

DOI: 10.3724/SP.J.1005.2008.01207

功能型分子标记(ISAP)的开发及评价

陆才瑞^{1,2}, 喻树迅¹, 于霁雯^{1,2}, 范术丽¹, 宋美珍¹, 王武¹, 马淑娟²

1. 中国农业科学院棉花研究所农业部棉花遗传改良重点实验室, 安阳 455004;
2. 华中农业大学国家作物遗传改良重点实验室, 武汉 430070

摘要: 分子标记在图谱构建, QTL 分析, 基因定位以及标记辅助育种中起着越来越重要的作用。研究者都期望一个分子标记位点代表一个特定的基因, 甚至与某种性状联系起来, 这样, 通过对某个分子标记的筛选即能对性状进行筛选, 此即功能型分子标记。而目前广泛使用的基于 PCR 基础的分子标记如 RAPD、SSR、AFLP 等或是扩增非编码区域, 或是随机在基因组中扩增, 得到的位点一般与目标性状基因距离较远, 这使得分子标记在应用上与其目标有一定的偏差。研究建立了一种基于基因中内含子序列的功能型分子标记, 试图使标记位点与基因序列联系起来以达到其功能型的目的。它利用内含子剪接位置的保守一致序列作为其引物的核心序列, 其上下游引物均为 18 bp, 上下游引物间通过组合配对的方式作为扩增的引物对, 对真核生物的基因序列进行扩增。为了验证 ISAP 的功能性, 研究设计了 17 条引物(9 条上游引物, 8 条下游引物, 共计 72 个引物组合)对棉花 F₂ 群体进行扩增并构建遗传连锁图谱, 其中 67 个显示了多态性, 共得到 212 个位点。我们用此 212 个位点连同 164 个 SRAP 位点构建了一张包含 276 个位点的遗传连锁图谱, ISAP 标记在整个连锁群中分布比较均匀, 部分区域呈现标记高饱和现象, 可能为编码序列富集区。另外对 20 个片段进行测序的结果表明, 85% 的序列显示了与已公布 EST 序列的同源性, 说明扩增是跨越了外显子进行的, 得到的序列与表达序列紧密连锁。结果显示, ISAP 标记是简单, 可靠, 具有较高多态性, 并且扩增基因区域的一种功能型分子标记。同时, 还使用 ISAP 标记对其他植物进行了扩增, 取得了良好的效果。

关键词: ISAP; 功能型分子标记; 开发

Development and appraisalment of functional molecular marker: intron sequence amplified polymorphism (ISAP)

LU Cai-Rui^{1,2}, YU Shu-Xun¹, YU Ji-Wen^{1,2}, FAN Shu-Li¹, SONG Mei-Zhen¹, WANG Wu¹, MA Shu-Juan²

1. China Cotton Research Institute, Key Laboratory of Cotton Genetic Improvement of Ministry of Agriculture of China, Anyang 455004, China;
2. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

Abstract: Molecular markers are playing an increasingly important role in map construction, QTL analysis, gene mapping and marker-assisted selection. Researchers hope the target gene and locus are as close as possible, one locus can present one gene, or linked with some important trait, then, individuals with useful trait can be selected through molecular markers se-

收稿日期: 2007-11-05; 修回日期: 2008-02-02

基金项目: 国家重点基础研究项目(编号: 2004CB117306)资助[Supported by National Basic Research Program of China (No. 2004CB117306)]

作者简介: 陆才瑞(1980-), 男, 河南人, 硕士, 实习研究员, 研究方向: 棉花功能基因组学。Tel: 0372-2562375; E-mail: lucr99@126.com

通讯作者: 喻树迅(1953-), 男, 湖北人, 博士, 研究员, 研究方向: 棉花遗传育种。Tel: 0372-2562201; E-mail: yu@cicaas.com.cn

致谢: 衷心地感谢梁文科副研究员, 闫俊副研究员惠赠玉米和小麦材料, 感谢袁有禄研究员和杜雄明研究员对本文的指导和帮助。

lecting, and it's the functional molecular marker. PCR-based molecular markers such as RAPD, SSR, AFLP amplified non-coding regions, or the whole genome randomly, the locus is far away from the gene of targeted trait, this limit the application of these molecular markers. This study established a kind of functional molecular markers based on intron of gene sequence, trying to link loci with gene sequence to achieve the purpose of its function. It used the conservative consistent sequence of intron splicing sites as its core sequence of amplification. ISAP is a PCR-based marker system, it has two kinds of primers: forward primer and reverse primer, both primers are 18 bases. Any of the primers can be used to construct a primer combination with the other kind of primers. Seventeen primers, 9 forward and 8 reverse, were used to construct 72 primer combinations, 67 of them showed polymorphism in a *G. hirsutum* cv. CCRI36 \times *G. barbadense* cv. H7124 F_2 population and a total of 212 loci were obtained. Together with 164 SRAP loci, these 212 loci were used to construct a genetic linkage map. ISAP markers distributed evenly in the entire linkage group, part of the region had a high saturation, might be the coding sequence-rich region. Sequencing results of 20 fragments showed that 85% of the sequences announced homology with published EST sequence stored in the NCBI which indicated that they were amplified adjacent to expressed sequences. These results showed that ISAP marker system was simple, efficient, reliable, and had a relatively high polymorphism, furthermore, it directly targeted gene sequence, was a functional molecular marker system. ISAP was also used to amplify other plants and good results were achieved.

Keywords: ISAP; functional molecular marker; development

DNA 分子标记广泛应用于植物分子遗传连锁图谱的构建、植物遗传多样性分析与种质鉴定、重要农艺性状基因定位与图位克隆、转基因植物鉴定和分子标记辅助育种等方面,取得惊人的成绩,具有广阔的应用前景。

标记位点距离目标基因的远近决定了其应用的可能性,研究者都希望这个距离越近越好,而目前广泛使用的基于 PCR 基础的分子标记如 RAPD、SSR、AFLP 等或是扩增非编码区域,或是随机在基因组中扩增,得到的位点一般与目标性状基因距离较远,这使得分子标记在应用上与其目标有一定的偏差,如重要农艺性状基因定位与图位克隆以及分子标记辅助育种等都要求标记位点与目标性状基因间的距离较近,而目前使用的这些分子标记普遍距离编码序列较远或很难在目标性状基因附近存在标记位点。

研究人员不断地尝试多种方法使分子标记与编码序列联系起来,以缩短标记位点与目标性状基因间的距离。真核生物中大多数基因都是由外显子和内含子组成,外显子进化较慢,比较保守,内含子进化较快,变异大。基因转录时剪接掉内含子序列,加工产生成熟的 mRNA,即编码序列。利用真核生物中基因序列的特点,研究者先后开发了多种与编码序列相联系的分子标记,如 EST 分子标记, Sequence related amplified polymorphism (SRAP) 标记, Target

region amplification polymorphism (TRAP) 标记, Intron fragment length polymorphisms (IFLP) 标记和 single-strand conformational polymorphism (SSCP)-SNP 标记等。目前基于 EST 开发的分子标记包括 EST-PCR、EST-SSR、EST-SNP、EST-RFLP 和 EST-AFLP。以 EST 为基础的分子标记具有开发简便、信息量高和通用性好等特点^[1],已经被广泛的应用于比较作图^[2],基因定位及寻找基因等^[3],但开发成本较高。芸苔属植物中开发出来的 SRAP 标记利用独特的引物设计对 ORFs 进行扩增^[4],上游引物长 17 bp,对外显子进行特异扩增,下游引物长 18 bp,对内含子区域、启动子区域进行特异扩增,因个体不同以及物种的内含子、启动子与间隔长度不等而产生多态性。TRAP 是利用生物信息工具和 EST 数据库信息,产生目标候选基因区多态性标记。TRAP 技术采用两个 18 核苷酸引物产生标记,一个为固定引物,依据 EST 序列设计;另一个为随机引物,针对外显子和内含子的特点,设计为分别富含 GC 或 AT 核心区的任意序列^[5]。但 SRAP 仅能触及 60% 的表达序列^[4],而 TRAP 更少(私人交流)。IFLP^[6]和 SSCP-SNP^[7]则是针对特定的基因在内含子两侧设计引物扩增目标内含子,以达到特定基因定位的目的。

基于以上各种分子标记的启发,本研究利用真核生物基因中拼接位点序列的保守性,统计其连续一致序列,设计上下游引物,对基因组进行扩增,

并研究其扩增产物序列的特点,旨在揭示基因序列中遗传多态性的方法,研究分子标记与表达序列紧密连锁的途径,以期获得一种便于分子标记的高效应用的功能型分子标记方法。

1 材料和方法

1.1 实验材料

构建棉花海陆杂交 F_2 群体的两个亲本为海岛棉 7124 和短季棉品种中棉所 36。2002 年在武汉华中农业大学试验地配制杂交组合得 F_1 代, 2003 年对 F_1 自交得 F_2 代种子, 2003 年冬季 F_2 代植株种植于中棉所海南植物园, 选取其中 186 个长势较好的单株构建群体。用于检测标记通用性的小麦 8 个品种由中国农科院棉花研究所小麦研究室提供, 为科农 199、良基 99、邯 7086、周 98100、汶农 6 号、烟 2070、衡 4338、科麦 1 号; 玉米 4 个品种由中国农科院棉花研究所玉米研究室提供, 为东农 250、中单 9409、农大 108、京玉 7 号; 辣椒 3 个品种为羊角椒、天鹰椒、农蕾 2 号; 烟草、花生、高粱、吊兰和一串红均采自中棉所温室中。

1.2 方法

1.2.1 基因组 DNA 的分离

DNA 的提取参照张金发等^[8]提取 DNA 的方法, 并作了一些优化。用紫外分光光度法检测 DNA 的纯度和浓度, 将高纯度的 DNA 稀释成 20 ng/ μ L 作为工作液, 4℃ 保存。

1.2.2 ISAP 标记体系

ISAP 采用上游和下游 2 种引物组合的方式进行扩增。ISAP 引物设计方法为: 随机选取 NCBI 数据库中的棉花或水稻 BACs, 在 BACs 中选择 cDNA 序列验证过的基因序列, 再从此基因序列中选择符合如下规则的内含子边界序列, 即供体端为 5'-A/CA-GGTAA-3' 或受体端为 5'-TTGCAG-3', 且受体端序列的 5' 上游的 7 个碱基至少含有 4 个嘌呤。然后提取该内含子的边界序列, 使引物的总长度为 18 个碱基, 剔除含有二级结构以及其他不适合作为引物的序列, 选取其中符合引物设计原则的引物作为上下游引物。引物由上海生工生物工程技术有限公司合成(表 1)。

1.2.3 ISAP 反应

20 μ L 的 PCR 反应体系为: 40 ng DNA 模板, 引物

Table 1 ISAP primers used in this study

正向引物 Forward primers (5' 3')	反向引物 Reverse primers (5' 3')
F1: CGATATAAGCAAAGGTAA	R1: CTGCAATTAAGCAAGAAC
F2: GCATGAATGCAAAGGTAA	R2: CTGCAATGTCCCATAGAT
F3: ATGGAAGCTCGCAAGGTAA	R3: CTGCAACAAGATCTCAGA
F4: ACGAAGATGGAAAGGTAA	R4: CTGCAAGTGAGAACACCC
F5: TAGCCGGTATCAAGGTAA	R5: CTGCAAAATTCAATAGTT
F6: CGTCCGATGAAAAGGTAA	R6: CTGCAAATGTTAAACCCA
F7: ATCAGCTGCTGCAGGTAA	R7: CTGCAAGGGTTAACCAGT
F8: AGCCGTTTATACAGGTAA	R8: CTGCAATAACCACATGAA
F9: CATCTCACTTTCAGGTAA	

各 0.25 μ mol/L, dNTPs 200 μ mol/L, 1× PCR buffer, 15 mmol/L $MgCl_2$, 1 U 的 *Taq* 酶, 不足部分由 ddH₂O 补足。

扩增反应在 PTC-100 PCR 仪上进行, PCR 反应程序如下: 94℃ 变性 5 min, 1 个循环; 94℃ 变性 1 min, 35℃ 复性 1 min, 72℃ 延伸 1 min, 5 个循环; 94℃ 变性 1 min, 50℃ 复性 1 min., 72℃ 延伸 1.5 min, 35 个循环; 72℃ 延伸 5 min, 4℃ 保存。

20 μ L 扩增产物中加入 10 μ L Loading buffer (95%去离子甲酰胺, 10 mmol/L EDTA, 0.25%溴酚蓝, 0.25%二甲苯青), 95℃ 变性 5 min, 上样 5 μ L。扩增产物用变性聚丙烯酰胺凝胶电泳(4%, 7 mol/L 尿素)分离, 电泳缓冲液为 0.5×TBE。电泳时先用 1 800 V 电压预电泳至电流为 30 mA, 上样后用 2 000 V 恒压电泳 1.5~2 h 至二甲苯青到胶板 3/4 处, 电泳过程中确保胶板温度不高于 50℃ 以免玻璃板破裂。

1.2.4 目标 DNA 片段的回收、连接、转化、测序

选择在目标分子量范围内, 扩增产物浓度高的 DNA 带作为目标片段。用手术刀片从聚丙烯酰胺凝胶上将目标 DNA 片段小心切下, 放到装有 50 μ L 灭菌 TE 的 0.5 mL 离心管中, 100℃ 煮沸 15 min 以回收 DNA。然后取回收好的 DNA 12 μ L 进行 PCR 重扩增。扩增产物直接取 2 μ L 用于后续的连接, 剩余扩增产物用于 1.2%琼脂糖电泳检测, 检测 仍为目标 DNA 的样品继续进行连接反应。

将 5 μ L 连接产物加到装有 50 μ L 感受态细胞的 1.5 mL 的离心管中, 轻轻敲打, 然后冰浴 30 min, 42℃ 热激 90 s, 继续冰浴 2 min。每管加入 600 μ L 灭菌的 LB 液体培养基, 温育 1.5 h。吸取 300 μ L 菌液混合 X-gal (20 mg/mL) 和 IPTG (200 mg/mL) 涂板, 待平板表面的菌液被培养基吸收后, 将平板倒置于培养箱中, 37℃ 培养 12~16 h。每个样品挑取 5

个白色克隆, 到装有 2 mL 灭菌 LB 液体培养基的 5 mL 离心管中, 37 °C, 240 r/min 培养 12 ~ 16 h。取培养好的菌液做模板, 用原引物和测序载体上的引物进行两轮 PCR 检测, 取结果仍为原 DNA 片段的菌液送上海生物公司测序。

表 2 本研究中使用的 SRAP 引物

Table 2 SRAP primers used in this study

正向引物 Forward primers (5' - 3')	反向引物 Reverse primers (5' - 3')
me1: TGAGTCCAAACCGGATA	em1: GACTGCGTACGAATTAAT
me2: TGAGTCCAAACCGGAGC	em2: GACTGCGTACGAATTTGC
me3: TGAGTCCAAACCGGAAT	em3: GACTGCGTACGAATTGAC
me4: TGAGTCCAAACCGGACC	em4: GACTGCGTACGAATTTGA
me5: TGAGTCCAAACCGGAAG	em5: GACTGCGTACGAATTAAC
me6: TGAGTCCAAACCGGTAG	em6: GACTGCGTACGAATTGCA
me7: TGAGTCCAAACCGGTTG	em7: GACTGCGTACGAATTATG
me8: TGAGTCCAAACCGGTGT	em8: GACTGCGTACGAATTAGC
me9: TGAGTCCAAACCGGTCA	em9: GACTGCGTACGAATTACG
	em10: GACTGCGTACGAATTTAG
	eml 1: GACTGCGTACGAATTTTCG
	em12: GACTGCGTACGAATTGTC
	em13: GACTGCGTACGAATTGGT
	em14: GACTGCGTACGAATTCAG
	em15: GACTGCGTACGAATTCTG
	em16: GACTGCGTACGAATTCGG
	em17: GACTGCGTACGAATTCCA

1.2.6 数据读取及分析方法

所有的多态性条带均统计为显性标记。分子标记的记录采用常规的Mapmaker软件^[10]记录方法。具有亲本中棉所 36 扩增带型的记为“D”, 无扩增带型的记为“B”; 具有亲本海 7124 扩增带型的记为“C”, 无扩增带型的记为“A”。缺失数据用“-”表示。采用引物组合命名法对得到的位点进行命名, 同一对引物产生多个位点, 在名字末尾按照扩增片段的大小依次添加字母a, b, c, 依次类推。M2E15 表示该标记来源于SRAP中 Em2 和Me15 的引物组合, F1R3 表示该标记来源ISAP中F1 和R3 的引物组合, F2R4b表示该标记为F2 和R4 的引物组合产生的第二大片段位点。

1.2.7 连锁分析

在Windows98 下运行Mapmaker/Exp3.0 作图软件^[10]进行分子标记的连锁分析和位点排序。首先设置软件的运行环境, 如“Error Detection on ”、“Use Three Point”、“Triple Exclusion Criteria 4.0”、“Triple Ermr Detection on”等。LOD值最小为 3.0, 最大遗传距离为 50 cM。具体分析时, 先用“Group”命令进行分组。组内少于 8 个标记时, 用“Compare”命令对标记进行两点测验排序, 构画连锁群。对于组内标记

1.2.5 SRAP 反应

153 个(me1~me9, em1~em17)SRAP引物组合中(me1~me6, em1~em6)是Li和Quiros^[4]等人发表的, (me7~me9, em7~em17)是华中农业大学张献龙教授自行设计的^[9], 由上海生工生物工程技术服务有限公司合成(表 2)。

数多于 8 个的组, 用“Three Point ”进行三点测验, 用“Order”命令确定标记顺序。对未进入连锁群的标记用“Try ”命令连到相应连锁群上, 最后用“Ripple”命令确定最佳排列顺序, 构建棉花分子标记连锁图。连锁图绘制采用MaPDraw2.1^[11]。

2 结果与分析

2.1 引物长度

本实验尝试了 14~20 个碱基的引物长度, 发现引物碱基长度与扩增效果具有明显的相关性。在引物长度为 14~17 个碱基时, 扩增结果显示了相对较高的多态性, 同时也表现出较低的可重复性; 而当引物长度在 19~20 个碱基时, 扩增结果显示了较高的可重复性, 但同时也表现出较低的多态性, 而且有时会出现背景强但带很糊的情况。当引物长度为 18 个碱基的时候扩增效果最好。因此最终设定 ISAP 引物的长度为 18 个碱基。

2.2 引物序列

真核生物中大多数基因都是由外显子和内含子组成, 基因转录时需要剪接掉内含子序列, 才能加工产生成熟的mRNA分子。作为非编码序列的内含

子要比作为编码序列的外显子进化速率快^[12,13],且外显子与内含子交界的边界序列是高度保守的,因此本研究决定以内含子边界序列为引物对基因组进行扩增,以达到标记位点联系表达序列并具备一个相对较高多态性的目的,即内含子序列扩增多态性(Intron sequence amplified polymorphism)。

Chambon等^[14]在调查了编码蛋白基因的 90 个供体位点和 85 个受体位点后得到了一个一致序列:供体位点一致序列为 5' -A/CAGGTA/GAGT-3',受体位点一致序列为 5 -PyPyPyPyPyPyXCAGGG/T-3。关于人类基因组和粘液菌的调查也显示了类似的结果,仅仅在长短内含子序列之间有所差别^[15]。

他们的研究是针对固定位置某个碱基出现的频率,没有研究多个碱基连续出现的频率,而这对于引物的设计至关重要。本实验抽取了 30 个GenBank里棉花和水稻的BACs分析其内含子边界序列。首先筛选符合GT-AG原则^[14]的内含子,然后统计紧邻GT-AG碱基出现的频率,确定该碱基以后,再统计紧邻该碱基的碱基出现的频率,仅留下出现频率最高的碱基相应的序列,依此类推。结果显示在内含子供体端的连续一致序列为 5 -A/CAGGTAA-3,占调查序列的 21%;受体端的连续一致序列为 5 -TTGCAG-3,且此一致序列的 5 上游的 7 个碱基应含有至少 4 个嘌呤,占调查序列的 44%。

本研究利用内含子边界中的连续一致序列设计 ISAP 引物,其多态性来自于品种间内含子进化速度的不同。使用 A/CAGGTAA 作为上游引物的 3 端是为了与内含子的 5 边界序列配对,使用 CTGCA-A 及 4 个嘌呤在下游引物的 5 端是为了与内含子的 3 边界序列配对。为了使引物的组成更接近自然状态,本研究直接利用 BAC 库中符合 ISAP 引物原则的内含子边界序列作为引物。

2.3 PCR 环境

将一种标记付诸应用时,首先必须建立一套稳定可靠、扩增效果好、可重复的反应体系,才能确保应用的持续性和结果的可靠性。通过对ISAP引物的扩增效果研究发现,引物长度对于扩增结果具有明显的作用效果。ISAP标记与SRAP标记^[4]具有近乎相同的长度,因此本研究借鉴了SRAP分析的引物用量及其PCR反应体系并按照电泳结果做了相应的改动,取得了很好的扩增效果。20 μ L的 PCR 反应体系为: 40 ng DNA模板,引物各 20 ng, dNTPs 200 μ mol/L, 1 \times PCR buffer, 15 mmol/L MgCl₂, 1U的*Taq*

酶,不足部分由diH₂O补足。起始 5 个循环复性温度为 35 ,较低的复性温度会使 2 种引物在部分错配的情况下完全结合到基因组DNA上,随后的 35 个循环复性温度增加到 50 ,这将保证DNA产物的稳定性和可靠性。我们也尝试过保持复性温度在 35 和 50 的情况下扩增,前者表现低重复性,而后者产生的条带很弱,难以分辨。

2.4 ISAP 标记的图谱构建

一种优良的分子标记除了具备稳定、可靠等特性以外,还必须具备较高的多态性并在染色体上分布相对均匀。为了检测 ISAP 标记的这些特点,本研究以海 7124 和中棉所 36 为父母本构建了 F₂ 群体并进行了图谱构建工作。结果表明 ISAP 标记在两亲本之间表现出较强的多态性,每个引物组合可产生 20~100 个清晰可辨的条带。使用了 9 条前端引物,8 条后端引物,共计 72 个引物组合。用 72 个引物组合进行群体分析,其中 67 个引物组合产生多态性,共得到 212 条多态性条带,每组合的多态性条带数从 1~8 不等。引物组合多态性频率为 93.1%,平均每个引物组合产生 3.16 个多态性条带(图 1)。

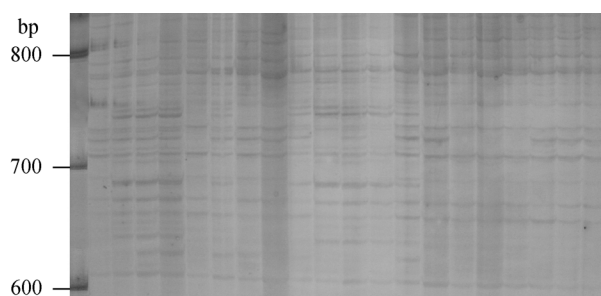


图1 ISAP 引物组合 F2R4 在棉花 F₂ 群体中的扩增
Fig. 1 Amplification of F2R4 in cotton F₂ population

为了确定ISAP标记在棉花基因组中的分布情况,我们使用得到的 212 个位点和 165 个SRAP位点对海陆群体进行的图谱构建工作,初步构建了一个包括 276 个位点的分子标记连锁图。其中ISAP位点 138 个,SRAP位点 138 个。该遗传连锁图共有 32 个连锁群(*LOD* 3.0),每个连锁群包含 2~31 个位点。根据连锁群长度从大到小顺序对各连锁群进行命名,最小的连锁群遗传距离为 5.2 cM,最大的连锁群遗传距离为 179.3 cM,平均标记距离为 8.9 cM。图谱全长为 2 179.4 cM,占据棉花基因组总长度的 46.8%^[16]。ISAP标记在整个连锁群中分布比较均匀,

部分区域呈现标记高饱和现象, 可能为基因组中的 基因聚集区(图 2, 表 3)。

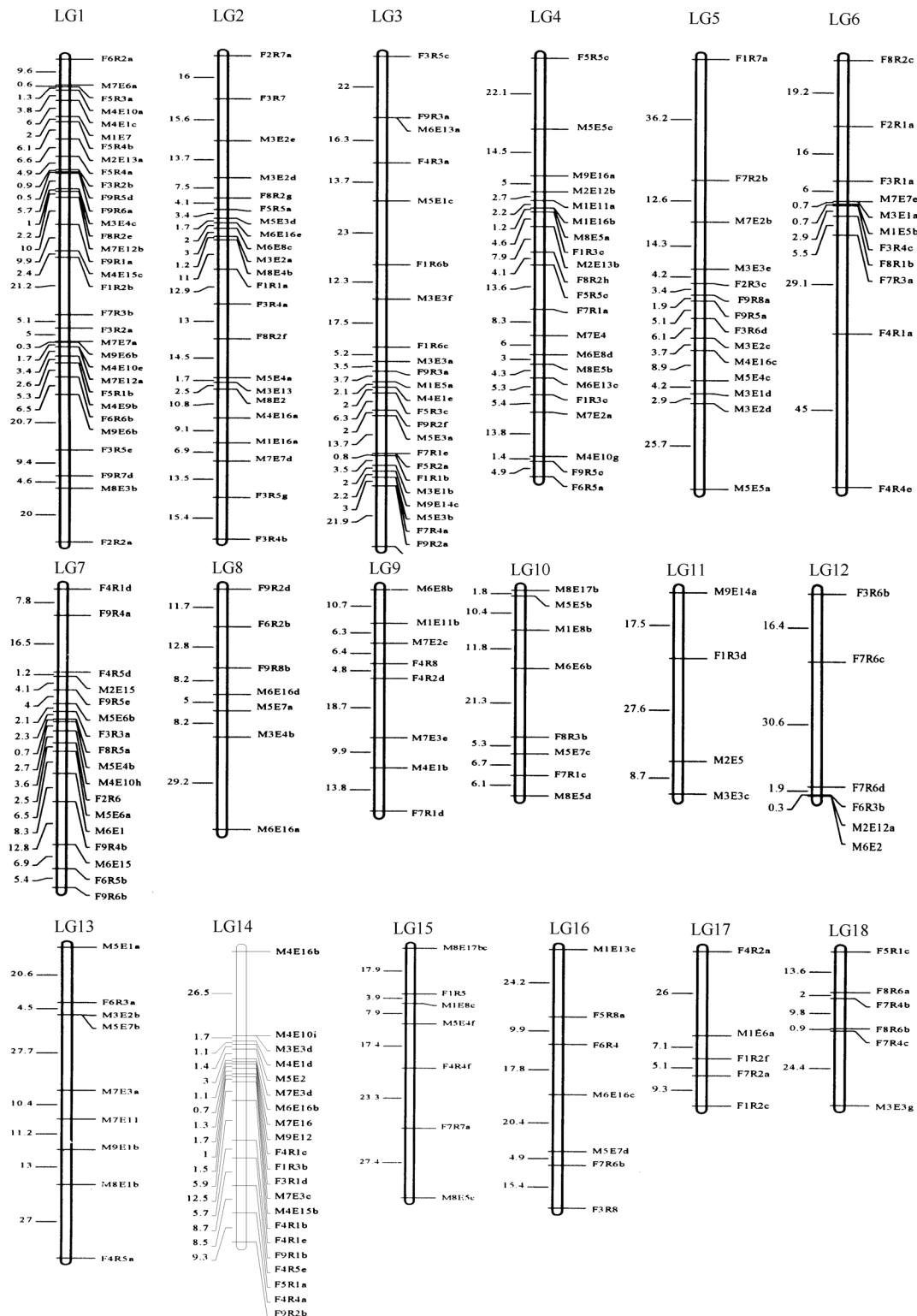


图 2 利用 ISAP 和 SRAP 标记对中棉所 36×H7124 F_2 群体构建遗传连锁图谱

LG: 连锁群。

Fig. 2 Construct genetic linkage map of CCRI36×H7124 with ISAP and SRAP markers

LG: Linkage group.

表 3 ISAP 和 SRAP 标记在棉花 32 个连锁群上的分布情况
Table 3 Distribution of ISAP and SRAP markers in 32 cotton linkage groups

分子标记 Molecular marker	连锁群 Linkage groups										
	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	LG10	LG11
ISAP	17	9	14	8	6	8	10	3	3	2	2
SRAP	15	13	10	13	8	3	7	4	5	6	6
	LG12	LG13	LG14	LG15	LG16	LG17	LG18	LG19	LG20	LG21	LG22
ISAP	4	5	10	3	4	4	5	4	2	2	2
SRAP	2	1	11	4	3	1	1	3	0	0	4
	LG23	LG24	LG25	LG26	LG27	LG28	LG29	LG30	LG31	LG32	
ISAP	1	1	0	3	3	3	0	2	0	2	
SRAP	1	2	3	2	0	1	4	0	2	0	

LG: 连锁群。
LG: Linkage group.

2.5 PCR 产物检测

图位克隆和分子标记辅助育种等均希望标记距离目标基因较近,如果分子标记本身就位于目标基因内,则可以大大提高分子标记辅助选择的精度并减少图位克隆后期的工作量。为了检测 ISAP 引物扩增产物的情况,我们随机抽取了棉花 20 个条带进行测序。

ISAP 引物扩增得到的片段分布于 50~2 000 bp,其中大部分片段集中在 600~1 000 bp。研究显示,大

部分外显子长 30~40 bp, 内含子则一般在 40~125 bp^[15]。由于扩增产物片段相对于一个内含子来说普遍较长, 推测其扩增可能是跨越了一个以上外显子发生的, 我们对测序结果我们进行了BLASTN检测,发现 17 个(85%)与已公布EST序列相似(表 4), 部分序列有明显的内含子现象(图 3), 这说明大部分扩增发生在基因序列区域, ISAP标记与表达序列联系紧密。

表 4 棉花中分离的 20 个条带测序与比对分析
Table 4 Sequence and blast of separated bands from cotton

标记名称 Marker name	片段大小 Size (bp)	比对长度 Length of similarity(bp)	比得分值 BLASTN score(bits)	E 值 E value	GenBank 序列号 GenBank accession number
F1R2a	522	114	105	1 × e ⁻¹⁹	DR455763
F1R3b	511	60	77	2 × e ⁻⁸	CV942815
F1R5	592	50	63	2 × e ⁻⁶	DT595915
F2R1a	560	33	42.8	2 × e ⁻³	DN577022
F2R1b	551	47	64	7 × e ⁻⁶	DW513951
F2R2c	433	138	158	1 × e ⁻³²	DT468253
F2R5	569	54	78.8	2 × e ⁻⁸	AU083383
F2R7a	469	0			
F3R2a	525	45	46.6	4 × e ⁻⁴	EB696786
F3R2b	565	27	41	0.71	BF253909
F3R3a	549	42	44.6	6 × e ⁻⁶	CX089680
F3R4d	560	41	44.6	4 × e ⁻⁵	CD151663
F3R5a	534	0			
F4R1b	566	50	44.6	8 × e ⁻⁶	CN930986
F4R1f	556	39	44.6	2 × e ⁻⁵	EC956607
F5R3a	553	30	41	0.13	BP917661
F5R3c	558	45	6.8	7 × e ⁻⁶	EE068098
F5R4a	548	25	41	1.3	CK085724
F5R4c	478	0			
F5R5c	560	33	42.8	2 × e ⁻⁴	CD575434

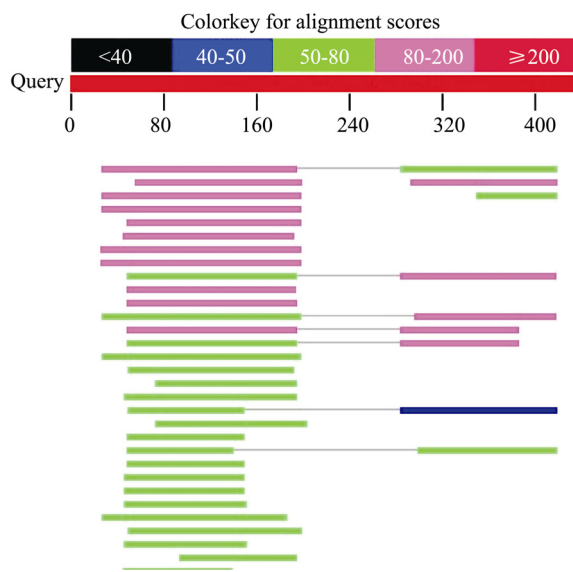


图 3 ISAP 标记 F2R2c 在 NCBI 上的比对图

Fig. 3 Blast in NCBI of ISAP marker F2R2c

2.6 ISAP 标记的通用性

ISAP 引物开发于基因转录编码区, 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, 因此, ISAP 标记有可能在物种间通用。有鉴于此, 我们随机抽取 8 对引物对 8 种植物进行了通用性分析, 包括小麦、玉米、辣椒、烟草、花生、高粱、吊兰和一串红, 其中玉米 4 个品种, 辣椒 3 个品种。所有引物在所有物种中均有扩增产物, 且在同一物种不同品种中均显示较高的多态性, 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 4), 但有部分条带在不同物种中也是相同的, 这说明不同物种间在很多基因序列上是保守的, 且 ISAP 标记在基因组中扩增的位置稳定一致。

3 讨论

3.1 功能型分子标记的开发

研究者都期望一个分子标记位点代表一个特定的基因, 甚至与某种性状联系起来, 这样, 通过对某个分子标记的筛选即能对性状进行筛选, 此即功能型分子标记。EST 标记, SRAP 标记, TRAP 标记, IFLP 标记, SSCP-SNP 标记等的开发就是基于这个目的。但 EST 标记普遍多态性低, SRAP 标记和 TRAP 标记触及的表达序列较少, IFLP 和 SSCP-SNP 标记则是根据已知基因序列开发来的, 成本较高,

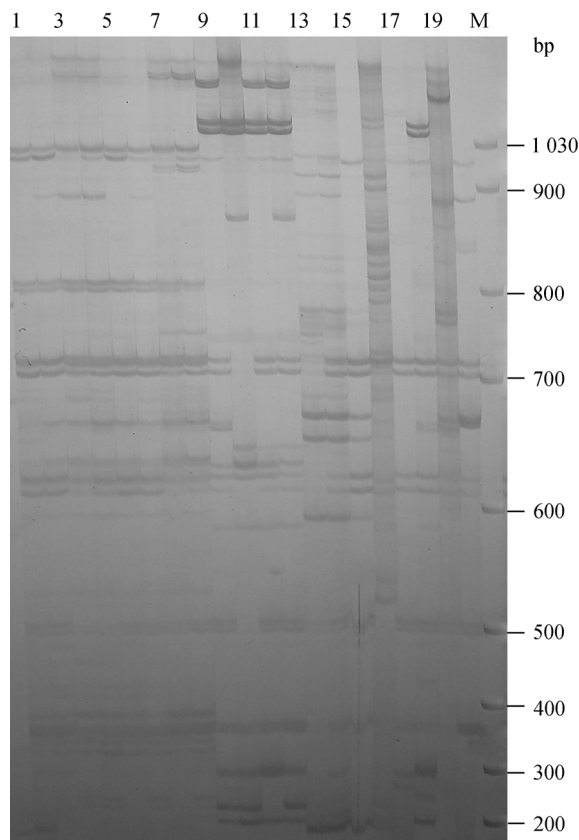


图 4 ISAP 标记在 F3R3 在 8 种植物中的扩增

1~8: 小麦; 9~12: 玉米; 13~15: 辣椒; 16: 烟草; 17: 花生; 18: 高粱; 19: 吊兰; 20: 一串红。

Fig. 4 Amplification of ISAP marker F3R3 in 8 plants

1~8: ISAPs amplified by primer combination F3R3 with wheat; 9~12: Corn; 13~15: Capsicum; 16: Tobacco; 17: Peanut; 18: Jowar; 19: Chlorophytum capense; 20: Salvia splendens.

用途专一。本研究利用基因序列中拼接位点序列的保守性, 统计其连续一致序列, 设计上下游引物组合, 开发了一种基于基因序列扩增的功能型分子标记 ISAP。PCR 产物检测结果显示标记位点与表达序列紧密联系, 并具有相对较高的多态性。

在 ISAP 标记的开发过程中, 我们发现通过对基因序列的分解可以开发出多种多样的功能型分子标记, 如扩增启动子, 5' UTR, 3' UTR 等区域的分子标记。本研究利用真核生物启动子的标志序列 CAATCT 与 SRAP 引物组合开发的 PRAP 标记取得了良好的扩增效果。同样, 通过进一步了解基因组的其他区域也可以开发出相应的分子标记, 这主要取决于我们对生物基因组了解的深浅。本研究还尝试了 SRAP+ISAP、SRAP+PRAP、PRAP+ISAP 等不同类型引物相组合的方式对棉花基因组进行扩增, 也取得了良好的效果(数据未列出)。这说明不同

类型的功能型分子标记引物之间在基因序列内部具有共通性, 可以相互组合以产生更多的多态性条带。

3.2 ISAP 标记的评价

ISAP 标记是基于 PCR 的标记系统, 因而操作简便, 使用 18 个碱基的引物以及 50 的退火温度, 保证了扩增结果的稳定性。上游引物与下游引物之间可自由组配, 因而用少数的引物可进行多种组合, 大大减少了合成引物的费用, 同时也大大提高了引

物的使用率。由于 ISAP 标记的引物由真核生物基因的内含子拼接序列开发而来, 使得到的位点与表达序列紧密连锁, 为它的应用提供了一个广阔的前景。目前应用比较广泛的 DNA 标记主要是 RFLP、SSR、RAPD、AFLP 等, 这些几乎都是扩增非编码区或是在基因组中随机扩增, SRAP 和 TRAP 标记触及了一些表达序列, 但仍然较少。相对而言, ISAP 标记是一种简便、稳定、快捷、经济与表达序列紧密连锁的分子标记(表 5)。

表 5 几种分子标记方法比较
Table 5 Comparison of 6 kinds of molecular markers

	RFLP	RAPD	SSR	AFLP	SRAP	ISAP
遗传特性 Genetic character	共显性 Codominant	显性 Dominant	共显性 Codominant	共显性/显性 Codominant/dominant	共显性/显性 Codominant/dominant	显性 Codominant/dominant
检测技术 Detection technique	分子杂交 Molecular hybridization	随机扩增 Random amplification	特异扩增 Specific amplification	选择性扩增 Selective amplification	选择性特异扩增 Selective specific amplification	选择性特异扩增 Selective specific amplification
检测基因组部位 Detected site in genome	单/低拷贝区 Single/low copy area	全基因组 Whole genome	重复序列 Repeat sequence	全基因组 Whole genome	全基因组 Whole genome	全基因组 Whole genome
单次检测位点数 Loci of Single detection	1~4	1~10	1	100~200	20~100	20~100
重复性 Repeatability	高 High	低 Low	高 High	高 High	高 High	高 High
DNA 质量要求 DNA quality requirement	高 High	低 Low	低 Low	高 High	低 Low	低 Low
同位素使用 Isotope using	必须 Must	不需要 Not need	不需要 Not need	非必须 Not necessary	非必须 Not necessary	非必须 Not necessary
技术难度 Technical difficulty	难 Difficult	易 Esay	易 Esay	中等 Moderate	易 Esay	易 Esay
费用 Cost	高 High	低 Low	低 Low	高 High	低 Low	低 Low
时间消耗 Time consumption	多 More	少 Less	少 Less	多 More	少 Less	少 Less
与基因联系程度 Degree of relation with genes	不紧密 Not close	不紧密 Not close	不紧密 Not close	不紧密 Not close	略紧密 Close slightly	很紧密 Quite close

本实验室应用 3 种 PCR 标记, 即 SSR、AFLP 和 SRAP 进行图谱构建, 约 4 000 对 SSR 引物、64 个 AFLP 引物组合和 153 个 SRAP 引物组合共揭示了 1 200 个多态性位点。平均每对 SSR 引物产生 0.21 个位点, 每条 AFLP 引物产生 4.34 个位点, 每个 SRAP 引物组合产生 2.7 个位点, 而平均每个 ISAP 引物组合产生 3.16 个位点, 证明 ISAP 有中等产率。ISAP 最重要的特点是与表达序列紧密连锁。由于其

引物是从内含子边界序列开发而来, 与其他分子标记相比更有可能靠近表达序列。本研究抽取部分多态性条带的测序结果也证实了扩增发生在与表达序列紧密连锁的部位。

3.3 ISAP 标记的及应用前景

ISAP 有别于以往任何一种分子标记, 它通过对基因序列的扩增而与表达序列紧密连锁。它的这种

特性使得它可以被广泛而高效地应用于高密度图谱的构建, QTL 检测, 基因定位, 图位克隆以及分子标记辅助育种等。

利用新类型的 ISAP 标记能增加遗传图谱位点的数目, 填补原有图谱在基因富集区的空缺, 促进连锁图谱的加密, 使分子标记在图谱上的分布更加均匀。基因定位, QTL 检测, 图位克隆以及分子标记辅助育种等均要求分子标记与目标性状候选基因紧密连锁或者是将基因定位于一个较小的区间, 目的基因与分子标记的连锁紧密不仅可以提高基因定位和 QTL 检测的效率, 还能在图位克隆和分子标记辅助育种中提高精度从而节省后期的工作。目前所使用的分子标记如 SSR、RAPD、AFLP 等, 几乎都是扩增非编码区或是在基因组中随机扩增, 一般距离目标候选基因较远, 不利于我们筛选与目标性状紧密相关的分子标记。ISAP 标记则可以促进此类进程。由于 ISAP 标记与表达序列紧密连锁, 在图谱构建时, 分子标记所在的位置也预示着表达序列的存在。而且基因组中基因序列通常串连存在于染色体上, 在进行基因定位和 QTL 检测时, 极有可能检测到的 QTL 与分子标记非常接近或者该 QTL 就包括扩增得到的表达序列片段, 这样就非常有利于图位克隆和分子标记辅助育种的操作。

4 结论

本研究统计了内含子拼接序列碱基连续出现的频率, 得到了一个拼接位点连续一致序列, 并以该序列作为核心序列建立了一种基于基因中内含子序列的功能型分子标记 ISAP。它具有简单、可靠、多态性相对较高、并且与表达序列联系紧密等特点。实验结果显示 ISAP 标记在基因组中分布均匀, 部分区域呈现标记高饱和现象, 且扩增确实发生在与表达序列近邻的区域。目前, QTL 定位及进一步的 ISAP 标记的功能型验证正在进行当中, 我们期望通过 ISAP 标记的应用能够促进目标性状候选基因的筛选, 加快分子标记辅助育种的进程。同时, 使用 ISAP 标记对其他植物进行扩增也取得了良好的效果。

参考文献(References):

- [1] Schubert R, Starck GM, Riegel R. Development of EST-PCR markers and monitoring their intrapopulation genetic variation in *Picea abies* (L.) Karst. *Theor Appl Genet*, 2001, 103: 1223–1231. [\[DOI\]](#)
- [2] Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci*, 2005, 168: 195–202. [\[DOI\]](#)
- [3] Han Zhi-Guo, Wang Chang-Biao, Song Xian-Liang, Guo Wang-Zhen, Gou Jin-Ying, Li Chun-Hong, Chen Xiao-Ya and Zhang Tian-Zhen. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor Appl Genet*, 2006, 12: 430–439. [\[DOI\]](#)
- [4] Li G, Quiros CF. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theor Appl Genet*, 2001, 103: 455–461. [\[DOI\]](#)
- [5] Hu J, Vick BA. Target region amplification polymorphism: a novel marker technique for plant genotyping. *Plant Mol Biol Rep*, 2003, 21: 289–294. [\[DOI\]](#)
- [6] Mary BS, Ju-Kyung Yu. Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms. *Plant Biotechnol J*, 2003, 167–185.
- [7] Bertin I, Zhu JH, Gale MD. SSCP-SNP in pearl millet—a new marker system for comparative genetics. *Theor Appl Genet*, 2005, 110: 1467–1472. [\[DOI\]](#)
- [8] Zhang Jing-Fa, Stewart JM. Economical and rapid method for extracting cotton genomic DNA. *Journal of Cotton Science*, 2000, 4: 193–201.
- [9] Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, Stewart JM. Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. *Plant Breed*, 2005, 124: 180–187. [\[DOI\]](#)
- [10] Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps on experimental and natural populations. *Genomics*, 1987, 1: 174–181. [\[DOI\]](#)
- [11] LIU Ren-Hu, MENG Jin-Ling. MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Hereditas(Beijing)*, 2003, 25(3): 317–321.
刘仁虎, 孟金陵. MapDraw 在 Excel 中绘制遗传连锁图的宏. *遗传*, 2003, 25(3): 317–321.
- [12] Small RL, Wendel JF. Copy number liability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics*, 2000, 155: 1913–1926.
- [13] Hughes AL, Yeager M. Comparative evolutionary rates of introns and exons in murine rodents. *J Mol Evol*, 1997, 45: 125–130. [\[DOI\]](#)
- [14] Breathnach R, Chambon P. Organization and expression of eucaryotic split gene coding for proteins. *Annual Rev of Biochem*, 1981, 50: 349–383. [\[DOI\]](#)
- [15] Michael D, Manyuan L. Intron–exon structures of eukaryotic model organisms. *Nucleic Acids Res*, 1999, 27: 3219–3228. [\[DOI\]](#)
- [16] Reinisch AJ, Dong JM, Brubaker CL, Stelly DM, Wendel JF, Paterson AH. A detailed RFLP map of cotton *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics*, 1994, 138: 829–847.