

基于陆地棉茎尖转录组序列 genic-SSR 标记的开发与评价

高 阳^{1,2}, 王芙蓉¹, 刘国栋¹, 张景霞¹, 张传云¹, 张 军^{1,2}

(¹农业部黄淮海棉花遗传改良与栽培生理重点实验室/山东棉花研究中心, 济南 250100; ²山东师范大学生命科学学院, 济南 250014)

摘要: 利用软件 MISA 和 SSR Locator 对陆地棉 (*Gossypium hirsutum* L.) Coker312 茎尖转录组测序得到的 73515 条 Unigene 序列进行分析, 查找到 4507 个 SSR 位点, 分布于 4039 条转录本序列中, SSR 位点出现频率为 5.5%, 非编码区的 SSR 位点要显著高于编码区 (2853/1654)。SSR 重复基元中, 三核苷酸重复基元占主导地位 (51.03%), 且 AAG/CTT (20.08%) 类型最多; 其次是二核苷酸重复基元 (28.76%), 主要为 AG/CT (55.47%)。利用引物批量设计软件共开发 1569 对 SSR 引物, 在陆地棉 TM-1、海岛棉 3-79 (*G. barbadense* L.)、阿非利加棉 (*G. herbaceum* L.)、雷蒙德氏棉 (*G. raimondii* Ulbrich.) 4 个棉种中进行初步评价, 其中 1117 对引物能在 4 个棉种间扩增出稳定的条带, 扩增率为 71.2%。选择 650 对能稳定扩增出条带的引物, 在涉及 5 个棉种 (增加了亚洲棉 (*G. arboreum* L.)) 的 13 份材料中进一步筛选, 83 对引物能在 13 份材料中扩增出特异性条带, PIC 变幅在 0.121~0.648 之间, 平均值为 0.422, 其中 80 对引物能在 7 份陆地棉材料中扩增出特异性条带, 平均 PIC 值为 0.336; 挑选 5 对在本实验室作图亲本鲁棉研 22 和鲁原 343 间有明显多态的引物进行连锁分析, 其中 4 对成功连锁到本实验室构建的遗传图谱上。本研究丰富了棉属 genic-SSR 的数量, 为棉属遗传作图、性状关联分析提供了更多引物选择。

关键词: 陆地棉; 茎尖转录组; genic-SSR; 多态性

Development and Evaluation of genic-SSR Markers Based on Shoot Tip Transcriptome Information of *Gossypium hirsutum* L.

GAO Yang^{1,2}, WANG Fu-rong¹, LIU Guo-dong¹,
ZHANG Jing-xia¹, ZHANG Chuan-yun¹, ZHANG Jun^{1,2}

(¹Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Ministry of Agriculture/Shandong Cotton Research Center, Ji'nan 250100; ²Academy of Biological Sciences, Shandong Normal University, Ji'nan 250014)

Abstract: A total of 4507 SSRs distributed in 4039 transcript sequences were identified from 73515 unigenes in *Gossypium hirsutum* shoot tip transcriptome. The frequency of unigene sequences containing SSR was 5.5% and the number of SSRs was significantly higher in non-coding regions than that in coding regions (2853/1654). The SSR repeat motifs dominated by trinucleotide repeats (51.03%), among which AAG/CTT (20.08%) accounted for the highest proportion. Dinucleotide repeats motifs (28.76%) were the next, and AG/CT (55.47%) were the dominant type. Using primers batch design software, 1569 pairs of genic-SSR primers were developed. Preliminary evaluation was conducted on four *Gossypium* species, TM-1 (*Gossypium hirsutum* L.), 3-79 (*G. barbadense* L.), *G. herbaceum* L. var. *africanum* and *G. raimondii* Ulbrich. A total of 1117 primer pairs amplified stable bands and the amplification rate was 71.2%. Six hundred and fifty pairs of primers were randomly selected and further screened in 13 accessions of five cotton species (plus *G. arboreum* L.). Eighty-three of these primers could amplify special bands,

收稿日期: 2014-07-23 修回日期: 2014-09-10 网络出版日期: 2015-06-11

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20150611.0936.005.html>

基金项目: 国家自然科学基金项目 (31171598); 转基因生物新品种培育重大专项 (2014ZX08005-003-004); “十二五”农村领域国家科技计划课题 (2013BAD01B03-03); 山东省农业良种工程项目农业生物资源创新利用课题

第一作者研究方向为植物发育遗传。E-mail: jlg0909@163.com

通信作者: 张军, 研究方向为分子育种。E-mail: scrczj@saas.ac.cn; zj0928@126.com

the *PIC* value ranged from 0.121 to 0.648, with an average of 0.422. Eighty pairs of primers displayed polymorphisms within *G. hirsutum*, and the average *PIC* value was 0.336. Five among the 83 genic-SSR primer pairs which distinctly revealed polymorphism between the laboratory mapping parents of Lumianyan22 and Luyuan343 were selected and carried on linkage analysis, four of the SSRs were successfully integrated into intraspecific genetic linkage map previously constructed in our lab. The results enriched the quantity of cotton genus genic-SSR for mapping and association analysis and related research on cotton species.

Key words: *Gossypium hirsutum* L.; shoot tip transcriptome; genic-SSR; polymorphisms

棉花是重要的经济作物之一,也是最重要的纺织纤维作物。棉属包含 52 个种,其中栽培种 4 个,包括 2 个二倍体棉种,即草棉 (*Gossypium herbaceum* L.) 和亚洲棉 (*G. arboreum* L.) 以及 2 个四倍体棉种,即陆地棉 (*G. hirsutum* L.) 和海岛棉 (*G. barbadense* L.)^[1]。陆地棉因产量高、适应性强而成为生产上的主要栽培种,但由于陆地棉遗传基础狭窄,产量和品质进一步改良的难度较大。而棉属其他种内蕴含优异纤维、抗黄萎病等有益基因资源,但由于种间杂交困难或连锁累赘等问题,仅依赖传统育种技术很难将外源的优异基因引入到陆地棉基因库中。

随着基因组学的发展,分子标记成为改善这一问题的首选方法。SSR 标记因其在基因组中分布广泛且具有共显性遗传、可重复性、多态性丰富等特点而被认为是最好的分子标记之一,广泛应用于种质资源的多样性分析、遗传连锁图谱构建、基因定位及分子标记辅助育种等研究^[2]。SSR 标记的开发成本较高,在很多作物中的应用受到了限制。近年来兴起的高通量测序技术,已在许多作物中开发出大量的转录组数据,不仅可以用于新基因的鉴定和表达谱分析,还可以用于新 SSR 标记的开发。与传统 SSR 标记相比,基于转录组数据开发的 SSR 标记,均与功能基因相关联,在分子标记辅助育种中具有重要意义。萝卜 (*Raphanus sativus* L.)^[3]、辣椒 (*Capsicum annuum* L.)^[4]、芝麻 (*Sesamum indicum* L.)^[5] 等作物中转录组数据的挖掘,为其识别和开发新标记提供了可靠性。利用棉花转录组序列开发分子标记已有报道,来德勇^[6] 利用陆地棉发育相关的转录组数据发掘了 670 个新的 EST-SSR 标记;陈浩东^[7] 从达尔文氏棉转录组序列中开发了 780 对 EST-SSR 引物,并详细分析了 EST-SSR 位点的特征,引物的多态性及可转移性;吕远大等^[8] 基于海岛棉纤维发育 EST 序列开发了 380 对棉花 SSRs 引物;S. N. Jena 等^[9] 在草棉转录组 99780 个 Unigene 中,查找出 12471 个 SSR 位点,分布于 8900 条 EST 中。目前,棉花中基于转录组开发的 SSR 位点主要与纤维发育基因相关,与棉花早期形态

建成相关的基因 SSR 标记开发尚未见报道。

茎端组织与植物叶原基、芽原基的发生以及植物顶端生长优势密切相关。本研究以陆地棉 Coker312 茎尖转录组 73518 条 Unigene 序列为材料,分析了陆地棉转录组 SSR 位点的分布特征,开发出 1569 对 genic-SSR 引物,在 4 个棉种间进行初步筛选的基础上,随机选取稳定扩增的引物在包括陆地棉 (*G. hirsutum*)、海岛棉 (*G. barbadense*)、草棉 (*G. herbaceum*)、雷蒙德氏棉 (*G. raimondii* Ulbrich.)、亚洲棉 (*G. arboreum*) 等 5 个种共 13 份材料中进行多态性评价,以期对种质资源遗传多样性的分析、品种纯度和真实性鉴定、分子标记辅助育种等研究提供新的数据。

1 材料与方法

1.1 材料

选取 5 个棉种共 13 份材料(表 1),用于评价新开发的 genic-SSR 引物,分析引物的有效性、通用性、多态性等。供试材料中阿非利加棉和雷蒙德氏棉活体保存于山东棉花研究中心温室,其他材料的种子均保存于山东棉花研究中心分子育种课题组种质资源库。

表 1 用于 SSR 分析的不同棉种及品种
Table 1 Cotton varieties and wild species for SSR analysis

编号 No.	材料名称 Accession	种名 Species
1	TM-1	<i>G. hirsutum</i>
2	Coker312	<i>G. hirsutum</i>
3	Acala SJ-5	<i>G. hirsutum</i>
4	中棉所 12 号	<i>G. hirsutum</i>
5	中棉所 10 号	<i>G. hirsutum</i>
6	鲁棉研 22	<i>G. hirsutum</i>
7	鲁原 343	<i>G. hirsutum</i>
8	3-79	<i>G. barbadense</i>
9	海 7124	<i>G. barbadense</i>
10	阿须莫尼	<i>G. barbadense</i>
11	石系亚 1 号	<i>G. arboreum</i>
12	阿非利加棉	<i>G. herbaceum</i>
13	雷蒙德氏棉	<i>G. raimondii</i>

1.2 方法

1.2.1 高通量测序 以陆地棉 Coker312 为材料,使用 RNAprep Pure 植物总 RNA 提取试剂盒 (TIANGEN) 提取 5~6 片真叶期茎尖组织中的总 RNA。用反转录试剂盒 PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time) (TAKARA) 合成 cDNA 第 1 链, -80 °C 保存。送往深圳华大基因构建测序文库,使用 Illumina HiSeq™ 2000 进行测序。

1.2.2 DNA 的提取、PCR 扩增及电泳检测 取成年植株的幼嫩叶片 100 mg,放入液氮中保存。采用改良的 CTAB 法^[10] 提取棉花基因组 DNA。PCR 反应体系 10 μL,包括 1.0 μL 10 × Taq Buffer (含 Mg²⁺), 0.2 μL (10 mmol/L) dNTP, 0.1 μL (2.5U) Taq 酶, 10 μmol/L 上下游引物各 0.6 μL, 2.0 μL (20~50 ng/μL) DNA 模板,加无菌蒸馏水至 10 μL。反应程序为:94 °C 预变性 3 min;94 °C 变性 45 s, (T_m) °C 退火 45 s, 72 °C 延伸 45 s, 32 个循环, 72 °C 终延伸 7 min。采用 8% 的聚丙烯酰胺凝胶电泳及参照张军等^[11] 银染方法进行检测。

1.2.3 转录组 EST-SSR 位点的发掘及引物设计

基于对陆地棉茎尖组织进行转录组高通量测序结果中的 73518 条 Unigene,利用软件 MISA 和 SSR Locator 查找 SSR 位点,查找条件为:二核苷酸重复次数 ≥ 6,三核苷酸重复 ≥ 5,四核苷酸重复 ≥ 4,五核苷酸重复 ≥ 4,六核苷酸重复 ≥ 4,七至十核苷酸重复均 ≥ 3。以所得的 SSR 序列为中心,提取 SSR 上下游各 200 bp 序列,如果上游或下游序列不足 200 bp 则提取上游或下游的全部序列设计引物。以新设计的 SSR 引物为引物,以 CMD (<http://www.cottonmarker.org/>) 的 EST 序列为模板运行 ePCR,获得了无重复的 1569 对新 EST-SSR 引物,命名为 SCRCXXX (Shandong Cotton Research Center)。设计的主要参数为:引物长度 18~21 bp,最适长度 20 bp,PCR 产物长度 100~395 bp;最适 T_m 值 58 °C。引物由上海生工生物技术有限公司合成。

1.2.4 预测编码蛋白框 将陆地棉 Coker312 茎尖转录组中的 Unigene 序列分别于蛋白库 nr、Swiss-Prot、KEGG 和 COG 的优先级顺序做 blastx 比对,比对参数为 E-value < 0.00001。如果某个 Unigene 序列比对上高优先级数据库中的蛋白,则不进入下轮比对,否则自动跟下一个库做比对,如此循环直到跟所有蛋白库比对完。取

blast 比对结果中 rank 最高的蛋白确定该 Unigene 的编码区序列。对于和以上 4 个库皆比不上的 Unigene,用软件 ESTScan^[12] 预测其编码区并确定序列的方向。

1.3 多态性引物的筛选

利用新合成的 genic-SSR 引物首先在 4 个棉种 (Tm-1、3-79、阿非利加棉、雷蒙德氏棉) 间进行初步筛选,从中随机选出部分能扩增出稳定条带的引物,然后再用 13 份材料对这些多态性引物进行深入分析,最终选择出多态性明显、扩增清晰稳定的引物。

1.4 数据统计及多态性信息含量分析

有扩增条带赋值为“1”,无扩增条带赋值为“0”统计数据。多态性信息含量 (PIC, polymorphism information content) 被用来评估所设计引物的潜在利用率^[13]。本试验的每个 genic-SSR 引物的多态性信息含量,按公式 $PIC = 1 - \sum P_i^2$ 进行计算,其中 P_i 为第 i 个等位基因出现的频率^[14-15]。

1.5 染色体定位

利用本实验室前期构建的 (鲁棉研 22 × 鲁原 343) 重组自交系群体进行染色体初步定位;采用 Jion-Map 4.0 软件构建分子遗传连锁图谱 (LOD 值 = 6.5),以 Kosambi 函数作图;采用 Map-Chart 2.2 绘图软件绘制遗传图谱。

2 结果与分析

2.1 转录组 genic-SSR 标记的发掘及分布特征

2.1.1 genic-SSR 标记的发掘 转录组序列经拼接去重后共获得 73518 条 Unigene,总长度 52.3 Mb,利用软件 MISA 和 SSR Locator 查找出 4507 个 SSR 位点,位于 4039 条转录本序列中,其中超过 1 个以上 SSR 位点的转录本序列有 113 条 (2.79%),genic-SSR 出现频率为 5.5%,平均每 12.9 kb 出现 1 个 SSR。

2.1.2 genic-SSR 中重复基元的类型特征 在 4507 个 SSR 位点中,三核苷酸占主导地位,其次是二核苷酸,分别为 51.03% 和 28.76%;二、三、四、五、六核苷酸为常见核苷酸重复,分别以 AG/CT、AAG/CTT、TAAA/TTTA、CCGAA/TTCCG、TGCTCC/GGAGCA 类型最多 (图 1,表 2)。2~10 bp 的各种重复基元的 SSR 数目、频率及在各种类型中的优势基元所占的比例如表 2 所示。其中 2~6 bp 核苷酸重复占 96.76%,而 7~10 bp 的各种类型重复基元仅占 3.24%。

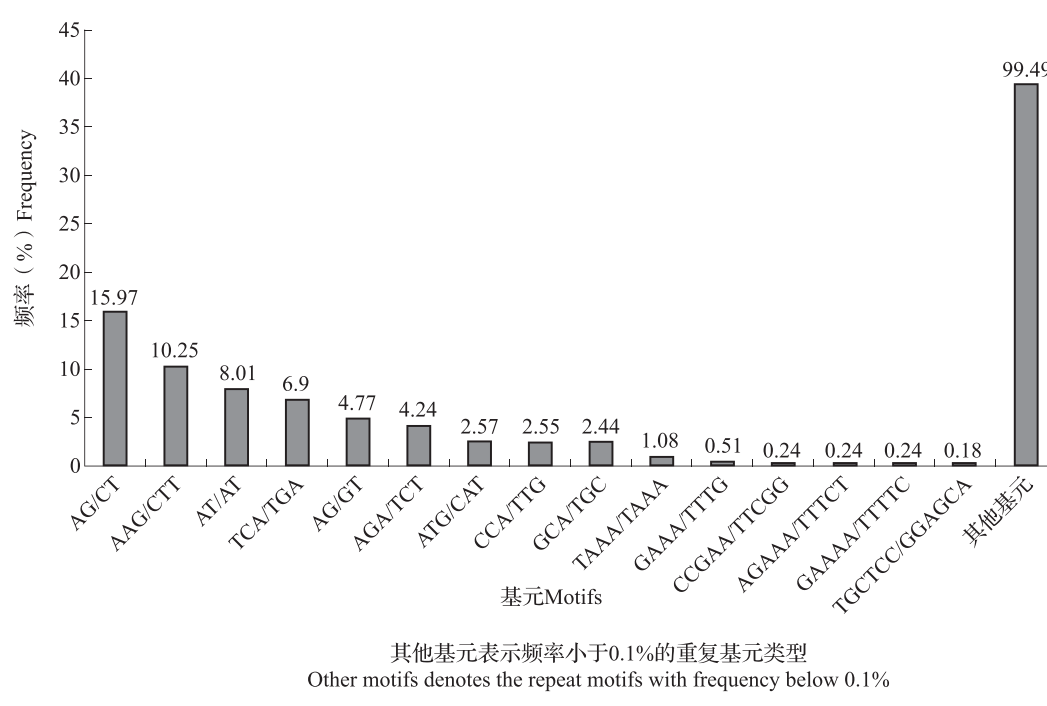


图 1 基于重复基元类型的 genic-SSR 分布
Fig. 1 The genic-SSR distribution on different repeat motifs

表 2 陆地棉茎尖转录组中不同 genic-SSR 重复基元出现的频率
Table 2 Occurrence frequency of differentgenic-SSR motifs of *G. hirsutum* shoot apex in transcriptome

重复单元类型 Repeat type	数量 Number	频率(%) Frequency	最多的重复单元及其数量和频率 Number and frequency of maximum repeat motif
二核苷酸 Dinucleotide repeats	1296	28. 76	AG/CT (719 ,55. 47%)
三核苷酸 Trinucleotide repeats	2300	51. 03	AAG/CTT(462,20. 08%)
四核苷酸 Tetranucleotide repeats	375	8. 32	TAAA/TTTA (49,13. 01%)
五核苷酸 Pentanucleotide repeats	194	4. 30	CCGAA/TTTCG (11,5. 67%)
六核苷酸 Hexanucleotide repeats	196	4. 35	TGCTCC/GGAGCA (8,4. 08%)
七核苷酸 Heptanucleotide repeats	110	2. 44	CGTATAC/GTATACG (8,7. 27%)
八核苷酸 Octanucleotide repeats	10	0. 22	AAAAAAGA/TCTTTTTT (2,20%)
九核苷酸 Nonanucleotide repeats	23	0. 51	TGCTGTTGC/GCAACAGCA (11,47. 83%)
十核苷酸 Decanucleotide repeats	3	0. 07	CGGCCACCGT/ACGGTGGCCG (1,33. 33%)
总计 Total	4507	100	-

2. 1. 3 SSR 位点在陆地棉茎尖转录组编码区中的分布特征 陆地棉茎尖转录组序列中,73518 条 Unigene 序列功能注释成功,其中能预测出编码框 (CDS) 的序列有 56619 条,占 77. 01%。共发现 4507 个 SSR 位点,其中位于非编码区的有 2853 个,位于编码区的 SSR 位点仅有 1654 个,非编码区 SSR

的数量比编码区高 72. 5%。在基因编码区中出现频率最高的是三核苷酸重复的 SSR 位点 1127 个, 占总编码区的 68. 14% ,二核苷酸仅占 17. 83% ;在非编码区中出现较多的 SSR 为二核苷酸重复基元和三核苷酸重复基元,其中三核苷酸重复基元仅比二核苷酸重复基元多 6. 03% (表 3)。

表 3 陆地棉茎尖转录组中蛋白编码区和非编码区中 SSR 的分布频率

Table 3 Occurrence frequency of SSR in protein coding regions and non-coding regions in *G. hirsutum* shoot tip transcriptome

重复单元类型 Repeat type	编码区 Protein-coding region		非编码区 Non-coding region	
	SSR 数量	百分比(%)	SSR 数量	百分比(%)
	No. of SSR	Percentage	No. of SSR	Percentage
二核苷酸 Dinucleotide repeats	295	17.83	1001	35.08
三核苷酸 Trinucleotide repeats	1127	68.14	1173	41.11
四核苷酸 Tetranucleotide repeats	74	4.47	301	10.55
五核苷酸 Pentanucleotide repeats	32	1.94	162	5.68
六核苷酸 Hexanucleotide repeats	82	4.96	114	3.99
七至十核苷酸 From Hepta-to Decanucleotide repeats	44	2.66	102	3.58
总计 Total	1654		2853	

2.2 genic-SSR 引物的筛选

为了检测新发掘 SSR 引物的可靠性,本研究首先选用 4 个不同棉种中具有代表性的材料: Tm-1 (*G. hirsutum*)、3-79 (*G. barbadense*)、雷蒙德氏棉 (*G. raimondii*)、阿非利加棉 (*G. herbaceum*), 对 1569 对新开发的引物进行初步评价。其中 1117 对引物能成功在 4 份材料之间扩增出清晰、稳定的条带, 引物扩增率为 71.2%。

2.3 genic-SSR 引物的通用性及多态性分析

选取 650 对扩增效果较好的引物, 在 13 份材料中进行通用性及多态性分析。其中 511 对能在 3 份海岛棉材料中扩增出稳定的条带, 通用性为 78.6%; 575 对引物能在亚洲棉 (石系亚) 中有效扩增, 扩增率为 88.5%; 650 对引物在雷蒙德氏棉和阿非利加棉中均能有效扩增。将 650 个标记按微卫星在基因上所处的位置进行划分: 252 个标记位于基

因编码区, 398 个标记位于非编码区; 其中 241 个位于基因编码区中的标记成功在海岛棉材料中扩增出稳定的条带, 通用性为 95.6%。在非编码区中有 270 个标记能在海岛棉材料中进行有效扩增, 通用性为 67.8%。

上述 650 对引物中, 83 对在 13 份材料之间存在明显差异, 多态率为 12.8%。多态性评价结果显示, 13 份材料的 PIC 值变幅为 0.121 ~ 0.648, 平均 PIC 值为 0.422, 其中 5 个棉种间的平均 PIC 值为 0.424, 明显高于陆地棉与海岛棉种间的平均 PIC 值 (0.395), 陆地棉种内的 PIC 值 (0.336) 高于海岛棉种内的 PIC 值 (0.282)。利用不同棉种和品种评价 83 对 SSR 引物时, 在海岛棉种内只有 23 对引物能扩增出差异性条带 (表 4), PIC 值变幅为 0.106 ~ 0.477。而只有 3 对 SSR 引物在 7 份陆地棉材料中没有多态性。

表 4 部分 genic-SSR 标记在 13 份材料种间及种内扩增的 PIC 值

Table 4 PIC-value of partial polymorphic primers amplified with cotton materials

引物名 Primer name	13 份材料 13 cotton materials	陆地棉种内 <i>G. hirsutum</i>	海岛棉种内 <i>G. barbadense</i>	陆地棉与海岛棉种间 <i>G. hirsutum</i> and <i>G. barbadense</i>	5 个物种之间 Among five species
SCRC0027	0.450	0.319	0.255	0.300	0.435
SCRC0033	0.331	0.225	0.318	0.337	0.375
SCRC0052	0.648	0.162	0.477	0.642	0.673
SCRC0056	0.587	0.441	0.106	0.561	0.535
SCRC0082	0.121	0.219	0.382	0.300	0.311
SCRC0100	0.451	0.403	0.382	0.456	0.435
SCRC0119	0.339	0.299	0.318	0.305	0.125
SCRC0135	0.412	0.427	0.212	0.428	0.224
SCRC0208	0.434	0.376	0.318	0.402	0.462
SCRC0219	0.329	0.205	0.259	0.234	0.375
SCRC0240	0.350	0.402	0.259	0.403	0.375

表 4(续)

引物名	13 份材料	陆地棉种内	海岛棉种内	陆地棉与海岛棉种间	5 个物种之间
Primer name	13 cotton materials	<i>G. hirsutum</i>	<i>G. barbadense</i>	<i>G. hirsutum</i> and <i>G. barbadense</i>	Among five species
SCRC0262	0.555	0.564	0.424	0.595	0.391
SCRC0291	0.355	0.403	0.424	0.374	0.400
SCRC0371	0.428	0.455	0.212	0.435	0.224
SCRC0584	0.526	0.455	0.212	0.536	0.391
SCRC0633	0.469	0.299	0.255	0.337	0.587
SCRC0878	0.588	0.592	0.106	0.584	0.558
SCRC0970	0.342	0.168	0.106	0.167	0.417
SCRC1005	0.365	0.336	0.423	0.370	0.324
SCRC1006	0.240	0.150	0.259	0.370	0.393
SCRC1022	0.502	0.403	0.255	0.399	0.587
SCRC1024	0.508	0.282	0.318	0.387	0.587
SCRC1327	0.387	0.137	0.212	0.167	0.558
平均值 Mean	0.422	0.336	0.282	0.395	0.424

2.4 陆地棉种内作图群体的多态性分析及染色体初步定位

5 对 genic-SSR 引物 SCRC0056、SCRC0371、SCRC0584、SCRC0878、SCRC1022 在陆地棉品种鲁棉

研 22 和鲁原 343 之间存在明显差异(图 2)。利用这 5 对多态性引物分析(鲁棉研 22 × 鲁原 343)重组自交群体的标记基因型,4 个标记成功连锁到图谱上,分别定位于 Chr. 16、Chr. 23、Chr. 18 和 Chr. 2(图 3)。

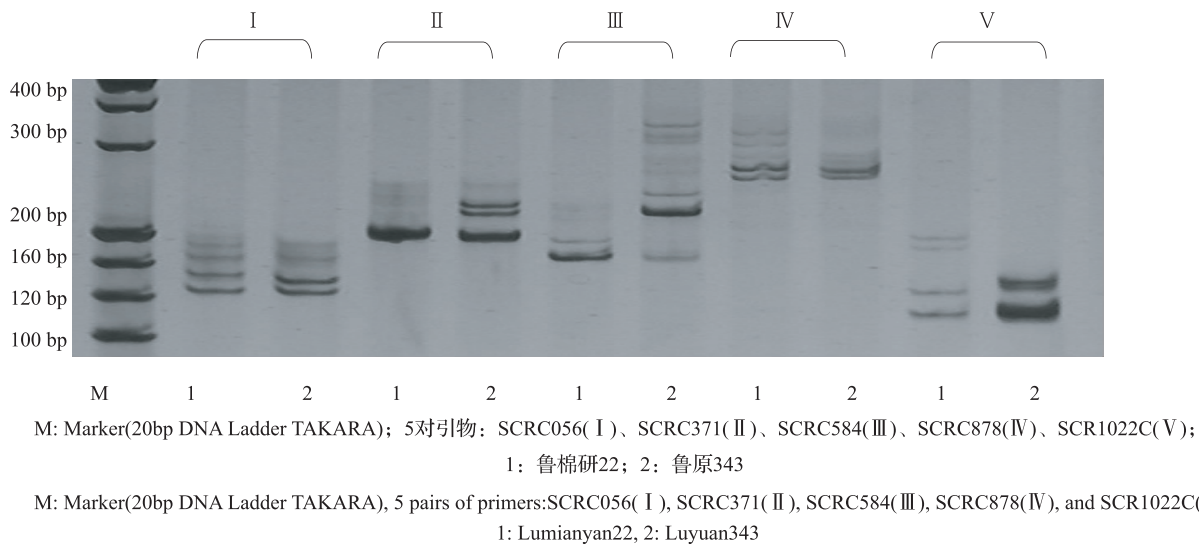


图 2 5 对引物在鲁棉研 22 和鲁原 343 上的扩增电泳图

Fig. 2 PCR results of five primer pairs amplified with Lumianyan22 and Luyuan343

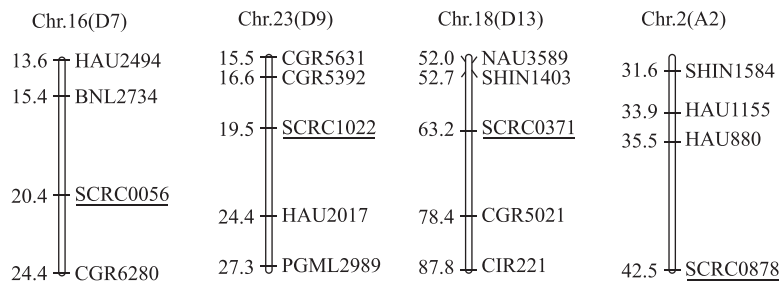


图 3 4 个多态性 genic-SSR 标记在染色体上的定位

Fig. 3 Preliminary chromosome mapping of 4 polymorphic markers

3 讨论

genic-SSR 具有特异性和高度保守性的优点,因此成为有潜力用于物种多样性分析和遗传图谱定位的新型分子标记。随着高通量技术的快速发展,转录组数据不断积累,为大量挖掘 genic-SSR 提供了资源。本研究首次利用陆地棉茎尖转录组开发了与棉花早期形态建成相关的基因 SSR 标记,根据转录组中的 73518 条 Unigene 进行 SSR 位点搜索,筛选出 4507 个 SSR 位点,分布于 4039 条 EST 序列,出现频率为 5.5%,平均每 12.9 kb 出现 1 个 genic-SSR,陆地棉转录组新标记的出现频率高于大麦^[16]、拟南芥^[17]、甘蔗^[18]及葡萄^[19] ESTs 中 SSR 位点频率(3.6%、3%、2.8% 和 2.5%),这种差异可能与 SSR 搜索标准、数据库大小及物种等有关。

各物种开发的 SSR 标记中,一般三核苷酸频率最高^[20],少数几种双子叶植物的 EST-SSR 以二核苷酸重复为主要类型^[21]。本研究中查找了 2~10 bp 核苷酸重复类型,其中最为丰富的是三核苷酸重复(51.03%),主要基元为 AAG/CTT,与海岛棉、陆地棉、雷蒙德氏棉^[8,15,22]中(AAG)_n为主要的三核苷酸类型一致;其次,二核苷酸重复占 28.76%,主要基元为 AG/CT,与王为等^[15]、吕远大等^[8]报道的相一致,与王长彪等^[22]报道的雷蒙德氏棉二核苷酸(AT)_n出现频率最多有差异。这种差异可能与植物本身基因组成分及开发新标记的序列来源和数量等有关。

本研究首次对陆地棉转录组蛋白编码区和非编码区 SSR 位点的分布情况进行分析,结果显示,发掘出的 4507 个 SSR 位点中,有 2853 个位于非编码区,仅有 1654 个位于编码区,位于非编码区 SSR 的数量比编码区高出 72.5%。这一结果与茶树(*Camellia sinensis* L.)花转录组中微卫星序列主要分布在非编码区,转录组编码区序列所含微卫星数量远低于基因非编码区的结论相符^[23]。在基因编码区中出现频率最高的是三核苷酸重复的分子标记,在非编码区中出现较多的标记分别为二核苷酸重复基元和三核苷酸重复基元,其中三核苷酸重复基元仅比二核苷酸重复基元多 6.03%,原因是编码区三联体密码子由三核苷酸组成,因此导致编码区三核苷酸重复基元出现频率最高,非三核苷酸重复面临移码突变的选择压^[24]。

本研究根据 4039 条转录本序列设计引物经 ePCR 筛选之后获得了 1569 对新的 genic-SSR 引

物,在陆地棉(Tm-1)、海岛棉(3-79)、雷蒙德氏棉、阿非利加棉 4 个不同棉种间进行初步筛选,其中 1117 对引物具有稳定的扩增性,有效扩增率为 71.2%,前人所设计引物的有效扩增率应在 60%~90% 之间^[15],出现一部分引物不具有扩增性的可能原因是所设计引物跨越 mRNA 剪切位点及扩增产物包含的内含子过大^[13]。

SSR 标记的多态性在分子标记研究中具有重要意义。*PIC* 值可用来估计每对引物扩增的等位基因的变异程度,而 *PIC* 值的大小是基于所检测到的等位基因的数目及分布频率^[24],所以在一定意义上 *PIC* 值也是衡量基因变异程度的指标。根据前人研究标准,将 *PIC* 值划分为 3 种,分别表示 SSR 标记的多态性;当 *PIC* > 0.5 时,该 SSR 标记为高度多态性位点;当 0.25 < *PIC* < 0.5 时,为中度多态性位点;当 *PIC* < 0.25 时,为低度多态性位点^[25]。M. W. Blair 等^[26]研究表明 EST-SSR 标记的 *PIC* 值低于基因组 SSR。本研究中 13 份材料之间及其中 5 个棉种之间多态性均比来自于陆地棉基因组 SSR(0.12~0.85,平均 0.4)^[27]和亚洲棉 EST-SSR(0.12~0.73,平均 0.42)^[28]的 *PIC* 值高,而其他棉种间及种内的多态性均低于陆地棉和海岛棉^[15,29]的研究结果,可能是由于所用材料中个别棉种测试材料的数目有限,所检测到的种内及种间的遗传变异还不够全面,因此导致多态性略低,但其可作为通用性标记在不同棉种中应用。此次开发的引物中有 4 对成功定位到染色体上,其中 3 个标记在 D 基因组(Chr. 6、23、18),1 个标记在 A 基因(Chr. 2)组上。可见本研究开发的 genic-SSR 在遗传图谱构建、后续功能标记研究方面不失为新的宝贵资源。

参考文献

- [1] Fryxell P A. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae) [J]. *Rheeda*, 1992, 2(2): 108-165
- [2] Chen M L, Wu J, Wang L F, et al. Development of mapped simple sequence repeat markers from common bean (*Phaseolus vulgaris* L.) based on genome sequences of a Chinese landrace and diversity evaluation [J]. *Mol Breeding*, 2014, 33(2): 489-496
- [3] Zhai L L, Xu L, Wang Y, et al. Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (*Raphanus sativus* L.) [J]. *Mol Breeding*, 2014, 33(3): 611-624
- [4] 刘峰,王运生,田雪亮,等. 辣椒转录组 SSR 挖掘及其多态性分析 [J]. *园艺学报*, 2012, 39(1): 168-174
- [5] Zhang H Y, Wei L W, Miao H M, et al. Development and validation of genic-SSR markers in sesame by RNA-seq [J]. *BMC Genomics*, 2012, 13: 316
- [6] 来德勇. 陆地棉花发育 EST 测序分析及其相关基因筛选与微卫星标记开发 [D]. 武汉: 华中农业大学, 2011
- [7] 陈浩东. 达尔文氏棉早肋道转录组测序、EST-SSR 开发及高密度遗传图谱构建 [D]. 北京: 中国农业科学院, 2013

- [8] 吕远大,蔡彩平,王磊,等. 海岛棉 EST-SSRs 分布特征及新标记的开发与利用[J]. 科学通报,2010(19):1886-1890
- [9] Jena S N, Srivastava A, Rai K M, et al. Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.) [J]. Theor Appl Genet, 2012, 124(3): 565-576
- [10] 宋国立,崔荣霞. 改良 CTAB 法快速提取棉花 DNA[J]. 棉花学报,1998,10(5):273-275
- [11] 张军,武耀廷,郭旺珍,等. 棉花微卫星标记的 PAGE/银染快速检测[J]. 棉花学报,2000,12(5):267-269
- [12] Iseli C, Jongeneel C V, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences[J]. Proc Int Conf Intell Syst Mol Biol, 1999:138-148
- [13] 魏利斌,张海洋,郑永战,等. 芝麻 EST-SSR 标记的开发和初步研究[J]. 作物学报,2008,34(12):2077-2084
- [14] Zhao L, Cai C P, Zhang T Z, et al. Fine mapping of the red plant gene R1 in upland cotton (*Gossypium hirsutum*) [J]. Chin Sci Bull, 2009, 54(9):1529-1533
- [15] 王为,王长彪,刘方,等. 棉花非冗余性 EST-SSR 新标记的开发及其评价[J]. 作物学报,2012,38(8):1443-1451
- [16] Kota R, Varshney R K, Thiel T, et al. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.) [J]. Hereditas, 2001, 135(2-3):145-151
- [17] Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants[J]. Genetics, 2000, 156(2):847-854
- [18] 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, 160(6):1115-1123
- [19] Scott K D, Eggler P, Seaton G, et al. Analysis of SSRs derived from grape ESTs [J]. Theor Appl Genet, 2000, 100(5): 723-726
- [20] Liang X, Chen X, Hong Y, et al. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species [J]. BMC Plant Biol, 2009, 9:35-48
- [21] Kumpatla S P, Mukhopadhyay S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species[J]. Genome, 2005, 48(6):985-998
- [22] 王长彪,郭旺珍,蔡彩平,等. 雷蒙德氏棉 EST-SSRs 分布特征及开发与利用[J]. 科学通报,2006,51(3):316-320
- [23] 王丽鸳,康韦,张成才,等. 茶树花转录组微卫星分布特征[J]. 作物学报,2014,40(1):80-85
- [24] Metzgar D, Bytof J, Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA[J]. Genome Res, 2000, 10(1):72-80
- [25] Yadav H K, Ranjan A, Asif M H, et al. EST-derived SSR markers in *Jatropha curcas* L. development, characterization, polymorphism, and transferability across the species genera[J]. Tree Genet Genomes, 2011, 7(1):207-219
- [26] Blair M W, Giraldo M C, Buendía H F, et al. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.) [J]. Theor Appl Genet, 2006, 113(1):100-109
- [27] Frelichowski J J, Palmer M B, Main D, et al. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends[J]. Mol Genet Genomics, 2006, 275(5):479-491
- [28] Park Y H, Alabady M S, Ulloa M, et al. Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred cotton population [J]. Mol Genet Genomics, 2005, 274(4):428-441
- [29] 张艳欣,林忠旭,李武,等. 海岛棉 EST-SSR 引物的开发与利用研究[J]. 科学通报,2007,52(15):1779-1787