

杜仲基因组微卫星特征及 SSR 标记开发

吴敏^{1,2}, 杜红岩^{1,2,3*}, 乌云塔娜^{1,2,3}, 刘攀峰^{1,2,3}, 荆腾^{1,2}

(1. 中国林业科学研究院经济林研究开发中心, 河南 郑州 450003; 2. 国家林业局杜仲工程技术研究中心, 河南 郑州 450003;
3. 国家林业局泡桐研究开发中心, 河南 郑州 450003)

摘要:为全面解读杜仲的遗传背景,本研究在基因组测序(数据未公布)的基础上,通过 MISA 软件搜索杜仲基因组(26 947/854 758 160 bp)中的完整型(1~7 核苷酸重复)及复合型 SSR 序列,共查找出 25 694 个 Scaffolds 含有 488 592 个 SSR 位点,占总 Scaffolds 的 95.3%。从 SSR 位点分布密度上讲,平均每 1 749 bp 出现 1 个微卫星,其中单核苷酸重复单元的 SSR 含量最多,约占总数的 54.34%,其次为二核苷酸(20.47%),而复合模式、三核苷酸、四核苷酸、五核苷酸、六核苷酸和七核苷酸重复分别占 20.29%、23.89%、0.77%、0.13%、0.10% 和 0.01%,并发现 SSR 中均以含 A、T 的重复类型占主导地位。根据 SSR 位点设计并合成引物 290 对,162 对 SSR 引物扩增出清晰的目的条带,其中 16 对引物多态性高、稳定性好,在 8 份杜仲资源中共检测到 84 个等位基因,平均每个 SSR 位点检测到 5.25 个等位基因。本研究对于进一步利用 SSR 分子标记分析杜仲遗传多样性、遗传图谱构建、杜仲雌雄株早期鉴定等具有较高的应用价值。

关键词: 杜仲;基因组;SSR;标记开发;多态性

中图分类号:S718.46

文献标识码:A

Characterization of Genomic Microsatellites and Development of SSR Markers of *Eucommia ulmoides*

WU Min^{1,2}, DU Hong-yan^{1,2,3}, WUYUN Tana^{1,2,3}, LIU Pan-feng^{1,2,3}, JING Teng^{1,2}

(1. Non-timber Forestry R&D Center, Chinese Academy of Forestry, Zhengzhou 450003, He'nan, China;
2. Eucommia Engineering Research Center of State Forestry Administration, Zhengzhou 450003, He'nan, China;
3. Paulownia R&D Center of China of State Forestry Administration, Zhengzhou 450003, He'nan, China)

Abstract: Based on genome sequencing of *Eucommia ulmoides*, the perfect SSR (1~7 nucleotide repeats) and compound SSR sequences were searched from *Eucommia* genome (26,947/854,758,160 bp) by using MISA. 488,592 SSR loci were found from 25,694 Scaffolds, accounting for 95.3% of the total Scaffolds. The results indicated that every 1 749 bp contained one SSR microsatellite on the average. The SSRs which had mononucleotide repeats were the most of all, accounting for 54.34%, and the dinucleotide repeats accounted for 20.47%. The percentage of compound SSR, trinucleotide repeats, tetranucleotide repeats, pentanucleotide repeats, hexanucleotide repeats, and heptanucleotides repeats were 20.29%, 23.89%, 0.77%, 0.13%, 0.10% and 0.01%, respectively. Meanwhile, the A/T repeats were predominant type among these SSRs. A total of 290 pairs of primer was designed and synthesized. The primers were verified by using 8 *Eucommia* clones, among which 162 pairs of primers were able to amplify products and 16 pairs of them were polymorphic and stable. 84 alleles were detected in these *Eucommia* clones, (5.25 alleles per SSR locus in average). The results are applicable for analyzing genetic diversity con-

收稿日期:2014-04-30

基金项目:国家林业公益性行业科研专项:杜仲育种群体建立与综合利用技术研究(201004029)

作者简介:吴敏,硕士研究生,研究方向:经济林栽培和育种. E-mail:wminspring@163.com

* 通讯作者:研究员,博士生导师,主要从事杜仲育种、栽培与综合利用的研究. dhy515@126.com

structing genetic map, and identifying early sex status by using SSR molecular markers on *Eucommia ulmoides*.

Key words: *Eucommia ulmoides*; genome; SSR; marker development; polymorphism

SSR (simple sequence repeats) 序列是指以 1~6 个, 甚至更多核苷酸为基本单位组成的串联重复序列片段, 又称微卫星 DNA (microsatellite DNA), 其最常见的重复片段为 1~6 个核苷酸的重复序列, 也有少数 7 个核苷酸甚至更多核苷酸的重复序列, 其长度大多在 200 bp 以内, 普遍存在于真核生物和原核生物基因组中^[1,2], 分布于编码区和非编码区^[3]。研究表明, SSR 在植物种间、种内居群间乃至个体间均存在高度多态性, 且不同物种的 SSR 序列在组成、长度、突变率及在染色体上的分布情况表现出高度的多态性, 反映了高度的等位基因多样性^[4]。不同的遗传材料在某一 SSR 位点重复次数上的差异可导致该 SSR 位点长度的多态性, 且在 SSR 位点两侧有一段相当保守的单拷贝序列, 根据此保守序列设计 PCR 引物能够精确检测出 SSR 位点的长度多态性。SSR 标记主要包括基因组 SSR (genomic SSR) 和表达序列标签 SSR (EST-SSR)。SSR 分子标记技术具有共显性、数量多、范围广、多态性高等优点, 并在近几十年的不断发展中得到了广泛的应用, 是基因定位、构建遗传连锁图谱和指纹数据库、品种鉴定、群体遗传学研究及分子标记辅助育种等工作的理想工具^[5-7]。

杜仲 (*Eucommia ulmoides*, $2n = 34$), 属杜仲科 (Eucommiaceae) 杜仲属 (*Eucommia*), 为单科单属单种^[8], 雌雄异株、多年生落叶乔木, 是我国特有的名贵经济树种。杜仲既是十分重要的优质天然橡胶资源, 又是名贵药材和工业原料, 同时也能够改善生态环境、增加碳汇储量^[9]。我国拥有世界 99% 的杜仲

资源, 具有独特的资源优势。目前杜仲的遗传背景并不完全清晰, 这给种质资源保存、杜仲良种选育、杂交育种等带来很大困难。AFLP、RAPD、ISSR 等分子标记的应用在一定程度上促进了杜仲遗传多样性、品种改良和种质创新等研究^[10-14], 在利用 SSR 分子标记研究方面, Deng 等利用 FIASCO 法开发了 19 对具有多态性的 SSR 引物, 及中国林业科学研究院经济林研发中心开发了 20 对 EST-SSR 引物, 经验证其中仅 10 对 SSR 引物得到成功扩增, 难以满足研究需要^[15]。为此, 本研究在采用 HiSeq、MiSeq 及第三代测序技术 PacBio 对杜仲基因组深度测序的基础上, 对数据进行系统分析, 开发了杜仲 genomic SSRs 引物, 为杜仲种质资源鉴定、遗传多样性分析及分子标记辅助育种等研究提供更为丰富详尽的背景资料。

1 材料与方法

1.1 植物材料

1.1.1 基因组测序材料 叶片样本采自湖北神农架一株树龄 200 多年的杜仲雄株 (编号 12001X), 疑似野生杜仲^[12], 液氮速冻后送武汉未来组生物科技有限公司 (Nextomics) 进行基因组测序。

1.1.2 引物筛选试材料 本研究引物筛选所用的 8 份杜仲样品来自中国林业科学研究院经济林研究中心杜仲基因库 (表 1), 于 2013 年 6 月初采集健壮、无病虫害的嫩叶, 液氮保存带回实验室, -80°C 保存以备提取 DNA。

表 1 杜仲试验材料

序号	样品编号	来源	经度/E	纬度/N	海拔/m	气候类型
1	12001X	湖北神农架	110°10'46"	31°24'17"	1 163	亚热带季风气候
2	Hong mu	河南原阳	113°46'28"	34°55'36"	106	暖温带大陆性季风气候
3	10600C	河南洛阳	112°25'14"	34°39'56"	159	暖温带大陆性季风气候
4	10590X	河南三门峡	110°39'37"	34°16'18"	1 072	暖温带大陆性季风气候
5	10586C	河南洛阳	112°25'14"	34°39'56"	159	暖温带大陆性季风气候
6	10587X	湖南江垭	110°77'23"	29°51'17"	159	亚热带季风性湿润气候
7	10601X	河南洛阳	112°25'14"	34°39'56"	159	暖温带大陆性季风气候
8	10585X	河南南阳	111°58'23"	32°48'16"	218	季风大陆半湿润气候

1.2 研究方法

1.2.1 基因组 DNA 提取及浓度测定 采用天根生物公司 (北京) 新型植物基因组提取试剂盒 (DP320

-02) 提取杜仲基因组 DNA, 1.5% 琼脂糖凝胶电泳检测 DNA 的质量, DNA 样品上样量 3 μL , 电泳结束后用 Bio-Red Gel Doc XR + 凝胶成像系统进行观察、

拍照;紫外分光光度计检测 DNA 含量及纯度, A_{260}/A_{280} 值在 1.8 ~ 2.0 范围内的样品,统一浓度至 $40 \text{ ng} \cdot \mu\text{L}^{-1}$ 置于 -20°C 保存备用。

1.2.2 杜仲基因组 SSR 位点的查找 利用 MISA 软件对杜仲基因组序列进行 SSR 位点查找, MISA 软件能识别出序列中的单核苷酸 SSR 和复合 SSR 及所在的位点。查找标准:单、二、三、四、五、六、七核苷酸重复单元重复次数分别大于 10、6、5、5、5、5、5 次,并且 SSR 侧翼序列长度大于等于 50 bp。

1.2.3 引物设计和评价 针对整个基因组中出现次数相对较少的 SSR 位点,利用 Primer3 (<http://primer3.ut.ee/>) 进行引物设计,并利用 Oligo7 对引物进行评价。引物设计主要参数为:1) 引物长度:18 ~ 25 bp;2) GC 含量:40 ~ 60%;3) 预期扩增产物长度:100 ~ 300 bp;4) 3' 端 ΔG 绝对值不超过 9,而 5' 端和中间 ΔG 值相对较高;5) 引物二聚体和发夹结构的 ΔG 小于 $4.5 \text{ kcal} \cdot \text{mol}^{-1}$ 。

1.2.4 PCR 扩增 引物由生工生物工程(上海)股份有限公司合成,引物扩增所用的 $2 \times \text{Taq PCR Master Mix}$ 购自于天根生化科技(北京)有限公司。PCR 反应体系为:Master Mix $8 \mu\text{L}$, primer1、primer2、模版 DNA 各 $0.8 \mu\text{L}$,加入 ddH_2O 将体系补至 $20 \mu\text{L}$ [15]。利用 Bio-Rad 基因扩增仪(T100)进行 PCR 扩增,循环体系为:1) 94°C 预变性 3 min;2) 94°C 变性 30 s;3) $T_m^\circ\text{C}$ 退火 30 s;4) 72°C 延伸 1 min;5) 返回第二步,重复 29 个循环;6) 4°C 保存,并在 24 小时内电泳检测。

1.2.5 聚丙烯酰胺电泳检测 利用 Bio-Rad 迷你小型垂直电泳槽(165-8001)进行 6% 非变性聚丙烯酰胺凝胶电泳检测 PCR 扩增产物,180 V 电泳 40 ~ 50 min,电泳结束后用 Gelred 荧光染料(美国)染色 30 ~ 60 min, Bio-Red Gel Doc XR + 凝胶成像系统进行观察拍照。

1.2.6 数据统计与分析 统计稳定且易于区分的目的条带(忽略杂带和弱带)。用 POPGENE 32 软件计算具有多态性的 SSR 位点的等位基因数(na),有效等位基因数(ne), Nei' S 基因多样性(h)和 Shannon 信息指数(I)。

2 结果与分析

2.1 杜仲基因组 SSR 位点的查找

本研究对杜仲基因组中 1 ~ 7 核苷酸重复完整型和复合型 SSR 进行查找分析。当前组装的杜仲基

因组(26 947/854 758 160 bp),平均每条 scaffold 长度为 31 720 bp,共查找出 25 694 个 Scaffold 含有 SSR 位点(488 592 个),占总 Scaffold 的 95.3%,平均每一条 Scaffold 含有 18 个 SSR 位点,只含有 1 个 SSR 位点的 Scaffold 有 10 706 条(表 2)。488 592 个 SSR 位点中复合模式的 SSR 为 99 164 个,其中部分微卫星信息见表 3。

表 2 杜仲基因组 SSR 分布统计

	Number/Size
Total examined sequences	26 947/854 758 160 bp
Total identified SSR	488 592
Sequences contain SSR(= 1)	10 706
Sequences contain SSR(> 1)	14 988

表 3 杜仲基因组 SSR 数据库的部分结果

scaffold ID	SSR 类型	重复单元	SSR 长度
scaffold3_size706409	单核苷酸	(A)10	10
scaffold1487_size130798	单核苷酸	(T)58	58
scaffold1968_size265695	二核苷酸	(TC)6	12
scaffold2582_size105070	二核苷酸	(TA)57	114
scaffold3_size706409	三核苷酸	(GAA)5	15
scaffold332_size899754	三核苷酸	(AAT)79	237
scaffold3_size706409	四核苷酸	(AAAT)5	20
scaffold662_size342166	四核苷酸	(AATG)64	192
scaffold8_size598530	五核苷酸	(ATTTT)5	25
scaffold1854_size107307	五核苷酸	(CACAC)9	45
scaffold2_size713950	六核苷酸	(GAAGGG)5	30
scaffold674_size199824	六核苷酸	(AAATTA)13	78
scaffold662_size342116	七核苷酸	(AAAAAAT)5	35
scaffold3_size706409	七核苷酸	(TTTAAAT)6	42
scaffold6_size1426843	复合模式	(T)10(A)10	20
scaffold2103_size95718	复合模式	(ATA)11atc(ATA)16	84

2.2 杜仲基因组微卫星组成分析

在杜仲基因组数据库中共查找出 488 592 个 SSR 位点,从微卫星分布密度上讲,平均 1 749 bp 出现 1 个微卫星,SSR 的分布频率(SSR 的个数与总 scaffold 的比值)为 18.12。其中单碱基重复单元的 SSR 含量最多,约占总数的 54.34%,其次为二核苷酸(20.47%)、复合模式(20.29%)、三核苷酸(3.89%)、四核苷酸(0.77%)、五核苷酸(0.13%)、六核苷酸(0.10%)、七核苷酸(0.01%)(表 4)。每种重复单元包含不同种类的重复碱基,单核苷酸重复 SSR 由 4 种不同的重复碱基组成,其中以 A/T 为主,占单核苷酸重复的 98.74%,且 A(131 775),T(130 394)接近 1:1;二核苷酸重复 SSR 组成类型(12 种)中 AT/TA 最多,共 45 970 个,占 45.97%,其

次是 AG/TC (20 075) 占 20.08%。;三核苷酸重复 SSR 组成类型 (60 种) 中 AAT/TTA (4 333) 占 22.83%, ATT/TAA (2 104), 占 11.08%, ATA/TAT (2 120) 占 11.17%;四核苷酸重复 SSR 组成类型 (128 种) 中 AAAT/TAAA (1 507), ATTT/TAAA (517) 分别占 40.30%, 13.83%;五核苷酸重复 SSR 组成类型 (172 种) 中 AAAAT/TTTTA, ATTTT/TA-AAA 占主导地位, 共 402 个, 占 65.26%;六核苷酸重复 SSR 组成类型 (299 种) 中 AAAAAAT/TTTTTA, GGTTT/CCAAA, 共 34 个, 占 11.20%;七核苷酸重复 SSR 组成类型 (46 种) 中 AAAAAAT/TTTTTTA (8 个), 占 12.31%。

表 4 杜仲基因组中 SSR 重复单元的分布特征

重复类型	SSR 数量	所在比例/%	发生频率	主要重复单元
单核苷酸	265 523	54.34	9.85	A/T,G/C
二核苷酸	99 991	20.47	3.71	AT/TA,AG/TC,CT/GA,AC/TG
三核苷酸	18 985	3.89	0.70	AAT/TTA,ATT/TAA,ATA/TAT, TTC/AAG,GAA/CTT TTG/AAC, TCT/AGA
四核苷酸	3 739	0.77	0.14	AAAT/TAAA,ATTT/TAAA, AATA/TTAT,TATT/ATAA
五核苷酸	616	0.13	0.02	AAAAT/TTTTA,ATTTT/TAAAA, TTGGG/AACCC,TGTGG/ACACC
六核苷酸	509	0.10	0.02	AAAAAT/TTTTTA,GGTTT/ CCAAA
七核苷酸	65	0.01	0.00	AAAAAAT/TTTTTTA,AAAAATA/ TTTTTA
复合模式	99 164	20.29	3.68	
合计	488 592	100	18.12	

2.3 杜仲基因组微卫星长度分布及变异

本研究中杜仲 genomic SSR 序列长度变化范围是 10~237 bp, 平均长度为 15.17 bp, 不同类型 SSR 长度变化范围差异很大。从图 1 中可以看出, 长度为 10 bp 的微卫星所占比例最大 (23.60%), 其次是长度为 12 bp 的微卫星 (14.95%), 29 bp 的微卫星仅占 0.11%, 大于等于 31 的微卫星占 4.38%。

微卫星长度的变异情况揭示了微卫星获得或失去重复单元速率的快慢程度。由图 2 可知, 不同重复 SSR 类型中的重复长度不同, 且长度的变异较大。其中单核苷酸 SSR 重复单元长度变化范围最大 (10 bp~237 bp), 二核苷酸重复 SSR 次之 (12 bp~196 bp), 变异程度最低的为五核苷酸重复微卫星, 仅有 5 种不同微卫星变化长度。每种重复 SSR 类型都随着重复长度的递增, 微卫星丰度呈递减趋势, 即 SSR 长度越长, 微卫星出现频率越低。

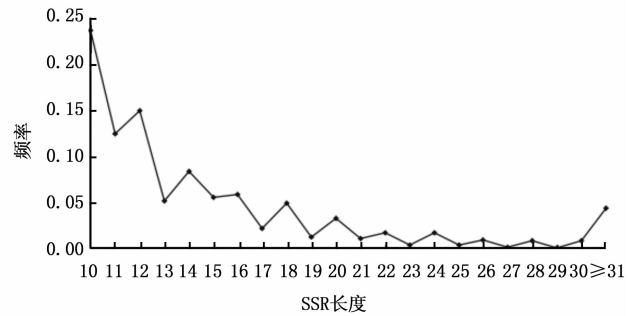


图 1 杜仲基因组不同长度微卫星出现频率

2.4 杜仲基因组 SSR 引物的有效性检测

选择 8 份差异较大的杜仲样品进行引物筛选, 合成的 290 对 genomic SSRs 引物中共有 162 对能扩增出明显的主条带, 其中 5 对引物扩增出的条带大于目的片段。从 162 对引物中进一步筛选出了主条带清晰、多态性高、稳定性好的引物 16 对 (见表 5), 部分引物 (GEU09、GEU037、GEU191、GEU223) 的扩增效果见图 3。

2.5 引物多态性分析

筛选出的 16 对引物在 8 份杜仲样品中检测到 84 个等位基因, SSR 位点检测到的等位基因数在 3~7 之间, 平均每个 SSR 位点检测到 5.25 个等位基因。目的片段差异最大的是 GEU158, PCR 产物片段长度 87~135 bp, 目的片段差异最小是 GEU271, PCR 产物片段长度 164~182 bp。(表 5)。

利用 16 个 SSR 位点对 8 份杜仲样品进行 PCR-SSR 检测表明: 各位点有效等位基因数分布在 2.169 5~5.818 2 之间, 平均每个位点的有效等位基因数为 3.488 0, Nei's 指数 (h) 的分布范围在 0.570 3~0.773 4 之间, 平均每个位点的 h 值为 0.694 3。 $I = 1 - \sum P_i^2$, P_i 为产物条带的表型, 平均每个位点的 I 值为 1.310 0, I 值最高为 1.840 7 (GEU045), I 值最低为 0.864 7 (GEU056) (表 6)。

3 讨论

杜仲基因组微卫星种类丰富, 数量多, 1~7 核苷酸的重复及复合型 SSR 均有出现, 其中单核苷酸和二核苷酸重复占主导地位, 分别占 SSR 总数的 54.34%, 20.47%, 七核苷酸重复的微卫星最少, 仅占 0.10%。有研究报道基因组越小, 单碱基重复序列越多, 随着物种基因组的增大, 其他类型微卫星逐渐增多^[16-17]。在本研究中杜仲单核苷酸 SSR 达到 54.34%, 这可能与目前杜仲基因组数据不完全有很

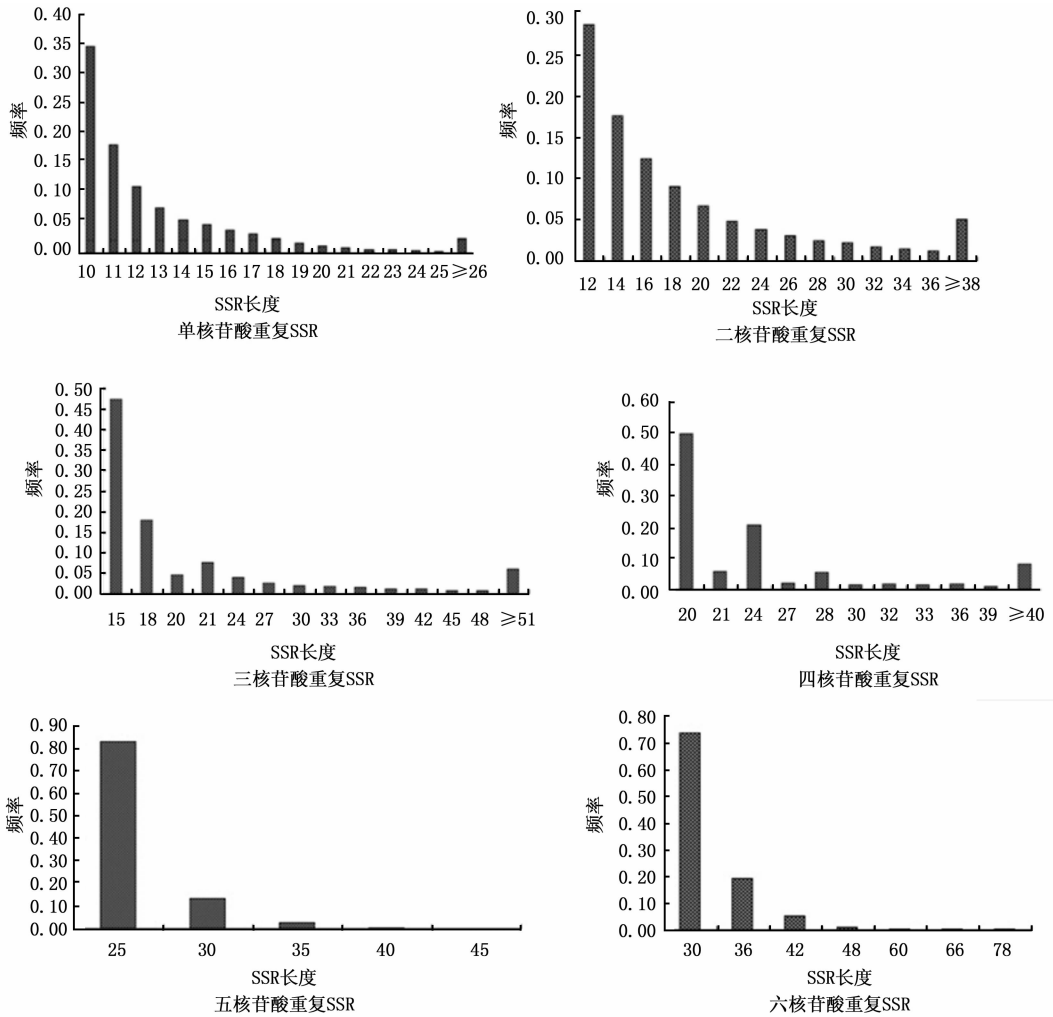


图2 不同长度重复单元微卫星变异情况

表5 G-SSR 引物信息

引物	重复单元	SSR 类型	SSR 长度 /bp	正向引物(5'-3')	反向引物(5'-3')	T _m /°C	产物范围 Size/bp
GEU09	(TAGA)10	p4	40	GCCTGAAGGCATGTTGAAAT	GGCACCGGTACATCTTCTCT	60	155 ~ 188
GEU026	(ATTCCTA)6	p7	42	CAGCCACTCAATACCGAATC	CACCAGAGGCCACAAGCATAA	60	170 ~ 197
GEU037	(GTTTTG)7	p6	42	TGGTCTTGGACATGCTCTG	ACGCAATTCGATTACGTGGT	60	264 ~ 298
GEU045	(GATA)11	p4	44	CATCGTCATCTTCTCCTTGC	CACGACACACTGAAGATCCA	58	272 ~ 308
GEU052	(TTTCAC)6	p6	36	TGCTATATGTTCTTCAGGCTCT	TGCGTATTGGAAGATTCTCG	58	263 ~ 298
GEU053	(TTTAAA)6	p7	42	GGTATGAGCATGAACGATGG	AACTTCTTCGGTCAACCTTGT	57	215 ~ 249
GEU056	(GCCA)7	p4	28	TGAAGGAGGTTGGTGACAGA	AAGATGTCCTTCATCCTCTATTGC	60	250 ~ 281
GEU150	(ATATAC)8	p6	48	CGCACCGAAGCATAGAACTT	CATGGCAATGCAAGCTAGAC	60	271 ~ 312
GEU155	(TAC)15	p3	45	CTGCCGATTCTCAGCCTATG	TTCGTGCACTTFCGAGTATT	60	236 ~ 266
GEU158	(GCCATC)8	p6	48	TCTCCAAGCTACACGCAATG	CTCCACAGTCCATTCAACGA	60	87 ~ 135
GEU191	(TCT)12	p3	36	CCAATCTACGAAGTTGCCAAG	TTGACCGAAGCACTAGGTGA	60	143 ~ 180
GEU219	(AAC)10	p3	30	ATCGATCTGAAGCGTGAT	ATGGTCTTGTATCGTCTATT	60	227 ~ 266
GEU223	(ACA)10	p3	30	TGC AACACTTGATCTCCAC	AACGGATTCTGTAATCGACCT	60	160 ~ 186
GEU239	(CTC)8	p3	24	ACTTGATTGAGCCGGAACG	CTTGAACCAGACAACCACCA	60	249 ~ 269
GEU271	(AGA)11	p3	33	TGCTCTCTGCTAACTGCAA	GGTCCGAACAGAGAATCGTC	60	164 ~ 182
GEU289	(AAAAT)6	p5	30	TCAAGTCAAGATCACGCTTAGC	CAATCTGTCTCAATTAACCTCA	60	249 ~ 274

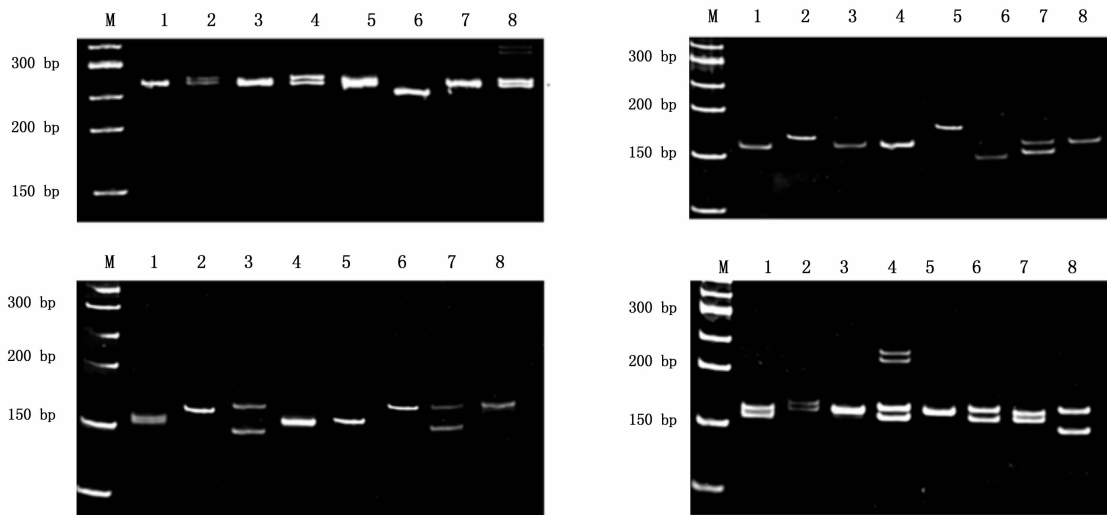


图3 引物 GEU-09、GEU037、GEU191、GEU223 在 8 个优良无性系间的多态性

表 6 杜仲 SSR 位点基因的多样性分析

位点名称 Loci	等位基因数 n_a	有效等位基因数 n_e	Nei's 基因多样性指数 h	Shannon 信息指数 I
GEU09	6.000 0	3.555 6	0.718 8	1.488 0
GEU026	5.000 0	4.000 0	0.750 0	1.474 1
GEU037	5.000 0	2.327 3	0.570 3	1.160 2
GEU045	7.000 0	5.818 2	0.828 1	1.840 7
GEU052	3.000 0	2.327 3	0.570 3	0.947 1
GEU053	4.000 0	3.282 1	0.695 3	1.282 0
GEU056	3.000 0	2.169 5	0.539 1	0.864 7
GEU150	5.000 0	3.657 1	0.726 6	1.424 4
GEU155	5.000 0	4.266 7	0.765 6	1.511 1
GEU158	7.000 0	4.413 8	0.773 4	1.688 7
GEU191	7.000 0	3.878 8	0.742 2	1.628 6
GEU219	4.000 0	2.844 4	0.648 4	1.162 7
GEU223	7.000 0	3.878 8	0.742 2	1.628 6
GEU239	5.000 0	3.200 0	0.687 5	1.331 7
GEU271	7.000 0	3.282 1	0.695 3	1.526 9
GEU289	4.000 0	2.909 1	0.656 2	1.213 0
Mean	5.250 0	3.488 0	0.694 3	1.310 0

大关系,也可能与杜仲是孑遗树种,在长期的进化过程中出现的变异有关。微卫星的形成可能与 DNA 复制过程中出现的滑脱、碱基交替以及重组不平衡有一定的关系,微卫星重复序列(如 GA、CA、GT 等重复)可以通过影响 DNA 的结构而影响 DNA 重组^[2,18],因此微卫星中的重复单元碱基组成在很大程度上会影响生物的生命活动。杜仲 SSR 在非编码区的含量要比在编码区的含量多^[19],genomic SSRs 出现的频率为 18.12,相比杜仲 EST-SSRs 出现的频率(2.69%)较高^[15]

SSR 序列显示出显著的碱基偏好性,A/T 的频率显著高于 G/C。其中 A/T 占单核苷酸 SSR 总数的 98.74%,二到七核苷酸重复的 SSR 中也均以含 A、T 的重复类型占主导地位。从总体来看,各种重

复类型的 SSR 位点都存在着长度越长出现的频次越少的趋势。本研究中若不统计单核苷酸重复的 SSR,杜仲基因组 SSR 也是以二、三核苷酸重复单元为主,这和大多数植物一致。但优势类型在不同物种间的分布有较大差异,在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、高粱(*Sorghum bicolor*)、油茶(*Camellia oleifera*)、玉米(*Zea mays*)、枣(*Ziziphus jujuba*)、二穗短柄草(*Brachypodium distachyon*)中即有相似现象,其中油茶的优势重复类型为二碱基重复,水稻、高粱的优势重复类型为三碱基重复,而玉米、枣、二穗短柄草基因组的优势重复为六碱基重复^[20-24]。杜仲转录组中二核苷酸重复占 69.90%,其次是三核苷酸(22.82%)^[15],在杜仲基因组数据中,若不统计单核苷酸和复合模式的 SSR,杜仲基因组

其结果是二核苷酸重复(80.71%)和三核苷酸重复(15.32%)占主导地位,这与杜仲转录组中SSR的优势重复相似。

290对引物中有5对引物扩增片段与预期产物片段大小不符,大于目的片段,这可能与扩增片段中插入内含子有关,属于无效扩增^[25]。16对扩增条带清晰、多态性高、稳定性好的引物的扩增片段大小在87~308 bp,其中具有多态性的SSR主要是三核苷酸重复的SSR,另外也有四核苷酸至七核苷酸重复的SSR,这在一定程度上验证了低级单元SSR的多态性普遍比高级单元的高这一推断^[26]。杜仲genomic SSR平均有效等位基因数达到3.488 0,相比杜仲EST-SSR(1.751 5)的多态性高^[15,27],这表明来源于基因组的SSR比来源于EST的SSR具有更丰富的多态性^[28]。

本研究对杜仲SSR标记的开发以及物种进化、遗传多样性等方面的研究将提供重要信息,并对于进一步利用SSR分子标记分析杜仲遗传多样性、遗传图谱构建、杜仲雌雄株早期鉴定等具有一定的实际应用价值。

参考文献:

[1] Mrazek J, Guo X, Shah A. Simple sequence repeats in prokaryotic genomes [J]. PNAS, 2007, 104(20): 8472-8477.

[2] Tóth G, Góspári Z, Jurka J. Microsatellites in different eukaryotic genomes[J]. survey and analysis. Genome Research, 2000, 10(7): 967-981.

[3] Li C Y, Korol A B, Fahima T, et al. Microsatellites with in genes: structure, function, and evolution[J]. Mol Biol Evol, 2004, 21(6): 991-1007.

[4] 敖日格乐, 贾晓, 葛台明, 等. SSR分子标记的开发策略概述[J]. 湖北民族学院学报, 2009, 27(4): 462-467.

[5] Hazan J, Dubay C, Pankowiak M C, et al. A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers [J]. Genomics, 1992, 12(2): 183-189.

[6] Copeland N G, Jenkins N A, Gilbert D J, et al. A genetic linkage map of the mouse. : Current applications and future prospects[J]. Science, 1993(262): 57-66.

[7] 刘列钊, 林 呐. 油菜简单重复序列SSR研究进展[J]. 生命科学, 2004, 16(3): 173-176.

[8] 张芝玉, 王伏雄, 钱南芬. 等. 杜仲花粉形态的研究[J]. 植物分类学报, 1988, 26(5): 367-370.

[9] 杜红岩, 胡文臻, 俞 锐. 杜仲产业绿皮书: 中国杜仲橡胶资源与产业发展报告[M], 社会科学文献出版社(2013): 001-009.

[10] 张 檀, 张康健, 杨吉安, 等. 杜仲脂酶同工酶地理递变及生长预测的研究[J]. 西北林学院学报, 1993, 8(4): 20-23.

[11] 王瑗琦, 黄璐琦, 邵爱娟, 等. 子遗植物杜仲的遗传多样性 RAPD 分析和保护策略研究[J]. 中国中药杂志, 2006, 31(19): 1583-1586.

[12] 邓建云, 李建强, 黄宏文. 一株具有特异 AFLP 指纹图谱的杜仲古树[J]. 武汉植物学研究, 2006, 24(6): 509-513

[13] Kalia R, Rai M, Kalia S, et al. Microsatellite markers. : an overview of there cent progress in plants[J]. Euphytica, 2011, 177(3): 309-334.

[14] Wu Mingqian, Chen Shuanglin, Wang Meixia, et al. An analysis of the genetic diversity and genetic structure of *Eucommia ulmoides* using inter-simple sequence repeat (ISSR) markers [J]. African Journal of Biotechnology, 2011, 10(84): 19505-19513.

[15] 杜红岩, 乌云塔娜, 刘攀峰. 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] Katti M V, Ranjekar P K, Gupta V S. Differential distribution of simple sequence repeats in eukaryotic genome sequences [J]. Mol Bio Evol, 2001, 18(7): 1161-1167.

[17] Karaoglu H, Lee C M Y, Meyer W. Survey of simple sequence repeats in completed fungal genomes [J]. Molecular Biology and Evolution, 2005, 22(3): 639-649.

[18] Biet E, Sun J S, Dutreix M. Conserved sequence preference in DNA binding among recombination proteins. : an effect of ssDNA secondary structure [J]. Nucleic Acids Research, 1999, 27(2): 596-600.

[19] Ellegren H. Microsatellites. : simple sequences with complex evolution [J]. Nature Reviews Genetics, 2004, 5(6): 435-445.

[20] Wei Wenliang, Qi Xiaoqiong, Wang Linhai, et al. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers [J]. BMC Genomics, 2011, 12(1): 451

[21] Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156: 847-854.

[22] 郑燕, 张耿, 吴为人, 禾木科植物微卫星序列的特征分析和比较 [J]. 基因组学与应用生物学, 2011, 30(5): 513-520.

[23] 马秋月, 戴晓港, 陈赢男, 等. 枣基因组的微卫星特征 [J]. 林业科学, 2013, 49(12): 81-87.

[24] 史 洁, 尹俊明, 管宏伟, 等. 油茶基因组微卫星特征分析 [J]. 南京林业大学党报, 2012, 36(2): 47-51.

[25] Saha M C, Mian M A R, Eujayl I, et al. Tall fescue EST-SSR markers with transfer ability across several grass species [J]. Theor Appl Genet, 2004, 109(4): 783-791.

[26] Dreisigacker S, Zhang P, Warburton M L, et al. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different mega environments [J]. Crop Science, 2004, 44(2): 381-388.

[27] 黄海燕, 杜红岩, 乌云塔娜, 朱高浦. 基于SSR分子标记的杜仲遗传多样性体系建立 [J]. 林业科学研究, 2013, 26(6): 795-799.

[28] Cho Y G, Ishii T, Temnykh S, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.) [J]. Theor Appl Genet, 2000, 100(5): 713-722.