

基于转录组序列的长柄双花木 SSR 标记开发

孟艺宏¹, 徐 璿², 徐刚标¹

(¹中南林业科技大学林木遗传育种实验室, 长沙 410004;

²四川大学生命科学学院, 成都 610064)

摘要: 双花木属 (*Disanthus* Maxim.) 是金缕梅科 (Hamamelidaceae) 最原始的单种属, 为东亚地区特有属。长柄双花木 (*D. cercidifolius* var. *longipes* H. T. Chang) 是日本特有植物双花木 (*D. cercidifolius* Maxim.) 的变种, 仅分布于我国南方地区, 为国家Ⅱ级保护植物, 被列入我国濒危植物种名录, 在研究金缕梅科系统发育和东亚植物区系地理演化等方面具有重要的科学价值。由于可利用的 SSR 标记引物数量不足, 阻碍了长柄双花木遗传学研究。本研究对长柄双花木转录组序列进行高通量测序, 采用 de novo 组装方法, 共获得 32325 条 unigene。利用 MISA 软件, 搜索到 13779 个 SSR 位点, SSR 位点发生频率为 42.63%, 位点分布频率为 1/2.95 kb。不同重复基序类型中, 二核苷酸重复基序数量最多, 占总数 44.10%; 单核苷酸重复基序次之, 占总数 37.90%; 三核苷酸重复基序, 占总数 16.71%。二核苷酸重复基序中, AG/CT 重复基序数量最多; 三核苷酸重复基序中, AAG/CTT 重复基序数目最多, 其次是 ATC/ATG、ACC/GGT 和 AGC/CTG。从不同种群中任取 1 株构成 8 株无亲缘关系的随机样本, 对随机合成的 60 对 SSR 引物的有效性进行琼脂糖电泳检测的结果表明, 46 对引物扩增出目的条带, 扩增率达 76.67%。进一步利用 TP-M13-SSR 技术基因分型结果显示, 30 对引物扩增产物显示出多态性, 多态位点百分比为 65.22%。这表明, 基于转录组序列开发的长柄双花木 SSR 位点多态性较高。本研究结果有助于后续的长柄双花木种群遗传学及系统进化等方面研究。

关键词: 长柄双花木; 转录组; TP-M13-SSR; 引物开发

Development of SSR Markers Based on Transcriptome Sequences of *Disanthus cercidifolius* var. *longipes*

MENG Yi-hong¹, XU Xun², XU Gang-biao¹

(¹The Laboratory of Forestry Genetics, Central South University of Forestry and Technology, Changsha 410004;

²College of life science, Sichuan University, Chengdu 610064)

Abstract: *Disanthus* Maxim. belongs to a small eastern Asian genus of the family Hamamelidaceae. *D. cercidifolius* Maxim. is found in Japan, and this species is the relative of *D. cercidifolius* var. *longipes* H. T. Chang that is particularly distributed in the south part of China. *D. cercidifolius* var. *longipes* H. T. Chang is listed in the Red List of Endangered Plant Species and Second Grade of the List of Wild Plants Under State Protection in China. *D. cercidifolius* var. *longipes* H. T. Chang is of significant interest in phylogeny study and floristic geography in East Asia. However, the limited number of available SSR molecular markers inhibited the current genetic research. In this study, by unlocking the transcriptome datasets by RNA-seq, we obtained a total of 32325 uni-genes. 13779 SSR loci were found using MISA, with the occurrence frequency of 42.63%. That included 37.90% of mono-nucleotide repeat motifs, 44.10% of di-nucleotide repeat motifs and 16.71% of tri-nucleotide repeat motifs. The AG/CT was the most abundant di-nucleotide repeat motifs, while the AAG/CTT was the most abundant tri-nucleotide repeat motif, followed by ATC/ATG, ACC/GGT, and AGC/CTG. 60 pairs of random-selected primers were used to test its availability in the random sample of 8 unrelated individuals from different populations using agarose electrophoresis. Forty-six pairs of

收稿日期: 2018-01-10 修回日期: 2018-01-31 网络出版日期: 2018-05-17

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20180517.0925.002.html>

基金项目: 国家“十三五”重点研发计划课题(2016YFC0503103); 中南林业科技大学研究生科技创新基金(20183002)

第一作者研究方向为森林遗传学。E-mail: 1187885817@qq.com

通信作者: 徐刚标, 主要从事森林遗传学与林木遗传改良研究。E-mail: gangbiaoxu@163.com

primers amplified the target fragments, including 30 pairs that were found to be polymorphic using TP-M13-SSR technology. The rate of PCR amplification and polymorphism were 76.67% and 65.22%, respectively. Thus, the large number of SSR markers developed in this study might become valuable in population genetics and phylogeny analysis of *D. cercidifolius* var. *longipes* H. T. Chang, and also this work might provide an example in development of SSR markers in other species.

Key words: *Disanthus cercidifolius* var. *longipes* H. T. Chang; transcriptome; TP-M13-SSR; primer development

双花木属 (*Disanthus* Maxim.) 是金缕梅科 (Hamamelidaceae) 最原始属, 为东亚地区特有, 一种一变种。双花木 (*Disanthus cercidifolius* Maxim.) 为日本特有种。长柄双花木 (*D. cercidifolius* Maxim. var. *longipes* H. T. Chang) 为我国特有种, 仅零星分布于湖南、江西、浙江、广东等地海拔 1300 m 以下的山区。长柄双花木作为中国-日本植物区系的替代种, 在研究金缕梅科系统发育和东亚植物区系地理演化等方面具有重要的科学价值^[1]。由于长柄双花木分布区内人类活动频繁, 人为干扰和破坏严重, 长柄双花木现存种群及个体数量少, 已被列入世界自然保护联盟 (IUCN) 濒危物种红色名录 (<http://www.iucnredlist.org>), 以及我国濒危植物种名录和 II 级保护植物种^[2-3]。

遗传标记是开展遗传学分析、系统发育研究的基本工具。肖宜安等^[4]、谢国文等^[5]先后基于等位酶和 ISSR 标记技术开展了长柄双花木局部地区种群遗传结构分析, 为长柄双花木遗传资源保护策略制定提供了参考。但是, 等位酶标记只能检测编码蛋白质基因, 无法检测到基因组中非结构基因以及不改变蛋白质分子迁移率的变异; ISSR 标记为显性标记, 不能区分种群中显性纯合体与杂合体。因此, 基于这些标记方法估算的种群遗传多样性都会产生偏差。SSR 标记具有全基因组覆盖、共显性遗传、重复性好、多态性高等特性^[6-7], 已成为遗传多样性分析、系统发育研究、遗传图谱构建、种质鉴定及基因定位的首选标记^[8]。但是, SSR 标记引物具有物种特异性, 开发 SSR 标记是开展目的物种 SSR 标记分析的前提。Gao 等^[9]利用磁珠富集微卫星分离法开发了 9 对长柄双花木 SSR 引物, 这远不能满足进一步开展种群遗传分析、系统发育研究的需要。

传统的 SSR 标记开发方法主要包括富集微卫星分离法、PCR 分离法、同源序列借鉴法和基因组文库测序法, 但都存在试验步骤繁琐、耗时费力、位点在基因组中覆盖率低等问题。利用 NCBI、EMBL、DDBJ 等公共基因数据库中的共享序列开发 SSR 标记, 但有限的基因序列资源仍然是共享基因序列不

足的物种 SSR 标记开发的瓶颈^[10]。近年来, 第二代高通量测序技术的发展, 极大降低了测序成本, 诸多未知基因组序列的物种较易获得包含 EST 序列的转录组数据, 能挖掘、开发出丰富的 SSR 标记位点, 提供更为全面的序列信息, 为基因组信息缺乏物种的 SSR 引物开发带来了机遇^[11-14]。本研究基于长柄双花木高通量转录组测序数据, 采用 de novo 组装方法构建其转录组参考序列, 运用生物信息学方法分析其转录组 unigenes 中 SSR 位点, 利用 TP-M13-SSR (simple sequence repeat with tailed primer M13) 技术验证开发出的 SSR 位点多态性, 旨在为后续的长柄双花木种群遗传学分析和系统发育进化研究提供实用可行的标记工具。

1 材料与方法

1.1 材料

长柄双花木当年生幼嫩叶片分别采自江西井冈山、玉山、南丰, 浙江开化, 湖南新宁、宜章、莽山, 广东连县 8 个自然种群。野外采集的样本, 放入盛有硅胶的密封袋中, 带回实验室备用。其中, 莽山种群中, 采集 1 株叶片样本, 立即用液氮快速处理, 干冰保存, 用于转录组测序。

1.2 研究方法

1.2.1 转录组数据产生 采用 Trizol 法提取 RNA, SMART 法构建 cDNA 文库, 利用 Illumina HiSeq4000 测序仪进行测序, 获得的 RNA-seq 测序数据进行质控。RNA-Seq 转录组测序委托上海美吉生物医药科技有限公司完成。利用 Trinity (<http://trinityrnaseq.sourceforge.net/>) 软件, 采用 de novo 法拼接、组装质控后高质量片段, 产生转录组数据。

1.2.2 SSR 位点挖掘与引物设计 采用 MISA 软件 (<http://pgrc.ipk-gatersleben.de/misa/>) 对长柄双花木转录组序列组装产生的 unigene 进行 SSR 位点定位、识别。MISA 软件参数为: 一、二、三、四、五、六核苷酸分别对应最少重复次数为 10、6、5、5、5、5。

基于引物长度为 16 ~ 30 bp、产物大小为 120 ~

500 bp、退火温度为 57 ~ 62℃、引物中 G + C 含量为 45% ~ 55%、4 种碱基随机分配的原则,利用 MISA 软件提供的用于批量转换微卫星两侧序列的 Perl 脚本 p3_in.pl 和 p3_out.pl,将 MISA 结果转换成 Primer 3.0 软件可识别格式,运行 Primer 3.0 软件程序批量开发、设计 SSR 引物。

1.2.3 DNA 提取与 PCR 扩增 采用李金璐等^[15]改良的 CTAB 法提取长柄双花木叶片 DNA,于 -20℃ 冰箱中保存待用。

从批量设计的引物中随机选取 60 对,在正向引物的 5'端加上 M13 尾巴(TGTAAAACGACGGC-CAGT),形成 TP-M13-SSR 正向引物。M13 引物采用 3 种荧光 ROX (red)、HRX (green) 和 FAM (blue) 标记。引物合成及 M13 引物荧光标记由上海 Sangon 公司完成。用 ABI PCR 扩增仪进行 PCR 扩增。第 1 轮扩增体系为:模板 DNA (10 ng/μL) 8 μL,TP-M13-SSR 正向引物 (2 μmol/L) 0.24 μL,反向引物 (2 μmol/L) 1.2 μL,10 × PCR buffer (Mg²⁺ free) 2 μL,Mg²⁺ (25 mmol/L) 1.6 μL,dNTPs (25 mmol/L) 2 μL,Taq 酶 (5 U/μL) 0.48 μL,ddH₂O 4.48 μL,共 20 μL。PCR 反应程序为:94℃ 预变性 5 min,94℃ 变性 45 s,53 ~ 57℃ 退火 45 s (不同引物的退火温度不同)、72℃ 延伸 45 s,35 个循环,72℃ 延伸 10 min,4℃ 保存待用。第 2 轮扩增以第 1 次 PCR 产物为模板,加入荧光标记 M13 引物 (2 μmol/L) 1.8 μL,10 × PCR buffer (Mg²⁺ free) 0.1 μL,Taq 酶 (5 U/μL) 0.16 μL,ddH₂O 0.12 μL。PCR 反应程序为:94℃ 预变性 5 min,94℃ 变性 30 s,53℃ 退火 30 s,72℃ 延伸 30 s,16 个循环,72℃ 最后延伸 10 min。0.8% 琼脂糖凝胶电泳检测 PCR 扩增产物。

1.2.4 TP-M13-SSR 基因分型与 SSR 引物筛选

将 3 种颜色荧光标记的 PCR 扩增产物混合,在 96 孔上机板每孔中加入 1 μL PCR 扩增混合物和 9 μL 去离子甲酰胺 (含 LIZ-500 分子量内标),离心、震荡混匀后,再离心。95℃ 变性 5 min 后,迅速降到 4℃ (冰浴),用毛细管自动荧光电泳系统 ABI 3730XL 检测、分型。

从每个种群中任取 1 株,组成 8 株没有亲缘关系的随机样本,用于 SSR 引物筛选及其有效性检测。每株 DNA 样本,PCR 扩增检测 3 次,验证其可重复性。使用 Gene Marker V 1.91 软件,判读扩增片段长度大小。

1.3 数据分析

SSR 发生频率为 SSR 的 Unigene 数目与总 Uni-

gene 数目之比,SSR 位点分布频率为 SSR 数目与 Unigene 总碱基数目之比,扩增率为扩增出条带引物数目与参试的引物数目之比,多态位点百分比为多态位点数目与位点总数目之比。

2 结果与分析

2.1 长柄双花木转录组中 SSR 位点开发

采用 Trinity 软件中 de novo 法,拼接、组装质控后的长柄双花木 cDNA 文库片段,共获得 32325 条 unigene,碱基总数量为 40626279 bp。其中,最长拼接长度为 15528 bp,最短拼接长度为 201 bp。采用 MISA 软件对 unigene 进行搜索,共搜索到 13779 个 SSR 位点,SSR 位点发生频率为 42.63%,平均分布频率为 1/2.95 kb。9919 条 unigene 分布有 SSR 位点,发生频率为 30.69%,其中,6118 条为单一型 SSR 位点,2813 条 unigene 含多个 SSR 位点,988 条 unigene 含有复合型 SSR 位点。采用 Primer 3.0 软件程序,基于 SSR 引物设计的基本原则,共开发出的 SSR 引物 5018 对。

2.2 长柄双花木转录组 SSR 位点分布特征

2.2.1 重复基序次数分布特征 长柄双花木转录组中重复基序及数量见表 1。由表 1 可知,长柄双花木重复基序种类比较丰富,从单碱基到六碱基重复均有分布,共有 193 种,但不同重复基序的 SSR 位点数量相差较大。四核苷酸重复基序最多,为 64 种,其次是三核苷酸重复基序,为 60 种。单核苷酸、二核苷酸、五核苷酸、六核苷酸重复基序数分别为 4、12、20、33 种。二核苷酸重复基序的 SSR 位点数量最多 (6076 个),占 SSR 位点总数的 44.10%,发生频率为 18.80%,平均长度为 15 bp;其次为单核苷酸重复类型 (5222 个),占 37.90%,平均长度为 12 bp;再次为三核苷酸重复类型 (2302 个),占 SSR 16.71%,平均长度为 17 bp。四、五、六核苷酸重复基序,SSR 位点数目较少,分别占 0.92%、0.15% 和 0.24%,平均长度分别为 21 bp、26 bp 和 35 bp。

在长柄双花木转录组的所有重复基序中,A/T 为优势重复基序,共 5131 个,占重复基序总数 37.24%;其次是 AG/CT,4765 个,占 34.58%。三核苷酸重复基序中,AAG/CTT 重复基序最多,有 683 个,占重复基序总数的 4.96%。四、五、六核苷酸重复基序数目较少,分别占重复基序总数 0.92%、0.15% 和 0.24%。

表 1 SSR 重复基序类型及数量

Table 1 Type and number of repeat motifs

重复类型 Repeat type	数目 Number	基序 Repeat motif	长度 (bp) Length	平均长度 (bp) Average	基序数目 Number	发生频率 (%) Frequency	比例 (%) Proportion
单核苷酸 Mononucleotide	4	A/T	10 ~ 23	12	5131	15.87	37.24
		C/G			91	0.28	0.66
二核苷酸 Di-nucleotide	12	AC/GT	12 ~ 24	15	548	1.70	3.98
		AG/CT			4765	14.74	34.58
		AT/AT			745	2.30	5.41
		CG/CG			18	0.06	0.13
三核苷酸 Tri-nucleotide	60	AAG/CTT	15 ~ 36	17	683	2.11	4.96
		ACC/GGT			271	0.84	1.97
		AGC/CTG			250	0.77	1.81
		ATC/ATG			296	0.92	2.15
		其他类型 Other types			802	2.48	5.82
四核苷酸 Tetra-nucleotide	64	AAAC/GTTT	20 ~ 36	21	16	0.05	0.12
		AAAG/CTTT			22	0.07	0.16
		AAAT/ATTT			29	0.09	0.21
		ACAT/ATGT			16	0.05	0.12
		其他类型 Other types			43	0.13	0.31
五核苷酸 Penta-nucleotide	20	所有类型 All types	25 ~ 35	26	20	0.06	0.15
六核苷酸 Hexa-nucleotide	33	所有类型 All types	30 ~ 48	35	33	0.10	0.24
总数 Total	193				13779	42.63	100

2.2.2 长柄双花木转录组中 SSR 位点分布特征

长柄双花木转录组中基序的重复次数分布统计结果见图 1。由图 1 可知,SSR 位点数目与基序重复次数有关。随着基序重复次数增加,SSR 位点数目逐渐减少(图 1A)。转录组中基序的重复次数主要在 5 ~ 15 次之间,SSR 位点 13128 个,占总数

95.27%。其中,低重复次数(5 ~ 10)的 SSR 位点 10267 个(74.51%),较低重复次数(11 ~ 15)的 SSR 位点 2861 个(20.76%)。中度重复次数(16 ~ 20)的 SSR 位点 588 个(4.27%),较高重复次数(>20)的 SSR 位点 63 个(0.46%)。

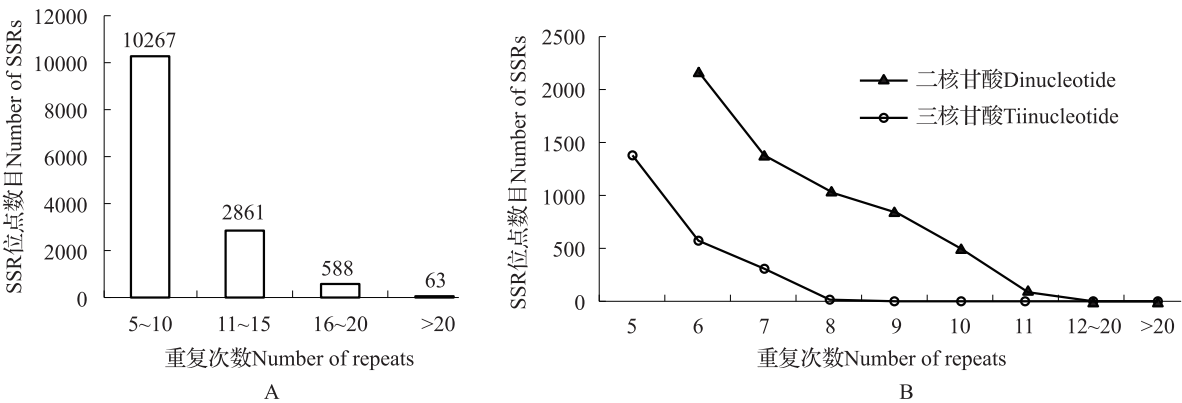


图 1 基元重复次数分布图

Fig. 1 Distribution and number of repeats in motif

长柄双花木转录组中,二、三核苷酸重复基序的 SSR 位点数目,随着基序重复次数的增加而减少,且基序重复数少的 SSR 位点数目占主导(图 1B)。二核苷酸重复基序的 SSR 位点数目,普遍高于三核苷酸重复基序。二核苷酸重复基序中,基序重复 6 次的 SSR 位点数目最多(2164 个),其次是重复 7 次(1402 个),重复次数 11~20 次的 SSR 位点数目为 112 个;三核苷酸重复基序中,重复 5 次的 SSR 位点数最多(1383 个),其次是重复 6 次(585 个),重复 11~20 次的 SSR 位点数目为 3

次。二、三核苷酸重复基序中,重复次数较高(>20)时,没有 SSR 位点分布。

长柄双花木转录组中重复基序长度分布见表 2。由表 2 可知,基序长度主要集中在 10~21 bp,分布有 11686 个 SSR 位点,占 SSR 位点总数 91.36%。其中,基序长度 12 bp,SSR 位点数目最多(2365 个),占总位点数 18.49%;其次是基序长度 10 bp(1548 个),占 12.10%;再次是基序长度 14 bp(1465 个),占 11.45%。基序长度大于 21 bp,SSR 位点数目占总数比例均低于 1.00%。

表 2 重复基序长度与 SSR 位点分布

Table 2 Bioinformatic prediction of SSR motifs

基序长度(bp)	数量	比例(%)	基序长度(bp)	数量	比例(%)	基序长度(bp)	数量	比例(%)
Motif length	Number	Proportion	Motif length	Number	Proportion	Motif length	Number	Proportion
10	1548	12.10	24	64	0.50	38	15	0.12
11	915	7.15	25	24	0.19	39	19	0.15
12	2365	18.49	26	12	0.09	40	9	0.07
13	370	2.89	27	11	0.09	41	9	0.07
14	1465	11.45	28	9	0.07	42	16	0.13
15	1429	11.17	29	4	0.03	43	8	0.06
16	1084	8.47	30	24	0.19	44	15	0.15
17	138	1.08	31	13	0.10	45	10	0.08
18	1374	10.74	32	17	0.13	46	8	0.06
19	71	0.56	33	17	0.13	47	11	0.09
20	610	4.77	34	15	0.12	48	9	0.07
21	317	2.48	35	13	0.10	49	12	0.09
22	122	0.95	36	30	0.23	50	10	0.08
23	22	0.17	37	16	0.13	51	7	0.05

2.3 长柄双花木转录组 SSR 引物位点多态性

遗传标记是否具有多态性是决定其实用性的基本要求。为了验证开发的 SSR 引物的实用性,从批量开发的 SSR 引物中随机选取 60 对引物对来自 8 个种群 8 株随机样本进行 PCR 扩增。0.8% 琼脂糖凝胶电泳检测的结果表明,46 对引物可重复扩增出清晰的目的条带,扩增率为 76.67%。扩增出目的条带的 PCR 产物进行毛细管自动荧光电泳系统检测的结果表明,16 对引物扩增产物的峰值完全一样,30 对引物的扩增产物显示有多态性(表 3),多态位点百分比为 65.22%。没有扩增出目的条带的引物,可能与扩增片段中有内含子或设计的引物特异性有关。遗传标记揭示的多态性位点数目及多态性高低可能与采集的样本容量以及材料是否存在亲缘关系有关。随着种群数目和样本容量增大,检测

到的多态性位点数目可能会增加,多态位点百分比也可能会有变化。与转录组数据进行比对发现,多态性 SSR 引物位点中,7 个 SSR 位点位于 3'非编码区,11 个 SSR 位点位于 5'非编码区,6 个位于编码区,6 个位置尚未确定。

多态性引物 DC101 扩增产物的琼脂糖凝胶电泳和毛细管自动荧光电泳系统检测、分型的结果见图 2。由图 2 可知,琼脂糖凝胶电泳(图 2A)仅能检测到 PCR 扩增的目的条带,很难人工分辨出扩增片段大小的差异;毛细管自动荧光电泳系统的分型结果(图 2B)不仅能区分 PCR 扩增产物片段的大小,而且能直观地将样本区分为纯合体(217/217 bp)和杂合体(214/226 bp)以及 SSR 位点上的等位基因数目。

表 3 SSR 引物及重复类型

Table 3 Information of SSR primers

引物 Primer	引物序列 Primer sequence(5'-3')		基序 Motif	产物长度(bp) Product length	位置 Position	等位基因数量 Number of alleles
DC55	F:GAATGAGGCTCTCCGATTT	R:CGAAAACGAACGATGACTGA	(TC) ₁₀	204	5'-UTR	3
DC57	F:GCTACAACATTTTCAGCGCA	R:CTCTCCTTCCCAAGCTTCT	(GT) ₁₀	220	3'-UTR	3
DC60	F:TGGGTTTTCTGAAATGGGTG	R:CTGATGCACTATCCCCACCT	(GA) ₁₀	265	5'-UTR	4
DC67	F:ATATTGCTGTTTCAGGCCGAC	R:AGTTGCCCTTTAGCTCTCCA	(AG) ₉	169	5'-UTR	3
DC68	F:TTGGAGTGTGGCTAGCAATG	R:ACGCTTCACTTGAGTTGGCT	(AG) ₉	173	undetermined	2
DC70	F:GACACAGAGAGGCAATCGG	R:AGCAGAAGCTTGAGGGTGG	(AG) ₉	178	CDS	4
DC71	F:TCTCTTGGTTTGGTGAAGGG	R:GCTGAAAACAGTGCAGGTGA	(AT) ₉	186	5'-UTR	4
DC73	F:CTTGTTGCTGGCACATTTTG	R:CTACTACCCTCTCCCCACC	(CA) ₉	200	undetermined	6
DC74	F:CCCCGAATGAACAAAGAGAAR	R:TCCACAACAGCCCTAAATCTT	(TA) ₉	200	3'-UTR	4
DC79	F:GCTTTTGCGCCAGTTTAAAG	R:CTCTTCGCTGCTGAATACCC	(GA) ₉	217	5'-UTR	3
DC82	F:AGCAGTTTTGCTCTTGGGA	R:AAAAAGGGGAATCAACCCAG	(AC) ₉	241	3'-UTR	2
DC85	F:TCTGGCATCTTTGATTGTGC	R:AAACACACCCCTGTCTGCTC	(GA) ₉	245	CDS	2
DC87	F:GGGACAGAGGTGTAATGCGC	R:GGACAGAGAGATGAGGCG	(GA) ₉	258	5'-UTR	5
DC89	F:TACCGATCTGCAACAGTGG	R:AGATCACAGTCCCAAGCACC	(TC) ₉	267	undetermined	2
DC90	F:TTGGGACTGAAATGGGAGAG	R:TATTTTCTTTCCCAACCCCT	(GA) ₉	271	5'-UTR	2
DC91	F:ATCTGTTTGAGCCCCCTTTT	R:ATTGCTCGAGGGAGAAACAA	(AT) ₉	272	undetermined	4
DC93	F:ATCCCCACCAAAATCAAAAT	R:ATCAAAAACGCAAAATGGAAC	(CT) ₉	275	5'-UTR	3
DC95	F:GCATAAGCTGGACCCATTTT	R:GCATTTCTTTTGTGATGCT	(AC) ₉	276	5'-UTR	3
DC96	F:TGATTCCGATTACCGAAGAGA	R:CTGCGATATCGTTTGGGAGT	(AT) ₉	277	5'-UTR	5
DC101	F:CATCTGGCAGTTGGACTTGA	R:CGCTCCAATACTGCAAGTGA	(TAG) ₁₂	202	3'-UTR	8
DC102	F:AAACACTTGGCGTGATTTCC	R:ACCGGAAGGAGAAAAAGAA	(ATC) ₁₁	186	CDS	3
DC104	F:TGTCAAACCTTTTGGATGCG	R:GGAGAAGGTTATGGATTGGGA	(ATT) ₈	272	3'-UTR	2
DC111	F:TTCGAGAGTTTGAATGGGG	R:CCAAGCACTTCAATGCCTTT	(ATA) ₅ (AGA) ₆	233	CDS	2
DC115	F:TGCCGATCATGAACAAAAAG	R:TTTCTCAGCAACCAACGGT	(CG) ₆ (CA) ₁₀	170	undetermined	8
DC13	F:ACTTCAAACCCAAGTCACCG	R:GTTACAAGAGAAGGCGGCAG	(CAGGCT) ₈	121	CDS	2
DC14	F:GCAGAACTGCCTCCTCATCT	R:ACCGGAAGCAGCTGAATAGA	(TC) ₇ G(CA) ₁₀	222	3'-UTR	2
DC22	F:CTTCTTGCTGACCATGGGAT	R:GTGGCTGATTGCTGTGAAGA	(GCATG) ₆	228	3'-UTR	2
DC24	F:AGCCACTGTGTCAGCAAATG	R:AGCATGTCAACAACACTTGAAA	(GGAA) ₈	196	5'-UTR	5
DC40	F:TAGGATGTTGGGTGCATTTG	R:GGTGCCATGTAAGCACACAC	(CTTT) ₅	253	undetermined	2
DC47	F:TCAATGGGAAAGCCTACTGG	R:TCGCGCTTTTGAAGAGTTT	(CACCAA) ₈	233	CDS	3

5'-UTR 指 SSR 位点位于 5'非编码区,3'-UTR 指 SSR 位点位于 3'非编码区,CDS 指 SSR 位点位于编码区,undetermined 指 SSR 位点位置尚未确定

5'-UTR refers to the SSR loci situated 5'-Untranslated region,3'-UTR refers to the SSR loci in the 3'-Untranslated region,CDS refers to the SSR loci in the coding sequence,undetermined refers to the SSR loci position has not been determined yet

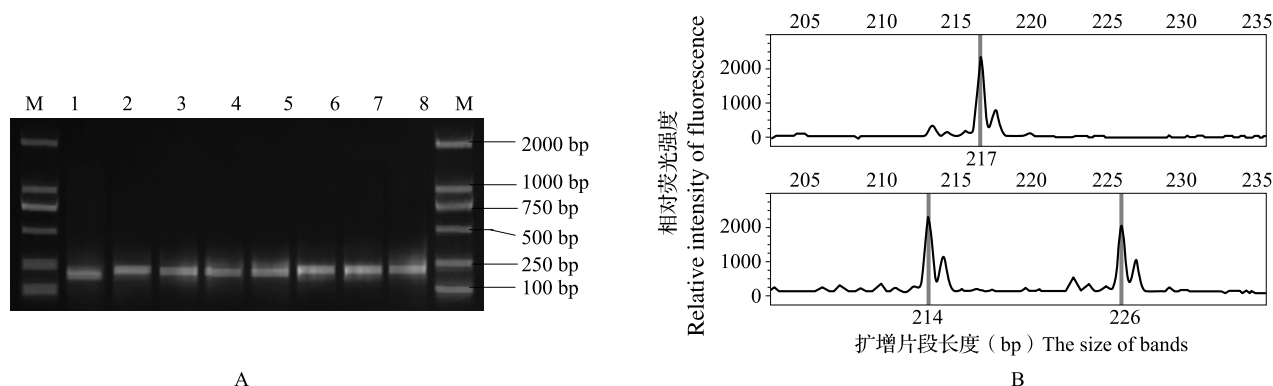


图 2 引物 DC101 在长柄双花木样本中的 PCR 扩增电泳图

Fig. 2 Electrophoresis of PCR fragments by primer DC101 in *D. cercidifolius* var. *longipes* H. T. Chang

3 讨论

对长柄双花木 32325 条 Unigens 进行 SSR 搜索,筛选出 13779 个 SSR 位点,SSR 位点发生频率为 42.63%,显著高于濒危植物南方红豆杉 *Taxus chinensis* var. *mairei* (Lemée & H. Lév.) W. C. Cheng & L. K. Fu (2.24%)^[16],稍高于杜仲 *Eucommia ulmoides* Oliv. (33.99%)^[17]。长柄双花木转录组中 SSR 分布频率为 1/2.95 kb,高于银杏 *Ginkgo biloba* L. (1/12.02 kb)^[18]、北美鹅掌楸 *Liriodendron tulipifera* L. (1/8.50 kb)^[19]、茶树 *Camellia sinensis* (L.) Kuntze (1/3.68 kb)^[20],低于橡胶树 *Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg (1/0.92 kb)^[21],说明长柄双花木转录组中 SSR 位点十分丰富。不同物种间转录组中 SSR 位点发生频率和分布频率存在较大差异,可能与物种本身特性、转录组测序方法、数据量大小和数据处理方法有较大的关系^[22]。

长柄双花木转录组中二核苷酸重复基序类型 GA/CT 4765 个,占 SSR 标记数目的比例最大(见表 1)。这在木本植物红松(*Pinus koraiensis* Siebo. & Zucc.)^[23]、沙棘(*Hippophae rhamnoides* L.)^[24]、刺梨(*Rosa roxburghii* Tratt.)^[25],以及大麦(*Hordeum vulgare* L.)、玉米(*Zea mays* L.)、水稻(*Oryza sativa* L.)、高粱(*Sorghum bicolor* (L.) Moench)、小麦(*Triticum aestivum* L.)等主要农作物^[26]研究中,也出现类似结论。转录组中三核苷酸重复基序 AAG/CTT 最丰富(表 1),与大多数双子叶植物^[27-29]、裸子植物南方红豆杉^[16]、杉木(*Cunninghamia lanceolata* (Lamb.) Hook.)^[30]的研究结果相一致。基因组中占主导地位的重复杂序,可能与其被翻译成多种氨基酸以及该物种中由这些氨基酸合成的蛋白质多肽链含量较高有关。对于二核苷酸重复基序,可能是由于阅读框改变,产生滑脱链误配(slipped strand mispairing)的差错,GA/CT 基序可以代表 mRNA 中 GAG、AGA、UCU 和 CUC 密码子,分别翻译为精氨酸、谷氨酸、丝氨酸和亮氨酸,导致 GA/CT 重复基序在转录组及 EST 文库中比例较高^[26]。三核苷酸重复基序 AAG/CTT,代表 mRNA 中 AAG、CUU 密码子,分别翻译为谷酰胺酸和亮氨酸,这暗示着长柄双花木细胞组织结构中,含有谷酰胺酸和亮氨酸的多肽链蛋白质比例较高。

一般认为,在 SSR 序列复制的过程中 DNA 聚合酶的滑动以及酶对被滑动链的错误修复,是 SSR 序列长度变异的主要机制,SSR 突变率取决于重复

基序类型、次数及基序长度等内在特征。其中,重复基序重复次数对 SSR 突变有负效应^[31]。有研究发现,高级重复基序(四、五、六核苷酸)的 SSR 位点多态性低于二核苷酸重复基序类型,多态性 SSR 引物一般为基序长度较长的二、三核苷酸重复基序类型^[32]。长柄双花木转录组中,随着基序重复次数增加,SSR 位点数目逐渐减少,高级重复基序的 SSR 位点数量仅占总数 1.31%(表 1),证实了前人的研究结论。考虑到重复基序长度大于 20 bp,可能引起独特的突变动力学,产生较高突变率^[31],本研究筛选的多态性 SSR 引物主要是考虑基序长度大于 18 bp 的二、三核苷酸重复基序类型(表 2)。

毛细管自动荧光电泳系统仅标记带尾巴的荧光引物,与荧光标记 SSR 引物相比,SSR 标记技术与毛细管自动荧光电泳检测技术相结合的 TP-M13-SSR 技术,成本低廉得多^[33];与传统的高密度琼脂糖凝胶电泳、聚丙烯酰胺凝胶电泳方法的 SSR 标记技术相比,TP-M13-SSR 技术的毛细管自动荧光电泳系统能够辨别出仅有 2 个 bp 之差的 SSR 扩增片段^[34],利用识别分析软件自动化记录电泳数据,克服了传统人工判读条带准确性不好的不足,极大地降低了系统误差,为 SSR 标记分析提供了高效、精确、灵敏的工具^[35]。目前,TP-M13-SSR 技术多应用于植物种质鉴定及遗传多样性研究^[36-37],但用于 SSR 引物筛选的报道较少。本研究采用 TP-M13-SSR 技术对随机选取的 60 对 SSR 引物进行验证的结果表明,30 个 SSR 位点显示有多态性,多态位点百分比为 65.22%。由此可见,基于转录组序列开发长柄双花木 SSR 标记引物是一种切实可行、高效率的方法。本研究结果极大地丰富了长柄双花木可用的 SSR 标记数量,采用的 TP-M13-SSR 技术能够检测出供试材料较高的多态性,适合于进一步开展长柄双花木种群遗传学分析、系统进化研究。

参考文献

- [1] 高浦新,李美琼,周赛霞,刘洁,牛艳丽,杜娟,丁剑敏. 濒危植物长柄双花木(*Disanthus cercidifolius* var. *longipes*)的资源分布及濒危现状. 植物科学学报,2013,31(1):34-41
- [2] 缪绅裕,曾庆昌,陈志明,刘嘉俊,李晓杰,黄金玲. 南岭大东山长柄双花木群落物种组成与种群结构特征分析. 植物资源与环境学报,2014,23(1):51-57
- [3] 傅立国. 中国珍稀濒危植物. 上海:上海科技出版社,1989:179
- [4] 肖宜安,何平,邓洪平,时明芝. 井冈山长柄双花木种群遗传多样性与遗传分化. 西南师范大学学报:自然科学版,2003,28(3):444-449
- [5] 谢国文,赵俊杰,李荣华,李海生,郭培国,陈雅丽,黄爱平. 南岭国家濒危植物长柄双花木遗传多样性研究. 广州大学

- 学报:自然科学版,2014(4):47-52
- [6] McGregor C E, Lambert C A, Greyling M M, Louw J H, Warnich L. A comparative assessment of dna fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica*, 2000, 113(2):135-144
- [7] Kalia R K, Rai M K, Kalia S, Singh R, Dhawan A K. Microsatellite markers; an overview of the recent progress in plants. *Euphytica*, 2011, 177(3):309-334
- [8] Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 2008, 27(4):617-631
- [9] Gao P X, Yang A H, Yao X H, Huang H. Isolation and characterization of nine polymorphic microsatellite loci in the endangered shrub *Disanthus cericidifolius* var. *longipes* (Hamamelidaceae). *Molecular Ecology Resources*, 2009, 9(3):1047-1049
- [10] Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation; A review. *Molecular Ecology*, 2002, 11(1):1-16
- [11] Wang Z, Gerstein M, Snyder M. RNA-Seq: are evolutionary tool for transcriptomics. *Nature Reviews Genetics*, 2009, 10(1):57
- [12] Davey J W, Hohenlohe P A, Etter P D, Boone J Q, Catchen J M, Blaxter M L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 2011, 12(7):499-510
- [13] Marioni J C, Mason C E, Mane S M, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 2008, 18(9):1509
- [14] Mortazavi A, Williams B A, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 2008, 5(7):621
- [15] 李金璐, 王硕, 于婧, 王玲, 周世良. 一种改良的植物 DNA 提取方法. *植物学报*, 2013, 48(1):72-78
- [16] 李炎林, 杨星星, 张家银, 黄三文, 熊兴耀. 南方红豆杉转录组 SSR 挖掘及分子标记的研究. *园艺学报*, 2014, 41(4):735-745
- [17] 林开勤, 李岩, 赵德刚. 杜仲转录组 SSR 发掘及标记开发. *分子植物育种*, 2016, 14(6):1548-1558
- [18] 樊洪泓, 李廷春, 李正鹏, 林毅, 蔡永萍. 银杏 EST 序列中微卫星的分布特征. *基因组学与应用生物学*, 2009, 28(5):869-873
- [19] Xu M, Sun Y, Li H. EST-SSRs development and paternity analysis for *Liriodendron* spp. *New Forests*, 2010, 40(3):361-382
- [20] 杨华, 陈琪, 韦朝领, 史成颖, 方从兵, 宛晓春. 茶树转录组中 SSR 位点的信息分析. *安徽农业大学学报*, 2011, 38(6):882-886
- [21] 甘霖, 覃碧, 刘实忠, 王秀全, 范高俊. 巴西橡胶树转录组中 SSR 位点的信息分析. *广东农业科学*, 2014, 41(16):142-146
- [22] Varshney, Rajeev K, Graner, Andreas, Sorrells, Mark E. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, 2005, 23(1):48-55
- [23] 张振, 张含国, 莫迟, 张磊. 红松转录组 SSR 分析及 EST-SSR 标记开发. *林业科学*, 2015, 51(8):114-120
- [24] Li H, Ruan C J, Wang L, Ding J, Tian X J. Development of RNA-Seq SSR Markers and Application to Genetic Relationship Analysis among Sea Buckthorn Germplasm. *Journal of the American Society for Horticultural Science*, 2017, 142(3):200-208
- [25] 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. *园艺学报*, 2015, 42(2):341-349
- [26] Kantety R V, Rota M L, Matthews D E, Sorrells M E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology*, 2002, 48(5-6):501-510
- [27] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, 2002, 30:194-200
- [28] 时小东, 朱学慧, 盛玉珍, 庄国庆, 陈放. 基于转录组序列的楠木 SSR 分子标记开发. *林业科学*, 2016, 52(11):71-78
- [29] 李响, 杨楠, 赵凯歌, 陈玉星, 唐锐君, 陈龙清. 蜡梅转录组 EST-SSR 标记开发与引物筛选. *北京林业大学学报*, 2013, 35(51):25-32
- [30] 文亚峰, 韩文军, 周宏, 徐刚标. 杉木转录组 SSR 挖掘及 EST-SSR 标记规模化开发. *林业科学*, 2015, 51(11):40-49
- [31] Kelkar Y D, Tyekucheva S, Chiaromonte F, Makova K D. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research*, 2008, 18(1):30
- [32] Dreisigacker S, Zhang P, Warburton M L, Van Ginkel M, Hoisington D, Bohn M, Melchinger A E. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different mega environments. *Crop Science*, 2004, 44(2):381-388
- [33] Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 2000, 18(2):233
- [34] 李莉, 杨剑波, Mackill D J, Colowit P M. 应用自动荧光测序系统分析不同水稻基因型 SSLPs. *作物学报*, 2000, 26(5):565-569
- [35] 刘志斋, 王天宇, 黎裕. TP-M13-SSR 技术及其在玉米遗传多样性研究中的应用. *玉米科学*, 2007(6):10-15, 31
- [36] 高源, 刘凤之, 王昆, 王大江, 龚欣, 刘立军. 基于 TP-M13-SSR 指纹图谱的中国原产苹果属植物分子身份证的建立. *植物遗传资源学报*, 2015, 16(6):1290-1297
- [37] Cao Y, Tian L, Gao Y, Liu F. Genetic diversity of cultivated and wild Ussurian Pear (*Pyrus ussuriensis* Maxim.) in China evaluated with M13-tailed SSR markers. *Genetic Resources & Crop Evolution*, 2012, 59(1):9-17