

从棉花 ESTs 数据库中筛选微卫星标记的初步研究

李华盛¹, 范术丽², 沈法富^{1*}

(1. 山东农业大学农学院, 山东泰安 271018; 2. 中国农业科学院棉花研究所
农业部棉花遗传改良重点开放实验室, 安阳 455000)

摘要:利用生物信息学方法在含有 94709 条棉花 EST 序列的数据库中进行微卫星标记筛选, 共发现微卫星序列 4396 个, 占整个 EST 数据库的 4.64%。其中双碱基重复序列 1282 个、三碱基序列 2411 个, 分别占在 EST 数据库中发现微卫星序列总数的 29.27% 和 54.8%。根据筛选到的微卫星序列设计并合成引物 25 对, 其中 24 对引物有扩增产物, 17 对产物条带比较清晰, 对这些条带比较清晰的产物作了一系列的分析。

关键词:棉花; 表达序列标签; 微卫星; 通用性分析

中图分类号: S562.035.3 **文献标识码:** A

文章编号: 1002-7807(2005)04-0211-06

Screening of microsatellite markers from Cotton ESTs

LI Hua-sheng¹, FAN Shu-li², SHEN Fa-fu^{1*}

(1. Department of Agronomy, Shandong Agricultural University, Taian, Shandong 271018, China; 2. Cotton Research Institute, CAAS; Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang, Henan 455000, China)

Abstract: Approximately 94709 EST sequences were searched for identification of SSRs using the Simple Sequences Repeat Identification Tool (SSRIT). 4396 microsatellite sequences were obtained from 94709 EST sequences, about 4.64% of the whole database a total of 1282(29.2%) dinucleotide repeats and 2411(54.8%) trinucleotide repeats were found in these microsatellite sequences. 25 primer pairs were designed from microsatellite. Among the 25 primer pairs, 17 can achieve clear PCR products by electrophoresis. The SSR markers reported in this study were highly polymorphic between *Gossypium hirsutum* and *Gossypium arboreum* but lower in *Gossypium hirsutum*. In order to evaluate transferability of EST-SSRs, we also selected six species: *Gossypium hirsutum*, *Zea mays*, *Lycopersicon esculentum*, *Brassica napus*, *Beta vulgaris*, *Abutilon theophrasti*, to detect the product. Among the product the primers Cri1, Cri3, Cri4, Cri7, Cri9, Cri14 and Cri15 have products in all of these six species. Cri16, Cri17 have no product in *Z. mays*, *L. esculentum* and *B. vulgaris*. And Cri13 has no product in *L. esculentum* and *B. vulgaris*.

Key words: cotton; EST; microsatellite; cross-species transferability

微卫星(microsatellite)又称 SSR,指的是以 1~6 个核苷酸为基本单元的串联重复序列。它普遍存在和随机分布于大多数真核生物的基因组中,重复单位的数目高度变异,但某一特定微卫星 DNA 的侧翼序列通常是高度保守的单一序列,根

据侧翼序列设计引物,即可通过 PCR 技术分析微卫星的多态性。由于 SSR 标记具有数量丰富、分布广泛、多态性高、共显性遗传和具有复等位性等优点,已在棉属进化^[1]、遗传多样性分析^[2]和分子遗传图谱构建^[3]与基因定位^[4-5]等方面得到广泛

收稿日期:2005-03-17

作者简介:李华盛(1981-),男,在读硕士研究生; *:通讯作者

基金项目:国家 863 计划项目(2004AA241080)

应用。

ESTs 即表达序列标签 (expressed sequence tags), 是来自随机选取的特定处理的 cDNA 克隆末端序列, 它能反映 mRNA 的信息, 在功能基因组研究中具有重要意义。传统基因组来源的 SSR 标记的开发需要构建基因组文库、探针杂交和克隆测序等复杂的操作程序。不仅费时费力, 而且开发成本很高。ESTs 是基因表达序列, 与构建小片段基因组文库进行微卫星序列筛选相比, 从 EST 数据库中获得微卫星序列要经济得多, 而且用 EST-SSRs 构建遗传连锁图, 就相当于定位功能已知的基因, 这将可能对决定重要表型的等位基因进行直接鉴定, 同时作物的比较基因组及功能基因组研究有重要意义。从 EST 筛选微卫星标记的工作, 在许多物种如小麦^[6]、云杉^[7]、苜蓿^[8]、甘蔗^[9] 等中已见报道。

目前, 虽然开发出的棉花 SSR 引物已经超过 1000 对, 但是从 EST 中开发 SSR 引物还未见报道。本研究从棉花 ESTs 数据库中进行微卫星序列的查找和分析, 设计引物, 筛选微卫星标记, 并对有清晰主带的引物做了通用性分析。

1 材料和方法

1.1 棉花 ESTs 中微卫星序列的查找

研究中的棉花 ESTs 序列下载于 NCBI 的 dbEST 数据库, 共搜索到 94709 条 ESTs 序列。使用 SSRIT 和 Repeatmasker 等软件从这些 EST 中进行微卫星序列的查找。查找的标准是 7 次以上重复的双碱基序列, 5 次以上重复的三碱基序列, 4 次以上重复的四碱基序列和 3 次以上重复的五碱基序列与六碱基序列。

1.2 微卫星引物设计

从获得的微卫星序列中, 选取合适的序列利用软件 Primer primer5.0 进行引物设计, 选择的

标准是: 10 次以上重复的双碱基重复序列, 7 次以上重复的三碱基序列, 5 次以上重复的四碱基和五碱基序列和 4 次以上重复的六碱基序列。引物长度在 18~24 bp 之间, 最好是 20 bp; T_m 值控制在 58 °C 左右, 最好是 60 °C; 并且尽量避免二级结构, 产物长度控制在 125~300 bp 之间。引物由上海生工生物工程技术服务有限责任公司合成。

1.3 模板 DNA 的提取

选取陆地棉 (*Gossypium hirsutum*) (TM-1)、海岛棉 (*Gossypium barbadense*) (海 7124)、亚洲棉 (*Gossypium arboreum*) (石系亚 1 号) 和非洲棉 (*Gossypium herbaceum*) (阿非利加棉) 四个棉种, 模板 DNA 提取材料来自棉花叶片, 参考沈法富等的方法提取^[10], 并做了一些优化。同时还提取玉米 (*Zea mays*)、番茄 (*Lycopersicon esculentum*)、油菜 (*Brassica napus*)、甜菜 (*Beta vulgaris*) 和苘麻 (*Abutilon theophrasti*) 的 DNA。使用前经紫外分光光度计检测浓度, -20 °C 保存备用。

1.4 PCR 扩增与电泳检测

PCR 反应体系 (10 μl) 包括 1 个单位 Taqase, 1 × PCR Buffer, 20 ng 基因组 DNA, 3 mmol MgCl₂ 200 μmol 的 dNTP, 0.2 mmol 引物。PCR 反应程序为: 95 °C 变性 10 min, 然后 94 °C 变性 50 s, 58 °C 变性 50 s, 72 °C 延伸 90 s, 共 40 个循环, 之后 72 °C 延伸 10 min, 4 °C 保存。

各对引物的最适温度通过梯度 PCR 结果而定, 对 PCR 产物用 6% (w/v) 聚丙烯酰胺凝胶 (6 mol · L⁻¹ 尿素) 电泳和银染检测其多态性。

2 结果与分析

2.1 从 ESTs 数据库中筛选微卫星序列

在 94709 条 ESTs 序列中, 经过计算机筛选结合人工选择共发现微卫星序列 4396 个, 占整个

表 1 从棉花 EST 数据库中筛选所得的微卫星序列

Table 1 Microsatellite sequences isolated from ESTs database of cotton

重复类型	数量	百分比	重复类型	数量	百分比
ct	90	2.06	ctc	62	1.41
tc	157	3.58	acc	58	1.32
ta	304	6.92	ctt	125	2.84
ga	86	1.96	tct	122	2.78
at	429	9.76	tga	90	2.05
ag	80	1.82	ttc	148	3.37
cat	86	1.96	ata	74	1.69
aag	97	2.21	gat	97	2.21
cac	76	1.73	tca	126	2.87
tcaggc	73	1.66			

EST 数据库的 4.64%, 其中双碱基重复序列 1282 个、三碱基序列 2411 个, 分别占在 EST 数据库中发现微卫星序列总数的 29.3% 和 54.8%。在这些微卫星序列中 (AT)_n 形式在双碱基重复中最常见, (TTC)_n 形式在三碱基重复中最常见, 分别占微卫星序列总数的 9.76% 和 3.37% (表 1)。

对有扩增产物的 EST 序列进行查找, 发现所设计 EST-SSR 引物对应 ESTs 序列的 mRNA 均取自棉花开花后十天内, 可能与纤维发育有关, 同

时用 BLAST 软件进行多核苷酸序列比对, 其中, 引物 Cri1 对应的 EST 与编码棉花 E6 基因的序列有关, 有 98% 的碱基相同。

2.2 引物设计和结果检测

根据以上的微卫星序列, 共选取 25 条微卫星序列进行了引物设计并合成, 其中 4 对引物来源于双碱基重复序列, 16 对来源于三碱基重复序列, 3 对来源于六碱基重复序列, 还有 2 对来源于四碱基重复序列 (表 2)。

表 2 棉花 EST-SSR 引物及其评价

Table 2 The microsatellite primers of cotton and their evaluation

引物编号	引物序列(5'~3')	核心序列	温度/°C	多态性信息含量
Cri1	AAC AAg AgC CAA ggT TCA TC TgA CAT Agg gTT CAT Agg TTT	(tcaggc) ₄	58	0.48
Cri2	CAG TCT CCg TgA gCg TTT g AgC CTg CTTC AAT ACT CCA A	(aaagga) ₄	58	0.44
Cri3	Cgg Agg AgA AAC TCT gAC gA ACA CCC ACA ACA gCA AgCA g	(aaaaat) ₄	60	0
Cri4	CgC TTA TCg TTT CTT CgC CAA TgC CAT CCA AAT CCC	(tct) ₁₅	56	0
Cri5	TgA AAC AAg CCA CTg TAA CTC AAg ATC CCT CCA TAA CCA AA	(ttc) ₁₀	58	0.44
Cri6	gCT CAg CCA ATA ACA CCA AAg ggC TAT TAA gCA ACT CA	(gaa) ₁₀	58	0
Cri7	AAg gAg ggC gAT AAA CAT AA TCg ggT AgT gCT ggA gATT	(cac) ₁₁	58	0.75
Cri8	CAA TgT AAA CCA Agg AgC Ag TgA gCA AAg ggT ATA gCA gT	(gca) ₈	58	0.75
Cri9	TTC ACC CAC CAC CCA gAC CA ACA ACA CCg CCA gCA gCA AT	(ctt) ₇	58	0.44
Cri10	CAC AAT CTC CAA TATC CCA CC TgA TgC TgA TgT AgA CCC AAT	(cat) ₇	58	0
Cri11	gTT ggA gTT ggg TCT Tgg TT AgA Agg gAA ATC TTT AgC AgT C	(tgg) ₈	58	0.44
Cri12	TCC TgC CgA CCC TAT TCC T CCA CCg ACC TgA CCg TTT T	(cag) ₇	60	0
Cri13	TgT TTC TgT ATC TTC ACA ACC CTA CgT CgC TTC TgA gAT TCg T	(gcc) ₇	58	0.48
Cri14	ggA CAT TgT ggA ATC ggT AT AgA ACA gCC TTg AgT TgA gC	(atg) ₇	58	0.44
Cri15	TCA CCC ACT ACC TTC TgC CTT AC gAC AgC Agg CTA TTT CTT gAC C	(ca) ₁₇	58	0.44
Cri16	CAA Tgg CAA Tgg TgA TAg g gAC AAC ggT gAA ATg AAg C	(ta) ₁₁	58	0.75
Cri17	TTC ATT ggC TgC TTT CgT AT gCT Tgg TgA CTC CAG gTT TT	(tc) ₁₂	58	0.44

经过 PCR 扩增,聚丙烯酰胺凝胶电泳和银染检测,有 24 对引物有产物,其中有 17 对引物产物清晰。将有清晰产物的引物在四个棉种中进行 PCR 检测,均有清晰产物,扩增出的条带数目均

为一条或两条。但是,有的同一个引物扩增产物的条带不同,在陆地棉中扩增出两条带,在亚洲棉中只扩增出一条带,初步推测可能与染色体倍数有关(图 1)。

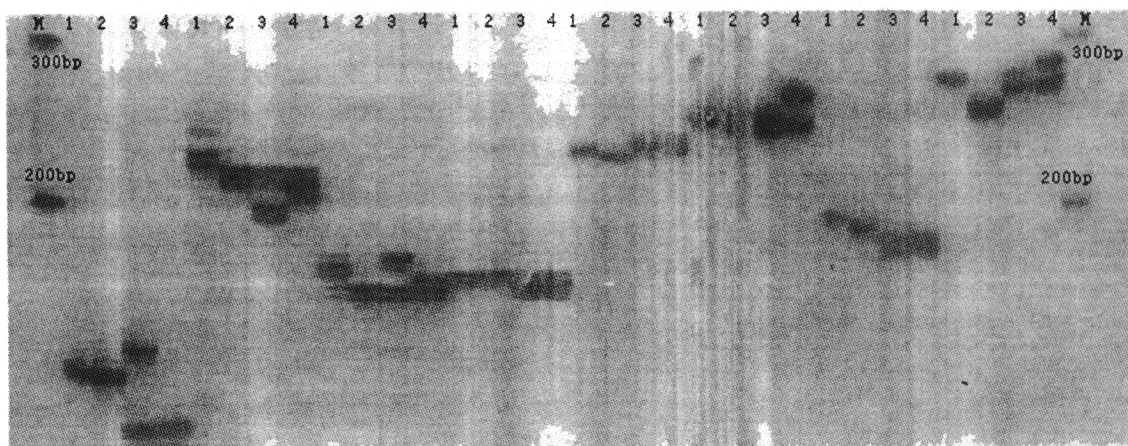


图 1 部分微卫星位点电泳图片

Fig. 1 The silver-stained polyacrylamide gel image of the SSR alleles amplified in cotton

注:从左到右引物号分别为 Cri5、Cri7、Cri8、Cri10、Cri13、Cri14、Cri16、Cri17;

其中 1 是亚洲棉,2 是非洲棉,3 是陆地棉,4 是海岛棉。

Note: From left to right, the No. of primer is Cri5、Cri7、Cri8、Cri10、Cri13、Cri14、Cri16、Cri17.

1. *G. arboreum*, 2. *G. herbaceum*, 3. *G. hirsutum*, 4. *G. barbadense*

2.3 通用性分析

左开井等曾经利用玉米 SSR 引物对棉花做了通用性分析,结果表明有 91% 的引物能扩增出产物^[11]。EST-SSRs 来源于基因转录编码区,而一些参与生物基本生理生化代谢、基本生命过程的基因在物种间具有很高的保守性。有鉴于此,我们用在 4 个棉种中均有清晰条带产生的引物对

油菜、苘麻、玉米、番茄、甜菜等进行了通用性分析,大部分引物有扩增产物,但是同一引物在不同物种中的扩增带数和产物长度有较大差异(图 2)。其中引物 Cri1、Cri14 在各物种间没有多态性,Cri16、Cri17 在玉米、番茄和甜菜中均没有产物,Cri13 也在番茄和甜菜上没有扩增产物,各引物的扩增情况见表 3。

表 3 棉花 EST-SSR 引物的通用性分析

Table 3 The analysis of EST-SSRs' cross-species transferability

	油菜	苘麻	玉米	番茄	甜菜	陆地棉	是否有多态性
Cri1	2	2	2	2	2	2	无
Cri3	5	3	2	2	5	2	有
Cri4	3	4	3	2	2	2	有
Cri7	1	1	1	2	2	2	有
Cri19	1	2	2	1	1	2	有
Cri13	1	1	1	0	0	1	无
Cri14	2	2	2	2	2	2	无
Cri15	2	2	2	2	2	2	有
Cri16	1	1	0	0	0	1	无
Cri17	2	2	0	0	0	2	有

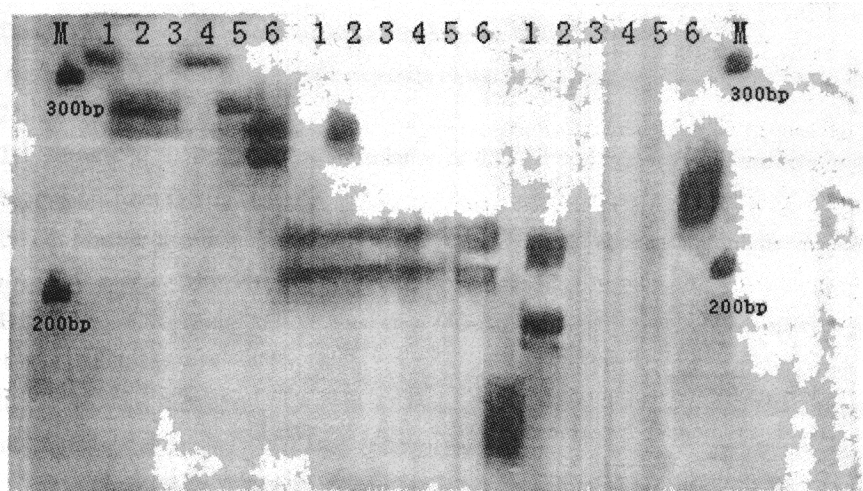


图 2 部分通用性分析电泳图片

Fig. 2 The silver-stained polyacrylamide gel image of some primers' cross-species transferability

注:从左向右引物分别为 Cri9、Cri14、Cri17;其中 1 油菜, 2 苘麻, 3 玉米, 4 番茄, 5 甜菜, 6 陆地棉。

Note: From left to right, the NO. of primer is Cri9, Cri14, Cri17; 1 *B. napus*, 2 *A. theophrasti*, 3 *Z. mays*, 4 *L. esculentum*, 5 *B. vulgaris*, 6 *G. hirsutum*.

3 结论与讨论

Eujay 等对苜蓿 (*Medicago spp.*) EST 数据库搜索, 在 147000 个 ESTs 中发现了 4384 个微卫星序列, 约占序列总数的 2.98%^[8]。Gupta 等在 15000 条面包小麦 (*Triticum aestivum* L.) 的 ESTs 中筛选出了 900 条微卫星列, 比例为 6.00%^[12]。Saha 等的研究表明棉花 cDNA 数据库中 1.1%~4.8% 的 EST 中含有 SSR 重复序列^[1]。徐鹏等在中国对虾 10000 多个 EST 中找到 229 个 SSRs, 约占 2.19%^[13]。Kantey 等发现在各个物种中的 ESTs 序列中, 简单重复序列的含量在 1%~5% 左右。本实验共在 94709 条棉花 EST 序列中发现各种类型的微卫星序列 4396 个, 占整个 EST 数据库的 4.64%。这个比例与其它物种筛选出的微卫星比例大体相符。

实验中, 最普遍的微卫星序列是 (AT)_n 序列, 这与 Morgante 和 Olivieri 的查询结果一致^[14]。但是本实验中重复序列的比例中三碱基最高, 约占 54.8%, 而不是双碱基序列。这虽然和传统的观点有点不附, 但与 Eujay (苜蓿)^[8] 和 Gupta (面包麦)^[12] 的研究结果一致。这说明在植物 ESTs 中分布最广泛的微卫星形式可能是三碱基重复, 而不是双碱基形式。

为了避免高丰度表达基因的 ESTs 被重复选中, 每一种形式的微卫星重复单元只选择一条 EST 序列进行引物设计, 使设计的引物更具代表

性; 同时选择具有五碱基或六碱基的 EST 进行引物设计, 以保证有较高的多态性。在引物设计中, 引物碱基数被设定在 18~24 之间, 各引物退火温度均在 58℃ 左右, 引物控制在 125~300 bp 之间, 这些都保证了 PCR 扩增的稳定性和准确性。

虽然在棉花 EST 数据库中, 微卫星资源并不丰富, 且大量的序列不完整, 大部分不适合引物设计, 但是从 EST 中发掘 SSR 标记可以节省大量时间和经费, 同时也是对 EST 数据库资源的充分利用, 为功能基因组和比较基因组的研究打下基础。

参考文献:

- [1] SAHA S, Kanaca M, Jenkins J N, et al. Simple sequence repeats as useful resources to study transcribed genes of cotton[J]. *Euphytica*, 2003, 130:355-364.
- [2] BASU S, Johnie N J, Douglas B S. Use of markers for analysis of genetic relationship among improved cotton cultivars and its association with F₂ hybrid performance[A]. *Plant & animal genome IX conference*, 2001.
- [3] NGUYEN T-B, Giband M, Brottier P, et al. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers [J]. *Theor Appl Genet*, 2004 109:167-175.
- [4] BEZAWADA C, Saha S, Jenkins J N, et al. SSR Marker(s) associated with root knot nematode resistance gene(s) in Cotton[J]. *The journal of cotton sci-*

- ence, 2003,7:179-184.
- [5] KARACA M, Saha S, Jenkins J N, et al. Simple sequence repeat (SSR) markers linked to the Ligon lintless (Li(1)) mutant in cotton[J]. *J hered*, 2002 May-Jun,93(3):221-224.
- [6] ENJAUL 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.
- [7] SCHUBERT R, Müller-Starck G, Riegel R. Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea bies* (L.) Karst[J]. *Theor appl genet*, 2001, 103:1223-1231.
- [8] ENJAUL I, Sledge M K, Wang L, et al. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp[J]. *Theor appl genet*, 2004, 108(3):414-422.
- [9] 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 sci*, 2001 May, 160(6):1115-1123.
- [10] 沈法富,于元杰,刘凤珍,等. 棉花核 DNA 的提取及其 RAPD 分析[J]. *棉花学报*,1996,8(5):246-249.
- [11] 左开井. 转基因抗虫棉 Bt 基因的遗传效应及定位研究[D]. 华中农业大学, 2000.
- [12] GUPTA P K, Rustgi S, Sharma S, et al. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat [J]. *Mol gen genomics* (2003) 270: 315-323.
- [13] 徐 鹏,周令华,田丽萍,等. 从中国对虾 ESTs 中筛选微卫星标记的研究[J]. *水产学报*,2003,27(3): 213-218.
- [14] MORGANTE M, Olivieri A M. PCR amplified microsatellite as markers in plant genetics[J]. *Plant J*, 1993,3:175-182. ●