

文章编号:1001-1498(2011)02-0171-05

## 基于牡丹 EST 信息的滇牡丹 SSR 标记开发

张艳丽<sup>1</sup>, 王 雁<sup>2</sup>, 李正红<sup>1\*</sup>, 马 宏<sup>1</sup>

(1. 中国林业科学研究院资源昆虫研究所, 国家林业局资源昆虫培育与利用重点实验室, 云南 昆明 650224;

2. 中国林业科学研究院林业研究所, 北京 100091)

**摘要:**对 NCBI(美国国立生物技术信息中心)中牡丹的 ESTs(expressed sequence tags)的序列进行分析,结果表明:在所分析的 2 204 条序列中,仅 324 条分布有 SSRs(simple sequence repeats),占全部 ESTs 序列的 14.70%;SSR 的出现频率为 15.20%,共计 335 个,其中,二核苷酸重复比例 84.18%,三核苷酸重复比例为 15.22%,四和六核苷酸重复比例为 0.30%。在此基础上,利用软件(serafer 1.3)设计了 51 对备选 SSR 引物,以 6 个滇牡丹不同花色类群 DNA 为模板对引物进行筛选,其中,10 对引物有扩增产物;用这些引物进一步在 10 个类群 50 个 DNA 模板进行多态性测试,结果显示:上述 10 对 SSR 引物均有多态性,且同一类群内不同模板 DNA 间也存在多态性。本研究结果证明:基于牡丹 EST 信息建立 SSR 标记是一种有效而可行的方法,有助于滇牡丹遗传多样性分析及基因组学方面的研究。

**关键词:**牡丹;滇牡丹;EST;SSR;标记

中图分类号:S685.11

文献标识码:A

### SSRs Marking of *Paeonia delavayi* Based on Peony EST Data

ZHANG Yan-li<sup>1</sup>, WANG Yan<sup>2</sup>, LI Zheng-hong<sup>1</sup>, MA Hong<sup>1</sup>

(1. Research Institute of Resource Insects, Chinese Academy of Forestry, Key Laboratory of Breeding and Utilization of Resource Insects of State Forestry Administration, Kunming 650224, Yunnan, China;

2. Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China)

**Abstract:** Through sequences analysis of *Paeonia suffruticosa* ESTs (expressed sequence tags) deposited in NCBI, only 335 SSRs (simple sequence repeats) were found. These SSRs are distributed in 324 ESTs out of the 2 204 ESTs examined, accounting for 14.70% of the total. The frequency of SSR was 15.20%, in which, the repetitive proportions of dinucleotide, trinucleotide, tetranucleotide and hexanucleotide were 84.14%, 15.22%, 0.30% and 0.30%, respectively. 51 candidate SSR primers were designed according to the ESTs sequences that containing SSRs, by using serafer 1.3 software. The 51 candidate SSR primers were screened against genomic DNA of 6 different flower color groups in *P. delavayi*, of which, 10 primer pairs amplified visible bands with polymorphism among 50 genomic DNAs from 10 color groups in *P. delavayi*. The results proved that the development of EST-SSR markers based on ESTs in *P. suffruticosa* is an effective and feasible approach, which will conduce to study on genetic diversity and genomics in *P. delavayi*.

**Key words:** *Paeonia suffruticosa*; *Paeonia delavayi*; EST; SSR; marker

随着分子生物学技术的发展,出现了许多基于 PCR 方法的 DNA 标记技术,大大提高了科研人员对植物进行遗传分析研究的能力,其中的 SSR (simple sequence repeats) 标记,因其广泛存在于真

收稿日期:2010-10-09

基金项目:国家林业局 948 项目(2006-4-C07);国家 863 计划(2006AA100109)

作者简介:张艳丽(1984—),女,在读硕士研究生,主要从事花卉种质改良方面研究工作。

\* 通讯作者:研究员,主要从事资源植物的引种驯化研究。

核生物基因组中,并具有数量丰富,多态性高、多等位性、共显性等特点<sup>[1-2]</sup>,是目前遗传学研究中令人信赖的分子标记之一;但 SSR 引物的开发需要的周期较长,且费用高,效率较低<sup>[3]</sup>。近年来 EST (expressed sequence tags) 发展十分迅速,以每周 10 000 个的速度从大量 cDNA 文库中被开发出来<sup>[4]</sup>,因而 EST-SSR 标记的开发相对快速、简便。EST-SSR 不仅具有 SSR 标记的多态性高、共显性、重复性好等特点,而且在种属间具有良好的通用性<sup>[5-6]</sup>,目前已被广泛应用于遗传连锁图谱的构建、遗传多样性分析以及比较基因组学等研究<sup>[7-9]</sup>。

滇牡丹(*Paeonia delavayi* Franch.)<sup>[10]</sup>是芍药属分布最南的一个牡丹类群,是中国西南地区特有种,在芍药属植物的起源、演化和地理分布研究上有着重要的学术价值<sup>[11]</sup>。出于其地理分布区狭小及其

研究和应用价值的考虑,滇牡丹已被认为是渐危种而列为国家三级保护植物<sup>[12]</sup>。目前,有关滇牡丹 EST-SSR 分子标记研究尚未见报道。本文从 NCBI (美国国立生物技术信息中心) 数据库中搜索到牡丹(*Paeonia suffruticosa* Andrew.) ESTs 序列 2 204 条,并进行了 SSRs 的查找和分析。在此基础上,设计扩增 SSR 位点的特异引物,通过对合成引物的多态性检测,开发滇牡丹 EST-SSR 分子标记。本研究旨在为滇牡丹 EST 资源的进一步开发利用、遗传图谱绘制以及目标性状标记的辅助选择提供一种实用、高效的研究工具。

## 1 材料和方法

### 1.1 材料

本实验材料为采自云南的 10 种具不同花色类型的滇牡丹,每类群 5 株。具体材料信息见表 1。

表 1 滇牡丹花色类群及来源

类群编号	花色类群	采集地名	海拔/m	经度(E)	纬度(N)
DM1	绿色变态	德钦明永冰川	2 520 ~ 2 691	98°46' ~ 98°47'	28°27' ~ 28°32'
DM2	绿色	德钦明永冰川	2 520 ~ 2 691	98°46' ~ 98°47'	28°27' ~ 28°32'
DS1	纯黄色	德钦升平一中后山	3 175 ~ 3 190	98°55' ~ 98°48'	28°28' ~ 28°51'
DS2	黄色基部红斑	德钦升平一中后山	3 175 ~ 3 190	98°55' ~ 98°48'	28°28' ~ 28°51'
DY	淡黄色基部红斑	德钦云岭红坡	2 681 ~ 2 691	98°54'	28°19'
ZH1	紫红色中部黄斑	中甸滑雪场	3 318 ~ 3 333	99°30' ~ 99°45'	27°42' ~ 27°57'
ZH2	橙黄色	中甸滑雪场	3 318 ~ 3 333	99°30' ~ 99°45'	27°42' ~ 27°57'
ZH3	紫红色	中甸滑雪场	3 318 ~ 3 333	99°30' ~ 99°45'	27°42' ~ 27°57'
LT	黄绿色基部紫红斑	丽江太安乡	3 138 ~ 3 144	100°02'	26°43'
LY	暗红色	丽江玉龙牦牛坪	3 202 ~ 3 523	100°14' ~ 100°16'	27°04' ~ 27°10'

### 1.2 方法

1.2.1 基因组 DNA 提取 采用 CTAB<sup>[13]</sup>法提取供试材料基因组 DNA。

1.2.2 牡丹 EST-SSR 序列查找 从 GenBank/dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) 下载牡丹 EST 序列,共获得 2 204 条牡丹 ESTs 序列,均为花芽 cDNA 文库测序产物。

应用 SSRIT (simple sequence repeat identification tool) 软件 (<http://www.gramene.org/db/searches/ssrtool>) 在线搜索 EST-SSR。搜索标准为:二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸重复序列次数大于或等于 5。

1.2.3 EST-SSR 引物设计开发 应用 Serafer 1.3 软件设计 EST-SSR 引物,设计引物时设置的主要参数为:GC 含量 20% ~ 80%,退火温度( $t_m$ ) 57 ~ 63 °C,上下游引物复性温度相差不大于 5 °C,引物长约

20 bp,预期扩增产物长度 120 ~ 500 bp。本实验从中挑选 51 对引物由上海生工生物技术公司合成以用于扩增。

1.2.4 PCR 扩增和聚丙烯酰胺凝胶电泳 PCR 反应 20  $\mu$ L 体系中:10  $\times$  buffer 2.0  $\mu$ L;1 U  $\cdot$   $\mu$ L<sup>-1</sup> 的 Taq DNA 聚合酶 2  $\mu$ L;25 mmol  $\cdot$  L<sup>-1</sup> 的 MgCl<sub>2</sub> 2  $\mu$ L;10  $\mu$ mol  $\cdot$  L<sup>-1</sup> 的引物各 1  $\mu$ L;约 30 ng  $\cdot$   $\mu$ L<sup>-1</sup> 的模板 DNA 1.5  $\mu$ L;10 mmol  $\cdot$  L<sup>-1</sup> 的 dNTPs 0.5  $\mu$ L;ddH<sub>2</sub>O 11  $\mu$ L。扩增程序如下:95 °C 预变性 7 min,进入第一阶段循环,95 °C 变性 15 s,62.5 ~ 67 °C 退火 30 s,72 °C 延伸 60 s,共 18 个循环,每循环退火温度下降 0.6 °C;第二阶段,95 °C 变性 15 s,51.5 ~ 57 °C 复性 30 s,72 °C 延伸 60 s,共 30 个循环,最后 72 °C 保持延伸 7 min。PCR 产物在 8% 非变性聚丙烯酰胺凝胶中电泳,银染后对多态性进行检测。

## 2 结果与分析

### 2.1 牡丹 EST-SSR 分布、频率及特点

在搜索的 2 204 个 ESTs 中,共发现了分布于 324 条 ESTs 中的 335 个 SSRs,其中,含有 SSR 的 EST 占全部 EST 的 14.70%,SSR 出现的频率是 15.20%,这表明牡丹 EST 中的 SSR 数量很丰富。在 324 条含有 SSR 的 EST 中,只有 1 个 SSR 的 EST 有 314 条,含有 2 个 SSR 的有 11 条。

分析结果(表 2)表明:牡丹的 EST-SSR 主要有 4 种类型,即二核苷酸重复、三核苷酸重复、四核苷酸重复和六核苷酸重复,但出现频率各不相同(表 2);其中,二核苷酸重复最为常见,共有 282 个,在全部 EST 中出现的频率为 12.79%,占 SSR 比例达 84.18%;其次为三核苷酸重复,出现频率为 2.31%,占 SSR 比例为 15.22%;四和六核苷酸重复出现频率均为 0.05%,占 SSR 比例均为 0.30%。由此可见,在牡丹 EST-SSR 中,二核苷酸重复占主导地位。

表 2 牡丹 EST 中 SSR 的出现频率

重复类型	重复基元种数	SSR 数量	占全部 SSR 比例/%	在 EST 中出现频率/%
二核苷酸	11	282	84.18	12.79
三核苷酸	22	51	15.22	2.31
四核苷酸	1	1	0.30	0.05
六核苷酸	1	1	0.30	0.05
总计	35	335	100	15.20

在搜索出的牡丹 EST-SSR 中,共观察到 35 种核苷酸重复(表 3),其中,二核苷酸重复中,ag、tg、tc、ga 出现的次数较多,分别占二核苷酸重复的 50.35%、12.06%、10.28%、10.28%;在三核苷酸基

元中,aga、acc、gat 出现的次数较多,分别占三核苷酸重复的 31.37%、7.84%、7.84%;四核苷酸重复为 atgt;六核苷酸重复为 agaggg。

表 3 牡丹 EST-SSR 中不同重复基元

重复类型	重复基元	数量	所占比例/%	重复基元	数量	所占比例/%
二核苷酸	tg	34	12.06	ct	22	7.80
	ag	142	50.35	at	3	1.06
	tc	29	10.28	ca	2	0.71
	ta	12	4.26	ac	2	0.71
	ga	29	10.28	gc	1	0.35
	gt	6	2.13			
三核苷酸	gaa	2	3.92	ttg	1	1.96
	gea	1	1.96	agc	1	1.96
	cag	2	3.92	tat	1	1.96
	tgt	1	1.96	ggt	1	1.96
	tta	1	1.96	caa	1	1.96
	tgc	2	3.92	gct	1	1.96
	agg	1	1.96	ggc	2	3.92
	acg	3	5.88	cca	1	1.96
	atg	3	5.88	acc	4	7.84
	gac	1	1.96	aga	16	31.37
tct	1	1.96	gat	4	7.84	
四核苷酸	atgt	1	100			
六核苷酸	agaggg	1	100			

### 2.2 EST-SSR 引物的筛选

用 DM2、DS1、DY、ZH2、LT、LY 类群共 6 个 DNA 模板对 51 对 EST-SSR 引物进行筛选。有 10 对引物

产生了扩增产物,占合成引物的 19.61%。

### 2.3 滇牡丹 EST-SSR 的多态性分析

利用筛选出的 13 对引物对 10 个不同类群的滇

牡丹进行多态性实验。结果有 10 对引物稳定且具多态性,占可扩增出产物引物的 100%,且在同一类群不同 DNA 模板间存在多态性,整体呈现出较高的

多态性。

图 1 为引物 P4、P9 多态性扩增情况,结果显示,利用 EST 建立滇牡丹多态性 SSR 标记是可行的。

表 4 10 对 EST-SSR 引物信息

引物名称	重复基元	引物序列	退火温度/℃	预期产物片段大小/bp
P1	(CTC) <sub>4</sub>	F: TGGCGCTCCACACGGTATCTT	58.24	228
		R: ATCGTCTTCGGTTCGCCGGT	59.08	
P2	(GCA) <sub>4</sub>	F: TCGATGCCGGCAGGTTTCT	58.42	381
		R: ACAGGTCTTGTCGGCCTTGT	58.01	
P3	(AAG) <sub>4</sub>	F: TCCAAAAGAGGGCGGTGCT	58.25	306
		R: TCCCTCCGAAACGGCATGT	58.05	
P4	(TTAA) <sub>3</sub>	F: ATGGCGAGATTGCCCGTCT	58.20	228
		R: TGCAGTCACCACCATCCGTCA	58.10	
P5	(TCTCGC) <sub>3</sub>	F: CGTCATCTGAACGCCGCCGA	59.85	208
		R: TCACGTTGGATGTTCATGCCCA	58.76	
P6	(AGA) <sub>5</sub>	F: CGGCGTGCCTTTGAGACGTG	58.91	157
		R: TGCCTCGAAAACTCGCCTTCTCC	59.48	
P7	(CATT) <sub>3</sub>	F: AGCGTGAAGCAACAAGCCGTG	58.82	167
		R: ACTGCGTTTCACGGCGAGGA	59.01	
P8	(ACTGA) <sub>3</sub>	F: AGCAGAGAAGGTAGGCGGCG	58.59	171
		R: AGCGCTGAAGCCAGTCATGG	59.45	
P9	(CAC) <sub>4</sub>	F: GGGGACTCAAATCCTTGCGAAAACCA	59.68	189
		R: AGGCCTAGTTTTGGTCTGGGCG	58.83	
P10	(GGT) <sub>4</sub>	F: CAGCTGGCTTCCCACGCAGT	59.63	177
		R: TAGCCATGCCGCCTCCACCT	59.97	

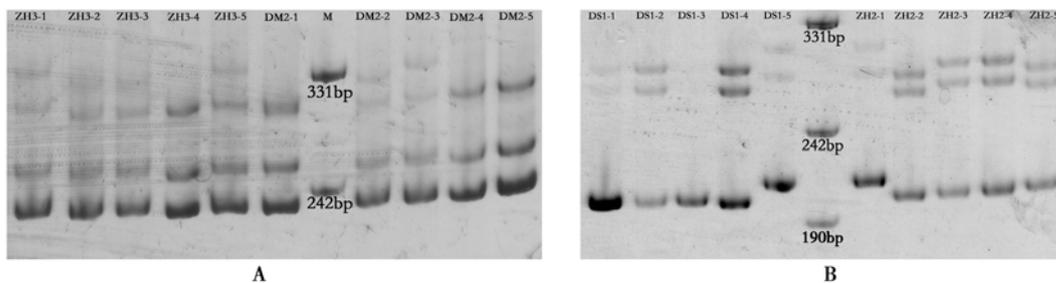


图 1 引物 P4(A)、P9(B)在不同滇牡丹花色类群间多态性

其它不同引物的扩增及其显示多态性情况见表 5。

表 5 10 对引物在 10 个滇牡丹花色类群的扩增情况和多态性

引物编号	扩增产物	多态性	引物编号	扩增产物	多态性
P1	AB	P	P6	AB	P
P2	AB	P	P7	AB	P
P3	A	P	P8	B	P
P4	AB	P	P9	B	P
P5	AB	P	P10	B	P

注:A:预期大小片段;B:非预期大小片段;AB:同时出现预期大小片段和非预期大小片段;P:有多态性。

### 3 讨论

本实验在 2 204 条 EST 序列中搜索 SSR 出现的频率为 15.20%;而现有文献报道的 SSR 出现频率各不相同,如甘蔗为 2.9%<sup>[14]</sup>、水稻为 4.7%<sup>[15]</sup>、油菜为 15.58%<sup>[16]</sup>、百合为 5.98%<sup>[17]</sup>、橡胶树为 11.42%<sup>[18]</sup>。经前人研究表明,若 EST 数足够大且无偏倚性,即 4 种不同的碱基随机组合将产生单核苷酸 2 种、二核苷酸 4 种、三核苷酸 10 种、四核苷酸 33 种、五核苷酸 102 种和六核苷酸 350 种基本重复类型<sup>[19]</sup>;而从目前报道来看,大多数植物的 EST-

SSR 主要属于二、三核苷酸重复类型,不同物种间主导重复基元的类型有所差异。牡丹 EST-SSR 分析结果中二核苷酸重复为主导类型,所占比例高达 84.18%,大多数二核苷酸重复均有出现,而其它核苷酸重复多数未出现,表现出明显的偏倚性。在二核苷酸重复中,ag 占主导地位,出现比例达 50.35%,这一现象与多数植物中报道的情况相同<sup>[20]</sup>;三核苷酸重复则以 aga 重复为主。牡丹 SSR 分布及核苷酸出现频率表现出来的这些差异是其自身特点的真实反应,还是因 EST 数量较少而造成,还有待进一步研究。

上述引物虽以严格的标准设计完成,但能扩增出产物的引物比例并不高,可能原因有:①所利用的设计软件本身特点导致或反应体系仍需进一步优化等;② 2 个引物或引物之一的序列跨越在两个外显子上,或两引物间有一相当长的内含子序列<sup>[2]</sup>。根据 EST 序列设计的引物对基因组 DNA 进行 PCR 扩增时,扩增产物部分会比目标 EST 片段长,这在以往报道中也是存在的,主要是由于内含子的存在而导致,且在前人的研究中也出现了引物扩增出多于 2 条条带的现象<sup>[21-22]</sup>。

目前,大规模 EST 的产生和 EST 数据库的建立,为新开发的 SSR 标记的发展提供了宝贵资源。本实验表明,从牡丹 EST 序列中开发 SSR 标记是可行的,这对于加速牡丹目标性状分子标记筛选、基因定位、遗传连锁图谱构建,实现特定性状的辅助选择和进行比较基因组学研究,都具有重要意义。

#### 参考文献:

- [1] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats [J]. Trends in Plant Science, 1996, 1: 215 - 222
- [2] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plant: features and applications [J]. Trends in Biotechnology, 2005, 23(1): 48 - 55
- [3] 谢 皓, 陈学珍, 杨 柳, 等. EST - SSR 标记的发展和在植物遗传研究中的应用 [J]. 北京农学院学报, 2005, 20(4): 73 - 76
- [4] 周延清. DNA 分子标记技术在植物研究中的应用 [M]. 北京: 化学工业出版社, 2005: 186 - 187
- [5] Yu J K, Rota M L, Kantety R V, et al. EST derived SSR markers for comparative mapping in wheat and rice [J]. Molecular Genetics and Genomics, 2004, 271: 742 - 751
- [6] Poncet V, Rondeau M, Tranchant C, et al. SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus [J]. Molecular Genetics and Genomics, 2006, 276: 436 - 449
- [7] 闫 娟, 楚海家, 李建强, 等. 用 EST-SSR 标记分析中国北部和中部地区天蓝苜蓿的遗传多样性和遗传结构 [J]. 生物多样性, 2008, 16(3): 263 - 270
- [8] 李小白, 崔海瑞, 张明龙. EST 分子标记开发及在比较基因组学中的应用 [J]. 生物多样性, 2006, 14(6): 541 - 547
- [9] 刘 振, 姚明哲, 王新超, 等. 基于 EST - SSR 的福建地区茶树资源遗传多样性和亲缘关系分析 [J]. 中国农业科学, 2009, 42(5): 1720 - 1727
- [10] 中国科学院昆明植物研究所. 云南植物志: 第 11 卷 [M]. 北京: 科学出版社, 2000: 291 - 294
- [11] 龚 询, 潘跃芝, 杨志云. 滇牡丹多样性和现状评估 [J]. 西北植物学报, 2003, 23(2): 218 - 223
- [12] 冯国楣. 黄牡丹 [C] // 傅立国. 中国植物红皮书—稀有濒危植物(第一册). 北京: 科学出版社, 1992: 530 - 531
- [13] Clark M S. 植物分子生物学 - 实验手册 [M]. 顾洪雅, 瞿礼嘉译. 北京: 高等教育出版社, 1998: 4 - 7
- [14] Cordeiro G M, Casu R, McIntyre C L, et al. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum [J]. Plant Science, 2001, 160(6): 1115 - 1123
- [15] Kantety R V, Rota M L, 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): 501 - 510
- [16] 李小白, 张明龙, 崔海瑞. 油菜 EST 资源的 SSR 信息分析 [J]. 中国油料作物学报, 2007, 29(1): 20 - 25
- [17] 杨素丽, 明 军, 刘 春, 等. 基于 EST 信息的百合 SSR 标记的建立 [J]. 园艺学报, 2008, 35(7): 1069 - 1074
- [18] 安泽伟, 赵彦宏, 程 汉, 等. 橡胶树 EST - SSR 标记的开发与应用 [J]. 遗传, 2009, 31(3): 311 - 319
- [19] Rota L R, Kantety R V, Yu J K. Nonrandom distribution and frequencies of genomic and EST - derived microsatellite markers in rice, wheat, and barley [J]. BMC Genomics, 2005, 6(1): 23
- [20] 李永强, 李宏伟, 高丽锋, 等. 基于表达序列标签的微卫星标记 (EST-SSR) 研究进展 [J]. 植物遗传资源学报, 2004, 5(1): 91 - 95
- [21] Kononov F, Toshchakova E, Gostimsky S. A CAPS marker set for mapping in linkage group III of pea (*Pisum sativum* L.) [J]. Cellular & Molecular Biology Letters, 2005, 10(1): 163 - 171
- [22] Jiang D, Zhong G Y, Hong Q B. Analysis of Microsatellites in Citrus Unigenes [J]. Acta Genetica Sinica, 2006, 33(4): 345 - 353