doi: 10.11707/j.1001-7488.20130807

# 基于 454 测序的油茶 DNA 序列微卫星观察与分析*

温 强[1,2] 徐林初[2] 江香梅[2] 李 江[2] 顾胤聪[1] 徐立安[1] 黄敏仁[1]
(1. 南京林业大学林木遗传与生物技术省部共建教育部重点实验室 南京 210037; 2. 江西省林业科学院 南昌 330032)

**摘 要:** 采用 MISA 软件挖掘比较各 1/4 个 454 高通量测序反应获得的普通油茶、浙江红山茶和短柱茶 EST 序列及普通油茶基因组序列中的微卫星信息。结果显示: 3 个种的 EST 序列微卫星出现频率大小相近,普通油茶 EST 序列的微卫星出现频率高于其基因组序列的; 在所有被检索序列的二至六碱基微卫星中,均以二碱基微卫星最多( >55% ),并以( AG )$_n$类型为主,但三碱基微卫星在所有 EST 序列中以( AAG )$_n$类型最多,而在普通油茶基因组序列中则以( AAT )$_n$最多; 除六碱基微卫星外,二至五碱基微卫星均表现为不同微卫星重复单元的丰度随着微卫星碱基长度增加而减少。在转录组序列中,除六碱基微卫星之外,不同微卫星单元重复数的变异与重复单元长度呈负相关,推测二碱基微卫星理论多态性最高,而五碱基微卫星理论多态性最低; 对在无冗余独立基因各区域中的微卫星进行统计,显示分布比例依次为 3′UTR > CDS > 5′UTR,在相同区域内比较,UTR 区域中二碱基微卫星占总数量比例最大,CDS 区域内三碱基微卫星的比例最高,此外 5′UTR 区域的三碱基微卫星的分布比例均高于 3′UTR,初步推断 5′UTR 区域的微卫星相对 3′UTR 要保守些。

**关键词:** 山茶属; 454 高通量测序; 微卫星
**中图分类号:** S718.46 **文献标识码:** A **文章编号:** 1001-7488(2013)08-0043-08

## Survey and Analysis of Microsatellites from DNA Sequences in *Camellia* Species Using 454 Pyrosequencing

Wen Qiang[1,2] Xu Linchu[2] Jiang Xiangmei[2] Li Jiang[2] Gu Yincong[1] Xu Li'an[1] Huang Minren[1]
(1. *Key Laboratory of Forest Genetics and Biotechnology of Ministry of Education* *Nanjing Forestry University* *Nanjing* 210037;
2. *Jiangxi Forestry Academy* *Nanchang* 330032)

**Abstract:** Abstract: By using 454 pyrosequencing, the simple sequence repeats (SSRs) from DNA sequences in three *Camellia* species, including *C. oleifera*, *C. brevistyla* and *C. chekiangoleosa*, were tapped with repeats of 2 to 6 bp by MISA. In all cases, the frequency of repeat motifs in the EST was similar, while the proportion in the EST of *C. oleifera* was higher than that of its genome. Comparisons of all SSRs, the dinucleotide repeat microsatellites (DNRs) were most dominant ( > 55% ) and the most common DNRs was (AG)$_n$. The most common trinucleotide repeat microsatellites (TNRs) was (AAG)$_n$ in EST, while the most common TNRs was (AAT)$_n$ in the genome of *C. oleifera*. Apart from the hexaucleotide repeat microsatellites (HXNRs), the abundance of the other four classes of microsatellites was negatively correlated with their repeat motif lengths. In the EST, apart from the HXNRs, diversification of microsatellite lengths (repeat number variation) was also negatively correlated with their repeat motif lengths, and we predicted that DNRs was the highest polymorphism and pentaucleotide repeat microsatellites (PTNRs) was the lowest. Subsequently, statistical analysis on SSRs within coding regions (CDS) and untranslated regions (UTR) in the Unigene showed that the ratio of motifs distribution was ranked as the following: 3′UTR > CDS > 5′UTR. The DNRs were the dominant microsatellite in UTR, while the TNRs were dominant in CDS. The proportion of TNRs in the 5′UTR was higher than that in the 3′UTR, indicating that the SSRs in the 5′UTR were more conservative.

**Key words:** *Camellia*; 454 pyrosequencing; microsatellites

山茶属(*Camellia*)植物绝大部分都具有重要的经济价值与观赏价值,其中重要的油用物种如油茶(*Camellia oleifera*)称之为"东方的橄榄树"而闻名于世(Zhang *et al.*, 2008)。目前在生产上通称的油茶

收稿日期: 2013-03-11; 修回日期: 2013-04-23。
基金项目: 国家自然科学基金项目(31260184); 江西省自然科学基金项目(2009GQN0047); 江西省科技支撑项目(20122BBF60125)。
* 徐立安为通讯作者。

是指种子含油率较高、有栽培及食用价值的山茶属多个种的总称，包括如油茶、浙江红山茶（*C. chekiangoleosa*）、短柱茶（*C. brevistyla*）等。当前全国油茶选育工作已进入新品种选育及种质创新阶段（庄瑞林，2010），从大量的油茶遗传资源中选育并丰富油茶良种迫在眉睫，因而研究油茶遗传变异规律、开展油茶分子标记辅助育种势在必行。

微卫星标记作为共显性分子标记，具有稳定、高效的特点，正被广泛运用于林木遗传变异分析、遗传连锁图谱构建以及遗传育种等研究中。该标记来源于广泛散布于基因组中的较短重复单元组成的微卫星序列，有研究认为这些序列是在基因组进化过程中导致并维持数量性状变异的重要因素（Kashi *et al.*, 2006），同时是基因组中变异率相对较高的一类DNA序列（Christian, 2000）。微卫星的密度与分布特点在不同物种的整个基因组序列中呈现丰富多样性（Dieringer *et al.*, 2003），其中分布于DNA序列的编码区域微卫星重复单位的类型及微卫星长度的变异影响不同调节蛋白在翻译过程中的相互关系（Lue *et al.*, 1989），甚至可能影响上游激活序列的转录活力（Martienssen *et al.*, 2001）。微卫星重复单位的高度变异表现为微卫星数目的整倍性变异或重复单元序列中的不完全相同序列（Gow, 2005），从而造成多个位点的多态性，使得利用微卫星开发的SSR（simple sequence repeat）标记具有较高的遗传多态性。同时，从在植物组织中直接参与功能表达的编码序列中开发的SSR标记，与其他标记相比，可更多地将标记与表型相关联（Varshney *et al.*, 2005），势必包含更加丰富的生物学信息。

山茶属中茶（*C. sinensis*）的SSR分析报道较多（金基强等，2007；Sharma *et al.*, 2009），而山茶属中重要油用物种的SSR分析仅见史洁等（2012）对浙江红山茶基因组开展的SSR研究。近年来，454测序技术引领的高通量测序为在大量DNA序列基础上开展多物种的微卫星比较分析乃至大量SSR标记的开发成为可能。基于前述，本研究在对普通油茶、浙江红山茶、短柱茶3个典型油茶树种454高通量深度测序获得的大量基因组与转录组序列的基础上，充分发掘微卫星信息，分析比较油茶DNA序列中微卫星重复序列的组成与分布特征，为山茶属不同树种基因组的分化提供重要参数，同时也为油茶多态性高的微卫星标记的开发提供生物信息学参考。

## 1 材料与方法

### 1.1 基因组及转录组序列来源

普通油茶、浙江红山茶与短柱茶转录组序列及普通油茶基因组序列来源于各1/4个454 GS FLX高通量测序反应的结果。其中普通油茶花芽总基因组经纯化后直接上机测序，测序得到的基因组序列用454测序仪自带的Newbler软件进行处理和拼接；其他转录组序列的获得及拼接处理参见陈英等（2011）的报道。

### 1.2 454序列中微卫星的挖掘与统计

采用MISA软件（Thiel *et al.*, 2003）对454序列中的2~6核苷酸重复类型SSR进行检索，检索标准同时包括精确型（perfect）及复合型（compound）SSR重复单元（Weber, 1990），含二（dinucleotide repeat microsatellites, DNRs）、三（trinucleotide repeat microsatellites, TNRs）、四（tetranucleotide repeat microsatellites, TTNRs）、五（pentanucleotide repeat microsatellites, PTNRs）和六核苷酸重复微卫星类型（hexanucleotide repeat microsatellites, HXNRs）的基序（motif），检索标准依据Berube等（2007）的方法，最小长度为18 bp，最小重复数分别为9,6,5,5,4次。

Excel软件统计检索数据，包括微卫星各种重复单元所占百分比含量、微卫星的平均跨度和发生频率（含有微卫星的DNA序列与分析的独立序列总数之比）。在统计重复基序类型时，将基序的所有可能的+1移码及其互补序列都视为同一个基序类型。如对于二核苷酸重复基序AC而言，AC = CA = TG = GT，将这4种二核苷酸重复类型出现的次数之和作为重复基序AC出现的次数。按照这种方法，二核苷酸重复基序只有4种类型，三核苷酸重复基序只有10种类型（Jurka *et al.*, 1995；Cardle *et al.*, 2000）。

### 1.3 EST序列中不同长度微卫星的分布及变异

统计油茶3树种EST序列中长度≥18 bp的SSR，利用EXCEL的作图功能对不同长度重复单元构成的相同类型的微卫星分别绘制饼图，图中每一扇区对应不同长度的微卫星及其在同类型微卫星中的出现频率，扇区的多少反映了微卫星长度的变异情况，扇区越多说明对应类型的微卫星获得或失去重复单元的速率越快，因此对应类型的微卫星总体上会有更高的多态性。

### 1.4 微卫星在Unigene上的分布

采用ESTscan软件（Iseli *et al.*, 1999）以拟南芥（*Arabidopsis thaliana*）转录组数据库为参照［参数设置插入与缺失（indel）罚值为15］，获取EST序列的CDS区域，分析微卫星在无冗余独立基因（unigene）序列的编码（CDS）及非编码区域（UTR）的分布规

律，同时结果采用 NCBI 的 ORF Finder 软件（http://www.ncbi.nlm.nih.gov/gorf/gorf.html）进行验证，数据结果采用 Excel 软件进行统计并绘制饼图。此外，以精确型 Motif 为检索标准，进一步统计编码区域及非编码区域中 2～6 核苷酸的组成情况。

## 2　结果与分析

### 2.1　油茶 454 序列微卫星分布的基本特征

经 MISA 软件发掘 4 种序列来源的微卫星（表 1）。比较微卫星分布频率与跨度，转录组序列微卫星发生频率依次为：短柱茶（0.056）＞浙江红山茶（0.055）＞普通油茶（0.047），而普通油茶转录组序列比其基因组序列具有较高相对含量的微卫星。在所检索序列的所有微卫星重复单元中，二核苷酸均为主要的重复类型，所占比例依次为普通油茶基因组序列（62.40%）＞短柱茶 EST 序列（58.77%）＞浙江红山茶 EST 序列（55.53%）＞普通油茶 EST 序列（55.43%）；除六核苷酸重复单元外，其他 4 种核苷酸重复单元表现为重复单元的碱基长度与不同类型重复单元的丰度呈负相关；在所检测重复单元类型中，转录组序列中三、六核苷酸重复单元百分含量明显高于普通油茶基因组序列中的。几种主要的核苷酸重复单元占总量的百分含量见表 2。在二核苷酸重复基序类型中，以 AG 类型最多，而 CG 重复仅在普通油茶基因组序列及浙江红山茶转录组序列中极少量出现；而对于三核苷酸重复基序类型，在 3 树种转录组序列中以 AAG 类型最多，普通油茶基因组序列中三核苷酸序列则以 AAT 类型数量最多；另外在所有序列中四核苷酸重复类型所占比例均表现出 AAAT 类型含量最高。

**表 1　油茶 3 树种 454 序列中微卫星的数量及发生频率①**

**Tab.1　Number and distribution of SSRs in *Camellia* 454 sequences**

| 特征 Characters | 序列来源 Sequence resource | | | |
|---|---|---|---|---|
| | 转录组序列 Transcript sequence | | | 基因组序列 Genome sequence |
| | 普通油茶 *C. oleifera* | 浙江红山茶 *C. chekiangoleosa* | 短柱茶 *C. brevistyla* | 普通油茶 *C. oleifera* |
| 检测序列总数 Total number of sequences examined | 36 339 | 46 279 | 40 168 | 98 756 |
| 检测序列总长度 Total size covered by examined sequences/bp | 12 867 557 | 17 065 478 | 15 957 902 | 30 773 610 |
| 检测到的 SSRs 总数量 Total number of SSRs identified | 1 815 | 2 730 | 2 411 | 3 229 |
| 复合型重复单元数量 Number of compound microsatellites | 84 | 127 | 118 | 120 |
| 包含有微卫星重复单元的序列数量 Number of SSR-containing sequences | 1 696 | 2 554 | 2 248 | 3 065 |
| 总的微卫星序列发生频率 Total relative abundance | 0.047 | 0.055 | 0.056 | 0.031 |
| 微卫星序列平均跨度 Average distance/bp | 7 089.56 | 6 251.09 | 6 618.79 | 9 530.38 |
| 二核苷酸重复 Dinucleotide repeat | 1 006（55.43） | 1 516（55.53） | 1 417（58.77） | 2 015（62.40） |
| 三核苷酸重复 Trinucleotide repeat | 484（26.67） | 727（26.63） | 630（26.13） | 596（18.46） |
| 四核苷酸重复 Tetranucleotide repeat | 103（5.67） | 133（4.87） | 102（4.23） | 268（8.30） |
| 五核苷酸重复 Pentanucleotide repeat | 33（1.82） | 46（1.68） | 36（1.49） | 78（2.42） |
| 六核苷酸重复 Hexanucleotide repeat | 189（10.41） | 308（11.28） | 226（9.37） | 163（5.05） |

① 括号前数值为各 SSR 重复类型的数量，括号内数值为对应的百分比含量（%）。The value in front of parentheses is the number of each SSR type, numbers in parentheses show percentage of total SSR content（%）.

### 2.2　EST 序列中微卫星长度分布及变异分析

微卫星 DNA 的多态性主要来源于串联数目的不同，因而了解序列中不同长度微卫星的分布及变化对于今后有针对性地开发高多态性的 SSR 标记意义重大。统计显示普通油茶、浙江红山茶、短柱茶的 EST 序列所含微卫星长度上存在较为丰富的变化，其中检索出来的重复基元序列长度大小不等，普通油茶为 18～70 个碱基，浙江红山茶为 18～72 个碱基，短柱茶为 18～170 个碱基，其中 170 个碱基长度的微卫星仅出现了 1 次；3 树种微卫星序列平均长度大致相同，依次分别为 23.272，23.054 及 23.123 个碱基。此外，对 EST 序列所含不同长度重复单元的微卫星的长度变异情况进行了分析，结果（图 1）显示，3 树种存在相似的规律，即对于 2～5 核苷酸重复单元，这些微卫星的长度变异与对应的重复单元类型的碱基长度呈反比，而六核苷酸重复单元则有别于前者，表现为相反的趋势。

表 2 主要的核苷酸重复单元在 4 种序列中的百分比

Tab. 2 Relative percentage of SSRs in *Camellia* 454 sequences

| 主要的重复单元类型 Repeat composition | 在不同序列类型中所占百分比 Relative percentage of SSRs (%) | | | |
|---|---|---|---|---|
| | 普通油茶基因组 Genome of *C. oleifera* | 普通油茶 EST EST of *C. oleifera* | 浙江红山茶 EST EST of *C. chekiangoleosa* | 短柱茶 EST EST of *C. brevistyla* |
| 二核苷酸 DNRs | | | | |
| AC | 3.90 | 3.64 | 3.63 | 4.27 |
| AG | 32.27 | 44.30 | 42.27 | 44.75 |
| AT | 26.14 | 7.49 | 9.60 | 9.75 |
| CG | 0.09 | 0 | 0.04 | 0 |
| 三核苷酸 TNRs | | | | |
| AAC | 1.98 | 2.53 | 2.49 | 2.74 |
| AAG | 3.59 | 8.10 | 7.33 | 6.55 |
| AAT | 7.68 | 4.13 | 4.36 | 3.82 |
| ACC | 1.73 | 4.24 | 4.32 | 3.98 |
| ACG | 0.37 | 1.76 | 1.39 | 1.70 |
| ACT | 2.51 | 3.69 | 4.87 | 4.44 |
| AGG | 0.46 | 1.87 | 1.54 | 2.49 |
| CCG | 0.12 | 0.33 | 0.33 | 0.41 |
| 四核苷酸 TTNRs | | | | |
| AAAC | 0.46 | 1.10 | 0.44 | 0.29 |
| AAAG | 0.43 | 0.72 | 0.33 | 0.37 |
| AAAT | 5.05 | 2.31 | 1.94 | 1.74 |
| AACT | 0.28 | 0.50 | 0.44 | 0.41 |
| AACG | 0.06 | 0.06 | 0.04 | 0.17 |
| AAGG | 0.06 | 0.06 | 0.11 | 0.21 |
| AAGT | 1.08 | 0.39 | 0.55 | 0.41 |
| AATT | 0.28 | 0.00 | 0.04 | 0.08 |
| AACC | 0.00 | 0.11 | 0.11 | 0.04 |
| ACCC | 0.06 | 0.00 | 0.04 | 0.08 |
| ACCG | 0.06 | 0.06 | 0.11 | 0.04 |
| ACGG | 0.03 | 0.06 | 0.04 | 0 |
| AGGG | 0.09 | 0.11 | 0.18 | 0.08 |
| AGGT | 0.00 | 0.00 | 0.00 | 0 |
| ATCC | 0.22 | 0.17 | 0.22 | 0.21 |
| ATCG | 0.12 | 0.06 | 0.29 | 0.08 |

### 2.3 微卫星在 Unigene 上的分布分析

油茶 3 树种 Unigene 上的 SSR 经 ESTscan 软件检索确定其分区情况,数据利用 Excel 软件绘制饼图(图 2),3 树种中除浙江红山茶有 16.83%、短柱茶有 13.52%、普通油茶有 16.52% 的 SSR 不能确定其在 Unigene 上的位置之外,其余的 SSR 均能明确其在 UTR 区域或者 CDS 区上的分布,并具有较为相似的分布规律。总水平上,3 树种表现为分布于 UTR 区域的 SSR 所占比例均高于分布于 CDS 区域的,而各分部区域比例总的表现为 3′UTR > CDS > 5′UTR。

此外,排除未知及横跨在 UTR 与 CDS 区域(文中称之为兼有)的微卫星重复单元,以精确型 Motif 为检索标准,进一步统计仅落在 CDS 及 UTR 区域中 2 ~ 6 核苷酸重复单元的组成情况。图 3 显示 3 树种不同长度重复单元的分布规律基本一致,在 UTR 区域,重复单元主要以二核苷酸重复单元为主,占分布其上的重复单元总数的 70% 以上,除六碱基重复单元其他 5 种重复单元呈现依次减少的趋势,同时,数据显示 3 树种分布于 5′UTR 区域的三碱基微卫星的比例均高于 3′UTR 的;而在 CDS 区域,重复单元则以三核苷酸重复单元为主,占分布其上重复单元总数的 50% 以上,其次为六核苷酸重复单元,占总数的 25% 以上。

## 3 讨论

### 3.1 总的微卫星重复类型分布特征

本研究 DNA 序列来自山茶属分属油茶组、红山茶组、短柱茶组的 3 个油用代表树种,同时在各 1/4 个 454 测序反应的结果中既包括了 3 树种的 EST 序列,又具有普通油茶基因组序列,检索结果较能全面地反映油茶 DNA 序列中微卫星的分布特征。从 MISA 软件总的检索结果来看,所有序列中 SSR 的发生频率在 0.031 ~ 0.067 之间,出现跨度在

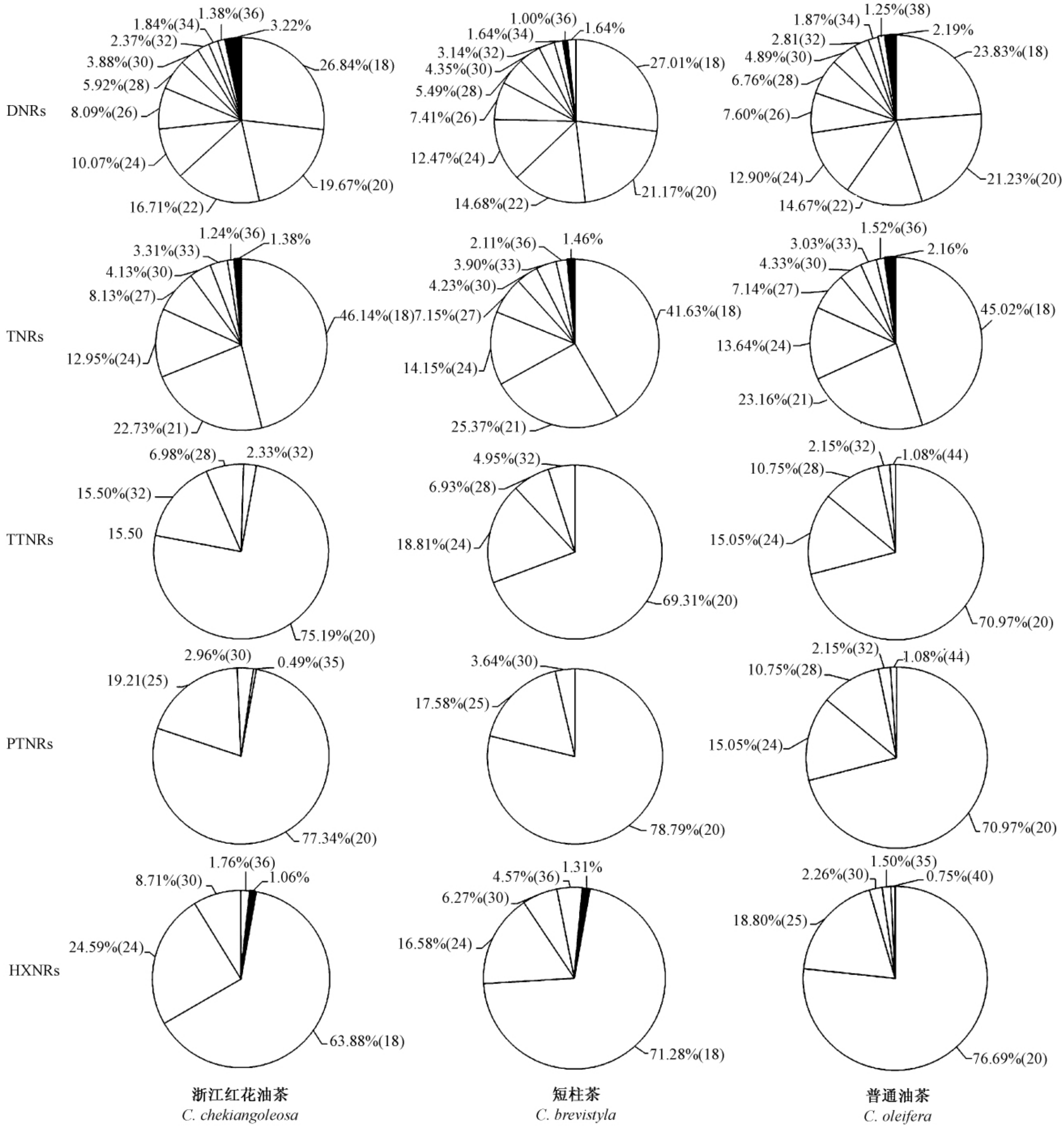


图 1　油茶 3 树种 EST 序列不同长度重复单元微卫星变异

Fig. 1　Length diversification of the microsatellites in 454-EST sequences of *Camellia*

对频率≤1%的微卫星进行了合并，每一扇区对应不同长度的微卫星，而该长度微卫星所占的比例用连线标注在对应的扇区一侧，除合并的扇区之外，括号内为其对应的微卫星长度。The microsatellites with percentage less than ≤ 1% are combined and shown in the black slice. The slice sizes in each pie chart are scaled according to the percentage of microsatellites in different lengths, numbers in parentheses show the length of the microsatellite except the black slice.

6 251.09 ~ 9 530.38 bp 之间，约小于 Kantety 等(2002)禾谷类作物 SSR 发生频率(0.07 ~ 0.10)，在许玉兰等(2012)统计的近年来文献报道的 13 种林木中 EST 序列的发生频率(0.015 9 ~ 0.145 9)范围之内。短柱茶与浙江红山茶的微卫星频率与跨度较为接近，均高于普通油茶基因组及转录组序列的观测值，造成这一结果的主要原因可能在于前二者为二倍体，而普通油茶为六倍体，多倍体植物核中存在的重复基因及重复基因组使测序效率下降，同时多倍体基因组重组过程中重复序列易出现基因入侵现象(杨继，2001)，这也降低了重复序列的复杂性。同时，油茶 454 序列检索显示，油茶 3 树种的 EST 序列与普通油茶基因组序列相比，后者微卫星含量相对最少，Morgante 等(2002)和 Lawson 等(2006)对基

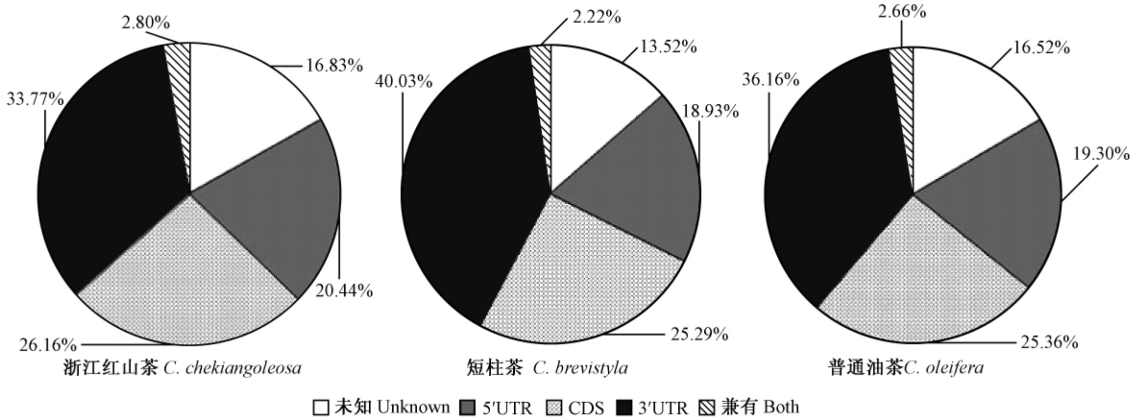


图 2 油茶 3 树种转录组微卫星的分布区域

Fig. 2 The distribution of SSR loci from 3 species of *Camellia* on Unigenes

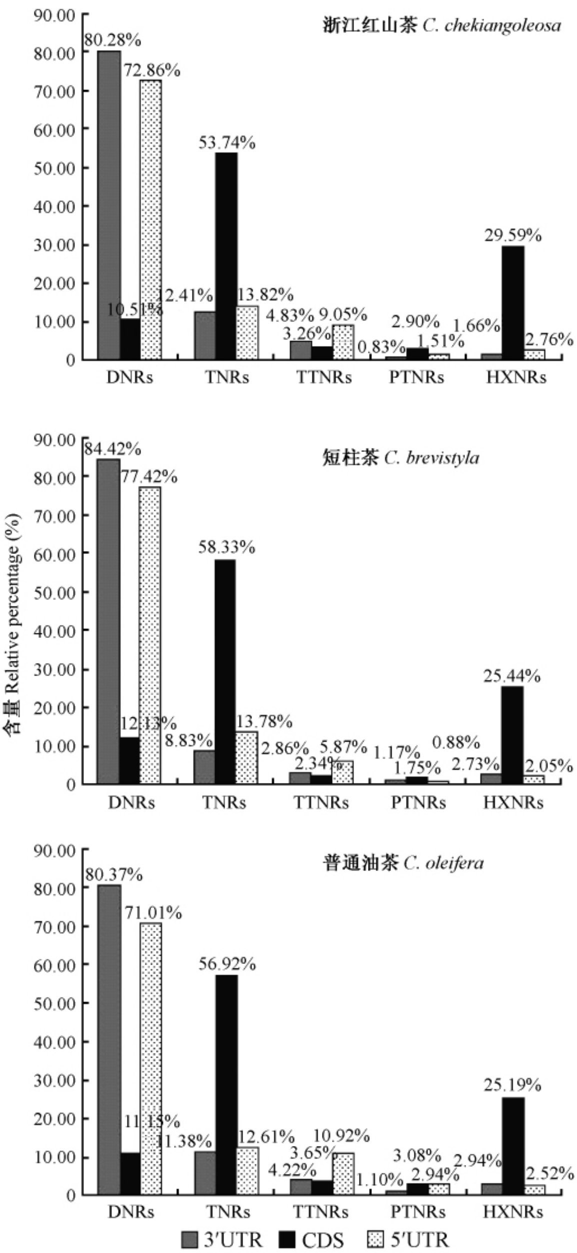


图 3 油茶 3 树种不同长度微卫星在 Unigene 上的分布

Fig. 3 Distribution of different size motifs for EST-SSR in three species of *Camellia*

因组全序列的分析和比较发现拟南芥等植物的基因组 DNA 比 EST 中的微卫星分布密度要低，这与本文的结果一致。在所观测的序列中微卫星重复单元中二、三核苷酸重复单元居多，其中以二核苷酸重复单元最为丰富，这与 Kantety 等(2002)、Varshney 等(2002)对禾谷类作物中的微卫星研究结果三核苷酸类型最多不同。从近年来林木微卫星研究报道来看，模式树种杨树(*Populus*)基因组中三核苷酸重复类型含量最多(Li *et al.*, 2007)，但在其他树种如麻疯树(*Jatropha curcas*)(Yadav *et al.*, 2011)、橡胶树(*Hevea brasiliensis*)(Feng *et al.*, 2009)、银杏(*Ginkgo biloba*)(樊洪泓等, 2009)、北美鹅掌楸(*Liriodendron tulipifera*)(胥猛等, 2008)EST 序列中重复单元则是以二核苷酸重复单元最多，可见油茶的这一特点非特例现象。

此外，油茶 454 序列的二核苷酸重复单元类型中均以$(AG)_n$类型为最多，而$(AC)_n$相对稀少，四核苷酸以$(AAAT)_n$类型最多，而检索到三核苷酸重复类型在油茶 3 树种的 EST 序列中最多的为$(AAG)_n$，这与杨树、桉树(*Eucalyptus*)EST 序列检索结果一致(阎毛毛等, 2011)，同时也与山茶属茶树 EST 序列中微卫星的分布特征(金基强等, 2007; Sharma *et al.*, 2009)相一致。Li 等(2004)认为 AAG 重复是植物微卫星中最主要的三核苷酸重复单元，但本研究检索到三核苷酸重复类型在普通油茶基因组序列中则以$(AAT)_n$最多，而史洁等(2012)检索基因覆盖度在 10% 的浙江红山茶基因组序列微卫星分布中三核苷酸重复类型也以$(AAT)_n$最多，初步推断这是山茶属树种基因组序列微卫星有别于其 EST 序列的主要特点。另外，Morgante 等(2002)研究了单子叶植物中的$(CG)_n$重复单元，认为这些基序增加了微卫星序列中 GC 含量的同时也可能是某

些特定的氨基酸序列；Li 等(2007)在对杨树基因组微卫星的研究中也提到，SSR 序列中的 GC 含量显著相关于微卫星的功能。本研究中普通油茶基因组序列与浙江红山茶转录组序列中也同样包含极个别的$(CG)_n$重复单元，而在三核苷酸重复中也存在少量的$(CCG)_n$类型，这些序列是否与某些特定功能相关有待今后进一步研究。

### 3.2　微卫星长度变异及在 Unigene 的分布

SSR 虽然是一些简单序列的重复，但它却有着复杂的进化机制(Ellegren, 2004)。在对 SSR 的研究中，提出了很多解释 SSR 进化或突变的模型，但这些模型都具有一定的局限性，不能完全解释 SSR 的特征。从油茶 3 树种 EST 序列检索的微卫星长度的分布来看，长度≥18 bp 的微卫星其序列长短不一，但平均值均在 23 bp 左右，显示短序列的重复单元占多数且相对集中，而长序列则分散而稀疏。由于微卫星序列越长，其稳定性越差，尤其在基因区，微卫星序列越长则基因越不稳定(李淑娴等, 2010)。根据这一分析，判断油茶表达基因所含的微卫星可能会受到强烈趋同选择的压力，从而使这些重复序列向较短的序列进化。另外，微卫星重复单元长度的变化情况可以反映微卫星位点获得(或失去)重复单元的活跃程度。本研究通过对 EST 不同长度重复单元的微卫星长度进行分析，结果显示除六核苷酸重复微卫星之外，这些微卫星的长度变异与所含重复单元的长度变化呈反比。故总体而言，除六核苷酸微卫星外，五核苷酸重复微卫星理论多态性最低，而二核苷酸重复微卫星理论多态性最高。

目前的研究已经表明，SSR 偏向分布于转录区，其在转录区的分布密度要比非转录区高得多，即使在转录区 SSR 的分布也不是均匀的随机分布，其在 UTR 的密度要显著高于 CDS 区(Li *et al*., 2002)。本研究中转录组的微卫星无论是频度还是跨度均高于基因组区域，也显示 SSR 共同的分布规律。SSR 位点在 Unigene 的位置会决定或者影响到其功能和变异。通常位于 CDS 区的 SSR，当其长度发生变异时会影响到蛋白质的翻译，甚至产生新的表型性状，这使得位于 CDS 区的 SSR 比位于 UTR 区的 SSR 面临着更大的选择压力(Li *et al*., 2004)，而油茶 EST 序列的 SSR 位点，位于 CDS 区的数量要少于 UTR 区总的数量，这与前述事实相符。此外，对 Unigene 上的精确型不同类型重复单元在各自分布区域的分布含量进行统计的结果显示，UTR 区域的二核苷酸重复单元数量占相当高比例，而 CDS 区域内三核苷酸重复单元的比例远远高于其他重复单元类型，这一结果与 Morgante 等(2002)和 Li 等(2004)的研究结果一致。另外，本研究中油茶 3 树种 5′UTR 区域的三核苷酸微卫星的分布均高于 3′UTR，这一结果与拟南芥的相关研究(Morgante *et al*., 2002)一致。部分 5′UTR 区域内的某些 SSR 发生变异时，该特定变异区域会通过影响转录和翻译来调控基因的表达，即在基因表达过程中起到一定的辅助作用(Kalia *et al*., 2011)，而基因内的 SSR 比基因组其他区域的 SSR 面临着更大的选择压力，因而可以预计 5′UTR 区域的 SSR 相对 3′UTR 要保守些。

## 参　考　文　献


陈　英，江香梅，张　露，等. 2011. 基于油茶 57 万条 EST 序列的转录组学研究. 林业科学，47(2)：161－163.

樊洪泓，李廷春，李正鹏，等. 2009. 银杏 EST 序列中微卫星的分布特征. 基因组学与应用生物学，28(5)：869－873.

金基强，崔海瑞，龚晓春，等. 2007. 用 EST-SSR 标记对茶树种质资源的研究. 遗传，29(1)：103－108.

李淑娴，张新叶，王英亚，等. 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.

史　洁，尹佟明，管宏伟，等. 2012. 油茶基因组微卫星特征分析. 南京林业大学学报：自然科学版，36(2)：47－51.

许玉兰，蔡年辉，康向阳，等. 2012. EST-SSR 标记的开发及其在木本植物中的分布特点. 中国农学通报，28(4)：1－7.

胥　猛，李火根. 2008. 鹅掌楸 EST-SSR 引物开发及通用性分析. 分子植物育种，6(3)：615－618.

阎毛毛，戴晓港，李淑娴，等. 2011. 松树、杨树及桉树表达基因序列微卫星比对分析. 基因组学与应用生物学，30(1)：103－109.

杨　继. 2001. 植物多倍体基因组的形成与进化. 植物分类学报，39(4)：357－371.

庄瑞林. 2010. 我国油茶育种选育工作的历史回顾与展望. 林业科技开发，24(6)：1－5.

Berube Y, Zhuang J, Rungis D, *et al*. 2007. Characterization of EST-SSRs in loblolly pine and spruce. Tree Genetics & Genomes, 3(3)：251－259.

Cardle L, Ramsay L, Milbourne D, *et al*. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, 156(2)：847－854.

Christian S. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma, 109(6)：365－371.

Dieringer D, Schlötterer C. 2003. Two distinct modes of microsatellite mutation processes：evidence from the complete genomic sequences of nine species. Genome Research, 13(10)：2242－2251.

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

Feng S P, Li W G, Huang H S, *et al*. 2009. Development, application and cross-species/genera transferability of EST-SSR makers for rubber tree (*Hevea brasiliensis*). Molecular Breeding, 23(1)：85－97.

Gow J L. 2005. A high incidence of clustered microsatellite mutations revealed by parent-offspring analysis in the African freshwater snail Bulinus forskalii (*Gastropoda pulmonata*). Genetica, 124 (1): 77 - 83.

Iseli C, Jongeneel C V, Bucher P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proceedings of the International Conference on Intelligent Systems Molecular Biology, 138 - 148.

Jurka J, Pethiyagoda C. 1995. Simple repetitive DNA sequence from primates: compilation and analysis. Journal of Molecular Evolution, 40(2): 120 - 126.

Kalia R K, Rai M K, Kalia S, *et al.* 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica, 177 (3): 309 - 334.

Kantety R V, La R M, Matthews D E, *et al.* 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology, 48(5/6): 501 - 510.

Kashi Y, King D G. 2006. Simple sequence repeats as advantageous mutators in evolution. Trends in Genetics, 22(5): 253 - 259.

Lawson M J, Zhang L. 2006. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biology, 7 (2): R14 doi: 10.1186/gb-2006-7-2-r14.

Li Shuxian, Yin Tongming. 2007. Map and analysis of microsatellites in the genome of *Populus*: the first sequenced perennial plant. Science in China Series C: Life Sciences, 50(5): 690 - 699.

Li Y C, Abraham B, Fahima K T, *et al.* 2004. Microsatellites within genes: structure, function, and evolution. Molecular Biology Evolution, 21(6): 991 - 1007.

Li Y C, Korol A B, Fahima T, *et al.* 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology, 11(12): 2453 - 2465.

Lue N F, Buchman A R, Kornberg R D. 1989. Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element in vitro. Proceedings of the National Academy of Sciences of USA, 86 (2): 486 - 490.

Martienssen R A, Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science, 293(5523): 1070 - 1074.

Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics, 30(2): 194 - 200.

Sharma R K, Bhardwaj P, Negi R, *et al.* 2009. Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). BMC Plant Biology, 9(1): 53.

Thiel T, Michalek W, Varshney R K, *et al.* 2003. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, 106(3): 411 - 422.

Varshney R K, Thiel T, Stein N, *et al.* 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Molecular Biology Letter, 7(2A): 537 - 546.

Varshney R K, Graner A, Sorrells M E. 2005. Genic microsatellite markers in plants: features and applications. Trends in Biotechnology, 23(1): 48 - 55.

Weber J L. 1990. Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics, 7(4): 524 - 530.

Yadav H K, Ranjan A, Asif M H, *et al.* 2011. EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species genera. Tree Genetics & Genomes, 7(1): 207 - 219.

Zhang D L, Stack L, Zhang R Q, *et al.* 2008. Teaoil camellia—eastern "olive" for the world. Acta Horticulturae, 769: 43 - 48.

(责任编辑 徐 红)