

皂荚 EST-SSR 分子标记开发与分析评价

林富荣¹, 邢俊连^{1,2}, 孟艳琼², 黄平¹, 郑勇奇¹

(¹中国林业科学研究院林业研究所/国家林木遗传育种重点实验室/国家林业局植物新品种分子测试实验室/国家林业局林木培育重点实验室, 北京 100091; ²安徽农业大学林学与园林学院, 合肥 230036)

摘要:本研究旨在开发皂荚 EST-SSR 分子标记, 为今后皂荚种质资源评价与分析提供基础。首先通过对已公布的皂荚转录组数据进行拼接得到 41003 个 Unigenes, 总长度 70.4 Mb, 平均长度 1716 bp, N50 为 2533 bp。进一步在 7009 个 Unigenes 中检测到 8494 个 EST-SSR 位点, 其中 1200 条 Unigenes 含有 2 个及以上的 SSR 位点, 复合型 SSRs 有 369 个。针对所有 EST-SSR 位点设计得到 6494 条特异性引物, 随机挑选 60 个位点进行试验验证与分析, 其中 44 对引物可扩增出特异性片段, 17 对具有多态性, PIC 值范围为 0.195~0.742, 均值为 0.501, 且大部分多态性位点位于 UTR 区域。通过对皂荚近缘种进行跨种 PCR 扩增试验, 结果显示在开发的 44 对有效引物中, 9 个近缘种美国皂荚、日本皂荚、绒毛皂荚、滇皂荚、华南皂荚、小果皂荚、野皂荚、肥皂荚和美国肥皂荚分别有 32、31、23、24、7、40、25、18 和 18 对引物可获得有效扩增片段, 说明 EST-SSR 标记在皂荚近缘种之间具有良好的通用性。研究表明利用转录组数据挖掘的 EST-SSR 位点具有扩增稳定、多态性良好、近缘种间通用等优点, 是林木物种开发分子标记的有效途径之一。

关键词: 皂荚; 转录组; EST-SSR; 多态性; 通用性

Development and Assessment of EST-SSR for *Gleditsia sinensis* Lam. Based on Transcriptome Sequences

LIN Fu-rong¹, XING Jun-lian^{1,2}, MENG Yan-qiong², HUANG Ping¹, ZHENG Yong-qi¹

(¹Research Institution of Forestry, Chinese Academy of Forestry/State Key Laboratory of Tree Genetics and Breeding/Lab of Molecular Identification of Plant Varieties/Key Laboratory of Silviculture of The State Forestry Administration, Beijing 100091; ²College of Forestry and Landscape Architecture, Anhui Agriculture University, Hefei 230036)

Abstract: The objective of this study is to develop EST-SSR markers for further molecular evaluation of genetic resource for *G. sinensis*. Based on the transcriptome RNA-seq data from NCBI, a total of 41003 Unigenes were obtained by assembling the RNA-seq data of *G. sinensis*. The total length, average length and N50 of the Unigenes were 70.4 Mb, 1716 bp and 2533 bp, respectively. 8494 EST-SSR loci were detected in 7009 Unigenes, among which 1200 Unigenes contained two or more SSRs and 369 Unigenes contained compound SSRs. For all EST-SSR loci, 6494 primers were designed and 60 from them were randomly selected to validate their effectiveness. Consequently, 44 primer pairs could amplify target products, among which 17 primer pairs showed polymorphism and most polymorphism markers were identified to be located in UTR region. Furthermore, the PIC value ranged from 0.195 to 0.742, with an average of 0.501. Finally, we validated whether the primers could amplify target products in relatives of *G. sinensis*, among the 44 effective primers, 32, 31, 23, 24, 7, 40, 25, 18, and 18 primer pairs could amplify target products in *G. tricanthos*, *G. japonica* Miq., *G. japonica* var. *velutina* L. C. Li., *G. japonica* var. *delavayi* (Franch.), *G. fera* (Lour.) Merr., *G. australis* Hemsl., *G. microphylla* D. A. Gordon ex Y. T. Lee, *Gymnocladus chinensis* Baill., *Gymnocladus dioica* (L.) K. Koch, indicating the high transferability of our developed primers. The results demonstrated that mining EST-SSR loci using transcriptome data is one of the efficient ways for developing molecular

收稿日期: 2016-02-29 修回日期: 2016-07-14 网络出版日期: 2016-12-14

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20161214.1415.022.html>

基金项目: 林业公益性行业专项(201204307); 国家林木(含竹藤花卉)种质资源平台(2005DKA21003)

第一作者研究方向为林木种质资源。E-mail: linfr888@163.com

通信作者: 郑勇奇, 研究方向为林木遗传育种。E-mail: zyzq8565@126.com

markers in tree species, with the advantages of steady amplification, high polymorphism and high transferability.

Key words: *Gleditsia sinensis* Lam.; transcriptome sequences; EST-SSR; polymorphism; transferability

皂荚 (*Gleditsia sinensis* Lam.) 为豆科皂荚属植物 ($2n=28$), 具有悠久的栽培历史, 是一种多用途生态经济型乡土树种, 广泛应用于园林绿化、生态造林。同时因荚果富含皂甙, 具有发泡性和去污能力成为洗涤化妆品的天然原料, 且皂刺含有丰富黄酮类、甾类等化学成分而成为医药的重要来源^[1]。皂荚适应性极强, 在石灰质、盐碱地甚至黏土砂土均能正常生长。皂荚在我国分布广泛, 北起河北、山西, 南达福建、广东、广西, 西至陕西、宁夏、甘肃、四川、贵州、云南, 东及山东、江苏、浙江等省(区)^[2]。

目前关于皂荚的研究主要集中在育苗栽培技术、有效化学成分分析与利用^[1]、种源试验分析^[3-4]。分子标记的缺乏限制了皂荚种质资源分子鉴定和遗传多样性研究, 影响了皂荚种质资源保存策略的制定。随着分子生物学的兴起, 分子标记手段已被应用于植物分类、系统进化、品种鉴定和遗传多样性分析等研究^[5]。常见的分子标记有 RAPD、AFLP、ISSR、SSR、SNP 等, 而 SSR 分子标记与其他标记相比具有共显性、信息量高、覆盖广、易操作、多态性高等优点, 而被广泛应用^[6-7]。

SSR (simple sequence repeats) 由 1~6 个核苷酸串联重复单元组成, 是广布于真核生物和原核生物基因组中的遗传标记。传统的利用基因数据开发 SSR 分子标记具有开发成本高、试验步骤繁琐、开发位点数量少等缺点。近年来, 随着测序成本降低, 转录组测序已经成为分子生物学研究的重要方法之一^[8], 越来越多的转录组数据被上传至公共数据库。在海量的转录组数据中蕴藏着大量的遗传信息, 但并未被充分利用, 仍有待科研人员深度挖掘。此外, 生物信息学的快速发展为转录组信息挖掘提供了技术保障, 利用生物信息学方法对林木树种进行 EST-SSR 标记开发的研究报道也越来越多^[9-11]。

本研究利用 NCBI 中公布的皂荚转录组数据, 对其进行二次挖掘, 进行 EST-SSR 开发与引物筛选, 找到适用于皂荚 EST-SSR 标记研究的引物, 并证明其在本皂荚和美国皂荚中的通用性, 将为以后利用 EST-SSR 分子标记进行皂荚遗传多样性分析、皂荚属亲缘关系鉴定、遗传图谱构建及功能基因挖掘等提供有效的分析方法和手段。

1 材料与方法

1.1 植物材料与 DNA 提取

选取来源于 6 个不同地点的 12 份皂荚叶片 (每地点 2 份) 作为分析材料, 对引物进行有效性筛选和多态性分析, 另选取皂荚 9 个近缘种叶片材料, 利用筛选出的有效引物进行引物通用性检测 (表 1)。

表 1 用于引物开发与检测引物通用性的试验材料

Table 1 The materials used for SSR primer developing and transferability validation

编号 No.	种名 Species	采集地 Source
1	皂荚 <i>G. sinensis</i> Lam.	北京植物园 Beijing Botanical Garden
2	皂荚 <i>G. sinensis</i> Lam.	山东枣庄 Zaozhuang, Shandong
3	皂荚 <i>G. sinensis</i> Lam.	山东费县 Feixian, Shandong
4	皂荚 <i>G. sinensis</i> Lam.	湖北宜昌 Yichang, Hubei
5	皂荚 <i>G. sinensis</i> Lam.	湖南城步 Chengbu, Hunan
6	皂荚 <i>G. sinensis</i> Lam.	陕西蓝田 Lantian, Shanxi
7	美国皂荚 <i>G. triacanthos</i>	山东枣庄 Zaozhuang, Shandong
8	日本皂荚 <i>G. japonica</i> Miq.	北京植物园 Beijing Botanical Garden
9	绒毛皂荚 <i>G. japonica</i> var. <i>velutina</i> L. C. Li	武汉植物园 Wuhan Botanical Garden
10	滇皂荚 <i>G. japonica</i> var. <i>delavayi</i> (Framch.) L. C. Li	昆明植物园 Kunming Botanical Garden
11	华南皂荚 <i>G. fera</i> (Lour.) Merr.	广州起义烈士陵园 Guangzhou Insurrectional Martyr Cemetery Park
12	小果皂荚 <i>G. australis</i> H. emsl.	华南植物园 South China Botanical Garden
13	野皂荚 <i>G. microphylla</i> D. A. Gordon ex Y. T. Lee	北京植物园 Beijing Botanical Garden
14	肥皂荚 <i>Gymnocladus chinensis</i> Baill.	武汉植物园 Wuhan Botanical Garden
15	美国肥皂荚 <i>Gymnocladus dioica</i> (L.) K. Koch	北京林业大学 Beijing Forestry University

所用材料的 DNA 均采用天根生物公司 (北京) 新型植物基因组 DNA 提取试剂盒按照说明书进行提取,

并采用多功能酶标仪(SpectraMax i3, 上海)检测所提 DNA 浓度和 OD_{260}/OD_{280} 值, 检测结果在 1.6~1.9 之间的 -20 °C 保存备用, 检测不合格的重新提取。

1.2 转录组数据来源

皂荚转录组数据来源于 NCBI 公共数据库 SRA (<http://www.ncbi.nlm.nih.gov/sra>, SRX 365131), 该转录组数据是基于皂荚的嫩芽、嫩叶、花蕾和枝刺的全转录组, 使用 Illumina HiSeq 2000 进行高通量测序, 原始测序短序列(reads) 压缩文件大小为 5 Gb^[12]。首先通过 Sratoolkit 软件将 SRA 文件转化成 fastq 数据, 进一步通过数据过滤去掉低质量和长度短于 50 bp 的序列。

1.3 EST-SSR 搜索与引物设计

皂荚转录组数据采用 Trinity 软件进行拼接^[13], 拼接命令参数为 Trinity.pl --seqType fq --JM 50G --left SRX365131_1.fq --right SRX365131_2.fq --output GS_unigenes --CPU 48 --jaccard_clip --inchworm_cpu 48 --bflyCPU 48, 使用 Cap3 (<http://seq.cs.iastate.edu/cap3.html>) 进行进一步组装获得 Unigenes 库, Cap3 采用的参数为软件默认。对皂荚的 Unigenes 库利用 MISA 程序使用默认参数进行 SSR 位点搜索, 基因编码结构采用 Trinity 软件包中的 Transdecoder 进行检测, 参数为 TransDecoder -t GS_unigenes.fa -m 30 --CPU 16。

调用 Primer 3.0 引物批量设计程序对含有 SSR 位点的 Unigenes 序列设计引物, 设置 SSR 位点侧翼序列长度 ≥ 50 bp。引物设计的主要参数标准为: (1) 退火温度(T_m) 在 57~63 °C, 最佳 60 °C, 上下游引物 T_m 相差 ≤ 5 °C; (2) PCR 扩增产物大小在 100~300 bp, 以 150 bp 为最佳长度, 引物长度在 18~23 bp, 以 20 bp 为最合适; (3) GC 含量在 40%~60%, 最佳值为 50%, 且避免复合型 SSR、引物二聚体和二级结构出现。

1.4 PCR 扩增体系与 SSR 引物筛选

随机选择基序长度为 2~6 bp 的 60 对引物, 由北京美吉桑格生物公司合成, $2 \times Taq$ PCR MasterMix (10 mM Tris-HCl, pH 8.3; 50 mM $MgCl_2$, 250 μM Dntp each, Taq 酶 0.5 U/ μL) 购自北京艾德莱生物公司。PCR 选用 20 μL 体系: DNA (20 ng/ μL) 2 μL , 灭菌去离子水 7 μL , $2 \times Taq$ PCR MasterMix 10 μL , 正反引物各 0.5 μL (相当于 5 pmol)。PCR 扩增均采用 Touchdown 模式: 94 °C 预变性 5 min; 94 °C 变性 30 s, 63 °C 退火 45 s (每个循环降低 1 °C), 72 °C 延伸 45 s, 10 个循环; 94 °C 变性 30 s, 55 °C 退火 45 s, 72 °C 延伸

45 s, 20 个循环; 72 °C 延伸 7 min。所得扩增产物用 6% 非变性聚丙烯酰胺凝胶进行检测, 170 V 电压电泳 100 min, 银染显色。选择扩增条带清晰, 具有多态性位点的引物, 进而合成 FAM 荧光引物, 对 12 份皂荚 DNA 进行 PCR 扩增, 用毛细管电泳仪 ABI 3730XL 对产物进行毛细管电泳, 采用 Genemarker 2.2.0 软件进行数据整理分析, 用 Popgene32 计算等位基因频率和 Shannon 信息指数, 用 PIC-calc 0.6 计算多态性信息量 (PIC)。利用从皂荚筛选出的有效引物对日本皂荚和美国皂荚进行扩增, 并用聚丙烯酰胺凝胶电泳检测引物的通用性。

2 结果与分析

2.1 SSR 在转录组中的分布特点

通过文件转化共得到 8.2 Gb 的 fastq 格式的 reads 数据, 过滤去掉低质量和长度短于 50 bp 的序列, 获得 5.6 Gb 的数据。拼接共获得 41003 条 Unigenes (http://bioinformatics.caf.ac.cn/resource/GS_unigenes.zip), 总覆盖长度 70.4 Mb, 序列平均长度 1716 bp, N50 值为 2533 bp, 表明拼接结果的可靠性, 同时对拼接得到的 Unigenes 进行了功能注释 (http://bioinformatics.caf.ac.cn/resource/function_annotation_desc.xls)。对其进行 SSR 位点搜索, 在其中的 7009 条 Unigenes 中检测到 8494 个 EST-SSR 位点, SSR 分布频率 (SSR 数量与总 Unigene 数量之比) 为 20.7%, 平均 8 kb 长度就有一个 SSR 位点。1200 条 Unigenes 序列中含有 2 个及以上的 SSR 位点, 发生频率为 2.93%, 其中复合型 SSRs 有 369 个。

2.2 EST-SSR 类型及特征

从皂荚 EST-SSR 核苷酸基序类型来看, 8494 个 SSR 位点共包含 651 种重复单元, 1~6 核苷酸重复分别有 4、10、59、67、194 和 317 种。EST-SSR 主要重复类型为单核苷酸重复和二核苷酸重复, 分别占总量的 30.3% 和 24.0%, 其次为三核苷酸, 占总数的 18.1%, 四核苷酸重复次数最少, 只占 5.5% (表 2)。

从 SSR 重复频率来看, 单核苷酸主要分布在 12~15 次, 其中以 A/T 重复最多, 占单核苷酸总数的 97.1%; 二核苷酸主要分布在 7~11 次, 以 AG/CT 为主, 占二核苷酸总数的 69.5%; 三核苷酸 98.7% 分布在 5~8 次, 以 AAG/CTT 最多, 占三核苷酸的 28.54%; 四核苷酸 99.6% 分布在 4~5 次, 其中 AAAG/CTTT 和 AAAT/ATTT 最多, 占其基序类型的 23.4% 和 39.7%; 五、六核苷酸几乎分布在

表 2 皂荚转录组 SSR 重复单元的分布特征
Table 2 Distribution of the SSR motifs in *G. sinensis* transcriptome

基序长度 (bp) Motif length	总计 Total	占总数百分 比(%) Ratio	主要 SSR 单元 Main repeat motif
1	2575	30.3	A/T、C/G
2	2037	24.0	AG/CT、AC/GT、AT/AT、
3	1538	18.1	AAG/CTT、AGC/CTG、AAT/ ATT、AGG/CCT
4	466	5.5	AAAG/CTTT、 AAAT/ATTTT、ACAT/ATGT
5	976	11.5	AAAAG/CTTTT、AAAAT/ ATTTT、AAAAC/GTTTT
6	902	10.6	AAAAAG/CTTTTT、 AAAAAC/GTTTTT、 AAAAAT/ATTTTT
总计 Total	8494	100	

3 ~ 4 次,以 3 次重复最多,五核苷酸主要有 AAAAG/CTTTT、AAAAT/ATTTT、AAAAC/GTTTT,分别占其基序类型的 28.3%、18.6% 和 6.27%,六核苷酸 AAAAAG/CTTTTT、AAAAAC/GTTTTT、AAAAAT/ATTTTT 最多,合计占六核苷酸总数的 25.6% (表 2,图 1)。

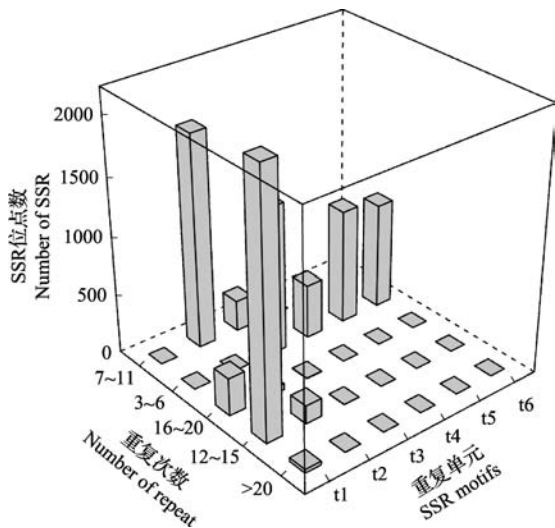


图 1 SSR 重复单元及重复次数的分布

Fig.1 The distribution of SSR motifs and repeat number

皂荚转录组 SSR 序列平均长度为 16.19 bp,一、二、三、四、五、六核苷酸重复的平均长度分别为

13.64 bp、17.90 bp、16.90 bp、16.93 bp、15.85 bp、18.88 bp。其中分布在 12 ~ 15 bp 的最多,占总数的 53.5%,16 ~ 19 bp 占 28.3%,20 ~ 23 bp 占 9.1%,24 ~ 27 bp 占 3.6%,28 ~ 31 bp 占 1.2%,32 bp 以上占 4.3% (图 2)。

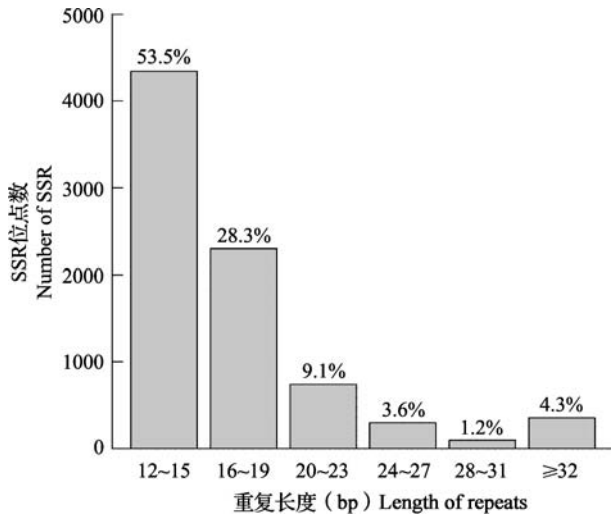


图 2 SSR 重复长度分布

Fig.2 Distribution of the SSR repeat length

2.3 EST-SSR 引物的有效性、多态性及通用性检测

对 7009 条 Unigenes 的 8494 个 EST-SSR 位点进行引物设计,得到了 6494 对 SSR 引物。随机选择合成 60 对引物,以 12 份皂荚材料为 DNA 模板,共有 44 对引物可扩增出目的特异性条带,有效扩增率为 73.3%。其中 17 对引物的扩增产物具有多态性,占有有效引物的 38.6%。根据对 17 对引物扩增位点所在基因位点的分析发现,6 个位点位于 3'非编码区 (3'-UTR, 3'-Untranslated region), 8 个位点位于 5'非编码区 (5'-UTR, 5'-Untranslated region), 3 个位于编码区 (CDS, coding sequence) (表 3)。

毛细管电泳结果显示,12 个样本在 17 个 SSR 位点检测到 60 个等位基因,平均每对引物检测到 3.53 个等位基因,多态性信息量 (PIC) 为 0.195 ~ 0.742,平均为 0.501,其中 PIC 是衡量基因遗传变异程度的指标^[14],根据 D. Botstein 等^[14]对多态性水平的定义,在开发出的 17 个标记中有 10 个标记为高度多态性位点 (PIC > 0.5), 6 个标记为中度多态性位点 (0.25 < PIC < 0.5), 1 个标记为低度多态性位点 (0 < PIC < 0.25) (表 4)。图 3 为多态性引物 GS-30 分别扩增 6 个不同地点皂荚模板 DNA 的毛细管电泳图。

表 3 17 对 SSR 引物信息

Table 3 Information of 17 pairs of SSR primers

编号 No.	上游引物 Forward primer	下游引物 Reverse primer	位置 Position	SSR 类型 SSR motif	预期大小(bp) Expected size
GS-1	CTGGTGACAAGCACGCTAAA	AGGGAGAATGTTATTGGGG	3'-UTR	(CT)10	292
GS-3	TGTGGCGTCGTTTGATCTAC	CTATGGCTCCCAGACACCAT	3'-UTR	(GTT)5	249
GS-5	CAGCAAAGCTCATCACCAAA	CTCCCGAATCCTGAAACAAA	3'-UTR	(TTTC)4	285
GS-8	TGGGGGATGTTTATTTTGA	GAAGCATTGAAGACAACGCA	5'-UTR	(TTTGG)3	137
GS-9	CAACACATGCAGGCTTATGG	TGTACAGCAGGTTGGAGCTG	CDS	(CCTCAA)3	202
GS-16	GCGAGACCCAGTCTCTCAAC	CCACAAAGCAAACAGCAAAA	5'-UTR	(TC)10	256
GS-18	AGTCCACATGACTCTTGCC	GGAAGCCCAATGAAGAAACA	3'-UTR	(AT)10	151
GS-28	TGCAACGGCAGGATACATAA	TTTGATTGCTCTCGTCCCTC	CDS	(ATC)7	120
GS-29	CGCACAGTCTTCTGCTCTTG	CCTTCGTCGTCTTCTTCGTC	5'-UTR	(GAA)7	253
GS-30	GTATGCCCATACCCCTTCTT	TGAAGCTGAAGCAGAAGCAA	5'-UTR	(TTC)8	231
GS-31	GGTCAGCTTTTGAACAACCC	CTGTCTCTGAAGACATTGTGTC	5'-UTR	(AAAG)5	171
GS-34	TCATCGCAGCTATTCGTTTG	AGTAGAGGGCTGAGGAAGCC	3'-UTR	(CTTC)4	186
GS-36	AATGTGAACCTTCGGCTCAG	CCCTTAAACCCAGTTGCAGA	5'-UTR	(AGAT)4	203
GS-37	GAACCTGGCACCTATCAAA	GAAGGGGCAAATTTGTCAAG	3'-UTR	(AAAT)5	281
GS-38	ATCTTAGTGGCCGTGTTTGG	ACGCATTGTGAGGTTTTTCC	5'-UTR	(CCTT)5	138
GS-40	GGGAAAGTGCCATTAACCCCT	TATGCACACTGAAGAAGGCG	5'-UTR	(ATTT)5	259
GS-53	GGGAACGATGTTGAGGAAAA	TCCCTGTGACCTTAGTGCC	CDS	(TGAAGA)4	277

表 4 SSR 引物的扩增产物及多态性分析

Table 4 The SSR amplification products and their polymorphism

SSR 引物 SSR primer	等位基因数量 Overall allele number	等位基因频率 Overall allele frequency	Shannon 信息指数 Shannon's information index (<i>I</i>)	多态性信息量 PIC
GS-1	4	0.042 ~ 0.500	1.137	0.573
GS-3	2	0.250 ~ 0.750	0.562	0.305
GS-5	2	0.375 ~ 0.625	0.662	0.359
GS-8	4	0.083 ~ 0.458	1.191	0.595
GS-9	4	0.042 ~ 0.750	0.815	0.386
GS-16	5	0.083 ~ 0.375	1.455	0.697
GS-18	6	0.042 ~ 0.292	1.596	0.742
GS-28	4	0.042 ~ 0.458	1.176	0.597
GS-29	4	0.042 ~ 0.542	1.038	0.515
GS-30	4	0.125 ~ 0.417	1.298	0.654
GS-31	4	0.208 ~ 0.292	1.379	0.699
GS-34	2	0.417 ~ 0.583	0.679	0.368
GS-36	2	0.375 ~ 0.625	0.662	0.359
GS-37	3	0.042 ~ 0.625	0.792	0.408
GS-38	4	0.042 ~ 0.583	1.072	0.535
GS-40	4	0.042 ~ 0.583	1.053	0.523
GS-53	2	0.125 ~ 0.875	0.377	0.195
平均 Mean	3.53		0.997	0.501

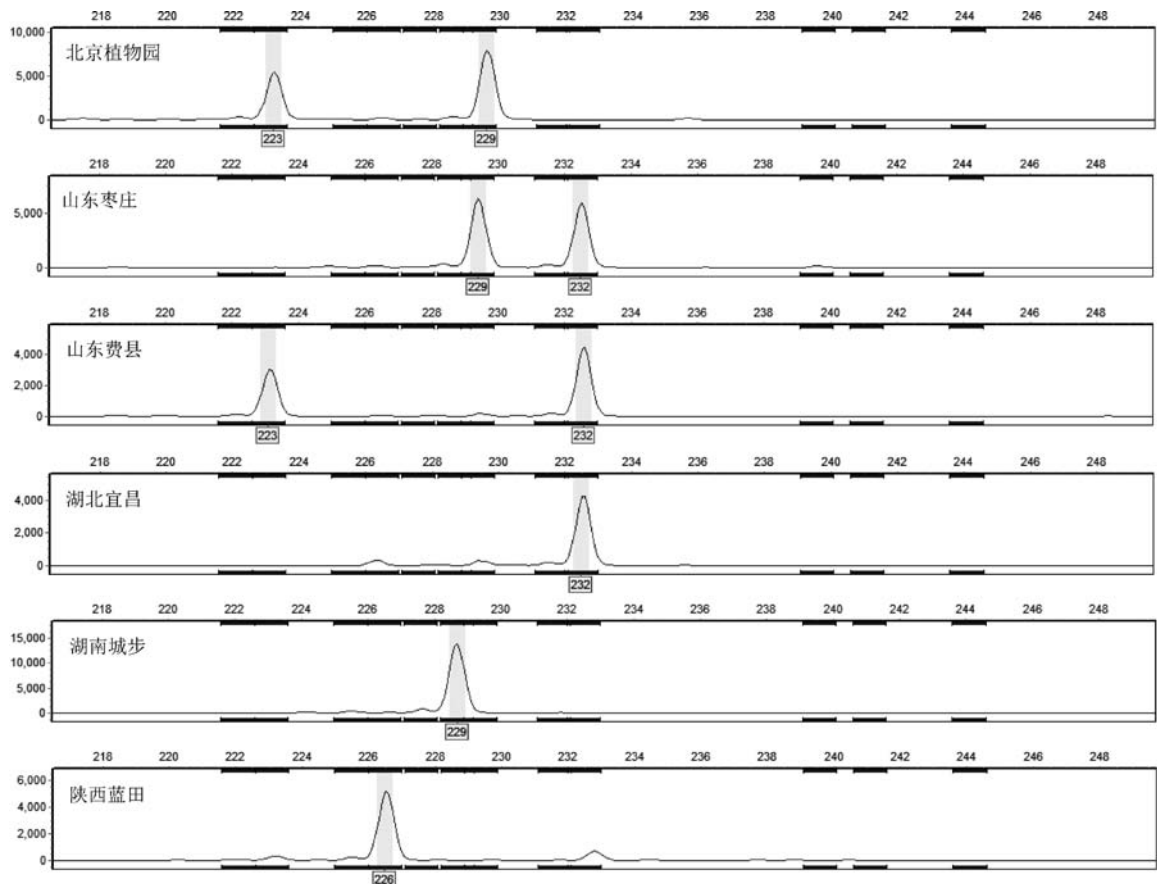


图3 6个样本在GS-30 SSR位点的基因型

Fig. 3 Genotypes of 6 samples at No. GS-30 SSR locus

利用在皂荚植物中开发的44对有效引物对其近缘种进行PCR扩增,聚丙烯酰胺凝胶电泳结果显示,分别有32、31、23、24、7、40、25、18和18对引物可在美国皂荚、日本皂荚、绒毛皂荚、滇皂荚、华南皂荚、小果皂荚、野皂荚、肥皂荚和美国肥皂荚上扩增出清晰目的条带,引物的转化率分别达到72.7%、70.5%、

52.3%、54.5%、15.9%、90.9%、56.8%、40.9%和40.9%。其中小果皂荚转化率最高,高达90.9%,华南皂荚转化率最低,只有15.9%。图4为有效引物G-13、G-26和G-27分别扩增美国皂荚、日本皂荚、绒毛皂荚、滇皂荚、华南皂荚、小果皂荚、野皂荚、肥皂荚和美国肥皂荚的非变性聚丙烯酰胺凝胶电泳图。

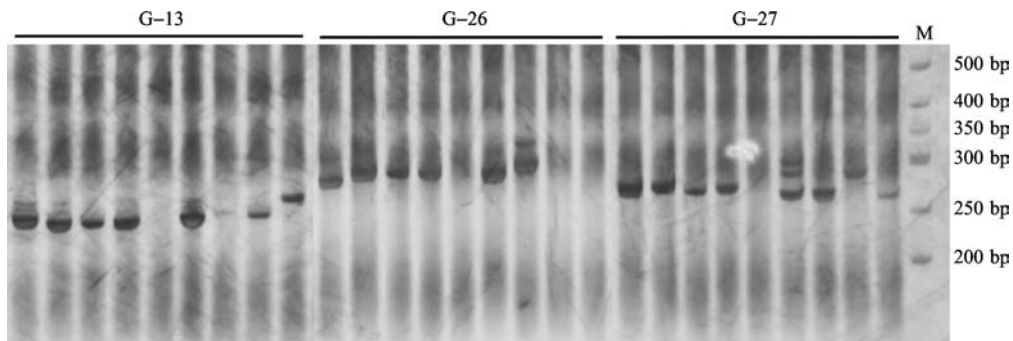


图4 3对引物对皂荚近缘种通用性检测图

Fig. 4 The transferability of 3 pairs of primers for *G. sinensis* relatives

3 讨论

研究表明,EST-SSR重复基序类型及比例依树种不同有所差异,除单核苷酸外,最常见的重复基序类

型是二核苷酸和三核苷酸,如杜仲(*Eucommia ulmoides* Oliv.)^[15]、木麻黄(*Casuarina equisetifolia* L.)^[16]、橡胶树(*Hevea brasiliensis* (Willd. ex A. Juss.) Mü ll. Arg.)^[10]、腊梅(*Chimonanthus praecox* (L.)

Link)^[17]、柠条锦鸡儿(*Caragana korshinskii* kom.)^[18]等以二核苷酸重复基序为最多,而云南松(*Pinus yunnanensis* Franch.)^[19]、杉木(*Cunninghamia lanceolata* (Lamb.) Hook.)^[20]、红松(*Pinus koraiensis* Siebold & Zucc.)^[21]、喜马拉雅红豆杉(*Taxus wallichiana* Zucc.)^[22]等以三核苷酸重复基序为最多,这两种核苷酸重复的多态性普遍高于其他重复基序类型^[14-15]。

本研究发现,除单核苷酸重复基序外,皂荚最多的重复基序类型是二核苷酸,其次是三核苷酸。这种不同树种间重复基序数量的差异可能与基序的基因表达程度、进化水平或者突变频率等因素的差异有关^[19,23]。皂荚转录组的二核苷酸重复基序和三核苷酸重复基序分别以 AG/TC 和 AAG/CTT 分布最多,这一结果与杜仲^[15]、橡胶树^[10]、腊梅^[17]、杉木^[20]和柠条锦鸡儿^[18]的研究结果相似。另外,CG/GC 在二核苷酸重复基序中十分罕见,在南方红豆杉^[22]、杜仲^[15]和甘蓝^[24]中也仅检测到一次,而本研究在皂荚转录组中也没有检测到这个重复基序,这些发现都为引物设计提供了重要依据。

本研究对多态性信息量分析结果显示,所开发的皂荚 EST-SSR 标记具有高度多态性,表明这些标记可以用于皂荚种质资源遗传多样性评价及相关研究。对 17 个基因位点的分析显示,大部分位点位于 3'-UTR 和 5'-UTR 区域,少量位于 CDS 区域,由于 UTR 区域相对 CDS 区域具有更大的遗传变异性,在种内个体间具有更好的多态性,也证明了多态性引物的可靠性。在木豆(*Cajanus cajan* (L.) Huth) SSR 研究中也发现,具有多态性的 SSR 中,分布在 3'-UTR 区域的数量比 5'-UTR 区域的数量多,二者数量均多于 CDS 区域^[25]。因此,在设计引物时,可根据需要针对性选择合适区域来开发。

本研究结果显示,利用皂荚转录组开发的 EST-SSR 在皂荚属内种间转移率高达 90.9%,除华南皂荚外,均超过 50%,在肥皂荚属植物肥皂荚和美国肥皂荚中转移率也高达 40.9%。这说明利用皂荚转录组开发的 EST-SSR 在其近缘种间具有很高的通用性,可为皂荚种质亲缘关系鉴定和皂荚属种间的系统进化提供新途径。很多物种也证实基于转录组开发的 EST-SSR 标记能够在近缘种间通用^[26-27],在缺乏转录组数据时,可以利用 EST-SSR 标记的通用性,开发其近缘种的 EST-SSR。

参考文献

[1] Zhang J P, Tian X H, Yang Y X, et al. *Gleditsia* species: An ethnomedical, phytochemical and pharmacological review [J]. J

- Ethnopharmacol, 2016, 178: 155-171
- [2] 顾万春,孙翠玲,兰彦平. 世界皂荚(属)的研究与开发利用[J]. 林业科学, 2003, 39(4): 127-133
- [3] Lai P, Du J R, Zhang M X, et al. Aqueous extract of *Gleditsia sinensis* Lam. fruits improves serum and liver lipid profiles and attenuates atherosclerosis in rabbits fed a high-fat diet [J]. J Ethnopharmacol, 2011, 137(3): 1061-1066
- [4] 李伟,林富荣,郑勇奇,等. 皂荚南方天然群体种实表型多样性[J]. 植物生态学报, 2013, 37(1): 61-69
- [5] 张金波,苗昊翠,王威,等. 应用 SSR 标记分析鹰嘴豆种质资源遗传多样性[J]. 植物遗传资源学报, 2015, 16(5): 1027-1034
- [6] 兰孟焦,吴胜胜,王瑞珍,等. 甘薯与牵牛 EST 资源的 SSR 信息分析[J]. 植物遗传资源学报, 2013, 14(5): 954-959
- [7] Pocai P, Varga I, Laos M, et al. Advances in plant gene-targeted and functional markers: a review [J]. Plant Methods, 2013, 9(1): 6
- [8] Zalapa J E, Cuevas H, Zhu H, et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences [J]. Am J Bot, 2012, 99(2): 193-208
- [9] Russell J R, Hedley P E, Cardle L, et al. tropiTree: an NGS-based EST-SSR resource for 24 tropical tree species [J]. PLoS One, 2014, 9(7): e102502
- [10] Li D, Deng Z, Qin B, et al. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.) [J]. BMC Genomics, 2012, 13: 192
- [11] 林元震,郭海,刘纯鑫,等. EST-SSR 标记在木本植物中的开发和应用[J]. 植物生理学通讯, 2009, 45(12): 1221-1225
- [12] Zhu L, Zhang Y, Guo W, et al. *Gleditsia sinensis* Lam.: transcriptome sequencing, construction, and application of its protein-protein, interaction network [J]. Biomed Res Int, 2014: 404578
- [13] Zhao Q Y, Wang Y, Kong Y M, et al. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study [J]. BMC Bioinformatics, 2011, 12(14): 2
- [14] Botstein D, White R L, Skolnick M, et al. Construction of a genetic linkage maps in man using restriction fragment length polymorphisms [J]. The American Journal of Human Genetics, 1980, 32(3): 314-331
- [15] 黄海燕,杜红岩,乌云塔娜,等. 基于杜仲转录组序列的 SSR 分子标记的开发[J]. 林业科学, 2013, 49(5): 176-181
- [16] 刘海. 基于高通量测序的木麻黄转录组分析[D]. 福州: 福建农林大学, 2014
- [17] 李响,杨楠,赵凯歌,等. 蜡梅转录组 EST-SSR 标记开发与引物筛选[J]. 北京林业大学学报, 2013(1): 25-32
- [18] Long Y, Wang Y, Wu S, et al. De novo assembly of transcriptome sequencing in *Caragana korshinskii* Kom. and characterization of EST-SSR markers [J]. PLoS One, 2015, 10(1): e0115805
- [19] 蔡年辉,许玉兰,徐杨,等. 云南松转录组 SSR 的分布及其序列特征[J]. 云南大学学报: 自然科学版, 2015, 37(5): 770-778
- [20] 文亚峰,韩文军,周宏,等. 杉木转录组 SSR 挖掘及 EST-SSR 标记规模化开发[J]. 林业科学, 2015(110): 40-49
- [21] 张振,张含国,莫迟,等. 红松转录组 SSR 分析及 EST-SSR 标记开发[J]. 林业科学, 2015, 51(8): 114-120
- [22] 李炎林,杨星星,张家银,等. 南方红豆杉转录组 SSR 挖掘及分子标记的研究[J]. 园艺学报, 2014, 41(4): 735-745
- [23] 李淑娴,张新叶,王英亚,等. 桉树 EST 序列中微卫星含量及相关特征[J]. 植物学报, 2010, 45(3): 363-371
- [24] 曲存民,卜海东,刘晓兰,等. 黄籽甘蓝型油菜核心种质的 SSR 标记分析[J]. 西南大学学报: 自然科学版, 2015, 37(1): 1-6
- [25] Sharma A, Chauhan R S. Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha curcas* [J]. Comp Funct Genomics, 2011: 286089
- [26] 杨彦伶,张亚东,张新叶. 杨树 SSR 标记在柳树中的通用性分析[J]. 分子植物育种, 2008, 6(6): 1134-1138
- [27] 许玉兰,康向阳,张瑞丽,等. 木本植物中 EST-SSR 的通用性及其多态性研究[J]. 西南林业大学学报, 2012, 32(4): 94-99