

榆树转录组 EST-SSR 引物开发 及遗传多样性分析

左力辉, 张 双, 梁海永, 王进茂, 杨敏生

(河北农业大学林学院/林木种质资源与森林保护重点实验室, 保定 071000)

摘要: 目前榆树已有的分子标记资源匮乏, 无法满足榆树种质资源评价及分子标记辅助育种等相关研究需要。本研究对白榆叶片转录组数据进行 EST-SSR 标记的检测和开发, 检测引物在不同榆树资源中的可用性, 并对不同榆树资源的多样性进行分析。本研究共检测到 8828 个精确型和 569 个复合型 SSR 位点, SSR 序列长度主要以 10~22 bp 的短序列为主。SSR 重复单元比例最大的为 A/T (3330, 40.18%), 其次为 AG/CT (1211, 14.61%) 和 AAG/CTT (568, 6.85%)。随机挑选 90 对 EST-SSR 引物进行验证, 有效扩增率为 51.11% (46 对), 其中有 63.04% 的引物 (29 对) 为高多态性引物, 多态性信息含量 *PIC* 在 0.054~0.683 间变化, 极大地丰富了榆树的 SSR 引物资源。聚类分析表明, 绝大部分榆树无性系均按其起源聚类, 从侧面也证明了本研究开发的 EST-SSR 引物的有效性。本研究开发的 SSR 引物可将绝大部分榆树资源进行区分, 与传统分类学相吻合, 为榆科植物的分类提供了分子依据。

关键词: 榆科; 转录组; EST-SSR; 遗传多样性

EST-SSR Primer Development of Elm Transcriptome and Genetic Diversity Analysis

ZUO Li-hui, ZHANG Shuang, LIANG Hai-yong, WANG Jin-mao, YANG Min-sheng

(College of Forestry, Hebei Agricultural University/The Key Laboratory of
Germplasm Resources of Forest Trees and Forests Protection, Baoding 071000)

Abstract: At present, the available molecular markers of elm have been seriously deficient, which can not meet the needs of molecular marker assisted breeding and evaluation of germplasm resources. In this study, transcriptome data of *Ulmus pumila* L. were used as materials to develop and verify EST-SSR markers of elms. The availability of primers were checked in different elm resources. Moreover, the genetic diversity of different elm resources were analysed by using the developed primers. In this study, 8828 perfect SSR and 569 compound SSR loci were detected in 36609384 bp sequences of Unigene sequences by using the transcriptome data of *Ulmus pumila* L. The length of perfect SSR sequence were mainly based on short sequence of 10~22 bp. The largest proportion of SSR repeat units were A/T (3330, 40.18%), followed by AG/CT (1211, 14.61%) and AAG/CTT (568, 6.85%). In this study, we selected 90 pairs of EST-SSR primers randomly for validation analysis, the effective amplification rate was 51.11% (46 pairs) of which 63.04% primers (29) were high polymorphism primers. The polymorphism information content *PIC* were ranged from 0.054~0.683, which greatly enriched the resources of the elm SSR primers. This study confirmed the availability of SSR primers in different elm resources. Cluster analysis showed that the majority of elm clones were separated from their origin, which proved the validity of the EST-SSR primers that developed in this study from the other side. The SSR primers developed in this study can distinguish most of the elm resources, and

收稿日期: 2017-06-05 修回日期: 2017-07-28 网络出版日期: 2017-10-10

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20171010.1455.002.html>

基金项目: 国家自然科学基金项目 (31370664); 河北省科技计划项目 (16226309D-2); 在读研究生优秀课题创新项目 (1099009)

第一作者研究方向为林木遗传育种。E-mail: zuolihui001@163.com

通信作者: 王进茂, 研究方向为林木遗传育种。E-mail: swjs224@aliyun.com

杨敏生, 研究方向为林木遗传育种。E-mail: yangms100@126.com

coincide with the traditional taxonomy, which provides the molecular basis for the classification of *Ulmus* plants.

Key words: ulmaceae; transcriptome; EST-SSR; primer exploitation

简单序列重复 (SSR, simple sequence repeats) 是一类由不同核苷酸 (1 ~ 6 个) 为重复单元组成的串联重复序列^[1-2]。SSR 因其两侧位点为相对保守的单拷贝序列, 具有较高的通用性, 在生物系统进化、种质资源评价、目标基因的标定及指纹图谱的绘制等方面的研究发挥着重要作用。根据来源可将 SSR 分为基因组 g-SSR 和表达序列标签 EST-SSR。由于 EST-SSR 来源于基因转录区域, 因此 EST-SSR 比 g-SSR 具有更高的通用性^[3-4]。EST-SSR 的多态性可能与基因的功能存在联系, 可以更加准确地分析不同物种间的遗传进化关系。测序技术的发展极大地提升了 EST 序列数据库的容量, 已成为 EST-SSR 引物开发的重要来源。目前, 基于转录组 EST 序列开发 EST-SSR 标记的相关研究, 在其他物种中已经有大量报道^[5-9]。

榆树为榆科 (Ulmaceae) 榆属 (*Ulmus* L.) 植物, 是我国分布最为广泛和重要的树种之一。榆树起源演化历史悠久, 是我国古老的乡土树种, 是第三纪地层的古老树种之一。我国对榆树的研究较早, 早在西汉时期《范胜之书》中, 就有关于榆树物候观察和应用等方面的记载。目前全世界大约有 40 余种榆树, 主要分布在我国北部地区及苏联中亚、哈萨克斯坦、外贝加尔、西伯利亚、蒙古、朝鲜等国家, 是典型的北温带落叶阔叶树种^[10-11]。榆树材质优良, 具有良好的耐旱、耐寒、耐盐碱和抗风能力, 是生态环境恶劣区最重要的保土、固沙的优良树种^[12]; 同时由于榆树枝叶繁茂、耐修剪, 观赏性较高, 被广泛用于园林植物造型^[13-14]。

近年来测序技术和生物信息学快速发展, 转录组测序已经成为科学研究的重要分析手段。目前利用转录组数据开发 SSR 标记仍然是较为经济、高效的 DNA 分子标记开发策略, 测序技术的发展极大地提高了 EST-SSR 引物的开发效率^[15-17]。然而, 作为第三纪后更早的古老树种之一, 迄今为止, 对榆科植物的遗传学相关研究背景十分薄弱。目前关于榆树的研究还仅限于表型、生理学及少量的分子生物学探讨, 可供使用的分子标记严重匮乏, 无法满足目前研究需要。而且关于榆科植物的分类一直是很多分类学家争论的焦点, 榆科植物的分类相对混乱, 尚无统一的定论。到目前为止, 未见任何关于榆树 EST-SSR 标记开发的相关报道。本研究以白

榆 (*U. pumila* L.) 叶片转录组测序数据为分析材料, 对 1 kb 以上 Unigene 序列进行 EST-SSR 标记的开发, 对 EST-SSR 位点的组成、分布等基本特征进行分析, 并检测引物在不同榆树资源中的可用性, 开发榆树 EST-SSR 引物, 进一步为榆树种质资源评价、连锁图谱构建等相关研究提供理论基础和指导。

1 材料与方法

1.1 试验材料

以正常生长的白榆为试验材料, 分别采集不同时期的新鲜叶片, 提取 RNA, 检测合格后, 送至北京百迈克生物信息科技有限公司进行建库, 采用 HiSeq 2500 进行高通量测序。本研究中各样品 Q30 碱基百分比均在 92.08% 以上, 测序质量较高, 满足后续试验要求。组装后共获得 62945 条 Unigenes, 其中长度在 1 kb 以上的 Unigenes 有 17624 条。

为了验证开发引物在不同类型榆树间的有效性, 本研究以榆属 4 个种 1 个变种共 34 个优良无性系为验证材料, 对筛选的 EST-SSR 标记进行验证, 分别包括白榆 7 个 (编号 B1 ~ B7)、裂叶榆 (*U. laciniata* (Trautv.) Mayr.) 2 个 (编号 Q1 ~ Q2)、黑榆 (*U. davidiana* Planch.) 7 个 (编号 H1 ~ H7)、春榆 (*U. davidiana* Planch. var. *japonica* (Rehd.) Nakai) 14 个 (编号 C1 ~ C14)、大果榆 (*U. macrocarpa* Hance) 4 个 (编号 D1 ~ D4)。所有材料均采自张家口黑龙山地区天然野生植株当年生枝的顶端嫩叶, 低温保存带回实验室, 储藏于 -70 °C 冰箱, 用于后续 DNA 提取及 SSR 标记分析。

1.2 试验方法

1.2.1 植物 DNA 的提取 采用改良 CTAB 法^[18]提取所有榆树的总 DNA, 检测合格后, 将所有样品 DNA 统一稀释至 30 ng/μL, 放入 -70 °C 冰箱中保存备用。

1.2.2 转录组 EST-SSR 位点的筛选 为保证序列组装和后续分析的准确性, 首先对测序原始数据进行过滤, 去除测序引物、接头等人工序列和质量值较低的数据, 得到高质量 Reads。经过测序质量控制, 共得到 40.25 Gb 干净数据 (Clean Data), 各样品 Q30 碱基百分比均不小于 92.08%, 测序质量较高, 满足后续试验要求。采用 Trinity 软件将测序 Reads 进行拼接, 得到 Transcript 和 Unigene 序列。使用

MISA^[19]程序(<http://pgrc.ikp-gatersleben.de/misa>)对 1 kb 以上 Unigene 序列进行 SSR 位点搜索,搜索标准为:单核苷酸至六核苷酸最低重复次数分别为 10、6、5、5、5、5 次。SSR 引物设计采用 Primer Premier 5 软件进行,引物长度设置在 18 ~ 25 bp 之间,40% ~ 55% 的 GC 碱基含量,退火温度(T_m)在 55 ~ 65 ℃ 之间,且正向和反向引物退火温度之差在 5 ℃ 以内,PCR 扩增产物大小在 100 ~ 400 bp 之间,引物之间无引物二聚体和二级结构的出现。所有 SSR 引物均由生工生物工程(上海)股份有限公司合成。

1.2.3 引物筛选与 PCR 扩增验证 PCR 反应采用 10 μL 体系,各成分及比例如下:2 × Taq Master Mix 4 μL,正引物、反引物(10 μmol/L)各 0.5 μL,DNA 模板 1 μL,去离子水 4 μL。PCR 反应程序:95 ℃ 预变性 5 min;95 ℃ 变性 1 min,50 ~ 60 ℃ 退火 1 min,72 ℃ 延伸 1 min,30 个循环;72 ℃ 延伸 7 min,4 ℃ 保存 5 min。PCR 反应采用 Biometer 公司生产的 TG 型 PCR 循环仪,PCR 产物用 8% 非变性聚丙烯酰胺凝胶进行电泳分析。

1.3 统计与分析

统计所有 Unigenes 序列的 SSR 出现频率和分布规律,并分析 SSR 位点的数量及类型。对设计的

引物进行验证分析,将得到的扩增图谱进行 2 进制编码,形成 0/1 数据矩阵。采用 PIC Calc 0.6 软件计算各引物的多态性信息含量。采用 DPS 7.05 和 MEGA 5.0 软件,计算不同榆树间的遗传距离并进行聚类分析。采用 POPGENE32 软件对不同类群榆树的多态性位点百分率(P)、 $Nei's$ 基因多样性(H)、 $Shannon's$ 指数(I)和有效等位基因数(N_e)等进行遗传多样性数据计算和对比分析。

2 结果与分析

2.1 EST-SSR 位点的数量与分布规律

通过分析可知,在 17624 条 Unigenes 序列 36609384 bp 长的序列中,总共检测到 8288 个精确型 SSR 和 569 个复合型 SSR 位点,发生频率分别为 47.03% 和 3.23% (表 1)。本研究中榆树的 SSR 位点类型丰富,单核苷酸至六核苷酸重复类型均存在。对各组分数量分析表明,单核苷酸重复、三核苷酸重复和二核苷酸重复类型的数量较多,占主导地位,3 种类型的 SSR 位点总数占精确型全部位点的 98.90%,说明榆树的 EST-SSR 主要以单、二和三碱基重复单元为主。

表 1 EST-SSR 引物检测结果
Table 1 Detection result of EST-SSR primer

SSR 检测结果 SSR searching result	数量 Number
序列总数 Total number of examined sequences	17624
序列长度(bp) Total size of examined sequences	36609384
SSR 位点总数 Total number of identified SSR	8288
包含 SSR 序列数 Number of SSR containing sequences	5780
包含 1 以上 SSR 位点序列数 Number of sequences containing more than 1 SSR	1689
复合 SSR 位点数 Number of SSR present in compound formation	569
单核苷酸 Mononucleotide	3344
二核苷酸 Dinucleotide	1987
三核苷酸 Trinucleotide	2866
四核苷酸 Tetranucleotide	76
五核苷酸 Pentanucleotide	11
六核苷酸 Hexanucleotide	4

2.2 转录组 SSR 基序重复类型和频率特征

对所有精确型 SSR 序列长度分析可知,榆树的 SSR 序列长度在 10 ~ 50 bp 间变化,主要以 10 ~ 22 bp 的短序列为主,占全部的 99.49% (图 1),其中 15 bp 长度的 SSR 数量最多,达到 25.29%,且主

要以三碱基重复单元为主。

2.3 SSR 重复单元的组成分析

榆树转录组的精确型 SSR 位点总共含有 142 种重复单元,重复单元种类最多的是三核苷酸重复单元,共 60 种,其余依次为四核苷酸(51)、二核苷

酸(12)、五核苷酸(11)、单核苷酸(4)和六核苷酸(4)。对所有重复单元的数量进行分析可知(表2),占比例最大的为 A/T(3330,40.18%),其次为 AG/CT(1211,14.61%)和 AAG/CTT(568,6.85%)。在二核苷酸重复单元中 AC/GT、AG/CT 和 AT/TA 占二核苷酸总数的 99.85%。三核苷酸

重复单元以 AAG/CTT、AAC/GTT、AAT/ATT、AGC/CTG 和 ATC/ATG 为主,占有三核苷酸重复单元的 78.79%。四核苷酸重复单元的类型较少,主要以 AAAC/GTTT、AAAT/ATTT 和 AAAG/CTTT 重复单元为主;五核苷酸和六核苷酸重复单元总共出现 15 次,出现次数最少。

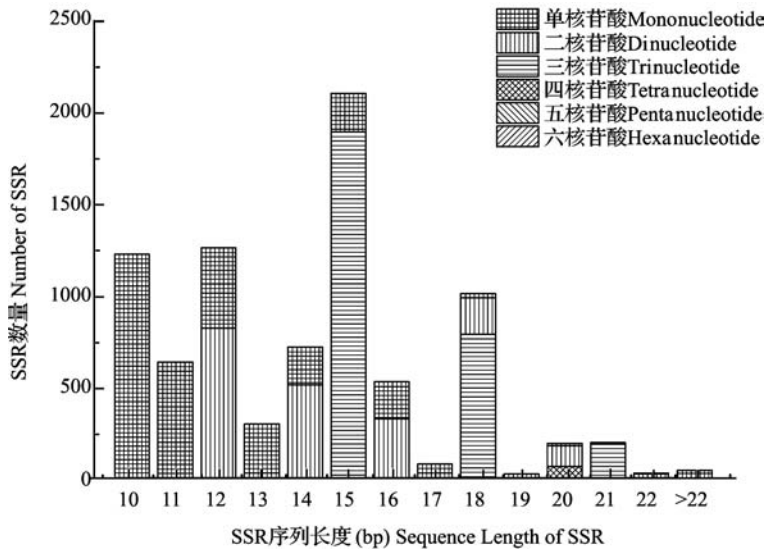


图 1 SSR 序列长度分布
Fig. 1 Distribution of SSR sequence length

表 2 不同 EST-SSR 重复单元结构组成

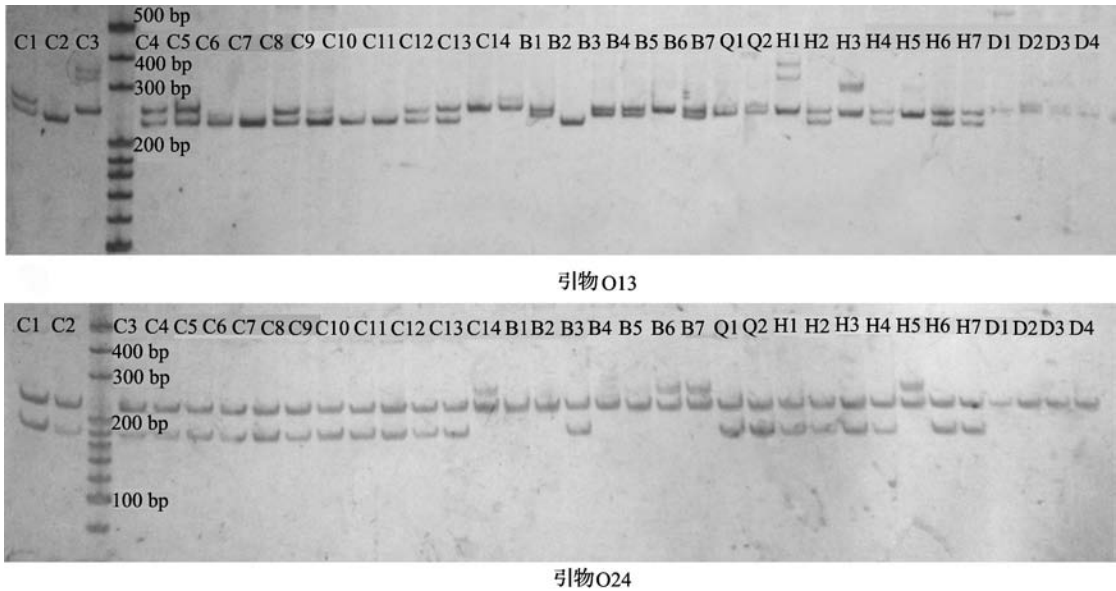
Table 2 Structure of different EST-SSR repeat units

重复单元 Repeats	重复次数 Number of replication												总数 Total
	5	6	7	8	9	10	11	12	13	14	15	> 15	
A/T						1208	630	435	295	206	211	345	3330
C/G						10	3	1					14
AC/GT		89	57	42	14	10	5	1					218
AG/CT		479	298	191	139	89	15						1211
AT/AT		249	152	90	46	15	3						555
CG/CG			1	1		1							3
AAC/GTT	332	153	28	1									514
AAG/CTT	391	144	32		1								568
AAT/ATT	267	147	23										437
ACC/GGT	98	28	17										143
ACG/CGT	55	13	2										70
ACT/AGT	77	34	16	2									129
AGC/CTG	264	109	17	2									392
AGG/CCT	123	38	12	2									175
ATC/ATG	217	94	34	2									347
CCG/CGG	61	24	5	1									91
其他 Other	72	15	2	1		1							91
总计 Total	1957	1616	696	335	200	1334	656	437	295	206	211	345	8288

2.4 EST-SSR 引物验证分析

为了验证本研究开发引物的有效性,根据 SSR 信息随机挑选 90 对 EST-SSR 引物对不同的榆树资源进行验证分析。通过分析最终有 46 对引物能够

扩增出预期大小的条带,有效扩增率为 51.11%。图 2 分别为引物 O13 和 O24 对 34 个榆树品种的扩增结果,由图可知,不同引物的扩增结果存在较大差异,即不同引物的效率值不同。



从左到右依次为春榆(C1~C14)、白榆(B1~B7)、裂叶榆(Q1~Q2)、黑榆(H1~H7)和大果榆(D1~D4)
From left to right were *Ulmus davidiana* var. *japonica* (Rehd.) Nakai (C1-C14), *Ulmus pumila* (B1-B7), *Ulmus laciniata* (Trautv.) (Q1-Q2), *Ulmus davidiana* (H1-H7), *Ulmus macrocarpa* (D1-D4)

图 2 引物 O13、O24 对所有榆树品种的扩增结果

Fig. 2 Amplification results of all elm species by primer O13 and O24

2.5 引物多态性分析

对所有引物扩增结果进行统计分析可知(表 3),46 对引物总共扩增出 110 个条带。不同引物的扩增总条带数在 1~6 个间变化,平均为 2.39 个,引物 J14 最多(6 个)。多态性条带数在 0~6 个

间变化,其中有 17 对引物(36.96%)为非多态性,说明这些位点所在的基因相对保守,变异程度较低。各引物的多态性信息含量 *PIC* 在 0.054~0.683 间变化,引物 O19 最低,引物 O3 最高,整体多态性水平较低。

表 3 不同引物在榆树中的扩增结果

Table 3 Amplification results of different primers in elms

引物 Primer	正向引物序列 Forward primer sequence (5'-3')	反向引物序列 Reverse primer sequence (5'-3')	重复单元 Repeat unit	扩增带数 Number of bands	多态性 带数 Number of polymorphic bands	多态性 信息含量 <i>PIC</i>
J1	CGTAACCTTCTTCACGCCTC	ATAGGGGACAAAATCGAGGG	(TTC) ₆	3	3	0.566
J5	TGTGCTTTTGTGAGGCTG	AAGAAGGGAGCAAAGAAGGC	(GCC) ₇	2	1	0.369
J6	AGAAATGGGGTTTGTCTGTG	GCAGAAGAGCTTTGGGTTTG	(TTC) ₆	3	3	0.524
J7	TGAAAGTGGCAAAGCTAGGG	TCGTTCTACGGCAAAGATGA	(TTA) ₆	4	4	0.477
J9	TACCTTCTCTCAACTCAGCG	ACACCCACATCGTTTCCTC	(TC) ₆	1	0	—
J10	ACCCTCTCTCCGGTGCTAAT	AGCCGATTCCCTGAAATTCT	(AAG) ₆	1	0	—
J11	CCTCATCTGCCTTGTGGAT	GTGGATAGACTGCTGCCAT	(TCT) ₆	1	0	—

表 3(续)

引物 Primer	正向引物序列 Forward primer sequence (5'-3')	反向引物序列 Reverse primer sequence (5'-3')	重复单元 Repeat unit	扩增带数 Number of bands	多态性 带数 Number of polymorphic bands	多态性 信息含量 <i>PIC</i>
J12	AGGAATGAACTCCTCGCTCA	AGGAGAGGGCATCACCTTTT	(CTT) ₅	3	2	0.277
J13	CTTATCTGCTTGGCCACCAT	TCTCTCTCCCTCCCTTCCTC	(GAT) ₇	4	3	0.225
J14	CACATTCTAATGAATCGCCA	CGCCATGTCCGTAACCTATC	(A) ₁₅ * (T) ₁₀	6	6	0.458
J15	TGGATAAAGGTCGGAAAACG	GGAGAGGCGTCTCAAGTGTC	(AGA) ₆	4	4	0.232
J16	GGAGACGATGATACGGAGGA	CTCTTGGACTCCCTCACAGC	(GGC) ₅	3	3	0.4
J17	GACGTCGCTCTGGGAAAATGT	CTCGAAGAACTGGACGAAGG	(TCC) ₆	3	3	0.397
J18	TTCATTGCGGCAATCATCAA	TCCCAAACTCACAAGGGAAG	(ATAG) ₆	1	0	—
J20	AACTGGGCCAACAGTTTCAG	GAAGCTGCTGTTCGGAGTGT	(ATC) ₅	1	0	—
J21	CCCATTTCAGGGATACCGAT	GGCTGCCATTTTGTGTTTT	(GCT) ₅	1	0	—
J22	CCATCAAAGGCATTAATTTTCC	TCCAAGCGGAGCTAGAAAAG	(TCT) ₇	1	0	—
J24	GCTTTCCAAGGAAGATGCAA	GGCAAACGCTTACGAAAAGA	(AT) ₇	1	0	—
J25	ACACATCAATTCTCGGGCTC	AATAGCGGCATCAAAAATGC	(A) ₁₁ * (T) ₁₀	1	0	—
F1	GGCAAACGATTGTAACACTTG	ATCGGAGCTGAAGAACAAGC	(AAT) ₅	3	3	0.336
F3	TCCTTTTCGATTGGTCTGACA	TTTTGGGGTCATATTCAGCC	(TTG) ₅	2	2	0.37
F6	TGTGTCTGAGTCTGCTTGGA	TCTTGCAAACACGGAATGAA	(CT) ₆	2	1	0.311
F7	AGCAGCACCAACAACAACAG	TCGTGAAGTTGGCTTCAATG	(AGG) ₅	1	0	—
F8	GGAAAGCCCGCTTTATTTCT	AGACGCTTCGAGAGCTTGAG	(CTT) ₅	1	0	—
F9	TTTCCCGTACAAAGTTTGCC	GCTCCATTGTCAAGCAACAA	(AG) ₇	3	3	0.257
F10	TCGATTTCCGCTTTGCTAGT	ATTAGCCTTTGGGAAGCCAT	(TTC) ₅	4	4	0.549
F12	TGGTCAAGGGGAACAAACTC	CCACCATAAGCACCACTCT	(TGG) ₇	3	2	0.146
F14	GGTAGGGAACCGAAGTAGCC	GCCATGTATTGGGGAGAAGA	(GAA) ₅	1	0	—
F17	GGAGGTGGGACTGCTAATGA	TACCGGTGCCTAATCCAGAC	(CAA) ₅	2	1	0.171
F19	TGGTCTGGCTACTCTGACC	TCCATAGAGGAGTATGGCCG	(CCT) ₆	4	4	0.524
F20	AGATGGATCAACCTGTTCCG	TCATATAAAGCTCCAAAACCATGA	(GAA) ₅	2	1	0.361
F22	CGTTCATCACAAATTTCCAA	CGTTTTTCTCTCTTGTGCGC	(TTG) ₅ * (TGC) ₅	1	0	—
F23	TTCCATCAAGGCAACAACAA	ATGATCACCACTCCTCCAG	(CAA) ₆	1	0	—
F25	GAAGGCAGAGGACCAAGAGA	TGGATTTGGGTGTGTGCTA	(ATTT) ₅	2	1	0.199
O2	AATTTCCCATTTCCGTCCTC	AGGGAAAACAGCGAAGAACA	(TC) ₆	2	2	0.141
O3	AAAGCTCAAAACCAAGCAA	CAACTGCATTGAAGAGGCAA	(CT) ₈	5	5	0.683
O8	GCCCTGCAACTGAACTTCTC	GCAAAGGGTGGACCTGTAAA	(GCC) ₆	3	3	0.333
O10	AGAAGAACCCTTTTGGATT	TGGAGTCCAAAACAACAAAAA	(T) ₁₂ * (T) ₁₀	3	3	0.455
O11	ATTTTCCCAATGGAAAGCTG	TTGAAAATCTTGCCAATCCC	(CAAAA) ₅	4	4	0.515
O13	ATTCCCAAGGAGGAGAGAA	TTGAGCCAAAAAGAAAAAGCA	(TG) ₆	4	4	0.608
O17	CCGCCGTGAAAGCTACTATT	ACCGGATTCTCTTCTCTCGT	(CGG) ₆	4	3	0.417
O19	CACCCCTTTGTTGAAAGCAT	TGCTTCTCTCTCCAGGAA	(CT) ₈	2	1	0.054
O21	AAATCTTAGGCCTTCTCCCG	TCAAGGAGATCGAAACCACC	(AAT) ₆	1	0	—
O24	CCTCGTCTCTTCGTAGTCG	GCTCACACCCACCGTAATCT	(GCG) ₅	3	2	0.45
O20	CAAAGAGAGCTCATCCCTGC	GCCTAGCCTCTTTTGTCCC	(TCT) ₆	2	0	0.375
O25	CATGGACGATGTCTGATTGA	GCGAATTGGACGATTGTGTT	(TA) ₇	1	0	—

* 表示为复合型 SSR 位点

* Means compound SSR loci

2.6 不同群体遗传多样性对比分析

通过分析可知,黑龙江地区不同榆树类群的遗传多样性存在较大差异(表4)。不同榆树类群多态性位点百分比在 27.59%~79.31% 间变化,春榆和黑榆较高,裂叶榆最低。不同类群观测等位基因数 N_a 同样为春榆和黑榆最高(均为 2.1034),裂叶榆最

低(1.2759)。有效等位基因数 N_e 、Shannon's 指数 I 、Nei's 基因多样性 H 、总体期望杂合度 H_e 均为黑榆最高,裂叶榆最低。观测杂合度 H_o 为春榆最高(0.3350),裂叶榆最低(0.2069)。研究表明,黑龙江调查范围内,春榆和黑榆的遗传多样性较高,裂叶榆的遗传多样性较低。

表 4 不同榆树类群遗传多样性分析

Table 4 Genetic diversity of different elm populations

类群 Population	$P(\%)$	N_a	N_e	I	H	H_o	H_e
春榆 <i>U. davidiana</i> Planch. var. <i>japonica</i> (Rehd.) Nakai	75.86	2.1034	1.5296	0.4390	0.2745	0.3350	0.2847
白榆 <i>U. pumila</i> L.	68.97	1.8621	1.4908	0.4108	0.2590	0.2759	0.2789
裂叶榆 <i>U. laciniata</i> (Trautv.) Mayr.	27.59	1.2759	1.2483	0.1822	0.1293	0.2069	0.1724
黑榆 <i>U. davidiana</i> Planch.	79.31	2.1034	1.6575	0.5018	0.3082	0.3005	0.3319
大果榆 <i>U. macrocarpa</i> Hance	58.62	1.7586	1.5414	0.4131	0.2705	0.3276	0.3091

P :多态性位点百分比; N_a :观测等位基因数; N_e :有效等位基因数; I :香农指数; H :Nei's 基因多样性; H_o :观测杂合度; H_e :期望杂合度,下同
 P :Percentage of polymorphic loci, N_a :Observed allele number, N_e :Effective number of alleles, I :Shannon's index, H :Nei's gene diversity, H_o :Observed heterozygosity, H_e :Expected heterozygosity,the same as below

2.7 基因位点多样度及遗传结构分析

对所有多态性位点的遗传多样度及遗传结构分析可知(表5),不同位点的观测杂合度 H_o 存在较大差异,变化范围为 0~0.9412,位点 J1 最低,J17 最高。不同位点的总体期望杂合度 H_e 变化范围为

0.0579~0.7037,019 位点最低,03 位点最高。不同位点的 Nei's 基因多样性 H 变化范围为 0.0571~0.6933,019 最低,03 最高,所有位点整体多样度较低,平均仅为 0.3876,这也是 EST-SSR 位点多态性信息含量 PIC 值较低的主要原因。

表 5 不同位点杂合度与基因多样度分析

Table 5 Heterozygosity and gene diversity of different loci

位点 Locus	H_o	H_e	H	Ave_He	位点 Locus	H_o	H_e	H	Ave_He
O2	0.0588	0.1124	0.1107	0.1203	J14	0.2941	0.4456	0.4390	0.2270
O3	0.0882	0.7037	0.6933	0.3730	J15	0.0882	0.1945	0.1916	0.2427
O8	0.0294	0.3648	0.3594	0.1097	J16	0.0882	0.3824	0.3767	0.1754
O10	0.1471	0.6361	0.6267	0.3136	J17	0.9412	0.5211	0.5134	0.5106
O11	0.1765	0.5035	0.4961	0.3476	F1	0.2353	0.3494	0.3443	0.1469
O13	0.3824	0.5737	0.5653	0.3893	F3	0.0294	0.4965	0.4892	0.1444
O17	0.8529	0.5206	0.5130	0.4837	F6	0.3529	0.2950	0.2907	0.1219
O24	0.7647	0.5180	0.5104	0.3903	F9	0.0882	0.1918	0.1890	0.1204
O19	0.0588	0.0579	0.0571	0.0703	F10	0.6176	0.6273	0.6181	0.4102
J1	0	0.6514	0.6419	0.2306	F12	0.0882	0.0865	0.0852	0.1062
J5	0.7353	0.4719	0.4650	0.3668	F17	0.1176	0.1124	0.1107	0.1000
J6	0.1176	0.5737	0.5653	0.0760	F19	0.9118	0.6093	0.6003	0.5614
J7	0.5294	0.5013	0.4939	0.3244	F20	0.6176	0.4333	0.4269	0.2832
J12	0.2059	0.1927	0.1899	0.2143	F23	0.1471	0.1383	0.1362	0.1077
J13	0.1471	0.1418	0.1397	0.1330	平均 Mean	0.3073	0.3933	0.3876	0.2483

Ave_He :平均期望杂合度
 Ave_He :Average expected heterozygosity

遗传一致度与遗传距离分析表明(表6),不同榆树间的遗传距离在0.1170~0.4566间变化,白榆和黑榆间遗传距离最近,春榆和大果榆间的遗传距

离最远。遗传距离矩阵结果表明,白榆、黑榆和春榆间的遗传关系较近,而裂叶榆和大果榆与其他榆树亲缘关系较远。

表6 不同榆树类群遗传一致度(上三角)与遗传距离(下三角)

Table 6 Nei's genetic identities(upper triangular) and genetic distances(lower triangle) among different elm populations

类群 Population	春榆 <i>U. davidiana</i> Planch. var. <i>japonica</i> (Rehd.) Nakai	白榆 <i>U. pumila</i> L.	裂叶榆 <i>U. laciniata</i> (Trautv.) Mayr.	黑榆 <i>U. davidiana</i> Planch.	大果榆 <i>U. macrocarpa</i> Hance
春榆 <i>U. davidiana</i> Planch. var. <i>japonica</i> (Rehd.) Nakai		0.8565	0.6974	0.8599	0.6334
白榆 <i>U. pumila</i> L.	0.1549		0.6817	0.8896	0.7042
裂叶榆 <i>U. laciniata</i> (Trautv.) Mayr.	0.3604	0.3832		0.7699	0.6421
黑榆 <i>U. davidiana</i> Planch.	0.1509	0.1170	0.2616		0.7633
大果榆 <i>U. macrocarpa</i> Hance	0.4566	0.3507	0.4430	0.2700	

2.8 不同榆树个体聚类分析

为了进一步验证开发引物的有效性,根据开发引物的扩增结果,利用DPS和MEGA 5.0软件采用类平均法对黑龙江地区的5类榆树的天然野

生种质资源进行聚类分析。由图3可知,所有34个榆树无性系被100%区分开,且可以分为5类。第1类包括所有大果榆无性系(4个),大果榆与其他榆树亲缘关系较远,单独聚为一类;第2类包括

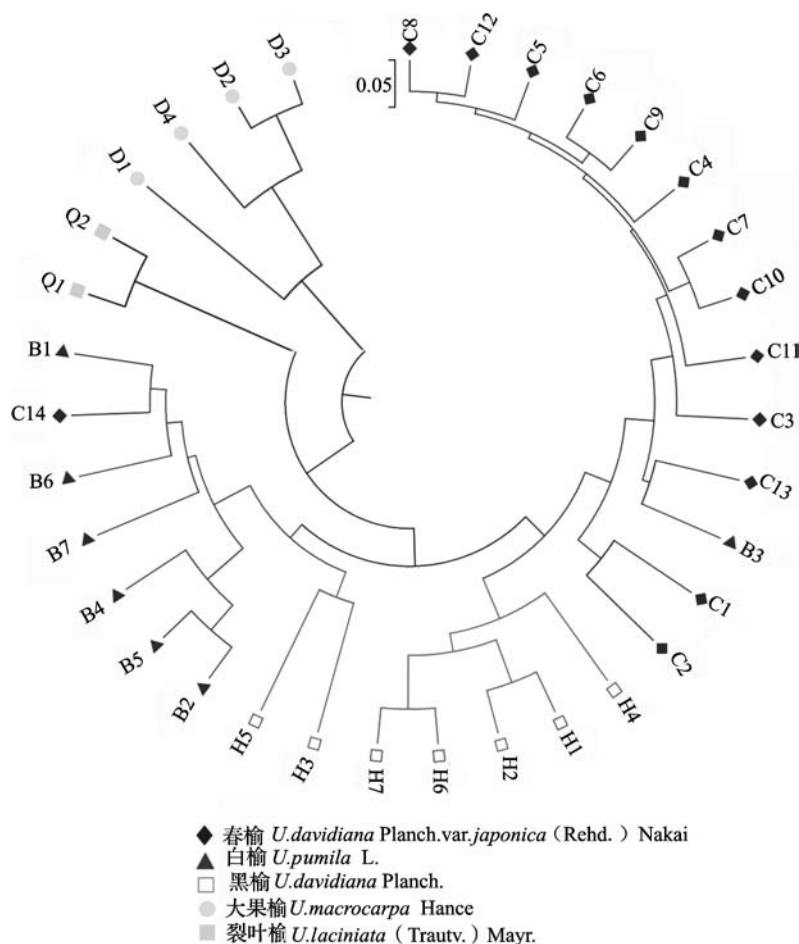


图3 基于EST-SSR标记的榆树品种聚类图

Fig. 3 Dendrogram of the *Ulmus* L. based on EST-SSR markers

2 个裂叶榆无性系;第 3 类主要包括 6 个白榆无性系、1 个春榆无性系和 2 个黑榆无性系,共 9 个无性系;第 4 类全部为黑榆无性系,共 5 个无性系;第 5 类主要为春榆,包括 13 个春榆和 1 个白榆,共 14 个无性系。绝大部分榆树无性系均按其起源分开。由聚类图可知,黑榆和春榆的亲缘关系较近,与传统分类学结果一致。传统分类学认为,春榆为黑榆的一个变种,本研究为榆树的分类提供了分子依据,也从侧面证明了本研究开发引物的有效性。

3 讨论

近年来测序技术和生物信息学迅速发展,极大地推进了分子生物学和林木遗传育种工作的研究进程。高通量测序具有快速、高效、数据量大等特点,是以往研究方法所不具备的,已经成为目前科学研究的一个重要手段。高通量测序数据的增加为各种分子标记的开发提供了大量的资源,加速了引物开发效率^[20-21]。EST-SSR 在不同物种间具有较好的通用性,被广泛应用于遗传多样性分析、种质资源鉴定、遗传连锁图谱的构建等相关研究。同时,EST-SSR 也普遍存在多态性较低的缺点^[22],但转录组数据开发 EST-SSR 标记的效率较高,大量的 EST-SSR 标记可以弥补其多态性较低的问题。

本研究总共检测到 8288 个精确型和 569 个复合型 SSR 位点,主要以单、二和三碱基重复单元为主。随机挑选 90 对 EST-SSR 引物进行验证分析,有效扩增率为 51.11% (46 对),与刺梨^[23]的结果 (53.76%) 相近,普遍存在有效扩增率偏低的问题。目前利用转录组数据开发 EST-SSR,仍然是比较经济有效的手段,可以同时开发大量的引物。然而,在引物验证过程中存在一个难题,就是存在大量无效扩增现象,从而加大引物开发的工作量和难度。究其原因,可能存在以下几个方面:(1)测序误差,目前二代测序的读长较短,在测序及序列拼接过程中容易出现错误,降低序列的准确性,进而导致引物设计存在误差,引物无法与 DNA 模板结合,造成无效扩增;(2)同源序列干扰,在基因组其他区域存在与引物位点区域相似的同源片段,无法得到理想的目的片段;(3)内含子干扰,SSR 引物的扩增片段一般在 100~500 bp 之间,由于引物是根据 mRNA 序列信息开发,无法判断目的片段是否含有内含子,倘若目的片段跨内含子,会导致目的片段偏大,造成无效扩增;(4)RNA 编辑干扰,植物中(叶绿体、线粒体)广泛存在 RNA 编辑现象,即在转录过程中,遗传物

质发生改变的现象, RNA 编辑可能会导致引物设计误差,不能与 DNA 结合,造成无效扩增。

引物开发的目的是筛选能够扩增出清晰、多态性高、重复性好的 SSR 标记,然而在有效引物的验证、筛选过程中,普遍存在多态性引物比例低的现象,如在杏^[24] (62.5%)、杨树^[25] (58%)、海岛棉^[26] (43.42%)、杉木^[27] (40%) 和红豆杉^[28] (38.71%) 等物种的引物开发中,多态性引物的比例均较低。本研究中,多态性引物比例仅为 63.04%,比例同样偏低。在有效扩增引物中,不同引物的效率值(多态性)存在较大差异,其原因可能是不同基因受到的环境选择压力不同导致的。EST-SSR 位点多态性水平可以间接反应出所在基因的多样性水平,而环境选择压力与基因的多样性水平存在直接关系,如 Y. H. Li 等^[29] 对大豆的野生品种和栽培品种的基因组进行对比发现,栽培大豆受到正向选择的基因多是与抗旱、产油等相关,而野生大豆受到正向选择的基因大多是与物种适应性相关。基因的进化速率越快,EST-SSR 位点的多态性水平就越高;相反,如果 EST-SSR 位点所在基因非常保守,则位点多态性较低,而这些非多态性位点在种质资源评价、遗传多样性分析等研究中的参考价值极低。在植物指纹图谱构建中,多态性 EST-SSR 引物的开发是目前引物开发的重点。在 EST-SSR 引物开发研究中,应把变异程度较大的基因作为重点对象,可能会提高多态性引物的开发效率。在植物指纹图谱构建中,由于不同引物的效率值差异较大,不同的引物组合对试验结果有较大影响,导致聚类结果偏移、混乱,出现与其起源或传统分类不符等现象,从而使结果的可靠性降低。因此,选用携带信息量大、代表性强、分布均匀的引物组合,才能准确、快速进行植物亲缘关系及遗传多样性分析。

本研究开发的引物在不同榆树资源中具有良好的通用性,利用开发的 SSR 引物可以将不同类型的榆树资源很好地区分开。聚类结果均表明,春榆、黑榆和白榆间亲缘关系较近,而裂叶榆和大果榆与其他榆树间的亲缘关系较远。相对于其他榆树,黑榆和春榆的亲缘关系最近,与传统分类学(春榆为黑榆的变种)吻合;大果榆的表型与其他榆树存在较大差异,如叶片较其他榆树粗糙、厚,且枝条存在两条规则的木栓翅,基于这些特征,传统分类学认为大果榆与其他榆树亲缘关系较远^[30]。本研究中的 SSR 聚类结果也显示为大果榆与其他榆树的亲缘关系较远,与传统分类吻合,说明 SSR 结果的准确性。本研究为榆树的分类提供了分子依据和科学指导。

目前本研究验证的 SSR 标记的范围和数量相对较少,在以后的工作中应继续扩大引物筛选范围,开发大批量的 SSR 引物,为榆科植物的分类、分子标记辅助育种、遗传图谱构建等相关研究提供更多、更有效的分子标记。

参考文献

- [1] Jia D M. Survey and analysis of simple sequence repeats (SSRs) in three genomes of *Candida* species[J]. *Gene*, 2016, 548(2): 129-135
- [2] Qu J B, Huang C Y, Zhang J X. Genome-wide functional analysis of SSR for an edible mushroom *Pleurotus ostreatus* [J]. *Gene*, 2016, 575(2): 524-530
- [3] Eujayl I, Sorrells M, Banm M, et al. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat [J]. *Theor Appl Genet*, 2002, 104(2): 399-407
- [4] Nicolai M, Pisani C, Bouchet J P, et al. Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*) [J]. *Genet Mol Res*, 2012, 11(3): 2295-2300
- [5] 徐小彪,姜春芽,廖娇,等. 中华猕猴桃矮型性状 EST-SSR 连锁标记的筛选[J]. *园艺学报*, 2010, 37(4): 553-558
- [6] 董清华,王西成,赵密珍,等. 草莓 EST-SSR 标记开发及在品种遗传多样性分析中的应用[J]. *中国农业科学*, 2011, 44(17): 3603-3612
- [7] Zhang M Y, Fan L, Liu Q Z, et al. A novel set of EST-derived SSR markers for pear and cross-species transferability in Rosaceae [J]. *Plant Mol Biol Rep*, 2014, 32(1): 290-302
- [8] 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
- [9] 马庆华,李京璟,赵天田,等. 基于 EST-SSR 标记的平欧杂种榛品种鉴定[J]. *植物遗传资源学报*, 2017, 18(5): 952-959
- [10] Sutton D. A new section of *Linaria* (Scrophulariaceae: Antirrhineae) [J]. *Bot J Linn Soc*, 1980, 81(3): 169-184
- [11] 张兴. 部分榆属种质资源亲缘关系及白榆辐射诱变的研究[D]. 哈尔滨:东北林业大学, 2012
- [12] 续九如,宋婉,邹受益,等. 榆属树种遗传改良研究现状及思考[J]. *北京林业大学学报*, 2000(6): 95-99
- [13] 成海钟. 园林植物栽培养护[M]. 北京:高等教育出版社, 2002
- [14] 史宝胜,任子蓓,杨露,等. 金叶榆叶片黄色素的提取及性质研究[J]. *北方园艺*, 2014(16): 133-136
- [15] Han X F, Ling Q F, Li C J, et al. Characterization of pikeperch (*Sander lucioperca*) transcriptome and development of SSR markers[J]. *Biochem Syst Ecol*, 2016, 66: 188-195
- [16] Guo L N, Zhao X L, Gao X F. De novo assembly and characterization of leaf transcriptome for the development of EST-SSR markers of the non-model species *Indigofera szechuensis* [J]. *Biochem Syst Ecol*, 2016, 68: 36-43
- [17] 林富荣,邢俊连,孟艳琼,等. 皂荚 EST-SSR 分子标记开发与评价[J]. *植物遗传资源学报*, 2017, 18(1): 148-154
- [18] 左力辉,韩志校,梁海永,等. 不同产地中国李资源遗传多样性 SSR 分析[J]. *园艺学报*, 2015, 42(1): 111-118
- [19] Guo S, Zheng Y, Joung J G, et al. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types [J]. *BMC Genomic*, 2010, 11: 384-391
- [20] Ueno S, Wen Y F, Tsumura Y. Development of EST-SSR markers for *Taxus cuspidata* from publicly available transcriptome sequences[J]. *Biochem Syst Ecol*, 2015, 63: 20-26
- [21] Li C J, Ling Q F, Ge C, et al. Transcriptome characterization and SSR discovery in large-scale loach *Paramisgurnus dabryanus* (Cobitidae, Cypriniformes) [J]. *Gene*, 2015, 557(2): 201-208
- [22] Eujayl I, Sorrells M, Baum M, et al. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs[J]. *Euphytica*, 2001, 9: 597-602
- [23] 鄢秀芹,鲁敏,安华明. 刺梨转录组 SSR 信息分析及其分子标记开发[J]. *园艺学报*, 2015, 42(2): 341-349
- [24] 上官凌飞,李晓颖,宁宁,等. 杏 EST-SSR 标记的开发[J]. *园艺学报*, 2011, 38(1): 43-54
- [25] 张新叶,宋丛文,张亚东,等. 杨树 EST-SSR 标记的开发[J]. *林业科学*, 2009, 45(9): 53-59
- [26] 张艳欣,林忠旭,李武,等. 海岛棉 EST-SSR 引物的开发与应用研究[J]. *科学通报*, 2007, 52(15): 1779-1787
- [27] 徐阳,陈金慧,李亚,等. 杉木 EST-SSR 与基因组 SSR 引物开发[J]. *南京林业大学学报:自然科学版*, 2014, 38(1): 9-14
- [28] 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
- [29] Li Y H, Zhou G Y, Ma J X, et al. De novo assembly of soybean wild relatives for pan genome analysis of diversity and agronomic traits[J]. *Nat Biotechnol*, 2014, 32(10): 1045-1052
- [30] 傅立国. 中国榆属的研究[J]. *东北林业大学学报*, 1980(3): 1-40