

榧树转录组 SSR 信息分析及其分子标记开发

张敏^{1,2}, 周彩红¹, 陈焘¹, 戴文圣¹, 张迟^{1,2*}

(¹浙江农林大学·省部共建亚热带森林培育国家重点实验室, 浙江临安 311300;

²景宁嘉树农业科技开发有限公司, 浙江景宁 323500)

摘要:【目的】基于‘细榧’(*Torreya grandis* ‘Xifei’)和‘圆榧’(*T. grandis* ‘Dielsii’)种子转录组数据, 开发 SSR 分子标记, 为榧属植物遗传多样性研究奠定基础。【方法】利用 MISA 软件对筛选得到的 1 kb 以上的 Unigene 做 SSR 分析, 利用 Primer 3.0 设计引物, 随机选择 49 对引物进行多态性扩增。【结果】在 22 976 条评估数目中, 包含 SSR 位点 5 458 个, 共鉴定出 6 种类型的 SSR。单核苷酸重复、三核苷酸重复和二核苷酸重复分别占 SSR 总数的 64.9%、18.56% 和 14.77%。对 5 458 个 SSR 位点设计出 4 633 对 SSR 引物, 随机选择 49 对在 10 份榧属植物中进行多态性扩增, 其中 16 对引物表现出多态性, 各个位点的等位基因数为 2~6 个, 平均等位基因数 3 个, 多态信息含量 PIC、观测杂合度 H_o 、期望杂合度 H_e 的平均值分别是 0.464 4、0.283 7 和 0.500 6。【结论】榧树转录组测序数据能够提供丰富的 SSR 位点, 用于快速高效地开发 SSR 引物。所开发的引物能为榧树遗传多样性研究、品种指纹图谱构建和分子标记辅助选择育种提供标记选择。

关键词: 榧树; 转录组; SSR

中图分类号: S664.5

文献标志码: A

文章编号: 1009-9980(2017)10-1258-08

Analysis of SSR information in transcriptome and development of molecular markers in *Torreya grandis*

ZHANG Min^{1,2}, ZHOU Caihong¹, CHEN Tao¹, DAI Wensheng¹, ZHANG Chi^{1,2*}

(¹State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; ²Jingning Jiashu Agricultural Science and Technology Development Company, Jingning 323500, Zhejiang, China)

Abstract: 【Objective】The genus *Torreya* is an important member of Taxaceae with only 9 species distributed in China, the United States of American, Japan and Korea. *T. grandis* Fort. ex Lindl. has been famous of its nut seeds with high oil for economic usage and as food for thousands years in China. Breeding of *T. grandis* is currently limited on selection of the wild resources derived from spontaneous mutation and occasional seedling. The genetic characteristics of the wild resources are great concern for the efficiency of improvement of *T. grandis*. To get insight of hereditary information for commercial utilization and cultivar discrimination of genetic resource of this species, microsatellite loci were exploited based on compared transcriptomic profiles de novo assembled for RNA-seq data of *T. grandis*. 【Methods】Seeds of *T. grandis* ‘Xifei’ and *T. grandis* ‘Dielsii’ were sampled from a orchard (Fuyang, Zhejiang province, China). Based on the biochemistry analysis of fatty acids biosynthesis, seeds at initiation stage (449 days after pollination, DAP), fastigium stage (469 DAP) and steady stage (512 DAP) were separately mixed as sample pools for RNA extraction; followed by comparative transcriptomic analysis through RNA-Seq. De novo transcriptome assembly was achieved using Trinity. Microsatellite mining was performed using the software MicroSATellite (MISA). Primers (20 base pairs) were designed using Primer 3 were used. 49 pairs of primers were randomly selected and were used for morphological amplification with product sizes ranging from 100 to 300 bp. Ten individuals of each species, including *T. nucifera*, *T. fargesii*, *T. jackii*, *T. grandis* and *T.*

收稿日期: 2017-02-13 接受日期: 2017-06-26

基金项目: 林业公益性行业科研专项(201504708); 浙江省团队科技特派员服务计划项目(浙科发农[2013]215-122); 浙江省科技计划项目(2012C22095); 浙江省农业新品种选育重大科技专项(2016C02052-12)

作者简介: 张敏, 博士, 研究方向为经济林遗传育种。Tel: 0571-63732764, E-mail: mzhang@zafu.edu.cn

*通信作者 Author for correspondence. Tel: 0571-63732764, E-mail: zhangchi1978@zafu.edu.cn

grandis var. *jiulongshanensis*, were selected for primary assess on the microsatellite polymorphism. Genomic DNA was extracted from young leaves by the cetyltrimethylammonium bromide (CTAB) modified method. 【Results】All high quality reads were assembled into 142 213 unigene. A total of 5 458 SSRs containing inserts with various microsatellite motifs (di- to penta nucleotides) were identified from 22 976 unigenes (> 1 kb). The distribution density was 105.4 SSRs per Mb. Mono nucleotide was the most abundant type of repeat motif with a frequency of 64.95%, followed by the tri-nucleotide (18.56%), di (14.77%), tetra (0.97%), hexa (0.42%) and pentanucleotides (0.33%). The AT/AT was the most frequent repeat motif (3.85%), followed by TA/TA (3.57%) and AG/CT (2.29%); Among the tri-nucleotide repeats, the dominant repeat motif in EST-SSRs was GAA/TTC (2.36%), followed by CTT/AAG (1.32%) and AAT/ATT (1.26%). The frequency of tetra- and hexanucleotide were much lower than that of di- and tri-nucleotide. Numbers of each repetitive unit ranged from 5 to 21. Units repeated less than or equal to 10 times and greater than 10 times were predominantly at mono nucleotide type with 1 990 (36.46%) and 1 555 (28.49%), and followed by 5 (12.90%), 6 (11.95%), 7 (5.42%) and 8 (2.53%) times among di- to hexanucleotide type. Repeat number was mainly 6, 7 and 8 in Di nucleotide unit as well as 5, 6 and 7 in tri-nucleotide unit. A total of 4 633 prime pairs were designed, among them 49 primer pairs were selected randomly to test in *Torreya*. From these, 16 polymorphic SSRs were characterized. The number of alleles of each locus ranged from 2 to 6, which was averagely 3 per locus. The average value of polymorphism information content (PIC), observed heterozygosity (*Ho*) and expected heterozygosity (*He*) were 0.464 4, 0.283 7 and 0.500 6 respectively. 【Conclusion】The transcriptome sequencing data of *Torreya* can provide abundant SSR locus, which can be used into develop SSR primer rapidly. The primer developed can provide marker selection for *Torreya* genetic diversity, the construction of fingerprint pattern of variety and molecular mark assisted breeding.

Key words: *Torreya grandis*; RNA-Seq; SSR

榧属 (*Torreya*) 于 1938 年创立^[1], 佛罗里达榧 (*Torreya taxifolia* Arn) 为该属的模式植物。据文献记载, 现存榧属植物仅有 7 个种 2 个变种, 间断分布于东亚和北美洲。其中, 佛罗里达榧 (*T. taxifolia*) 和加州榧 (*T. californica* Torrey) 产于北美, 日本榧 [*T. nucifera* (L.) Sieb. et Zucc.] 野生分布于日本和朝鲜半岛, 我国部分城市作为庭院树引种栽培。中国自然分布有巴山榧 (*T. fargesii* Franch.)、长叶榧 (*T. jackii* Chun)、榧树 (*T. grandis* Fort.ex Lind)、四川榧 (*T. parvifolia* Yi, L Yang et Long)^[2]、云南榧 (*T. fargesii* Franch. var. *yunnanensis*) 和九龙山榧 (*T. grandis* Fort.ex Lind. var. *jiulongshanensis*)。还有一种榧属植物 *T. clarnensis* 已灭绝, 在美国俄勒冈州中北部地区发现了它的种子化石, 这批种子化石位于始新世中期的黑硅石中^[3]。

榧树 (*T. grandis*) 通常雌雄异株, 群体内变异复杂, 籽形从圆到长, 从大到小呈正态分布^[4]。会稽山区的榧树类型最为丰富, 不同类型的民间称谓有: 细榧、芝麻榧、米榧、茄榧、獠牙榧、旋纹榧、大圆榧 (炭髻榧)、圆榧、小圆榧、寸金榧、羊角榧、长榧和冲杠榧

等。在安徽省黄山地区有神仙榧、花生榧、和尚榧、糯米榧、转筋榧、苹果榧、牛卵子榧、圆榧、小圆榧和米榧等品种类型的称谓。2011 年, ‘细榧’ (*T. grandis* ‘Xifei’) 通过国家林业局林木品种审定委员会审定, 良种编号: 国 S-SV-TG-024-2011。近年来, 浙江省林木品种审定委员会审(认)定了 ‘珍珠榧’ ‘大长榧’ ‘象牙榧’ ‘东白珠’ ‘脆仁榧’ ‘朱岩榧’ ‘丁山榧’ ‘东榧 1 号’ ‘东榧 2 号’ ‘东榧 3 号’ 和 ‘大叶种细榧’ 等品种。随着榧树新品种的不断选育和香榧产业的发展壮大, 榧树各居群的实生种质需要进行遗传多样性评价, 新品种也急需通过构建 DNA 指纹图谱来加以鉴定和保护知识产权。

简单序列重复 (simple sequence repeat, SSR) 又称微卫星序列, 是广泛存在于真核和原核生物基因组中的遗传标记, 具有共显性、覆盖性广、信息量大、操作简单等优势^[5], SSR 标记已广泛应用于油茶^[6]、陆地棉^[7]、油棕^[8]、花生^[9]、梨^[10]、平榛^[11]等植物的遗传多样性评价和品种指纹图谱构建。传统的 SSR 引物主要从基因组文库中获得, 其步骤复杂、工作量大, 开发成本高。近年来, 高通量测序技术迅猛发展, EST-

SSR是ESTs测序计划的副产物,开发成本较低,共显性遗传,多态性高,重复性及稳定性好,是目前最适合用于种质鉴定分析的标记之一。在榿属植物分子标记开发方面,金则新等^[12]建立了长叶榿ISSR-PCR的反应体系,梁丹等^[13]建立了香榿AFLP试验体系并利用AFLP进行了榿树雌雄株的鉴定,刘浩凯等^[14]利用SRAP标记研究了雌性榿树4个居群的遗传多样性,戴正等^[15]用RAPD和SCAR标记鉴定了香榿幼苗的性别,张党权等^[16]利用RAPD将榿树的几个栽培品种进行了分类。Yi等^[17]基于1棵香榿树的叶片转录组数据,从4 600条Unigene中获得1 713个SSR位点,设计了108对引物,其中37对引物有扩增产物,19对引物在供试样品中检测到多态性。

笔者基于‘细榿’和‘圆榿’种子转录组数据,对获得的SSR数据进行分析,筛选出扩增条带清晰且多态性高的SSR引物,为榿属植物遗传多样性分析、品种指纹图谱构建和分子标记辅助选择育种奠定基础。

1 材料和方法

1.1 转录组数据来源

转录组数据来源于本课题组2014年对‘细榿’和‘圆榿’种仁进行Illumina高通量测序的结果。测序时采取3个发育时期的种仁:种子脂肪酸开始合成期(授粉后第449天)、脂肪酸合成高峰期(授粉后第469天)、脂肪酸合成平缓期(授粉后第512天),提取RNA后委托北京百迈克生物科技公司完成RNA-Seq测序,并通过De Novo方法组装。得到142 213条Unigene,作为背景数据。

1.2 植物材料及其DNA提取

用于SSR引物筛选和可用性评价的材料采自本课题组收集保存的榿属5个种的10个单株,分别是细榿、榿树雄株、大圆榿、九龙山榿雌株、九龙山榿雄株、长叶榿雌株、长叶榿雄株、巴山榿雌株、巴山榿雄株、日本榿雌株。每个单株采集新鲜的叶片置于冰盒中,立即带回实验室,采用改良CTAB法^[18]提取基因组总DNA,最后将所提取的DNA稀释成 $50 \text{ mg} \cdot \text{L}^{-1}$, $-4 \text{ }^{\circ}\text{C}$ 保存备用。

1.3 转录组SSR位点鉴别及SSR引物设计

转录组数据采用Trinity软件组装,获得转录本序列。组装共得到246 862条Transcript。利用MISA程序(MicroSATellite identification tool, <http://pgrc.ipk-gatersleben.de/misa/misa.html>)对筛选得到的1

kb以上的Unigene做SSR位点搜索。

用Primer3程序对含有SSR位点的Unigene序列设计引物。引物设计的主要原则:引物长度控制在20 bp;退火温度 $58 \sim 62 \text{ }^{\circ}\text{C}$;上下游引物的 T_m 值相差不大于 $2 \text{ }^{\circ}\text{C}$;PCR产物长度为 $100 \sim 300 \text{ bp}$;GC含量为 $40\% \sim 60\%$;每个位点产生3对候选引物。

1.4 EST-SSR引物筛选

随机选取49对EST-SSR引物交由生工生物工程(上海)股份有限公司合成。选取榿树、巴山榿、长叶榿、日本榿进行引物筛选。PCR扩增体系为 $10 \mu\text{L}$, $10 \times \text{Buffer}(\text{Mg}^{2+}) 1 \mu\text{L}$, $\text{dNTP mix } 0.8 \mu\text{L}$,上下游引物各 $0.5 \mu\text{L}$, $rTaq$ 酶 $0.05 \mu\text{L}$,模板 $0.5 \mu\text{L}$, $\text{ddH}_2\text{O } 6.65 \mu\text{L}$,混合均匀后置于PCR仪中扩增。扩增程序为: $95 \text{ }^{\circ}\text{C}$ 预变性5 min;然后进行34次循环,每个循环包括 $94 \text{ }^{\circ}\text{C}$ 变性45 s, $54 \sim 58 \text{ }^{\circ}\text{C}$ (退火温度因不同引物而异)退火30 s, $72 \text{ }^{\circ}\text{C}$ 延伸1 min;最后 $72 \text{ }^{\circ}\text{C}$ 延伸3 min, $4 \text{ }^{\circ}\text{C}$ 保存。

PCR产物用1.2%的琼脂糖凝胶电泳检测,选取扩增出单一明亮条带的或2条明亮带的引物保留,用于多态性分析。

初次筛选好的PCR产物在8%的非变性聚丙烯酰胺凝胶中电泳分离。电泳缓冲液为 $1 \times \text{TBE}$,电压 180 V ,电泳90 min。聚丙烯酰胺凝胶在银染液和显色液作用下显现条带,在可见光灯箱上观察并拍照。

1.5 数据统计

用SSR出现频率和SSR平均分布距离来描述EST-SSR。计算公式为:(1)SSR出现频率, $f_c/\% = c/n \times 100$, c 为搜索到的SSR数量, n 为无冗余EST数量;(2)SSR平均分布距离, $f_N = N/c$, N 为无冗余EST数量的总碱基数。

聚丙烯酰胺凝胶电泳结果采用人工读带的方式,相同迁移的位置上,有带记为1,无带记为0,缺失记为9,统计扩增结果,通过参考Smith等^[19]的方法计算多态信息含量(PIC)值。应用POPGENE Version 1.31软件,对标记数据进行遗传参数分析。

2 结果与分析

2.1 榿树转录组中SSR的分布及结构特点

‘细榿’和‘圆榿’种子转录组测序共获得142 213条Unigenes,其中大于1 kb的Unigene为22 976条(序列总长约 $51\,784\,615 \text{ bp}$),利用MISA软

件对筛选得到的1 kb以上的Unigene做SSR分析,发现其中4 548条Unigene序列中包含5 458个SSR位点,其中包含2个或2个以上SSR位点的序列有175条,占3.8%。总体上,SSR发生频率为23.76%,平均每9.49 kb出现1个SSR。SSR的类型十分丰富,单核苷酸至六核苷酸重复类型均存在。单核苷酸所占比例最大,占64.94%,其次二、三核苷酸重复出现频率占主要优势,分别占总SSR的14.77%和18.56%,四、五、六核苷酸重复类型比较少,分别占

0.97%、0.33%、0.42%。

转录组SSR重复单元的重复次数为5~21次,单核苷酸重复单元重复次数在10次以内(包括10次)的有1 990个,占36.46%;重复次数大于10次的有1 555个位点,占28.49%。二核苷酸至六核苷酸重复单元中,5、6次重复所占比例最大,分别占12.90%、11.95%;其次是7次与8次,占总SSR的5.42%和2.53%。通过分析,发现二核苷酸重复的次数主要为6次、7次、8次;而三核苷酸的重复次数主要为5次、6次、7次(表1)。

表 1 榿树种子转录组中不同 EST-SSR 的数量、类型以及频率

Table 1 The number of different EST-SSR types and frequency in *T. grandis* seeds transcriptome

重复单元类型 Repeat motif type	重复次数 Repeat number							总计 Total	百分比 Percentage/%
	5	6	7	8	9	10	>10		
单核苷酸 Mononucleotide	0	0	0	0	0	1 990	1 555	3 545	64.95
二核苷酸 Dinucleotide	0	373	185	127	83	29	9	806	14.77
三核苷酸 Trinucleotide	635	261	106	11	0	0	0	1 013	18.56
四核苷酸 Tetranucleotide	42	10	1	0	0	0	0	53	0.97
五核苷酸 Pentanucleotide	17	0	0	0	0	0	1	18	0.33
六核苷酸 Hexanucleotide	10	8	4	0	0	1	0	23	0.42
总计 Total	704	652	296	138	83	2 020	1 565	5 458	100.00
百分比 Percentage/%	12.90	11.95	5.42	2.53	1.52	37.01	28.67	100	

2.2 转录组SSR基序重复类型和频率特征

榿树种子转录组的SSR核苷酸基序类型十分丰富(表2),二核苷酸主要以AT/AT、TA/TA、AG/CT为主,其中AT/AT所占比例最大,为3.85%,TA/TA占3.57%,

AG/CT占2.29%;三核苷酸重复的主要类型较二核苷酸的类型多,以GAA/TTC、CTT/AAG、AAT/ATT为主,GAA/TTC在三核苷酸基序中占2.36%,其次是CTT/AAG(1.32%);四、五核苷酸基序类型较二、三核苷酸基

表 2 榿树种子 EST 中二核苷酸和三核苷酸重复基元的类型及频率

Table 2 Dinucleotide and trinucleotide EST-SSR repeat motifs and their frequency in ESTs of *T. grandis* seeds

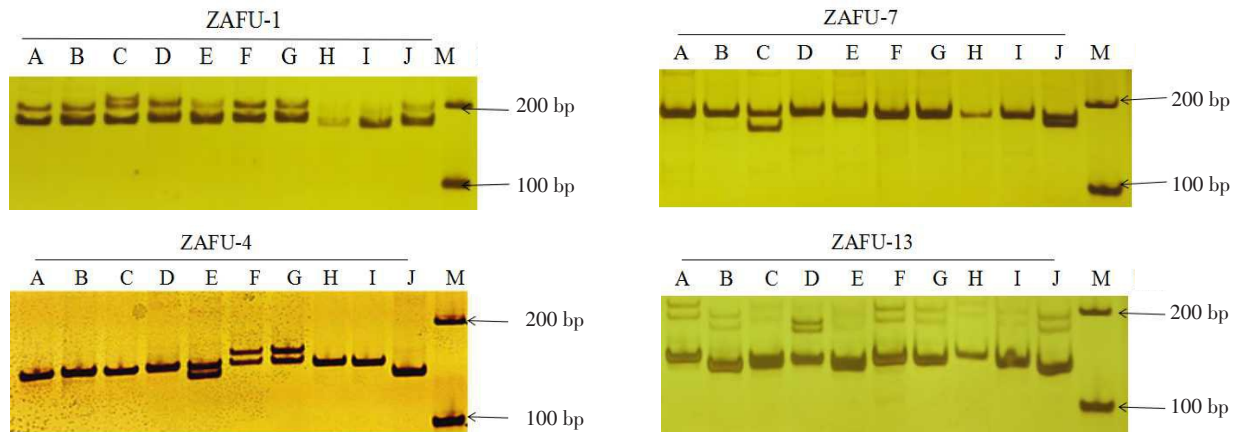
重复单元类型 Repeat motif type	重复次数 Repeat number							总计 Total	百分比 Percentage/%	重复单元类型 Repeat motif type	重复次数 Repeat number							总计 Total	百分比 Percentage/%
	5	6	7	8	9	10	>10				5	6	7	8	9	10	>10		
AC/GT	29	23	14	9	1	1	77	1.41	AAC/GGT	2	2	1					5	0.09	
AG/CT	64	32	11	10	4	4	125	2.29	AAT/ATT	47	16	6					69	1.26	
AT/AT	98	35	46	23	6	2	210	3.85	ACA/TGT	7	3	1					11	0.20	
CA/TG	54	24	10	8	3	1	100	1.83	ACC/GGT	5	4						9	0.16	
GA/TC	46	21	13	11	7		98	1.80	ACT/AGT	2							2	0.04	
TA/TA	82	49	33	22	8	1	195	3.57	AGA/TCT	41	13	6					60	1.10	
GC/GC			1				1	0.02	AGC/GCT	34	9	5					48	0.88	
CAT/ATG	26	11	7				44	0.81	ATA/TAT	21	10	4					35	0.64	
CCA/TGG	13	9	3				25	0.46	ATC/GAT	21	8	2					31	0.57	
CCG/CGG	10	1	2	1			14	0.26	CAA/TTG	22	4	2					28	0.51	
CCT/AGG	27	16	5	1			49	0.90	CAC/GTG	10	5	1	1				17	0.31	
CGA/TCG	5	2					7	0.13	GCA/TGC	39	13	4	1				57	1.04	
CGC/GCG	6	3		1			10	0.18	GCC/GGC	14	5	5					24	0.44	
CGG/CCG	5		1				6	0.11	GGA/TCC	30	22	13	1				66	1.21	
CGT/ACG	1						1	0.02	GTA/TAC	2	2						4	0.07	
CTA/TAG	1						1	0.02	GTT/AAC		1						1	0.02	
CTC/GAG	35	11	6				52	0.95	TAA/TTA	19	10	6	1				36	0.66	
CTG/CAG	36	14	10	1			61	1.12	TGA/TGA	22	7	4					33	0.60	
CTT/AAG	33	32	6	1			72	1.32	总计 Total	635	634	291	138	83	29	9	1 819	33.32	
GAA/TTC	96	27	5	1			129	2.36	百分比	11.63	11.62	5.33	2.53	1.52	0.53	0.16	33.32		
GAC/GTC	3	1	1	1			6	0.11	Percentage/%										

序类型少, AAGG/CCTT、AGGA/TCCT、TGCCC/GG-GCA、AAATT/AATTT是主要的几个类型,但所占比重低。

2.3 榧树种子 SSR 引物的有效性及多态性表现

随机选取 49 对引物,其中包括二核苷酸、三核苷酸、四核苷酸、五核苷酸及六核苷酸等 6 种重复基元的 SSR 位点,进行引物有效性验证。

以榧属不同种的 10 个样品进行 PCR 扩增,筛选。结果显示,其中 35 对引物能扩增出比较理想的产物,有效扩增率为 71.43%,同时也说明 35 对引物具有良好的通用性,能在榧属的部分种间进行扩增。对榧属的 5 个种进行多态性分析,共筛选出多态引物 16 对,多态率为 32.65%。说明利用榧树种子 EST 序列开发的 SSR 引物是高效可行的。图 1 为部



M. Marker; A. 细榧雌株; B. 榧树雄株; C. 大圆榧雌株; D. 九龙山榧雌株; E. 九龙山榧雄株; F. 长叶榧雌株; G. 长叶榧雄株; H. 巴山榧雌株; I. 巴山榧雄株; J. 日本榧雌株。

M. Marker; A. *T. grandis* 'Xifei' (female plant); B. *T. grandis* (male plant); C. *T. grandis* 'Majus' (female plant); D. *T. grandis* var. *jiulongshanensis* (female plant); E. *T. grandis* var. *jiulongshanensis* (male plant); F. *T. jackii* (female plant); G. *T. jackii* (male plant); H. *T. fargesii* (female plant); I. *T. fargesii* (male plant); J. *T. nucifera* (female plant).

图 1 部分位点在 10 个榧属个体中的多态性

Fig. 1 The polymorphism about some loci in 10 *Torreya* individuals

分引物在 10 份榧属样中的扩增情况。

2.4 SSR 引物评价

利用验证的 35 对引物对 10 份榧属植物进行扩增及多态性的评价,16 对 SSR 引物在榧属中表现出多态性,共获得了 48 个等位基因,平均等位基因数 3 个,其中 ZAFU-1, ZAFU-4 和 ZAFU-5 的等位基因数分别为 6、5、5 个。观望杂合度均值与期望杂合度均值分别是 0.283 7 和 0.500 6。PIC 值是衡量群体变异的一个重要的参数,16 个位点的 PIC 值为 0.105 0~0.957 5,平均值为 0.464 4,其中高度多态的位点有 8 个 ($PIC \geq 0.5$)^[20],表 3 为 16 个位点的多态性扩增情况。

3 讨 论

随着高通量测序技术的迅猛发展,通过转录组测序来开发 EST-SSR 引物的成本大为降低,对于缺乏基因组信息的物种,该技术提供了前所未有的机遇,已在植物中得到了广泛的应用。笔者对 2 个品

种 3 个不同发育时期的榧树种子进行转录组测序,组装共得到 246 863 条 Transcript 和 142 213 条 Unigene。利用 MISA 程序对筛选得到的 1 kb 以上的 Unigene 做 SSR 位点搜索,共获得 5 458 个 SSR 位点。榧树 SSR 的发生频率为 23.76%,比梨高 16.56%^[21],比云南松高 20.69%^[22],比金线莲高 11.54%^[23],但比萝卜低 0.03%^[24]。Yi 等^[17]基于 1 株榧树的叶片转录组数据,组装得到 46 000 条 Unigene,搜索获得 1 713 个 SSR 位点,与本研究结果存在着差异。这可能是 SSR 的发生频率与转录组测序所选取的植物器官、数据分析工具的差异存在着一定的关系,在枇杷^[25-26]SSR 的研究中也曾有类似报道。

榧树转录组中的 SSR 存在一至六核苷酸的重复类型,其中一至三核苷酸重复为主要的重复类型,占 SSR 总数的 98.23%,四至六核苷酸重复类型所占比例较低,仅占总 SSR 数的 1.77%,在红松^[27]、刺梨^[28]和桑树^[29]的研究中也存在类似的结果。另外,分析榧

表3 榿属16个EST-SSRs标记的特征
Table 3 The characteristics of 16 EST-SSRs markers in *Torreya*

引物编号 Primer No.	SSR 基序重复 Repeat motif	引物序列 Primer sequence (5'-3')	温度 Temperature/°C	预期扩增片 段长度 Expected size/bp	实际扩增 片段长度 Actual size/bp	等位 基因数 <i>Na</i>	观测 杂合度 <i>Ho</i>	期望 杂合度 <i>He</i>	多态信 息含量 PIC
ZAFU-1	(TG)9	F:GGCTATGCTACACCCAAAGAA R:GGGGCACCACCTATTGTATG	59	167	160~200	6	0.100 0	0.836 8	0.795 0
ZAFU-2	(TG)9	F:TCAAAGTGCAACCGGTACAA R:CAACAGGCCAACATGGAGTA	59	118	110~140	2	0.111 1	0.111 1	0.105 0
ZAFU-3	(CAG)8	F:GGGTTACCCCTTGCTTTAT R:CCCTACTTTATTTCGGTGCG	59	254	250~280	3	0.200 0	0.415 8	0.957 5
ZAFU-4	(TAA)8	F:GAATTCCCATTCCCATTGTG R:ACCCCTTCTGCTCTGATTT	59	152	140~170	5	0.300 0	0.763 2	0.172 5
ZAFU-5	(TTC)8	F:AATGAATGCCGTGTTACGCTG R:TTGGAGCGGAAGGAATAATG	59	174	170~200	5	0.600 0	0.789 5	0.667 5
ZAFU-6	(TTCT)5	F:CCAATTTGTGGAGCGTTTCT R:TGTGGAAAGGTGGTGAACAA	59	192	190~210	2	0.000 0	0.442 1	0.380 0
ZAFU-7	(TATT)5	F:TTTTCCAACCTCAACCCCTTG R:ATGTTTGGGGTTGACGTGTT	59	171	170~190	2	0.000 0	0.189 5	0.159 6
ZAFU-8	(TAGCA)5	F:AATTGGCCCTTCATTCAACA R:CTAGTGGGTGCATTTGAGCC	60	229	250~700	3	0.571 4	0.615 4	0.591 9
ZAFU-9	(GAGTC)5	F:GGAGCGAGCGTAGCGTATAG R:GTTGCTGCATCTTCCTCCTC	59	214	210~230	3	0.125 0	0.508 3	0.511 7
ZAFU-10	(TC)9	F:GCCTGGTCTGCTTCTTGTTC R:AGGCAAAGGCTGCAATAGAA	59	161	160~200	2	0.000 0	0.263 7	0.209 8
ZAFU-11	(AT)8	F:ACATCTGCAAGGCAAGGTTTC R:TTGAATTTTACCAGGCTCC	59	174	170~180	2	0.000 0	0.303 0	0.277 8
ZAFU-12	(TA)9	F:TGCTAATGTTTGCTTTGAATGG R:AACAACCTAAACTTTTGCCGAA	60	153	150~180	3	0.300 0	0.689 5	0.717 5
ZAFU-13	(ACC)6	F:GAAATTAACCGGTTACAGGG R:CATTGCTGTCCGTCTCGTAA	59	133	130~150	2	0.500 0	0.394 7	0.375 0
ZAFU-14	(ATTG)5	F:CACAATCCTCCATTTCATCCA R:GCCACTGGTGTGATTCTCAA	59	208	205~215	2	0.285 7	0.439 6	0.345 7
ZAFU-15	(GGACG)5	F:ATTTGATACGACGGAAACGG R:GGTGGTCAATATCCCATGCT	59	118	110~130	3	0.571 4	0.538 5	0.500 0
ZAFU-16	(TGAGCC)7	F:AAGGTTGCCACCTCAGTCAC R:ACAGAACGTCTCCAACCGAC	60	246	220~260	3	0.875 0	0.708 3	0.664 1
平均值	Average					3	0.283 7	0.500 6	0.464 4

榿树转录组数据发现,除单核苷酸重复类型外,三核苷酸重复类型所占比重最大,达18.56%。在榿树的研究中,研究者也发现三核苷酸重复所占比重大,可能是因为密码子的选择作用^[30]。密码子是三核苷酸为一个功能单位,三核苷酸的位移对于一个基因的表达不会有太大的影响,自然选择机制对于三核苷酸基序具有积极的选择作用^[31]。因此,三核苷酸重复类型被认为是常见转录组SSR重复基序类型。

转录组SSR的核苷酸重复序列类型多样,在本研究中,分析二核苷酸重复和三核苷酸重复序列的碱基组成发现,二核苷酸重复序列主要以AT/AT、TA/TA为主,三核苷酸重复序列以GAA/TTC、CTT/AAG、AAT/ATT为主。A/T优势基元在榿树核苷酸重复序列中有较好的呈现,这与打破AT键的能量低

于GC键,AT的波动较GC容易有关^[32]。另外,榿树中出现的高频重复基元和油桐^[33]、南方红豆杉^[34]中出现的高频重复基元基本吻合。

本研究开发的16对引物用于5种榿属植物的多态性分析结果表明,平均等位基因数为3.4,多态信息含量(PIC)平均值为0.464 4。PIC>0.5为高度多态位点,0.25<PIC<0.5为中度多态位点,PIC<0.25为低度多态性位点^[20-21]。本研究所选的16对SSR引物多态性较好,为榿树遗传多样性研究、品种指纹图谱构建和分子标记辅助选择育种奠定了基础。

参考文献 **References:**

[1] 黎章矩,戴文圣. 中国香榿[M]. 北京:科学出版社,2007:188.
LI Zhangju, DAI Wensheng. *Torreya grandis* in China [M]. Bei-

- jing: Science Press, 2007: 188.
- [2] 易同培, 杨林, 隆廷伦. 榿属(红豆杉科)一新种——四川榿[J]. 植物研究, 2006, 26(5): 513-515.
YI Tongpei, YANG Lin, LONG Tinglun. *Torreya parvifolia*, a new species of the Taxaceae from Sichuan, China[J]. Bulletin of Botanical Research, 2006, 26(5): 513-515.
- [3] MANCHESTER S R. Fruits and seeds of the Middle Eocene nut beds flora, Clarno Formation, Oregon[M]. New York: Paleontological Research Institution, 1994: 1-205.
- [4] 程晓建, 黎章矩, 戴文圣, 喻卫武, 曾燕如. 榿树种质资源调查与评价[J]. 果树学报, 2009, 26(5): 654-658.
CHENG Xiaojian, LI Zhangju, DAI Wensheng, YU Weiwu, ZENG Yanru. Investigation and evaluation of *Torreya grandis* germplasm resource[J]. Journal of Fruit Science, 2009, 26(5): 654-658.
- [5] POWELL W, MORGANTE M, ANDRE C, HANAFEY M, VOGEL J, TINGEY S, RAFALSKI A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis[J]. Molecular Breeding, 1996, 2(3): 225-238.
- [6] 周文才, 温强, 杨军, 王建文, 徐立安, 徐林初. 油茶栽培品种 SSR 指纹图谱构建及聚类分析[J]. 分子植物育种, 2017, 15(1): 238-249.
ZHOU Wencai, WEN Qiang, YANG Jun, WANG Jianwen, XU Li'an, XU Linchu. Establishment of DNA fingerprints and cluster analysis for oil camellia cultivars based on SSR markers[J]. Molecular Plant Breeding, 2017, 15(1): 238-249.
- [7] 多力坤, 李雪源, 华金平. 新疆自育陆地棉品种 SSR 遗传多样性分析[J]. 作物学报, 2014, 40(2): 369-379.
DUO Likun, LI Xueyuan, HUA Jinping. Genetic diversity analysis on local upland cotton cultivars in Xinjiang based on SSR markers[J]. Acta Agronomica Sinica, 2014, 40(2): 369-379.
- [8] 周丽霞, 吴翼, 肖勇. 基于 SSR 分子标记的油棕遗传多样性分析[J]. 南方农业学报, 2017, 48(2): 216-221.
ZHOU Lixia, WU Yi, XIAO Yong. Genetic diversity of *elaeis guineensis* Jacq. based on SSR markers[J]. Journal of Southern Agriculture, 2017, 48(2): 216-221.
- [9] 任小平, 郑艳丽, 黄莉, 陈玉宁, 周小静, 陈伟刚, 雷永, 晏立英, 万丽云, 廖伯寿, 姜慧芳. 花生 SSR 核心引物筛选及育成品种 DNA 指纹图谱构建[J]. 中国油料作物学报, 2016, 38(5): 563-571.
REN Xiaoping, ZHENG Yanli, HUANG Li, CHEN Yuning, ZHOU Xiaojing, CHEN Weigang, LEI Yong, YAN Liying, WAN Liyun, LIAO Boshou, JIANG Huifang. Selection of core SSR markers and identification of fingerprint on peanut cultivars[J]. Chinese Journal of Oil Crop Sciences, 2016, 38(5): 563-571.
- [10] 薛华柏, 杨健, 王龙, 王苏珂, 张慧蓉, 乔玉山, 章镇, 李秀根. 29 个梨品种 SSR 特征指纹数据表的构建[J]. 果树学报, 2015, 32(6): 1028-1035.
XUE Huabai, YANG Jian, WANG Long, WANG Suke, ZHANG Huirong, QIAO Yushan, ZHANG Zhen, LI Xiugen. Construction of SSR characteristic fingerprinting data table for 29 pear cultivars [J]. Journal of Fruit Science, 2015, 32(6): 1028-1035.
- [11] 赵爽, 苏淑钗, 陈志钢, 盛淑艳, 王忠. 基于 SSR 的辽宁铁岭地区平榛遗传多样性与居群遗传结构分析[J]. 果树学报, 2016, 33(1): 24-33.
ZHAO Shuang, SU Shuchai, CHEN Zhigang, SHENG Shuyan, WANG Zhong. An assessment of the genetic diversity and population genetic structure concerning the *Corylus heterophylla* Fisch., grown in the Tieling district of Liaoning province, using SSR markers[J]. Journal of Fruit Science, 2016, 33(1): 24-33.
- [12] 金则新, 李钧敏. 濒危植物长叶榿 ISSR-PCR 反应体系的建立[J]. 西北林学院学报, 2006, 21(5): 94-97.
JIN Zexin, LI Junmin. Establishment of ISSR-PCR reaction system of endangered plant *Torreya jackii*[J]. Journal of Northwest Forestry University, 2006, 21(5): 94-97.
- [13] 梁丹, 吴勇, 曾燕如, 程晓建, 吴松, 童再康. 香榿 AFLP 实验体系的建立[J]. 福建林业科技, 2007, 34(2): 93-96.
LIANG Dan, WU Yong, ZENG Yanru, CHENG Xiaojian, WU Song, TONG Zaikang. Establishment of an AFLP system in *Torreya grandis*[J]. Journal of Fujian Forestry Science and Technology, 2007, 34(2): 93-96.
- [14] 刘浩凯, 胡树恒, 董雷鸣, 张辉, 曾燕如, 喻卫武, 戴文圣. 雌性榿树 4 个居群遗传多样的 SRAP 分析[J]. 经济林研究, 2014, 32(3): 30-35.
LIU Haokai, HU Shuheng, DONG Leiming, ZHANG Hui, ZENG Yanru, YU Weiwu, DAI Wensheng. SRAP analysis of genetic diversity of four female population *Torreya grandis*[J]. Nonwood Forest Research, 2014, 32(3): 30-35.
- [15] 戴正, 陈力耕, 童品璋. 香榿性别鉴定的 RAPD 和 SCAR 标记[J]. 果树学报, 2008, 25(6): 856-859.
DAI Zheng, CHEN Ligeng, TONG Pinzhang. RAPD and SCAR molecular markers for sex identification in *Torreya grandis*[J]. Journal of Fruit Science, 2008, 25(6): 856-859.
- [16] 张党权. 香榿主要栽培品种的 RAPD 分子鉴别[D]. 长沙: 中南林业科技大学, 2001.
ZHANG Dangquan. RAPD molecular identification of main cultivars of *Torreya grandis*[D]. Changsha: Central South University of Forestry and Technology, 2001.
- [17] YI G M, ZHAO T, YIN Z, TANG P, QIU Y J, YAO X H. Mining and characterizing EST-SSR markers from ESTs of *Torreya grandis*[J]. Biochemical Systematics and Ecology, 2016, 65: 124-128.
- [18] 沈登锋. 榿树种质资源的收集、评价与核心种质的确定[D]. 杭州: 浙江农林大学, 2011.
SHEN Dengfeng. The collection, evaluation and determination of core germplasm in *Torreya grandis*[D]. Hangzhou: Zhejiang A & F University, 2011.
- [19] SMITH J S C, CHIN E C L, SHU H, SMITH O S, WALL S J, SENIOR M L, MITCHELL S E, KRESOVICH S, ZIEGLE J. An eval-

- uation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparison with data from RFLPs and pedigree [J]. *Theoretical and Applied Genetics*, 1997, 95(1): 163–173.
- [20] NAGY S, POCZAI P, CERNAK I, GORJI A M, HEGEDUS G, TALLER J. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies [J]. *Biochemical Genetics*, 2012, 50(9): 670–672.
- [21] YUE X, LIU G, ZONG Y, TENG Y, CAI D. Development of genic SSR markers from transcriptome sequencing of pear buds [J]. *Journal of Zhejiang University—Science B (Biomedicine & Biotechnology)*, 2014, 15(4): 303–312.
- [22] 蔡年辉,许玉兰,徐杨,邓丽丽,周丽,王大玮,何承忠,段安安. 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.
- CAI Nianhui, XU Yulan, XU Yang, DENG Lili, ZHOU Li, WANG Dawei, HE Chengzhong, DUAN An'an. Distribution and Sequence characteristics of SSR in *Pinus yunnanensis* Franch. transcriptome[J]. *Journal of Yunnan University (Natural Sciences Edition)*, 2015, 37(5): 770–778.
- [23] 姜福星,魏丕伟,寇亚平,贾茵,陈其兵. 台湾金线莲转录组特性研究[J]. *分子植物育种*, 2015, 13(12): 2743–2753.
- JIANG Fuxing, WEI Piwei, KOU Yaping, JIA Yin, CHEN Qibing. Study on the characterization of the Transcriptome of *Anoectochilus formosanus* Taiwan[J]. *Molecular Plant Breeding*, 2015, 13(12): 2743–2753.
- [24] WANG S, WANG X, HE Q, LIU X, XU W, LI L, GAO J, WANG F. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish[J]. *Plant Cell Reports*, 2012, 31(8): 1437–1447.
- [25] 郑婷婷,魏伟淋,杨向晖,林顺权. 基于枇杷转录组序列的 SSR 分子标记引物开发[J]. *亚热带植物科学*, 2015, 44(4): 274–278.
- ZHENG Tingting, WEI Weilin, YANG Xianghui, LIN Shunquan. Development of SSR molecular markers based on transcriptome sequencing of *Eriobotrya japonica* [J]. *Subtropical Plant Science*, 2015, 44(4): 274–278.
- [26] 孙淑霞,陈栋,李靖,涂美艳,江国良. 枇杷转录组 EST-SSR 和 SNP 标记的发掘利用[J]. *园艺学报*, 2015, 42(s1): 2642.
- SUN Shuxia, CHEN Dong, LI Jing, TU Meiyao, JIANG Guoliang. The using of transcriptome SNP markers and EST-SSR in loquat [J]. *Acta Horticulturae Sinica*, 2015, 42(s1): 2642.
- [27] 张振,张含国,莫迟,张磊. 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.
- ZHANG Zhen, ZHANG Hanguo, MO Chi, ZHANG Lei. Transcriptome sequencing analysis and development of EST-SSR markers for *Pinus koraiensis* [J]. *Scientia Silvae Sinicae*, 2015, 51(8): 114–120.
- [28] 鄢秀芹,鲁敏,安华明. 刺梨转录组 SSR 及其分子标记开发[J]. *园艺学报*, 2015, 42(2): 341–349.
- YAN Xiuqin, LU Min, AN Huaming. Analysis on SSR information in transcriptome and development of molecular markers in *Rosa roxburghii* [J]. *Acta Horticulturae Sinica*, 2015, 42(2): 341–349.
- [29] 姚刘慧,唐翠明,戴凡炜,王振江,罗国庆. 基于桑树转录组测序的 SSR 标记开发和引物筛选[J]. *蚕业科学*, 2015, 41(1): 18–27.
- YAO Lihui, TANG Cuiming, DAI Fanwei, WANG Zhenjiang, LUO Guoqing. Development and primer screening of SSR markers based on mulberry transcriptome sequencing[J]. *Science of Sericulture*, 2015, 41(1): 18–27.
- [30] 李淑娴,张新叶,王英亚,尹佟明. 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.
- LI Shuxian, ZHANG Xinye, WANG Yingya, YIN Tongming. Content and characteristics of microsatellites detected in expressed sequence tag sequences in *Eucalyptus* [J]. *Chinese Bulletin of Botany*, 2010, 45(3): 363–371.
- [31] GARG R, PATEL R K, TYAGI A K, JAIN M. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification[J]. *DNA Research*, 2011, 18(1): 53–63.
- [32] 童治军,肖炳光. 3 种烟草基因组 SSR 位点信息分析和标记开发[J]. *西北植物学报*, 2014, 34(8): 1549–1558.
- TONG Zhijun, XIAO Bingguang. Survey of SSR loci information in three tobacco genomes and development of SSR markers[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2014, 34(8): 1549–1558.
- [33] 孙颖,谭晓风,罗敏,李建安. 油桐花芽 2 个不同发育时期转录组分析[J]. *林业科学*, 2014, 50(5): 70–74.
- SUN Ying, TAN Xiaofeng, LUO Min, LI Jian'an. The sequencing analysis of transcriptome of *Vernicia fordii* flower buds at two development stages[J]. *Scientia Silvae Sinicae*, 2014, 50(5): 70–74.
- [34] 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[J]. *园艺学报*, 2014, 41(4): 735–745.
- YAN Lin, YANG Xingxing, ZHANG Jiayin, HUANG Sanweng, XIONG Xingyao. Studies on SSR molecular markers based on transcriptome of *Taxus chinensis* var. *mairei* [J]. *Acta Horticulturae Sinica*, 2014, 41(4): 735–745.