

## 研究论文

## Research Article

# PCR 产物直接测序和混合克隆测序进行桉树 EST-SSR 标记开发

周长品<sup>✉</sup>, 李发根<sup>✉</sup>, 翁启杰<sup>✉</sup>, 于晓丽<sup>✉</sup>, 李梅<sup>✉</sup>, 甘四明<sup>✉</sup>

中国林业科学研究院热带林业研究所, 广州, 510520

✉ 通讯作者及电子邮件: Siming.Gan@ritf.ac.cn; ✉ 作者

分子植物育种, 2010 年, 第 8 卷, 第 1 篇 DOI: 10.5376/mpb.cn.2010.08.0001

收稿日期: 2010 年 7 月 12 日

接受日期: 2010 年 9 月 10 日

发表日期: 2010 年 9 月 12 日

这是一篇开放获取的论文, 其论文发布和传播接受《Creative Commons Attribution License》所有条款。只要对本原作有恰当的引用, 版权所有人允许和同意第三方无条件的使用、传播以及任何媒介的复制或再制作。

建议最佳引用格式:

周长品等, 2010, PCR 产物直接测序和混合克隆测序进行桉树 EST-SSR 标记开发, 分子植物育种 Vol 8 No 1 (DOI: 10.5376/mpb.cn.2010.08.0001)

**摘要** EST-SSR 标记开发中, 合适的测序方案是保障序列质量、控制实验成本和提高实验效率的关键。本研究利用 1 株细叶桉(P<sub>2</sub>)对 PCR 产物大小不同的 32 个桉树候选 EST-SSR 标记进行了 PCR 产物直接测序和混合克隆测序的比较。PCR 产物直接测序成功的标记有 21 个(65.6%), 但有 4 个标记因简单重复序列邻近测序引物未能确认重复单位数、8 个标记因在 P<sub>2</sub> 中等位片段长度杂合而在重复单位后面的测序峰图中显示双峰。在 8 个、16 个、24 个和全部 32 个 EST-SSR 对 P<sub>2</sub> 的 PCR 产物的混合克隆测序中, 测序成功的标记分别为 6 个(75.0%)、11 个(68.8%)、11 个(45.8%)和 14 个(43.8%)。综合考虑克隆转化与测序的效率、成本和时间, 推荐 16 个 PCR 产物为混合克隆测序的最佳数量。本研究中共有 27 个桉树 EST-SSR 标记成功测序, 与源 EST 比对的序列一致性介于 90.23% 和 100% 之间, 包括在 P<sub>2</sub> 中片段长度杂合的标记 10 个、纯合的 17 个; 在长度纯合的 17 个标记中, 有 8 个在 P<sub>2</sub> 中存在 1 个以上的个体内 SNP。

**关键词** 桉树, EST-SSR, 直接测序法, 混合克隆测序法

## Comparison between direct sequencing and pool-cloning-based sequencing of PCR products in EST-SSR marker development in *Eucalyptus*

Changpin Zhou<sup>✉</sup>, Fagen Li<sup>✉</sup>, Qijie Weng<sup>✉</sup>, Xiaoli Yu<sup>✉</sup>, Mei Li<sup>✉</sup>, Siming Gan<sup>✉</sup>

Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou, 510520, China

✉ Corresponding author, siming.gan@ritf.ac.cn; ✉ Authors

**Abstract** An appropriate sequencing protocol is critical in terms of sequence quality, cost effectiveness and efficiency. In the present study, a total of 32 candidate EST-SSR markers varying in PCR product size against an *Eucalyptus tereticornis* individual(P<sub>2</sub>) were used for PCR product direct sequencing and pool-cloning-based sequencing. In PCR product direct sequencing with P<sub>2</sub>, 21 (65.6%) markers were successfully sequenced. Four of them had only partial repeat motifs validated due to their closeness to the (forward) sequencing primer, and eight of them which were length-heterozygous showed double peaks behind the repeats in the sequencing traces. Pool-cloning of 8, 16, 24 and 32 EST-PCR products resulted in 6 (75.0%), 11 (68.8%), 11 (45.8%) and 14 (43.8%) markers reliably sequenced, respectively. Considering the efficiency as well as cost and time inputs, it is recommended that pool-cloning of 16 EST-PCR products is the best option. Totally 27 eucalypt EST-SSR markers were validated in this study, with an identity with the original EST between 90.23% and 100%, out of which 10 were length-heterozygous in P<sub>2</sub> and 17 were homozygous, including 8 each containing at least one within-individual SNP.

**Keywords** *Eucalyptus*, EST-SSR, Direct sequencing, Pool-cloning-based sequencing

## 研究背景

桉树是桃金娘科(Myrtaceae)桉属(*Eucalyptus*)、杯果木属(*Angophora*)和伞房桉属(*Corymbia*)树种的简称, 共有 808 种和 137 个变种(Hill and Johnson,

1995)。桉树的天然分布主要是澳大利亚、巴布亚新几内亚和印度尼西亚及其附近岛屿, 因其具有速生、丰产、适应性广和用途多样等特点, 在热带和亚热带地区广为引种, 全球人工桉树林面积已达 1

780 万公顷以上(FAO, 2000)。

简单序列重复(simple sequence repeats, SSR), 或称微卫星, 在基因组中数量众多, 作为分子标记具有重复性好、共显性和变异性高等特点(Tauz, 1989; Powell et al., 1996), 在生物学有关研究中应用广泛(方宣钧等, 2001)。目前, 桉树上已开发的 SSR 标记相对较少, 包括 367 个基因组 SSRs (Brondani et al., 2006)、35 个叶绿体 SSRs (Steane et al., 2005) 和 20 个表达序列标签(expressed sequence tag, EST) SSRs, 或称 EST-SSR(Faria et al., 2010)。其中, EST-SSR 因与基因相关, 在近缘物种中的通用性较高(Varshney et al., 2005; Ellis and Burke, 2007), 与功能基因的连锁不平衡也可能较强(Ayers et al., 1997)。公共数据库中日益增多的 EST 序列也为 EST-SSR 开发提供了便利。截止 2010 年 7 月 12 日 GenBank 中已有桉属 EST 序列 37 751 条(<http://www.ncbi.nlm.nih.gov/nucest/?term=Eucalyptus>)。分析表明, 桉属 EST 序列约 15% 含有 SSR (李淑娴等, 2010), 非冗余 EST 序列中 25% 以上含有 SSR (Rabello et al., 2005; Ceresini et al., 2005)。

虽然, 在许多物种中开展了 EST-SSR 标记开发研究, 但极少确认聚合酶链式反应(polymerase chain reaction, PCR)产物与源 EST 序列的一致性。据我们所知, 只有 3 篇报道进行了部分 EST-SSRs 的序列确认, 包括鹰嘴豆(*Cicer arietinum* L.) 60 个 EST-SSR 标记中的 5 个(Choudhary et al., 2009)、生菜(*Lactuca sativa* L.) 61 个标记中的 30 个(Simko, 2009)和茶树(*Camellia sinensis* L.) 61 个标记中的部分标记(Sharma et al., 2009)。由于引物的低特异性可能导致 PCR 错误扩增, 从而产生非目的片段或假阳性片段(Cha and Thilly, 1995; Kunkel and Bebenek, 2000; 张晓红等, 2009), 因此, 对于包括 SSR 在内的序列标签位点(sequence tagged site, STS)的序列确认是非常必要的。

PCR 产物一般可以通过两种方法测序, 即直接测序和克隆后的菌落测序。前者操作上相对快捷, 但当 PCR 产物不纯时测序结果较难判读(张晓红等, 2009); 后者针对单菌落测序, 模板单一, 可以获得理想的测序结果, 尤其是可以获得 PCR 产物的完整序列, 但需要进行克隆转化, 增加了实验的成本以

及工作量。因此, 如何选择合适的测序方案以保障测序质量、控制实验成本和提高实验效率, 是 EST-SSR 标记开发和其它 PCR 产物测序的一个关键问题。

本研究以 1 株细叶桉(*E. tereticornis* Smith)为材料, 比较了桉树 EST-SSR 标记开发中 PCR 产物直接测序和混合克隆测序的效果, 以期获得高效、经济的 EST-PCR 产物测序方案。同时, 开发了一批桉树 EST-SSR 标记, 有助于增加桉树的分子标记资源。

## 1 结果与分析

### 1.1 桉树 32 对 EST-SSR 引物的 PCR 扩增

为了在混合克隆后可以通过单菌落 PCR 产物查找长度对应的 EST-PCR 产物, 并且避免对插入片段相同、菌落 PCR 产物长度亦同的菌落的重复测序, 候选 32 对桉树 EST-SSR 引物对 1 株细叶桉(P<sub>2</sub>) DNA 扩增的产物大小尽量不同, 介于 120 bp 和 1 500 bp 之间(图 1)。PCR 产物基本都是单一、明亮的谱带, 少量 PCR 产物具有较弱的杂带(图 1 中 27、29 和 31 号)。另外, 32 对 EST-SSR 引物中, 有的 PCR 产物大小与设计的 EST 长度相符, 有的明显大于设计的 EST 长度(表 1)。

### 1.2 桉树 32 对 EST-SSR 引物的 PCR 产物的直接测序

32 个 EST-SSR 对 P<sub>2</sub> 的 PCR 产物直接测序中, 成功测序的有 21 个(表 1), 比例为 65.6%。其中, 4 个 EST-SSR (03、08、20 和 22 号)的简单重复序列邻近测序引物, 未能确认完整的重复单位数, 因毛细管测序中引物序列后一般有 30~40 bp 不能测出。成功测序的 21 个 EST-SSR 中, P<sub>2</sub> 的 2 个等位片段间在长度上纯合的标记有 13 个, 杂合的有 8 个(08、11、13、17、18、20、22 和 28 号), 其中 18 号的杂合性是由于单个碱基重复数的变异所致(表 1)。对于长度杂合的 SSR 标记, PCR 产物直接测序峰图中重复单位变异后出现双峰(单峰也是双峰叠加的结果)。图 2 和图 3A 分别显示了 P<sub>2</sub> 中等位片段长度纯合标记 EUCeSSR1135 (06 号)和杂合标记 EUCeSSR-0946 (13 号)的重复单位及其与源 EST 的对比。

另外, 片段长度纯合的 13 个标记中, 有 7 个(02、03、04、05、10、14 和 23 号)在 P<sub>2</sub> 的 2 个等位片段间存在至少 1 个单核苷酸多态性(single nucleotide polymorphism, SNP), 或称个体内 SNP。

### 1.3 EST-PCR 产物的混合克隆测序

在8个、16个、24个和全部32个EST-SSR对细叶桉 $P_2$ 扩增产物的混合克隆测序中, 测序成功的标记数分别为6个(75.0%)、11个(68.8%)、11个(45.8%)和14个(43.8%) (表1)。可见, 随着混合克隆的EST-PCR产物数量的增加, 虽然测序成功的标记数量呈增加的趋势, 但比例却逐渐下降。其中, 16个PCR产物混合克隆的成功测序的标记数量较大、比

例也较高。不同数量PCR产物混合克隆共成功测序19个标记(表1)。与PCR产物直接测序相比, 有6个标记(15、16、26、27、29和30号)只在混合克隆测序中被测到, 包括2个片段长度杂合标记(15和29号)和1个长度纯合、含有个体内SNP的标记(26号); 但有8个PCR产物直接测序成功的标记(5、9、10、11、14、17、18和19号)在混合克隆后没有被测到。

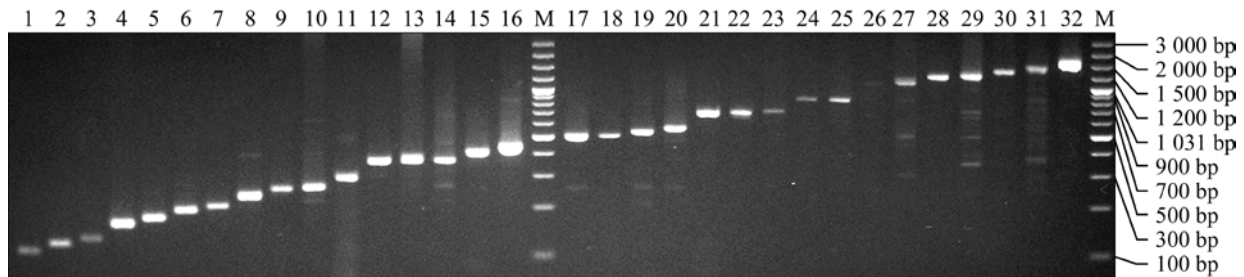
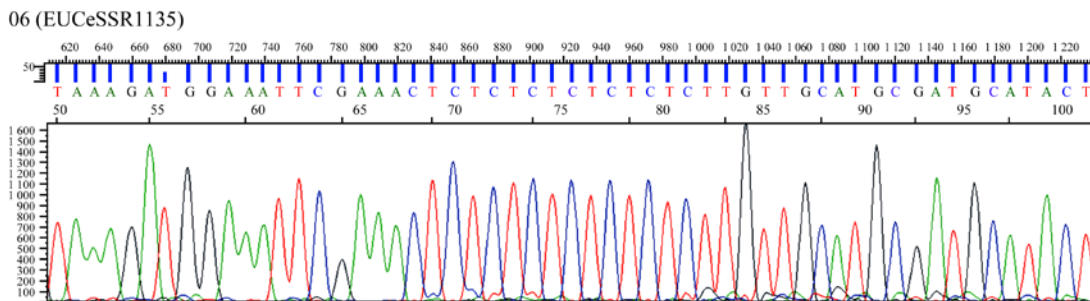


图1 桉树32对EST-SSR引物对1株细叶桉( $P_2$ )的PCR扩增产物

注: 01~32: EST-SSR引物(见表1); M: 100bp DNA ladder

Figure 1 PCR products amplified from the tested *Eucalyptus tereticornis* individual ( $P_2$ ) using 32 EST-SSR primer pairs

Note: 01~32: EST-SSR primer pairs (see Table 1); M: 100bp DNA ladder



Seq: 50 TAAAGATGGAAATTCGAAACCTCTCTCTCTCTCTCTCTCTGTGTGTCATGCATGCATACAT

EST: 235 TAAA GATGGAAATTCGAAA (CT)<sub>14</sub> TGTGTCATGCGATGCATACAT

图2 在 $P_2$ 中片段长度纯合的标记EUCeSSR1135 (06号)的部分序列及其与源EST的对比

注: 阴影部分的序列为EST-SSR的重复单位

Figure 2 Partial sequences of the length-homozygous EST-SSR marker EUCeSSR1135 (No. 06) within  $P_2$  and their comparison to the original EST

Note: The bases in grey shadow show the repeat motif of the EST-SSR marker within  $P_2$

混合克隆的单菌落测序的峰图比较“干净”(图3B和C), 有利于查找SSR的位置和数量; 但对于长度杂合的EST-SSR标记, 单菌落测序只能测到一个等位片段。本研究中, 由于有的杂合标记参加了不同数量的PCR产物混合克隆, 刚好不同的等位片

段均被测序, 如8、13、20、22和29号(表1)。

混合克隆测序中也发现了一些“误扩增”片段, 即引物序列同某一个EST, 但序列的长度与相应的PCR产物不同、且与源EST的序列一致性很低。该类片段可能为非特异性扩增所致, 因长度刚好类似



于另一个 EST-PCR 产物而被测序。

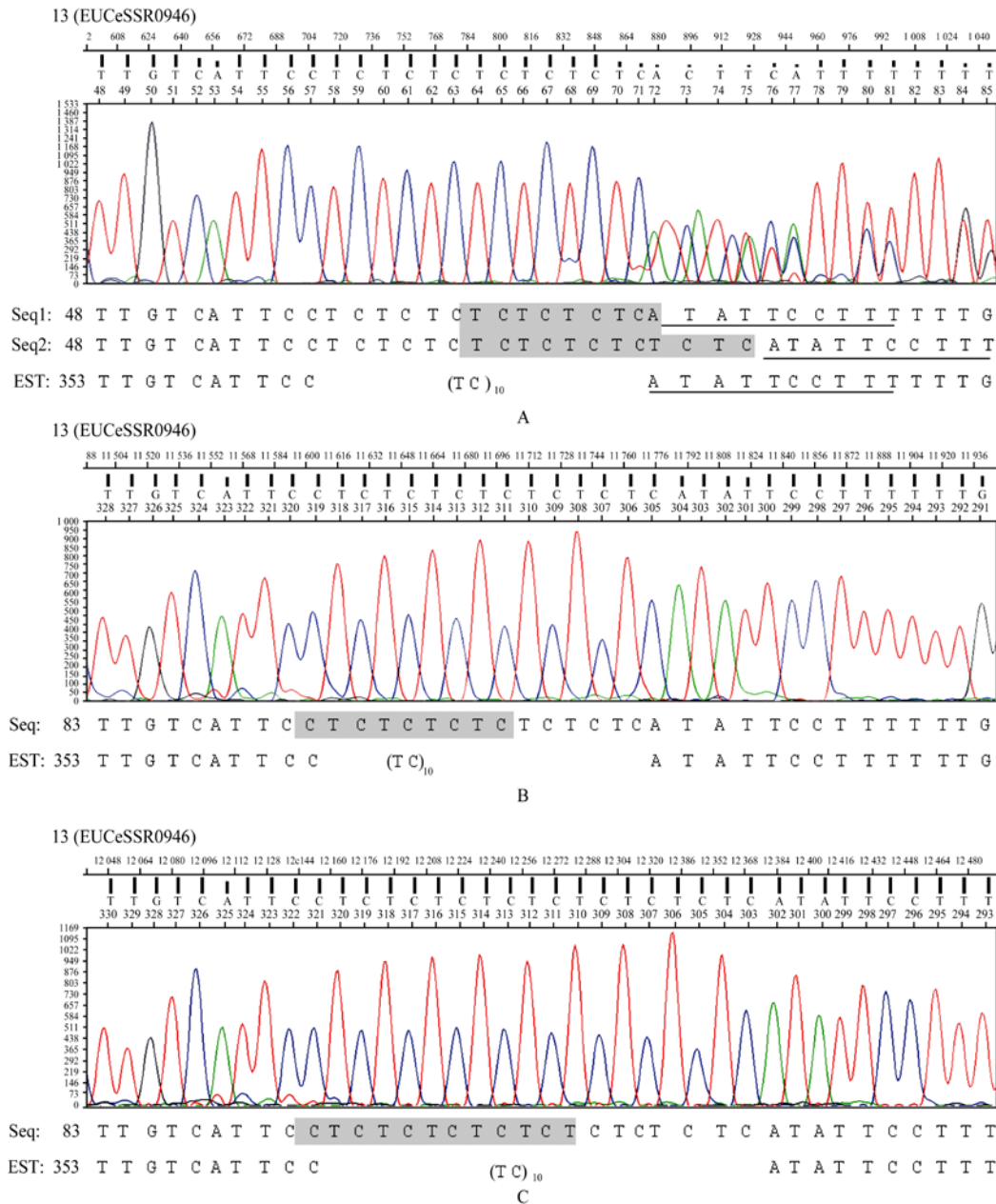


图 3 在 P<sub>2</sub> 中片段长度杂合的标记 EUCeSSR0946 (13 号)的部分序列及其与源 EST 的对比

注: A: PCR 产物直接测序; B: 16 个 PCR 产物混合克隆中的测序; C: 32 个 PCR 产物混合克隆中的测序; Seq1 对应的测序峰一般稍高于 Seq2; 灰色阴影部分的序列为 EST-SSR 的重复单位; 下划线表示重复单位后的对应序列

Figure 3 Partial sequences of the lengthheterozygous EST-SSR marker EUCeSSR0946 (No. 13) within P<sub>2</sub> and their comparison to the original EST

Note: A: Direct PCR product sequencing; B: pool cloning with PCR products of 16 EST-SSRs; C: Pool cloning with PCR products of 32 EST-SSRs; Seq1 corresponding mostly to the higher traces in the sequencing profile and Seq2 to the lower ones; The bases in grey shadow show the repeat motifs of the EST-SSR marker within P<sub>2</sub>, while those underlined in (A) show the identical bases following the repeats

#### 1.4 桉树 EST-SSR 标记的开发

结合对细叶桉 P<sub>2</sub> 的 PCR 产物直接测序和混合克隆测序, 共确认了 27 个桉树 EST-SSR 的序列, 与源 EST 比对的一致性介于 90.23% 和 100% 之间(表 1)。其中, 在 P<sub>2</sub> 中片段长度杂合的标记 10 个、纯合的 17 个; 在长度纯合的 17 个标记中, 有 8 个(02、03、04、05、10、14、23 和 26 号)存在 1 个以上的个体内 SNP。

## 2 讨论

目前, 新一代的测序技术已有广泛的应用(Mardis, 2008), 但传统的末端终止法仍是普遍使用的测序技术(周德贵等, 2008), 尤其是针对小规模的目的序列的测序, 如 EST 标记开发中的 EST-PCR 产物测序。因此, 如何经济、快捷和高效地进行末端终止法测序仍然具有极为重要的应用价值。

PCR 产物直接测序具有快速、简便和低成本的优点(徐祖元等, 2002; 张晓红等, 2009); 并且, 对于二倍体的 DNA 模板, 可以同时测出 2 个等位片段(虽然存在插入/缺失时后面测序峰会呈现双峰)。但是, 因技术限制, PCR 产物直接测序时前面 30~40 bp 左右往往遗失(苑克俊等, 2007), 这不利于完整序列的确认, 尤其是目的序列靠近测序引物位置时, 如本研究中 SSR 重复单位未能完整判读的 4 个标记; 并且, 当 DNA 模板的引物结合序列发生变异时, 引物特异性较差将导致非特异性 PCR 产物的产生, 测序的峰图较乱, 影响测序结果判读的准确性(张晓红等, 2009), 这也是本研究中尚有 11 个(34.4%)扩增较好的 EST-SSR 标记未能成功测序的主要原因。因此, PCR 产物直接测序主要应用于扩增特异性好、测序引物后面的几十个碱基不需完整确认的序列。

克隆测序可以克服 PCR 产物直接测序的不足, 获得完整序列(毛伟华, 2005)。但是, 以往的克隆测序多是针对单个 PCR 产物, 这在待测样品较多时需要花费大量的经费和人力进行克隆转化, 因此多个 PCR 产物混合克隆应是一个不错的选择。本研究中混合克隆测序的成功率为 43.8~75.0%, 并且有 6 个标记只在混合克隆中被成功测序, 证明了混合克隆测序的有效性。

比较 8 个、16 个、24 个和 32 个 EST-PCR 产物混合克隆后的测序成功率, 兼顾克隆转化与测序的

效率、成本和时间, 我们推荐 16 个 EST-PCR 产物混合克隆为最好的选择。混合克隆的 PCR 产物太少, 克隆转化的成本较高、耗时也较多。PCR 产物太多, 非特异性扩增的谱带数增加(虽然切胶回收法可以解决这一问题, 但增加成本和人力), 菌落 PCR 后通过片段长度查找对应的 PCR 产物时易与长度近似的目的片段混淆, 而造成测序的浪费, 如本研究 32 个 EST-PCR 产物的混合克隆中 32 次测序只确认了 14 个标记; 并且, PCR 产物增多, 产物间长度差别相对较小, 菌落 PCR 后通过琼脂糖凝胶电泳不易准确查找对应长度的 EST-PCR 产物(尤其对于长片段)。当然, 在实际实验中, 再挑取更多的菌落测序可以获得更多的标记, 但增加了菌落 PCR 与测序的工作量和成本。

目前, 已发表的桉树 EST-SSR 标记只有 20 个, 且未进行测序确认(Faria et al., 2010)。本研究共开发了与源 EST 序列一致性好的 27 个桉树 EST-SSR 标记(表 1), 这将显著增加桉树的分子标记资源。并且, 27 个标记中, 在 P<sub>2</sub> 中存在片段长度杂合的标记有 10 个、存在个体内 SNP 的标记有 8 个, 这些标记将在以 P<sub>2</sub> 为父本的遗传图谱构建群体中分离, 从而有效用于遗传图谱构建。这也表明在桉树中开发基于 EST 的分子标记具有极大潜力。

## 3 材料与方法

### 3.1 实验材料与 DNA 提取

1 株细叶桉(P<sub>2</sub>)是 Gan 等(2003)遗传图谱构建群体的父本。DNA 提取参考 CTAB 法(Doyle and Doyle, 1991), 稍做改进, 即在 CTAB 提取液中加入 5% 的聚乙烯吡咯烷酮(PVP)和 2% 的 β-巯基乙醇。pGEM®-T Easy 载体及 JM109 大肠杆菌感受态细胞购自 Promega 公司。

### 3.2 EST 序列、引物序列和 PCR

桉树 EST 序列下载自 GenBank (<http://www.ncbi.nlm.nih.gov/projects/dbEST/>), 截止 2009 年 1 月 2 日共有 35 643 条。利用 PHRAP 软件(<http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>)拼接后, 通过 MISA 软件(<http://pgrc.ipk-gatersleben.de/misa/>)查找简单序列重复, 标准为: 2、3 和 4 碱基重复的最

短长度为 12 bp, 5 和 6 碱基为 15 bp。包含 SSR 的序列用于 EST-SSR 引物设计, 软件为 Primer Premier 5.0 (<http://www.premierbiosoft.com/primerdesign/index.html>), 标准为: 引物长度 18~22 bp (最佳 20 bp),  $T_m$  55~65°C (最佳 60°C), GC 含量 40~60%, 目的片段长度 100~500 bp。引物合成委托上海 Invitrogen 公司完成。

利用细叶桉( $P_2$ ) DNA 进行 EST-SSR 引物的 PCR 条件优化, 主要针对 PCR 体系的  $Mg^{2+}$  浓度和 PCR 程序的退火温度。PCR 优化后, 选择 PCR 产物长度不同的 32 个 EST-SSR 进行测序实验, 引物序列和 PCR 产物大小见表 1。PCR 体系和 PCR 程序参考张晓红等(2009), 退火温度均为 56°C, 但 10×Buffer 中  $Mg^{2+}$  浓度改为 15 或 20 mmol/L。

### 3.3 PCR 产物纯化和直接测序

PCR 产物经乙醇-醋酸钠法纯化后直接测序(张晓红等, 2009), 测序一般利用前向引物, 对较长片段也利用后向引物测序。测序采用 20  $\mu$ L 体系, 参考 Bigdye Terminator V3.1 操作手册 (Applied Biosystems Co.), 但测序试剂 Bigdye 用量为 0.5  $\mu$ L。测序仪为 ABI 3130xl, 数据收集利用软件 Data Collection 2.0, 序列查看和分析利用软件 Sequencing Analysis 5.2。所测序列与目标 EST 序列的比对利用软件 DNAMAN V5.2.2 (Lynnon Biosoft, Quebec, Canada)。

### 3.4 PCR 产物的混合克隆和测序

对于候选 32 个桉树 EST-SSR 标记, 按照对细叶桉  $P_2$  扩增产物长度的相互差别尽量较大的原则, 分别选取 8 个(04、06、08、12、15、17、22 和 26 号)、16 个(01、04、06、08、10、11、13、16、17、20、22、23、24、26、28 和 29 号)、24 个(去除 02、05、09、13、14、18、22 和 32 号)以及全部 32 个标记的 PCR 产物进行等摩尔混合, 并利用载体 pGEM®-T Easy (Promega) 进行克隆、通过热激法转化 JM109 感受态细胞, 转化细胞在含有 X-gal/IPTG 的 LB 固体培养基上进行蓝白斑筛选。操作过程按照试剂盒的说明。

从培养平板上挑取 6 倍于原 PCR 产物数量的阳性菌落进行 PCR, 引物为 pUC/M13-F

(5'-GTTTTCCAGTCACGAC-3') 和 pUC/M13-R (5'-TCACACAGGAAACAGCTATGAC-3')。PCR 体系参考张晓红等(2009), 但 10×Buffer 中  $Mg^{2+}$  浓度为 20 mmol/L; PCR 程序为 95°C 5 min, 再 30 个循环: 94°C 30 s、55°C 30 s、72°C 3 min, 最后 72°C 10 min。根据菌落 PCR 产物大小(减去载体序列 253 bp) 确定最可能对应的原 EST-PCR 产物, 针对各个最可能插入了原 EST-PCR 产物的菌落进行测序, 即各批测序的克隆数等于混合的标记数量。测序过程参考前 3.3。所测序列与目标 EST 序列的比对利用软件 DNAMAN V5.2.2 (Lynnon Biosoft)。

### 作者贡献

周长品和李发根是本研究的实验设计和实验研究的执行人; 翁启杰和于晓丽及李梅参与实验执行以及结果分析; 甘四明是项目的构思者及负责人, 指导实验设计, 数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

### 致谢

本研究由国家自然科学基金(31070592)、广东省自然科学基金(10151052001000000)、国家高技术研究发展计划(863 计划)重点项目(2006AA100109)和中央级公益性科研院所基本科研业务费专项(RITFKYYW2008-4 和 2010-04)共同资助。

### 参考文献

- Ayers N.M., McClung A.M., Larkin P.D., Bligh H.F.J., Jones C.A., and Park W.D., 1997, Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm, *Theor. Appl. Genet.*, 94(6-7): 773-781
- Brondani R.P.V., Williams E.R., Brondani C., and Grattapaglia D., 2006, A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus, *BMC Plant Biol.*, 6: 20
- Ceresini P.C., Silva C.L.S.P., Missio R.F., Souza E.C., Fischer C.N., Guilherme I.R., Gregorio I., da Silva E.H.T., Cicarelli R.M.B., da Silva M.T.A., Garcia J.F., Avelar G.A., Neto L.R.P., Marcon A.R., Junior M.B., and Marini D.C., 2005, Satellypus: analysis and database of microsatellites from ESTs of *Eucalyptus*, *Genet. Mol. Biol.*, 28(3s): 589-600
- Cha R.S., and Thilly W.G., 1995, Specificity, efficiency, and fidelity of PCR. In: Dieffenbach C.W. and Dveksler G.S. (eds.), PCR primer: a laboratory manual, Cold Spring



- Harbor Press, New York, USA, pp.37-52
- Choudhary S., Sethy N.K., Shokeen B., and Bhatia S., 2009, Development of chickpea EST-SSR markers and analysis of allelic variation across related species, *Theor. Appl. Genet.*, 118(3): 591-608
- Doyle J.J., and Doyle J.L., 1991, Isolation of plant DNA from fresh tissue, *Focus*, 12(1): 13-15
- Ellis J.R., and Burke J.M., 2007, EST-SSRs as a resource for population genetic analyses, *Heredity*, 99(2): 125-132
- Fang X.J., Wu W.R., and Tang J.L., eds., 2002, *Molecular marker assistant breeding in crop*, Science Press, Beijing, China, pp.10-21 (方宣钧, 吴为人, 唐纪良, 编著, 2002, 作物DNA标记辅助育种, 科学出版社, 中国, 北京, pp.10-21)
- FAO, 2000, *Global forest resources assessment 2000 – main report*, FAO Forestry Paper 140, FAO, Rome, Italy
- Faria D.A., Mamani E.M.C., Pappas M.R., Pappas G.J., Grattapaglia D., 2010, A selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of *Eucalyptus*, *J. Hered.*, 101(4): 512-520
- Gan S., Shi J., Li M., Wu K., Wu J., and Bai J., 2003, Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers, *Genetica*, 118(1): 59-67
- Hill K.D., and Johnson L.A.S., 1995, Systematic studies in the *Eucalyptus*. 7. A revision of the bloodwoods, genus *Corymbia* (Myrtaceae), *Telopea*, 6(2-3): 185-504
- Kunkel T.A., and Bebenik K., 2000, DNA replication fidelity, *Annu. Rev. Biochem.*, 69: 497-529
- Li S.X., Zhang X.Y., Wang Y.Y., and Yin T.M., 2010, Content and characteristics of microsatellites detected in expressed sequence tag sequences in *Eucalyptus*, *Zhiwu Xuebao* (Chinese Bulletin of Botany), 45(3): 363-371 (李淑娴, 张新叶, 王英亚, 尹佟明, 2010, 桉树EST序列中微卫星含量及相关特征, *植物学报*, 45(3): 363-371)
- Mao W.H., 2005, Establishment of direct sequencing method with colony PCR products and its application in rice gene sequencing, *Zhongguo Shuidao Kexue* (Chinese Journal of Rice Science), 19(5): 463-466 (毛伟华, 2005, 菌落PCR产物直接测序方法的建立及在水稻基因测序中的应用, *中国水稻科学*, 19(5): 463-466)
- Mardis E.R., 2008, Next-generation DNA sequencing methods, *Annu. Rev. Genomics Human Genet.*, 9: 387-402
- Powell W., Machray G.C., and Provan J., 1996, Polymorphisms revealed by simple sequence repeats, *Trends Plant Sci.* 1(7): 215-222
- Rabello E, de Souza AN, Saito D, Tsai S-M (2005) *In silico* characterization of microsatellites in *Eucalyptus* spp.: abundance, length variation and transposon associations, *Genet. Mol. Biol.*, 28(3s): 582-588
- Sharma R.K., Bhardwaj P., Negi R., Mohapatra T., and Ahuja P.S., 2009, Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.), *BMC Plant Biol.*, 9: 53
- Simko I., 2009, Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.), *J. Hered.*, 100(2): 256-262
- Steane D.A., Jones R.C., and Vaillancourt R.E., 2005, A set of chloroplast microsatellite primers for *Eucalyptus* (Myrtaceae), *Mol. Ecol. Notes*, 5(3): 538-541
- Tauz D., 1989, Hypervariability of simple sequences as a general source for polymorphic DNA markers, *Nucleic Acids Res.*, 17(16): 6463-6471
- Varshney R.K., Graner A., and Sorrells M.E., 2005, Genic microsatellite markers in plants: features and applications, *Trends Biotechnol.*, 23(1): 48-55
- Xu Z.Y., Bao Q.Y., and Niu Y.X., 2002, Factors that influence direct sequencing of PCR products, *Yichuan* (Hereditas), 24(5): 548-550 (徐祖元, 包其郁, 牛宇欣, 2002, PCR产物直接测序技术中影响因素的研究, *遗传*, 24(5): 548-550)
- Yuan K.J., Liu Q.Z., Ai C.X., and Wang J.H., 2007, A method for isolation of genomic DNA in fruit tree and sequencing its PCR product, *Fenzi Zhiwu Yuzhong* (Molecular Plant Breeding), 5(6s): 173-176 (苑克俊, 刘庆忠, 艾呈祥, 王绎辉, 2007, 果树基因组DNA的提取和PCR扩增测序方法, *分子植物育种*, 5(6s): 173-176)
- Zhang X.H., Li F.G., Wang Y., Xu L.A., Li M., and Gan S.M., 2009, An optimized protocol for sequencing EST-PCR products in *Eucalyptus*, *Jiyinzuxue Yu Yingyongsheng wuxue* (Genomics and Applied Biology), 28(3): 535-543 (张晓红, 李发根, 王宇, 徐立安, 李梅, 甘四明, 2009, 桉树EST-PCR产物测序方案的优化, *基因组学与应用生物学*, 28(3): 535-543)
- Zhou D.G., Zhao Q.Y., Fu C.Y., Li H., Cai X.F., Luo D., and Zhou S.C., 2008, The next generation sequencing and its effect on the rice molecular design breeding, *Fenzi Zhiwu Yuzhong* (Molecular Plant Breeding), 6(4): 619-630 (周德贵, 赵琼一, 付崇允, 李宏, 蔡学飞, 罗达, 周少川, 2008, 新一代测序技术及其对水稻分子设计育种的影响, *分子植物育种*, 6(4): 619-630)

表 1 候选 32 个桉树 EST-SSR 的引物以及对细叶桉(P<sub>2</sub>)的 PCR 序列确认

Table 1 Primer pairs and sequence validation for the 32 candidate EST-SSRs of *Eucalyptus*

序号 No.	EST-SSR 标记 EST-SSR locus	EST 号 GenBank accession	引物对(5'-3') Primer pair (5'-3')	EST 设计长度 (bp) EST target size (bp)	PCR 产物长 度 <sup>a</sup> (bp) PCR product size <sup>a</sup> (bp)	EST 的重复单位 Repeats in EST	测序确认的 P <sub>2</sub> 的重复单位 Repeats in P <sub>2</sub> as revealed in sequencing		序列一致 性(%) Sequence identity (%)
							PCR 产物直接测序 <sup>b</sup> Direct PCR product sequencing <sup>b</sup>	混合克隆测序 <sup>c</sup> Pool cloning of PCR products <sup>c</sup>	
1	EUCeSSR1085	CU403532	F:GTTGTA CTCTGCCCTAA R:CAGCCTCGGATTGTTCA	118	111 (113)	(CT) <sub>8</sub>	(CT) <sub>6</sub>	(CT) <sub>5</sub> : [16]	97.30
2	EUCeSSR1078	CU402447	F:GTGGCATT CACAAGAGCA R:GGAAGCCTTTCGGGATAA	126	129	(CAA) <sub>5</sub>	(CAA) <sub>5</sub>	(CAA) <sub>5</sub> : [32]	100.00
3	EUCeSSR0984	CU395760	F:TCTTACGCACCGTCTCT R:AGCCATTCGCTCCTCC	130	139	TCGTCA (TCG) <sub>6</sub>	(TCG) <sub>(2+n)</sub> T(C/A)T (TCC) <sub>2</sub>	(TCG) <sub>8</sub> (TCT) (TCC) <sub>2</sub> : [32]	97.69
4	EUCeSSR1110	GO247750	F:ATCTCCGCACCCACTCTT R:TGTCATTCTCGTT CAGC	185	173	(CCAATC) <sub>3</sub>	CCAATT	CCAATT: [24/16/8]	90.23
5	EUCeSSR0925	CU402880	F:GAAGGCAGAGGCTTACGA R:CTTATTTTCTCCCCCAA	222	209	(ATC) <sub>11</sub>	(ATC) <sub>6</sub>		99.32
6	EUCeSSR1135	CU402600	F:GAAGTGTCTGCTGCCGAATG R:TGAAGCCCAAGGAAATGC	217	205	(CT) <sub>14</sub>	(CT) <sub>8</sub>	(CT) <sub>8</sub> : [32/24/16/8]	100.00
7	EUCeSSR0980	EW688281	F:TGACCCTTCGGCTTCCTT R:GGGCTCGCCAATCTCCAT	217	214	(GAG) <sub>6</sub>	(GAG) <sub>5</sub>	(GAG) <sub>5</sub> : [24]	99.07
8	EUCeSSR0985	CU396018	F:CTTGTTCCGGTTTCTTTCA R:CACTGAGTAACAGGCAAG	264	242/244/246	(AG) <sub>20</sub>	(AG) <sub>(8+n)(9+n)</sub>	(AG) <sub>9/10</sub> : [32/24/16/8]	99.19
9	EUCeSSR0955	GO248042	F:CATTGGAGTAGCACGAGAA R:ACTCCATACTCTCTGTGTCTCT	257	263	(GAC) <sub>7</sub> + (GA) <sub>7</sub>	(GAC) <sub>7</sub> + (GA) <sub>10</sub>		99.04
10	EUCeSSR1046	CU396352	F:TCAAACCCAGCGATACCC R:AGCCCGATGATCCAAATAA	264	266	(TGC) <sub>5</sub>	(TGC) <sub>5</sub>		98.61
11	EUCeSSR1049	CU397549	F:GGTGATGAGGTGGACAG R:GAGTGGAGTTTCTTATGG	294	312/314	(GA) <sub>7</sub>	(GA) <sub>11/12</sub>		94.76
12	EUCeSSR1075	CU402010	F:TGCGTGTTCGTCAGGAT R:TGACATTGCCCTTCGAT	379	368	(GA) <sub>18</sub>	(GA) <sub>5</sub> (GGGA) <sub>2</sub> GAGGACAA	(GA) <sub>5</sub> (GGGA) <sub>2</sub> GAG GACAA: [32]	96.73
13	EUCeSSR0946	CU403671	F:ACTCCGTCTACATAGCATC R:CCAAGACCACCAAACATA	372	366/370	(TC) <sub>10</sub>	(TC) <sub>7/9</sub>	(TC) <sub>7/9</sub> : [32/16]	98.92
14	EUCeSSR1002	CU395531	F:GAGGAGGATGACGAGATGAG R:CGCAGATAAGTAAGAAGGCTAA	370	370	(CAC) <sub>5</sub>	(CAC) <sub>5</sub>		93.15



序号 No.	EST-SSR 标记 EST-SSR locus	EST 号 GenBank accession	引物对(5'-3') Primer pair (5'-3')	EST 设计长度 (bp) EST target size (bp)	PCR 产物长 度 <sup>a</sup> (bp) PCR product size <sup>a</sup> (bp)	EST 的重复单位 Repeats in EST	测序确认的 P <sub>2</sub> 的重复单位		序列一致 性(%) Sequence identity (%)
							Repeats in P <sub>2</sub> as revealed in sequencing		
							PCR 产物直接测序 <sup>b</sup> Direct PCR product sequencing <sup>b</sup>	混合克隆测序 <sup>c</sup> Pool cloning of PCR products <sup>c</sup>	
15	EUCeSSR1014	CU395155	F:GGGACTGAGCTCTGATCCTACT R:GACCGCTGGTACTTGGCTG	235	396/402	(AG) <sub>3</sub> AA(AG) <sub>3</sub> G AAGCAAGAGA T(AG) <sub>15</sub> GAAAG	AC(AG) <sub>9/12</sub> : [24/8/32]	97.42	
16	EUCeSSR1083	CU403250	F:CGGGGTCTCTCGAAGCGTCT R:CCGACGGCATCGTAAACC	314	435	(GGAAGC) <sub>3</sub>	(GGAAGC) <sub>3</sub> : [24/16]	97.45	
17	EUCeSSR1136	CU396171	F:GATGGCAGGACATAAAGA R:GCTAAACAAACCAGGAAA	373	490/492	(GA) <sub>8</sub> AA(GA) <sub>9</sub>	[(G/A)A](GA) <sub>5</sub> [(G/A)A](GA) <sub>20/21</sub>	98.79	
18	EUCeSSR1091	CU404650	F:GTTTCATCCGCCCTTACAG R:ACCCTTGCTCCTTCTTCG	402	~490	(TGTTT) <sub>3</sub> +T <sub>14</sub> G <sub>3</sub>	TGTT(TGTTT) <sub>2</sub> + (T <sub>9</sub> G <sub>17</sub> )/(T <sub>8</sub> G <sub>12</sub> A <sub>3</sub> G <sub>4</sub> )	97.51	
19	EUCeSSR0894	EW688426	F:AGTCCTTCAGCAGGTTC R:CCGCCGGTATAGTAGTTG	145	518	(CAC) <sub>6</sub>	(CAC) <sub>6</sub>	92.78	
20	EUCeSSR0961	CU396545	F:CCCCTTCACTGCTCTTCT R:ACCAACCGCGAGTATGTC	242	547	(GCT) <sub>6</sub>	(GCT) <sub>n/(n+1)</sub> (GCT) <sub>3</sub> [(G/A)CT] (GCT) <sub>2/3</sub> : [32/24/16]	98.76	
21	EUCeSSR1087	CU403690	F:TGAGGCACAACAAGGTAA R:CTGGCTGTAGATTGAAGAGT	439	~650	(AGC) <sub>10</sub>			
22	EUCeSSR1054	CU398509	F:CTTCTCGTCTTCTTCTGC R:TAAGGTCACCCAAGTTCTG	241	665/671	(CTG) <sub>2</sub> (CCG) (CTG) <sub>6</sub> : (CTG) <sub>6</sub>	(CTG) <sub>(n+3)/(n+5)</sub> (CTG) <sub>1/2</sub> (CCG) (CTG) <sub>9/10</sub> : [32/16/8]	99.59	
23	EUCeSSR1122	GO248349	F:ACAACAATCTTGCCTTCT R:TCACAATAAACGAGTCCC	361	687	(CTG) <sub>6</sub>	(CTG) <sub>6</sub> (CTG) <sub>6</sub> : [32/24/16]	99.59	
24	EUCeSSR1011	CU396518	F:GCAGAGGAGCGGAAACTA R:GCTACTTTGGGTGIGTCTT	440	~850	(TC) <sub>9</sub>			
25	EUCeSSR1150	CU397973	F:ACCCTCCTCCTCCGTATT R:CACATCCCATCTCCTTCC	440	~850	(TC) <sub>6</sub>			
26	EUCeSSR0970	CU402791	F:GGCCCGCCTTTCCAATTC R:TTGTCTCGGTTGGGTCTT	225	997	(TC) <sub>10</sub>	(TC) <sub>10</sub> : [32/16/8]	91.11	
27	EUCeSSR0914	CU397331	F:CCGGGGGCTTTGCTTICA R:TACCTCCATTGCTTCCTT	220	~1200	(AG) <sub>10</sub>	(AG) <sub>10</sub> : [32]	96.36	
28	EUCeSSR1069	CU400565	F:ACCTCCATCCTAAGACGA R:ATCCCAAGACCCATT	318	1175	(GA) <sub>5</sub> CA(GA) <sub>4</sub> (CA)(GA) <sub>5</sub> CA (GA) <sub>3</sub>	(GA) <sub>15/16</sub> (GA) <sub>15</sub> : [32/24]	97.36	
29	EUCeSSR1102	EW688317	F:TTCTTCTTGGGTTTCAGT R:ACAAGCAATGCTTTTCTC	371	~1300	(GG) <sub>2</sub> (AG) <sub>7</sub> AA(A G) <sub>3</sub>	GG(GA) <sub>10/12</sub> : [32/24/16]	98.36	

序号 No.	EST-SSR 标记 EST-SSR locus	EST 号 GenBank accession	引物对(5'-3') Primer pair (5'-3')	EST 设计长度 (bp) EST target size (bp)	PCR 产物长 度 <sup>a</sup> (bp) PCR product size <sup>a</sup> (bp)	EST 的重复单位 Repeats in EST	测序确认的 P <sub>2</sub> 的重复单位		序列一致 性(%) Sequence identity (%)
							Repeats in P <sub>2</sub> as revealed in sequencing		
							PCR 产物直接测序 <sup>b</sup> Direct PCR product sequencing <sup>b</sup>	混合克隆测序 <sup>c</sup> Pool cloning of PCR products <sup>c</sup>	
30	EUCeSSR1060	CU398944	F:CTCAGCTTCCGTATCTTG R:TTGATTACTGCGATTTAGAG	420	1274	GG(GA) <sub>7</sub>		(GG) <sub>2</sub> (GA) <sub>13</sub> : [24]	98.57
31	EUCeSSR1034	CU399256	F:CGGGGGCGTTTCGGAGATT R:CCACTCTGACAGCCCTTT	240	1450	(AGAAG) <sub>3</sub>			97.62 (部分序列)
32	EUCeSSR1028	GO248406	F:GCCATTGGGTATTCTTCT R:CTCGCTTCCTTTCACT	299	1500	(GA) <sub>7</sub>			

注: a: PCR 产物长度优先根据克隆测序的结果确定, 次则根据 PCR 产物直接测序与源 EST 的比对结果推测, 测序未成功者根据图 1 确定; b: PCR 产物直接测序中, 因部分 SSR 的重复单位靠近(前向)测序引物, 只有部分重复单位可见, n 表示未知的重复单位数; c: 混合克隆测序中, 中括号内数字表示成功测序的混合克隆的 PCR 产物数量  
 Note: a: PCR product size was determined sequentially with pool cloning of PCR products, direct sequencing of PCR products or manually estimated from Figure 1; b: n denotes an unknown number of SSR motif repeats missed in sequencing due to the closeness of SSR motifs to the (forward) sequencing primer site; and c: Numerals in the brackets show the number of EST-SSRs pooled for cloning in which the marker was successfully sequenced