

辣椒 EST-SSR 标记的开发与应用

吴智明¹, 刘伟强², 唐鑫², 崔俊杰², 程蛟文², 胡开林²

(1 仲恺农业工程学院 园艺园林学院, 广东 广州 510225; 2 华南农业大学 园艺学院, 广东 广州 510642)

摘要:通过对数据库中 32 295 条非冗余的辣椒 EST 序列进行搜索,发现 3 396 个 SSR 分布于 3 277 条 EST 序列中, EST-SSR 频率为 10. 52%, 平均分布距离为 4. 46 kb. 在辣椒 EST-SSR 中, 二核苷酸和三核苷酸重复基元占主导地位, 分别占总 SSR 的 43. 02% 和 37. 84%. 优势重复基元为 GA/TC、AG/CT 和 AT, 分别占 15. 99%、11. 98% 和 11. 37%. 用 Primer Premier 5. 0 软件设计了 420 对引物, 对辣椒抗疫病自交系‘B072’和感疫病自交系‘B088’进行 PCR 扩增, 403 对引物有扩增产物, 引物有效扩增率为 95. 96%, 其中 76 对引物有多态性, 多态性引物占可扩增引物的 18. 86%. 试验证明, 利用辣椒 EST 序列开发 SSR 标记是可行的.

关键词:辣椒; EST-SSR; 标记开发
中图分类号:S641. 3 **文献标志码:**A **文章编号:**1001-411X(2012)02-0171-04

Development and Application of EST-SSR Markers in Pepper

WU Zhi-ming¹, LIU Wei-qiang², TANG Xin², CUI Jun-jie², CHENG Jiao-wen², HU Kai-lin²

(1 College of Horticulture and Landscape Architecture, Zhongkai University of
Agriculture and Engineering, Guangzhou 510225, China;

2 College of Horticulture, South China Agricultural University, Guangzhou 510642, China)

Abstract: A total of 32 295 non-redundant expressed sequence tags (ESTs) in pepper were screened by using bioinformatics software to search for SSR motifs, 3 396 SSRs were sought out, distributing in 3 277 ESTs, corresponding to one SSR in every 4. 46 kb of the ESTs. Dinucleotide and trinucleotide repeats were major types among the obtained unigenes, accounting for 43. 02% and 37. 84%, respectively. GA/TC (15. 99%), AG/CT (11. 98%) and AT (11. 37%) were the most abundant motifs. Based on the flanking sequences of these 3 396 SSRs, 420 primer pairs were designed by using Primer Premier 5. 0 software. 403 SSRs (95. 96%) were successfully amplified and 76 of them (18. 86%) showed polymorphism between ‘B072’ (the resistance inbred line) and ‘B088’ (the susceptible inbred line). These results proved that it is an effective and feasible approach to developing SSR markers based on ESTs in *Capsicum annuum* L.

Key words: pepper; EST-SSR; marker developing

SSR(Simple sequence repeat,简单重复序列)又称为微卫星 DNA(Microsatellite DNA),是一种由 2~6 个核苷酸作为基本重复单元(基元)串联组成的重复序列,这些序列广泛存在于真核细胞的基因组中.与其他分子标记相比,SSR 标记具有多态性高、检测容易、数量丰富、呈共显性遗传等优点^[1],是一种比较理想的分子标记. SSR 可分为传统的基因组 SSR

和表达序列标签 SSR(Expressed sequence tag SSR, EST-SSR).传统的基因组 SSR 标记开发需要构建基因组文库、制备探针和分子杂交等过程,费时、费力,且开发成本较高;EST-SSR 标记开发是通过搜索含有 SSR 的 EST 开发标记的方法.随着测序技术的发展,生物信息数据库中不同物种的 EST 数量也在飞速增长,使得 EST-SSR 标记的开发显得更加快捷、简

收稿日期:2011-07-10
作者简介:吴智明(1981-),男,副教授,博士;通信作者:胡开林(1963—),男,博士,教授,E-mail:hukailin@scau.edu.cn
基金项目:广东省自然科学基金团队项目(S2011030001410);广东省自然科学基金(10451022501005594)

便和高效. 近年来, 开发与应用 EST-SSR 已在小麦^[2-3]、棉花^[4]、大豆^[5]、黄瓜^[6]、甜瓜^[7]、甘蓝^[8]、梨^[9]等多种作物中报道.

辣椒 *Capsicum annuum*. L 是我国最重要的蔬菜作物之一. 目前辣椒上已公开报道可利用的 SSR 标记仅有 500 多个^[10-16], 与水稻、小麦等大田作物相比, 数量明显偏少, 远不能满足辣椒分子育种研究的需要. 本研究基于 DFCI 数据库中公开报道的 11 万多条辣椒相关 ESTs, 开发 EST-SSR 标记并分析其多态性, 为辣椒种质资源多样性分析、遗传图谱构建和分子标记辅助育种等提供参考.

1 材料与方法

1.1 材料

供试的辣椒材料为经过多代自交纯化的 2 个自交系, 分别为抗疫病自交系 ‘B072’ 和感疫病自交系 ‘B088’, 由华南农业大学园艺学院蔬菜学系提供.

1.2 方法

1.2.1 辣椒基因组 DNA 的提取 取辣椒苗期嫩叶, 采用改良 CTAB 法提取基因组 DNA^[17], 用 8 g/L 的琼脂糖凝胶电泳检测 DNA 质量, 并根据对照 λ DNA 的质量估计、定量辣椒基因组 DNA 浓度.

1.2.2 含 SSR 的辣椒 EST 序列的检索 从哈佛大学 DFCI 植物 EST 数据库 (<http://compbio.dfci.harvard.edu/tgi/plant.html>) 下载经整理的辣椒 EST 序列 (第 4 版, 截止 2009 年 5 月 21 日, 最新) 共 32 295 条, 其中包括组装 (Tentative consensus sequence, TC sequence) 的序列 14 727 条和单序 (Singleton) 17 568 条. 利用 SSR 在线分析工具 (<http://www.icugi.org/cgi-bin/ICuGI/tool/SSR.cgi>) 对 EST 序列进行 SSR 搜索. 搜索标准为 SSR 长度 ≥ 18 bp, 即二、三、四、五和六核苷酸基序 (Motif) 的最少重复次数分别 ≥ 9 、 ≥ 6 、 ≥ 5 、 ≥ 4 、 ≥ 4 , 其中不包含单核苷酸类型.

1.2.3 SSR 引物设计 根据 SSR 在 EST 序列中的位置, 去除 SSR 位点距离序列两端 ≤ 30 bp 的序列. 根据 SSR 位点两端的保守区域, 利用 Primer Premier 5.0 软件设计引物, 引物设计基于下列原则: 引物长度 18 ~ 25 bp, 预测扩增率 $\geq 80\%$; 预期扩增产物大小 100 ~ 300 bp; 引物退火温度 45 ~ 65 $^{\circ}\text{C}$; 尽量避免引物二聚体、发夹结构和错配等情况的发生. 设计好的引物由上海生工生物工程技术有限公司合成.

1.2.4 SSR 扩增 PCR 扩增体系 20 μL , 其中含 10 $\times$ PCR buffer 2 μL , 25 mmol/L MgCl_2 1.5 μL , 10 mmol/L dNTPs 0.4 μL , 1 $\mu\text{mol/L}$ SSR 引物 2.8 μL , *Taq* DNA 聚合酶 1 U, 模板 DNA 10 ng. 扩增反应程序

为: 94 $^{\circ}\text{C}$ 预变性 5 min, 94 $^{\circ}\text{C}$ 1 min, 55 $^{\circ}\text{C}$ 30 s, 72 $^{\circ}\text{C}$ 1 min, 36 个循环, 72 $^{\circ}\text{C}$ 延伸 5 min. 扩增产物在 60 g/L 的非变性聚丙烯酰胺凝胶上电泳 (电泳缓冲液 0.5 $\times$ TBE, 恒定电压 250 V, 时间 2.5 h). 银染显色.

2 结果与分析

2.1 辣椒 EST-SSR 频率

采用 SSR 在线分析工具从 32 295 条辣椒非冗余 ESTs (总长度 15.15 Mb) 中共搜索到 3 396 个 SSR, 分布在 3 277 条 EST 中, SSR 频率为 10.52%, 平均每 4.46 kb 出现 1 个 SSR (表 1). 在含有 SSR 的 3 277 条 EST 中, 3 172 条 EST 仅含有 1 个 SSR, 105 条 EST 含 2 个或 2 个以上 SSR.

表 1 辣椒 ESTs 中各基元长度 SSR 频率
Tab.1 Occurrence frequency of SSRs with different motif lengths in pepper ESTs

重复类型	基元种数	SSR 数	发生频率 ¹⁾ /%	占全部 SSR 比例/%	平均距离 ²⁾ /kb
二核苷酸	10	1 461	4.52	43.02	10.37
三核苷酸	54	1 285	3.98	37.84	11.79
四核苷酸	21	139	0.43	4.09	108.99
五核苷酸	44	155	0.48	4.56	97.74
六核苷酸	62	356	1.10	10.48	42.56
合计	191	3 396	10.52	100	4.46

1) 发生频率 = 检出的 SSR 数/无冗余 EST 总数; 2) 平均距离 = 无冗余 EST 总长度/SSR 数量.

2.2 辣椒 EST-SSR 特点

在搜索出的 3 396 个 SSR 中, 二、三、四、五、六核苷酸重复基元的 SSR 数量分别是 1 461、1 285、139、155、356 个, 占总 SSR 的比例分别为 43.03%、37.84%、4.09%、4.56% 和 10.48%. 可见, 在辣椒 EST-SSR 中二核苷酸和三核苷酸重复基元占主导地位, 其次为六核苷酸重复基元 (表 1).

检测到的重复基元共有 191 种, 各核苷酸重复基元种数如表 1 所示. 二核苷酸 EST-SSR 序列平均长度为 37.20 bp, 优势重复基元为 GA/TC, 其次为 AG/CT 和 AT, 出现频率分别为 1.68%、1.26% 和 1.20%, 分别占 SSR 总数的 15.99%、11.98% 和 11.37%; 三核苷酸 EST-SSR 序列平均长度为 21.30 bp, 其中没有特别明显的优势重复基元, 出现频率最高的 2 种重复基元分别是 AGA/TCT 和 GAA/TTC, 出现频率分别为 0.51% 和 0.43%; 四核苷酸、五核苷酸和六核苷酸 EST-SSR 序列平均长度分别为 22.01、21.08 和 26.47 bp, 各优势重复基元分别为 TAAT、ATATA/TATAT 和 CTGCTC, 出现频率分别为 0.15%、

0.17% 和 0.34% (表 2), 分别占相应核苷酸 SSR 总数的 33.81%、34.84% 和 30.90%.

表 2 辣椒 EST-SSR 主要重复基元与特征
Tab.2 The major motifs and their percentages in pepper ESTs

重复类型	重复基元	数量	所占比 例 ¹⁾ /%	总长度/ bp	平均长 度/bp	出现频 率/%
二核苷酸		1 461		54 349	37.20	4.52
	AG/CT	407	11.98	13 916	34.19	1.26
	AT	386	11.37	10 151	26.30	1.20
	CA/TG	77	2.27	1 824	23.69	0.24
	GA/TC	543	15.99	27 442	50.54	1.68
三核苷酸		1 285		27 366	21.30	3.98
	AAG/CTT	82	2.41	1 815	22.13	0.25
	AAT/ATT	85	2.50	1 788	21.04	0.26
	AGA/TCT	165	4.86	3 330	20.18	0.51
	ATA/TAT	89	2.62	2 004	22.52	0.28
	CAA/TTG	82	2.41	1 674	20.41	0.25
	CAG/CTG	68	2.00	1 968	28.94	0.21
	GAA/TTC	139	4.09	2 791	20.08	0.43
	TAA/TTA	75	2.21	1 638	21.84	0.23
四核苷酸		139		3 060	22.01	0.43
	AAAT/ATTT	20	0.59	428	21.40	0.06
	TAAT	47	1.38	944	20.09	0.15
	TTTA	21	0.62	540	25.71	0.07
五核苷酸		155		3 267	21.08	0.48
	ATATA/TATAT	54	1.59	1 110	20.56	0.17
六核苷酸		356		9 424	26.47	1.10
	ATCAGC/GCTGAT	21	0.62	504	24.00	0.07
	CCAAAG	22	0.65	528	24.00	0.07
	CTGCTC	110	3.24	2 790	25.36	0.34
	CTTGTT	21	0.62	744	35.43	0.07

1) 比例小于 0.5% 的基元未列出.

2.3 EST-SSR 引物设计及筛选结果

根据 SSR 在 EST 序列中的位置, 去除 SSR 位点距离序列两端≤30 bp 的序列. 同时通过相似性比对去除文献已报道的 EST 序列, 最终利用 Primer Premier 5.0 软件设计了 420 对 SSR 引物. 利用 420 对引物同时对辣椒抗疫病自交系 ‘B072’ 和感疫病自交系 ‘B088’ 进行 PCR 扩增. 结果表明, 除去无扩增条带 (9 对) 和条带模糊难读 (8 对) 的引物, 共有 403 对引物有较好的扩增产物, 扩增片段在 100 ~ 300 bp, 基本符合预期片段大小, 引物有效扩增率为 95.96%. 76 对引物检测到多态性, 多态性引物占可扩增引物的 18.86%, 所检测到的多态性差异明显 (图 1). 说明利用 EST 开发辣椒 SSR 标记是高效、可行的.



图 1 12 对多态性 EST-SSR 引物在 ‘B072’ 和 ‘B088’ 中的扩增带谱
Fig.1 Polymorphisms showed between ‘B072’ and ‘B088’ by 12 EST-SSR primers

3 讨论

本研究分析了来源于辣椒 EST 中的 SSR 分布规律, 并探讨了新开发的 EST-SSR 在辣椒抗/感疫病自交系多态性检测中的可行性. 在对 32 295 条辣椒 EST 进行搜索时发现了 3 396 个 SSR, 其发生频率为 10.52% (1SSR/4.46kb, 下同), 与 Yi 等^[12] 在辣椒上报道的发生频率 10.2% 基本一致. 该比率高于甘蓝的 6.22% (11.48 kb)^[8]、甜瓜的 8.36% (8.72 kb)^[7] 和西瓜的 6.89% (7.49 kb)^[18] 以及支莉等^[16] 在辣椒上的报道 (1.92%). 但低于黄瓜的 12.79% (3.80 kb)^[6] 和油菜的 15.58% (4.34 kb)^[19], 尤其是低于蓖麻的 40.85% (1.77kb)^[20] 和木薯的 25.12% (2.42 kb)^[21]. 通常来说, SSR 发生频率与距离成反比. 这种频率的差异可能源于各物种在 GenBank 中可用来分析的 EST 数量不同、或用来搜索 SSR 的软件算法的不同及所设定参数的不同, 也可能源于物种间真实的 SSR 信息差异. 相信随着 EST 数据库的容量不断增大, 各个物种间的 SSR 频率比较将有一个统一的标准.

在搜索出的 3 396 个辣椒 EST-SSR 中, 二核苷酸和三核苷酸重复基元占主导地位, 两者的总和占全部 SSR 的 80.87%, 与大多数作物如小麦^[22] (二、三核苷酸重复基元总和占 85.70%, 下同)、大豆^[23] (78.97%)、辣椒^[12,16] (84%, 79.7%) 等报道的一致. 但不同的物种间优势重复基元的类型有所差异. 理论上, 如果 EST 数足够大且无偏倚性, 则 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^[24]. 辣椒 EST-SSR 分析的结果显示, 不同重复基元出现与否及其频率高低表现出明显的偏倚性. 本研究搜索获得的辣椒 EST-SSR 中二核苷酸和三核苷酸优势重复基元分别为 GA 和 AGA, 与前人^[12,14-16] 的报道一致.

本研究共设计了 420 对辣椒 EST-SSR 引物, 通

过对辣椒抗/感疫病 2 个自交系进行扩增,发现共有 403 对引物有较好的扩增产物,引物有效扩增率为 95.96%,其中 76 对引物检测到多态性,占可扩增引物的 18.86%。可见,辣椒与梨^[25]、核桃^[26]等其他植物相比,EST-SSR 引物有效扩增率高于梨的 64.58% 和核桃的 90.9%。说明本研究开发的辣椒 EST-SSR 引物的有效性较高。而多态性引物所占比例比梨的 56.25% 和核桃的 70.00% 都低,可能源于所用的材料少或材料自身的遗传差异不明显。仍有极少数开发的引物没有扩增产物,可能是因为用于引物设计的 EST 序列所对应的基因组序列含有内含子。

从本研究结果可以看出,辣椒 EST-SSR 出现频率高,而且类型丰富。从多态性潜能的角度考虑,搜索到的这些 EST-SSR 也具有较高的可用性。研究的结果对于加速辣椒 EST 资源的开发利用、丰富其分子标记类型、遗传资源评价、遗传图谱构建、实现特定性状的分子标记辅助选择和进行比较基因组学研究均有一定的参考价值。

参考文献:

- [1] POWELL W, MACHRAY G C, PROVAN J. Polymorphism revealed by simple sequence repeats [J]. Trends Plant Science, 1996, 7: 215-222.
- [2] EUJAYL I, SORRELLS M E, BAUM M, et al. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat [J]. Theor Appl Genet, 2002, 104: 399-407.
- [3] YU J K, DAKE T M, SINGH S, et al. Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat [J]. Genome, 2004, 47: 805-818.
- [4] HAN Zhi-guo, WANG Chang-biao, SONG Xian-liang, et al. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton [J]. Theor Appl Genet, 2006, 112: 430-439.
- [5] 洪彦彬,陈小平,刘海燕,等. 源于大豆 EST 的花生属 (*Arachis*) 同源 SSR 标记的开发及利用[J]. 作物学报, 2010, 36(3): 410-421.
- [6] 胡建斌,李建吾. 黄瓜基因组 EST-SSRs 的分布规律及 EST-SSR 标记开发[J]. 西北植物学报, 2008, 28(12): 2429-2435.
- [7] 胡建斌,李建吾. 甜瓜 EST-SSR 位点信息及标记开发[J]. 园艺学报, 2009, 36(4): 513-520.
- [8] 陈琛,庄木,李康宁,等. 甘蓝 EST-SSR 标记的开发与应用[J]. 园艺学报, 2010, 37(2): 221-228.
- [9] 王西成,姜淑苓,上官凌飞,等. 梨 EST-SSR 标记的开发及其在梨品种遗传多样性分析中的应用评价[J]. 中国农业科学, 2010, 43(24): 5079-5087.
- [10] HUANG San-wen, ZHANG Bao-xi, DAN M, et al. Devel-

opment of pepper SSR markers from sequence databases [J]. Euphytica, 2000, 117: 163-167.

- [11] LEE J M, NAHM S H, KIM Y M, et al. Characterization and molecular genetics mapping of microsatellite loci in pepper [J]. Theor Appl Genet, 2004, 108: 619-627.
- [12] YI G, LEE J M, LEE S, et al. Exploitation of pepper EST-SSRs and an SSR-based linkage map [J]. Theor Appl Genet, 2006, 114: 113-130.
- [13] MINAMIYAMA Y, TSURO M, HIRAI M. An SSR-based linkage map of *Capsicum annuum* [J]. Mol Breeding, 2006, 18(2): 157-169.
- [14] NAGY I, STÁGEL A, SASVÁRI Z, et al. Development, characterization, and transferability to other *Solanaceae* of microsatellite markers in pepper (*Capsicum annuum* L.) [J]. Genome, 2007, 50(7): 668-88.
- [15] 李晶晶,王述彬,刘金兵,等. 辣椒 EST-SSR 标记的开发[J]. 分子植物育种, 2008, 6(6): 1219-1222.
- [16] 支莉,王述彬,刁卫平,等. 辣椒 EST-SSRs 分布特征及其应用[J]. 江苏农业学报, 2010, 26(6): 1323-1328.
- [17] 王关林,方宏筠. 植物基因工程[M]. 2 版. 北京:科学出版社, 2002: 742-744.
- [18] 赵胜杰,刘文革,阎志红,等. 西瓜 EST 资源的 SSR 信息分析[J]. 中国蔬菜, 2009, 24: 47-51.
- [19] 李小白. 油菜 *Brassica napus* L. EST-SSR 标记的开发及应用研究[D]. 杭州:浙江大学, 2007.
- [20] QIU Li-jun, YANG Chun, TIAN Bo, et al. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean *Ricinus communis* L. [J]. BMC Plant Biology, 2010, 10: 278-287.
- [21] KUNKEAW S, YOOCHA T, SRAPHET S, et al. Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihot esculenta* Crantz) [J]. Mol Breeding, 2011, 27: 67-75.
- [22] 陈军方,任正隆,高丽锋,等. 从小麦 EST 序列中开发新的 SSR 引物[J]. 作物学报, 2005, 31(2): 154-158.
- [23] 常玮,赵雪,李侠,等. 大豆 EST-SSR 标记开发及与 Genomic-SSR 的比较研究[J]. 中国油料作物学报, 2009, 31(2): 149-156.
- [24] 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.
- [25] 王西成,姜淑苓,上官凌飞,等. 梨 EST-SSR 标记的开发及其在梨品种遗传多样性分析中的应用评价[J]. 中国农业科学, 2010, 43(24): 5079-5087.
- [26] 朱艳,郝艳宾,王克建,等. 核桃 EST-SSR 信息分析与标记的初步建立[J]. 果树学报, 2009, 26(3): 394-398.