

DOI:10.13275/j.cnki.lykxyj.2019.03.013

基于锦绣杜鹃花蕾转录组的 SSR 标记开发及应用

王书珍, 张羽佳, 黄诗颖, 罗炎炎, 金正强, 李志良, 金卫斌*

(大别山特色资源开发湖北省协同创新中心;黄冈师范学院 生命科学院,湖北 黄冈 438000)

摘要: [目的] 探究锦绣杜鹃 (*Rhododendron pulchrum* Planch.) EST-SSR 的类型、分布频率和分布特征, 开发有效的 SSR 标记, 并验证其在遗传多样性研究和跨物种转移中的应用潜力。 [方法] 采用 RNA-seq 技术对锦绣杜鹃‘紫鹤’品种的花蕾进行转录组测序, MISA 软件对组装的 unigenes 内部的 SSR 位点进行检索, 分析其类型、分布频率。利用 Primer 3.0 软件设计引物, 并对锦绣杜鹃群体和其近缘种映山红群体进行多样性检测, 利用 POP-GENE-PC 2.2 软件计算等位基因数、有效等位基因数、多态性位点百分率、Shannon's 信息指数、Nei 氏多样性指数、观察杂合度、期望杂合度等参数。 [结果] 锦绣杜鹃‘紫鹤’花蕾转录组共有 49 527 个 unigenes (43 766 249 bp), 从中筛选出 16 120 个 SSR 位点 (24.46%), 发生频率为 $1 \cdot (2.7 \text{ kb})^{-1}$ 。微卫星的重复次数主要集中在 5~24 次之间, 二核苷酸发生频率最高 (9 624 个, 59.70%), 其次是单核苷酸 (3 738 个, 23.19%), 频率最低的为五核苷酸 (42 个, 0.26%)。频率最高的重复基序有 A/T、AG/CT、AAG/CTT、AGG/CCT、ACC/GGT、AGC/GCT、AAAG/CTTT、AAGAG/CTCTT、AGAGGG/CCCTCT 等。选用 13 个多态性 SSR 标记对锦绣杜鹃群体进行 PCR 扩增, 共得到 71 个等位基因, 平均每个位点 3~9 个; 有效等位基因数在各 SSR 位点之间变化范围为 1.684~5.930; 观察杂合度 H_o 和期望杂合度 H_e 的变化范围分别为 0.000~1.000 和 0.433~0.848, 平均值分别为 0.696 ± 0.426 和 0.705 ± 0.129 ; Shannon's 信息指数 I 和 Nei 氏多样性指数 h 的变化范围分别为 0.736~1.961 和 0.406~0.831, 平均值分别为 1.376 ± 0.339 和 0.683 ± 0.131 。此 13 个标记在映山红群体中的跨物种扩增成功率为 100%, 并反映出丰富的遗传多样性。2 个群体间的遗传差异 93.4% 存在于群体内部。 [结论] 基于 AG/CT 重复基序开发的 13 个 SSR 标记具有高度的多态性, 为后续杜鹃花属植物的遗传多样性研究、遗传图谱构建、基因定位、分子标记辅助育种等研究奠定了基础。大别山野生映山红资源具有较高的遗传多样性, 且杂合度过剩, 遗传变异主要发生在群体内部。

关键词: 锦绣杜鹃; 转录组测序; 微卫星; 遗传多样性; 跨物种扩增

中图分类号: S718.46

文献标识码: A

文章编号: 1001-1498(2019)03-0097-08

杜鹃花属 (*Rhododendron*) 包含近 1 000 个物种, 是杜鹃花科 (Ericaceae) 最大的属, 北温带植物区系中的大属, 主产于东亚和东南亚, 是重要的园林绿化植物^[1-3]。锦绣杜鹃 (*Rhododendron pulchrum* Planch.) 是常用的栽培种常绿杜鹃, 品种繁多, 是杜鹃花园艺分类中毛鹃的代表, 优良的田园植物, 主要产于广东、广西、湖北、湖南、江西、福建、浙江、江苏等地^[4]。目前, 锦绣杜鹃中缺乏特异性微卫星标记和遗传连锁图谱, 其遗传改良和优良品种选育又是杜鹃花产业的重要任务, 因此, 筛选可靠的分子标记

评价种质资源、遗传多样性分析、分子标记辅助育种等对杜鹃花良种选育和种质资源保护意义重大。

简单重复序列 (SSR), 又称微卫星, 包括基因组 SSR 和表达序列标签 SSR (EST-SSR), 是 1~6 个核苷酸重复基元组成的短串联重复序列, 因其具有共显性、多态性强、重复性好、分布广泛等特点而在花卉遗传学中被广泛应用^[5-6]。近年来, 基于二代测序平台建立的转录组测序技术不仅能够挖掘特定组织在特定时期的表达基因, 也具有快速、高效、高通量开发分子标记的优点, 已经广泛用于萝卜 (*Rapha-*

收稿日期: 2018-03-08 修回日期: 2019-03-22

基金项目: 国家自然科学基金 (NSFC 31500995); 经济林木种质改良与资源综合利用湖北省重点实验室开发基金 (2017BX06)。

* 通讯作者: 金卫斌, 主要从事生态基因组研究。E-mail: 381355814@qq.com

nus sativus L.)、大白菜(*Brassica rapa* L. ssp. *pekinensis*)、康乃馨(*Dianthus caryophyllus* L.)、大蒜(*Allium sativum* L.)、砂梨(*Pyrus pyrifolia* Nakai.)等经济作物的 SSR 标记开发等研究^[7-11]。

杜鹃属现有的 SSR 标记非常有限,很难满足杜鹃花遗传多样性和群体结构研究、遗传改良及新品种选育的需要。因此,本研究对锦绣杜鹃花蕾组织进行转录组测序,基于组装的 unigenes 分析 SSR 的组成、频率和分布特征,开发 SSR 引物,同时对 SSR 标记的多态性和跨物种扩增效率进行评价。本研究基于微卫星标记特征,从批量开发合成的 856 对引物中筛选多态性强的 SSR 引物进行有效性和跨物种转移的验证,以期后续杜鹃花种质资源的遗传多样性研究、遗传图谱构建、基因定位、遗传改良、分子标记辅助新品种选育等研究奠定基础。

1 材料与方法

1.1 试验材料

2016 年 3 月初自黄冈师范学院植物资源圃采摘锦绣杜鹃的花蕾组织(表 1),液氮冻存后置于 -80 °C 冰箱,供后续总 RNA 提取和转录组测序。2016 年 5 月至 2016 年 9 月分别自黄冈市遗爱湖公

园和大别山主峰五脑山采集栽培的锦绣杜鹃群体和野生的映山红群体,摘取顶端健康叶片,用于基因组 DNA 的提取和遗传多样性检测(表 1)。样品采集按照种群遗传学的原理和方法,遵循代表性、可对照性原则取样(样本间距 1 ~ 1.96 m)。

1.2 总 RNA 提取与转录组测序

将 3 份锦绣杜鹃花蕾组织混合,采用高盐 Trizol 试剂和 RNA 结合柱的方法提取总 RNA,重复 3 次。采用链特异性纯化试剂盒对总 RNA 进行 mRNA Oligo(dT)磁珠富集、双链 cDNA 合成、末端修复、接头连接,再用 USER 酶对 cDNA 第二条链进行降解,保留真实转录的 mRNA 第一链,随后进行 PCR 扩增和 2% 琼脂糖凝胶电泳检测,回收 300 ~ 500 bp 的片段作为测序文库。采用 Agilent 2100 Bioanalyzer 和 ABI StepOnePlus Real-time PCR System 对文库进行质检,合格后使用 Illumina HiSeq 2500 PE125 平台进行测序(2 × 150 bp),数据过滤采用 FASTX (http://hannonlab.cshl.edu/fastx_toolkit/)和 CUTADAPT (<https://pypi.python.org/pypi/cutadapt/1.4.2>)软件,采用 Cufflinks 软件进行数据组装。本研究的转录组测序数据已经上传到 NCBI 数据库。

表 1 转录组测序和多样性检测材料采集地信息

Table 1 Information of sampling populations used for RNA-seq and genetic diversity analysis

种质 Germplasm	样品来源 Sample source	经度 Longitude (E)	纬度 Latitude (N)	海拔 Altitude/m	样品个数 Numbers of sample
锦绣杜鹃	黄冈师范学院资源圃	114°55'22.62"	30°26'56.82"	22	3
锦绣杜鹃	黄冈遗爱湖公园	114°53'28.21"	30°27'21.67"	21	30
映山红	大别山主峰五脑山	114°59'57.47"	31°14'10.95"	104	30

1.3 SSR 位点的检测、引物设计及筛选

利用 MISA (<http://pgrc.ipk-gatersleben.de/misa>)软件对组装的 unigene 序列内部的 SSR 位点进行检测。单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸最少重复次数分别设置为 10、7、5、5、5、5 次。对含有 SSR 位点且侧翼序列大于 150 bp 的 unigenes 进行筛选,用于 SSR 引物的设计。挑选 856 个富含 (AG/CT)_n 基序的 unigenes,采用 primer 3.0 软件 (<http://primer3.ut.ee/>)设计保守引物,SSR 位点距离侧翼序列长度在 50 ~ 300 bp 范围内,退火温度为 50 ~ 65 °C 之间,扩增产物大小为 80 ~ 300 bp,长度 18 ~ 28 bp,GC 含量 40% ~ 60%,并且避免产生引物二聚体以及错配。引物在 Unigene 库中进行 BLAST 比对验证,去除无效引物。挑

选 110 对引物委托南京金斯瑞生物科技有限公司合成。采用温度梯度 PCR 确定每个引物的最适合退火温度。随机选取 5 份锦绣杜鹃材料快速筛选核心引物。

1.4 SSR-PCR 扩增和遗传多样性分析

采用改良的 CTAB 法提取锦绣杜鹃和映山红 2 个群体的基因组 DNA^[12]。采用核心引物对栽培种锦绣杜鹃和野生种映山红 2 个群体进行 PCR 扩增,以验证引物的有效性并对 2 个居群的遗传多样性进行分析。选用 10 μL 的 PCR 扩增体系:10 × Taq Reaction Buffer 1 μL,2.5 mmol · L⁻¹ dNTPs 0.25 μL,5 U · μL⁻¹ Taq DNA Polymerase 0.15 μL,10 μmol · L⁻¹ 上下游引物各 0.15 μL,1 μL 模板 DNA (100 ng · μL⁻¹),用灭菌的去离子水补足剩余体积。PCR

扩增程序为:94℃预变性 10 min,35 个扩增循环(94℃变性 40 s,最适退火温度下退火 30 s,72℃延伸 40 s),然后 72℃延伸 7 min。

PCR 扩增产物用 6% 聚丙烯酰胺凝胶电泳分离,并用硝酸银染色法检测扩增产物。按照 PCR 扩增条带的电泳位置统计每个个体的基因型,构建基因型数据矩阵。采用 POPGENE-pc 2.2 软件统计各个群体内的等位基因数(N_a)、有效等位基因数(N_e)、Nei 氏多样性指数(h)、观察杂合度(H_o)、期望杂合度(H_e)、Shannon's 信息指数(I)、种群内近交系数(F_{is})、总近交系数(F_{it})、种群间遗传分化系数(F_{st})、基因流(Nm)等遗传参数^[13]。

2 结果与分析

2.1 锦绣杜鹃花蕾转录组中 SSR 的分布特点

锦绣杜鹃花蕾组织转录组测序序列总长度为 43 766 249 bp, Cufflinks 软件组装共得到 49 527 个 unigenes, 其中, 含有 SSR 位点的 unigene 序列有 12 113 个, 共发现 16 120 个 SSR 位点。平均每 2.7 kb 就有 1 个, SSR 位点的出现频率为 24.46%。仅含有 1 个 SSR 位点的 unigene 序列有 8 943 条, 占 73.83%; 含有 2 个或者 2 个以上 SSR 位点的 unigene 有 3 170 条, 占 26.17%。1 695 个 unigenes 中含有复合型 SSR 位点, 占总 unigenes 的 13.99%。

2.2 锦绣杜鹃花蕾转录组中 SSR 的类型和频率

锦绣杜鹃花蕾转录组内发现的 SSR 序列共有 6 种类型, 分别是单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸, 且 SSR 位点的个数大体上随着重复基序长度的增加而减少(表 2)。二核苷酸出现频率最大, 有 9 624 个(59.70%), 其次是单核苷酸(3 738 个, 23.19%)和三核苷酸(2 589 个, 16.06%)(表 2)。出现频率最低的为五核苷酸, 仅有 42 个, 所占比例为 0.26%; 其次是六核苷酸, 有 50 个, 所占比例为 0.31%。

表 2 不同 SSR 重复类型的个数

Table 2 Distribution of SSR repeat types

SSR 类型 Types	数量 Numbers	比例 Percent/%	基序类型 Types of motifs
单核苷酸/Mono-nucleotides	3 738	23.19	2
二核苷酸/Di-nucleotides	9 624	59.70	4
三核苷酸/Tri-nucleotides	2 589	16.06	10
四核苷酸/Tetra-nucleotides	77	0.48	18
五核苷酸/Penta-nucleotides	42	0.26	17
六核苷酸/Hexa-nucleotides	50	0.31	24

在杜鹃花的花蕾组织转录组序列中, 共发现 75 种 SSR 重复类型。单核苷酸有 A/T 和 G/C 两种, 主要类型是 A/T, 共有 3 669 个, 占总 SSR 重复类型的 22.761%(表 3)。二核苷酸有 4 种类型, 分别是 AG/CT(9 047 个, 56.123%)、AC/GT(343 个, 2.128%)、AT/AT(226 个, 1.402%)、CG/CG(8 个, 0.050%)。三核苷酸有 10 种类型, 出现最多的是 AAG/CTT, 其次是 AGG/CCT、ACC/GGT、AGC/GCT。四核苷酸有 18 种, 主要为 AAAG/CTTT(23 个, 0.143%)、ACAT/ATGT(9 个, 0.056%)、AGCG/CGCT(6 个, 0.037%)、ATCC/GGAT(5 个, 0.031%)。五核苷酸有 17 种类型, 其中, AAGAG/CTCTT 最多(7 个, 0.043%), 其次是 AAAAG/CTTTT(6 个, 0.037%)和 AGAGG/CCTCT(6 个, 0.037%)。六核苷酸有 24 种, 主要是 AGAGGG/CCCTCT(8 个, 0.050%), 其次是 AAGAGG/CCTCTT(5 个, 0.031%)和 AGCAGG/CCTGCT(5 个, 0.031%)。

表 3 主要 SSR 重复类型的百分比含量

Table 3 Percent of the main SSR repeat types

重复类型 Repeats	数量 Numbers	比例 Percent/%
A/T	3 669	22.761
AG/CT	9 047	56.123
AAG/CTT	758	4.702
AGG/CCT	440	2.730
ACC/GGT	398	2.469
AGC/GCT	223	1.383
AAAG/CTTT	23	0.143
AAGAG/CTCTT	7	0.043
AGAGGG/CCCTCT	8	0.050

6 种不同长度类型的 SSR 位点个数均随重复次数的增加而减少(表 4)。单核苷酸重复次数最多的为 9~12 次(1 996 个, 53.4%), 其次是 13~16 次(864 个, 23.11%), 大于 24 次重复的个数仅仅含有 179 个(4.79%)。二核苷酸重复次数最多的为 5~8 次(3 425 个, 35.59%), 其次是 9~12 次(2 737 个, 28.44%), 仅 210 个重复次数大于 24(2.18%)。三核苷酸的重复次数主要集中在 5~8 次(2 484 个, 95.94%), 重复次数在 9~12 次的有 89 个, 在 13~16 次的有 14 个, 而重复次数在 17~20 和 21~24 次的各有 1 个(0.04%), 没有重复次数大于 24 的三核苷酸重复基序。四核苷酸和六核苷酸的重复次数均集中在 5~8 次。五核苷酸的重复次数主要在 5~8 次(95.24%), 重复次数在 9~12 和 13~16 次的各有 1 个。

表4 不同 SSR 重复类型的分布频率

Table 4 Frequency distribution of different SSR repeat types

SSR 类型 Types	5~8	9~12	13~16	17~20	21~24	>24
单核苷酸/Mono-nucleotides	0	1 996	864	353	346	179
二核苷酸/Di-nucleotides	3 425	2 737	1 548	1 250	454	210
三核苷酸/Tri-nucleotides	2 484	89	14	1	1	0
四核苷酸/Tetra-nucleotides	77	0	0	0	0	0
五核苷酸/Penta-nucleotides	40	1	1	0	0	0
六核苷酸/Hexa-nucleotides	50	0	0	0	0	0

2.3 EST-SSR 引物筛选与有效性检测

选取 856 个 SSR 位点长度大于 20 bp 且两端侧翼序列长度均大于 150 bp 的 unigenes 设计引物, 占含 SSR 位点 unigenes 总序列的 8.62%。经 BLAST 比对后, 去除无效引物 115 对。在有效引物中随机挑选 110 对进行合成, 选用转录组测序用的锦绣杜鹃嫩叶基因组 DNA 进行 PCR 扩增验证, 产物经 3% 的琼脂糖凝胶电泳检测。共筛选出 68 对 SSR 引物

能够扩增出与预期大小一致, 且条带清晰的目的 DNA 产物, 占总挑选引物数的 61.82%。设置温度梯度 PCR 程序, 对 68 对 SSR 引物的最适合退火温度进行筛选。选取 5 份锦绣杜鹃种质基因组 DNA 对此 68 对引物进行核心引物的筛选, 获得多态性强的引物 40 对, 选用 13 对多态性最强、无明显影子带的引物进行后续的群体遗传多样性的检测(表 5)。

表5 多态性 SSR 引物序列及基序信息

Table 5 Characteristic of Polymorphic SSR primer pairs

引物 Locus	上游引物序列(5'-3') Forward primer sequence	下游引物序列(5'-3') Reverse primer sequence	基序 motif	片段大小 Size range/bp	退火温度/°C Annealing Temperature
Rp01	CTTGCCACTTTGAGTTTGAG	AGAGTAATTTGGAGGAAGCG	(CT)28	209~221	58
Rp02	TGAACCCCTCTTCTTCTTCC	TTTGATTGAAGGGTGGAGTG	(TC)23	250~258	60
Rp03	CTCTCTCTCTCCTTCCTTCA	GGATTCTTACTCGTGTCTGG	(CT)27	217~229	58
Rp04	GCAGCACACGGATATTTAAG	CTCAATACCACACTACACCC	(CT)29	201~209	60
Rp05	TAGCTGCTTACTGTTGAAGG	AGTAAAAGGGCTGAAACTGT	(AG)22	150~160	56
Rp06	TCTTCTTCTTCTTCTTCTCC	GGGGAAGGAGAAGAGAAAG	(CT)25	179~195	58
Rp07	GCCCTATCCCTCAACTTTAC	GAGGAGCGTGGTTAGTAATT	(TC)21	230~240	58
Rp08	GTATGGGACCTGTGATTTCC	CTCCAAGTACTACTCCAAC	(GA)24	229~239	58
Rp09	GAAATCTCGAATCACCTCCA	AAGGTGTTGGTGGACTAATC	(TC)21	150~166	58
Rp10	TTGAAGAACACTCAAGTTGC	ACGTAGAACATGCTTTCCT	(GA)21	187~203	60
Rp11	CCCTTCTCTTCTCAAAATCC	CGTCAATTTACACACAGAG	(CT)20	174~190	62
Rp12	CTCTCCAAAATGACCGAT	GAAITGGCTGTTGGATGATG	(CT)21	234~250	60
Rp13	AGAAAAGTGGGAGATGTCTC	AGGTGATCATCTTTCATGT	(CT)21	245~258	56

2.4 锦绣杜鹃和映山红两个群体多样性分析

选用的 13 个 SSR 标记在 2 个群体内扩增的 DNA 片段范围为 150~258 bp, 在栽培种锦绣杜鹃和野生映山红群体中分别扩增出 71 个和 74 个等位基因, 平均每个 SSR 位点扩增出的等位基因数为 5.462 和 5.692 个(表 6)。2 个种群内近交系数 F_{is} 变化范围为 -0.515~1.000, 平均值为 0.014。总近交系数 F_{it} 变化范围为 -0.315~1.000, 平均值为 0.080。遗传分化系数 F_{st} 变化范围为 0.010~0.159, 平均值为 0.066, 即 6.6% 的遗传变异发生在群体间, 发生在群体内的遗传变异高达 93.4%。基因流 N_m 变化范围为 1.324~24.762, 平均值为 3.514 (表 6)。

表6 13 个微卫星位点的遗传分化和基因流信息

Table 6 Summary of genetic variation and gene flow

位点 Locus	种群内近交系数 F_{is}	总近交系数 F_{it}	分化系数 F_{st}	基因流 N_m
Rp01	1.000	1.000	0.141	1.519
Rp02	-0.515	-0.297	0.144	1.491
Rp03	1.000	1.000	0.067	3.457
Rp04	-0.285	-0.265	0.015	16.032
Rp05	-0.292	-0.226	0.050	4.708
Rp06	-0.232	-0.212	0.016	15.044
Rp07	-0.382	-0.289	0.067	3.478
Rp08	-0.216	-0.180	0.030	8.128
Rp09	0.286	0.362	0.107	2.098
Rp10	0.877	0.881	0.032	7.512
Rp11	-0.328	-0.315	0.010	24.762
Rp12	0.016	0.172	0.159	1.324
Rp13	-0.223	-0.191	0.027	9.169
平均 Mean	0.014	0.080	0.066	3.514

锦绣杜鹃群体中,等位基因数 N_a 在各 SSR 位点之间变化范围为 3~9,有效等位基因数变化范围为 1.684~5.930, Rp03 标记扩增的等位基因最少(3个),Rp08 扩增的等位基因最多(9个)。观察杂合度 H_o 和期望杂合度 H_E 的变化范围分别为 0.000~1.000 和 0.433~0.848,平均值分别为 0.696 ± 0.426 和 0.705 ± 0.129 。Shannon's 信息指数 I 和 Nei 氏多样性指数 h 的变化范围分别为 0.736~1.961 和 0.406~0.831,平均值分别为 1.376 ± 0.339 和 0.683 ± 0.131 。Shannon's 信息指数 I 、期望杂合度 H_E 、Nei 氏多样性指数 h 呈现相似的变化趋势,最大值均出现在 Rp08 处,最小值在位点 Rp03

处(表7)。

在映山红群体内,各 SSR 位点有效等位基因数 N_e 变化范围为 1.667~6.922,平均值为 4.241。观察杂合度 H_o 和期望杂合度 H_E 的变化范围分别为 0.000~1.000 和 0.408~0.871,平均值分别为 0.699 和 0.747。Shannon's 信息指数 I 和 Nei 氏多样性指数 h 的变化范围分别为 0.794~2.049 和 0.400~0.856,平均值分别为 1.508 和 0.730。与锦绣杜鹃类似,Shannon's 信息指数 I 、期望杂合度、Nei 氏多样性指数也呈现相似的变化趋势,然而其最大值出现在 Rp06 处,最小值在 Rp09 处(表7)。

表7 SSR引物在2个群体中的遗传多样性

Table 7 Genetic diversity of two populations revealed by SSR markers

位点 Locus	锦绣杜鹃(<i>R. pulchrum</i>)						映山红(<i>R. simsii</i>)					
	N_a	N_e	H_o	H_E	I	h	N_a	N_e	H_o	H_E	I	h
Rp01	4	2.909	0.000	0.700	1.213	0.656	4	3.922	0.000	0.764	1.376	0.745
Rp02	6	2.586	1.000	0.633	1.157	0.613	5	3.412	1.000	0.719	1.361	0.707
Rp03	3	1.684	0.000	0.433	0.736	0.406	5	3.600	0.000	0.754	1.424	0.722
Rp04	4	3.982	1.000	0.775	1.384	0.749	7	5.200	1.000	0.824	1.777	0.808
Rp05	5	3.193	0.882	0.708	1.273	0.687	5	4.360	1.000	0.789	1.533	0.771
Rp06	8	4.308	1.000	0.782	1.718	0.768	9	6.922	1.000	0.871	2.049	0.856
Rp07	5	3.767	1.000	0.752	1.429	0.735	4	3.485	1.000	0.742	1.308	0.713
Rp08	9	5.930	1.000	0.848	1.961	0.831	7	5.365	1.000	0.830	1.775	0.814
Rp09	6	4.624	0.765	0.808	1.633	0.784	4	1.667	0.080	0.408	0.794	0.400
Rp10	4	2.579	0.071	0.623	1.091	0.612	4	2.344	0.074	0.584	1.008	0.573
Rp11	5	4.000	1.000	0.767	1.485	0.750	6	4.093	1.000	0.769	1.576	0.756
Rp12	6	1.862	0.333	0.490	1.038	0.463	7	5.521	0.929	0.834	1.825	0.819
Rp13	6	5.729	1.000	0.842	1.767	0.825	7	5.244	1.000	0.824	1.800	0.809
平均值 Mean	5.462	3.627	0.696	0.705	1.376	0.683	5.692	4.241	0.699	0.747	1.508	0.730
标准差	1.664	1.330	0.426	0.129	0.339	0.131	1.600	1.416	0.459	0.125	0.350	0.122

3 讨论

SSR 分子标记已经广泛应用于物种的遗传变异分析、物种起源进化、基因分型、指纹鉴定、遗传图谱构建、法医科学和动植物遗传育种等研究。目前,从杜鹃属植物已经成功开发出一些有用的 SSR 标记,但仍不能满足杜鹃属物种遗传学研究的需要^[14-18]。传统开发 SSR 标记的方法包括构建基因组文库进行 DNA 杂交鉴定并测序、富集微卫星分离法、基于 PCR 的分离法和同源序列鉴定,但都只能获得少量的有效标记^[14]。海量研究工作表明,通过转录组测序可以挖掘和开发出丰富的 SSR 标记^[7,20]。

本研究经转录组测序组装得到 49 527 个 unigenes,其中,24.46% 的 unigenes 含有 SSR 位点,高于菜薹(19.26%)、萝卜(23.79%)、绿豆(20.4%)、

大白菜(16.6%)、玉米(1.5%)等物种内的比例,但是低于水稻内 EST-SSR 的比例(27%)^[8,21-24]。锦绣杜鹃花蕾转录组平均每 2.7 kb 就有 1 个 SSR 位点,高于魔芋($1 \cdot (3.63 \text{ kb})^{-1}$)、水稻($1 \cdot (3.4 \text{ kb})^{-1}$)、小麦($1 \cdot (5.4 \text{ kb})^{-1}$)、大豆($1 \cdot (7.4 \text{ kb})^{-1}$)、拟南芥($1 \cdot (14 \text{ kb})^{-1}$)、棉花($1 \cdot (20 \text{ kb})^{-1}$)^[25-27]。造成 SSR 出现频率不同的原因不仅与物种的差异有关,也与转录组测序得到的 unigene 长度和数量、数据组装和分析采用的软件、SSR 位点筛选的条件有关。锦绣杜鹃花蕾转录组中,二核苷酸重复基序出现频率最大,达到 59.70%,此结果与魔芋、云杉、西葫芦、苦瓜、木豆等的研究一致^[2,20,25,28-29]。(AG/CT)_n 重复基序是锦绣杜鹃主要的微卫星类型,占总搜索到的 SSR 位点的 56.123%,与多种植物的研究结果一致^[2,20,28-29]。

为了开发多态性强的 SSR 标记,笔者从 856 个适合 SSR 引物设计的序列中,随机挑选 110 个含有 (AG/CT)_n 基序且重复次数为大于 19 次的 unigenes 进行引物设计与合成,其中,68 对引物能够扩增出与预期大小一致,且条带清晰的目的 DNA 产物。本研究结果进一步支持了多态性标记与 SSR 位点中重复序列的长短和重复次数正相关的结论^[30]。本研究所合成的 20 对引物在基因组 DNA 内扩增出来的产物远大于预期片段大小,可能的原因是基因组 DNA 序列内存在较大的内含子和无效等位基因,这些引物位点也可能经历了 RNA 可变剪接。22 个标记始终没有扩增产物出现,可能的原因是 cDNA contig 序列组装时误差造成的。40 对有效引物具有高度的多态性,占 68 对有效引物的 58.82%,该扩增比例高于绿豆(33%,31 份资源)、菜薹(40%,4 份资源)和萝卜(58.2%,32 份资源),但是远远低于在大白菜内的扩增效率(70.9%,24 份资源)^[7-8,21-22,31]。

Takezaki 等^[32]认为,SSR 位点的期望杂合度在 0.3~0.8 的范围内则种群具有较高的遗传多样性,本研究的 2 个群体期望杂合度平均值分别为 0.705 和 0.747,即群体内均有丰富的遗传多样性。映山红和锦绣杜鹃属于杜鹃属的 2 个近缘种,本研究从锦绣杜鹃中开发的 SSR 引物在映山红中的跨种扩增成功率高达 100%,表明此近缘种间亲缘关系较近(遗传距离仅为 0.3806)。然而,野生映山红群体的等位基因数、有效等位基因数、观察杂合度、期望杂合度、Shannon's 信息指数 *I*、Nei 氏多样性指数都比栽培的锦绣杜鹃群体高,即野生映山红群体遗传多样性更高。杜鹃花野生种群主要以种子繁殖为主,种群内存在着新生的种子苗,野外观察中发现蜜蜂等昆虫的授粉行为,蜜蜂的长距离飞行极大促进了基因交流,从而使得野生映山红群体在自然选择中保留了更加丰富的遗传多样性。栽培锦绣杜鹃种群根系发达、分株易成活,主要以无性繁殖为主,在长期驯化和人工选育过程中遗传变异下降降低,可能是其遗传多样性相对较低的主要原因。李乃伟等^[33]对南方红豆杉的多样性研究中,也发现野生种群的遗传多样性($h = 0.2058$, $I = 0.3259$)略高于迁地保护的栽培种群($h = 0.1807$, $I = 0.2750$)。潘庆杰等^[34]发现,地涌金莲野生型种群的遗传多样性($h = 0.1598$, $I = 0.2402$, 平均多态位点比率为 49.97%)显著高于栽培种群($h = 0.0898$, $I = 0.1351$, 平均多态位点比率为 27.09%)。王雪凤

等^[35]研究发现,9 个栽培蒙古黄芪种群多样性(基因多样性 $H_T = 0.314$, 种群内的基因多样性 $H_S = 0.235$) 低于 6 个野生黄芪种群($H_T = 0.383$, $H_S = 0.240$)。依据 Nagylaki^[36]的理论, F_{IS} 和 F_{IT} 平均值均为正数,表明 2 个群体在检测的 SSR 位点杂交现象不严重,尤其是 Rp01、Rp03、Rp09、Rp10、Rp12 等位点存在着明显的近交现象。基因流是影响种群分化的重要因素^[37],本研究检测的 13 个位点基因流为 3.514,足以抵抗遗传漂变造成的遗传变异。

本研究对锦绣杜鹃的转录组测序和数据分析,不仅丰富了现有的 SSR 标记数据库资源,也将有助于后续野生杜鹃花资源的遗传多样性研究、遗传图谱构建、基因定位、分子标记辅助育种、转录组的比较分析等研究。基于栽培品种与野生资源的遗传变异,大别山野生杜鹃花资源可以作为新品种选育的优良材料,更加利于拓宽现有杜鹃栽培品种的遗传背景。

4 结论

锦绣杜鹃‘紫鹤’花蕾内微卫星位点的发生频率为 $1 \cdot (2.7 \text{ kb})^{-1}$,微卫星的重复次数主要集中在 5~24 次,二核苷酸发生频率最高,其次是单核苷酸,频率最低的为五核苷酸。频率最高的重复基序有 A/T、AG/CT、AAG/CTT、AGG/CCT、ACC/GGT、AGC/GCT、AAAG/CTTT、AAGAG/CTCTT 和 AGAGGG/CCCTCT 等。设计合成的 13 个 SSR 标记多态性较强,在映山红群体中的跨物种扩增成功率为 100%。本研究丰富了杜鹃属现有的 SSR 标记数据库资源,也为后续杜鹃花属植物的遗传改良、遗传图谱构建、目标性状基因定位、分子标记辅助育种等研究奠定了基础。

参考文献:

- [1] 赵健,仇硕,李秀娟,等.不同激素对锦绣杜鹃的催花作用[J].广西植物,2009,29(1):92-95.
- [2] Wang S, Pan L, Hu K, et al. Development and characterization of polymorphic microsatellite markers in *Momordica charantia* (Cucurbitaceae) [J]. American Journal of Botany, 2010, 97(8): e75-e78.
- [3] 赵喜华,张乐华,王曼莹.11种杜鹃花 RAPD 分类学初步研究[J].江西农业大学学报,2006,28(4):544-547.
- [4] 唐源江,武晓燕.锦绣杜鹃盛花期叶的光合特性研究[J].北方园艺,2010(21):68-71.
- [5] Russell J R, Fuller J D, Macaulay M, et al. Direct comparison of levels of genetic variation among barley accessions detected by

- RFLPs, AFLPs, SSRs and RAPDs [J]. *Theoretical and Applied Genetics*, 1997, 95(4): 714–722.
- [6] Liao Y Y, Yue X L, Gituru WR, *et al.* Genotypic diversity and genetic structure of populations of the distylous aquatic plant *Nymphaoides peltata* (Menyanthaceae) in China [J]. *Journal of Systematics and Evolution*, 2013, 51(5): 536–544.
- [7] Zhai L, Xu L, Wang Y, *et al.* Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (*Raphanus sativus* L.) [J]. *Molecular Breeding*, 2014, 33(3): 611–624.
- [8] Ding Q, Li J, Wang F, *et al.* Characterization and development of EST-SSRs by deep transcriptome sequencing in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) [J]. *International Journal of Genomics*, 2015, 9: 1–11.
- [9] Yagi M, Yamamoto T, Isobe S, *et al.* Construction of a reference genetic linkage map for carnation (*Dianthus caryophyllus* L.) [J]. *BMC Genomics*, 2013, 14(1): 734.
- [10] Liu T M, Zeng L B, Zhu S Y, *et al.* Large-scale development of expressed sequence tag-derived simple sequence repeat markers by deep transcriptome sequencing in garlic (*Allium sativum* L.) [J]. *Molecular Breeding*, 2015, 35(11): 1–9.
- [11] Zhou H, Cai B H, Lü Z Q, *et al.* Development, characterization, and annotation of potential simple sequence repeats by transcriptome sequencing in pears (*Pyrus pyrifolia* Nakai) [J]. *Genetics & Molecular Research*, 2016, 15(3): gmr.15038683
- [12] 陈昆松, 李方, 徐昌杰, 等. 改良 CTAB 法用于多年生植物组织基因组 DNA 的大量提取 [J]. *遗传*, 2004, 26(4): 529–531.
- [13] Hu J, Pan L, Liu H, *et al.* Comparative analysis of genetic diversity in sacred lotus (*Nelumbo nucifera* Gaertn.) using AFLP and SSR markers [J]. *Molecular Biology Reports*, 2012, 39(4): 3637–3647.
- [14] Delmas C E, Lhuillier E, Pornon A, *et al.* Isolation and characterization of microsatellite loci in *Rhododendron ferrugineum* (Ericaceae) using pyrosequencing technology [J]. *American Journal of Botany*, 2011, 98(5): e120–e122.
- [15] Kameyama Y, Isagi Y, Nakagoshi N. Relatedness structure in *Rhododendron metternichii* var. *hondoense* revealed by microsatellite analysis [J]. *Molecular Ecology*, 2002, 11(3): 519–527.
- [16] Tan X X, Li Y, Ge X J. Development and characterization of eight polymorphic microsatellites for *Rhododendron simsii* Planch (Ericaceae) [J]. *Conservation Genetics*, 2009, 9(1): 326–329.
- [17] Wang X, Huang Y, Long C. Isolation and characterization of twenty-four microsatellite loci for *Rhododendron decorum* Franch. (Ericaceae) [J]. *HortScience*, 2009, 44(7): 2028–2030.
- [18] Hsieh Y C, Chung J D, Wang C N, *et al.* Historical connectivity, contemporary isolation and local adaptation in a widespread but discontinuously distributed species endemic to Taiwan, *Rhododendron oldhamii* (Ericaceae) [J]. *Heredity*, 2013, 111(2): 147–156.
- [19] 张涛, 路宏朝, 李新生. 微卫星进化属性及其功能的研究进展 [J]. *中国农学通报*, 2010, 26(1): 44–47.
- [20] Dutta S, Kumawat G, Singh B P, *et al.* Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea (*Cajanus cajan* (L.) Millspaugh) [J]. *BMC Plant Biology*, 2011, 11(1): 17.
- [21] 李荣华, 王直亮, 陈静芳, 等. 菜薹转录组中 SSR 信息与可用性分析 [J]. *园艺学报*, 2016, 43(9): 1816–1824.
- [22] Chen H, Wang L, Wang S, *et al.* Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers [J]. *PLoS ONE*, 2015, 10(4): e0120273.
- [23] Kantety R V, La Rota M, Matthews D E, *et al.* Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. *Plant Molecular Biology*, 2002, 48(5–6): 501–510.
- [24] 江东, 钟广炎, 洪棋斌. 柑橘 EST-SSR 分子标记分析 [J]. *遗传学报*, 2006, 33(4): 345–353.
- [25] Zheng X, Cheng P, Ying D, *et al.* Development of microsatellite markers by transcriptome sequencing in two species of amorphophallus (Araceae) [J]. *BMC Genomics*, 2013, 14(2): 414–416.
- [26] Peng J, Lapitan N L V. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers [J]. *Functional & Integrative Genomics*, 2005, 5(2): 80–96.
- [27] Varshney R K, Thiel T, Stein N, *et al.* In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species [J]. *Cellular & Molecular Biology Letters*, 2002, 7(2A): 537–546.
- [28] Rungis D, Berube Y, Zhang J, *et al.* Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags [J]. *Theoretical & Applied Genetics*, 2004, 109(6): 1283–1294.
- [29] Gong L, Stift G, Kofler R, *et al.* Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. [J]. *Theoretical & Applied Genetics*, 2008, 117(1): 37–48.
- [30] Singh H, Deshmukh R K, Singh A, *et al.* Highly variable SSR markers suitable for rice genotyping using agarose gels [J]. *Molecular Breeding*, 2010, 25(2): 359–364.
- [31] Wang S, Wang X, He Q, *et al.* Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish [J]. *Plant Cell Report*, 2012, 31(8): 1437–1447.
- [32] Takezaki N, Nei M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA [J]. *Genetics*, 1996, 144(1): 389–399
- [33] 李乃伟, 贺善安, 束晓春, 等. 基于 ISSR 标记的南方红豆杉野生种群和迁地保护种群的遗传多样性和遗传结构分析 [J]. *植物资源与环境学报*, 2011, 20(1): 25–30.
- [34] 潘庆杰, 李正红, 王雁, 等. 地涌金莲野生与栽培种群遗传多样性 RAPD 分析 [J]. *林业科学研究*, 2007, 20(5): 668–672.
- [35] 王雪凤, 梁颖, 刘德旺, 等. 内蒙古地区栽培及野生蒙古黄芪遗传多样性研究 [J]. *中国草地学报*, 2018, 40(1): 42–28.
- [36] Nagylaki T. Fixation indices in subdivided populations [J]. *Genetics*, 1998, 148(3): 1325–1332.
- [37] Whitlock M C, McCauley D E. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$ [J]. *Heredity*, 1999, 82(2): 117–125.

Development and Application of SSR Markers Based on Buds Transcriptomic Data of *Rhododendron pulchrum* Planch.

WANG Shu-zhen, ZHANG Yu-jia, HUANG Shi-ying, LUO Yan-yan, JIN Zheng-qiang, LI Zhi-liang, JIN Wei-bin

(Hubei Collaborative Innovation Center for the Characteristic Resources Exploitation of Dabie Mountains; College of Life Science,

Huanggang Normal University, Huanggang 438000, Hubei, China)

Abstract: [**Objective**] The types, distribution frequencies and characteristics of expressed sequence tag-simple sequence repeats (EST-SSRs) were clarified through transcriptome analysis of buds of *Rhododendron pulchrum* Planch (“Zihe” variety). Moreover, the development of polymorphic SSR markers were performed, as well as their usage in genetic diversity analysis and potential for cross-amplification in related species. [**Method**] RNA-seq data from buds of *R. pulchrum* were searched with MISA software. The characteristics of SSR loci were analyzed and SSR primers were designed with Primer 3.0. The effective primer pairs were used in genetic diversity analysis of *R. pulchrum* and *R. simsii* populations. The genetic parameters of population were calculated with POPGENE-pc 2.2 software. [**Result**] Totally, 49 527 unigenes (43 766 249 bp) were obtained, and 16 120 SSR loci 1. (2.7 kb)⁻¹ were searched, which accounted for 24.46% of the total unigenes. Repeat numbers of most SSR loci ranged between 5-24. Dinucleotide repeat was the most abundant type with a frequency of 59.70% (9 624), followed by mono-nucleotide repeat (3 738, 23.19%), and the least type was penta-nucleotide repeat (42, 0.26%). Moreover, the typical motifs were A/T, AG/CT, AAG/CTT, AGG/CCT, ACC/GGT, AGC/GCT, AAAG/CTTT, AAGAG/CTCTT, and AGAGGG/CCCTCT. The availability and polymorphism of the thirteen SSR markers selected were clarified in *R. pulchrum* population. A total of 71 alleles was scored, and the amount of allele (N_a) and effective number of allele (N_e) per locus ranged from 3-9 and 1.684 – 5.930, respectively. The observed heterozygosity (H_o) and expected heterozygosity (H_e) varied from 0.000 to 1.000 and from 0.433 to 0.848, with the mean values of 0.696 ± 0.426 and 0.705 ± 0.129 , respectively. Shannon’s information index (I) and Nei’s gene diversity (h) ranged from 0.736 to 1.961 and from 0.406 to 0.831, with the average values of 1.376 ± 0.339 and 0.683 ± 0.131 , respectively. All these SSR markers could be successfully cross-amplified in the related species *R. simsii*, which also showed high genetic diversity. The genetic variation existed mainly among populations. [**Conclusion**] The 13 polymorphic SSR markers based on unigenes containing (AG/CT)_n loci will benefit for following genetic diversity analysis, genetic map construction, gene mapping, and molecular marker assisted breeding of *Rhododendron* species. Natural *R. simsii* germplasm resources possess high genetic diversity. Moreover, excess heterozygosity is observed, and genetic variation is mainly maintained among populations.

Keywords: *Rhododendron pulchrum* Planch; RNA-seq; microsatellite; genetic diversity; cross-amplification

(责任编辑:张 研)