DOI: 10.3724/SP.J.1006.2010.00410

源于大豆 EST 的花生属(*Arachis*)同源 SSR 标记的开发及利用

洪彦彬 陈小平 刘海燕 周桂元 李少雄 温世杰 梁炫强*

广东省农业科学院作物研究所, 广东广州 510640

摘要: 通过拼接 394 370 条大豆 EST 共获得 82 614 条 Uni-EST, 其中 2 082 条包含 2 191 个 SSR 位点, 平均每 22.96 kb EST 出现 1 个 SSR。二核苷酸重复在大豆 EST-SSR 中所占比例最大(63.5%), 其次是三核苷酸重复(30.9%); AG/CT 在 SSR 基序(motif)中出现频率最高(35.8%), AT/AT (25.4%)次之。2 082 条 SSR-EST 共设计引物 685 对, 其中 582 对在 4 个大豆品种中得到有效扩增, 98 对检测出多态性。582 对可扩增引物在花生属中的可转移性分析表明, 大豆 EST-SSR 在花生属 9 大区组间的可转移性有所差异(12.4%~15.7%), 匍匐区组最高, 大根区组最低, 平均为 14.2%。79 对可转移性同源标记在花生区组中的多态性分析表明, 69 个 SSR 在 10 个野生种间检测出多态性, 仅 5 个在 22 个栽培品种中具多态性。对引物 ES-105 在大豆和花生区组中的扩增产物测序结果显示, 两个大豆品种间仅 SSR 位点存在 3 个 AT 重复差异, 其余序列均一致, 而大豆与花生区组间的扩增产物在序列上存在较大差异, 花生区组除缺失 SSR 位点外, 侧翼序列也存在频密的插入/缺失和置换。研究结果表明, 通过大豆 EST 开发花生属同源 SSR 标记具可行性。作为有功能的分子标记, 大豆 EST-SSR 在花生区组内多态性丰富, 可直接用于大豆-花生比较基因组研究。

关键词: 大豆; 花生属(*Arachis*); EST; SSR

Development and Utilization of Orthologous SSR Markers in *Arachis* through Soybean (*Glycine max*) EST

HONG Yan-Bin, CHEN Xiao-Ping, LIU Hai-Yan, ZHOU Gui-Yuan, LI Shao-Xiong, WEN Shi-Jie, and LIANG Xuan-Qiang*

Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

Abstract: Lack of sufficient molecular markers hindered current genetic research in peanuts (*Arachis hypogaea* L.). It is necessary to enrich molecular markers for potential use in peanut breeding programs. Recently, EST-SSRs have received much attention with the higher level of transferability to closely related species and it can be rapidly developed from EST database by data mining at low cost. In this study, we mined SSRs from the soybean EST and analyzed their transferability in peanuts (*Arachis hypogaea* L.) and its wild species. A total of 394 370 soybean ESTs were clustered to generate 82 614 Uni-ESTs, 2 082 ESTs of which contained 2 191 SSRs with 1 SSR in 22.96 kb EST. Di-nucleotide motif (63.5%), followed by tri-nucleotide (30.9%), were the most abundant in soybean EST-SSR. The top two motif sequence types with high frequency were AG/CT (35.8%) and AT/AT (25.4%). Based on the 2 082 SSR-ESTs, a total of 685 primer pairs were successfully designed and synthesized to test the amplification in four soybean cultivars. There were 582 primer pairs amplified effectively in soybean cultivars, 98 of which exhibited polymorphism. The cross-transferability of soybean EST-SSR was different among nine sections in *Arachis*, ranging from 12.4% to 15.7% with an average of 14.2%. Polymorphism analysis in *Arachis* section showed that 69 of the 79 orthologous SSR markers displayed polymorphism in ten wild species, while only five markers had polymorphism in 22 cultivars. The PCR bands amplified in soybean and *Arachis* section by primer ES-105 were cloned and sequenced. The results revealed that the polymorphism among two soybean cultivars was caused by little variation in SSR motif (AT)_n, while the polymorphism between soybean and *Arachis* section was contributed not only by the indel in SSR loci, but also by the frequency of the indel and substitution in flanking region. The work indicated that the development of orthologous SSR markers for *Arachis* from soybean EST is feasible. There was a high level of polymorphism in *Arachis* sections for soybean-derived EST-SSR, therefore, the research of comparative genomics of *Glycine* and *Arachis* can be performed due to the high transferability of functional molecular markers.

Keywords: Soybean; *Arachis*; EST; SSR

本研究由国家高技术研究发展计划(863 计划)项目(2006AA10Z156), 现代农业产业技术体系建设专项, 国家自然科学基金项目(30900907)和广东省自然科学基金项目(07117967)资助。

* 通讯作者(Corresponding author): 梁炫强, E-mail: liang804@yahoo.com; Tel: 020-87597315

第一作者联系方式: E-mail: hongyanbin1979@yahoo.com.cn; Tel: 020-87511794

Received(收稿日期): 2009-10-12; Accepted(接受日期): 2009-12-08.

花生属(*Arachis*)隶属豆科(Legume), 蝶形花亚科(Faboideae; *Papilionoideae*), 包含 9 大区组(section)、50~70 个种(species), 其中花生栽培种(*Arachis hypogaea* L.)是唯一在农业生产上种植的种。尽管 RFLP、RAPD、AFLP 等 DNA 标记在花生野生种间存在丰富的多态性^[1-3], 但在栽培品种间多态性极低^[3-5], 导致花生重要性状基因定位、标记辅助选择和比较基因组学等研究落后于其他作物。近年来系列研究表明, SSR 在花生栽培种间存在相对丰富的多态性^[6-7], 使其成为当前检测栽培种分子多态性的主要标记^[8-9]。由于对花生基因组来源的 SSR 的开发投入较大且费时费力, 目前国际上已开发的花生栽培种基因组 SSR 引物仅千余对, 难于满足花生分子标记及相关研究的需求。开发花生 SSR 被列为国际花生基因组计划的重要内容。近年来, 大量快速增长的 EST 数据成为 SSR 的重要来源, 并为 SSR 标记的开发提供了新的途径^[10]。来自 EST 的 SSR (EST-SSR)除具有一般 SSR 标记特点外, 还有以下特殊优势: (1) 开发简单、快捷、费用低。利用生物信息学方法能在短时间内完成 EST 数据的比对拼接和 SSR 位点的搜索^[11-12]。(2) 是功能标记。EST 本身就是表达基因的一部分, 通过与已知功能基因比对可预测 SSR-EST 的功能^[13]。(3) 通用性好。EST 来自转录区, 保守性较高, 具有较好的通用性, 适合作为近缘物种的同源标记^[14]。目前, 国内外学者已从水稻、小麦、玉米、甘蔗等 20 多种植物中开发了大量的 EST-SSR^[15-17], 并已在小麦、柑橘、咖啡等植物上证明 EST-SSR 在种/属间具有较高的可转移性^[18-20]。Varshney 等^[21]分析了 165 个小麦 EST-SSR 在小麦、黑麦、水稻上的具可转移性, 可转移率分别为 78.3%、75.0% 和 43.0%。Gutierrez 等^[22]采用 209 对蒺藜苜蓿(*Medicago truncatula*) EST-SSR 和 33 对 genomic-SSR 引物扩增蚕豆、鹰嘴豆和豌豆的基因组 DNA, 表明 EST-SSR (39%~43%)的可转移率几乎是 genomic-SSR (21%~24%)的两倍。He 等^[23]分析 432 个大豆 SSR 在花生栽培种中的可转移性, 表明大豆 EST-SSR(34.0%)在花生栽培种中可转移性高于 genomic-SSR (24.4%)。至今 GenBank 收录的大豆 ESTs 已达 1 386 618 条(截至 2009 年 4 月 10 日), 多数序列仍未用于开发 SSR, 而用于发展同源标记的更是寥寥无几。本文拟通过利用大豆 EST 数据开发 SSR, 采用同源扩增方法, 筛选在花生栽培种和野生种上具有可转移性的同源 SSR 标记。

1 材料与方法

1.1 试验材料

从 GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) 下载 394 370 条大豆 EST (截至 2007 年 12 月 30 日)。用于 EST-SSR 检测的大豆品种华夏 1 号、华夏 3 号、华春 1 号和华春 2 号, 由华南农业大学农学院提供; 花生野生种 20 个, 花生栽培品种 22 个, 均由广东省旱地作物资源库提供(表 1)。

1.2 基因组 DNA 提取

取每个品种苗期嫩叶, 采用 CTAB 法提取基因组 DNA^[24], 用 0.8%的琼脂糖凝胶电泳检测 DNA 质量, 并根据对照 λ DNA 的亮度估计花生基因组 DNA 浓度。

1.3 EST 序列拼接

采用 TIGR 基因指数聚类工具(TIGR Gene Indices Clustering Tools, 简称 TGICL, <http://compbio.dfci.harvard.edu/tgi/software/>)拼接大豆 EST。拼接标准是 EST 序列重叠碱基数 50 bp, 序列相似度 90%, 序列不匹配长度 20 bp。

1.4 SSR 位点搜索

采用 SSR 搜索脚本 MicroSATellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>) 在 Uni-EST 中搜索 SSR 位点。搜索标准为 SSR 长度 20 bp, 二、三、四、五、六核苷酸基序(motif)重复次数分别 10、7、6、5、5。将 MISA 配置文件(misa.ini)中代表 SSR 基序长度和最小重复次数的数据对设定为 2-10、3-7、4-6、5-5、6-5 (不包含单核苷酸类型)。

1.5 SSR 引物设计

根据 SSR 位点两端的保守区域, 采用 Primer premier 5 软件设计引物。设计引物基于下列原则: (1) 溶解温度(T_m) 50~60℃, 最适复性温度 55℃; (2) 预测扩增产物大小 100~500 bp; (3) 引物长度 18~24 bp, 预测扩增率 80%; (4) 引物 GC 含量 40%~60%。在 PE9700 热循环仪(美国 ABI 公司)上进行 PCR。

1.6 PCR 扩增

PCR 体系 20 μ L, 含 1 $\times$ PCR buffer (50 mmol L⁻¹ KCl, 10 mmol L⁻¹ Tris-HCl, pH 8.3, 1.5 mmol L⁻¹ MgCl₂, 0.1%明胶), 1 U *Taq* DNA 聚合酶, 50~100 ng 模板 DNA, 0.15 μ mol L⁻¹ 引物, 0.2 mmol L⁻¹ dNTP。反应程序为 94℃ 5 min; 94℃ 1 min, 55℃ 30 s, 72℃ 1 min, 35 个循环; 72℃ 5 min。扩增产物经 6.0%非变性聚丙烯酰胺凝胶电泳(电泳缓冲液 1 $\times$ TBE, 电压

表 1 用于大豆 EST-SSR 可转移性分析的花生栽培种和野生种
Table 1 Cultivars and wild peanut species used in transferable analysis of soybean EST-SSR

区组	种	亚种	品种	来源
Section	Species	Subspecies	Accession	Origin
大根区组 <i>Caulorrhizae</i> section	<i>A. pintoi</i> Krapov. & Gregory			巴西 Brazil
直立区组 <i>Erectoides</i> section	<i>A. paraguariensis</i> Chodat & Hassl.			巴西 Brazil
围脉区组 <i>Extranervosae</i> section	<i>A. prostrate</i> Benth.			巴西 Brazil
异形花区组 <i>Heteranthae</i> section	<i>A. pusilla</i> Benth.			巴西 Brazil
匍匐区组 <i>Procumbentes</i> section	<i>A. rigonii</i> Krapov. & Gregory			巴西 Brazil
根茎区组 <i>Rhizomatosae</i> section	<i>A. glabrata</i> Benth			巴西 Brazil
三叶区组 <i>Triseminatae</i> section	<i>A. guaranitica</i> Chodat & Hassl.			巴西 Brazil
三籽粒区组 <i>Trirectoides</i> section	<i>A. triseminata</i> Krapov. & Gregory			巴西 Brazil
花生区组 <i>Arachis</i> section	<i>A. monticola</i> Krapov. & Rigoni			阿根廷 Argentina
	<i>A. Stenosperma</i> Krapov. & Gregory			巴西 Brazil
	<i>A. correntina</i> Krapov. & Gregory			巴西 Brazil
	<i>A. cardenasii</i> Krapov. & Gregory			巴西 Brazil
	<i>A. magna</i> Krapov., Gregory & Simpson			巴西 Brazil
	<i>A. duranensis</i> Krapov. & Gregory			阿根廷 Argentina
	<i>A. Trinitensis</i> Krapov. & Gregory			巴西 Brazil
	<i>A. batizocoi</i> Krapov. & Gregory			玻利维亚 Bolivia
	<i>A. helodes</i> Mart. ex Krapov. & Rigoni			巴西 Brazil
	<i>A. villosa</i> Benth.			巴西 Brazil
	<i>A. hypogaea</i> L.	连续开花亚种	PI 393531	美国 America
		<i>ssp. fastigiata</i>	PI 390693	美国 America
			琼山花生 Qiongsshanhuasheng	中国 China
			辽宁四粒红 Liaoningsilihong	中国 China
			德豆 Dedou	中国 China
			广柳 Guangliu	中国 China
			三月拧 Sanyuening	中国 China
			粤油 20 Yueyou 20	中国 China
			Spancross	美国 America
			Tennessee Red	美国 America
		交替开花亚种	小硫球 Xiaoliuqiu	中国 China
		<i>ssp. hypogaea</i>	阳江铺地毯 Yangjiangpudizhan	中国 China
			细花勾豆 Xihuagoudou	中国 China
			耙豆 Padou	中国 China
			Bo-50	中国 China
			英德鸡豆仔 Yingdejidouzai	中国 China
			河源半蔓 Heyuanbanman	中国 China
			托克逊小花生	中国 China
			Toukexunxiaohuasheng	
			Sunoleic 97R	美国 America
			Tifrunner	美国 America
			Georgia Green	美国 America
			NC0940-22	美国 America

75 V, 时间 2.5 h), 0.1% AgNO₃ 染色, BIORAD 凝胶成像系统观察、照相、读带。

1.7 DNA 回收测序

采用 Biospin 聚丙烯酰胺 DNA 回收试剂盒回收扩增产物, 回收产物与克隆载体 pMD18-T 连接后转化大肠杆菌 DH5 α 感受态细胞, 重组子经鉴定后送

上海英骏生物技术有限公司测序。

2 结果与分析

2.1 EST 序列拼接

为了减少冗余序列, 提高 EST 序列的质量, 避免开发的 SSR 出现重叠, 采用 TGICL 工具对大豆

EST 序列进行拼接。394 370 条 EST (163.5 Mb)经拼接后共获得 82 614 条 Uni-EST (50.3 Mb), 其中 contigs 35 242 条, singletons 47 372 条。

2.2 EST-SSR 分布特点

采用 MISA 脚本从 82 614 条大豆 Uni-EST 中共搜索到 2 191 个 SSR, 分布在 2 082 条 EST 中, 其中 109 条 EST 含一个以上 SSR, 平均每 22.96 kb EST 出现 1 个 SSR。在搜索出的 SSR 中, 二至六核苷酸重复均存在, 但出现的频率差异较大。二核苷酸重复最高, 共 1 391 个, 占 63.5%, 其次是三核苷酸重复, 676 个, 占 30.9%。四、五、六核苷酸重复出现频率远低于二、三核苷酸重复, 只占 5.6% (表 2)。SSR 平均长度为 35.3 bp, 二、三、四、五、六核苷酸重复长度分别为 40.9、25.0、25.9、27.3 和 31.6 bp。

大豆 EST-SSR 种类较为丰富, 包含 78 种基序类型 (motif sequence type), 二、三、四、五、六核苷酸

重复分别含有 4、10、10、15 和 39 种基序类型。出现频率最高的 6 种基序类型为 AG/CT (35.8%)、AT/AT (25.4%)、AAG/GTT (8.5%)、AAT/ATT (6.5%)、AAC/GTT (4.5%)和 ACC/GGT (3.4%), 其余 72 种只占 15.9% (图 1)。

2.3 EST-SSR 引物设计及可扩增性验证

采用 Primer premier 5 软件对 2 082 条 SSR-EST 设计引物, 结果 685 条设计成功, 其余 1 397 条由于 SSR 侧翼序列太短或者结构复杂而无法设计。已设计引物的 SSR-EST 中, 分别含有二、三、四、五和六核苷酸重复的 330、345、19、12 和 26 个。

为验证引物的可扩增性, 采用全部引物扩增 4 个大豆品种 (华夏 1 号、华夏 3 号、华春 1 号和华春 2 号)。结果表明, 有 582 对引物能在 4 个品种中扩增出清晰条带, 其余 103 对无扩增产物或扩增带型微弱。在可扩增的引物中, 415 对扩增片段长度与预

表 2 大豆 EST 中 SSR 搜索结果
Table 2 Summary of mining SSR in soybean EST

序列类型 Type of sequence	叠连群 Contig (kb)	单序 Singleton (kb)	小计 Total (kb)
无冗余 ESTs 数目 Number of Uni-ESTs	35242 (27031)	47372 (23279)	82614 (50310)
检测的 SSR 数目 Number of identified SSRs	952	1239	2191
含 SSR 的 ESTs 数目 Number of ESTs containing SSR	877	1195	2082
含 1 个以上 SSR 的 ESTs 数目 Number of ESTs containing more than 1 SSR	75	34	109
复合型 SSR 的数目 Number of SSRs present in compound formation	29	42	71
二核苷酸 Di-nucleotide	516	875	1391 (63.5%)
三核苷酸 Tri-nucleotide	340	336	676 (30.9%)
四核苷酸 Tetra-nucleotide	27	17	44 (2.0%)
五核苷酸 Penta-nucleotide	13	11	24 (1.1%)
六核苷酸 Hexa-nucleotide	26	30	56 (2.5%)

图 1 大豆 EST-SSR 基序类型分布图
Fig. 1 Frequency distribution of soybean EST-SSR based on motif sequence type

期相近, 156 对大于预期, 11 对小于预期。此外, 有 98 对(16.9%)能在 4 个品种间检测出多态性, 其中 88 对检测出 2 个等位基因, 8 对检测出 3 个等位基因, 2 对检测出 4 个等位基因。

2.4 大豆 EST-SSR 在花生属中的可转移性

为研究大豆 EST-SSR 在花生属不同区组(section)间的可转移性, 用 582 对可扩增引物分析 9 个来自不同区组的野生种(表 3)。结果表明, 供试引物在不同区组间扩增出 3 种带型, I 为特异带, 通常包含 1~2 条主带; II 为 RAPD 带; III 为无扩增产物或者扩增带型微弱。将带型 I 和 II 的 SSR 作为可转移同源标记进行统计, 共有 96 个 SSR 至少在一个区组具可转移性, 其中 61 个在 9 个区组中均具有可转移性。不同区组间的可转移率有所差异(12.4%~15.7%), 匍匐区

组最高, 大根区组最低, 平均 14.2%。
2.5 同源标记在花生区组中的多态性
采用 79 对在花生区组具可转移性的同源标记引物扩增区组内 10 个野生种及 22 个栽培品种(表 4 和表 5)。结果表明, 全部引物在供试材料中均得到有效扩增, 其中 56 对扩增出带型 I, 23 对扩增出带型 II。具带型 I 的 SSR 引物中有 43 对能在野生种间检测出 48 个多态性位点, 每个位点等位基因数 2~6 个, 平均 3.1 个; 扩增出带型 II 的 SSR 引物在野生种间均检测出多态性, 每对引物扩增出多态性带 3~9 条, 平均 6.2 条。与野生种相反, 同源 SSR 标记在花生栽培品种间多态性偏低。79 个 SSR 仅 5 个能检测出多态性(图 1), 且均为带型 II, 多态性条带 2~5 条, 平均 3.2 条。

表 3 大豆 EST-SSR 在花生属 9 大区组中的可转移性
Table 3 Transferability of soybean EST-SSR in 9 sections of *Arachis*

区组 Section	物种 Species	可转移 EST-SSR Transferable EST-SSR		可转移率 Transfer frequency (%)
		带型 I Band pattern I	带型 II Band pattern II	
大根区组 <i>Caulorrhizae</i> section	<i>A. pinto</i> Krapov. & W.C. Gregory	56	16	12.4
直立区组 <i>Erectoides</i> section	<i>A. paraguariensis</i> Chodat & Hassl.	52	30	14.1
围脉区组 <i>Extranervosae</i> section	<i>A. prostrate</i> Benth.	53	25	13.4
异形花区组 <i>Heteranthae</i> section	<i>A. pusilla</i> Benth.	56	30	14.8
匍匐区组 <i>Procumbentes</i> section	<i>A. rigonii</i> Krapov. & W.C. Gregory	55	36	15.7
根茎区组 <i>Rhizomatosae</i> section	<i>A. glabrata</i> Benth	55	31	14.8
三叶区组 <i>Triseminatae</i> section	<i>A. guaranitica</i> Chodat & Hassl.	57	21	13.4
三籽粒区组 <i>Trirectoides</i> section	<i>A. triseminata</i> Krapov. & Gregory	51	39	15.5
花生区组 <i>Arachis</i> section	<i>A. monticola</i> Krapov. & Rigoni	56	23	13.6

图 1 引物 ES-291 在 22 个花生栽培品种中的扩增带型

Fig. 1 Amplification results of primer ES-291 in 22 peanut cultivars

M: marker; 1: PI393531; 2: PI390693; 3: 琼山花生; 4: 辽宁四粒红; 5: 德豆; 6: 广柳; 7: 三月柠; 8: 粤油 20; 9: Spancross; 10: Tennessee Red; 11: 小琉球; 12: 阳江铺地毡; 13: 细花勾豆; 14: 耙豆; 15: Bo-50; 16: 英德鸡豆仔; 17: 河源半蔓; 18: 托克逊小花生; 19: Sunoleic; 20: Tifrunner; 21: Georgia Green; 22: NC0940-22。
M: marker; 1: PI393531; 2: PI390693; 3: Qiongshanhuasheng; 4: Liaoningsilihong; 5: Dedou; 6: Guangliu; 7: Sanyuening; 8: Yueyou 20; 9: Spancross; 10: Tennessee Red; 11: Xiaoliuqi; 12: Yangjiangpudizhan; 13: Xihuagoudou; 14: Padou; 15: Bo-50; 16: Yingdejidouzai; 17: Heyuanbanman; 18: Tuokexunxiaohuasheng; 19: Sunoleic; 20: Tifrunner; 21: Georgia Green; 22: NC0940-22.

表 4 花生区组同源 EST-SSR (带型 I)的基本特征
Table 4 Characteristics of orthologous EST-SSR (band pattern II) among *Arachis* section

标记 Marker	正向引物序列 F primer (5'-3')	反向引物序列 R primer (5'-3')	SSR 基序 Motif	长度 Length (bp)	扩增长度 Amplified length (bp)	野生种等位基因数 Alleles No. in wild species
ES-20	AGGGTATGTGAGGAGCGAGAC	CGTGTTCGAACGGTATGA	(GA)36	263	120	3
ES-47	ATCGCCACCAACTATCACTTC	TAAGTTGCGGGCTTACAGGA	(AT)11	306	110	3
ES-96	ATCAGAGCACAAGCCAACCTG	CGAGTAAGGATAAGACGCTGAAGA	(TAT)8	238	140	4
ES-97	GGCAAACCTTGGTCATTACACA	CGGTCCAGGAGTATCAACAAG	(AG)16	338	150	3
ES-102	AGACGAAACCTTCCTTTACAC	TGCTGCCTCTTATTCATCCTA	(AT)11	212	140	2
ES-105	CCCCTTTGCTTCTGATGCTTT	GATTGGATGCTGGTGGGAGAT	(AT)27	252	170	6
ES-112	TTCCTTATGTTCAATTCCTCCAC	GTCAGTACGGACCAGCAGAAT	(AT)24	326	250, 470	2, 2
ES-147	ATTGGAACCTGAATTAAGGACCAC	CTGGCAACGACTAATCAATCACA	(AC)10	213	135	
ES-165	TAATAACCAGCAGGGTCAGG	TGTGGTGGTGATGGTGAGAC	(CCG)7	227	280	
ES-180	AGCGAAAGGTAGCACCCAAAT	TATCGGGATGTCGTCTTCGTC	(GAA)7	236	240	2
ES-186	ACTGGAAGGCGATGGAGTT	AAACCGAACTGTTCCCAGA	(AG)10	198	200	
ES-189	TCATGCTCACCTTCCCCT	AGCACTCGGAATCTCGTCAA	(GA)16,(AG)37	302	210	
ES-214	CCTTCGCCATGCTCCAAT	TAATGGTGGTTGCCTCCTCA	(CT)61	313	100, 160	3, 2
ES-221	CTCCACGAATCCATATCCCTC	CAGCATTACATGCCCTACGC	(GGC)7	276	620	
ES-224	GCTCCACCAACTAAGAACGG	TAGCTGCACCAAGGGAAACA	(TC)17	248	110	6
ES-235	TGGTTGGACTTGGTGGCATAG	CTGGTGCCGCCCTAACAA	(TGC GTT)5	141	90	4
ES-243	AAGTTTCTGGCAAACAACCTACCC	ATCATTCCTCAATCACTGCATCAG	(CT)37	292	500	
ES-245	TGGAAGACAACGGGAAGC	AGTTGGCGAAATGGTGGGT	(TCT)7	272	110	3
ES-254	GTTGTAAATTAAGGTGCAAGAAGC	ACTCCACCACTCTGCTCTTCC	(TA)15	183	350	5
ES-255	TCTCCTGTCGCCTTATTGAA	CCTGAAGTACCAGCAAGCAC	(TCA)7	272	185	2
ES-260	TAGTCGAGCCGGGAAAG	GGCACTGTGATATGGCAAGAG	(GGT)9	331	210	3
ES-264	TCGCTGGCTTCGCAAGTTAT	GAGGCATGGTTGAAATTAGGG	(GA)52	330	140	
ES-285	CCAAATGGTGACACGAGGAGA	ACATACGAGGGTGAGCTACTGC	(AT)24	294	150	3
ES-294	TCATCATTCTTTCTCGCACCA	CTGGCTCTTCTTCTGCTTGCT	(TTG)8	130	210	4
ES-296	GGCGTTTCACAAATGAGGTTC	CTTCAACGGTCTCCAATCCC	(ATA)7	228	450	
ES-299	TAGCGAGTATCAGCAGGTGGA	GCGCAATATGTGCAGAGGC	(TGC)8	197	140	4
ES-307	TGACGCCTCCACCAAAGC	AAACCCATCACTGCCAAAGTC	(TCT)7	170	175	2
ES-334	CACCTCTACATTGGCACGATT	TGAGATTGAGCCATTGGTGTA	(TGC)7	235	500	2
ES-350	CCTTTGGGTTTGCGTCTCAG	CGCCACCAAGTCTCCTTCATAT	(TGC)8	234	330	2

(续表 4)

标记 Marker	正向引物序列 F primer (5'-3')	反向引物序列 R primer (5'-3')	SSR 基序 Motif	长度 Length (bp)	扩增长度 Amplified length (bp)	野生种等位基因数 Alleles No. in wild species
ES-354	ACTGCCACTTCCTCCTCC	ATAAACTATCATACCCACATTCC	(GTT)9	151	200	3
ES-367	TGCCCTGACTTGAACTTG	ATGCCCAGAAATCATAAGC	(CTG)8	206	240	
ES-377	GGAACCCAAATTATGACAACG	CACCAGCACTCGCTCACTCTT	(GA)13	272	290	
ES-387	TCAACGCATCCAATTACCA	ACCAGCCAAACTTCAACTCA	(GA)19	195	130	2
ES-401	GAGTTCTCCGACCTTCTTG	ATCAGATGTCCACCTCAAACG	(CT)11	290	250	4
ES-403	GGCAAAGGAGGAGGGAGAA	GTGGGAAATAAGAACAGCAACT	(AAG)9	153	150, 320	2, 3
ES-405	GGCAAGGATAAGCGACCCA	GTACCAGGCAACTGACTACA	(ATT)13	205	140	
ES-470	ACAACCGCCGCTACGACT	TGATGAGGATGGAGTAGAGGAGC	(CCT)7	267	270	
ES-479	ACGACTACATTCCGAGCCTGAG	ACAGCCCTGCCATCACCTTC	(TCT)8	204	180	3
ES-500	TGTGGTTTGTTGCGATGC	TCAACATCCGAGCCACCATT	(GAATCA)5	147	150	4
ES-510	GTCACGAACGAAAGGGATAAT	CTTTGACCGATCTTCTGCTGT	(AAAG)6	251	250	3
ES-515	GTAGCTGCTGTGCTAGGAGATG	CACTTGAGATTCTAGCTTTTCG	(TGA)7	202	400	3
ES-520	TACCCTTGTTGATTGATTGG	CTCTGGGTTGGTAGGAGTGG	(TTG)12	189	180, 250	2, 3
ES-550	AGTCAAAGTTCTCACCCACCAT	AACGAAAGGAACAGAAGGTGC	(TTG)7	214	340	2
ES-551	CTTCCGTTTCCGATTATTCCA	TTATGTACCCGATTCTCCTCCC	(CGG)7	122	250	3
ES-566	GGACCCGCGAATGTCTCA	AACCCAGAAACAGAATCAACGA	(AG)16	222	160	3
ES-567	TTTGAGCTTAGACTTGGTTAT	CTCATCAGAAACCCGAATC	(TA)20	229	170	3
ES-570	CAGCTACTCATCGGAGACACTCG	CGGAGACTGGAGAAGCGAAGT	(ACG)8	242	160	3
ES-576	GAAAGCGGTGTTGCAGAGGTA	CAGAACCACCCTCATCCTCAA	(AT)16	301	310	4
ES-577	AGGCACAGTCTCAGAGCAACA	CGAAGCCGAAAGTGTAAGCC	(CCTAAC)9	198	250	3
ES-579	GCTTTGATTTCCTGCTGCTTC	ATTTGGAGCTGAGGCTTGCTC	(AT)13,(AT)10	184	340	
ES-593	TTGCTATCCGCTGCCTCC	CCTTGATGATGCTGCTGTCTA	(GGC)7	141	120	4
ES-618	TTTTCACCTTTTCGCACTCT	CTCAGGGAGGTTTCGATTG	(CCCAAG)5	211	260	3
ES-620	GGAAACTCAGTCCGAAACCA	GCGTGGTGAAGAAGGAGGA	(CCG)7	191	180	3
ES-625	GCTTCACTCATGCCACCACA	CCTGCTTGATGCGTATGTAGAA	(CCA)7	369	160, 410	3, 3
ES-637	CGCAATCTATCGTTGTGGC	ACGACATCTCGGTTTGGCTC	(TTA)22	244	250	4
ES-639	TTTCAACAATGGCAGTCTCAG	TCAGAGGACCCATCTTCACC	(GTGTG)5	232	160	3

2.6 EST-SSR 扩增产物的序列分析

图 2 和图 3 结果表明, 引物 ES-105 在两个大豆品种间只存在 3 个 AT 重复差异, 其余序列均一致, 而大豆与花生区组间存在较大差异(图 3)。华夏 1 号、华夏 3 号分别含有(AT)₁₂ 和(AT)₁₅ 重复序列, 但花生

栽培品种 Tifrunner 和野生种 *A. duranensis*、*A. trinitensis*、*A. batizocoi* 均缺失 AT 重复序列。除 SSR 位点缺失外, 侧翼序列在花生区组中存在插入/缺失和置换。例如在 BU544951 第 204 碱基和第 301 碱基处花生区组分别缺失(AAT)₂ 和(CAA)₃ 序列, 第

图 2 引物 ES-105 在大豆和花生区组中的扩增带型

Fig. 2 Amplification results of ES-105 in soybean and *Arachis* section

M: markers; 1: PI393531; 2: PI390693; 3: 琼山花生; 4: 辽宁四粒红; 5: 德豆; 6: 广柳; 7: 三月柠; 8: 粤油 20; 9: Spencross; 10: Tennessee Red; 11: 小硫球; 12: 阳江铺地毯; 13: 细花勾豆; 14: 耙豆; 15: Bo-50; 16: 英德鸡豆仔; 17: 河源半蔓; 18: 托克逊小花生; 19: Sunoleic; 20: Tifrunner; 21: Georgia Green; 22: NC0940-22; 23: *A. monticola*; 24: *A. stenosperma*; 25: *A. correntina*; 26: *A. cardenasii*; 27: *A. magna*; 28: *A. duranensis*; 29: *A. trinitensis*; 30: *A. batizocoi*; 31: *A. helodes*; 32: *A. villosa*; 33: 华夏 1 号; 34: 华夏 3 号; 35: 华春 1 号; 36: 华春 3 号。

M: marker; 1: PI393531; 2: PI390693; 3: Qiongsanhuasheng; 4: Liaoningsilihong; 5: Dedou; 6: Guangliu; 7: Sanyuening; 8: Yueyou 20; 9: Spencross; 10: Tennessee Red; 11: Xiaoliuqi; 12: Yangjiangpudizhan; 13: Xihuagoudou; 14: Padou; 15: Bo-50; 16: Yingdejidouzai; 17: Heyuanbanman; 18: Tuokexunxiaohuasheng; 19: Sunoleic; 20: Tifrunner; 21: Georgia Green; 22: NC0940-22; 23: *A. monticola*; 24: *A. stenosperma*; 25: *A. correntina*; 26: *A. cardenasii*; 27: *A. magna*; 28: *A. duranensis*; 29: *A. trinitensis*; 30: *A. batizocoi*; 31: *A. helodes*; 32: *A. villosa*; 33: Huaxia 1; 34: Huaxia 3; 35: Huachun 1; 36: Huachun 3.

图 3 ES-105 在大豆和花生区组间扩增产物序列比对

Fig. 3 Alignment of sequences obtained from SSR bands amplified in soybean and *Arachis* section by ES-105 primers

BU544951 为标记 ES-105 的原始 EST 序列 GenBank 登录号(<http://www.ncbi.nlm.nih.gov/nucest/22927812>); N=A, T, C, G.

BU544951 was the accession No. of original EST sequence of ES-105 in GenBank (<http://www.ncbi.nlm.nih.gov/nucest/22927812>); N=A, T, C, G.

表 5 花生区组同源 EST-SSR(带型 II)的基本特征
Table 5 Characteristics of orthologous EST-SSR (band pattern II) among *Arachis* section

标记 Marker	正向引物序列 F primer (5'-3')	反向引物序列 R primer (5'-3')	SSR 基序 Motif	长度 Length (bp)	扩增长度 Amplified length (bp)	多态性条带 Polymorphic band	
						野生种 Wild species	栽培种 Cultivar
ES-34	CCTGAAGAAACAACCCAACC	GTGGAAGAAAAGGGAAGGAAGT	(CAT)7	252	80-450	5	
ES-37	GTCATTGAAGCCTCCTCCTG	CCACCACCTTAACTCCAACC	(GTGCCC)6	179	90-600	6	
ES-101	CAGTGAGAAGGGACAGCATC	ATGATCCCTGCTTTCCTTTA	(TA)11	248	100-500	8	
ES-118	AGGCATGGTTGGTGGTAGTGT	TCGGGATGATATTCAGAGGTG	(GCT)8	138	100-370	7	
ES-169	GGCATGAGTGGACGCTATAAGA	ATCACAGGAAACAACGGAAAA	(CACT)6	155	80-600	7	2
ES-170	GAGCATAGTAACGGCGGAATC	GTGCCTCTTACTCGTCTCGTCTT	(GAGAA)5	244	80-500	8	4
ES-187	GTGGTATCATCACAGCGAAGC	ACTCTTCTCCTCCCTCGTCCT	(CT)10	116	110-420	3	
ES-200	GCTTCAGCGACGACTCCA	ACCATTCCCTCCAACCTCT	(GTT)7	292	160-400	3	
ES-204	CATCTTCAGCACTGCCATCA	CTCATCAGAACTTCTACCCAAAAC	(GA)24	206	80-600	6	
ES-251	AAGGAACAGCCAATGAGGATA	ATCATGTAATTCTTCTGCACCC	(ATC)7	266	120-570	5	
ES-256	GTAGGTATGAGGTGGTGAAGGC	CTACTATAAAGGAGTAACCACGATGA	(TC)57	228	180-1000	8	
ES-291	GACAGCGAAGTTGGTAGAGCA	TGCTTTGGCTTCACCCCTCT	(AGA)7	208	120-500	7	3
ES-293	CCTCGGTAGTCCCCTTCAC	GTAGCCCATAAACGCATTCC	(GA)10	252	130-700	9	5
ES-303	CTTCCATTGGCTGCCTTCA	ACACGCAGCACCCAATGAC	(TC)10	255	90-700	7	2
ES-331	CGTCGCAAAGAAGGTTTCAG	CTATCCACCGAGCCACCTC	(TTC)7	292	100-400	8	
ES-335	TGGGTCAGGTGGGAGCAGTA	CAAGTCAAAGCCAGGGAAGG	(CAA)7	229	140-530	5	
ES-389	TTTCGGTGCCTCTTCCTTCA	TCTAGCTGCTTCCCGTTTCC	(TAT)7	288	100-800	5	
ES-490	GCAAAGGCAGGCTTGGAGTA	AAGAGGCGTGATGAAAGGATTA	(TGT)8	221	100-800	6	
ES-531	CTTTCACAGCCATCACAGTCC	CCTCTTCCCTCCCTTCCTCT	(AT)14(AG)15	299	100-450	5	
ES-539	CCCTCCCTTTCCTCCACTG	CGTGTTTGGCCTTGTGTCT	(CAC)7,(CTC)8	337	100-800	6	
ES-559	GATGACCTGACGCCGAAGA	TTCTCGTTCATCTCCGCCAC	(TCC)7	143	100-600	6	
ES-592	GAAGGGACACCGTGGAAGA	CACAGATTGTTGGTAAAGGAGGA	(AAG)10	205	120-740	5	
ES-597	AGCCTCATCCACCATAAACC	AGTTTGGTTCATCTGGGTCAT	(AAG)7	247	90-300	7	

320 碱基处同时出现插入/缺失和置换[(TAC)₂TAG→(TAC)₁₋₆]。引物 ES-105 在花生区组内的多态主要是由第 320 碱基处(TAC)_n的插入/缺失造成。Tifrunner 两条主带的差异与野生种间的差异类似,说明两主带可能是 ES-105 在 A 和 B 基因组的扩展片段。由此可知,ES-105 在大豆和花生区组间的长度多态除与(AT)_n重复次数差异外,还与侧翼序列的插入/缺失有关。尽管单碱基置换无法通过聚丙烯酰胺凝胶电泳检测出来,但测序结果显示大豆与花生区组间存在频密的单碱基置换,依靠检测扩增片段长度不足以反映大豆与花生区组间的 DNA 多态性。

3 讨论

系列研究表明,植物不同物种间甚至同一物种内 EST-SSR 的频率、核苷酸重复类型存在一定的差异^[25-27]。Varshney 等^[25]分析大麦(75.2 Mb)、玉米(54.7 Mb)、水稻(43.9 Mb)、黑麦(3.7 Mb)、高粱(41.6 Mb)和小麦(37.5 Mb)的 EST,发现 EST-SSR 频率为 1 SSR/1~6 kb,其中三核苷酸重复占 54.0%~78.0%,二核苷酸重复占 17.1%~40.4%。Jayashree 等^[28]利用 SSR 搜索软件 SSRIT 扫描了 135.9 Mb 大豆 EST 序列,平均每 3.5 kb 检测到 1 个 SSR,其中二、三、四、五核苷酸重复比例为 10.3%、58.6%、24.1%和 6.9%。本研究中大豆 SSR 的频率为 1 SSR/22.4 kb,二、三、四、五核苷酸的比例为 63.5%、30.9%、2.0%和 1.1%,与 Jayashree 的研究结果存在很大差异。其原因除与物种间基因组差异、EST 数据来源、SSR 搜索软件有关外,还受 SSR 搜索标准的影响。本研究中设置的标准为 SSR 20 bp,如果按照 Jayashree 的标准(SSR 12 bp)进行搜索,SSR 的频率将变为 1 SSR/2.9 kb,二、三、四、五核苷酸重复比例为 21.4%、48.8%、22.6%和 7.2%,与 Jayashree 的结果类似。

引物设计受诸多因素影响,并非能够完全精确无误,与 genomic-SSR 一样,EST-SSR 引物常常存在部分(10%~40%)无效扩增^[16,29-32],本研究中有 20%的 EST-SSR 引物在大豆基因组 DNA 中无扩增产物。EST-SSR 无效扩增的原因可能是 SSR 一端或两端引物刚好位于内含子/外显子剪切位点,或基因组 DNA 序列中存在大内含子,或用于设计引物的 DNA 序列存在问题。据统计,大约 9%的禾本科植物 EST 序列存在质量问题^[33]。除无效扩增外,本研究中有 167 对引物扩增片段明显大于或小于预期长度,原因可能与基因组序列存在内含子,或者发生大的

插入/缺失有关。

ES-105 在大豆和花生上的扩增产物序列分析表明,SSR 序列在花生区组中存在缺失/缩短,类似现象也出现在大豆^[34]、拟南芥^[35]、番茄^[36]、小麦^[37]等植物基因组 SSR 上。Treuren 等^[35]等分析 10 个拟南芥 SSR 位点在拟南芥属其他种上的同源序列,9 个存在序列缩短现象。目前,关于此现象存在多种解释,其中测量偏倚(ascertainment bias)理论^[38-39]被学术界普遍接受。该理论认为在开发 SSR 过程中,通常倾向于选择多态性高的 SSR,其序列长度往往在目标物种中已接近最大值,显然,近缘物种上对应的序列长度必定较短。结合前人关于 SSR 序列长度影响多态性的研究结论^[40],初步判断大豆 EST-SSR 在花生栽培品种间多态性低的原因与 SSR 重复序列缩短有关。

多数研究表明,SSR 在属间的可转移性明显低于种间,而科间几乎没有转移性。且 EST-SSR 的可转移性通常比 genomic-SSR 更高。本研究中大豆 EST-SSR 在花生属中的可转移率为 12.4%~15.7%,与 Peakall 等^[34]的研究结果相符,但与 He 等^[23]的研究结果差异较大,原因除供试引物不同外,还可能与本研究未将扩增带型较弱的引物列入统计范围有关,如将其列入,可转移率将提高至 28.4%。另外,根据 ES-105 的测序结果可推测,大豆 EST-SSR 在花生属可转移性较低可能与 SSR 侧翼序列存在频密的插入/缺失和置换有关。如果引物序列在花生区组基因组对应的位置刚好存在插入/缺失或者置换,将导致扩增失败。尽管大豆 EST-SSR 在花生上的可转移率偏低,但由于大豆 EST 数据庞大,包含的 EST-SSR 数量多,利用生物信息学从大豆 EST 开发花生属同源 SSR 标记仍然较传统的方法省时省力。

4 结论

利用大豆 EST 开发花生同源 SSR 标记是可行的。作为有功能的分子标记,尽管大豆 EST-SSR 在花生栽培品种间检测的多态性偏低,但在花生区组内存在丰富的多态性,可直接用于大豆-花生比较基因组研究。

References

- [1] Kochert G, Halward R, Branch W D, Simpson C E. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor Appl Genet*, 1991, 81: 565-570
- [2] Stalker H T, Dhesi J S, Kochert G. Genetic diversity within the

- species *Arachis duranensis* Krapov. & W. C. Gregory, a possible progenitor of cultivated peanut. *Genome*, 1995, 38: 1201–1212
- [3] Raina S N, Rani V, Kojima T, Ogihara Y, Singh K P, Devarumath R M. RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. *Genome*, 2001, 44: 763–772
- [4] Subramanian V, Gurtu S, Rao R C N, Nigam S N. Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. *Genome*, 2000, 43: 656–660
- [5] Milla S R, Isleib T G, Stalker H T. Taxonomic relationships among *Arachis* sect: *Arachis* species as revealed by AFLP markers. *Genome*, 2005, 8: 1–11
- [6] Tang R H, Gao G Q, He L Q, Hang Z Q, Shan S H, Zhong R C, Zhou C Q, Jiang Q, Li Y R, Zhuang W J. Genetic diversity in cultivated groundnut based on SSR markers. *J Genet Genom*, 2007, 34: 449–459
- [7] Hong Y-B(洪彦彬), Liang X-Q(梁炫强), Chen X-P(陈小平), Lin K-Y(林坤耀), Zhou G-Y(周桂元), Li S-X(李少雄), Liu H-Y(刘海燕). Genetic diversity analysis in botanical varieties of the cultivated peanut (*Arachis hypogaea* L.) based on SSR polymorphism. *Mol Plant Breed* (分子植物育种), 2008, 6(1): 71–78 (in Chinese with English abstract)
- [8] Hong Y B, Liang X Q, Chen X P, Liu H Y, Zhou G Y, Li S X, Wen S. Construction of genetic linkage map based on SSR markers in peanut (*Arachis hypogaea* L.). *Agric Sci China*, 2008, 7: 915–921
- [9] Hong Y-B(洪彦彬), Liang X-Q(梁炫强), Chen X-P(陈小平), Liu H-Y(刘海燕), Zhou G-Y(周桂元), Li S-X(李少雄), Wen S-J(温世杰). Construction of genetic linkage map in peanut (*Arachis hypogaea* L.) cultivars. *Acta Agron Sin* (作物学报), 2009, 35(3): 395–402 (in Chinese with English abstract)
- [10] Liang X Q, Chen X P, Hong Y B, Liu H Y, Guo B Z. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol*, 2009, 9: 35
- [11] Benson G. Tandem repeats finder: A program to analyze DNA sequences. *Nucl Acids Res*, 1999, 27: 573–580
- [12] Castelo A T, Martins W, Gao G R. TROLL-tandem repeat occurrence locator. *Bioinformatics*, 2002, 18: 634–636
- [13] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol*, 2005, 23: 48–55
- [14] Choudhary S, Sethy N K, Shokeen B, Bhatia S. Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor Appl Genet*, 2009, 118: 591–608
- [15] Kantety R V, La Rota M, Matthews D E, Sorrells M E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*, 2002, 48: 501–510
- [16] Yu J K, Dake T M, Singh S, Benscher D, Li W, Gill B, Sorrells M E. Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. *Genome*, 2004, 47: 805–818
- [17] Rossi M, Araujo P G, Paulet F, Garsmeur O, Dias V M, Chen H, Van Sluys M A, D'Hont A. Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Mol Genet Genomics*, 2003, 269: 406–419
- [18] Castillo A, Budak H, Varshney R K, Dorado G, Graner A, Hernandez P. Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. *BMC Plant Biol*, 2008, 8: 97
- [19] Luro F L, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other citrus species and their effectiveness for genetic mapping. *BMC Genomics*, 2008, 9: 287
- [20] Aggarwal R K, Hendre P S, Varshney R K, Bhat P R, Krishnakumar V, Singh L. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet*, 2006, 114: 359–372
- [21] Varshney R K, Sigmund R, Borner A, Korzun V, Stein N, Sorrells M E, 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
- [22] Gutierrez M V, Patto M C V, Huguet T, Cubero J I, Moreno M T, Torres A M. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor Appl Genet*, 2005, 110: 1210–1217
- [23] He G H, Woullard F E, Marong I, Guo B Z. Transferability of soybean markers in peanut (*Arachis hypogaea* L.). *Peanut Sci*, 2006, 33: 22–28
- [24] Lu S-D(卢圣栋). Laboratory Techniques for Modern Molecular Biology (现代分子生物学实验技术). Beijing: Chinese Academy of Medical Sciences & Peking Union Medical College Press, 1999. pp 101–136 (in Chinese)
- [25] Varshney R K, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett*, 2002, 7: 537–546
- [26] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. *Nat Genet*, 2002, 30: 194–200
- [27] Gao L F, Tang J F, Li H W, Jia J Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breed*, 2003, 12: 245–261
- [28] Jayashree B, Punna R, Prasad P, Bantte K, Hash C T, Chandra S, Hoisington D A, Varshney R K. A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: Survey and evaluation. *In silico Biol*, 2007, 6: 607–620
- [29] Saha M C, Mian M A, Eujayl I, Zwonitzer J C, Wang L, May G D. Tall fescue EST-SSR markers with transferability across several grass species. *Theor Appl Genet*, 2004, 109: 783–791
- [30] Thiel T, Michalek W, Varshney R K, Graner A. Exploiting EST

- databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*, 2003, 106: 411–422
- [31] Cordeiro G M, Casu R, McIntyre C L, Manners J M, Henry R J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci*, 2001, 160: 1115–112
- [32] Gupta P K, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics*, 2003, 270: 315–323
- [33] Sreenivasulu N, Kavikishor P B, Varshney R K, Altschmied L. Mining functional information from cereal genomes—the utility of expressed sequence tags. *Curr Sci*, 2002, 83: 965–973
- [34] Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Mol Biol Evol*, 1998, 15: 1275–1287
- [35] Treuren V R, Kuittinen H, Karkkainen K, Baenagonzalez E, Savolainen O. Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol Biol Evol*, 1997, 14: 220–229
- [36] Provan J, Powell W, Waugh R. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theor Appl Genet*, 1996, 92: 1078–1084
- [37] Röder M S, Plaschke J, König S U, Börner A, Sorrells M E, Tanksley S D, Ganai M W. Abundance, variability and chromosomal location of microsatellites in wheat. *Mol Gen Genet*, 1995, 246: 327–333
- [38] Ellegren H, Primmer C R, Sheldon B C. Microsatellite ‘evolution’: Directionality or bias. *Nat Genet*, 1995, 11: 360–362
- [39] Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon B C. Microsatellite evolution: A reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol Biol Evol*, 1997, 14: 854–860
- [40] Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Genet*, 2004, 5: 435–445