

## 橄榄果实转录组 SSR 和 SNP/InDel 位点特征

赖瑞联, 沈朝贵, 冯 新, 陈义挺, 韦晓霞, 吴如健\*

福建省农业科学院果树研究所, 福建福州 350013

**摘要:**简单重复序列标记(simple sequence repeats, SSR)和单核苷酸多态性标记(single nucleotide polymorphism, SNP)具有高灵敏度和高特异性特征,挖掘不同类型橄榄果实性状相关分子标记可为其分子辅助育种提供参考。本研究以充分成熟的长营和惠圆橄榄果实为材料,经总 RNA 提取和 cDNA 文库构建后采用 Illumina Novaseq 平台进行转录组测序,并采用 MISA 1.0 和 GATK3 软件分析转录组的 SSR、SNP 和插入缺失标记(InDel)位点特征。结果在橄榄果实转录组的 10 124 条 unigenes 上鉴定到 13 935 个 SSR 位点,发生频率为 22.98%,平均每 1 kb 序列出现 0.25 个 SSR 位点,平均长度为 14.34 bp;其中,单碱基重复基元类型的 SSR 位点数量最多(占比 66.80%),长度为 10~64 bp,平均长度为 12.85 bp,重复基元的重复次数集中在 9~12 次,频率最高的基元为 A/T(占比 66.67%);六碱基重复基元类型的 SSR 位点数量最少(占比 0.47%),长度为 30~54 bp,平均长度 31.76 bp,基元重复次数集中在 5~8 次,频率最高的基元为 AGATGG/ATCTCC(占 0.04%)。在橄榄果实转录组中共检测到 284 992 个 SNP 位点,平均每 1 kb 序列含有 5.21 个;其中,转换类型的 SNP 位点数量为 166 162,包括 C/T 和 A/G 两种;颠换类型的 SNP 位点数量为 118 830,包括 A/T、A/C、T/G 和 C/G 四种。此外,橄榄果实转录组中共挖掘到 18 548 个 InDel 位点,平均每 1 kb 序列存在 2.95 个,其中含有 1 个 InDel 位点的 unigenes 数量最多(7853 条);通过比对预测发现,含有最多 InDel 位点的 unigene 可能是胍肌质合成酶基因。研究表明,通过转录组测序可有效开发橄榄 SSR 和 SNP/InDel 标记,不同性状橄榄果实中存在丰富多样的 SSR 位点且广泛分布 SNP/InDel 位点。该研究结果为橄榄果实性状鉴定标记的开发提供数据基础。

**关键词:**橄榄;转录组;简单重复序列;单核苷酸多态性;插入缺失标记

中图分类号: S667.5 文献标识码: A

## SSR and SNP/InDel Characteristics of Fruit Transcriptomic Data of *Canarium album*

LAI Ruilian, SHEN Chaogui, FENG Xin, CHEN Yiting, WEI Xiaoxia, WU Rujian\*

Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian 350013, China

**Abstract:** Simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) markers have been confirmed to be high sensitivity and specificity. Development of molecular markers related to different types of fruit quality traits of *Canarium album* (Lour.) Raeusch. can provide reference for its molecular assisted breeding to a considerable extent. The fully mature fruits of *C. album* cv. Changying and Huiyuan were collected to use as materials. After total RNA extraction and cDNA library construction, the transcriptome was sequenced on the Illumina Novaseq platform, and the SSR, SNP and InDel loci characteristics of the transcriptome were analyzed by MISA 1.0 and GATK3 software. Results showed that a total of 13 935 SSR loci were identified from 10 124 unigenes of *C. album* fruit transcriptome, the average 1 kb sequence appeared 0.25 SSR loci, the frequency and average length was 22.98% and 14.34 bp, respectively. Among them, the single base repeat type had the largest number of SSR loci (accounting for 66.80%), with a length of 10–64 bp, and an average length of 12.85 bp, the repeat times of repeat motifs were concentrated in 9–12, and the motif

收稿日期 2022-06-31; 修回日期 2022-07-18

基金项目 福建省属公益类科研院所基本科研专项(No. 2021R1028003, No. 2022R1028003); 农业农村部物种品种资源保护(热带作物)项目(No. 18220025)。

作者简介 赖瑞联(1990—),男,博士,助理研究员,研究方向:果树生物技术与遗传资源。\*通信作者(Corresponding author): 吴如健(WU Rujian), E-mail: wurujian@126.com。

with the highest frequency was A/T (accounting for 66.67%). The number of SSR loci of six base repeat motif type was the least (0.47%), the length was 30–54 bp, which average length was 31.76 bp, the number of motif repeats was concentrated in 5–8 times, and the motif with the highest frequency was AGATGG/ATCTCC (0.04%). A total of 284 992 SNP loci were detected in the transcriptome of *C. album* fruit, the average 1 kb sequence contained 5.21 SNP loci; Among them, the number of SNP loci of transformation type was 166 162, including C/T and A/G. The number of SNP loci of transversion type was 118 830, including A/T, A/C, T/G and C/G. In addition, 18 548 InDel loci were found in the transcriptome of *C. album* fruit, the average 1 kb sequence existed 2.95 InDel loci. The number of unigenes containing one InDel locus was the largest. It was predicted that the unigene containing the most InDel loci might be the callose synthase gene. These results showed that SSR and SNP/InDel markers could be effectively developed through RNA-seq. The SSR loci and SNP/InDel loci were widely distributed in *C. album* fruits with different quality traits. The results would provide a data basis for the development of identification markers of *C. album* fruit traits.

**Keywords:** *Canarium album*; transcriptome; simple sequence repeats; single nucleotide polymorphism; InDel

**DOI:** 10.3969/j.issn.1000-2561.2023.04.003

橄榄 (*Canarium album*) 是我国热带亚热带地区特色果树, 其果实富含多种营养和药用成分, 在我国福建、广东、四川、广西等地区广泛栽培。福州市是我国橄榄最主要的产地之一, “福州橄榄” 品牌于 2011 年获得农业部地理标志保护品牌, 2017 年品牌价值评估达 20.80 亿元人民币, 入选全国农产品百强<sup>[1]</sup>。福州市传统栽培的橄榄品种主要包括长营、惠圆、檀香、自来圆等。其中, 长营和惠圆的果实性状存在明显差异, 长营可食率约为 78.40%, 可供鲜食或加工, 在长营基础上选育出了一系列鲜食品种, 例如福榄 1 号(光甜)<sup>[2]</sup>、清榄 1 号<sup>[3]</sup>、福榄 2 号<sup>[4]</sup>等; 惠圆可食率约为 85.20%, 因鲜食品质较差, 主要以加工为主, 通过惠圆选育出了惠圆 1 号<sup>[5]</sup>、惠圆 3 号<sup>[6]</sup>等橄榄加工品种。作为我国特色果树产业, 品质性状提升是橄榄科研创新的重要命题。目前, 在橄榄品种选育<sup>[2, 7]</sup>、性状评价<sup>[8-9]</sup>、品质模型<sup>[10]</sup>、遗传背景<sup>[11]</sup>等方面都开展了大量研究。值得注意的是, 橄榄童期较长, 开发果实性状相关分子标记用于育种材料初步筛选可有效提高种质创新和育种效率。而现阶段, 橄榄果实性状形成分子机制研究及相关标记的开发仍处于空白, 限制了橄榄分子辅助育种进程。

DNA 分子标记鉴定是植物分子辅助育种重要技术手段, 尤其是具有高灵敏度、高特异性的简单重复序列标记 (simple sequence repeats, SSR) 和单核苷酸多态性标记 (single nucleotide polymorphism, SNP) 在许多植物中得到了有效应用。ŞAHİN 等<sup>[12]</sup>筛选了抗霜霉病相关的向日葵 (*Helianthus annuus*) SSR 标记并成功应用于向日葵辅助育种; GHARSALLAH 等<sup>[13]</sup>结合表型和

SSR 标记挖掘了番茄 (*Lycopersicon esculentum*) 品种耐盐特性相关位点, 有效提高了番茄耐盐胁迫育种效率; TERAKAMI 等<sup>[14]</sup>采用 SSR 标记对中国梨 (*Pyrus ussuriensis*) 黑斑病易感基因进行了定位, 为梨抗黑斑病辅助育种提供了重要途径。在 SNP 开发相关研究中, TAN 等<sup>[15]</sup>开发了小麦 (*Triticum aestivum*) 抗黑森瘿蚊基因的 SNP 标记, 可用于进一步指导小麦育种; KIM 等<sup>[16]</sup>通过转录组测序挖掘了萝卜 (*Raphanus sativus*) SNP 标记用于萝卜种子的鉴定、纯度检测和亲本组合的调整; FROUIN 等<sup>[17]</sup>鉴定了水稻 (*Oryza sativa*) 预防砷吸收和积累相关的 SNP 位点, 为水稻相关育种提供了参考依据。目前, 橄榄转录组测序相关研究也取得了一定进展, 为分子标记开发奠定了良好的基础<sup>[18]</sup>。在前期研究基础上, 本研究基于转录组数据对具有不同性状的橄榄品种长营和惠圆的果实进行 SSR 和 SNP/InDel 特征分析, 以期对橄榄果实性状相关分子标记开发提供数据基础。

## 1 材料与方法

### 1.1 材料

供试材料为长营和惠圆橄榄品种的果实, 取自福建省农业科学院果树研究所橄榄种质资源圃 (26° 07'40"N, 119°20'16"E)。以孔祥佳等<sup>[19]</sup>建立橄榄果实成熟度评价体系为参考标准, 采集不同品种成熟、健康、无机械病虫损伤的果实, 充分洗净后用液氮速冻, 并置于超低温冰箱中保存备用。

### 1.2 方法

1.2.1 RNA 提取与检测 采用植物多糖多酚试剂盒 E.Z.N.A.TM Plant RNA Kit 提取果实总 RNA, 采用 1.0% 琼脂糖凝胶电泳和 Agilent 2100 bio-

analyzer (Agilent Technologies, USA) 检测总 RNA 的完整性, 随后利用 NanoPhotometer spectrophotometer (IMPLEN, USA) 检测总 RNA 的纯度。

1.2.2 文库构建与质检 取 1  $\mu\text{g}$  橄榄果实总 RNA, 采用 NEBNext<sup>®</sup> Ultra<sup>™</sup> RNA Library Kit 试剂盒 (Illumina, USA) 进行建库。首先使用 Oligo (dT) 磁珠富集带有 polyA 尾的 mRNA, 使用二价阳离子在高温下在 NEBNext 第一链合成反应缓冲液 (5 $\times$ ) 中进行裂解, 使用随机寡核苷酸引物和 M-MuLV 逆转录酶体系合成 cDNA 第一链。随后使用 RNaseH 酶降解 RNA 链, 再通过 DNA 聚合酶 I 合成 cDNA 第二链。纯化后的 cDNA 进行末端修复、加 A 尾、连接测序接头, 随后使用 AMPure XP beads 筛选长度为 250~300 bp 的 cDNA 片段, 经 PCR 扩增后进行 PCR 产物纯化, 最终获得文库。采用 Qubit 2.0 Fluorometer (Invitrogen, USA) 和 Agilent 2100 bioanalyzer 进行文库质量检测, 随后采用 qRT-PCR 对文库进行定量, 筛选有效浓度高于 2.0 nmol/L 的文库后用于后续测序。

1.2.3 RNA 测序与数据组装 采用 TruSeq PE Cluster Kit v3 cBot HS (Illumina, USA) 对样本进行聚类, 随后在 Illumina Novaseq 平台上对制备的文库进行测序, 并产生 150 bp 配对末端读数。测序片段的图像数据经 CASAVA 转化为 fastq 格式的序列数据文件, 去除带测序接头、无法确定碱基信息以及低质量的 reads, 最终获得高质量的 clean reads。随后采用 TRINITY 软件<sup>[20]</sup>将 clean reads 拼接成 transcripts, 再采用 CORSET 软件<sup>[21]</sup>进行层次聚类后得到最长 cluster 序列作为 unigene 用于后续分析。每个样本进行 3 次生物学重复。

1.2.4 SSR 位点鉴定 采用 MISA 1.0 软件<sup>[22]</sup>筛选和识别 unigene, 并对转录组 SSR 位点进行分

析, 鉴定出单碱基重复、双碱基重复、三碱基重复、四碱基重复、五碱基重复和六碱基重复等类型的橄榄果实 SSR。采用 Microsoft Excel 2013 软件进行数据分析和图表制作。

1.2.5 SNP/InDel 分型与特征分析 采用 Picard-tools v1.41 和 Samtools v0.1.18 软件进行排序并删除重复数据, 同时合并每个样本的校准结果, 并采用 GATK3 软件 v3.4 版本<sup>[23]</sup>默认参数进行 SNP 位点检测和数据统计。

## 2 结果与分析

### 2.1 橄榄果实的 SSR 位点检测

2.1.1 SSR 位点数量 通过长营和惠圆橄榄品种果实转录组测序, 6 个测序样本平均获得 22 813 726 bp 的 raw reads, 过滤后的 clean reads 为 22 019 057 bp, 原始数据过滤后的平均碱基数为 6.61 Gb, 整体测序错误率约为 0.02%, Q20 和 Q30 分别为 98.11% 和 94.19%, 说明测序结果较好, 可用于后续分析。拼接后, 共获得 125 021 条转录本和 44 062 条 unigenes, 这些 unigenes 包含 54 735 280 bp 核苷酸。采用 MISA1.0 对获得的 unigenes 进行 SSR 位点鉴定, 总共在 10 124 条 unigenes 上鉴定到 13 935 个 SSR 位点, 发生频率为 22.98%, 平均每 1 kb 序列出现 0.25 个 SSR 位点。这些 SSR 分布的 unigenes 中, 2657 条序列包含 1 个以上的 SSR 位点, 1072 条序列包含复合型 SSR 位点。可见, 橄榄果实转录组中广泛分布 SSR 位点。

2.1.2 SSR 重复基元类型 橄榄果实 SSR 位点的重复基元类型如表 1 所示。SSR 重复基元类型较为丰富, 除了含有单碱基重复到六碱基重复 6 种类型, 还包含 977 种复杂重复类型。对单碱基重

表 1 SSR 重复基元类型  
Tab. 1 Types of SSR repeat motifs

重复基元类型 Repeat motif of type	SSR 位点数量 SSR number	比例 Proportion/%	频率 Frequency/%	平均分布距离 Mean distance/kb
单碱基	9308	66.80	21.12	5.88
双碱基	2480	17.80	5.63	22.07
三碱基	1830	13.13	4.15	29.91
四碱基	176	1.26	0.40	311.00
五碱基	76	0.55	0.17	720.20
六碱基	65	0.47	0.15	842.08
总计	13 935	100.00	31.63	3.93

复到六碱基重复的 SSR 类型进一步分析发现,单碱基重复的 SSR 类型最多,达到 9308 个,占全部 SSR 位点的 66.80%,随着重复基元碱基数的增加,类型数量逐渐减少,六碱基重复的 SSR 类型仅有 65 个,占全部 SSR 位点的 0.47%。从 SSR 位点的出现频率和分布距离来看,单碱基重复的 SSR 位点类型分别为 21.12%和 5.88 kb,而六碱基重复的 SSR 位点类型分别为 0.15%和 842.08 kb。结果可见,不同重复基元类型的 SSR 位点数目和分布存在较大差异。

2.1.3 SSR 序列长度分布 不同重复基元的橄榄果实 SSR 类型的序列长度为 10~320 bp,平均长度为 36.19 bp。对其中单碱基重复基元到六碱基重复基元的 SSR 序列进一步分析发现(表 2),整体上序列长度为 10~72 bp,平均长度为 14.34 bp。其中,长度差异最大的是单碱基重复的 SSR 类型,介于 10~64 bp 之间,平均长度为 12.85 bp;差异最小的为五碱基重复的 SSR 类型,介于 25~30 bp 之间,平均长度为 25.68 bp。

2.1.4 SSR 优势重复基元分布特征 橄榄果实 SSR 位点基元数量统计分析结果如图 1 所示。分布频率最高的基元分别是 A/T (9290 个,占比

**表 2 SSR 长度统计**  
Tab. 2 Length statistics of SSR

重复基元类型 Repeat motif of type	最小值 Minimum	最大值 Maximum	均值 Average	合计 Total
单碱基	10	64	12.85	105 020
双碱基	12	62	16.53	31 906
三碱基	15	72	17.82	27 105
四碱基	20	36	21.92	3156
五碱基	25	30	25.68	1695
六碱基	30	54	31.76	1620
合计	10	72	14.34	170 502

66.67%)、AG/CT (1304 个,占比 9.36)、AT/AT (857 个,占比 6.15%)、AAT/ATT (421 个,占比 3.02%)、AAG/CTT (392 个,占比 2.81%)。对不同 SSR 重复基元类型的出现频数进行分析发现,所有 SSR 位点共出现 105 种重复基元,单碱基到六碱基重复的基元类型分别为 2、4、10、21、29、39 种。单碱基重复基元到六碱基重复基元中的优势重复基元分别为 A/T、AG/CT、AAT/ATT、AAAT/ATTT (65 个,占比 0.47%)、AAGAG/CTCTT (10 个,占比 0.07%) 和 AGATGG/ATCTCC (5 个,占比 0.04%)。

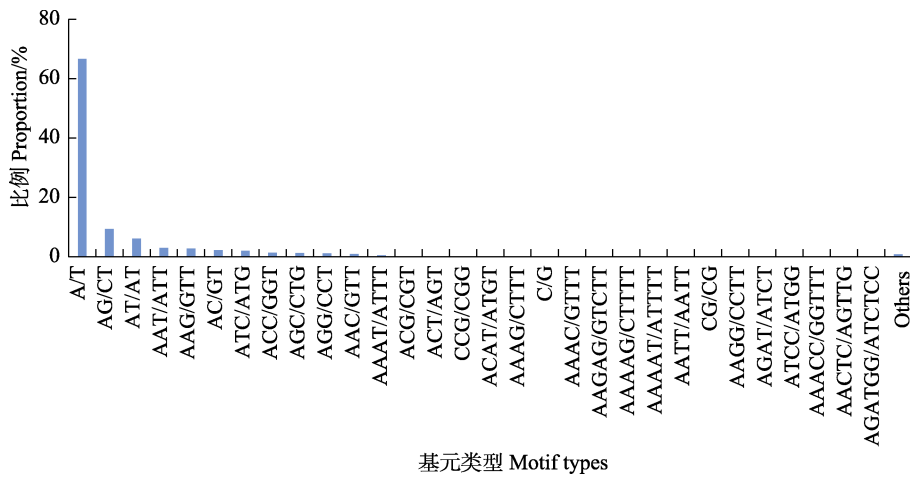


图 1 不同重复基元类型的 SSR 统计  
Fig. 1 Statistics of different SSR motifs

2.1.5 SSR 各基元类型重复次数 橄榄果实 SSR 位点不同基元类型的重复次数统计结果如图 2 所示。其中,单碱基重复基元的重复次数主要分布在 9~12 和 13~16 之间,分别占 SSR 位点总数的 41.69%和 17.22%,占单碱基重复基元总数的 62.42%和 25.77%;双碱基到六碱基重复基元的重复次数均主要集中在 5~8,分别占对应重复基元

总数的 68.67%、93.55%、98.30%、100%和 98.46%。不同重复基元类型的 SSR 位点数量随着重复次数的增多呈递减趋势。

2.2 橄榄果实的 SNP/InDel 的特征分析

基于转录组数据进行不同性状橄榄品种果实 SNP/InDel 挖掘。共获得 284 992 个 SNP 位点,平均每 1 kb 序列含有 5.21 个。其中,转换类型的

SNP 位点共 166 162 个，C/T 和 A/G 发生频率相近，每 1 kb 序列分别含有 1.51 个和 1.52 个。颠换类型的 SNP 位点共 118 830 个，每 1 kb 序列中 A/T、A/C、T/G 和 C/G 的平均个数分别为 0.63、0.53、0.54 和 0.47。转换类型的 SNP 位点（3.03 个）的发生频率明显高于颠换类型（2.18 个）。其中，第 1、2 和 3 位上发生 SNP 突变的密码子数量分别为 51 513、25 378 和 50 306。在这些 SNP 位点中，4479 条 unigenes 包含 1 个 SNP 位点，3096 条 unigenes 包含 2 个 SNP 位点，2419 条 unigenes 包含 3 个 SNP 位点，随着 SNP 位点数增多，unigenes 的数量逐渐减少（图 3A）。值得注意的是，其中的 14 条 unigenes 包含有 100 个以上的 SNP 位点（表 3），其中的 9 条 unigenes

得到了功能注释，仍有 5 条 unigenes 在所有数据库中未得到有效比对，这些存在高频 SNP 变异的 unigenes 可能在果实性状差异形成过程中发挥了重要功能。

基于转录组测序数据，共获得 18 548 个 InDel 位点，平均每 1 kb 序列含有 2.95 个，其分布变化趋势与 SNP 相似（图 3B），unigenes 中含有 1 个 InDel 位点的数量最多，达到 7853 条。InDel 位点最多的是 Cluster-4594.16496，达到 24 个 InDel 位点，通过比对预测，其可能是胍胍质合成酶。

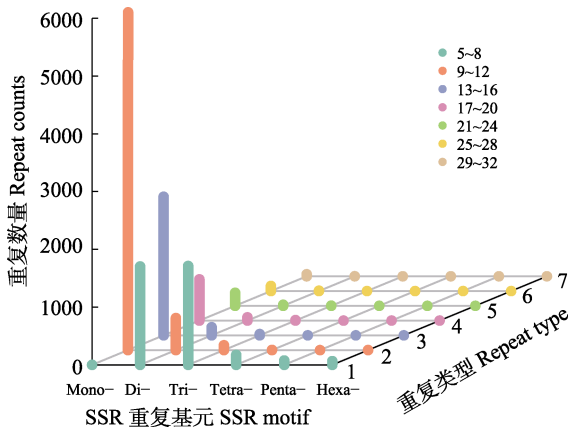


图 2 SSR 长度分布

Fig. 2 Length distribution of SSR

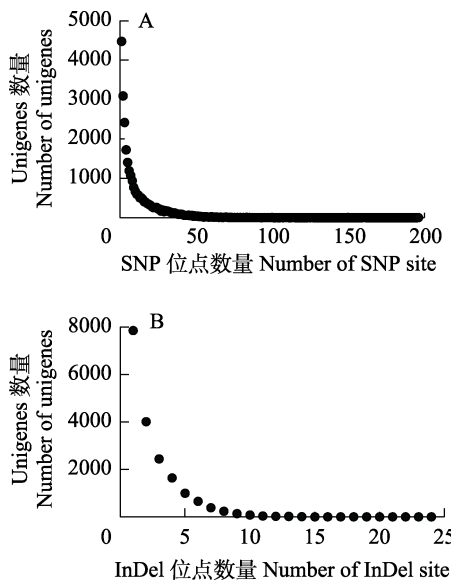


图 3 SNP 和 InDel 位点分布

Fig. 3 Distribution of SNP and InDel site

表 3 存在高频率 SNP 位点的基因序列信息

Tab. 3 Unigenes with high-frequency of SNP loci

基因 ID Gene ID	SNP 位点数量 Number of SNP site	基因功能预测 Gene functional prediction
Cluster-4594.14962	100	未知
Cluster-4594.16468	103	染色质重塑
Cluster-4594.22679	106	抗病蛋白
Cluster-4594.19172	106	未知
Cluster-4594.13384	108	未知
Cluster-4594.20246	109	DnaJ 同源亚家族
Cluster-4594.11465	110	ATP 依赖解旋酶
Cluster-4594.15912	113	Pre-mRNA 剪接因子
Cluster-4594.24976	114	抗病蛋白
Cluster-4594.18548	124	未知
Cluster-4594.15297	126	非光周期依赖早花相关基因
Cluster-4594.22117	160	未知
Cluster-4594.16377	161	生长素转运蛋白
Cluster-4594.15956	196	转化/转录结构域相关蛋白

### 3 讨论

#### 3.1 转录组测序可有效开发橄榄 SSR 和 SNP/InDel 标记

转录组测序是植物 SSR 和 SNP/InDel 标记开发有效的技术手段。VIDYA 等<sup>[24]</sup>从生姜 (*Zingiber officinale*) 转录组中挖掘了 16 790 个 EST-SSR 位点，其中 4597 个 SSR 位点分布在已预测的编码基因中；ZHOU 等<sup>[25]</sup>从楠木 (*Phoebe bournei*) 转录组中鉴定了 40 853 个 SSR 位点，并对其中 23 个多态性 EST-SSR 标记进行应用；TULSANI 等<sup>[26]</sup>从芫荽 (*Coriandrum sativum*) 转录组中鉴定到了 9746 个 SSR 位点，为其转录组草图和基因标记提供了重要信息；XU 等<sup>[27]</sup>基于转录组测序从草珊瑚

瑚 (*Sarcandra glabra*) 中挖掘了 726 476 个 SNP 位点和 42 939 个 InDel 位点, 为其资源开发奠定了基础。本研究采用 Illumina Novaseq 测序平台, 从不同性状的橄榄果实中分别挖掘了 13 935 个 SSR 位点、284 992 个 SNP 位点和 18 548 个 InDel 位点, 为后续橄榄分子标记开发和应用奠定了数据基础。

### 3.2 橄榄果实中存在丰富多样的 SSR 位点

SSR 标记具有多态性高、分辨率高、通用性好、稳定性强等优点, 是物种种质资源鉴定、遗传连锁图谱构建、基因定位、数量性状基因位点分析、分子标记辅助育种等研究的重要技术手段<sup>[28]</sup>。本研究通过转录组高通量测序, 在 10 124 条 unigenes 上鉴定到 13 935 个 SSR 位点, 发生频率为 22.98%。其中, 单碱基重复基元类型的 SSR 位点最多并且单碱基重复是橄榄果实优势重复基元, 随着重复碱基数量的增加, SSR 位点出现频率下降, 与油梨 (*Persea americana*)<sup>[29]</sup>、甜柿 (*Diospyros kaki*)<sup>[30]</sup>、香椿 (*Toona sinensis*)<sup>[31]</sup> 等物种的规律一致。橄榄果实中的优势重复基元主要包括 A/T、AG/CT、AT/AT、AAT/ATT、AAG/CTT 等, 尤其 A/T 的比例达到 66.67%, 与甜柿<sup>[30]</sup>、龙眼 (*Dimocarpus longan*)<sup>[32]</sup>、油梨<sup>[29]</sup> 等其他物种存在一定的相似性, 但仍有差异, 可见 SSR 位点分布存在物种特异性。此外, 橄榄果实 SSR 位点中也发现了 6 个 CG/CG 和 29 个 CCG/CGG 重复基元, 其在双子叶植物中较为少见<sup>[30]</sup>, 可能在橄榄进化过程中发挥特殊的生物学功能。从多态性上看, 通常认为长度大于 20 bp 的 SSR 位点具有较高的多态性<sup>[33]</sup>, 橄榄果实中四碱基、五碱基和六碱基重复的 SSR 位点长度均在 20 bp 以上, 具有更高的多态性, 可能具有较高的应用价值。

### 3.3 橄榄果实中广泛分布 SNP/InDel 位点

SNP/InDel 具有高效性、准确性, 并且与作物性状密切相关。本研究基于转录组测序, 共检测到橄榄果实 284 992 个 SNP 位点, 平均每 1 kb 序列含有 5.21 个 SNP 位点。其中, 转换和颠换类型的比例为 1.398, 与理论值 0.500 存在较大偏差, 说明橄榄进化过程中对不同类型的选择并非随机而可能存在偏向性<sup>[34]</sup>。此外, 本研究共获得橄榄果实 18 548 个 InDel 位点, 每 1 kb 序列平均包含 2.95 个 InDel 位点。后续还应进一步针对 SNP/InDel

位点进行深入研究, 开发橄榄果实性状相关的特异性分子标记, 为橄榄种质资源鉴定和分子辅助育种提供技术支撑。

### 参考文献

- [1] 许长同. 福州发展橄榄百亿产业的调查报告[J]. 东南园艺, 2019, 7(2): 44-48.  
XU C T. Investigation report on the development of ten billions of olive industry in Fuzhou[J]. Southeast Horticulture, 2019, 7(2): 44-48. (in Chinese)
- [2] 吴如健, 万继锋, 韦晓霞, 陈瑾, 潘少霖, 胡菡青, 张泽焯. 优质鲜食橄榄新品系‘光甜’的选育[J]. 热带作物学报, 2011, 32(8): 1445-1448.  
WU R J, WAN J F, WEI X X, CHEN J, PAN S L, HU H Q, ZHANG Z H. Selection and breeding of a new high-quality and fresh Chinese olive strain ‘Guangtian’[J]. Chinese Journal of Tropical Crops, 2011, 32(8): 1445-1448. (in Chinese)
- [3] 许长同, 肖振林, 朱宗良, 杨文荣. 橄榄鲜食新品种‘清榄 1 号’选育初报[J]. 中国南方果树, 2013, 42(3): 70-72.  
XU C T, XIAO Z L, ZHU Z L, YANG W R. Preliminary report on breeding of a new fresh Chinese olive variety ‘Qinglan 1’[J]. South China Fruit, 2013, 42(3): 70-72. (in Chinese)
- [4] 万继锋, 熊双伟, 吴如健, 韦晓霞, 陈瑾, 胡菡青, 林岳晶. 橄榄鲜食新品系‘福榄 2 号’的主要性状及栽培技术要点[J]. 中国南方果树, 2014, 43(4): 129-130, 134.  
WAN J F, XIONG S W, WU R J, WEI X X, CHEN J, HU H Q, LIN Y J. Main characters and cultivation techniques of a new Chinese olive fresh edible strain ‘Fulan 2’[J]. South China Fruit, 2014, 43(4): 129-130, 134. (in Chinese)
- [5] 陈思聪. ‘惠圆 1 号’橄榄品种选育报告[J]. 中国园艺文摘, 2013, 29(7): 1-2, 13.  
CHEN S C. Breeding report of Chinese olive variety ‘Huiyuan 1’[J]. Chinese Horticultural Abstracts, 2013, 29(7): 1-2, 13. (in Chinese)
- [6] 许长同, 余述. ‘惠圆 3 号’橄榄品种选育初报[J]. 中国南方果树, 2013, 42(6): 107-108.  
XU C T, YU S. Preliminary report on breeding of Chinese olive variety ‘Huiyuan 3’ olive variety[J]. South China Fruit, 2013, 42(6): 107-108. (in Chinese)
- [7] 肖维强, 赖多, 梁银浩, 匡石滋, 陈盖洵, 邵雪花, 苏燕钿, 林立红, 孙倍煜, 林洁瑶, 张力, 刘传和, 贺涵, 黄锐浩, 刘传滨. 鲜食橄榄新品种‘饶脆橄榄’[J]. 园艺学报, 2021, 48(S2): 2821-2822.  
XIAO W Q, LAI D, LIANG Y H, KUANG S Z, CHEN G X, SHAO X H, SU Y X, LIN L H, SUN B Y, LIN J Y, ZHANG L, LIU C H, HE H, HUANG R H, LIU C B. A New *Canar-*

- ium album* cultivar 'Raocui Ganlan'[J]. *Acta Horticulturae Sinica*, 2021, 48(S2): 2821-2822. (in Chinese)
- [8] 赖瑞联, 冯新, 陈瑾, 陈义挺, 吴如健. 福州 3 个鲜食橄榄品种(系)的果实品质特征与香气组分分析[J]. *食品科学*, 2020, 41(6): 231-238.
- LAI R L, FENG X, CHEN J, CHEN Y T, WU R J. Quality characteristics and aroma components of three fresh edible cultivars of *Canarium album* grown in Fuzhou[J]. *Food Science*, 2020, 41(6): 231-238. (in Chinese)
- [9] 赖瑞联, 陈瑾, 冯新, 池丽丽, 陈义挺, 吴如健. 福州 3 个鲜食橄榄品种(系)的氨基酸组成与品质评价[J]. *中国南方果树*, 2021, 50(5): 71-75.
- LAI R L, CHEN J, FENG X, CHI L L, CHEN Y T, WU R J. Amino acid composition and quality evaluation of three edible cultivars of *Canarium album* in Fuzhou[J]. *South China Fruit*, 2021, 50(5): 71-75. (in Chinese)
- [10] 池毓斌, 朱丽娟, 黄敏杰, 彭真汾, 叶清华, 张静芳, 陈清西, 许长同. 鲜食橄榄品质综合评价模型的建立与验证[J]. *果树学报*, 2017, 34(8): 1051-1060.
- CHI Y B, ZHU L J, HUANG M J, PENG Z F, YE Q H, ZHANG J F, CHEN Q X, XU C T. Establishment and verification of a comprehensive evaluation model for quality of fresh Chinese olive[J]. *Fruit Science*, 2017, 34(8): 1051-1060. (in Chinese)
- [11] 赖瑞联, 陈瑾, 冯新, 韦晓霞, 陈义挺, 沈朝贵, 吴如健. 橄榄 ISSR 和 RAPD 遗传多样性分析和核心种质构建[J]. *热带亚热带植物学报*, 2022, 30(1): 41-53.
- LAI R L, CHEN J, FENG X, WEI X X, CHEN Y T, SHEN C G, WU R J. ISSR and RAPD genetic diversity analysis and core germplasm construction of *Canarium album*[J]. *Journal of Tropical and Subtropical Botany*, 2022, 30(1): 41-53. (in Chinese)
- [12] ŞAHİN E Ç, KALENDEROĞLU A, AYDIN Y, EVCI G, UNCUOĞLU A A. SSR markers suitable for marker assisted selection in sunflower for downy mildew resistance[J]. *Open Life Science*, 2018, 13(1): 319-326.
- [13] GHARSALLAH C, BEN ABDELKRIM A, FAKHFAKH H, SALHI-HANNACHI A, GORSANE F. SSR marker-assisted screening of commercial tomato genotypes under salt stress[J]. *Breed Science*, 2016, 66(5): 823-830.
- [14] TERAKAMI S, ADACHI Y, TAKEUCHI Y, TAKADA N, NISHIO S, SAITO T, YAMAMOTO T. Development of an SSR marker set for efficient selection for resistance to black spot disease in pear breeding[J]. *Breed Science*, 2021, 71(2): 240-252.
- [15] TAN M K, EL-BOUHSSINI M, WILDMAN O, TADESSE W, CHAMBERS G, LUO S, EMEBIRI L. Development of SNP assays for hessian fly response genes, *Hfr-1* and *Hfr-2*, for marker-assisted selection in wheat breeding[J]. *BMC Genetics*, 2018, 19(1): 50.
- [16] KIM J, MANIVANNAN A, KIM D S, LEE E S, LEE H E. Transcriptome sequencing assisted discovery and computational analysis of novel SNPs associated with flowering in *Raphanus sativus* in-bred lines for marker-assisted backcross breeding[J]. *Horticulture Research*, 2019, 6(1): 120.
- [17] FROUIN J, LABEYRIE A, BOISNARD A, SACCHI G A, AHMADI N. Genomic prediction offers the most effective marker assisted breeding approach for ability to prevent arsenic accumulation in rice grains[J]. *PLoS One*, 2019, 14(6): e0217516.
- [18] LAI R, FENG X, CHEN J, ZHANG Y, WEI X, CHEN Y, CHENG C, WU R. *De novo* transcriptome assembly and comparative transcriptomic analysis provide molecular insights into low temperature stress response of *Canarium album*[J]. *Scientific Reports*, 2021, 11(1): 1-17.
- [19] 孔祥佳, 林河通, 周鹤, 林艺芬, 陈艺晖, 王慧. 鲜食橄榄果实的适宜采收期及其品质评价参数的研究[J]. *保鲜与加工*, 2016, 16(2): 6-14.
- KONG X J, LIN H T, ZHOU H, LIN Y F, CHEN Y H, WANG H. Study on the optimum harvesting date of fresh-eating Chinese olive fruit and its quality assessment parameters[J]. *Storage and Process*, 2016, 16(2): 6-14. (in Chinese)
- [20] GRABHERR M G, HAAS B J, YASSOUR M, LEVIN J Z, THOMPSON D A, AMIT I, ADICONIS X, FAN L, RAYCHOWDHURY R, ZENG Q, CHEN Z, MAUCELI E, HACOEN N, GNIRKE A, RHIND N, DI PALMA F, BIRREN B W, NUSBAUM C, LINDBLAD-TOH K, FRIEDMAN N, REGEV A. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. *Nature Biotechnology*, 2011, 29(7): 644-652.
- [21] DAVIDSON NM, OSHLACK A. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes [J]. *Genome Biology*, 2014, 15(7): 410.
- [22] BEIER S, THIEL T, MÜNCH T, SCHOLZ U, MASCHER M. MISA-web: a web server for microsatellite prediction[J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [23] VAN DER AUWERA G A, CARNEIRO M O, HARTL C, POPLIN R, DEL ANGEL G, LEVY-MOONSHINE A, JORDAN T, SHAKIR K, ROAZEN D, THIBAUT J, BANKS E, GARIMELLA KV, ALTSHULER D, GABRIEL S, DEPRISTO M A. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline[J]. *Current Protocols in Bioinformatics*, 2013, 43(1110): 11.
- [24] VIDYA V, PRASATH D, SNIGDHA M, GOBU R, SONA

- C, MAITI C S. Development of EST-SSR markers based on transcriptome and its validation in ginger (*Zingiber officinale* Rosc.)[J]. PLoS One, 2021, 16(10): e0259146.
- [25] ZHOU Q, ZHOU P Y, ZOU W T, LI Y G. EST-SSR marker development based on transcriptome sequencing and genetic analyses of *Phoebe bournei* (Lauraceae)[J]. Molecular Biology Reports, 2021, 48(3): 2201-2208.
- [26] TULSANI N J, HAMID R, JACOB F, UMRETIYA N G, NANDHA A K, TOMAR R S, GOLAKIYA B A. Transcriptome landscaping for gene mining and SSR marker development in Coriander (*Coriandrum sativum* L.)[J]. Genomics, 2019, 112(2): 1545-1553.
- [27] XU Y, TIAN S, LI R, HUANG X, LI F, GE F, HUANG W, ZHOU Y. Transcriptome characterization and identification of molecular markers (SNP, SSR, and Indels) in the medicinal plant *Sarcandra glabra* spp[J]. Biomed Research International, 2021, 2021: 9990910.
- [28] 杨梦婷, 黄洲, 干建平, 徐君驰, 庞基良. SSR 分子标记的研究进展[J]. 杭州师范大学学报: 自然科学版, 2019, 18(4): 429-436.
- YANG M T, HUANG Z, GAN J P, XU J C, PANG J L. Research progress of SSR molecular markers[J]. Journal of Hangzhou Normal University (Natural Sciences Edition), 2019, 18(4): 429-436. (in Chinese)
- [29] 郭俊, 朱婕, 谢尚潜, 张叶, 叶蓓蕾, 郑丽燕, 凌鹏. 油梨转录组 SSR 分子标记开发与种质资源亲缘关系分析[J]. 园艺学报, 2020, 47(8): 1552-1564.
- GUO J, ZHU J, XIE S Q, ZHANG Y, YE B L, ZHENG L Y, LING P. Development of SSR molecular markers based on transcriptome and analysis of genetic relationship of germplasm resources in avocado[J]. Acta Horticulturae Sinica, 2020, 47(8): 1552-1564. (in Chinese)
- [30] 王艺儒, 索玉静, 傅建敏. 小果甜柿果实转录组的 SSR、SNP 和 InDel 特征分析[J]. 西北农林科技大学学报(自然科学版), 2022, 50(7): 1-8.
- WANG Y R, SUO Y J, FU J M. SSR, SNP and InDel analysis based on transcriptome data of *Diospyros kaki* 'Xiaoguo-tianshi' fruit[J]. Journal of Northwest A&F University (Natural Science Edition), 2022, 50(7): 1-8. (in Chinese)
- [31] 于鹏飞, 孙晓健, 李晨晨, 张旭, 郑威, 刘常金. 香椿染色体核型分析及 SSR 分子标记开发[J]. 园艺学报, 2019, 46(6): 1172-1182.
- YU P F, SUN X J, LI C C, ZHANG X, ZHENG W, LIU C J. Chromosome karyotype analysis and development of SSR molecular markers in *Toona sinensis*[J]. Acta Horticulturae Sinica, 2019, 46(6): 1172-1182. (in Chinese)
- [32] 胡文舜, 陈秀萍, 郑少泉. 龙眼 EST-SSR 标记开发及无患子科 5 个属种质遗传多样性分析[J]. 园艺学报, 2019, 46(7): 1359-1372.
- HU W S, CHEN X P, ZHENG S Q. EST-SSR markers developed from *Dimocarpus longan* and their application in genetic diversity analysis of five genera of Sapindaceae[J]. Acta Horticulturae Sinica, 2019, 46(7): 1359-1372. (in Chinese)
- [33] 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.
- [34] WAKELEY J. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance[J]. Trends in Ecology & Evolution, 1996, 11(4): 158-162.