

蝴蝶兰 EST-SSRs 分析

张君毅*, 陈瑞凤

华侨大学生物工程与技术系, 福建厦门 361021

摘要: 对蝴蝶兰属 EST 序列进行了 SSR 分析。蝴蝶兰属 EST 总长为 4.5 Mb, 含有 609 个 SSR。SSR 出现频率 7.65%, 平均距离 7.39 kb, 平均长度为 22.17 bp。单碱基、二碱基和三碱基重复是主要重复类型, 分别占 EST-SSR 总数的 21.67%、40.39% 和 33.50%。A、AG 和 CCG 分别是单碱基、二碱基和三碱基重复中主导重复基元, 分别占 96.21%、58.54% 和 32.25%。设计引物及检测的结果表明, 蝴蝶兰 EST-SSR 标记对兰科其他属植物具有一定的通用性。

关键词: 蝴蝶兰; SSR; EST; 微卫星

EST-SSRs Analysis of Moth Orchids

ZHANG Jun-Yi*, CHEN Rui-Feng

Department of Bioengineering & Biotechnology, Huaqiao University, Xiamen, Fujian 361021, China

Abstract: Simple sequence repeats (SSRs) were investigated in expressed sequence tags (EST) of moth orchids. The total length of moth orchids non-redundant ESTs was 4.5 Mb, including 609 SSRs. The frequency of EST-SSRs was 7.65% and the mean distance was 7.39 kb. The mean length of SSRs was 22.17 bp. Mononucleotide, dinucleotide and trinucleotide repeats were major types, which accounted for 21.67%, 40.39% and 33.50% in all SSRs, respectively. A, AG and CCG were mainly frequent motifs in mononucleotide, dinucleotide and trinucleotide, accounting for 96.21%, 58.54% and 32.25%, respectively. Primers were designed and detected. The result showed that some of them could be transferable to other genera of orchid.

Key words: moth orchids; SSR; EST; microsatellite

蝴蝶兰属植物约 40 种, 分布于热带亚洲至澳大利亚, 具有较高的观赏价值和经济价值, 被誉为洋兰“皇后”。近 20 年来, 中国内地的蝴蝶兰市场需求旺盛, 促进了育种工作开展(乔永旭等 2008)。然而兰科植物因在自然界杂交的缘故, 给属内种的分类和整理带来了较大的困难, 给进一步杂交培育新品种带来不便。

简单序列重复(simple sequence repeats, SSR) 标记以其共显性、高多态性、分布广泛性等优点被广泛地应用于植物育种中(徐莉和赵桂仿 2002)。SSR 应用的前提是引物开发。用数据库搜索的方法开发 SSR 相对于筛选基因组文库法、微卫星富集法、省略筛库法等传统方法, 具有省时、省力、低成本等优点(常玉梅等 2005), 尤其是从表达序列标签(expressed sequence tags, EST)数据库中搜索的 SSR 还具有通用性高的特点(Holton 等 2002), 已经成为一种广为采用的方法(Cardle 等 2000; Thiel 等 2003)。本文通过分析 NCBI 数据库中蝴蝶兰属植物的 EST 等 DNA 序列, 挖掘简单重

复序列, 为蝴蝶兰属植物 SSR 引物开发提供依据, 进而为蝴蝶兰属资源的遗传变异、品种鉴别、基因标记及进化研究提供重要的信息。

材料与方法

1 EST 的获得与处理

从 NCBI 数据库下载蝴蝶兰属植物 EST 序列, 数据截止到 2009 年 8 月 1 日。采用 EST-trimmer 软件除去 3' 端 poly T 或 poly A 以及长度小于 100 bp 的序列。对于长度大于 700 bp 的 EST 则保留其 5' 端序列。通过软件 Phrap 进行片段重叠群分析和聚类以去除冗余序列。运用 BLSAT_X 程序对处理过的数据进行同源性比较, 采用 Kofler 等 (2007) 开发的 SciRoKo 3.1 软件对聚类后的 EST 搜

收稿 2010-03-01 修定 2010-04-06

资助 福建省农业科技重点项目(2009N0041)和泉州重点科技项目(2008N40)。

* 通讯作者(E-mail: zjy0054@yahoo.com.cn; Tel: 0595-22693329)。

SSR, 搜索单碱基最少重复次数为15次, 二碱基为8次, 三碱基为5次, 四碱基以上为4次。

2 SSR引物设计

用Primer 5.0软件对含有SSR的EST序列进行引物设计, 尽量避免引物二级结构的出现。引物设计主要参数为: 引物长17~23 bp; 引物退火温度50~60 °C, 正向和反向引物退火温度之差在5 °C以内; GC含量40%~70%; PCR预期产物长100~300 bp。

3 引物筛选与通用性分析

用蝴蝶兰属(*Phalaenopsis*)的‘台林红天使’(*Phalaenopsis Tai Lin Redangel* ‘V31’)等品种验证引物的有效性。此外, 选择其他属内及属间种或杂交种以检验引物在兰科植物的通用性, 包括‘国王’(*Dendrobium Red Emperor* ‘Prince’), ‘伊丽莎白’(*Bulbophyllum Elizabeth Ann* ‘Buckleberry’), ‘天母’(*Cattleya Orlade’s Grand* ‘TM’), ‘白仙女’(*Degarmoara Winter Wonderland* ‘White Fairy’), ‘小金鱼’(*Rhyncattleanthe Ahchung Yoyo* ‘Little Goldfish’), 腋唇兰[*Maxillariella tenuifolia* (Lindl.) M.A. Blanco & Carnevali]和万代兰(*Vanda coerulea* Griff.)等7种植物。采用王国鼎等(2007)的方法提取所有材料的DNA。PCR扩增反应体系及参数参考Huang等(2010)的报道。用6%非变性聚丙烯酰胺凝胶进行电泳和分离。

结果与讨论

1 EST-SSR的分布特点

通过检索从NCBI数据库中获得7963条蝴蝶兰属植物EST序列, 分别来源于桃红蝴蝶兰(*P. equestris*)花芽5604条和紫纹蝴蝶兰(*P. violacea*)花芽2359条。共有609条含有SSR, 出现频率7.65%,

略低于甜瓜的8.16% (胡建斌等2009)。SSR平均距离为7.39 kb, 与大豆(7.4 kb)基本相当(Cardle等2000), 高于拟南芥(14.9 kb)、小麦(15.6 kb)、棉花(20.0 kb)和番茄(11.1 kb)等作物(李永强等2004), 低于水稻3.4 kb (Cardle等2000)和大麦6.3 kb (Thiel等2003), 属于SSR出现频率相对较高的植物, 说明大规模开发EST-SSR标记的可能性较大。两种蝴蝶兰ESR-SSR的分布情况列于表1。从SSR数量上看, 桃红蝴蝶兰539个远高于紫纹蝴蝶兰的70个。SSR的平均长度和出现频率, 前者显著大于后者。

2 EST-SSR的基元类型与频率

基元统计时, 将所有可循环的及互补的序列归为一种标准基元。从单碱基到六碱基重复共有30种基元(表2)。单碱基基元以A为主, 共有127个, 占单碱基SSR总数的96.21%; 在二碱基基元中AG出现频率最高, 占该类SSR的58.54%; 在三碱基类型中, 以CCG为主, 占32.35%。这与在玉米、水稻和高粱等单子叶植物的结果基本一致(Kantety等2002)。从四碱基到六碱基基元共有16种, 占基元类型总数的一半以上, 但各自出现频率都很低。

桃红蝴蝶兰以二碱基重复为主(224个), 占其SSR总数的41.6%, 其次是三碱基重复168个, 占31.1%; 紫纹蝴蝶兰则以三碱基重复为主(36个), 占其SSR总数的51.4%, 其次是二碱基重复(22个), 占31.4%。从总体上看, 在所有EST-SSR中, 出现最多是二碱基基元, 共有246个, 占总数的40.39%; 其次是三碱基重复和单碱基重复, 分别占总数的33.50%和21.67%。由表2可见, 蝴蝶兰属EST-SSR主要类型是AG、A、AT和CCG四种, 合计占总量的71.92%。Dreisigacker等(2004)发现, 高级基元SSR多态性普遍比低级基元低, 说明本研究

表1 桃红蝴蝶兰和紫纹蝴蝶兰 EST-SSR 分布

Table 1 Distribution of EST-SSRs in *P. equestris* and *P. violacea*

物种	数量/条	EST 总长 /bp	SSR 数量	平均长度 /bp	出现频率*/%	平均距离**/kb
桃红蝴蝶兰	5 604	3.1×10 ⁶	539	22.47	9.62	5.75
紫纹蝴蝶兰	2 359	1.4×10 ⁶	70	11.58	2.97	20.00
共计	7 963	4.5×10 ⁶	609	22.17	7.65	7.39

* 出现频率 = 检出的 SSR 数目 / EST 总数; ** 平均距离 = EST 总长度 / SSR 总数。

表2 兰花 EST-SSR 基元类型与频率
Table 2 Motif of EST-SSRs in moth orchid and its frequency

SSR类型	标准基元	SSR数量	基元百分率/%	频率*	总百分率/%
单碱基	A	127	96.21	28.22	21.67
	C	5	3.79	1.11	
二碱基	AT	101	41.05	22.44	40.39
	AG	144	58.54	32.00	
三碱基	AC	1	0.41	0.22	33.50
	AAC	7	3.43	1.56	
	AAG	42	20.59	9.33	
	AAT	36	17.65	8.00	
	ACC	11	5.39	2.44	
	ACG	3	1.47	0.67	
	AGC	9	4.41	2.00	
	AGG	13	6.37	2.89	
	ATC	17	8.33	3.78	
	CCG	66	32.35	14.67	
四碱基	AAAC	3	23.08	0.67	2.13
	AAAT	7	53.85	1.56	
	AGCG	1	7.69	0.22	
	AGGG	1	7.69	0.22	
	ATAG	1	7.69	0.22	
五碱基	AAAAAT	1	33.33	0.22	0.49
	AAACC	1	33.33	0.22	
	ACCCC	1	33.33	0.22	
六碱基	AGCGGC	2	18.18	0.44	1.81
	AGCAGG	1	9.09	0.22	
	AGAGGG	1	9.09	0.22	
	ACTCCC	1	9.09	0.22	
	ACCGCC	1	9.09	0.22	
	AATATT	1	9.09	0.22	
	AAGGAG	1	9.09	0.22	
AAACCC	3	27.27	0.67		

*SSR 频率以每 1 Mb 碱基出现 SSR 的数量表示。

发掘的大部分 SSR 具有高多态性潜能。

3 EST-SSR的重复次数与长度

SSR重复次数的变异引起位点长度的变化是产生 SSR 多态性的主要原因。一般认为, SSR 位点的变异频率与基元重复数存在一定的正相关, 即重复次数越多, SSR 产生变异的可能性越大 (Schlotterer 2000)。当 SSR 基序长度大于或等于 20 bp 时多态性较高 (Temnykh 等 2001)。蝴蝶兰单碱基至三碱基类型 SSR 重复次数变异较大 (图 1), 分别为 SSR 15~69、8~42 和 5~16 次, 其基序长度也较大, 平均长度分别为 18.90、26.89 和 18.32

bp。总体 EST-SSR 平均长度为 22.17 bp。由此可见, 无论是重复次数还是基序长度都显示出所发掘的 SSR 可能具有良好的变异性, 尤其是二碱基和三碱基重复基元。

另外, 从图 1 可以看出, 三碱基类型 SSR 重复 4~16 次, 随着重复次数增加其数量呈现下降的趋势; 同样的趋势也出现在二碱基类型 8~15 次和单碱基 15~24 次中。对于违背这个趋势的 SSR 需要特别注意, 如二碱基类型 15~24 次重复以至 25 次及以上重复序列, 往往产生较大变异, 是获得有效 SSR 的活跃位点。

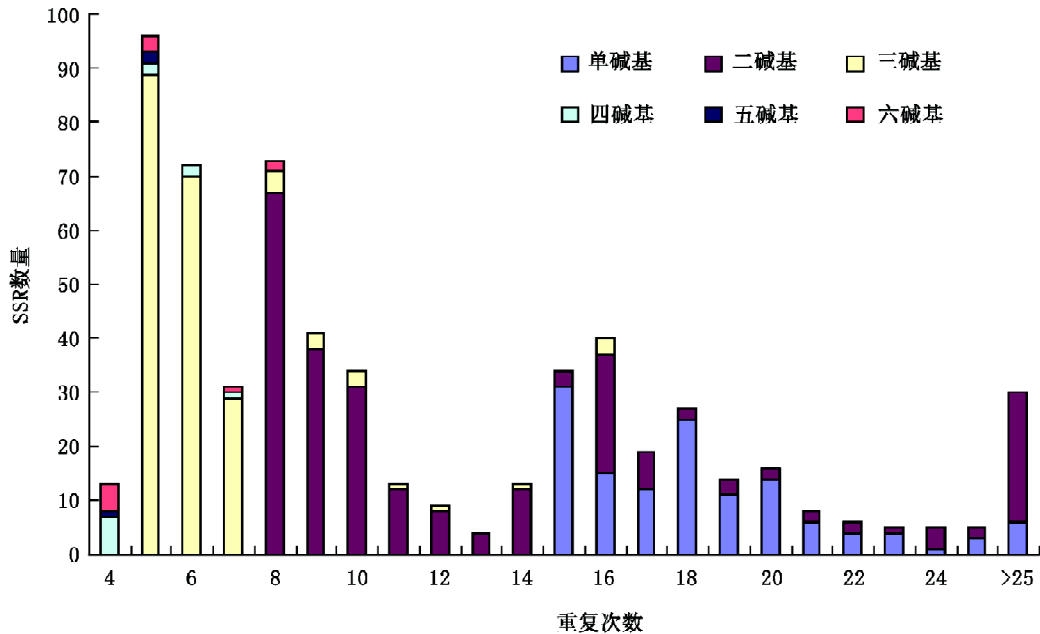


图1 不同基元长度的SSR数量

Fig.1 The numbers of SSRs in different motif length

4 EST-SSR有效性验证

随机挑选40条含有SSR的EST进行引物设计。由于微卫星侧翼序列过短、GC含量偏高或偏低等原因,获得28对引物。随机选择8对引物对‘台林红天使’等蝴蝶兰品种验证EST-SSR标记的有效性。结果发现,有5对均有扩增产物,产物大小为100~300 bp,基本符合预期大小,说明多态性主要由EST-SSR长度变化引起。每对引物能检测到2~3个等位基因,平均每对引物能检测到2.2个等位基因。

进而以这5对引物对包括蝴蝶兰在内的8种兰科不同属的植物进行PCR扩增,检测蝴蝶兰EST-SSR标记的通用性。图2为引物Estssr4扩增带谱,可见该引物并不适用于‘石豆兰’和腋唇兰。引物Estssr1能够扩增除石斛属‘国王’和卷瓣兰属‘石豆兰’外的其它6种不同属的植物基因组DNA,多态性位点有4个。引物Estssr5能够扩增所有参试材料,但多态性位点只有2个。有扩增产物的引物序列、预期产物大小、退火温度和所检测到的等位基因数见表3。另外,5对引物在万代兰中均得到扩增产物,说明蝴蝶兰EST-SSR标记在万代兰属中的通用性较高。由于EST-SSR可以直接标记遗

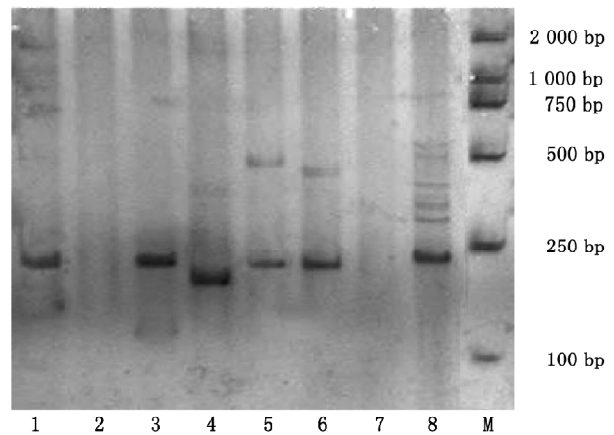


图2 引物Estssr4扩增带谱

Fig.2 PCR amplification pattern generated by primer pair Estssr4

1: ‘白仙女’; 2: ‘伊丽莎白’; 3: ‘天母’; 4: ‘国王’; 5: ‘台林红天使’; 6: ‘小金鱼’; 7: 腋唇兰; 8: 万代兰; M: DL2000 marker。

传性状,具有开发成本低、物种间通用性高、EST数目巨大等优点,在一定程度上弥补了相对基因组SSR标记多态性较低的不足。进行兰科唯一登录EST的蝴蝶兰属植物EST-SSR的大规模开发,必将大大促进该科植物在资源分类、辅助育种、遗传作图等方面的研究。

表3 有扩增产物的 EST-SSR 引物
Table 3 The primer pairs produced amplified products

引物编号	重复基元	引物序列(5' → 3')	$T_m/^\circ\text{C}$	多态位点*	预期产物大小 /bp	EST 来源
Estssr1	(TC) ₁₀	CTTGTTATTTCCCTCCTCG CCTTTAGCCCAAGTTCAGT	55	4(6)	151	CB033577.1
Estssr2	(CCG) ₆	GAGCCAGAAGTAGAACCAGC TTCGGTCTTAGTCTCCTCAAT	50	3(3)	237	CB034804.1
Estssr3	(CTT) ₆	CCTTCCCTCATACCATCC CTCTGCTTCCCACCATCAA	56	2(3)	233	CB031932.1
Estssr4	(GGC) ₇	TTTCCTCCCAAAGACCCT TCAAGCCCACATCCCATT	56	2(6)	242	CB034028.1
Estssr5	(AGAAGG) ₅	ATGGGAAAGGCAAGCAAGT CAATCTCGGACTCAAACAAC	55	2(8)	253	CB032583.1

* 括号中数字为参试 8 种植物中有扩增产物的种类数。

参考文献

- 常玉梅, 李绍戊, 梁利群, 孙效文(2005). 微卫星标记的制备策略. 中国生物工程杂志, S1: 210~214
- 胡建斌, 刘颖, 王兰菊, 李建吾(2009). 甜瓜 EST 序列中微卫星的分布特征. 植物生理学通讯, 45 (3): 258~262
- 李永强, 李宏伟, 高丽锋, 何蓓如(2004). 基于表达序列标签的微卫星标记(EST-SSRs)研究进展. 植物遗传资源学报, 5 (1): 91~95
- 乔永旭, 张永平, 王桂兰, 陈超, 王燕(2008). 蝴蝶兰原生质体提取方法的优化. 植物生理学通讯, 44 (6): 1177~1180
- 王国鼎, 文晓鹏, 季祥彪, 乔光, 胡鹏(2007). 11 种兰属植物 DNA 的提取及 RAPD-PCR 实验体系的建立与优化. 种子, 26 (3): 24~27
- 徐莉, 赵桂仿(2002). 微卫星 DNA 标记技术及其在遗传多样性研究中的应用. 西北植物学报, 22 (3): 714~722
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshal D, Waugh R (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 156: 847~854
- Dreisigacker S, Zhang P, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2004). SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. *Crop Sci*, 44 (2): 381~388
- Holton TA, Christopher J, McClure L, Harker N, Henry RJ (2002). Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. *Mol Breed*, 9: 63~71
- Huang Y, Li F, Chen KS (2010). Analysis of diversity and relationships among Chinese orchid cultivars using EST-SSR markers. *Biochem Syst Ecol*, 38 (1): 93~102
- Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*, 48: 501~510
- Kofler R, Schlotterer C, Lelley T (2007). SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, 23: 1683~1685
- Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109: 365~371
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res*, 11: 1441~1452
- Thiel T, Michalek W, Varshney RK, Graner A (2003). Exploiting EST database for the development and characterization of gendered SSR markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*, 106: 411~422