

基于柿雌雄花芽转录组测序的 SSR 和 SNP 多态性分析

杜改改^{1,2} 孙鹏^{1,2} 索玉静^{1,2} 韩卫娟^{1,2} 刁松锋^{1,2} 傅建敏^{1,2*} 李芳东^{1,2}

(1. 国家林业局 泡桐研究开发中心, 郑州 450003;

2. 中国林业科学研究院经济林研究开发中心, 郑州 450003)

摘要 为深入开发柿 SSR 和 SNP 分子标记,进而推动柿品种鉴定和遗传多样性等研究,以‘禅寺丸’柿的雌雄花芽转录组序列为基础,对 SSR 和 SNP 位点进行发掘。结果表明:转录组测序获得了 154 741 条 unigene,其中有 38.49% unigene 在数据库中得到注释。利用 MISA 软件进行 SSR 位点的搜索,共得到 44 304 个 SSR,包含 83 种重复单元,其中以 A/T 类型为主的单核苷酸重复所占的比例最高(20 006 个,占 47.63%),其次是以 AG/CT 类型为主的二核苷酸重复(16 055 个,占 38.23%),再次是以 AAG/CTT 类型为主的三核苷酸重复(5 402 个,占 12.87%)和以 AAAG/CTTT 类型为主的四核苷酸重复(500 个,占 1.19%)。在转录组得到的 unigene 中共发现 SNP 405 685 个,发生频率为 1/253 bp。6 种单核苷酸变异中,转换类型发生频率显著高于颠换类型,转换类型中的 C/T(31.51%)和 A/G(31.41%)发生频率最高,在颠换类型中 A/T 发生频率最高。柿 SSR 和 SNP 的位点是非常丰富的,可为柿遗传图谱构建、遗传多样性和亲缘关系研究提供丰富的基础数据信息。

关键词 柿;转录组测序;SSR 标记;重复单元;SNP

中图分类号 S665.2

文章编号 1007-4333(2017)10-0045-11

文献标志码 A

SSR and SNP polymorphism analysis based on persimmon (*Diospyros kaki*) transcriptome

DU Gaigai^{1,2}, SUN Peng^{1,2*}, SUO Yujing^{1,2}, HAN Weijuan^{1,2},
DIAO Songfeng^{1,2}, FU Jianmin^{1,2}, LI Fangdong^{1,2}

(1. China Paulownia Research Center, State Forestry Administration, Zhengzhou 450003, China;

2. Non-Timber Forest Research and Development Center, Chinese Academy of Forestry, Zhengzhou 450003, China)

Abstract The aim of this study was to develop SSR and SNP markers and promote the researches on varieties identification and genetic diversity in *Diospyros kaki*. The transcriptome of persimmon flowers were sequenced by RNA-Seq technology, and MISA and GATK3 software were used to explore the SSR and SNP loci. A total of 154 741 unigenes were detected, of which 38.49% were annotated according to databases. And 44 304 SSRs occurred in 59 570 unigenes were identified, including 83 SSR repeat motifs. Mononucleotide repeats(20 006, 47.63%) based on A/T accounted for great proportion in complete SSRs, dinucleotide repeats(16 055, 38.23%) mainly on AG/CT ranked the second; and trinucleotide(5 402, 12.87%) with the main type of AAG/CTT and tetranucleotide(500, 1.19%) with the main type of AAAG/CTTT, ranked third and forth separately. A total of 405 685 candidate SNP bit points were searched and the average appearance frequency was 1/253 bp, and the number of the conversion type was significantly higher than the transversion type. C/T and A/G belonged to conversion type occurred most frequently among 6 kinds of SNP types. Of all the transversion types, A/T occurred most frequently. SSR and SNP loci were abundant in persimmon, the development of molecular marker would provide the basic information for genetic map structure, analysis of genetic diversity and identification of relative relationships for *Diospyros kaki*.

Keywords *Diospyros kaki*; transcriptome sequencing; SSR marker; frequent motif; SNP

收稿日期: 2016-10-27

基金项目: 青年科学基金项目(31500559); 中国林科院青年项目配套类(CAFYBB2016QB017); 河南省基础与前沿技术基金项目(152300410135)

第一作者: 杜改改, 硕士研究生, E-mail: 772615428@qq.com

通讯作者: 傅建敏, 副研究员, 主要从事经济林栽培育种研究, E-mail: fim371@163.com

柿(*Diospyros kaki* Thunb.)为柿科(Ebenaceae)柿属(*Diospyros*)植物,果实风味香甜、营养丰富,具有一定的药用和保健价值,树姿优美,叶、果色泽艳丽,又具有良好的观赏价值^[1],是我国重要的经济栽培树种。柿主要分布于暖温带和亚热带地区,以中国、日本和韩国栽培最为广泛^[2]。我国也是世界上柿栽培历史最悠久、面积最广和产量最多的国家,现有柿种质1 058份,居世界首位^[3-5]。然而各品种间的亲缘关系不甚明了,同名异物和同物异名现象严重,品种的鉴定和分类工作尚未形成体系,严重阻碍了育种工作的进行^[6]。SSR和SNP分子标记是品种鉴定和亲缘关系研究的有力手段,SSR已被应用于柿种质的遗传多样性分析^[7-9]、品种鉴定^[10-11]、遗传图谱构建^[12]。

Soriano^[13]通过构建基因组文库的方法在柿中开发出了37个SSR标记位点,此方法需对每个克隆进行筛选鉴定,工作量大,花费大量的人力、物力和财力,而且效率较低,在植物中已报道的阳性克隆比率仅为2%~3%,成功获得引物的机率则更低^[14];郭大龙^[15]利用核酸数据库中序列信息筛选和设计了部分引物,建立起了柿属植物SSR技术体系,但数量有限,而且由于柿基因组信息非常有限,用于设计引物的多是与乙烯合成相关的序列信息,引物的多态性又低;Guo等^[16]利用Fiasco的方法对日本柿中微卫星进行了分离和鉴定,需要杂交富集微卫星序列,效率较低;郭大龙等^[17]介绍了一种利用ISSR开发SSR标记的方法,针对性强,简便易行,效率较高,但数量有限,仍未得到广泛应用。因此,传统的SSR分子标记开发方法还不完善,可用于品种鉴定的SSR位点偏少,还需开发更多的SSR位点以促进此类研究的深入。另外,SNP分子标记也被广泛应用于品种鉴定和亲缘关系的研究中^[18],其具有密度高、遗传稳定性强、易于实现自动化分析等^[19]优势,但该技术还未应用于柿属植物研究中,仍需进一步开发。通过转录组测序的方法能为鉴定柿多态性SSR和SNP分子标记提供重要数据源,而且可信度更高^[20]。日本学者Akagi^[21]对柿属植物君迁子(*Diospyros lotus* L.)进行了转录组测序获得了大量的序列信息,而关于柿转录组和基因组信息尚未有报道。

本研究利用Illumina的RNA-seq转录组测序技术对雌雄花混生的‘禅寺丸’进行转录组测序和数据组装,对基因进行功能注释和代谢通路分析,并充

分挖掘柿SSR和SNP的信息,通过SSR和SNP特征分析,以期对柿SSR和SNP标记的开发和利用提供生物信息学基础,为柿品种鉴定、遗传资源评价等研究提供基础的序列信息和高效的分子标记。

1 材料方法

1.1 研究材料

材料取自国家林业局泡桐研究开发中心实验基地(34°55′18″~34°56′27″N,113°46′14″~113°47′35″E),选取树龄15年,长势良好、无病虫害,能正常开花结实的雌雄同株的‘禅寺丸’柿为试验树。李加茹等^[22]研究发现,在4月17日雌花中雄蕊原基和雄花中雌蕊原基败育,单性花产生,因此确定此阶段为花芽分化的关键时期。此时期雌雄花芽可以从外部特征上辨别,而表型显现早于分子调控,因此选择在花芽分化关键时期(4月17日)之前的4月13日分别采集同一株‘禅寺丸’的雌、雄花芽各3份,液氮冷冻后在-80℃保存。

1.2 总RNA的提取

柿花芽总RNA的提取使用上海生工的EZ-10DNAaway RNA mini-prep Kit试剂盒,然后利用Agilent 2100Bioanalyzer检测提取的RNA是否达到RNA-Seq的试验标准,即总量 $\geq 3\mu\text{g}$ 、浓度 $\geq 50\text{ ng}/\mu\text{L}$ 、 $\text{OD}_{260/280} \geq 1.8$ 、RIN值 ≥ 6.5 、 $28\text{S}/18\text{S} \geq 1$ 。

1.3 转录组测序及数据组装和注释

样品检测合格后,用带有Oligo(dT)的磁珠富集真核生物mRNA,随后加入Fragmentation buffer将mRNA打断成短片段,以mRNA为模板,用六碱基随机引物(Random hexamers)合成一链cDNA,然后加入缓冲液、dNTPs、DNA polymerase I和RNase H合成二链cDNA,再用AMPure XP beads纯化双链cDNA。纯化的双链cDNA先进行末端修复、加A尾并连接测序接头,再用AMPure XP beads进行片段大小选择。最后进行PCR扩增,并用AMPure XP beads纯化PCR产物,得到最终的文库。文库构建完成后,先使用Qubit 2.0进行初步定量,稀释文库至 $1.5\text{ ng}/\mu\text{L}$,随后使用Agilent 2100对文库的insert size进行检测,Insert size符合预期后,使用Q-PCR方法对文库的有效浓度进行准确定量(文库有效浓度 $> 2\text{ nmol/L}$),以保证文库质量。库检合格后,把不同文库按照有效浓度及目标下机数据量的需求Pooling后进行Illumina HiSeq测序。

对 6 份试验样品(雌花芽、雄花芽各 3 份)测序得到的原始数据过滤掉里面含有带接头的、低质量的 Read 得到 Clean reads。利用 Trinity 软件对所有样品的 Clean reads 进行混合拼接成转录本序列,取每条基因中最长的转录本为 Unigene 作为后续分析的参考序列,最后将 Unigene 序列与 Nr、Nt、Pfam、KOG、SwissProt、KEGG 和 GO 数据库比对,获得 Unigenes 的注释信息。如果不同库之间的比对结果有矛盾,则按 Nr、SwissProt、KEGG 和 COG 的优先级确定 Unigene 的序列方向,比对不上的 Unigene 则用软件 ESTScan 预测其编码区并确定序列的方向,以确定对应核苷酸和蛋白质的注释信息和基因的功能信息。

1.4 SSR 和 SNP 的分析方法

SSR 位点搜索主要是利用 MISA 软件(<http://pgrc.ipk-gatersleben.de/misa/>) 搜索所得到的 unigene,参数设置为:单碱基、二碱基、三碱基、四碱基、五碱基、六碱基的最短重复次数分别为 10、6、5、5、5 和 5。

SNP 位点的搜索通过 Samtools 和 Picard-tools 等工具对对比结果进行染色体坐标排序、去掉重复的 Reads 等处理,最后通过变异检测软件 GATK3 进行 SNP Calling,并对原始结果进行过滤(过滤掉质量值>40,距离>2 的 SNP)。

2 结果分析

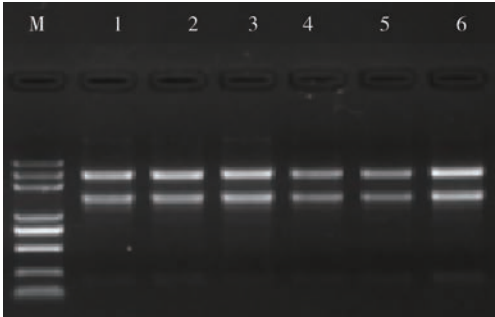
2.1 RNA 质量检验

提取的总 RNA 样品电泳检测结果如图 1,28 和 18S 条带明亮,且亮度为 2 : 1,Aligent2100 检测总 RNA 样品质量,RIN 值都在 8.0 以上,总 RNA 的浓度和总量等指标均已达到测序要求,可用于后续转录组测序等试验。

2.2 测序与序列组装

‘禅寺丸’的雌雄花芽转录组测序数据经去冗余处理后获得 54.47 G clean reads,GC 含量为 48.79%,Q30 值为 93.85%。对 Clean reads 进行拼接获得 221 537 条平均长度为 807 bp 的 transcript 序列,其 N50 为 1 385 bp(图 2(a)),对 transcript 序列进一步组装获得 154 741 条平均长度为 663 bp 的非冗余 Unigene,获得核酸信息 102 528 371 nt,其长度集中在 200~2 000 bp,其中长度为 200~500 bp 的序列最多(97 985 条),占总 unigene 序列的 63.32%,500~2 000 bp 的序列

(47 558 条)占总 unigene 序列的 30.73%,>2 000 bp 的序列占总 unigene 序列的 5.94%(图 2(b))。



M: 5 000bp marker; 1~6 样本 M1、M2、M3、F1、F2、F3。
M1、M2、M3 分别为 3 个雄花芽样品重复; F1、F2、F3 分别为 3 个雌花芽样品重复。
M1、M2、M3 represents three male samples; F1、F2、F3 represents three female samples seperatively.

图 1 总 RNA 琼脂糖电泳图

Fig. 1 Detection of total RNA by electrophoresis

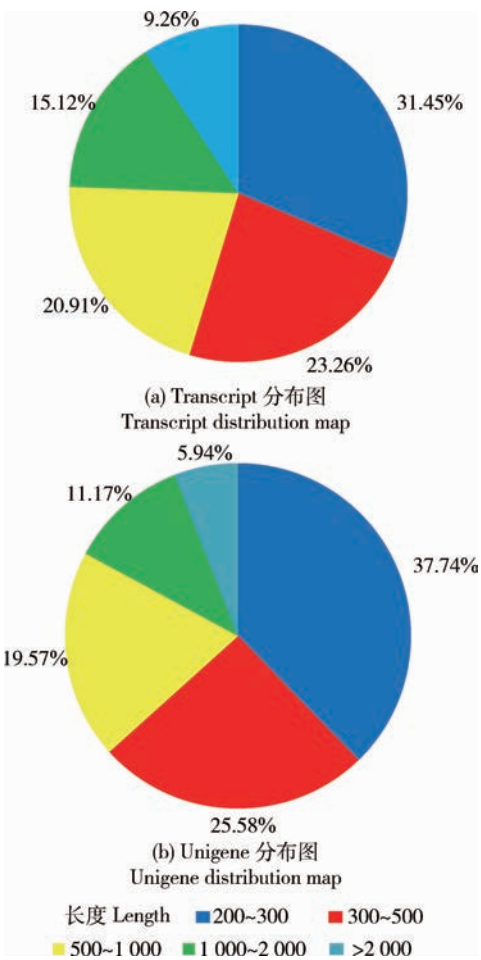


图 2 柿转录组 transcript 和 unigene 数据分类图

Fig. 2 Classification of transcript and unigene in the transcriptome of Diospyros kaki

2.3 Unigene 功能注释

利用 BLAST 软件将 unigene 序列分别与 NR、GO、KOG 和 KEGG 数据库比对分析后,统计每个库上注释到的 unigene 数目(表 1)。结果显示,有 38.49%(59 570 条)的 unigene 在数据库中得到注

释,在七大数据库中均得到注释的 unigene 有 7 100 条。Nr 数据库注释到的 unigene 数量最多(47435 条,30.65%),其次是 SwissProt 数据库(33 320 条,21.53%),KO 数据库注释到的数量最少,为 14 813 条(9.57%)。

表 1 Unigene 注释结果统计表
Table 1 Statistical results of functional annotations in *Diospyros kaki*

数据库 Database	注释基因数 Number of anotated unigenes	占注释基因百分数/% Percentage of anotated unigenes	占总 unigene 百分数/% Percentage of total unigenes
NR	47 435	79.63	30.65
NT	29 886	50.17	19.31
KO	14 813	24.87	9.57
SwissProt	33 320	55.93	21.53
PFAM	32 218	54.08	20.82
GO	32 646	54.80	21.09
KOG	16 098	27.02	10.40
总计 Total	59 570	—	38.50

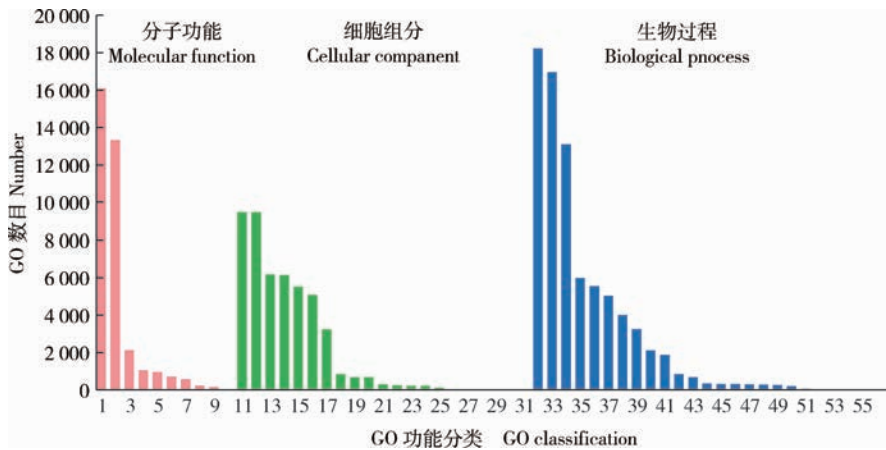
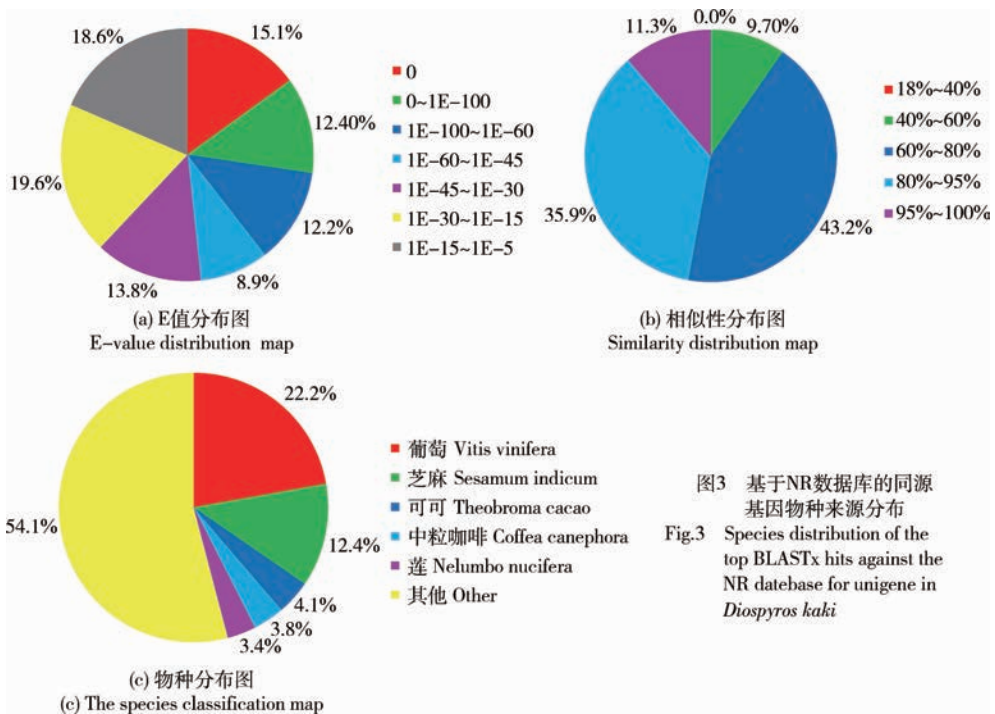
从比对序列同源性的 E-value 结果显示(图 3(a)),47 435 条 Unigene 在 NR 数据库中可比对出同源序列,E-value<1E-30 时,有 62.4%的序列与已知基因具有强同源性,E-value<1E-45 时,只有 48.6%的序列与已知基因有很强的同源性;对 Unigene 序列进行相似性分析,有 90.4%的序列与数据库中已知可匹配序列有>60%的相似性,只有 47.2%的序列与已知序列有大于 80%的相似性(图 3(b));对同源序列所属物种分布进行统计,与葡萄、芝麻和可可等 5 个物种的已知基因具有相似性的 unigene 数量占 NR 数据库注释的总 unigene 数量的 45.9%,其中与葡萄(*Vitis vinifera*)基因的同源序列相似性最高,达到 22.2%,其次依次是芝麻(*Sesamum indicum*)(12.4%)、可可(*Theobroma cacao*)(4.1%)、中粒咖啡(*Coffea canephora*)(3.8%)和莲(*Nelumbo nucifera*)(3.4%)(图 3(c))。

2.4 Unigene 分类

2.4.1 Unigene 的 GO 分类

32 646 条(21.09%)Unigene 比对到 GO 数据

库,可将其分成细胞组分(48 422 条,29.61%)、分子功能(35 251 条,21.56%)和生物过程(79 835 条,48.83%)3 大类别 56 个小类。分子功能又可分为 10 个亚类,其中与结合、催化活性和转运蛋白活性相关的 unigene 占总体的 83.44%,大部分均与分子活性有关,这说明植物体正在进行快速的生长代谢活动;在细胞组分类别中,将已注释的 unigene 分为 21 类,其中与细胞、细胞器、膜功能相关和大分子复合物的 unigene 占 93.10%;根据生物过程将其分为 25 类,其中与细胞过程相关的 unigene 数量最多(18 223 条),其次是与代谢过程相关的(16 952 条),另外 40.11%的 unigene 属于单有机体类别(13 093 条),说明了这些 unigene 的器官特异性(图 4)。生物过程中的代谢过程、细胞过程以及生物调节等基因将为柿花性别分化等研究提供基础数据。细胞组分中的细胞、细胞器、共质体等与花芽分化过程中茎生长点、花器官原基分化功能相关的基因,将为花芽分化分子调控等研究提供基础。



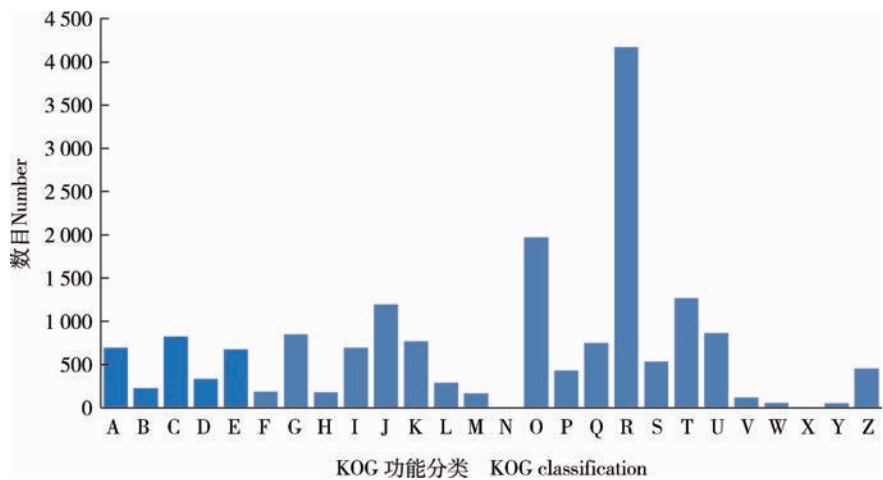
1. 分子功能 Molecular function: 1)结合 Binding;2)催化剂活性 Catalytic activity;3)转运蛋白活性 Transporter activity;4)结构分子活性 Structural molecule activity;5)核酸结合转录因子活性 Nucleic acid binding transcription factor activity;6)分子功能调节因子 Molecular function regulator;7)分子传感器活性 Molecular transducer activity;8)转录因子活性蛋白结合 Transcription factor activity,protein binding; 9)抗氧化活性 Antioxidant activity;10)金属分子伴侣活性 Metallochaperone activity;2. 细胞组分 Cellular component;11)细胞 Cell;12)细胞要素 Cell part;13)细胞器 Organelle;14)大分子复合物 Macromolecular complex;15)膜 Membrane;16)膜要素 Membrane part;17)细胞器部分 Organelle part;18)膜围绕的内腔 Membrane-enclosed lumen;19)病毒体 Virion;20) 病毒体要素 Virion part;21)细胞间区域 Extracellular region;22)细胞间区域要素 Extracellular region part;23)其他有机体 Other organism;24)其他有机体要素 Other organism part;25)细胞外液 Extracellular matrix;26)胞间连丝 Cell junction;27)突触体 Synapse part;28)突触 Synapse;29)细胞外液组分 Extracellular matrix component; 30)共质体 Symplast;31) 核质 Nucleoid;3. 生物过程 Biological process;32)细胞过程 Cellular process;33)代谢过程 Metabolic process;34)单有机体过程 Single-organism process;35)生物调控 Biological regulation;36)生物过程调控 Regulation of biological process;37)定位 Localization;38)刺激应答 Response to stimulus;39)细胞组分组织或生物合成 Cellular component organization or biogenesis;40)信号 Signaling;41)多有机体过程 Multi-organism process;42)多细胞组织过程 Multicellular organismal process;43)发育过程 Developmental process;44)繁殖 Reproduction;45)生物过程正调控 Positive regulation of biological process;46)生物过程负调控 Negative regulation of biological process;47)运动 Locomotion;48)免疫系统过程 Immune system process;49)生殖过程 Reproductive process;50)生物粘附 Biological adhesion;51)生长 Growth;52)细胞杀伤 Cell killing;53)行为 Behavior;54)生物阶段 Biological phase;55)节律过程 Rhythmic process;56)细胞聚集 Cell aggregation

图 4 Unigene 的 GO 功能注释及分类统计结果
Fig. 4 GO function annotation and classification of unigene

2.4.2 Unigene 的 KOG 分类

KOG 库对于预测单个蛋白的功能和整个新基因组中蛋白质的功能是有用的。‘禅寺丸’雌雄花转录组有 16 098 条 Unigene 可以在 KOG 数据库中得到功能注释,分为 26 个类别,其中与功能预测相关

的 Unigene 最多,为 4 166 条,占总数的23.49%,其次是翻译后修饰、蛋白质翻转、分子伴侣(1 971 条, 11.11%),信号转导机制(1 269 条,7.15%)和翻译、核糖体结构和合成(1 196 条,6.74%)等与生物过程相关的 Unigene(图 5)。



A. RNA 加工和修饰 RNA processing and modification;B. 染色质结构和动力学 Chromatin structure and dynamics;C. 能量产生和转化 Energy production and conversion;D. 细胞周期调控、细胞分裂和染色体分区 Cell cycle control, cell division, chromosome partitioning;E. 氨基酸转运和代谢 Amino acid transport and metabolism;F. 核酸转运和代谢 Nucleotide transport and metabolism;G. 碳水化合物转运和代谢 Carbohydrate transport and metabolism;H. 辅酶转运和代谢 Coenzyme transport and metabolism;I. 脂类转运和代谢 Lipid transport and metabolism;J. 翻译、核糖体结构和生物合成 Translation, ribosomal structure and biogenesis;(K)转录 Transcription;L. 复制、重组和修复 Replication, recombination and repair;M. 细胞壁/膜/被膜生物合成 Cell wall/membrane/envelope biogenesis;(N)细胞迁移 Cell motility;O. 翻译后修饰、蛋白翻转、分子伴侣 Posttranslational modification,protein turnover,chaperones;P. 无机离子转运和代谢 Inorganic ion transport and metabolism;Q. 次生代谢物的合成、转运和代谢 Secondary metabolites biosynthesis,transport and catabolism;R. 功能预测 General function prediction only;S. 功能未知 Function unknown;T. 信号转导机制 Signal transduction mechanisms;U. 细胞内运输、分泌和膜泡运输 Intracellular trafficking,secretion,and vesicular transport;V. 防御机制 Defense mechanisms;W. 细胞外结构 Extracellular structures;X. 未知蛋白 Unamed protein;Y. 核结构 Nuclear structure;Z. 细胞骨架 Cytoskeleton

图 5 KOG 分类图

Fig. 5 KOG function classification by BLASTx against COG databases

2.4.3 Unigene 的 KEGG 分类

依据 KEGG 注释信息共识别出 14 813 条 Unigene,包括碳水化合物代谢,翻译,折叠、分类和降解等共 19 个途径(表 2)。利用 KEGG 数据库对非冗余 Unigene 的 Pathway 进行富集性分析,结果表明所有的 Unigene 可归为 129 个信号途径,其中碳水化合物代谢途径的 Unigene 数量最多,有 1 714 个,占有所有非冗余 Unigene 的 13.39%,而异黄酮生物合成途径只有一条 Unigene,数量最少。

2.5 微卫星特征分析

2.5.1 柿转录组序列中 SSR 位点的分布

在柿花转录组的 154 741 条 unigene 序列中发现 44 304 个 SSR,其中包含 2 305 个混合型 SSR 和

41 999 个完整型 SSR 位点,完整型 SSR 占总 SSR 位点的 94.80%。包含 2 个及以上 SSR 位点的 unigene 共有 6 744 条。SSR 位点的出现频率为 0.43 个/kb,即每 2.3 kb 就出现 1 个 SSR。

完整型 SSR 位点共包含 83 种重复基元,单核苷酸至六核苷酸分别有 2、4、10、31、18 和 18 种。其中完整型 SSR 重复基元的重复次数均在 5~26 次,其中以重复 10 次(7 048 个)的最多,占总 SSR 的 16.78%;其次为重复次数 6、7、8、11、9 和 5 次,其 SSR 位点个数分别为 5 788、3 930、3 679、3 583、3 543 和 3 473。重复 5~10 次的 SSR 位点共有 27 461 个,占总 SSR 的 65.38%,重复 11~26 次的 SSR 位点有 14 538 个,占 34.62%(图 6)。

表 2 柿花芽转录组的 KEGG 代谢途径分析

Table 2 KEGG classification of flower bud in *Diospyros kaki*

编号 Number	代谢通路类别 KEGG Pathway	代谢通路名称 Name of metabolic pathway	Gene 数目 Number of gene	所占比例/% The percentage
1	细胞过程	运输和分解代谢	760	5.94
2	环境信息处理	膜运输	106	0.83
3		Signal transduction(信号转导)	482	3.77
4		Folding, sorting and degradation(折叠,分拣和降解)	1 226	9.58
5	遗传信息处理	Replication and repair(复制和修复)	180	1.41
6		Transcription(转录)	480	3.75
7		Translation(翻译)	1 455	11.37
8	代谢	Amino acid metabolism(氨基酸代谢)	919	7.18
9		Biosynthesis of other secondary metabolites(其他次生代谢产物的生物合成)	447	3.49
10		Carbohydrate metabolism(碳水化合物代谢)	1 714	13.39
11		Energy metabolism(能量代谢)	837	6.54
12		Glycan biosynthesis and metabolism(聚糖的生物合成与代谢)	218	1.7
13		Lipid metabolism(脂类代谢)	688	5.37
14		Metabolism of cofactors and vitamins(辅助因子和维生素代谢)	378	2.95
15		Metabolism of other amino acids(其他氨基酸代谢)	490	3.83
16		Metabolism of terpenoids and polyketides(萜类化合物和聚酮化合物的代谢)	325	2.54
17		Nucleotide metabolism(核苷酸的代谢)	418	3.27
18		Overview(概观)	1 113	8.69
19	有机系统	Environmental adaptation(环境适应)	565	4.41

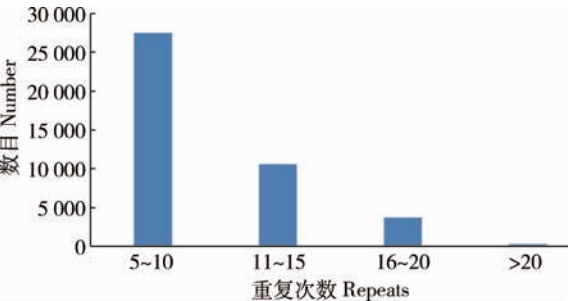


图 6 SSR 重复次数分布

Fig. 6 Number of repeats

2.5.2 柿转录组微卫星优势重复单元碱基的特征分析

完整型 SSR 位点共包含 83 种重复基元,单核苷酸至六核苷酸分别有 2、4、10、31、18、18 种。以 A/T 类型为主的单核苷酸重复所占的比例最高(20 006 个,47.63%);其次是以 AG/CT 类型为主的二核苷酸重复(16 055 个,38.23%);再次是以 AAG/CTT 类型为主的三核苷酸重复(5 402 个,12.87%)和以 AAAG/CTTT 类型为主四核苷酸重

复(500 个,1.19%);五和六核苷酸重复基元的类型较多,但数量较少,所占比例均为 0.04%,出现频率较低(图 7)。检测到的 SSR 位点分析显示,出现频率最高的 6 类基序依次为: A/T (17 748 个,

42.26%)、AG/CT(10 382 个,24.72%)、AT/AT (3 607 个,8.59%)、C/G (2 258 个,5.38%)、AC/GT(1 996 个,4.75%)和 AAG/CTT(1 919 个,4.57%)。

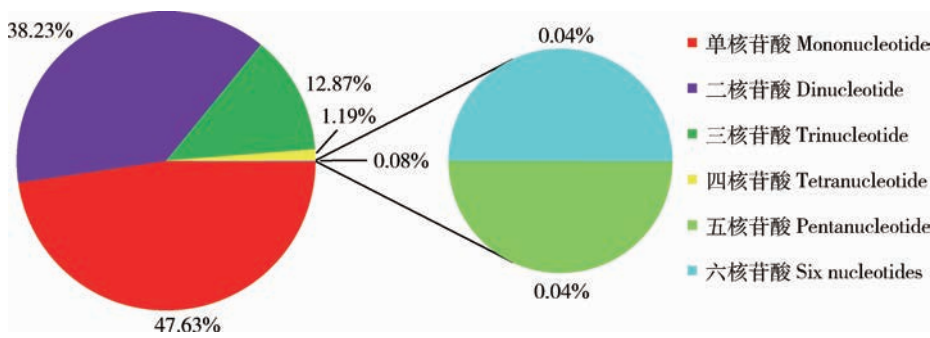


图 7 柿完整型 SSR 重复基元种类分布图

Fig. 7 Types and distribution of complete SSRs in *Dispyros kaki*

微卫星重复基元统计结果显示,单碱基重复基元中 A/T 为主要的重复基元,共有 17 748 个,占 88.71%,C/G 仅占 11.29%;二碱基重复单元有 4 种(AC/GT、AG/CT、AT/AT、CG/CG),其中 AG/CT 重复基元的数量最多(10 382 个),占总体的 64.67%,其次是 AT/AT(3 607 个)和 AC/GT(1 996 个),分别占 22.47%和 12.43%。最少的是 CG/CG(70 个,0.44%);三碱基重复单元类型主要有 10 个,AAG/CTT(1 919 个,35.52%)的数量最多,其次是 AAT/ATT(657 个,12.16%)、AGG/CCT(535 个,9.90%),再次是 ATC/ATG(485 个,8.98%)、AAC/GTT(480 个,8.89%)和 AGC/CTG(457 个,8.46%),其余重复类型相对较少,ACT/AGT(96 个,1.78%)最少;四碱基重复类型最多,有 31 种,AAAG/CTTT 的数量最多,有 86 个,占总体的 17.20%,其次是 AAAT/ATTT(72 个,14.4%)和 ACAT/ATGT(53 个,10.6%),再次是 AGAT/ATCT(37 个,7.4%)、ATCC/ATGG(30 个,6.0%)、AAAC/GTTT(24 个,4.8%)和 AATT/AATT(21 个,4.2%) AAGG/CCTT(19 个,3.8%)和 AGCC/CTGG(19 个,3.8%),其他重复类型则相对较少;五碱基和六碱基重复类型均有 18 种,但每种类型均只有 1 个。

2.5.3 微卫星的长度分布

微卫星长度也存在极显著变异,长度变化范围为 10~426 bp,平均长度为 17 bp (图 8)。以重复长度为 10~15 bp 的短序列最多,其次为长度在

16~19 bp(28.42%),长度大于 20 bp 的长序列仅占微卫星总数的 13.15%。

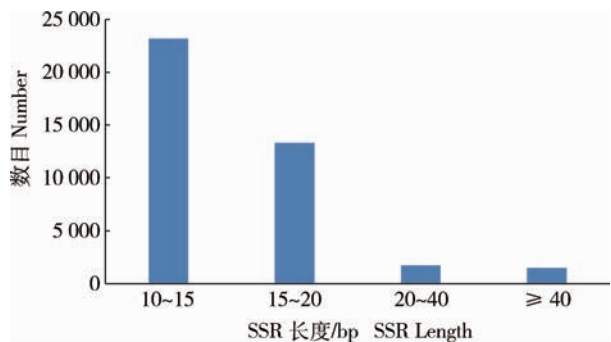


图 8 SSR 长度分布

Fig. 8 Distribution of SSR length

2.5.4 SNP 位点的特征

在转录组得到的 unigene 中共发现 SNP 405 685 个,发生频率为 1/253 bp,即每 253 bp 就会有 1 个 SNP 位点出现,其中转换 255 259 个(62.92%),颠换 150 426 个(37.08%)。6 种单核苷酸变异中以转换类型的 C/T 和 A/G 发生频率最高,分别达到总数的 31.51%和 31.41%,其他 4 种颠换类型的 SNP A/T、A/C、T/G、C/G 分别占到总数的 10.11%、8.84%、8.71%和 9.42%。其中转换类型显著高于颠换类型,在转换类型中,C/T 发生频率最高,这可能与 SNPs 在 CG 序列上出现的最为频繁,而 C(胞嘧啶)常以甲基化形式存在,脱氨后即成为 T(胸腺嘧啶)有关。

3 结论与讨论

本研究利用 Misa 软件在 33 639 条 unigene 中搜索到了 41 999 条完整型 SSR, 包含 SSR 的序列的发生频率为 21.74%, SSR 的出现频率为 0.43 个/kb, 即平均每 2.31 kb 出现 1 个 SSR。比茶树 (0.272 个/kb)^[23]、油茶 (0.417 个/kb)^[24]、桃 (0.313 个/kb)^[25] 的分布密度都要大, 这进一步表明本研究中柿的 SSR 标记的数量是极其丰富的, 有望在柿属植物的起源、进化等多个领域得到广泛应用。研究方法上采用从转录组测序的序列信息中开发 SSR 引物, 较其他的 SSR 标记开发的方法更为简便和精确^[13,15-17]。

微卫星序列会对基因功能产生重要的影响, 其序列特征是了解不同物种基因组差异的重要指标。研究表明, 柿花转录组微卫星种类较为丰富, 1-6 核苷酸重复基元均有出现, 优势重复基元为单核苷酸重复类型, 占 SSR 总数的 47.63%, 其次为二核苷酸重复类型 (38.23%) 和三核苷酸重复类型 (12.86%), 这与大多数植物中 SSR 的主要类型是二、三核苷酸重复类型不同^[23,26-27]。可能是由于物种不同和微卫星搜索条件设定并没有统一的标准, 而本研究中设定的搜索重复次数偏低。

柿优势碱基重复单元中所占比例最高的分别是 A/T (42.26%), 与油松^[28] 的研究一致。其次是分别在二、三碱基单元中比例最高的 AG/CT (24.72%) 和 AAG/CTT (4.57%), 这与王丽鸳^[25] 关于二、三碱基单元中的优势碱基的研究结果一致, 而在水稻、高粱、玉米^[29]、百合^[30] 中 GGC/CCG 为三碱基重复类型中数量最多的研究结果不同, 这可能是由于不同物种相应编码蛋白的使用频率不同有关。

SSR 的长度和重复次数是影响其分子标记的多态性的重要因素。Tenmykh^[31] 认为, 当 SSR 长度大于等于 20 bp 时多态性较高, 长度在 12~20 bp 的 SSR 多态性中等, 长度在 12 bp 以下时多态性极低。在本研究中, 微卫星特征分析得到的 SSR 序列长度在从 10~426 bp 不等, 平均长度为 17 bp。其中, 以重复长度小于 20 bp 的短序列最多, 重复长度大于 20 bp 的长序列仅占微卫星总数的 13.15%, 高于茶树 (12.13%) 低于杨树 (18.90%) 和桉树 (33.30%)^[32], 此类 SSR 位点具有高度多态性, 可能在柿 SSR 标记上有较高的利用

价值。单核苷酸多态性在植物基因组中广泛存在。本研究从柿转录组测序数据中获得了 405 685 个 SNP 位点, 其发生频率为 1/253 bp, 而水稻和玉米基因组中 SNP 的发生频率分别为每 89 bp 有一个 SNP^[33], 平均每 61 bp 有 1 个 SNP^[34]。可见, SNP 在柿中的发生频率较低。柿中 SNP 位点以 C/T 和 A/G 的碱基转换类型为主, 分别占总数的 31.51% 和 31.41%, 这与万志兵^[18] 对菊花中 EST-SNP 特征的研究结果一致。

参考文献 References

- [1] 梁玉琴, 韩卫娟, 张嘉嘉, 孙鹏, 梁晋军, 傅建敏. 河南省柿种质资源表型多样性研究[J]. 中国农业大学学报, 2015, 20(1): 74-85
Liang Y Q, Han W J, Zhang J J, SUN P, Liang J J, Fu J M. Phenotypic diversity of persimmon germplasm in Henan Province[J]. *Journal of China Agricultural University*, 2015, 20(1): 74-85 (in Chinese)
- [2] 杨勇, 阮小凤, 王仁梓, 李高潮. 柿种质资源及育种研究进展[J]. 西北林学院学报, 2005, 20(2): 133-137
Yang Y, Ruan X F, Wang R Z, Li G C. Advances in research of germplasm resources and breeding of *Dispyros kaki* [J]. *Journal of Northwest Forestry University*, 2005, 20(2): 133-137 (in Chinese)
- [3] 罗正荣, 蔡礼鸿, 胡春根. 柿属植物种质资源及其利用研究现状[J]. 华中农业大学学报, 1996, 15(4): 381-388
Luo Z R, Cai L H, Hu C G. Research development of germplasm resources of dispyros and their utilization [J]. *Journal of Huazhong Agricultural University*, 1996, 15(4): 381-388 (in Chinese)
- [4] 袁录霞, 张青林, 郭大勇, 罗正荣. 中国甜柿及其在世界甜柿基因库中的地位[J]. 园艺学报, 2011, 38(2): 361-370
Yuan L X, Zhang Q L, Guo D Y, Luo Z R. Characteristics of Chinese PCNA types Sinica [J]. *Acta Horticulturae Sinica*, 2011, 38(2): 361-370 (in Chinese)
- [5] 罗正荣, 李发芳, 蔡礼鸿. 部分中国原产甜柿种质的分子系统学研究[J]. 园艺学报, 1999, 26(5): 297-301
Luo Z R, Li F F, Cai L H. Molecular systematic of China native nonastringent persimmon based on random amplified polymorphic DNA [J]. *Acta Horticulturae Sinica*, 1999, 26(5): 297-301 (in Chinese)
- [6] 罗正荣, 米森敬三, 杉浦明. 应用 RAPD 技术研究柿品种间的亲缘关系[J]. 果树学报, 1998(4): 311-316
Luo Z R, Yonemori K, Shan P M. Study on the genetic relationship among persimmon cultivars by RAPD [J]. *Journal of Fruit Science*, 1998(4): 311-316 (in Chinese)
- [7] Naval M D M, Zuriaga E, Pecchioli S. Analysis of genetic diversity among persimmon cultivars using microsatellite

- markers[J]. *Tree Genetics & Genomes*, 2010, 6(5): 677-687
- [8] 邓立宝,何新华,卢家仕. 基于 SSR 分子标记的广西柿种质资源遗传多样性分析[J]. 南方农业学报, 2014, 45(11): 1903-1909
- Deng L B, He X H, Lu J S. Genetic diversity analysis of *Dispyros* L germplasm resources in Guangxi using SSR molecular marker[J]. *Journal of Southern Agriculture*, 2014, 45(11): 1903-1909 (in Chinese)
- [9] 耿攀,阮小凤,杨勇,赵红星. 柿属植物种质资源遗传多样性的 SSR 分析[J]. 西北农林科技大学学报: 自然科学版, 2010, 38(12): 190-196
- Geng P, Ruan X F, Yang Y, Zhao H. Analysis of genetic diversity of *Diospyros* spp germplasm resources by using SSR markers[J]. *Journal of Northwest Agriculture and Forestry University: Natural Science Edition*, 2010, 38(12): 190-196 (in Chinese)
- [10] 陈方永,倪海枝,何理坤,王引,卢国耀,王冬米,任正初. 浙江部分柿属植物鉴定及亲缘关系的 SSR 分析[J]. 果树学报, 2011(1): 46-50
- Chen F Y, Ni H Z, He L K, Wang Y, Lu G Y, Wang D M, Ren Z C. Identification and genetic relationships analysis of some oriental persimmon (*Diospyros kaki*) and related species in Zhejiang Province based on SSR analysis[J]. *Journal of Fruit Science*, 2011(1): 46-50 (in Chinese)
- [11] Liang Y, Han W, Sun P. Genetic diversity among germplasms of *Diospyros kaki* based on SSR markers [J]. *Scientia Horticulturae*, 2015, 186: 180-189
- [12] 艾呈祥,秦志华,陶吉寒,王长君. 32 个柿主栽品种 SSR 图谱构建及遗传变异分析[J]. 西北植物学报, 2011, 31(11): 2185-2191.
- Ai C X, Qin Z H, Tao J H, Wang C J. SSR fingerprints and genetic variations of the 32 persimmon major cultivars[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2011, 31(11): 2185-2191 (in Chinese)
- [13] Soriano J M, Pecchioli S, Romero C, Badenes M L. Development of microsatellite markers in polyploid persimmon (*Diospyros kaki*) from an enriched genomic library [J]. *Molecular Ecology Notes*, 2006, 6(2): 368-370
- [14] 阮小凤. 柿 SSR 富集 DNA 文库构建与筛选[D]. 杨凌: 西北农林科技大学, 2005
- Ruan X F. Construction and screening of SSR-enriched genomic library in persimmon[D]. Yangling: Northwest Agriculture and Forestry University, 2005 (in Chinese)
- [15] 郭大龙,罗正荣. 柿和君迁子 SSR 分析技术的建立[J]. 农业生物技术学报, 2004, 12(4): 386-389
- Guo D L, Luo Z R. Establishment of SSR technique of *Diospyros kaki* and *Dispyros lotus*[J]. *Journal of Agricultural Biotechnology*, 2004, 12(4): 386-389 (in Chinese)
- [16] Guo D L, Luo Z R. Microsatellite isolation and characterization in Japanese persimmon (*Diospyros kaki*) [J]. *Biochemical Genetics*, 2008, 46(5): 323-328
- [17] 郭大龙,罗正荣. 一种利用 ISSR 开发 SSR 引物的方法[J]. 植物生理学报, 2006, 42(2): 268-270
- Guo D L, Luo Z R. A method for development of SSR primers from ISSR [J]. *Plant Physiology Communications*, 2006, 42(2): 268-270 (in Chinese)
- [18] 万志兵,王言言,陈黎. 菊花 EST-SNP 的发掘与特征分析[J]. 东北林业大学学报, 2013, 41(5): 84-86
- Wan Z B, Wang Y Y, Chen L. Detection and characteristic analysis EST-SNP in chrysanthemum [J]. *Journal of Northeast Forestry University*, 2013, 41(5): 84-86 (in Chinese)
- [19] Chagne D, Gasic K R, Han Y. Development of a set of SNP markers present in expressed genes of the apple[J]. *Genomics*, 2008, 92(92): 353-358
- [20] 李小白,向林,罗洁,胡标林,田胜平,谢鸣,孙崇波. 转录组测序 (RNA-seq) 策略及其数据在分子标记开发上的应用[J]. 中国细胞生物学学报, 2013(5): 720-726
- Li X B, Xiang L, Luo J, Hu B L, Tian S P, Xie M, Sun C B. The strategy of RNA-seq, application and development of molecular marker derived from RNA-seq [J]. *Chinese Journal of Cell Biology*, 2013(5): 720-726 (in Chinese)
- [21] Akagi T, Henry I M, Tao R, Comai L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons [J]. *Science*, 2014, 346(6209): 646-650
- [22] 李加茹,孙鹏,韩卫娟,李芳东,傅建敏,刁松峰. 不完全甜柿‘禅寺丸’花性别分化形态学关键时期的研究[J]. 园艺学报, 2016, 43(3): 451-461
- Li J R, Sun P, Han W J, Li F D, Fu J M, Diao S F. Morphological key period study on floral sex differentiation in pollination-constant and non-astringent persimmon ‘Zenjimaru’ [J]. *Acta Horticulturae Sinica*, 2016, 43(3): 451-461 (in Chinese)
- [23] 杨华,陈琪,韦朝领,史成颖,方从兵,宛晓春. 茶树转录组中 SSR 位点的信息分析[J]. 安徽农业大学学报, 2011, 38(6): 882-886
- Yang H, Chen Q, Wei C L, Shi C Y, Fang C B, Wan X C. Analysis on SSR information in *Camellia sinensis* transcriptome [J]. *Journal of Anhui Agricultural University*, 2011, 38(6): 882-886 (in Chinese)
- [24] 史洁,尹冬明,管宏伟,戴晓港,陈金慧,施季森. 油茶基因组微卫星特征分析[J]. 南京林业大学学报: 自然科学版, 2012, 36(2): 47-51
- Shi J, Yin T M, Guan H W, Dai X G, Chen J H, Shi J S. Characteristic analysis of microsatellites of *Camellia spp* [J]. *Journal of Nanjing Forestry University: Natural Science Edition*, 2012, 36(2): 47-51 (in Chinese)
- [25] Wang L, Zhao S, Gu C. Deep RNA-Seq uncovers the peach transcriptome landscape [J]. *Plant Molecular Biology*, 2013, 83(4-5): 365-377
- [26] 马秋月,廖卓毅,张得芳,戴晓港,陈赢男,李淑娟. 碧桃花瓣转录组微卫星特征分析[J]. 南京林业大学学报: 自然科学版,

2015(3):34-38

Ma Q Y, Liao Z Y, Zhang D F, Dai X G, Chen Y N, Li S X. Deep sequenced-based transcriptome analysis of microsatellites in peach(*Prunus persica* cv. duplex) flowers[J]. *Journal of Nanjing Forestry University: Natural Sciences Edition*, 2015 (3):34-38 (in Chinese)

[27] 温强,徐林初,江香梅. 基于 454 测序的油茶 DNA 序列微卫星观察与分析[J]. 林业科学, 2013, 49(8): 43-50

Wen Q, Xu L C, Jiang X M. Survey and analysis of microsatellites from DNA sequences in Camellia species using 454 pyrosequencing[J]. *Scientia Sive Sinicae*, 2013, 49(8): 43-50 (in Chinese)

[28] 安俊,郑彩霞,姚阳,陈彬丽. 油松胚珠转录组微卫星特征分析[J]. 北京林业大学学报, 2016, 38(5): 77-83

An J, Zheng C X, Yao Y, Chen B L. Characterization of microsatellites of ovule transcriptome in pinus tabulaeformis [J]. *Journal of Beijing Forestry University*, 2016, 38(5): 77-83 (in Chinese)

[29] Kantety R V, Rota M L, Matthews D E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat[J]. *Plant Molecular Biology*, 2002, 48(5-6): 501-10

[30] 杜方. 百合不同器官转录组分析及 SSR 标记开发应用[D]. 杭州:浙江大学, 2014

Du F. Transcriptome analysis of different lily organs and development and applications of SSR markers[D]. Hangzhou: Zhejiang University, 2014 (in Chinese)

[31] Temnykh S, Declerck G, Lukashova A. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L)[J]. *Genome Research*, 2001, 11: 1441-1452

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

Yan M M, Dai X G, Li S X, Yin T M. Sequence analysis and comparison of EST-SSRs in pine, polar and *Eucalyptus* [J]. *Genomic and Applied Biology*, 2011(1): 103-109 (in Chinese)

[33] Nasu S, Suzuki J, Ohta R. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers [J]. *DNA Research*, 2002, 9(5): 163-71

[34] Ching A, Caldwell K S, Jung M. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines [J]. *BMC Genetics*, 2002, 3(7): 19

责任编辑：王燕华