

## 云锦杜鹃转录组分析

许蔷薇<sup>1,2</sup>, 楼雄珍<sup>1,2</sup>, 杨彬<sup>1,2</sup>, 林二培<sup>1,2</sup>, 童再康<sup>1,2</sup>

(1. 浙江农林大学 林业与生物技术学院, 浙江 杭州 311300; 2. 浙江农林大学 省部共建亚热带森林培育国家重点实验室, 浙江 杭州 311300)

**摘要:** 云锦杜鹃 *Rhododendron fortunei* 具有很高的园艺观赏价值。由于缺乏相关的遗传信息, 云锦杜鹃的多样性、分子辅助育种等工作进展较为缓慢。通过高通量测序技术(Illumina), 对云锦杜鹃的转录组进行了测定和分析。通过测序拼接和去冗余共获得 84 633 条单基因簇(unigene)序列, 平均长度为 691.4 bp。通过与公共数据库比对成功注释到了 35 526 条单基因簇, 其中与葡萄 *Vitis vinifera* 基因相似的序列最多。根据基因本体论数据库(GO)可将 23 215 个单基因簇归类于生物学过程、细胞组分及分子功能等三大类的 55 个功能组; 真核直系同源基因数据库(KOG)将 11 085 条单基因簇分为 26 个功能分类, 涉及一般功能预测的单基因簇最多, 多达 1 970 条; 以京都基因与基因组百科全书(KEGG)作为参考, 依据代谢途径可将 9 887 个单基因簇定位到 272 个代谢通路。进一步分析基因注释结果, 挖掘获得 24 个编码 MADS-box 基因的单基因簇分属于 10 个不同亚家族。此外, 检测到了 21 900 个简单序列重复(SSR)位点, 可用于 SSR 标记开发。图 4 表 5 参 27

**关键词:** 林木育种学; 云锦杜鹃; 转录组; 基因注释; 简单序列重复(SSR)

**中图分类号:** S722.3; Q753 **文献标志码:** A **文章编号:** 2095-0756(2019)06-1190-09

## Transcriptome sequencing and analysis of *Rhododendron fortunei*

XU Qiangwei<sup>1,2</sup>, LOU Xiongzen<sup>1,2</sup>, YANG Bin<sup>1,2</sup>, LIN Erpei<sup>1,2</sup>, TONG Zaikang<sup>1,2</sup>

(1. School of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China; 2. State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China)

**Abstract:** To enhance studies on genetic diversity and molecular assistance breeding of *Rhododendron fortunei*, one of the native *Rhododendron* species in China with a high horticultural and ornamental value, the transcriptome of *R. fortunei* was sequenced using Illumina technology. In addition, sequence assembly, public databases, Gene Ontology (GO) analysis, Eukaryotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and MADS-box transcription factor were used. Results of the sequence assembly obtained 84 633 unigenes with a mean length of 691.4 bp. Public databases were successfully searched annotating 35 526 unigenes, and GO analysis showed that 23 215 unigenes were potentially involved in 55 functional categories of biological processes, cellular components, and molecular functions. The KOG analysis indicated 11 085 unigenes classified into 26 functional groups; whereas, the KEGG analysis assigned 9 887 unigenes to 272 known pathways. Also, 24 unigenes were identified to encode MADS-box transcription factor belonging to 10 different subfamilies. In addition, 21 900 simple sequence repeat (SSR) loci were determined in the transcriptome of *R. fortunei*. In conclusion, this study could provide an important foundation for genetic diversity analysis, marker development, molecular assistant breeding, and functional gene isolation in *R. fortunei*, as well as future SSR marker development. [Ch, 4 fig. 5 tab. 27 ref.]

**Key words:** forest tree breeding; *Rhododendron fortunei*; transcriptome; gene annotation; SSR

收稿日期: 2018-11-07; 修回日期: 2019-03-29

基金项目: 浙江省农业新品种选育重大科技专项(2016C02052-12)

作者简介: 许蔷薇, 从事林木遗传育种研究。E-mail: 459262240@qq.com。通信作者: 童再康, 教授, 博士, 从事林木遗传育种等研究工作。E-mail: zktong@zafu.edu.cn

随着第 2 代测序技术的发展, 转录组测序(RNA-Seq)已成为植物分子生物学研究的重要手段, 广泛应用于功能基因挖掘、分子标记开发、代谢通路和调控机制研究等方面<sup>[1-3]</sup>。该技术具有成本低、数据量大、效率高、准确性高等优点<sup>[4]</sup>, 可对组织或者细胞中所有 RNA 进行测序, 并通过读段(reads)的拼接和丰度统计获得相应的转录本序列信息及其表达水平<sup>[5]</sup>。云锦杜鹃 *Rhododendron fortunei* 为杜鹃花科 Ericaceae 杜鹃花属 *Rhododendron* 植物, 为中国特有种, 主要分布于安徽、湖南、湖北、浙江等省<sup>[6]</sup>。云锦杜鹃叶形大, 花形美丽具清香, 具有较高的园艺观赏价值; 适应性较强, 易人工栽培, 也常作为杜鹃属种间杂交的亲本; 此外, 枝叶等组织中含有槲皮素、山柰酚、杨梅素等有效活性成分, 可开发入药<sup>[7-8]</sup>。可见, 云锦杜鹃作为一种兼具观赏价值和药用价值的优良木本花卉, 开发潜力巨大, 具有良好的产业化前景。目前, 云锦杜鹃的研究主要集中在野生资源调查<sup>[9]</sup>、群落结构<sup>[10]</sup>、无性繁殖技术<sup>[11]</sup>、有效成分分析<sup>[8,12]</sup>、光合特性<sup>[13]</sup>、菌根共生等方面<sup>[14]</sup>。由于缺乏遗传信息, 云锦杜鹃的遗传多样性、分子辅助育种等的研究较为滞后。本研究通过 RNA-seq 高通量测序技术对云锦杜鹃转录组进行测定, 通过序列拼接、功能注释和分析, 获取大量的序列信息, 旨在为云锦杜鹃分子标记开发, 遗传多样性分析, 以及功能基因挖掘、重要性状形成分子机制等研究提供序列信息。

## 1 材料与方法

### 1.1 植物材料

前期以天台华顶林场云锦杜鹃优株茎段为外植体, 通过组培获得云锦杜鹃无性系 ZL-01。以该无性系组培苗为材料, 取其嫩叶、嫩茎和根, 液氮冷冻后保存于-80 °C用于 RNA 提取。

### 1.2 RNA 提取与质控

采用 Trizol 试剂(Invitrogen), 按照试剂说明书分别提取组培苗根、茎、叶的总 RNA。采用 Agilent 2100 Bioanalyzer(Agilent), 对总 RNA 样品的纯度、浓度和完整性进行检测评估, 等量混合后用于测序文库构建。

### 1.3 测序文库构建

取检验合格的 RNA, 用带有 Oligo(dT)的磁珠富集 mRNA, 将得到的 mRNA 随机打断成短片段, 用六碱基随机引物合成一链 cDNA, 然后加入缓冲液、dNTPs、DNA polymerase I 和 RNase H 合成二链 cDNA。双链 cDNA 经纯化后进行末端修复, 加 ploy(A)并连接测序接头, 通过 PCR 扩增及纯化后得到测序文库。采用 Illumina HiSeq 2500 进行转录组的测序, 测序由北京诺禾致源生物公司完成。

### 1.4 转录组数据分析

针对测序得到的原始数据(raw data)进行接头和低质量测序片段(reads)去除等处理, 获得高质量的干净数据(clean data); 计算干净数据的 Q20 和 Q30, GC 含量和碱基错误率, 评估测序质量; 利用序列的重叠, 通过 Trinity 软件对干净数据数据进行序列组装, 将短测序片段延伸成较长的片段, 并得到片段集合, 利用 De Bruijn 方法<sup>[15]</sup>得到转录本(transcripts)和单基因簇(unigene)序列。

### 1.5 基因功能注释

利用 BLAST 和 HMMER 软件将云锦杜鹃单基因簇序列与公共数据库进行比对, 根据基因的相似性进行功能注释, 得到与给定单基因簇具有最高序列相似性的蛋白, 从而得到该单基因簇的蛋白功能注释信息, 其中 BLAST 参数  $E \leq 1e^{-5}$ , HMMER 参数  $E \leq 1e^{-10}$  作为筛选标准。比对采用的公共数据库包括美国生物信息中心(NCBI)非冗余蛋白数据库(Non-redundant protein database, Nr), NCBI 核酸序列数据库(Non-redundant nucleotide sequences, Nt), 蛋白质序列数据库(Swiss Prot protein database, SwissProt), 真核直系同源基因数据库(Eukaryotic orthologous groups, KOG), 蛋白质家族数据库(Protein families database, Pfam), 基因本体论数据库(Gene ontology, GO)及京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)。

### 1.6 SSR 分析

采用 MISA 软件对云锦杜鹃单基因簇进行简单序列重复(SSR)分析, 鉴定其中 6 种类型的 SSR; 各 SSR 类型重复次数设定为: 单核苷酸 SSR, 重复数  $\geq 10$  次; 二核苷酸 SSR, 重复数  $\geq 6$  次; 三至六核苷酸 SSR, 重复数  $\geq 5$  次。

## 2 结果与分析

### 2.1 测序结果与组装

分别提取无性系 ZL-01 组培苗根、茎和叶的 RNA。检测结果表明：根 RNA 的质量浓度为  $362 \text{ mg} \cdot \text{L}^{-1}$ ，28S/18S 为 1.6，RNA 完整性计数(RNA integrity number, RIN)为 9.6；茎 RNA 的质量浓度为  $515 \text{ mg} \cdot \text{L}^{-1}$ ，28S/18S 为 1.8，RIN 值为 9.8；叶 RNA 的质量浓度为  $485 \text{ mg} \cdot \text{L}^{-1}$ ，28S/18S 为 1.8，RIN 值为 9.7。这些指标均符合建库测序要求，因此，将这 3 个组织的 RNA 等量混合后进行测序。

测序数据经处理后得到 94 252 430 个干净测序片段(clean reads)，包含了 14.14 Gb 核苷酸序列信息，GC 含量为 47.58%。测序质量评估结果显示，碱基错误率为 0.02%，Q20 为 96.70%，Q30 为 91.85%。这些表明该转录组的数据量和质量均较高，为后续的序列组装提供了高质量的原始数据。

通过 Trinity 软件组装得到 112 777 个转录本，总长度为 91 621 682 bp，平均长度为 812.4 bp，N50 为 1 410，其中长度在 1 kb 以上的有 29 225 条，占 25.92%；2 kb 以上的 10 418 条，占 9.24%(表 1)。对转录本进行聚类 and 组装得到 84 633 个单基因簇，总长度为 58 517 298 bp，平均长度为 691.4 bp，N50 为 1 177 bp；其中超过 1 kb 的有 16 631 条，占 19.65%；超过 2 kb 的 5 760 条，占 6.81%(表 1)。

### 2.2 基因功能注释

将单基因簇序列与 Nr, Nt, SwissProt, GO, KOG, KEGG 和 Pfam 7 个数据库进行比对，共有 35 526 条单基因簇获得成功注释，占单基因簇的 41.97%(表 2)。其中，30 700 条单基因簇获得 Nr 数据库注释，占 36.27%；23 215 条单基因簇获得 GO 数据库注释，占 27.43%；22 882 条单基因簇获得 Pfam 数据库注释，占 27.03%；22 202 条单基因簇获得 Swiss-prot 数据库注释，占 26.23%；18 046 条单基因簇获得 Nt 数据库注释，占 21.32%；11 085 条单基因簇获得 KOG 数据库注释，占 13.09%；9 887 条单基因簇获得 KEGG 数据库注释占 11.68%(表 2)。

表 2 云锦杜鹃单基因簇功能注释

Table 2 Annotation of unigenes in transcriptome of *R. fortunei*

数据库	被注释的单基因簇/条	百分比/%	数据库	被注释的单基因簇/条	百分比/%	数据库	被注释的单基因簇/条	百分比/%
Nr	30 700	36.27	SwissProt	22 202	26.23	KOG	11 085	13.09
KEGG	9 887	11.68	Pfam	22 882	27.03	总计	35 526	41.97
Nt	18 046	21.32	GO	23 215	27.43			

在 Nr 库中，云锦杜鹃转录组注释到其他物种的单基因簇基因序列共 30 700 条。其中与葡萄 *Vitis vinifera* 基因序列相似的最多，所占比例为 24.39%；其次为中粒咖啡 *Coffea canephora*，所占比例为 7.39%；第 3 为可可 *Theobroma cacao*，占 6.77%；其他相似性序列数量大于 3% 的物种有烟草 *Nicotiana tomentosiformis* (6.19%)，毛果杨 *Populus trichocarpa* (4.09%)，甜橙 *Citrus sinensis* (3.82%)，麻疯树 *Jatropha curcas* (3.76%) 和梅花 *Prunus mume* (3.16%)，其他物种占 40.43%(图 1)。

### 2.3 单基因簇的 GO 功能分类

通过 GO 数据库的注释，共有 23 215 条单基因簇获注释信息，得到 119 389 个 GO 功能注释。由分类结果可知：生物学过程(biological process)最多 55 692 条，占 46.65%，其次是细胞组分(cellular component)，35 577 条，占 29.80%，最少的分子功能(molecular function)，有 28 120 条，占 23.55%；这三大功能分类又可分为 55 个亚类，其中生物学过程 23 个亚类，细胞组分 18 个亚类，分子功能 14 个亚类(图 2)。生物学过程中，涉及细胞过程、代谢过程和单一有机体进程的单基因簇较多，分别有 12 698，12 376 和 9 581 条；细胞组分中涉及较多的是细胞、细胞部分和大分子复合体，分别有

表 1 云锦杜鹃转录组组装结果

Table 1 Summary of transcriptome assembly for *R. fortunei*

长度范围/bp	转录本/个	单基因簇/个
<301	35 575(31.54)	32 101(37.93)
301~500	25 511(22.62)	20 759(24.53)
501~1 000	22 466(19.92)	15 142(17.89)
1 001~2 000	18 807(16.68)	10 871(12.84)
>2 000	10 418(9.24)	5 760(6.81)
总数	112 777	84 633
总长度	91 621 682	58 517 298
N50 长度	1 410	1 177
平均长度	812.4	691.4

说明：括号内数值为占总数的百分比(%)

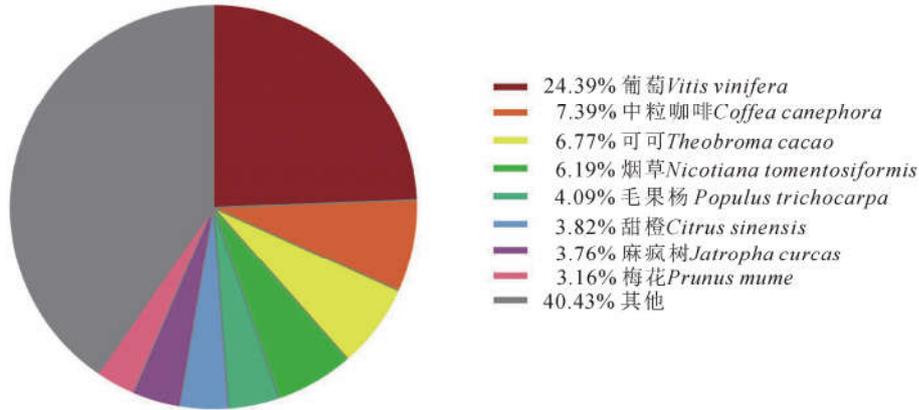


图 1 云锦杜鹃单基因簇 Nr 数据库比对相似性物种分布图

Figure 1 Similarity of unigenes of *R. fortunei* with those of other species in Nr database

7 216, 7 214 和 4 713 条; 分子功能中涉及较多的有结合功能和催化活性, 分别有 13 302, 10 524 条(图2)。

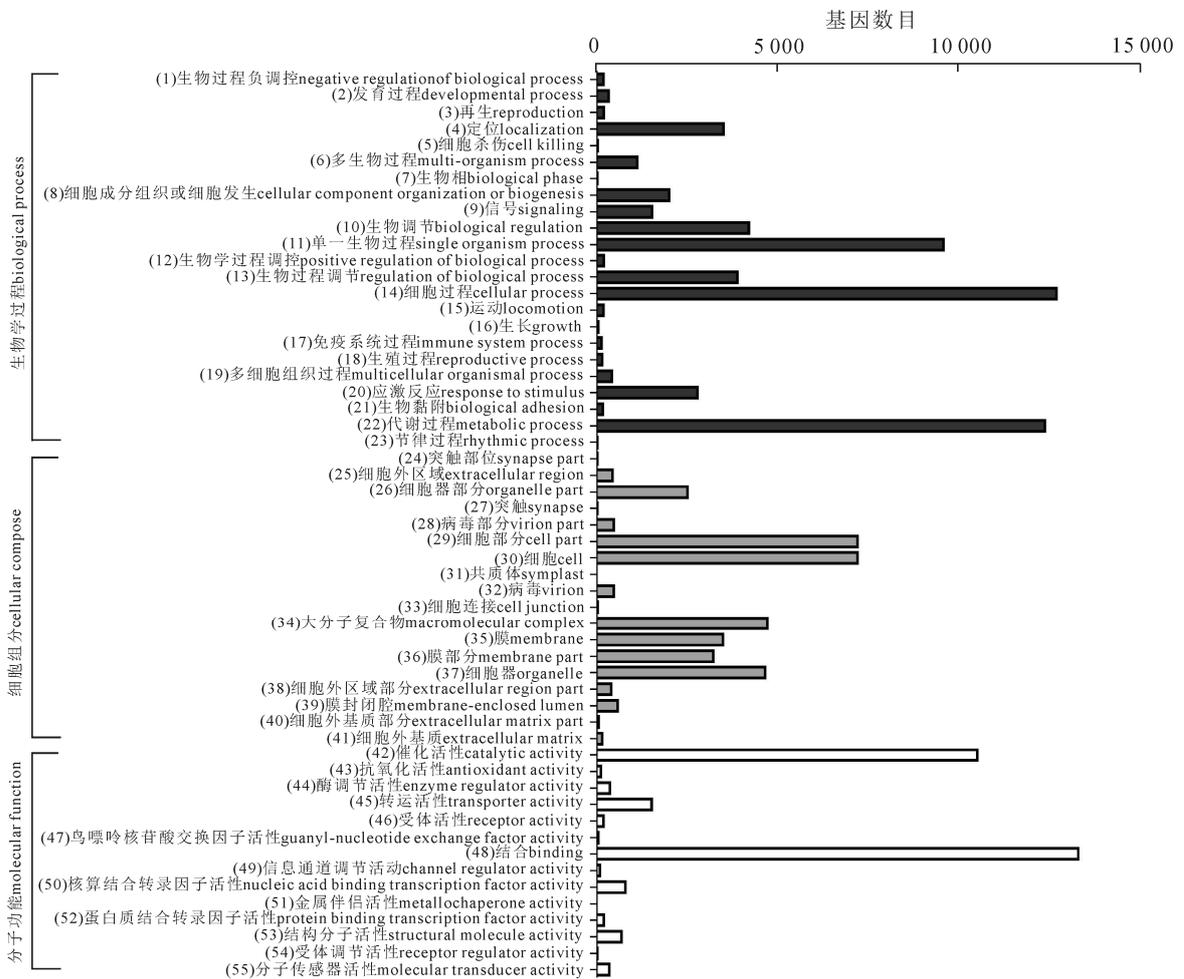


图 2 云锦杜鹃单基因簇的 GO 功能分类

Figure 2 GO functional categories of *R. fortunei* unigenes

### 2.4 单基因簇的 KOG 功能注释

通过 KOG 数据库对云锦杜鹃单基因簇进行注释, 结果显示有 11 085 条序列获得 12 475 个注释信息, 可分为 26 个功能分类(图 3)。从基因功能分类来看, 涉及一般功能预测的序列最多, 多达 1 970 条; 涉及翻译后修饰、蛋白翻转、分子伴侣功能的序列次之, 有 1 527 条; 而涉及核结构、胞外结构和细胞运动的序列很少, 仅有 37, 34 和 5 条(图 3)。

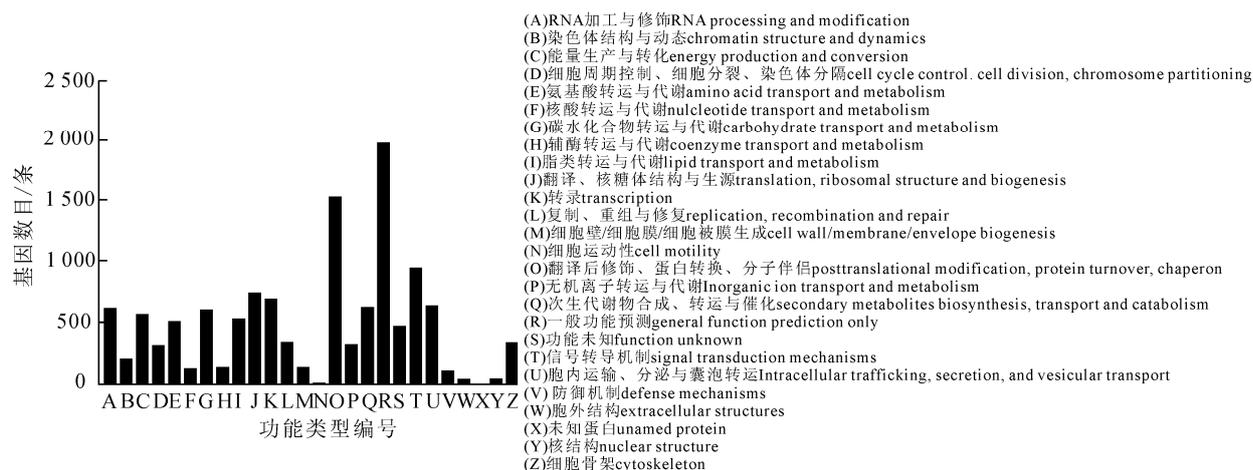


图3 云锦杜鹃单基因簇的 KOG 功能分类

Figure 3 KOG functional categories of *R. fortunei* unigenes

## 2.5 KEGG 通路注释

利用 KEGG 注释系统对云锦杜鹃单基因簇涉及的代谢途径进行分析, 结果显示: 9 887 条单基因簇得到 15 455 个注释, 归属于 272 条通路。按获得注释的基因数量进行排序, 取前 20 个途径, 发现含有 200 条单基因簇以上的通路有 10 个, 涉及碳代谢的单基因簇最多, 有 377 条; 其次是与核糖体相关的单基因簇, 有 364 条; 第 3 位是与氨基酸合成相关的单基因簇, 有 348 条; 涉及其他途径有内质网蛋白加工(319)、剪切体(256)、植物激素信号转导(244)、淀粉和蔗糖代谢(240)、RNA 转运(227)、氧化磷酸化(224)和植物病原物相互作用(210); 其他通路的单基因簇数量均在 200 以下(表 3)。

## 2.6 MADS-box 基因家族分析

MADS-box 基因家族在植物花分生组织形成、花器官发育等过程中发挥关键作用<sup>[16]</sup>。鉴定转录组中的 MADS-box 基因有助于进一步研究或调控云锦杜鹃成花过程。通过与 Nr, Nt 和 Swiss-prot 三大数据库

表 3 云锦杜鹃单基因簇的 KEGG 分析

Table 3 Summary KEGG pathway of *R. fortunei* transcriptome

KEGG 通路	通路 ID	单基因簇/条
碳代谢 carbon metabolism	ko01200	377
核糖体 ribosome	ko03010	364
氨基酸合成 biosynthesis of amino acids	ko01230	348
内质网蛋白加工 protein processing in endoplasmic reticulum	ko04141	319
剪切体 spliceosome	ko03040	256
植物激素信号转导 plant hormone signal transduction	ko04075	244
淀粉和蔗糖代谢 starch and sucrose metabolism	ko00500	240
RNA 转运 RNA transport	ko03013	227
氧化磷酸化 oxidative phosphorylation	ko00190	224
植物病原物互作 plant-pathogen interaction	ko04626	210
内吞作用 endocytosis	ko04144	194
嘌呤代谢 purine metabolism	ko00230	193
糖酵解途径 glycolysis / gluconeogenesis	ko00010	184
泛素介导蛋白降解 ubiquitin mediated proteolysis	ko04120	174
mRNA 监测通路 mRNA surveillance pathway	ko03015	171
苯丙素生物合成 phenylpropanoid biosynthesis	ko00940	169
细胞周期 cell cycle	ko04110	168
氨基糖和核苷酸糖代谢 amino sugar and nucleotide sugar metabolism	ko00520	160
嘧啶代谢 pyrimidine metabolism	ko00240	148
RNA 降解 RNA degradation	ko03018	146

说明: 注释数量总数为 15 455

比对, 共找出编码 MADS-box 基因的单基因簇序列 24 条, 分别属于 10 个不同的亚家族(表 4)。其中 AGL17 亚家族成员最多, 包含 c41091\_g4, c39737\_g1, c36841\_g1, c38538\_g1 和 c40334\_g4; SQUA 和 SVP 亚家族则各有 4 个成员, 分别为 c50592\_g1, c10093\_g1, c9572\_g1, c2583\_g1 与 c37773\_g1, c31351\_g1, c30064\_g1, c33061\_g2; TM3 亚家族有 3 个成员, 其他亚家族仅有 2 个或 1 个成员(表 4)。这些单基因簇的同源基因分别与花分生组织发育、花期调控、花器官发育、果实发育等重要生物学过程相关。

表 4 云锦杜鹃成花相关 MADS-box 基因鉴定

Table 4 Identification of Floral related MADS-box genes of *R. fortunei*

序列号	注释	功能	亚家族
c18022_g1	Agamous-like MADS-box protein AGL15	雄蕊, 心皮, 胚珠和果实发育; 花分生组织发育	AG
c70492_g1	Agamous-like MADS-box protein AGL12	花期转变	AGL12
c12164_g1	Agamous-like MADS-box protein AGL15	胚胎发育	AGL15
c41091_g4	MADS-box transcription factor 27	促进开花	AGL17
c39737_g1	MADS-box transcription factor 27	促进开花	AGL17
c36841_g1	Agamous-like MADS-box protein AGL21	促进开花	AGL17
c38538_g1	MADS-box transcription factor 27	促进开花	AGL17
c40334_g4	MADS-box transcription factor 27	促进开花	AGL17
c18716_g1	Agamous-like MADS-box protein AGL3	花被发育; 花分生组织发育	AGL2
c33061_g1	Agamous-like MADS-box protein AGL9 homolog	花被发育; 花分生组织发育	AGL2
c29354_g1	Floral homeotic protein DEFICIENS	花瓣和雄蕊发育	DEF
c32438_g1	Agamous-like MADS-box protein AGL31	花期调控; 成花抑制	FLC
c37758_g1	MADS-box protein FLOWERING LOCUS C	花期调控; 成花抑制	FLC
c50592_g1	Agamous-like MADS-box protein AGL8 homolog	萼片和花瓣发育; 花分生组织发育; 果实发育; 花期调控	SQUA
c10093_g1	Agamous-like MADS-box protein AGL8 homolog	萼片和花瓣发育; 花分生组织发育; 果实发育; 花期调控	SQUA
c9572_g1	Agamous-like MADS-box protein AGL8 homolog	萼片和花瓣发育; 花分生组织发育; 果实发育; 花期调控	SQUA
c2583_g1	Agamous-like MADS-box protein AGL8 homolog	萼片和花瓣发育; 花分生组织发育; 果实发育; 花期调控	SQUA
c37773_g1	MADS-box protein SVP	花期调控; 成花抑制和激活	SVP
c31351_g1	MADS-box protein SVP	花期调控; 成花抑制和激活	SVP
c30064_g1	MADS-box protein JOINTLESS	花期调控; 成花抑制和激活	SVP
c33061_g2	MADS-box protein SVP	花期调控; 成花抑制和激活	SVP
c73448_g1	MADS-box transcription factor 56	花期调控; 成花激活	TM3
c20770_g1	MADS-box transcription factor 50	花期调控; 成花激活	TM3
c5966_g1	MADS-box protein SOC1	花期调控; 成花激活	TM3

## 2.7 SSR 分析

利用 MISA 软件对云锦杜鹃转录组序列进行分析, 共发现 21 900 个 SSR 位点, 分布在 17 414 条单基因簇中, 其中有含有 1 个以上 SSR 位点的单基因簇有 3 606 条(表 5)。在所有 SSR 位点中, 双碱基重复 SSR 最多, 有 12 294 个, 占总数的 56.14%; 其次为单碱基重复 SSR, 有 6 448 个, 占总数的 29.44%; 三碱基重复 SSR 有 2 970 个, 占 13.56%; 四碱基重复 SSR 有 140 个, 占 0.63%; 五碱基和六碱基重复 SSR 分别仅有 25 个和 23 个(表 5)。进一步分析这些 SSR 重复基序, 可以发现, 在单碱基 SSR 中, A/T 发生频率最高; 双碱基 SSR 中, 发生频率最高的是 AG/CT, 其次是 AC/GT; 三碱基重复中发生频率最高是 AAG/CTT, 其次是 AGG/CCT(图 4)。

## 3 结论与讨论

中国是世界上杜鹃花属植物资源最为丰富的国家, 为世界杜鹃花育种做出了巨大贡献。但中国杜鹃花育种尤其常绿杜鹃育种水平较欧美、日本等国仍有较大差距, 资源开发利用水平低, 优良品种少。云锦杜鹃作为中国特有常绿杜鹃, 观赏价值高、抗性好, 野生资源也较为丰富, 具有良好的开发潜力。但

表5 云锦杜鹃转录组 SSR 分析结果

Table 5 Summary of SSR in *R. fortunei* transcriptome

搜索项目	数量/个
分析序列数量 total number of sequences examined	84 633
分析序列长度 total size of examined sequences (bp)	58 517 298
SSR 位点总数 total number of identified SSRs	21 900
包含 SSR 位点序列数 number of SSR containing sequences	17 414
包含 1 个以上 SSR 位点序列数 number of sequences containing more than 1 SSR	3 606
复合形式存在的 SSR 位点数 number of SSRs present in compound formation	1 437
单碱基重复 mono-nucleotide	6 448
二碱基重复 di-nucleotide	12 294
三碱基重复 tri-nucleotide	2 970
四碱基重复 tetra-nucleotide	140
五碱基重复 penta-nucleotide	25
六碱基重复 hexa-nucleotide	23

由于缺乏相关的遗传背景信息,云锦杜鹃的遗传多样性、杂交子代鉴定和优异基因型挖掘等遗传育种研究一直受到制约。近年来,随着高通量测序技术的发展,转录组测序已成为非模式生物遗传背景解析的重要手段,在标记开发、表达分析、功能基因挖掘等方面得到广泛应用<sup>[17-21]</sup>。本研究利用 Illumina 测序技术对云锦杜鹃组培苗的转录组进行测序和分析,以获得其转录组序列信息。云锦杜鹃转录组的测序数据分析结果表明:数据的 Q30 值为 91.85%,拼接后共获得 84 633 条单基因簇,平均长度为 691.4 bp, N50 值为 1 177 bp。一般认为 Q30 在 80% 以上就认为测序质量可靠;N50 值越大就表示长片段越多,且不小于 800 bp 就说明组装得到序列完整性较好<sup>[22]</sup>。上述结果表明本研究测序数据的质量和组装长度达到了转录组分析的基本要求,为进一步分析利用奠定了基础。

基因功能注释是转录组分析的重要内容,是进行重要功能基因挖掘的前提。因此,本研究利用 Nr 和 Swiss-prot 等七大数据库对云锦杜鹃转录组序列进行功能注释,结果表明:共有 35 526 条单基因簇获得注释信息,仍有约 5 万条序列没有获得注释。这与薏苡 *Coix lachryma-jobi*<sup>[20]</sup> 和岩穴蕨 *Monachosorum maximowiczii*<sup>[23]</sup> 的情况类似,可能是由于云锦杜鹃是未测序物种,在相关数据库中缺乏对应的功能注释信息,也可能是部分云锦杜鹃单基因簇序列本身太短造成的。GO 和 KOG 注释功能分类的结果显示,云锦杜鹃单基因簇的功能涉及了各类生命活动;KEGG 通路注释到 9 887 条单基因簇,涉及到 272 条代谢通路。这些结果表明:对于云锦杜鹃等这一非模式植物,转录组测序可以有效地解析遗传背景,获得大量序列信息。基因功能注释也是挖掘与特定途径或功能相关基因的有效手段。如在紫色黄秋葵 *Abelmoschus esculentus* 中,通过转录组的 KEGG 注释,获得与花色素苷、黄酮、类黄酮、二萜类和萜类骨架等生物合成相关的单基因簇<sup>[24]</sup>。本研究通过功能注释,也鉴定获得 24 个编码 MADS-box 基因的单基因簇,属于 10 个不同的亚家族,它们可能与花分生组织发育、花期调控、花器官发育等重要成花过程相关。

简单序列重复(SSR)又称微卫星序列,具有共显性、密度大、信息量丰富等优势,广泛应用于遗传图谱构建、遗传多样性分析、基因定位、分子标记辅助育种等方面<sup>[25]</sup>。利用转录组序列开发 SSR 标记具有通量高,成本低的优势,已在多种植物中获得成功<sup>[19,26-27]</sup>。在大王杜鹃 *R. rex* 转录组序列中鉴定获 15 314 个 SSR 位点,占比最高的为双碱基重复 SSR,其次为单碱基重复 SSR 和三碱基重复 SSR,且利用这些 SSR 位点开发了相应引物对 20 份大王杜鹃种质进行了遗传多样性评价<sup>[3]</sup>。本研究也在云锦杜鹃单

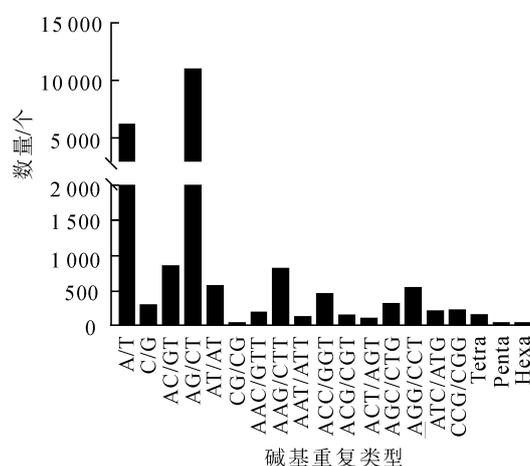


图4 云锦杜鹃 SSR 类型数量分布

Figure 4 Distribution of SSR motif number of *R. fortunei*

基因簇序列中鉴定获得 21 900 个 SSR 位点, 发现其中双核苷酸重复 SSR 最多, 达到 12 294 个; 其次为单核苷酸重复和三核苷酸重复 SSR, 这与大王杜鹃中发现的规律类似。这些结果将为云锦杜鹃 SSR 标记开发提供重要序列信息, 也为云锦杜鹃种质资源遗传多样性分析、功能基因挖掘以及分子辅助育种等工作提供了重要基础。

#### 4 参考文献

- [1] UNAMBA C I N, AKSHAY N, SHARMA R K. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants [J]. *Front Plant Sci*, 2015, **6**: 1074.
- [2] PARAMPREET K, KISHOR G. From genomes to GENE-omes: exome sequencing concept and applications in crop improvement [J]. *Front Plant Sci*, 2017, **8**: 2164.
- [3] ZHANG Yue, ZHANG Xue, WANG Yuehua, *et al.* De novo assembly of transcriptome and development of novel EST-SSR markers in *Rhododendron rex* Lévl. through Illumina Sequencing [J]. *Front Plant Sci*, 2017, **8**: 1664.
- [4] COSTA V, ANGELINI C, de FEIS I, *et al.* Uncovering the complexity of transcriptomes with RNA-Seq [J]. *J Biomed Biotechnol*, 2010: 853916. doi: 10.1155/2010/853916.
- [5] WANG Zhong, GERSTEIN M, SNYDER M. RNA-Seq: a revolutionary tool for transcriptomics [J]. *Nat Rev Genet*, 2009, **10**(1): 57 – 63.
- [6] WU Zhengyi, RAVEN P, HONG D Y. *Flora of China: Vol 14* [M]. Beijing: Science Press, 2005: 260 – 455.
- [7] 曾红, 钱慧琴, 梁兆昌, 等. 云锦杜鹃枝叶化学成分研究[J]. *中草药*, 2013, **44**(22): 3123 – 3126.  
ZENG Hong, QIAN Huiqin, LIANG Zhaochang, *et al.* Chemical constituents in twigs and leaves of *Rhododendron fortunei* [J]. *Chin Med Mat*, 2013, **44**(22): 3123 – 3126.
- [8] LAI Yongqi, ZENG Hong, HE Meijun, *et al.* 6,8-Di-C-methyl-flavonoids with neuroprotective activities from *Rhododendron fortunei* [J]. *Fitoterapia*, 2016, **112**: 237 – 243.
- [9] 徐晓锋, 潘德月, 陈江芳, 等. 浙江大盘山杜鹃花属植物资源及其保护与利用[J]. *中国野生植物资源*, 2017, **36**(1): 58 – 61.  
XU Xiaofeng, PAN Deyue, CHEN Jiangfang, *et al.* Study on exploitation and utilization of *Rhododendron* Linn. in Dapanshan, Zhejiang Province [J]. *Chin Wild Plant Resour*, 2017, **36**(1): 58 – 61.
- [10] 邓贤兰, 刘鹏, 吴杨, 等. 井冈山云锦杜鹃群落特征研究[J]. *亚热带植物科学*, 2011, **40**(4): 20 – 25.  
DENG Xianlan, LIU Peng, WU Yang, *et al.* Studies on the community characteristics of *Rhododendron fortunei* in Jinggang Mountain [J]. *Subtrop Plant Sci*, 2011, **40**(4): 20 – 25.
- [11] 王书胜, 单文, 张乐华, 等. 基质和 IBA 浓度对云锦杜鹃扦插生根的影响[J]. *林业科学*, 2015, **51**(9): 165 – 172.  
WANG Shusheng, SHAN Wen, ZHANG Lehua, *et al.* Effects of media and IBA concentrations on rooting of *Rhododendron fortunei* for cutting propagation [J]. *Sci Silv Sin*, 2015, **51**(9): 165 – 172.
- [12] 李钧敏, 金则新, 杨蓓芬. 云锦杜鹃总黄酮含量及成分分析[J]. *西北林学院学报*, 2004, **19**(1): 110 – 112.  
LI Junmin, JIN Zexin, YANG Beifen. The content and analysis of the components of flavonoids in *Rhododendron fortunei* [J]. *J Northwest For Univ*, 2004, **19**(1): 110 – 112.
- [13] 柯世省, 杨敏文. 水分胁迫对云锦杜鹃光合生理和光温响应的影响[J]. *园艺学报*, 2007, **34**(4): 959 – 964.  
KE Shisheng, YANG Minwen. Effects of water stress on photosynthetic physiological characteristics in leaves of *Rhododendron fortunei* and their response to light and temperature [J]. *Acta Horti Sin*, 2007, **34**(4): 959 – 964.
- [14] WEI Xiangying, CHEN Jianjun, ZHANG Chunying, *et al.* Differential gene expression in *Rhododendron fortunei* roots colonized by an ericoid mycorrhizal fungus and increased nitrogen absorption and plant growth [J]. *Front Plant Sci*, 2016, **7**: 1594.
- [15] GRABHEER M G, HAAS B J, YASSOUR M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biotechnol*, 2011, **29**: 644 – 652.
- [16] KRIZEK B A, FLETCHER J C. Molecular mechanisms of flower development: an armchair guide [J]. *Nat Rev Genet*, 2005, **6**(9): 688 – 698.
- [17] FENG Chao, CHEN Ming, XU Changjie, *et al.* Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq [J]. *BMC Genomics*, 2012, **13**(1): 19.

- [18] XU Wei, CUI Qinghua, LI Fei, *et al.* Transcriptome-wide identification and characterization of MicroRNAs from castor bean (*Ricinus communis* L.) [J]. *PLoS One*, 2013, **8**(7): e69995. doi: 10.1371/journal.pone.0069995.
- [19] CHEN Chunling, XU Meilong, WANG Wenwen, *et al.* Characterization of the *Lycium barbarum* fruit transcriptome and development of EST-SSR markers [J]. *PLoS One*, 2017, **12**(11): e0187738. doi: 10.1371/journal.pone.0187738.
- [20] 黄玉兰, 殷奎德, 向君亮. 蕹苡幼苗叶片转录组分析[J]. 农业生物技术学报, 2017, **25**(3): 386 – 396.  
HUANG Yulan, YIN Kuide, XIANG Junliang. Analysis transcriptome of coix (*Coix lachryma-jobi*) leaf at seedling stage [J]. *J Agric Biotech*, 2017, **25**(3): 386 – 396.
- [21] YANG Yanhui, LI Mingjie, YI Yanjie, *et al.* The root transcriptome of *Achyranthes bidentata* and the identification of the genes involved in the replanting benefit [J]. *Plant Cell Rep*, 2018, **37**(4): 611 – 625.
- [22] ZHANG Linsen, YANG Xiaoni, QI Xiangning, *et al.* Characterizing the transcriptome and microsatellite markers for almond (*Amygdalus communis* L.) using the Illumina sequencing platform [J]. *Hereditas*, 2018, **155**(1): 14.
- [23] 刘莉, 舒江平, 韦宏金, 等. 东亚特有珍稀蕨类植物岩穴蕨(碗蕨科)高通量转录组测序及分析[J]. 生物多样性, 2016, **24**(12): 1325 – 1334.  
LIU Li, SU Jiangping, WEI Hongjin, *et al.* De novo transcriptome analysis of the rare fern *Monachosorum maximowiczii* (Dennstaedtiaceae) endemic to East Asia [J]. *Biodivers Sci*, 2016, **24**(12): 1325 – 1334.
- [24] 张少平, 邱珊莲, 郑云云, 等. 紫色黄秋葵转录组功能基因测序及分析[J]. 核农学报, 2017, **31**(4): 643 – 653.  
ZHANG Shaoping, QIU Shanlian, ZHENG Yunyun, *et al.* The purple *Abelmoschus esculentus* transcriptome as a source for gene sequence information [J]. *Acta Agric Nucl Sin*, 2017, **31**(4): 643 – 653.
- [25] 罗冉, 吴委林, 张旻, 等. SSR 分子标记在作物遗传育种中的应用[J]. 基因组学与应用生物学, 2010, **29**(1): 137 – 143.  
RUO Ran, WU Weilin, ZHANG Yang, *et al.* SSR marker and its application to crop genetics and breeding [J]. *Genom Appl Biol*, 2010, **29**(1): 137 – 143.
- [26] CHEN Honglin, WANG Lixia, LIU Xiaoyan, *et al.* De novo transcriptomic analysis of cowpea (*Vigna unguiculata* L. Walp.) for genic SSR marker development [J]. *BMC Genet*, 2017, **18**(1): 65.
- [27] 张敏, 周彩虹, 陈焘, 等. 榿树转录组 SSR 信息分析及其分子标记开发[J]. 果树学报, 2017, **34**(10): 1258 – 1265.  
ZHANG Min, ZHOU Caihong, CHEN Tao, *et al.* Analysis of SSR information in transcriptome and development of molecular markers in *Torreya grandis* [J]. *J Fruit Sci*, 2017, **34**(10): 1258 – 1265.