

DOI:10.13275/j.cnki.lykxyj.2022.006.010

# 基于核桃参考基因组的 SSR 位点鉴定分析和单态性标记开发

贺君星, 马庆国\*, 裴东, 张俊佩

(林木遗传育种国家重点实验室, 国家林业和草原局林木培育重点实验室, 中国林业科学研究院林业研究所, 北京 100091)

**摘要:** [目的] 为了阐明核桃 (*Juglans regia* L.) ( $2n=2x=32$ ) 全基因组不同染色体上 SSR 位点数量及其分布特征, 开发并验证单态性 SSR 引物。[方法] 基于核桃品种‘中牧查一’的高质量参考基因组序列图谱, 使用 MISA 软件筛选并分析 SSR 位点, 利用 Primer 3.0 进行引物设计, 通过电子 PCR 进行引物多态性分析, 筛选并合成单态性 SSR 引物并通过真实 PCR 实验验证其有效性。[结果] (1) 在全基因组 16 条染色体上鉴定得到了共计 357 629 个 SSR 位点, 分布密度为  $662.28 \text{ SSRs}\cdot\text{Mb}^{-1}$ , 其中, 10~30 bp 长度的短序列占 95.00% 以上, 优势重复单元以 A/T 碱基为主; 不同染色体上 SSR 位点数量差异较大, 其中, Chr1 染色体上 SSR 位点数量最多, Chr16 染色体上最少, SSR 重复单元的数量及种类与染色体长度呈极显著正相关, 进一步统计分析获得的 644 种稀有 SSR 单元中六核苷酸重复单元占多数; (2) 基于不同染色体上 SSR 重复单元的类型和数量构建分布矩阵, 聚类分析发现, 以遗传距离 0.075 为阈值可将所有染色体分为 4 组, 其中, 第 4 组中成员最多 (11 条), 而第 1 组中仅有 Chr10 染色体, 总体看, Chr10 染色体自成一个主枝, 显示其可能经历了较为保守的进化过程; (3) 利用 SSR 位点侧翼的保守序列设计出 SSR 引物 303 009 对, 通过电子 PCR 筛选并随机合成了 32 对单态性引物进行验证, 其中, 30 对 (93.75%) 在 6 个核桃品种中扩增出了清晰的目标产物, 28 对 (87.50%) 的 PCR 扩增结果与电子 PCR 评估结果一致。[结论] 本研究鉴定了‘中牧查一’核桃参考基因组不同染色体序列中的 SSR 位点, 发现其数量和重复类型变化较大, 且与染色体长度呈极显著正相关, 单碱基重复的 SSR 是最常见的类型。建立了电子 PCR 与传统方法联用的引物开发流程, 同时验证了其有效性, 为核桃 SSR 引物的个性化快速开发提供了有效策略。筛选获得的 28 对单态性引物可为分子辅助育种中杂交子代“私生检测”等研究提供科学借鉴与参考。

**关键词:** 核桃; 微卫星; 分布规律; 分子标记

**中图分类号:** S664.1

**文献标志码:** A

**文章编号:** 1001-1498(2022)06-0089-12

核桃 (*Juglans regia* L.) 是在全世界范围内广泛栽培利用的重要经济树种, 中国是核桃的原产地之一。核桃种质资源极为丰富, 在我国华北、西北、中南、华东、四川以及西藏东南等地区均有分布, 种植历史达 3 000 余年<sup>[1]</sup>, 截至 2020 年底, 我国核桃种植面积达 782.22 万  $\text{hm}^2$ , 总产量 479.59

万 t, 居世界首位<sup>[2]</sup>。核桃雌雄同株异形异花, 育种周期长, 多数品种亲和力和很强。20 世纪 60 年代以来, 我国的引种和杂交育种工作不断推进, 涌现了大量的自主知识产权品种, 现有的国家审、认定良种 17 个, 生产中使用较多的省审定良种和有效期内的认定良种约计 239 个, 遍布全国各地<sup>[3]</sup>, 丰

收稿日期: 2022-04-15 修回日期: 2022-05-26

基金项目: 国家科技基础资源调查专项子课题 (2019FY100801\_02); 福建省属公益类科研院所基本科研专项 (2020R1009008); 核桃产业国家创新联盟 (NAWI)。

作者简介: 贺君星, 在读硕士。主要研究方向: 经济林生物技术。E-mail: 2648496184@qq.com

\* 通讯作者: 马庆国, 博士, 助理研究员。主要研究方向: 特色经济林种质资源。E-mail: mqg@caf.ac.cn

富的品种资源有力地支撑了我国核桃产业发展。

微卫星或简单重复序列 (SSR) 标记以其数量丰富、多态性高、重复性好、对基因组覆盖度广、易于扩增以及共显性等显著特点, 成为目前遗传学研究和植物品种鉴定中使用最便捷、应用最广泛的标记系统之一<sup>[4-6]</sup>, 国际植物新品种保护联盟 (UPOV) 生化 and 分子生物技术工作组 (BMT) 也将其作为用于品种鉴定和分子身份证构建的最佳分子标记<sup>[7]</sup>。SSR 通常是以 1~6 个核苷酸为重复单位的串联重复序列, 在基因组中广泛分布。在微生物<sup>[8]</sup>、植物<sup>[9-10]</sup>、动物<sup>[11]</sup> 和人类<sup>[12-13]</sup> 等不同物种中都开展过全基因组范围内的 SSR 引物开发工作。Woeste 等<sup>[14]</sup> 首次基于美国东部黑核桃基因组 DNA 开发了 SSR 标记, 这些引物在后续核桃属植物的遗传研究中也得到了应用<sup>[15]</sup>。Zhang 等<sup>[16]</sup>、Dang 等<sup>[17]</sup> 基于核桃 EST 序列各开发了 41 和 39 对 EST-SSR 引物。陈凌娜等<sup>[18]</sup>、Ikhsan 等<sup>[19]</sup> 和 Eser 等<sup>[20]</sup> 基于细菌人工染色体 (Bacterial artificial chromosome, BAC) 序列分别开发了 19、307 和 20 对 BES-SSR 引物。据不完全统计, 目前大约已开发出 770 对 SSR 引物, 多数是从 *J. regia* 这个种开发的, 为核桃的分子标记研究奠定了一定的基础<sup>[20]</sup>。但是, 由于核桃基因组的复杂性和越来越深入的研究工作对分子标记产生新的需求, 这些引物已经不能满足相关工作的需要, 而可用于子代纯度检测等分子辅助育种研究的核桃单态性 SSR 位点则未见报道<sup>[21]</sup>。

随着核桃基因组测序的开展和完善, 本课题组已组装完成并发表了一套染色体水平的高质量参考基因组<sup>[22]</sup>, 本研究基于该参考基因组序列分析其不同染色体上 SSR 位点的分布、重复单元数量及长度、稀有 SSR 碱基分布情况等特征, 利用电子 PCR 技术分析 SSR 引物的多态性, 并随机选取部分单态性 SSR 引物进行 PCR 实验验证, 旨在明确电子 PCR 方法在核桃 SSR 引物分析中的有效性, 有助于核桃 SSR 引物的快速批量化开发, 进而为核桃种质资源保护及开发利用、遗传研究提供支撑。

## 1 材料与方法

### 1.1 植物材料

选取 6 个核桃主栽品种用于 PCR 实验验证, 其名称和来源地等信息见表 1, 于生长季采其健

康、无病虫害、中等成熟度的叶片, 低温带回实验室, 立即提取其基因组 DNA。

表 1 供试核桃品种  
Table 1 Walnut cultivars used in this study

编号 No.	名称 Name	来源地或采样地点 Origin/Source
1	'强特勒' 'Chandler'	河北省易县
2	'辽宁1号' 'Liaoning 1'	辽宁省大连市
3	'清香' 'Qingxiang'	河北省易县
4	'香玲' 'Xiangling'	山东省泰安市
5	'扎343' 'Zha 343'	新疆阿克苏市
6	'新新2号' 'Xinxin 2'	新疆阿克苏市

### 1.2 数据来源

FASTA 格式的核桃品种 '中牧查一' 的染色体水平参考基因组序列下载自国家基因组科学数据中心 (CNCB-NGDC, 网址 <https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA002070>), 编号 PRJCA002070, 共包含 16 条染色体, 全长为 540 Mb, Contig N50 为 3.34 Mb<sup>[22]</sup>; 核桃品种 'Chandler'<sup>[23]</sup> 的染色体水平全基因组序列下载自 GigaDB (<http://gigadb.org/dataset/100735>)。使用 famap 和 fahash 软件将 2 套基因组数据转换为 hash 数据库, 以备引物多态性电子 PCR 模拟评估使用。

### 1.3 SSR 位点搜索和引物设计

利用 MicroSATellite (MISA, <http://pgrc.ipkga.terleben.de/misa/>) 程序对核桃全基因组不同染色体中的 SSR 位点进行检索和统计, 设定单核苷酸 (Mono-)、二核苷酸 (Di-)、三核苷酸 (Tri-)、四核苷酸 (Tetra-)、五核苷酸 (Penta-) 和六核苷酸 (Hexa-) 的最少重复分别为 10、6、5、5、5 和 5 次, 统计 SSR 位点的数量、长度和重复单元类型等信息, 分析不同染色体上各种类型 SSR 位点的频率分布。

利用 Primer 3.0 设计 SSR 引物, 采用的参数及筛选标准主要有: 引物长度为 18~28 bp, 20 bp 为佳; 产物长度 100~500 bp; 引物退火温度为 55~65 °C, 以 60 °C 为佳; 引物序列 GC 含量为 40%~60%, 以 50% 为佳; 避免引物二聚体、发夹结构及错配<sup>[6, 24-25]</sup>。

### 1.4 SSR 引物多态性的电子 PCR 检测

利用电子 PCR 程序中的 re-PCR 模块将筛选出的 SSR 标记在 1.2 中构建的基因组数据库中进

行模拟扩增, 主要参数为: re-PCR -S <hashfile> -n 0 -g 0 100-1000, 根据模拟扩增结果将引物分为单态和多态两类, 其中, 单态引物在基因组数据库中只能扩增出一个位点, 而多态引物可以扩增得到多个位点, 即如果扩增产物大小相差  $\geq 2$  bp, 则 SSR 被归类为多态性引物, 扩增产物大小相同则被视为单态, 而仅有 1 bp 差异的 SSR 位点则认为是不明确的, 并从分析中删除<sup>[26]</sup>。

### 1.5 单态性 SSR 引物的 PCR 验证

从每条染色体上随机选择单态性 SSR 引物各 2 对, 合成 TP-M13 引物 (上海生工), M13 尾巴序列为 TGTAACGACGGCCAGT。采用改良的 CTAB 法<sup>[15]</sup> 提取核桃叶片的基因组 DNA, 参照 Chen 等<sup>[27]</sup> 的方法进行扩增和毛细管电泳检测, 利用 GeneMarker v2.2.0 读取电泳条带, 然后使用 Excel 2016 软件进行数据统计和分析。

## 2 结果与分析

### 2.1 核桃全基因组 SSR 频率及总体分布特征

利用 MISA v2.1 软件对全长 540 Mb 的核桃全基因组序列中的 SSR 位点进行鉴定, 共得到 357 629 个 SSR 位点, 平均每 1.51 kb 出现 1 个 SSR 位点, SSR 序列总长度为 8 019 209 bp, 占基因组全长的 1.49%, 包括 4 005 种重复单元。其中, 单核苷酸重复单元占比最高, 达 58.93%; 二核苷酸到四核苷酸重复单元占比分别为 34.00%、5.19%、1.03%; 五核苷酸和六核苷酸重复单元的占比则不足 1.00%, 但碱基组合可选择性和随机性更强, 种类更加丰富, 且重复单元占自身核苷酸重复类型的比例分布更加均匀。如 ACTCCG/AGTCGG 占六核苷酸重复的比例为 8.96%, AAAAAT/ATTTTT 占比为 8.46% (图 1 a)。单核苷酸重复类型中 (A)<sub>n</sub> 最为丰富, 接近所有核苷酸重复比例的一半, 二核苷酸重复类型中 (AT)<sub>n</sub> 最多 (32.10%), 三核苷酸类型中 (AAT)<sub>n</sub> 最多 (16.50%), 而四核苷酸类型中 (AAAT)<sub>n</sub> (9.00%) 占比最高 (图 1 b)。无论是单碱基还是多碱基重复, 占比居前四位的重复基序中, 仅有 A、T 碱基出现, 说明核桃全基因组微卫星具有 A/T 丰富的特性。AGG、AAC、AGC、ACT、ACG 和 CCG 占比不足 1.00%, 可能属于稀有 SSR 单元的主要基序组成成员。核桃 SSR 序列长度在 10~297 bp 间变化, 平均为 73.30 bp, 以 10~30 bp 长度

的短重复序列为主 (95.93%), 而长度大于 30 bp 的仅占 4.07%, 不同长度的 SSR 序列所占比例存在较大差异, 其中, 10 bp 长度的 SSR 所占比例最大 (18.09%), 随着 SSR 序列长度的增加, 其占比呈下降趋势 (图 1 c)。

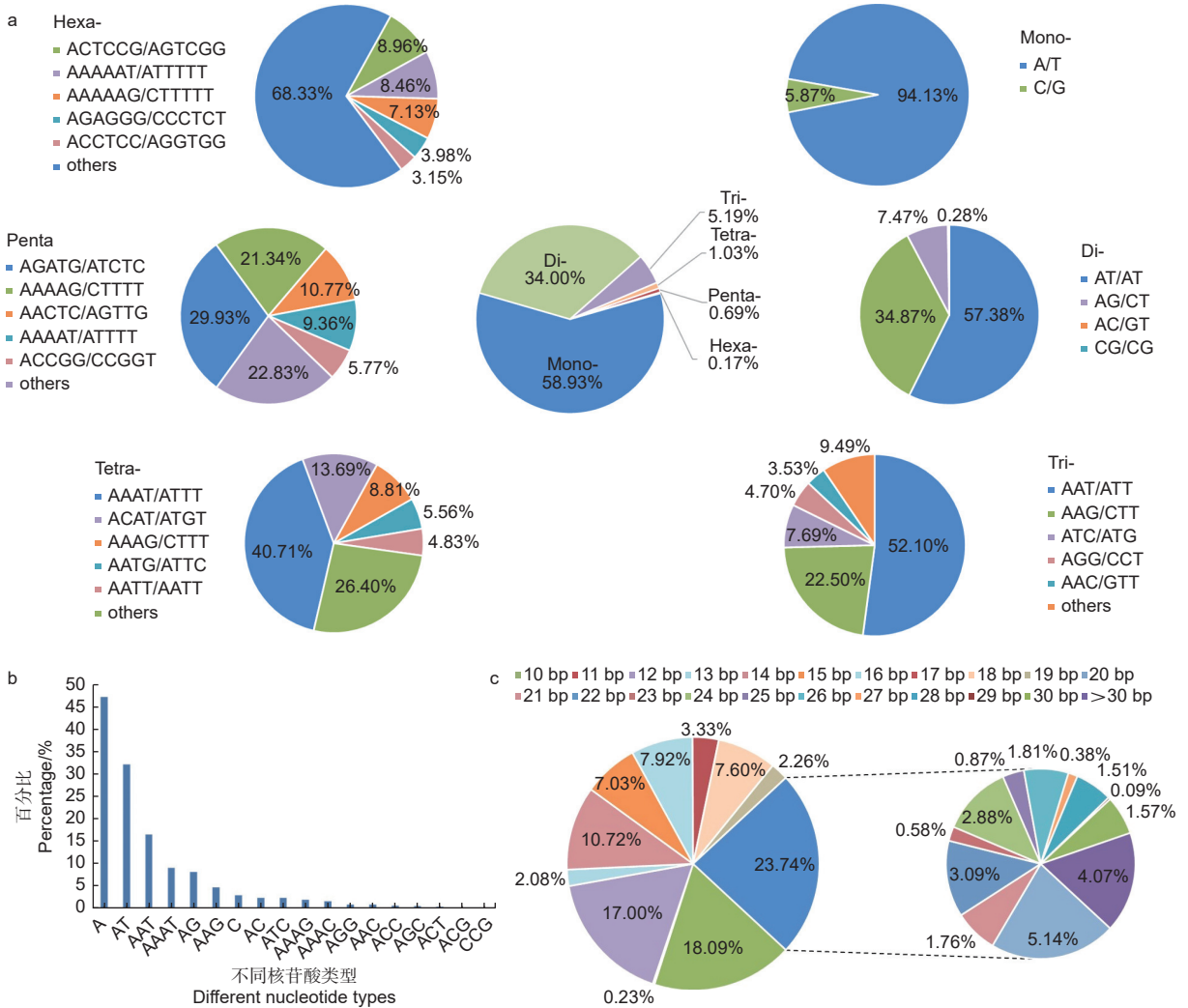
### 2.2 核桃不同染色体上 SSR 的数量及其分布

核桃参考基因组包含 16 条染色体, 即 Chr1~Chr16。不同染色体上 SSR 位点数量差异较大 (图 2 a), 其 Chr1 上数量最多 (34 749, 9.72%), Chr16 上数量最少 (13 666, 3.82%)。采用一元线性回归分析拟合发现, SSR 位点数量与染色体长度间的线性关系明显, 得到回归方程  $y=1\ 437x + 1\ 679\ 338$ , 决定系数为 0.96, 拟合效果较好, 即染色体长度越大, 相应的 SSR 位点数量越多 (图 2 b)。各染色体上 SSR 位点的分布密度相对稳定, 数量变化幅度为 1 427~1 685 个, 其中, Chr9 最低, Chr15 最高。同时, 进一步对 SSR 的数量、种类与染色体长度间的相关分析表明, 染色体长度与 SSR 数量 ( $r=0.982\ 0$ ,  $p<0.01$ )、种类 ( $r=0.900\ 3$ ,  $p<0.01$ ) 间均呈极显著正相关, 表明随着染色体长度的增加, 其 SSR 的种类与数量均呈增加趋势。

对 16 条染色体上不同重复类型的 SSR 位点数量进行相关性分析发现, 不同重复类型 SSR 位点间均呈极显著相关关系 ( $p<0.01$ ), 其中, 单核苷酸与二核苷酸位点数的相关系数 ( $r$ ) 最大 (0.986 3), 四核苷酸和六核苷酸的相关系数最小, 仅为 0.785 4 (表 2)。单核苷酸到四核苷酸间联系极紧密 ( $r>0.90$ ); 而 5~6 核苷酸重复类型的数量和其它重复类型间相关系数较低, 最大为 0.879 6, 最小仅为 0.785 4。

根据 SSR 位点不同重复类型和重复次数等, 构建不同染色体 SSR 位点  $280 \times 16$  阶分布矩阵, 并进行相关性分析建立相关系数矩阵, 通过相关系数矩阵, 对染色体进行聚类分析 (图 2 d)。以遗传距离 0.075 为阈值可将核桃 16 条染色体分为 4 组 (I~IV), 其中, 第 I 组只有 1 条染色体, 即 Chr10; 第 II 组包括 2 个成员, 即 Chr1 和 Chr3; 第 III 组包括 Chr14 和 Chr16 这 2 条染色体; 第 IV 组则囊括了其余 11 条染色体, 这个组又可分为 2 个亚组, 第 1 个亚组中有 Chr4、Chr7、Chr5 和 Chr11 这 4 条染色体, 第 2 个亚组包括 Chr6、Chr13、Chr2、Chr9、Chr15、Chr8 和





注: a. SSR 位点中不同类型重复单元的占比; b. 不同核苷酸类型 SSR 位点的数量分布; c. SSR 序列长度分布

Notes: a. The proportions of different SSR units; b. Distribution of different nucleotide types; c. Sequence length distribution

图 1 核桃全基因组 SSR 位点重复类型的分布特征

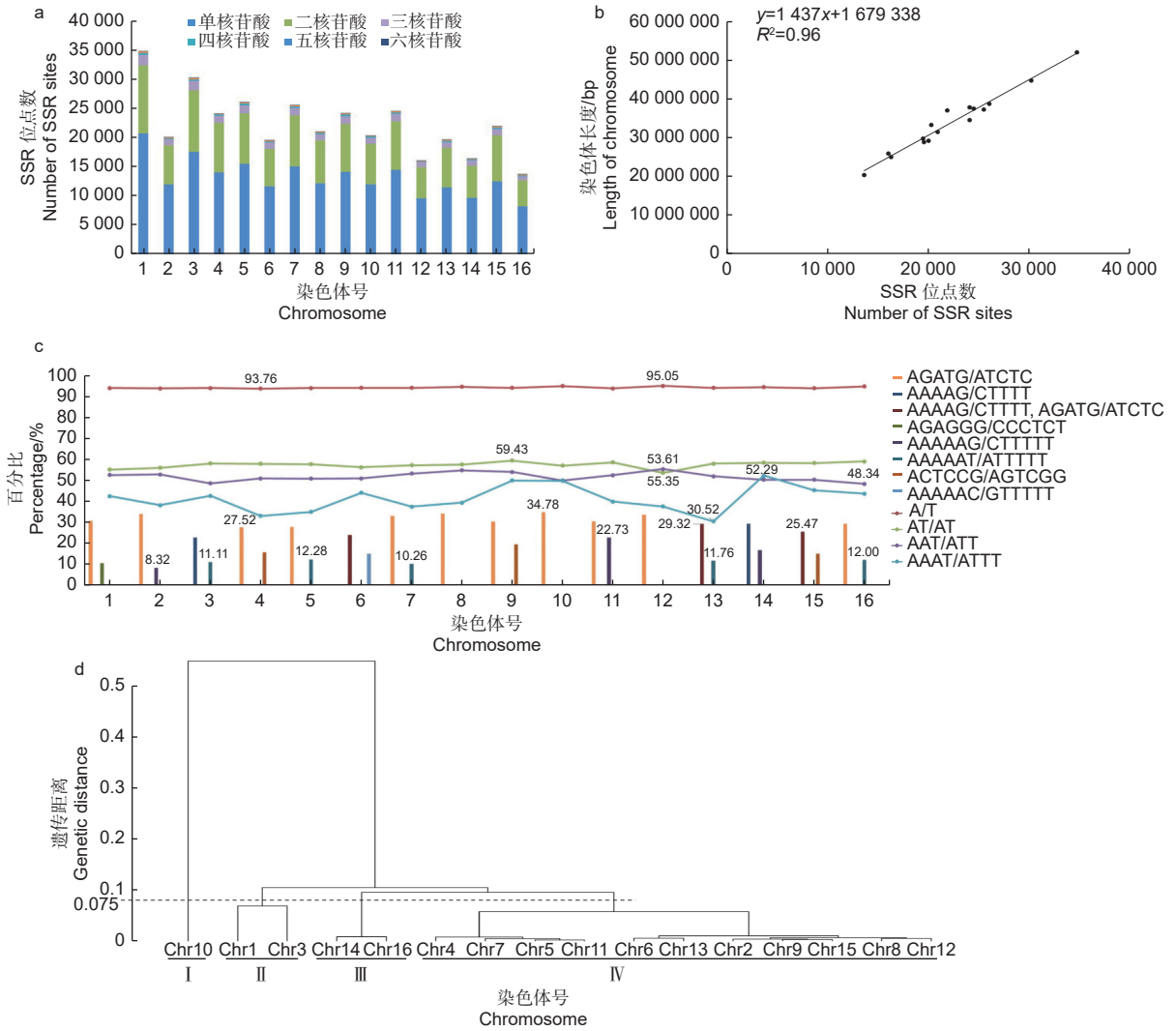
Fig. 1 Genome-wide distribution of SSR repeat types in walnut

Chr12 这 7 条染色体, 这 2 个亚组中的成员也是 SSR 分布模式最相似的染色体。总体上, 第 II、III 及 IV 组回溯到同一个主枝, 而第 I 组则归类于单独的一个主枝, 表明 Chr10 上 SSR 位点的数量、分布和频率等与其他 15 条染色体相比差异较大。

### 2.3 核桃不同染色体上 SSR 的重复单元类型和稀有 SSR 位点

不同染色体间重复单元数及重复碱基的种类存在一定的差异 (图 2 c), Chr1 染色体上最多 (406 种), 其次为 Chr3、Chr5 和 Chr11, 而 Chr16 染色体上最少 (188 种), 其中, 单核苷酸重复 SSR 序列均以 A/T 重复单元为主, 其含量在不同染色体上相对稳定, 其中, Chr4 上最低 (93.76%), Chr12 上最高 (95.05%); 二核苷

酸重复 SSR 序列均以 AT/AT 重复单元为主, 所占比例在 53.61% (Chr12) ~ 59.43% (Chr9) 间变化; 三核苷酸重复 SSR 序列的重复单元不同染色体均为 10 种重复单元类型, 主导单元均为 AAT/ATT, 其所占百分比在 48.34% (Chr16) ~ 55.35% (Chr12) 间变化; 四核苷酸 SSR 序列的主要重复单元是 AAAT/ATTT, 其所占百分比在 30.52% (Chr13) ~ 52.29% (Chr14) 间变化; 五核苷酸重复 SSR 序列在不同染色体上的重复单元类型数量为 18 (Chr12、13、16) ~ 36 (Chr3) 种, 其中, Chr3 (22.73%) 和 Chr14 (29.29%) 上的 SSR 主导单元为 AAAAG/CTTTT, Chr6 (23.91%)、Chr13 (29.32%) 和 Chr15 (25.47%) 染色体上为 AAAAG/CTTTT 和 AGATG/ATCTC 且



注: a. 不同重复 SSR 单元的含量; b. 染色体长度与 SSR 位点数一元回归方程; c. 主要重复类型 SSR 的分布; d. 基于 SSR 分布特征的核桃基因组不同染色体聚类图

Notes: a. The contents of different SSR units; b. One element regression equation of chromosome length and SSR locus; c. Distribution of main repeat types; d. Dendrogram of different chromosomes based on SSR loci characteristics

图 2 核桃基因组不同染色体 SSR 位点分布

Fig. 2 Distribution of SSR loci in different chromosomes of walnut genome

表 2 核桃不同重复类型 SSR 相关性分析

Table 2 Correlation coefficient of different SSR repeat types in Walnut

重复单元 Repeat units	单核苷酸 Mono-nucleotide	二核苷酸 Di-nucleotide	三核苷酸 Tri-nucleotide	四核苷酸 Tetra-nucleotide	五核苷酸 Penta-nucleotide	六核苷酸 Hexa-nucleotide
单核苷酸 Mono-nucleotide	1.000 0					
二核苷酸 Di-nucleotide	0.986 3**	1.000 0				
三核苷酸 Tri-nucleotide	0.985 6**	0.969 0**	1.000 0			
四核苷酸 Tetra-nucleotide	0.928 6**	0.938 7**	0.911 8**	1.000 0		
五核苷酸 Penta-nucleotide	0.879 6**	0.850 5**	0.870 2**	0.862 1**	1.000 0	
六核苷酸 Hexa-nucleotide	0.845 6**	0.796 3**	0.822 9**	0.785 4**	0.798 5**	1.000 0

注: \*\*表示  $p < 0.01$

Note: \*\*stand for  $p < 0.01$

占比相同, 其余 11 条染色体的 SSR 主导单元均为 AGATG/ATCTC, 所占百分比为 27.52% (Chr4) ~ 34.78% (Chr10); 六核苷酸重复单元的类型为 13 (Chr14) ~ 43 (Chr1) 种, 且大部分染色体都以 AAAAAT/ATTTTT 或 AAAAAG/CTTTTT 单元为主, 所占比例在 8.32% (Chr2) ~ 22.73% (Chr11) 间变化。

核桃基因组中存在 644 种稀有 SSR 单元 (即

该重复单元仅在单一染色体中存在), 不同染色体上存在的 SSR 稀有单元数存在较大差异, 其中, Chr4 中最少, 为 22 种, Chr5 中最多, 达 64 种 (表 3)。稀有 SSR 单元为 4~6 核苷酸重复, 其中以六核苷酸最多 (426 种), 四核苷酸最少 (38 种), 说明 SSR 单元组成的核苷酸越多, 其所占比例越低, 成为稀有 SSR 单元的概率越大。

表 3 核桃参考基因组不同染色体的稀有 SSR 重复单元

Table 3 Rare SSR units of different chromosomes in Walnut reference genome

染色体 Chromosome	四核苷酸重复 Tetra-nucleotide repeats	五核苷酸重复 Penta-nucleotide repeats	六核苷酸重复 Hexa-nucleotide repeats
Chr1	ATCA, CTGA, GTTC	AACCC, AATAC, ACTAA, CATAA, CGAGA, CGGGC, CTCGT, GAAAT, GAGAG, GATAT, GATTA, GGAAA, GGCTC, GGTTT, GTTTT	AAATTA, ACAGCA, ACCCCT, ACTCCA, AGTGGT, ATTTGG, CCAGGT, CCCTCT, CCGCCC, CCTACC, CGAAGA, CGGCTC, CTCAAA, CTCTGC, CTCTGG, CTGCGT, CTTCCC, CTTCC, GAAAA, GAAGCC, GACGCA, GATCGT, GCCAAA, GGAGGT, GGCCTG, GTCGGA, GTGAGA, GTTGGT, TCATCT, TCTCCC, TGGGCA, TTCACT, TTGTTT
Chr2	CGTT, TGAT	ACATG, AGATC, ATTTG, CCACT, CCTTC, CTAGC, GGGTG, GTCGG, TGGTC	AATATT, AGGAAG, ATGGAT, CAGCCA, CCATCA, CCCTCA, CCGACT, CGCAGT, CGGGAGT, CGGGTG, GAGAGT, GAGGAA, GAGTTC, GCCTCC, GGTGGA, TCAAAC, TGGGTA, TTGATG
Chr3	AGCG, CGGA, GTCA, TTCC	AACCA, ACAGA, ACATC, ATCCC, CCGAT, CGCCA, CTCTG, GAATT, GATTC, GTGCA, TCTCC, TGAAT, TGATG, TGCTG, TGCTT, TGTCG	AAATAT, AATTAA, AATTCA, ACACCC, ACCCTT, AGCAGG, ATATAA, ATCTCT, ATTCCA, CAATCA, CAGCTC, CATCAC, GCACAC, GCAGAG, GGAGAG, GTGAAC, TATCTT, TGCCCC, TGTTGC, TTGAAG, TTTTAT
Chr4	GCCA, GTCC, TCAA	AAGCC, ACCTC, AACT, CATCA, CCAGC, GCTCG, TGAAG, TTGAT	ATAGAT, CATGCA, CATTGC, CCGCCA, CCTCTA, GACAAA, GCCTCA, GGGGCT, TCAAGG, TCTTGG, TTCCTA
Chr5	AACT, GATT, GCGG, GGGT	ATAAG, ATGTG, CAAGC, CATAT, CTGGA, GAAAC, GATTT, GGTGG, TCCCT, TGAGG, TGTTA, TTAAT	AAAGGA, AAATAC, AACCCCT, AAGGTC, AATTCC, ACACAT, ACATAC, ACTGAT, AGAAGG, AGCCTG, AGGAGA, AGTCGG, ATAAAA, ATGCAT, CCAAAA, CCCCTT, CCTCCC, CGACTC, CTC AAG, CTCATT, CTCCT, TCCCGA, CTGATA, CTGTCT, GAAATG, GACTCC, GCCACT, GGC GGG, GGGCAG, GGGTGG, GTAGGT, GTTGCT, TACCCC, TATTAG, TATTTT, TCCACC, TCGAAC, TCGGAG, TCTCTG, TGAGAG, TGCTCG, TGGCGA, TGGGCC, TGGTAG, TGGTTC, TTCTTT, TTTGGT, TTTTAT
Chr6	CCTG, GCTC, GTAG	AGTCT, CACCA, CCCC, CCGCA, CTACC, GCACT, GCCGG, GCTGA, GGGGT, GGTTC, GTGAA, TATTA	AATCAG, ACAAAA, AGACAG, AGTAGG, AGTCGG, ATAGGC, ATTTTT, CCAACA, CCTCCG, CCTTTG, CGGAGG, CTCAT, CTCCGA, CTCCGG, CTCCTA, GACCTC, GCCTCT, GCTTCT, GGATCA, GGGTGT, GGTTAG, GTAGGT, GTTTTT, TATCAT, TCCGAC, TCGGAG, TCTCCA
Chr7	GGCT, TCGC, TGGG	AACAT, AGTAG, CCTAC, CTCAC, GATAA, GCCCG, TGAGC, TTGAA, TTGTT	ACCCAC, ACTAAT, ACTCCT, AGAAAC, AGAGAT, ATACAC, ATACTA, ATATAC, ATGTAT, ATTTG, CAGAAT, CCACAC, CCACCT, CCACG, CCCAAC, CCCTGT, CTTGAG, CGGTGG, CTCTCC, GACTCC, GAGTGA, GCGTGC, GGAGTG, GGGAGA, GGGTGT, TCTATC, TGGAGG, TGGTTG, TTATTT, TTGGGT
Chr8	CACT	AAGGA, ACAAC, AGATA, ATTTT, CAAAT, GTATA, GTGTG, GTTAG, TAATT, TCTGA, TTTGA	AGGCTC, AGGGTT, ATGGCA, ATTGTC, CCCCCA, CCCCTT, CCTCCG, CGGACT, CTCCGA, CTCTCC, CTCTTG, GAGCTA, GCAAAC, GGCTGT, GGGCTA, GGGTTA, GTGTAT, TAATTA, TCTTCC, TCTTTT, TGAGCA, TGTCCC, TGTCGG, TTCGCC
Chr9	CCTA, TCCG, TGAC	ACCCG, ATGGG, CCACT, CTAAT, GAGGT, GCAAT, GCCAT, GCCCC, GGCAT, GGGCT, GGGGA, GGTAA, TCCAG, TCGGC, TCTTA	ACGAAC, AGAGAC, AGCAGA, AGTCGG, CAACAC, CCCTCC, CTCCAA, CTCCCT, CTCCGA, CTCTAT, CTCTCA, CTTGGT, GAATAT, GACTCC, GAGGGA, GCAAAA, GGCCCA, GGTAGG, GTCGAG, TATATG, TATTAC, TCAGTT, TCGGAG, TCTGTC, TCTTTT, TGGTCC
Chr10	ACTG, CCGG, CTTG, TGGT	AGTCG, CATAG, CGTCC, CTTCC, GATCG, GCCCA, TCATG, TTAGC, TTGTG	AAGCCA, AAGTGG, AATTTG, ACACAT, ACCCGC, ACCCTC, AGGGAT, ATACT, CATATA, CATCGT, CCCCTC, CCCTCC, CCTCCA, CCTCTT, CGACTC, CGGAGG, CTGCTC, GACTCC, GAGGGA, GAGGTG, GCCCCT, GGTAG, GTGGGC, TAAAAA, TATATG, TATTTG, TCCGGT, TCGGAG, TGATCA, TGGAGG, TGGCAG, TTATTT, TTCTCC, TTGGGG

续表 3

染色体 Chromosome	四核苷酸重复 Tetra-nucleotide repeats	五核苷酸重复 Penta-nucleotide repeats	六核苷酸重复 Hexa-nucleotide repeats
Chr11	AGTC, CCGC, GGCC	AAGCA, ACCAA, CATCC, CCCGG, CCGCA, CGGGG, CAAGA, GATGT, GGAGG, GTATT, TGCTC	AAAGAA, AAATAC, AATACT, AGAGAT, AGAGGG, AGATAG, AGCTTA, ATTAAT, CACATG, CCACT, CCCCCT, CTCCAT, CTCCCCT, CTCCGA, CTGGTG, GACCCA, GATTTCG, GATTTT, GGAAGG, GGGTTA, GTAGGG, TATTAG, TCCGAC, TGAGCA, TGGAGG, TGTTTT, TTCCCC, T TCTTT, TTTAAC
Chr12	GCCC	AAACG, AAGTG, AATGC, ATTAC, CAATA, TTCCT	AAGAAC, ACCTCC, AGTCGG, ATAAAA, ATCAAC, ATTTTT, CAGGCT, CCAAGC, CCTCCA, CCTCTC, CTGCA, CTGTG, CTATTG, CTGGCT, CTTTTA, GAAGGA, GAGATG, GAGCTC, GATGAA, GATGGA, TATTAC, TATTAG, TCCGAC, TCTCAC, TCTGTC, TACTT, TTCTTT, TTTATT, TTTCTC, TTTGAT
Chr13	CGTG, GGCA	AGAGA, AGCCG, CAGTA, GGCCC, TATTG, TCACC	AAATTG, AACCAA, AAGAGA, ACGGCG, AGACAG, AGGGGG, AGTAGG, ATAAAA, ATCTAA, CAACAC, CAGCCC, CCTACA, CGACTC, CGCACA, CGGAGG, CTCCGA, CTTGGG, GAATTT, GCTGTG, GGCACA, GTGCGT, TCAAAA, TGAGCC, TGCATG, TTCATG, TTGATT, TTTAAA
Chr14	ACTT	AGAAG, ATCCG, CACAT, CAGTG, TAGTA	AGTTCA, CACACG, CCCTAA, CACTG, CTCTAT, CTTTTG, GAAGAG, GCACCG, GGGTTA, TAAAA, TCTTTT, TTATAT
Chr15	-	AACCT, ACACG, ACTAC, ATTGC, CAATC, GAGGA, GAGTT, TGTA	AAGCCC, AATGCC, ACCTCC, ACTACC, ACTCCG, AGGAGT, AGTAGG, AGTCGG, ATACAT, CCTCCG, CGAGCT, CTCCAC, CTCCCCT, CTCCGA, CTCTCG, GAGGGG, GCCCAA, GCGAGA, GTAGGA, GTGCTC, GTGGTC, TATTAC, TCCACC, TCCCTC, TCCGAC, TCGGAG, TCTTTT, TGATCG, TGGCTG, TTAGGG, TTTATT, TTTCTT, TTTGTG
Chr16	CTAC	ACTCG, ATAAC, GCATT, GGGTC, TAGTG, TTCAC	AAAATT, ACATAT, ATACGA, ATTAGT, ATTTGA, ATTTTT, CACGCA, CCCCCT, CTTTTT, CGGAGG, CTCCGA, CTTTTT, GACTCC, GATTTA, GCTCCG, TATATG, TCATCC, TCATCG, TCCCTG, TCTCTT, TCTTTT, TGGGTG, TTAGTA

## 2.4 核桃全基因组单态性 SSR 引物开发与 PCR 验证

利用 Primer 3.0 软件根据 SSR 位点侧翼的保守序列, 从 357 629 个 SSR 位点中共设计出 303 009 对 (91.51%) SSR 引物, 包括 6 种完全微卫星 (258 024, 85.15%)、不完全微卫星 (1 688, 0.56%) 和复合型微卫星 (43 297, 14.29%) 等 3 种类型。然后, 利用 re-PCR 将 2~6 核苷酸重复的完全型 SSR 引物比对到基因组 hash 数据库, 通过电子模拟扩增评价其多态性, 根据在不同基因组中的电子模拟扩增条带将其分为单态和多态两类, 条带大小差异仅 1 bp 的引物将被弃用。电子模拟扩增分析发现, 不同染色体上的单态性标记最少为 2 295 个 (Chr16), 最多为 10 881 个 (Chr3), 长重复单元 (>30 nt) 的引物中多态性引物的比例要高于短重复单元的引物, 这与 Biswas 等<sup>[26]</sup> 在甜橙基因组中的研究结果类似。

为了验证电子模拟扩增结果的可靠性和新开发的 SSR 标记的有效性, 从经 re-PCR 评估的单态引物中随机均匀地选择三碱基重复引物 32 对 (表 4), 合成 TP-M13 引物在‘强特勒 (Chandler)’等 6 个品种中扩增, 并利用毛细管电泳技术进行检测

(图 3)。所选 32 对引物中除 CAF36 和 CAF350 以外, 其余 30 对 SSR 引物 (93.75%) 均可以在供试样品中扩增出清晰的目标产物, PCR 扩增结果与电子模拟评估结果一致性较好。此外, 30 对 SSR 引物中有 4 对 (CAF11、CAF129、CAF271 和 CAF364) 在供试样品之间表现出多态性。

## 3 讨论

### 3.1 核桃全基因组 SSR 分布特征

SSR 在基因组中的出现主要源于进化过程中的突变, 如滑链错配、一个或多个重复基序的插入缺失等, 因此, 特定数量和长度的 SSR 可以作为进化过程中遗传变异的指标<sup>[24]</sup>。本研究从‘中牧查一’核桃参考基因组 16 条染色体中共鉴定出了 357 629 个 SSR 位点, 其密度为 662.28 SSRs·Mb<sup>-1</sup>, 低于石榴 (1 294.62 SSRs·Mb<sup>-1</sup>)<sup>[6]</sup> 和枣 (872.60 SSRs·Mb<sup>-1</sup>)<sup>[28]</sup> 等树种, 高于亚麻 (225.3 SSRs·Mb<sup>-1</sup>)<sup>[29]</sup>、茶树 (216.88 SSRs·Mb<sup>-1</sup>)<sup>[30]</sup> 和花生 (392.45 SSRs·Mb<sup>-1</sup>)<sup>[31]</sup> 等植物, 而与杨树 (667.9 SSRs·Mb<sup>-1</sup>)<sup>[32]</sup> 在基因组上的研究结果相当。这些数据可能反映了不同物种本身基因组大小的差异以及基因组序列中碱基排列组合方式的随

表 4 32 对核桃单态 SSR 引物信息

Table 4 Information of 32 primers from walnut genome

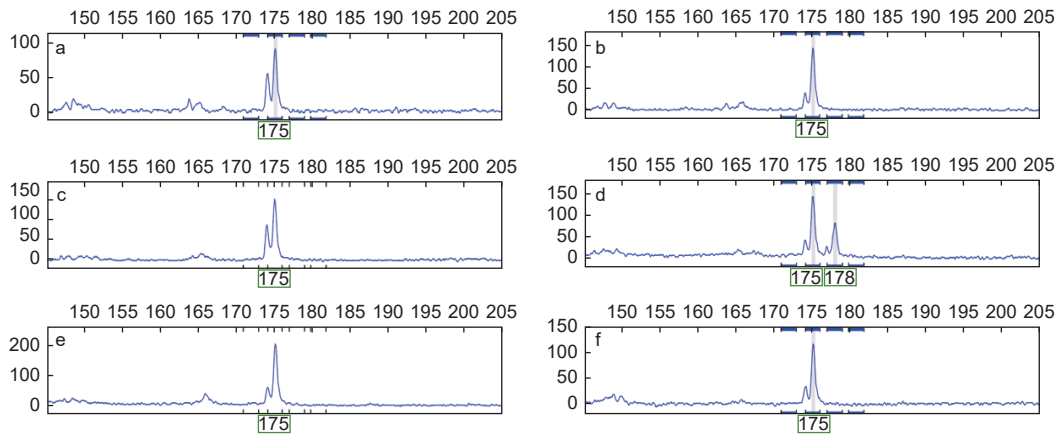
引物编号 Primer ID	染色体编号 Chromosome	重复基序 Repeat motifs	上游序列 (5'-3') Forward (5'-3')	下游序列 (5'-3') Reverse (5'-3')	退火温度 Tm/°C
CAF271	Chr1	(TTC) <sub>5</sub>	ATCGGATCGTTCGTCATGG	TTTGTGGCTTGTTTTGGCC	60
CAF100	Chr1	(GAG) <sub>5</sub>	GGCCAAGACATAACCCAGCT	AGATGGGTAATCAGGGGGCT	60
CAF364	Chr2	(TCT) <sub>6</sub>	GACTCCACCCCTTGTGTCATG	GTGAACAAGACCCGACCCAT	60
CAF169	Chr2	(ATG) <sub>5</sub>	GGGTGACGACGGTGACTATC	GATGCAGAAAGGCTCCCAGT	60
CAF36	Chr3	(TTC) <sub>5</sub>	TGGGTTTGCCTACCTGAGT	TGAAACTGATTGCACCCCA	60
CAF322	Chr3	(TGC) <sub>5</sub>	ATACCGCTGCTGTAGATGCC	TGGCCCTCATCACCTCATA	60
CAF293	Chr4	(AAG) <sub>6</sub>	GTGTTGAGAGGAAGGGCGAA	ACCCATGTGTGAGCATGCTT	60
CAF359	Chr4	(AGA) <sub>5</sub>	CTCGTCGTAGAAGTCGGAGC	TTGAGTCCGAGCACTTGCAT	60
CAF1	Chr5	(TCT) <sub>6</sub>	GCCAATCGGGGTAGAGGATG	GCGGAGGAGTTGATAGAGGC	60
CAF119	Chr5	(AGG) <sub>5</sub>	GAGAGCGAGAGCGTGATGG	CCTTGTGGAATCGTCAGCCT	60
CAF150	Chr6	(TAA) <sub>5</sub>	GCATGCAGCAGTCTTGGAAC	CTCCTTCTTCACTGCCAG	60
CAF186	Chr6	(TTA) <sub>6</sub>	ACGGAACAACGGCCTATGT	TTTCTTTTTCGCGGCTTCGG	60
CAF11	Chr7	(CCA) <sub>5</sub>	TTCCATCTCCTCTCCCCAG	TTATCTTCGCGATCGGCACA	60
CAF103	Chr7	(CTT) <sub>5</sub>	GCGACGGAGATATGCAGGAA	CGAGGGCAAAGTTGACTCT	60
CAF350	Chr8	(GCT) <sub>5</sub>	TAAGCTGGAACCCAATGGCC	AATCTCAACAGCAGCAGCAG	60
CAF288	Chr8	(TTC) <sub>5</sub>	AATGTCGTTCTGAGCACGT	TATCTAACGACCGCCGCTTC	60
CAF6	Chr9	(GAG) <sub>6</sub>	GAGGAGGAGGAGGAAGGGAG	TCTGCGCTCTCTGTTTTGT	60
CAF147	Chr9	(GAA) <sub>6</sub>	TGCTGGCTCCCTTCTAGAT	CCCCTCACACCTCATTTC	60
CAF16	Chr10	(AGA) <sub>6</sub>	TATCGCTGGCTTCCGATCC	GCGACATTGCAATCCAGGTC	60
CAF265	Chr10	(AAG) <sub>5</sub>	TGACGTCCTTGAGTTGTCG	AGTGCGTACATGACCAAGCA	60
CAF107	Chr11	(GAG) <sub>5</sub>	AGAAGCCGTGGAACCTGAGG	TTGGACAACGCGATCCAGAA	60
CAF108	Chr11	(TTA) <sub>6</sub>	CACACCCTTCCCGCTTTTG	TGTCACCGATCTGTCCGTTG	60
CAF15	Chr12	(AGA) <sub>5</sub>	GAGAGCGGCGAGTACTACAC	CCCTGCCTCCCTACTCTTCT	60
CAF208	Chr12	(TGA) <sub>5</sub>	GTCAGCCACGAACAACAC	GATCCTGGAAGTCGACAGCC	60
CAF151	Chr13	(AAG) <sub>5</sub>	AAACAGGGCGAGATTGGAGG	CCTTATCGCCATCGTCGTCA	60
CAF153	Chr13	(TCT) <sub>5</sub>	AATAGCAGAGCCCGTGATCG	CTGCCACAGTTGTAGAGCCA	60
CAF236	Chr14	(AGT) <sub>5</sub>	GCTGCAAAGCACACAATCCA	GCTTCTCCGCTGTTCTCT	60
CAF246	Chr14	(TAA) <sub>5</sub>	GCAACAGCAGGATTTTGCCA	AGGGGATGGGTCTGTAGCAT	60
CAF131	Chr15	(TTC) <sub>5</sub>	TGCTGTCAAAGGGGAATCAA	AGCTCTTCCAAGCACCATGT	60
CAF8	Chr15	(GCT) <sub>5</sub>	TCGCCAACGACACCATAACA	TTTCCAGCTGCTTTTGTGCG	60
CAF129	Chr16	(CTA) <sub>6</sub>	TCTCAGGCCCAACAACCATC	TAGAGAGAGATCCTCCGGCG	60
CAF536	Chr16	(CAG) <sub>5</sub>	TTCTCCACAAGTACCGCACC	CCTCTCTACTACCCAGCT	60

机性和变异性, 以往研究表明基因组大小会影响 SSR 的密度, 但并不是所有物种中二者之间都具有显著相关性<sup>[33-34]</sup>, 而本研究中核桃基因组 SSR 的数量、种类与染色体序列长度均呈极显著正相关。同时, SSR 的分布和密度变化很大, 也可能是由于搜索标准和数据库挖掘工具的不同导致的。

单碱基重复的 SSR 是核桃基因组每条染色体上最常见的类型, 2~3 碱基重复次之, 4~6 碱基重复则较少, 这在小麦<sup>[24]</sup>和甜橙<sup>[26]</sup>等物种的基因组中也有报道, 但与烟草<sup>[35]</sup>中以二碱基重复类型为主的情况不同。核桃的 SSR 基本组成在所有重

复类型中以 A 和 T 为主, 而 CG/CG 在二核苷酸重复序列中密度最低, 这与烟草<sup>[35]</sup>和四倍体野花生<sup>[36]</sup>等物种相似, 核桃 SSR 中最多的二核苷酸重复是 AT/AT, 其次是 AC/GT 和 AC/GT, 王玉龙等<sup>[36]</sup>报道, 四倍体野花生基因组中数量最多的重复单元依次是 A/T、AT/AT、AAT/ATT, AAAT/ATTT、AAAAT/ATTTT 和 AAAAAT/ATTTTT。Lu 等<sup>[31]</sup>对栽培花生的研究结果与此类似。尽管分布模式有所差异, 但对许多植物基因组而言, AAN、AAAN、AAAAN 和 AAAAN 比其他重复基序更常见。





注: a. '强特勒'; b. '辽宁 1 号'; c. '清香'; d. '香玲'; e. '扎 343'; f. '新新 2 号'

Notes: a. 'Chandler'; b. 'Liaoning 1'; c. 'Qingxiang'; d. 'Xiangling'; e. 'Zha343'; f. 'Xinxin 2'

图 3 CAF11 引物在 6 个核桃主栽品种中的毛细管电泳图谱

Fig. 3 The capillary electrophoresis patterns from primer CAF11 in 6 main cultivars of walnut.

研究指出, 单碱基或二碱基重复单元大量发生则表明该物种的进化水平较高<sup>[37]</sup>, 而核桃单核苷酸与二核苷酸数量的总和占有所有微卫星位点的 92.80%, 这可能说明核桃起源相对较晚同时容易发生 SSR 变异, 从而产生更多的碱基重复类型, 其中, 单碱基重复占 58.85%。有研究表明, 单碱基重复的数量随物种基因组大小不同而产生差异, 相比而言物种基因组增大, 其单碱基相对丰富增加<sup>[38]</sup>。Song 等<sup>[39]</sup>分析了 112 种植物 3 951 919 条基因序列中 SSR 的分布情况, 发现三碱基重复 SSR 数量超过 50%, 与核桃 SSR 数量分布比例有所不同, 这可能与密码子以 3 个碱基为组成单位有关。近期的一项研究将石榴基因组中的 SSR 分为三大类, 即 class I (> 30 nt), class II (20~30 nt) 和 class III (< 20 nt), 其中, 第一类 SSR 可能更易发生突变呈现不稳定状态<sup>[6]</sup>, 更早的研究中 Portis 等<sup>[40]</sup>、陈凌娜等<sup>[18]</sup>和 Temnykh 等<sup>[41]</sup>也强调了 SSR 片段长度在标记开发和育种中的重要性。本研究发现, 核桃 SSR 序列主要以 10~30 bp 的短重复序列为主, SSR 序列的出现频率呈现一定的规律性变化, 重复长度越长, 出现频率则越低。

### 3.2 核桃全基因组特异性单态 SSR 引物开发与验证

以往研究表明, 单核苷酸重复的 SSR 并不适宜于开发 SSR 引物<sup>[26, 42]</sup>, 因此, 本文在引物设计和开发时仅使用了 2~6 核苷酸重复的类型; 以往研究表明, 单态 SSR 标记可作为遗传分析和育种子代群体“私生检测”的有力工具<sup>[21, 43]</sup>, 也有将单态

性标记转换为多态性标记的报道<sup>[44-45]</sup>。在 re-PCR 检测到的单态性标记中, 二核苷酸基序 (77.28%) 最常见, 其中, Chr3 染色体含有最多的单态性标记, 其可能是核桃分子标记辅助育种的较好选择。随机选择 32 对单态性 SSR 引物在 6 个不同核桃主栽品种中进行 PCR 验证, 其中, 4 对预测为单态的引物扩增出了多个基因座, 这可能是由于电子 PCR 分析仪采用了 2 套基因组所致, 也有可能是核桃的全基因组尚有未完全揭示的序列, 而多达 28 对引物获得了与电子 PCR 分析相一致的 PCR 实验扩增结果。由此可见, 基于全基因组序列采用电子 PCR 结合传统分子标记筛选策略的方法进行 SSR 标记开发效果较好。

## 4 结论

核桃是世界范围广泛栽培的重要经济树种, 本研究从‘中牧查一’核桃参考基因组中鉴定了 357 629 个 SSR 位点, 这些位点在不同染色体上的数量和重复类型具有明显差异, 其中, 单碱基重复占比较高, A/T 碱基重复是其优势重复单元。在此基础上, 建立了联合应用电子 PCR 和传统引物筛选方法进行引物开发的新策略, 每条染色体上随机选取 2 对共计 32 对引物以 6 个核桃主栽品种为试材进行扩增, 结果显示其中 30 对引物的扩增效果较好, 28 对 (87.50%) 引物的扩增结果与电子 PCR 分析结果相一致, 从而验证了这一引物开发策略的有效性, 为核桃 SSR 引物的个性化快速开发提供了有效策略, 筛选获得的 28 对单态性引物

可为分子辅助育种中杂交子代“私生检测”等研究提供科学借鉴与参考。

## 参考文献:

- [ 1 ] 裴 东, 鲁新政. 中国核桃种质资源[M]. 北京: 中国林业出版社, 2011.
- [ 2 ] 国家林业和草原局. 中国林业统计年鉴(2020)[M]. 北京: 中国林业出版社, 2021.
- [ 3 ] 马庆国, 乐佳兴, 宋晓波, 等. 新中国果树科学研究70年—核桃[J]. 果树学报, 2019, 36 ( 10 ): 1360-1368.
- [ 4 ] Abd Ei-Moghny A M, Santosh H B, Raghavendra K P, *et al.* Microsatellite marker based genetic diversity analysis among cotton (*Gossypium hirsutum*) accessions differing for their response to drought stress[J]. Journal of Plant Biochemistry and Biotechnology, 2017, 26: 366-370.
- [ 5 ] Li J, Ye C. Genome-wide analysis of microsatellite and sex-linked marker identification in *Gleditsia sinensis*[J]. BMC Plant Biology, 2020, 20(1): 338.
- [ 6 ] Patil P G, Singh N V, Bohra A, *et al.* Comprehensive characterization and validation of chromosome-specific highly polymorphic SSR markers from Pomegranate (*Punica granatum* L.) cv. Tunisia genome[J]. Frontiers in Plant Science, 2021, 12: 645055.
- [ 7 ] UPOV. Guidelines for DNA-profiling: molecular marker selection and database construction (“BMT guidelines”) [EB/OL]. [https://www.upov.int/edocs/mdocs/upov/en/bmt\\_18/upov\\_inf\\_17\\_2\\_draft\\_2.pdf](https://www.upov.int/edocs/mdocs/upov/en/bmt_18/upov_inf_17_2_draft_2.pdf), 2019-10-10.
- [ 8 ] Mathema V B, Nakeesathit S, White N J, *et al.* Genome-wide microsatellite characteristics of five human *Plasmodium* species, focusing on *Plasmodium malariae* and *P. ovale curtisi*[J]. Parasite, 2020, 27(7): 34.
- [ 9 ] Dharajiya D T, Shah A, Galvadiya B P, *et al.* Genome-wide microsatellite markers in castor (*Ricinus communis* L.): Identification, development, characterization, and transferability in Euphorbiaceae[J]. Industrial Crops and Products, 2020, 151: 112461.
- [ 10 ] Zhu H, Song P, Koo D H, *et al.* Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis[J]. BMC Genomics, 2016, 17(1): 557.
- [ 11 ] Manee M M, Algarni A T, Alharbi S N, *et al.* Genome-wide characterization and analysis of microsatellite sequences in camelid species[J]. Mammal Research, 2019, 65(16): 359-373.
- [ 12 ] Subramanian S, Mishra R K, Singh L, *et al.* Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions[J]. Genome Biology, 2003, 4(2): 1-10.
- [ 13 ] Kelkar Y D, Tyekucheva S, Chiaromonte F, *et al.* The genome-wide determinants of human and chimpanzee microsatellite evolution[J]. Genome Research, 2008, 18(1): 30-38.
- [ 14 ] Woeste K, Burns R, Rhodes O, *et al.* Thirty polymorphic nuclear microsatellite loci from black walnut[J]. Journal of Heredity, 2002, 93(1): 58-60.
- [ 15 ] Wang H, Pan G, Ma Q G, *et al.* The genetic diversity and introgression of *Juglans regia* and *J. sigillata* in Tibet as revealed by SSR markers[J]. Tree Genetics & Genomes, 2015, 11(1): 804.
- [ 16 ] Zhang R, Zhu A, Wang X, *et al.* Development of *Juglans regia* SSR markers by data mining of the EST database[J]. Plant Molecular Biology Reporter, 2010, 28(4): 646-653.
- [ 17 ] Dang M, Zhang T, Hu Y H, *et al.* De novo assembly and characterization of bud, leaf and flowers Transcriptome from *Juglans regia* L. for the identification and characterization of new EST-SSRs[J]. Forests, 2016, 7(10): 247.
- [ 18 ] 陈凌娜, 马庆国, 张俊佩, 等. 核桃BES-SSR的开发及在遗传多样性分析中的应用[J]. 北京林业大学学报, 2014, 36 ( 6 ): 24-29.
- [ 19 ] Ikhsan A S, Topçu H, Sütyemez M, *et al.* Novel 307 polymorphic SSR markers from BAC-end sequences in walnut (*Juglans regia* L.): Effects of motif types and repeat lengths on polymorphism and genetic diversity[J]. Scientia Horticulturae, 2016, 213(1): 1-4.
- [ 20 ] ESER E, Topçu H, Kefayati S, *et al.* Highly polymorphic novel simple sequence repeat markers from Class I repeats in walnut (*Juglans regia* L.) [J]. Turkish Journal of Agriculture & Forestry, 2019, 43(2): 174-183.
- [ 21 ] Hama-Ali E O, Tan S G. Using monomorphic microsatellite markers in Oil Palm (*Elaeisguineensis* Jacq.) [J]. Research & Reviews: Journal of Botanical Sciences, 2014, 3(4): 1-6.
- [ 22 ] Zhang J P, Zhang W T, Ji F Y, *et al.* A high - quality walnut genome assembly reveals extensive gene expression divergences after whole - genome duplication[J]. Plant Biotechnology Journal, 2020, 18(9): 1848-1850.
- [ 23 ] Marrano A, Britton M, Zaini P A, *et al.* High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome[J]. GigaScience, 2020, 9(5): g100050.
- [ 24 ] Han B, Wang C, Tang Z, *et al.* Genomewide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.) [J]. PLoS ONE, 2015, 10: e0141540.
- [ 25 ] Xue H, Zhang P, Shi T, *et al.* Genome-wide characterization of simple sequence repeats in *Pyrus bretschneideri* and their application in an analysis of genetic diversity in pear[J]. BMC Genomics, 2018, 19(1): 473.
- [ 26 ] Biswas M K, Xu Q, Mayer C, *et al.* Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*) [J]. PLoS ONE, 2014, 9(8): e104182.
- [ 27 ] Chen L N, Dong R Q, Ma Q G, *et al.* Precocious genotypes and

- homozygous tendency generated by self-pollination in walnut[J]. *BMC Plant Biology*, 2018, 18: 323.
- [28] Fu P C, Zhang Y Z, Ya H Y, *et al.* Characterization of SSR genomic abundance and identification of SSR markers for population genetics in Chinese jujube (*Ziziphus jujuba* Mill.)[J]. *PeerJ*, 2016, 4(2): e1735.
- [29] Pan G, Chen A, Li J, *et al.* Genome-wide development of simple sequence repeats database for flax (*Linum usitatissimum* L.) and its use for genetic diversity assessment[J]. *Genetic Resources and Crop Evolution*, 2020, 67(4): 865-874.
- [30] Liu S R, An Y L, Li F D, *et al.* Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*) [J]. *Molecular Breeding*, 2018, 38(5): 59.
- [31] Lu Q, Hong Y, Li S, *et al.* Genome-wide identification of microsatellite markers from cultivated peanut (*Arachis hypogaea* L.) [J]. *BMC Genomics*, 2019, 20(1): 799.
- [32] Sonah H, Deshmukh R K, Sharma A, *et al.* Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium* [J]. *PLoS ONE*, 2011, 6: e21298.
- [33] Qi W H, Jiang X M, Du L M, *et al.* Genome-wide survey and analysis of microsatellite sequences in bovid species [J]. *PLoS ONE*, 2015, 10(7): e0133667.
- [34] Park S, Son S, Shin M, *et al.* Transcriptome-wide mining, characterization, and development of microsatellite markers in *Lychnis kiusiana* (Caryophyllaceae) [J]. *BMC Plant Biology*, 2019, 19(1): 1-12.
- [35] 童治军, 焦芳婵, 肖炳光. 普通烟草及其祖先种基因组SSR位点分析[J]. *中国农业科学*, 2015, 48(11): 2108-2117.
- [36] 王玉龙, 黄冰艳, 王思雨, 等. 四倍体野生种花生 *A. monticola* 全基因组 SSR 的开发与特征分析 [J]. *中国农业科学*, 2019, 52 ( 15 ): 2567-2580.
- [37] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: surveys and analysis [J]. *Genome Research*, 2000, 10(7): 967-981.
- [38] Karaoglu H, Li C M Y, Meyer W. Survey of simple sequence repeats in completed fungal genomes [J]. *Molecular Biology and Evolution*, 2005, 2(3): 639-649.
- [39] Song X M, Yang Q H, Bai Y, *et al.* Comprehensive analysis of SSRs and database construction using all complete gene-coding sequences in major horticultural and representative plants [J]. *Horticulture Research*, 2021, 8(1): 122.
- [40] Portis E, Lanteri S, Barchi L, *et al.* Comprehensive characterization of simple sequence repeats in eggplant (*Solanum melongena* L.) genome and construction of a web resource [J]. *Frontiers in Plant Science*, 2018, 9: 401.
- [41] Temnykh S. Computational and experimental analysis of microsatellites in Rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential [J]. *Genome Research*, 2001, 11(8): 1441-1452.
- [42] Cavagnaro P F, Senalik D A, Yang L, *et al.* Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.) [J]. *BMC Genomics*, 2010, 11: 569.
- [43] Nazareno A G, dos Reis MS. The same but different: monomorphic microsatellite markers as a new tool for genetic analysis [J]. *American Journal of Botany*, 2011, 98(1): e265-e267.
- [44] Ince A G, Karaca M, Onus A N. CAPS-microsatellites: use of CAPS method to convert non-polymorphic microsatellites into useful markers [J]. *Molecular Breeding*, 2010, 25(3): 491-499.
- [45] Karaca M, Ince A G. New non-redundant microsatellite and CAPS-microsatellite markers for cotton (*Gossypium* L.) [J]. *Turkish Journal of Field Crops*, 2011, 16(2): 172-178.

# Genome-Wide Analysis of Microsatellite Loci and Specific Monomorphism SSR Marker Development in Walnut (*Juglans regia* L.) Reference Genome

HE Jun-xing, MA Qing-guo, PEI Dong, ZHANG Jun-pei

(State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of the State Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China)

**Abstract:** [Objective] To identify the number and distribution of SSR loci on different chromosomes in the whole genome of walnut (*Juglans regia* L.,  $2n = 2x = 32$ ), and to develop and validate the monomorphic SSR primers. [Method] In this study, walnut whole genome sequencing data were used as experimental materials, and the whole genome microsatellites were screened and analyzed by bioinformatics software MISA. Primer 3.0 was employed to design monomorphic SSR primers. SSR primers were evaluated by electronic PCR and some of the monomorphic ones were synthesized randomly to detect their usefulness and verify the effectiveness of the method. [Result] (1) A total of 357 629 SSR loci were identified in the walnut genome, with a distribution density of 662.28 SSRs/MB. The dominant repeat units were mainly A/T bases, showing significant base preference. These SSR sequences were mainly short sequences with a length of 10~30 bp, up to more than 95.00%. The number of SSR loci on different chromosomes varied greatly. Among them, the number of SSR loci on chromosome 1 was the largest, and the number of SSR loci on chromosome 16 was the least. The number and type of SSRs showed positively correlated with chromosome sequence length. Most of the 644 rare SSR units were hexa-nucleotides. (2) Based on cluster analysis, all the 16 chromosomes could be divided into 4 groups, of which the number of members in group 4 was the most (11), and there was only chromosome 10 in group 1. In general, chromosome 10 forms a main branch, indicating that it may have experienced a relatively conservative evolutionary history. (3) 303 009 pairs of SSR primers were designed by using the conservative sequence flanking the SSR locus. And then 32 pairs of monomorphic primers clarified by electronic PCR were randomly screened and synthesized for wet PCR experiments, of which 30 pairs (93.75%) were amplified in 6 walnut varieties. The PCR amplification results of 28 pairs (87.50%) were consistent with that of electronic PCR. [Conclusion] In this study, SSR loci in different chromosome sequences of 'Zhongmucha-1' walnut reference genome are identified. Their amounts and repeat types are found to be highly variable among different chromosome sequences and show a highly significant positive correlation with chromosome length. Mono-nucleotide repeat SSRs are the most common type. A novel protocol combining electronic PCR and traditional screening methods are established and validated, which provide an effective strategy for the personalized and rapid development of walnut SSR primers. The developed 28 pairs of monomorphic primers can provide scientific basis for "Illegitimacy Testing" of hybrid offspring in molecular marker assisted breeding.

**Keywords:** *Juglans regia* L.; microsatellite; distribution characteristic; molecular marker

(责任编辑: 张 研)