00770	泛酸盐和辅酶 A 生物合成	1
ko00071	脂肪酸降解	2
ko00380	色氨酸代谢	1
ko00565	醚脂质代谢	7
ko00591	亚油酸代谢	2
ko00940	苯丙素生物合成	10
ko04130	SNARE 在水泡运输中的相互作用	3
ko00920	硫代谢	1
ko00945	二苯乙烯类和姜酚生物合成	2
ko04144	胞吞作用	11
ko00260	甘氨酸,丝氨酸和苏氨酸代谢	2

续表 4

编号	KEGG 功能注释	单基因数列数量 (个)
ko00410	β -丙氨酸代谢	2
ko01200	碳代谢	4
ko00130	泛醌和其他萜类醌生物合成	3
ko00040	戊糖和葡萄糖醛酸互相转化	35
ko00290	缬氨酸、亮氨酸和异亮氨酸的生物合成	1
ko00520	氨基糖和核苷糖代谢	2
ko00280	缬氨酸、亮氨酸和异亮氨酸降解	1
ko00960	托品烷、哌啶和吡啶生物碱的生物合成	1
ko04141	蛋白质在内质网中的加工	1
ko04626	植物病原体相互作用	24
ko00190	氧化磷酸化	3
ko00710	光合生物体中的碳固定	1
ko00270	半胱氨酸和蛋氨酸代谢	2
ko00053	抗坏血酸和新陈代谢	10
ko03015	mRNA 监视通路	1
ko00052	半乳糖代谢	2
ko03010	核糖体	1
ko04120	泛素介导的蛋白水解	1
ko01040	不饱和脂肪酸的生物合成	1
ko00592	α -亚麻酸代谢	6
ko00561	甘油酯代谢	5
ko03430	错配修复	1
ko00910	氮代谢	6
ko01210	2-氧代羧酸代谢	2
ko00904	二萜生物合成	1
ko00950	异喹啉生物碱生物合成	1
ko00360	苯丙氨酸代谢	7
ko01212	脂肪酸代谢	1
ko00630	乙醛酸和二羧酸代谢	2
ko00562	肌醇磷酸盐代谢	10
ko00250	丙氨酸、天冬氨酸和谷氨酸代谢	4
ko00600	鞘脂代谢	1

表 5 玫瑰茄花瓣、花萼花青素合成相关基因差异表达情况

基因 ID	名称	基因简称	倍数变化(log ₂ Fold change)	基因表达
c48234. graph_c0	查尔酮 - 黄烷酮异构酶基因	CHI	4.951 297 991 780 12	上调(up)
c64297. graph_c0	黄酮醇合成酶基因	FLS	-4.304 995 556 012 35	下调(down)
c69277. graph_c0	花青素还原酶基因	ANR	-3.500 692 726 516 15	下调(down)

(chalcone synthase, CHS) 催化下,生成查尔酮(chalcone),查尔酮又在 CHI 作用下形成柚皮素(naringenin,NAR),此过程花萼中 CHI 基因表达比花瓣显著,将大量查尔酮转化成花青素合成代谢所必需的 NAR;玫瑰茄花瓣、花萼中 NAR 在黄烷酮-3-羟基化酶(flavanone 3-hydroxylase,F3H)催化下产生二氢山奈酚(dihydrokaempferol,DHK),DHK 分别在类黄酮-3'-羟基化酶(flavonoid-3'-hydroxylase,F3'H)和类黄酮-3'5'-羟基化酶(flavonoid-3'5'-hydroxylase,F3'5'H)作用下生成二氢槲皮素(dihydroquercetin,DHQ)和二氢杨梅素(dihydromyricetin,DHM),DHK、DHQ 和 DHM 在二氢黄酮醇-4-还原酶(dihydroflavonol 4-reductase,DFR)作用下分别生成无色天竺葵苷元(leucopelargonidin)、无色矢车菊苷元(leucocyanidin)和无色飞燕草素苷元(leucodelphinidin),而后分别在花青素苷合成酶(anthocyanin synthase,ANS)作用下生成有色的天竺葵苷元(pelargonidin)、矢车菊苷元(cyanidin)和飞燕草素苷元(delphinidin),各种花青素苷元分别在葡萄糖基转移酶(glycosyl transferases,GT)等作用下,生成稳定的花青素苷。此过程中玫瑰茄花瓣中 FLS 和 ANR 基因显著表达,进入黄酮醇代谢途径和原花青素代谢途径,说明 FLS、ANR 可能通过底物竞争方式影响花青素合成,而玫瑰茄花萼 CHI 上调表达,FLS 和 ANR 下调表达,有利于花青素合成。本研究丰富了花青素相关研究,可为阐明玫瑰茄花青素合成机制提供理论依据。

参考文献:

[1] 佚名. 卫计委: 有关新食品原料,普通食品名单汇总[J]. 饮料工业,2014,17(8):58-62.

[2] Adeleke T O, Ashsye O A. Quality attributes of stored Roselle jam [J]. Int Food Res,2009,16(3):363-371.

[3] 耿 华. 关于玫瑰茄饮料的试验与研究[J]. 软饮料工业,1996(1):21-24.

[4] Iwalokun A B, Shittu M O. Effect of *Hibiscus sabdariffa* (Calyce) extract on biochemical and organoleptic properties of yogurt [J]. Pakistan Journal of Nutrition,2007,6(2):172-182.

[5] 孙军涛,肖付刚,祝鑫冉. 玫瑰茄苹果果酒的研制[J]. 酿酒科

技,2015(2):78-80.

[6] Hirunpanich V, Utaipat A, Morales N P, et al. Antioxidant effects of aqueous extracts from dried calyx of *Hibiscus sabdariffa* Linn. (Roselle) *in vitro* using rat low-density lipoprotein (LDL) [J]. Biological & Pharmaceutical Bulletin,2005,28(3):481-484.

[7] Akim A, Ling L C, Rahmat A, et al. Antioxidant and anti-proliferative activities of Roselle juice on Caov-3, MCF-7, MDA-MB-231 and HeLa cancer cell lines [J]. African Journal of Pharmacy and Pharmacology,2011,5(7):957-965.

[8] Ubani C S, Joshua P E, Oraeki A N. Influence of aqueous extract of *Hibiscus sabdariffa* calyces on lipid profile of phenobarbitone induces wistar albinorats [J]. Journal of Pharmacy Research, 2010(2):319-324.

[9] Gosain S, Ircchiaya R, Sharma P C, et al. Hypolipidemic effect of ethanolic extract from the leaves of *Hibiscus sabdariffa* L. in hyperlipidemic rats [J]. Acta Poloniae Pharmaceutica,2010,67(2):179-184.

[10] Fleschhut J, Kratzer F, Reckemmer G, et al. Stability and biotransformation of various dietary anthocyanins *in vitro* [J]. European Journal of Nutrition,2006,45(1):7-18.

[11] Kong J M, Chia L S, Goh N K, et al. Analysis and biological activities of anthocyanins [J]. Phytochemistry, 2003, 64(5):923-933.

[12] Dos Passos A D, Madrona G S, Marcolino V A, et al. The use of thermal analysis and photoacoustic spectroscopy in the evaluation of maltodextrin microencapsulation of anthocyanins from *juçara* palm fruit (*Euterpe edulis* Mart.) and their application in food [J]. Food Technol Biotechnol,2015,53(4):385-396.

[13] Lippert E, Ruemmele P, Obermeier F, et al. Anthocyanins prevent colorectal cancer development in a mouse model [J]. Digestion, 2017,95(4):275-280.

[14] Fernandes I, Marques C, Evora A, et al. Pharmacokinetics of table and Port red wine anthocyanins: a crossover trial in healthy men [J]. Food & Function,2017,8(5):2030-2037.

[15] Sheng L X, Xia W, Zang S, et al. Transcriptome-sequencing analyses reveal putative genes related to flower color variation in Chinese *Rosa rugosa* [J]. Acta Physiologiae Plantarum,2018,40(3):62-75.

[16] Wang J W, Guo M L, Li Y H, et al. High-throughput transcriptome sequencing reveals the role of anthocyanin metabolism in *Begonia semperflorens* under high light stress [J]. Photochemistry and Photobiology,2018,94(1):105-114.

[17] 孟亚南,张 琳,刘召强,等. 荷花转录组测序及花青素苷合成相关基因表达分析 [J]. 西南林业大学学报(自然科学版),2018,38(2):61-69.

[18] 蒋会兵,孙云南,李 梅,等. 紫娟茶树叶片不同发育期花青素积累及合成相关基因的表达 [J]. 茶叶科学,2018,38(2):174-182.