

DOI:10.12171/j.1000-1522.20200337

基于 RAD-seq 技术的金银花 SSR 标记开发及鉴定

李 慧^{1,2} 刘东超² 徐瑞瑞² 侯立娜^{1,2} 王天琪²
刘忠华^{1,2} 付 晓³ 李圣波⁴

(1. 北京林业大学林木遗传育种国家工程实验室, 林木花卉遗传育种教育部重点实验室, 国家林业局花卉育种与生物工程重点开放实验室, 北京 100083; 2. 北京林业大学生物科学与技术学院, 北京 100083; 3. 平邑县金银花果茶发展促进中心, 山东 平邑 273300; 4. 山东亚特生态技术有限公司, 山东 莒南 266071)

摘要:【目的】通过开发特异性高的多态性分子标记位点,以弥补金银花在简单重复序列(SSR)标记方面的匮乏,推动金银花在遗传资源管理及良种鉴定等方向的研究,为后续全基因组测序及组装策略奠定基础。【方法】运用 RAD-seq 技术对 2 份金银花样品进行简化基因组测序,利用 MISA 软件识别 contig 上的 SSR 序列并对各类型序列特征进行归纳分析,进而设计引物并利用生物信息学的方法进行引物筛选和有效性检测。【结果】对‘九丰一号’和‘亚特’金银花进行简化基因组测序,分别获得 27.805 Mbp 和 42.560 Mbp 干净读长。在组装的 45 850 个基因序列中共检测到 46 999 个 SSR 位点,其中单核苷酸重复单元比例最高(49.15%),六核苷酸重复单元最少(0.20%)。SSR 位点的重复基序以(A/T)_n 为主,具有偏向性。除单碱基和二碱基重复类型外,SSR 位点基序的重复次数集中在 5~6 次,且随重复次数增加,SSR 位点重复类型出现的频率呈下降趋势。金银花基因组 SSR 序列长度介于 10~310 bp 之间,随重复次数增加,各 SSR 序列出现的频率逐渐下降。利用 Primer3.0 成功设计出 38 507 对 SSR 引物,设计成功率为 81.93%。对挑选的 35 对 SSR 引物进行多态性分析,等位基因数、观测杂合度和多态性信息含量(PIC)的平均值分别为 5.057、0.363 和 0.568,其中高多态信息位点 26 个,中度多态信息位点 9 个,均未偏离哈迪-温伯格平衡。【结论】利用 RAD-seq 技术可实现 SSR 标记的大量开发和多态性 SSR 引物的筛选,并为金银花在遗传多样性分析和种质鉴定等相关方面的研究提供数据支持。

关键词: 金银花;简单重复序列(SSR);RAD-seq;多态性

中图分类号: S793.9;S567.9 **文献标志码:** A **文章编号:** 1000-1522(2021)06-0108-10

引文格式: 李慧,刘东超,徐瑞瑞,等. 基于 RAD-seq 技术的金银花 SSR 标记开发及鉴定 [J]. 北京林业大学学报, 2021, 43(6): 108-117. Li Hui, Liu Dongchao, Xu Ruirui, et al. Development and identification of SSR markers based on RAD-seq of *Lonicera japonica* [J]. Journal of Beijing Forestry University, 2021, 43(6): 108-117.

Development and identification of SSR markers based on RAD-seq of *Lonicera japonica*

Li Hui^{1,2} Liu Dongchao² Xu Ruirui² Hou Lina^{1,2} Wang Tianqi²
Liu Zhonghua^{1,2} Fu Xiao³ Li Shengbo⁴

(1. National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, The Tree and Ornamental Plant Breeding and Biotechnology Laboratory of State Forestry Administration, Beijing Forestry University, Beijing 100083, China;
2. School of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China;
3. Pingyi County Honeysuckle Nectar Development Promotion Center, Pingyi 273300, Shandong, China;
4. Shandong Yate Eco-Tech, Co., Ltd., Junan 266071, Shandong, China)

收稿日期: 2020-11-09 修回日期: 2020-12-16

基金项目: 国家林业局业务委托项目(2019104011), 北京林业大学一流学科建设项目(2019XKJS0308), 北京市园林绿化局计划项目(2019-STBHXFC-02-022)。

第一作者: 李慧。主要研究方向: 经济林品种鉴定。Email: hyacinth1107@163.com 地址: 100083 北京市海淀区清华东路 35 号北京林业大学生物科学与技术学院。

责任作者: 刘忠华, 博士, 副教授。主要研究方向: 植物生长发育与系统进化。Email: liuzh6@bjfu.edu.cn 地址: 同上。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

Abstract: [Objective] Through the development of polymorphic molecular markers to make up for the lack of *Lonicera japonica* in SSR markers, this paper aims to promote the research in genetic resource management and variety identification, and lay the foundation for future de novo assembling of *Lonicera japonica* genome. [Method] RAD-seq (restriction-site associated DNA sequencing) was used to conduct simplified genome sequencing for two *Lonicera japonica* samples. The SSR sequences were identified by MISA and the characterizations of SSR sequences and their polymorphisms were analyzed by bioinformatics methods. [Result] 27.805 Mbp and 42.560 Mbp clean reads were obtained in ‘Jiufengyihao’ and ‘Yate’ using restriction-site associated DNA (RAD). A total of 46 999 SSR loci were detected among the 45 850 gene sequences assembled, with the highest proportion of mono-nucleotide repeat SSRs (49.15%) and the lowest proportion of hexa-nucleotide repeat SSRs (0.20%). The repeated motifs at SSR loci were dominated by (A/T) n and showed a bias. Apart from mono-nucleotide and di-nucleotide repetition types, the repeat counts of SSR motifs mainly ranged from 5 to 6. The frequency of SSRs repetition types showed a downward trend with the increase of repeat number. The range of SSR length ranged from 10 to 310 bp, and the frequency of SSR sequence tended to decrease as the number of repetitions increasing. 38 507 pairs of SSR primers were successfully designed, with a design success rate of 81.93%. 35 pairs of SSR primers were analyzed for polymorphism, in which the average number of alleles, observed heterozygosity and PIC were 5.057, 0.363 and 0.568, among which there were 26 highly polymorphic loci and 9 moderately polymorphic loci, all of which did not deviate from the Hardy-Weinberg equilibrium. [Conclusion] The large-scale development of SSR markers and the screening of polymorphic SSR primers can be realized based on RAD-seq technology, and provide data support for the research of *Lonicera japonica* in genetic diversity analysis and germplasm identification.

Key words: *Lonicera japonica*; simple sequence repeat (SSR); RAD-seq; polymorphism

金银花(*Lonicera japonica*)别名忍冬、金银藤、银花等,为忍冬科(Caprifoliaceae)忍冬属(*Lonicera*)重要的经济树种^[1],山东、湖南、广西等地作为金银花的道地产区,具有丰富的种质资源。由于长期受新旧产区交互引种、气候环境变化以及栽培措施等方面的影响,金银花在长期栽培种植过程中出现了许多农家栽培品种或品系^[2]。近年来,随着金银花用量的增加以及在食品、化妆品等领域的旺盛需求,市场上作为金银花流通使用的品种混杂,混淆品及代用品在质量、产量和药性等方面的差异严重制约了金银花良种选育及产业发展^[3-4]。因此,开展金银花遗传多样性分析、指纹图谱构建和良种选育等方面的研究,对于金银花遗传资源的保护和发展具有重要意义。

分子标记鉴定是品种间在基因组水平上关于遗传多样性的直接反映,目前常用的分子标记主要包括限制性内切酶片段长度多态性(restriction fragment length polymorphism, RFLP)、简单重复序列区间扩增多态性(inter-simple sequence repeat, ISSR)、随机扩增多态性 DNA(randomly amplified polymorphic DNA, RAPD)、简单重复序列(simple sequence repeat, SSR)、序列标签简单重复序列(expressed sequence tag-

simple sequence repeat, EST-SSR)和扩增片段长度多态性(amplified fragment length polymorphism, AFLP)等。近年来,分子标记被逐渐应用于金银花种质资源保护和良种鉴定等领域中。徐石勇等^[5]利用从 GenBank 数据库中查询到的金银花核酸序列,结合 BatchPrimer3 在线软件,设计出 25 对多态性引物,遗传多样性分析的结果显示 SSR 标记技术具有快速、稳定的优点,具备金银花种质资源标准化分析及中药材市场化检测应用的价值。张贵生等^[4]从金银花转录组中获得 15 对有效 SSR 标记,并通过鉴定和进化树分析对 10 种金银花进行了亲缘关系比较。孙稚颖等^[6]利用 12 条引物对栽培金银花、野生忍冬和灰毡毛忍冬(*Lonicera macranthoides*)等 36 个金银花品种进行 ISSR 分析,发现金银花资源具有丰富的遗传多样性,为良种选育和资源开发提供了重要物质基础。在众多的分子标记技术中,SSR 标记因其共显性、特异性和高重复性,比 RAPD、EST-SSR 和 AFLP 等分子标记更具优势^[7],因而被广泛用于遗传图谱构建、目标基因标定和种质资源保护等研究^[8]。近些年,高通量测序技术的飞速发展使得人们开始利用全基因组数据进行 SSR 标记的大规模开发,并建立 SSR 检索数据库,以便更好地利用分子标记^[9-10]。

尽管如此,由于金银花全基因组数据库尚未建立,大部分与金银花基因相关的研究只是一个“近似值”^[3],因此与其他物种相比,现有的标记引物的数量、类型及鉴别能力仍十分有限。

RAD-seq (restriction-site associated DNA sequencing)是基于高通量测序技术发展而来的、极大程度覆盖全基因组变异信息的简化基因组测序技术。由于只对酶切片段进行测序,降低了基因组的复杂度,且不受参考基因组的限制,目前该技术被广泛应用于包括非模式生物在内的任何物种的分子标记开发、QTL定位、全基因组关联分析及分子育种等方向^[11-12]。本研究利用RAD-seq技术对金银花材料进行简化基因组测序,从位点数量、重复基序类型和重复次数等方面对SSR序列特征进行分析总结,并首次探讨了利用该技术进行金银花SSR标记大规模开发的可行性,以期为进一步开展金银花遗传多样性分析、指纹图谱构建及良种鉴定等研究提供参考。

1 材料与方法

1.1 材料

供试12份金银花样本中由山东平邑县金银花果茶发展促进中心和莒南县亚特生态技术有限公司提供,主要采自山东平邑县(‘九丰一号’‘台尔曼忍冬’‘北花一号’‘紫茎子’‘鹅翎筒’‘对花子’)、山东莒南县(‘亚特’‘九绿一号’)河南封丘县(‘大毛花’)、河北巨鹿县(‘细毡毛忍冬’‘叶里齐’)和湖南隆回县(‘华南忍冬’)等地,其中‘九丰一号’和‘亚特’金银花用于简化基因组测序。采集各植株幼嫩叶片,变色硅胶干燥,备用。

1.2 DNA提取

利用自动磨样机,结合改良后的CTAB法^[13]提取植物基因组DNA。经1%的琼脂糖凝胶电泳检测DNA的质量后,利用NanoDrop 2000分光光度计(thermo scientific)检测其浓度和OD₂₆₀/OD₂₈₀值,将检测结果在1.8~2.0之间的DNA样品的质量浓度稀释至20 ng/μL, -20℃保存备用。

1.3 简化基因组测序

利用RAD-seq技术,对‘九丰一号’和‘亚特’的基因组DNA进行简化测序。大致操作流程如下^[14]:限制性酶切样品基因组DNA,经过P1接头连接、混合液纯化、打断DNA、添加P2接头等操作构建RAD文库。采用Illumina Miseq平台对RAD文库进行高通量测序,对原始测序数据(raw data)进行过滤、评估等获取干净读长(clean reads)。过滤标准^[15]为:(1)去除序列两端的低质量区域,质量阈值为Q30;

(2)去除接头序列;(3)去除含模糊碱基N的序列;(4)去除长度小于60 bp的序列。对过滤后获得的高质量序列进行拼接、延伸得到contig序列,然后对拼接后的序列进行聚类,去除其中的冗余得到最终的拼接结果。具体实验流程由北京源宜基因科技股份有限公司完成。

1.4 SSR位点识别及引物设计

根据测序得到的基因组序列数据,利用MISA(<http://pgrc-ipk-gatersleben.de/misa/>)软件查找金银花基因组数据库中的SSR位点。标准如下:单、二、三、四、五、六核苷酸的最低重复次数为10、6、5、5、5、5次,相邻SSR之间的最大间隔区域长度为100 bp。

根据所查找的金银花SSR位点两侧的保守序列,借助Primer3.0(v2.3.7; <http://primer3.sourceforge.net/>)设计引物^[16],参数设置为:引物长度18~30 bp,最优为20 bp;退火温度为57~61℃,最优60℃;GC含量40%~60%;产物长度100~300 bp。

1.5 PCR扩增及SSR检测

本研究共需3种引物(北京睿博兴科生物技术有限公司)^[17]:5'端带有通用型M13序列(5'-TGTAACGACGCGCCAGT-3')的上游引物,下游引物及3'端具有荧光染料标记(FAM、HEX、TAMRA、ROX)的M13引物,通过与在引物5'末端添加的未标记M13互补,将标记的M13添加到PCR反应中以检测PCR扩增产物。20 μL反应体系包括:2 μL模板DNA(20 ng/μL),10 μL 2×Taq PCR Mix(天根生物公司),0.08 μL(10 μmol/L)上游引物,0.32 μL(10 μmol/L)下游引物,0.4 μL M13引物和7.2 μL的ddH₂O。

PCR扩增程序为:94℃ 5 min, 94℃ 30 s, 适宜的退火温度退火40 s, 72℃ 30 s, 34个循环, 72℃ 5 min, 4℃保存。PCR产物委托北京睿博兴科生物技术有限公司在ABI-3730XL基因分析仪(Applied Biosystems, Foster City, CA, USA)上进行毛细管电泳检测。考虑到M13序列(18 bp)的存在,通过从扩增片段的长度中减去18 bp来获得每个预期片段的长度。

1.6 数据统计与分析

应用GeneMarker 2.2.0(SoftGenetics, USA)软件^[18]对SSR片段长度进行整理分析;利用CONVERT version 1.31进行数据格式转换,POPGEN version 1.32和CERVUS version 3.0软件计算多态性位点的等位基因数(N_a)、有效等位基因数(N_e)、期望杂合度(H_e)、Shannon信息指数(I)、多态性信息含量(PIC)等信息^[19-20];OriginPro 2018软件用于Pearson相关分析^[21]。

2 结果与分析

2.1 金银花 RAD-seq 的产出及质量检测

简化基因组的测序结果(表 1)显示,‘九丰一号’和‘亚特’2 个样品获得的原始序列数分别为 31.929 Mbp 和 45.994 Mbp。对原始数据进行过滤、质控后得干

净读长分别为 27.805 Mbp 和 42.560 Mbp, GC 含量分别为 36.18% 和 36.75%, AT 含量高于 GC 含量,对结果影响不大;平均 Q20(质量得分不低于 20 的碱基数及比例)为 98.0%, 平均 Q30(质量得分不低于 30 的碱基数及比例)为 94.6%, 以上结果表明样本测序数据可靠, 建库成功。

表 1 RAD-seq 基因组测序数据统计

Tab. 1 Data statistics of genomic sequencing generated by RAD-seq

样本 Sample	过滤前 Before filter				过滤后 After filter				平均长度 Average length/bp
	原始读长 Raw read/Mbp	原始碱基总数 Total number of original bases/Gbp	Q20/Gbp	Q30/Gbp	干净读长 Clean read/Mbp	有效碱基总数 Total number of bpeffective base/Gbp	Q20/Gbp	Q30/Gbp	
‘九丰一号’ ‘Jiufengyihao’	31.929	4.789	4.478 (93.5%)	4.147 (86.6%)	27.805	4.034	3.935 (97.5%)	3.758 (93.2%)	145.1
‘亚特’ ‘Yate’	45.994	6.899	6.652 (96.4%)	6.367 (92.3%)	42.560	6.304	6.211 (98.5%)	6.047 (95.9%)	148.1

注: Q20. 质量得分不低于 20 的碱基数及比例; Q30. 质量得分不低于 30 的碱基数及比例。Notes: Q20, the number and proportion of bases with a quality score of not less than 20; Q30, the number and proportion of bases with a quality score of not less than 30.

2.2 金银花基因组 SSR 位点信息及其组成分析

运用 MISA 软件检测到‘九丰一号’和‘亚特’的总序列数分别为 221 771 个和 244 153 个, 总序列长度分别为 55 512 285 bp 和 62 681 845 bp; 对 DNA 序列中的 SSR 位点进行搜索发现, 在‘九丰一号’金银花和‘亚特’金银花中有 45 850 个含 SSR 的序列, 含 1 个以上 SSR 位点的序列数为 6 779(14.79%)个, SSR 位点总数为 46 999 个, 其中复合形式(compound)的 SSR 位点仅 5 810 个, 占位点总数的 12.36%; 完全重复形式(perfect)的 SSR 位点共 41 189(87.64%)个, 具体信息如表 2 所示。重复基序碱基变化为 1~6, 单核苷酸重复单元最多, 为 23 101 个, 占总重复单元的 49.15%; 六核苷酸重复单元最少为

94(0.20%)个。另外, 在 P 型(完全重复型)SSR 位点中, 六核苷酸重复型序列的平均长度最长(34.40 bp); 五核苷酸重复型的平均长度次之(27.99 bp); 而 C 型(复合型)SSR 位点的平均长度高达 70.01 bp。

对金银花 P 型 SSR 位点的基序重复类型进行统计, 共发现 319 种基序, 其中单核苷酸重复型 SSRs 以(A/T) n 为主, 共 21 767 个, 占总位点数的 46.31%; 二碱基重复型基序有 6 种, (AT/TA) n 占绝对优势, 共 8 802 个, 占此重复基序的 65.57%, 占全部 SSR 位点数的 18.73%, (CG/GC) n 最少, 仅占总位点数的 0.02%; 三核苷酸重复型 SSRs 以(AAT/ATT) n 最多, 共 534(1.14%); 四核苷酸重复型中, 出现频率较高的为(AAAT/ATTT) n 和(TAAA/TTTA) n , 占此

表 2 金银花基因组中 SSR 位点信息

Tab. 2 Information of SSR loci in *Lonicera japonica*

SSR 类型 SSR type	数量 Number	重复次数 Repeat time	平均长度 Average length/bp	百分比 Percentage/%	重复单元类型 Motif type	
					数量 Number	主要基序 Main motif
单核苷酸 Mono-nucleotide	23 101	10~82	11.83	49.15	2	A/T, C/G
二核苷酸 Di-nucleotide	13 424	6~45	16.89	28.56	6	AT/TA, AG/CT, AC/GT
三核苷酸 Tri-nucleotide	3 883	5~49	22.04	8.26	59	AAT/ATT, GAA/TTC, TAA/TAA, AAG/CTT, TCA/TGA
四核苷酸 Tetra-nucleotide	540	5~14	23.03	1.15	86	AAAT/ATTT, TAAT/ATTA, TAAA/TTTA, CTTT/AAAG, TTTC/GAAA, ATAC/GTAT, TATC/GATA
五核苷酸 Penta-nucleotide	147	5~18	27.99	0.31	92	TTTTT/GAAAA, TTTCT/AGAAA, AAAAT/ATTTT, GGGTT/AACCC, ATCCA/TGGAT
六核苷酸 Hexa-nucleotide	94	5~13	34.40	0.20	74	GGCAAT/ATTGCC, GGAGAT/ATCTCC, GGTGGA/TCCACC, GATATA/TATATG, CAATCA/TGATTG, GCATTT/AAATGC
复合型 Compound	5 810		70.01	12.36		
总计 Total	46 999		21.54	100.00	319	

重复基序的 10.00%，占全部 SSR 位点的 0.11%；(ATCCA/TGGAT) n 和 (GGCAAT/ATTGCC) n 分别为五核苷酸和六核苷酸重复型的主要基序。

2.3 金银花基因组 SSR 位点基序重复次数

金银花基因组不同类型的 SSR 位点的重复次数如图 1 所示。单核苷酸类型的重复次数主要集中在 10 ~ 16 次；二核苷酸基序重复次数集中在 6 ~ 11 次；三、四、五、六核苷酸基序的重复次数均集中

在 5 ~ 6 次。在 SSR 位点基序的重复次数中出现 10 次重复的 SSR 位点数最高，为 12 238 个 SSR 位点，占总 SSR 位点数 26.04%；11 次和 6 次重复次数分别为 5 276(11.23%)和 4 981(10.60%)个 SSR 位点；18 次重复次数的位点数最少，仅 342(0.73%)个。另外，从图中数据可以看出，除三核苷酸重复型和六核苷酸重复型外，各 SSR 位点重复类型出现的频率存在随重复次数的增加而逐渐下降的趋势。

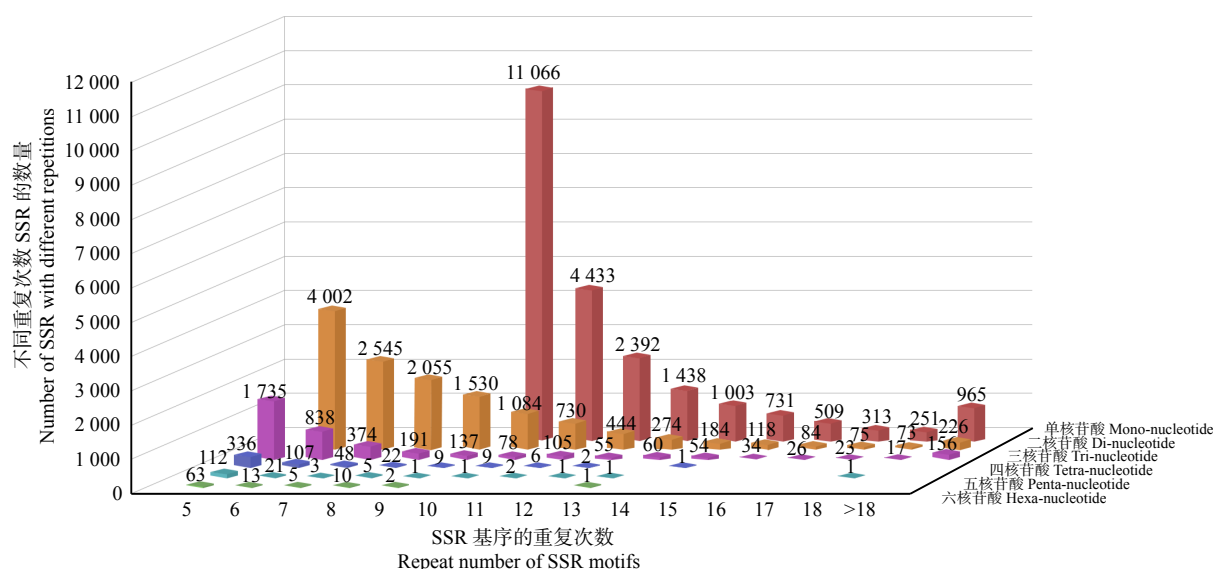


图 1 简化基因组中 SSR 位点的基序重复次数分布

Fig. 1 Distribution of repeat numbers of SSR motifs in simplified genome

2.4 金银花基因组 SSR 序列长度及变异分析

对金银花基因组 SSR 序列长度进行分析(图 2)，发现其长度在 10 ~ 310 bp 之间，平均长度为 21.54 bp。长度为 10 bp 的 SSRs 数量最多，占总位点数的 23.62%；其次为长度 12 bp 的 SSRs，占比 13.62%。对于 P 型 SSR 序列，不同类型重复单元的 SSR 序列长度也具有丰富的变异性。各 SSR 序列类型出现的频率均存在随重复次数增加而逐渐下降的趋势(图 3)。单核苷酸重复型 SSR 序列的长度变化最

大，为 10 ~ 82 bp，平均长度为 12.57 bp，其中(C/G) n 序列的平均长度高达 14.89 bp，(A/T) n 平均长度为 11.65 bp；二核苷酸重复型的平均长度为 16.89 bp，其中(AG/CT) n 序列的平均长度最大，为 17.67 bp，(CG/GC) n 平均长度最小，仅 13.00 bp。Pearson 相关结果表明重复单元长度与 SSR 平均长度呈正相关($P < 0.0005$)。

2.5 金银花基因组 SSR 引物设计及多态性分析

利用 46 999 个 SSR 位点成功设计出 38 507 对 SSR 引物，设计成功率为 81.93%。从所设计的引物中随机合成 288 对，用 12 个金银花样本对 SSR 引物的多态性进行验证，有 175 对可扩增出目的条带，其中 27 对引物的产物出现分裂峰、连续峰或影子峰，16 对引物的扩增产物未在目的片段范围内，因此扩增产物较稳定的引物有 132 对，有效扩增比率为 75.43%。经筛选，132 对引物中有 72 对能在 12 个金银花样品中扩增出多态性产物，从中选取 35 对峰型良好，多态性高，稳定好的引物作进一步验证(图 4、表 3)。35 对 SSR 引物的等位基因数在 3 ~ 10 之间，平均为 5.057；有效等位基因数在 1.419 ~ 6.698 之间，平均为 2.907；期望杂合度介于 0.308 ~ 0.888 之间；Shannon 信息指数介于 0.624 ~ 2.084 之间，平均

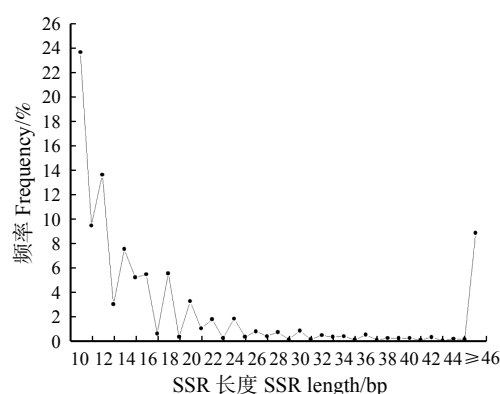


图 2 金银花基因组不同 SSR 序列长度的出现频率

Fig. 2 Frequency of SSR with different lengths in *L. japonica*

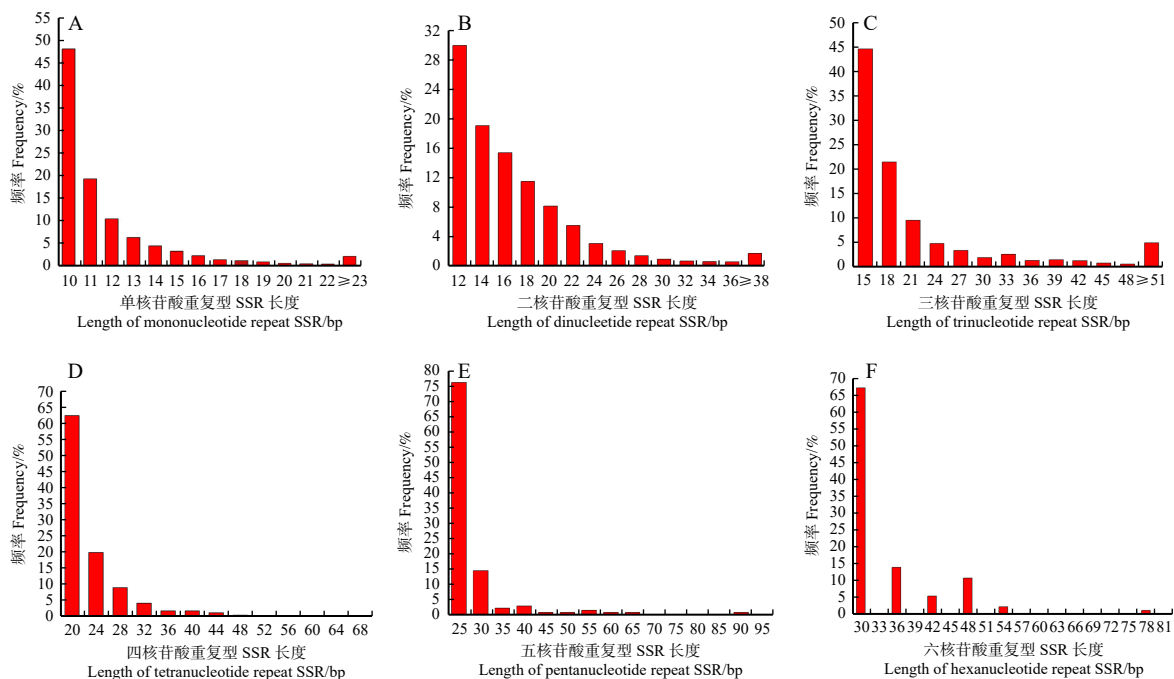


图 3 不同长度重复单元 SSR 序列的长度变异情况

Fig. 3 Variation in sequence length of different types of SSR

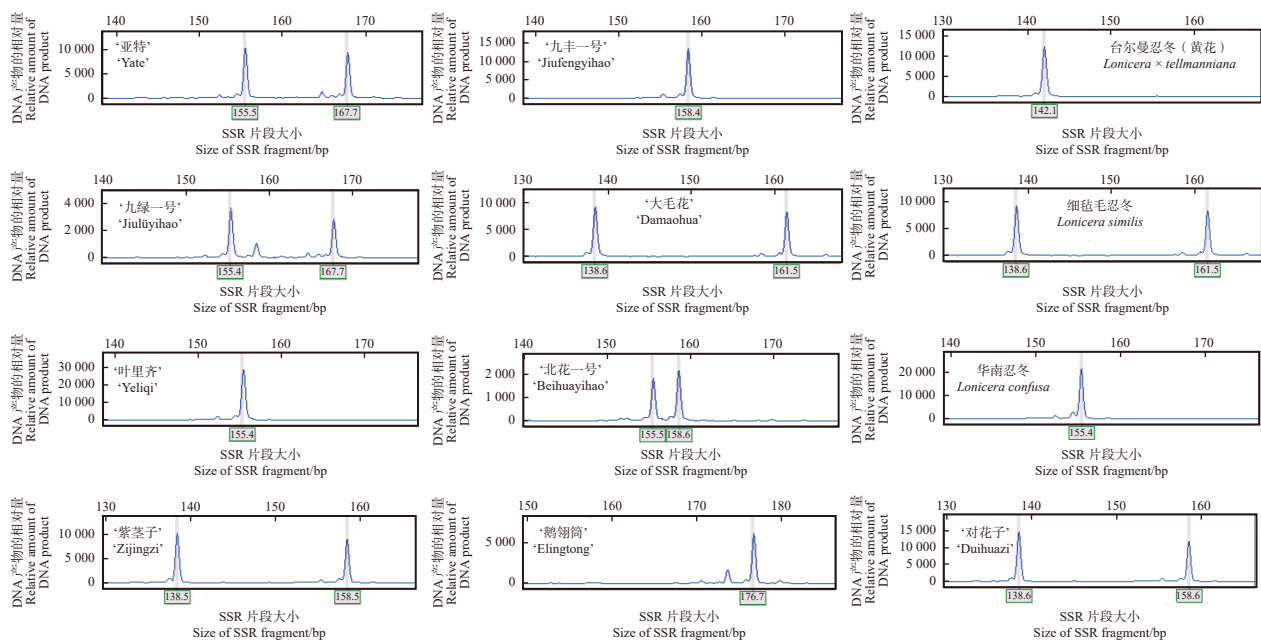


图 4 引物 jap5 在 12 个金银花样品中的毛细管电泳检测图

Fig. 4 Capillary electrophoresis detection of 12 samples of *L. japonica* with jap5 primer

为 1.217。本研究中 35 个多态性引物的 PIC 值在 0.281 ~ 0.835 之间, 平均值为 0.568。其中, 高多态信息位点(PIC > 0.5)26 个, 中度多态信息位点(0.5 > PIC > 0.25)有 9 个, 多态性较高。另外, 35 对多态性引物均未偏离哈迪-温伯格平衡。

3 讨 论

目前金银花尚未进行全基因组或转录组测序, 虽然可利用近源物种的可参考全基因组序列去开发

分子标记, 但这些标记存在成本高、数量少、耗时费力、准确性较差等问题。因此, 利用金银花分子标记建立一套统一的种质资源鉴定方法是目前急需解决的问题。迄今为止, RAD-seq 已广泛应用于 23 种以上有或没有参考基因组的物种^[22]。但是, 至今尚无运用 RAD-seq 技术开发金银花 SSR 标记的报道。本研究通过对金银花进行简化基因组测序, 对获得的核苷酸序列进行过滤、聚类 and 组装后, 在测序获得的基因组序列数据中, ‘九丰一号’和‘亚特’金银花的

表 3 35 个位点 SSR 引物多态性分析

Tab. 3 Polymorphism analysis of SSR primers at 35 loci

引物 Primer	引物序列(5'—3') Primer sequence (5'—3')	SSR类型 Type of SSR	重复基序 Repeat motif	N_a	N_e	H_e	H_o	H	I	PIC
jap2	F:AGATGGATAGTGGGCCATTGT R:TAGAGTCACCGGGAGCTACC	P3	(TGA)5	6.000	2.342	0.598	0.402	0.573	1.211	0.544
jap5	F:GCACGTGCACTCAGAAACAG R:GCAACCGTTGAGCTCCTTTG	P3	(GTA)7	8.000	6.000	0.870	0.130	0.833	1.932	0.814
jap6	F:GCATAGACTCTGCTGGCCAA R:AGCAGTGGCAGATATGGCTG	P3	(TCA)6	4.000	3.236	0.721	0.279	0.691	1.258	0.635
jap15	F:TGAAAAGAAGGGAATAGGATTAGGGT R:CCTCTTGGTGAGACCGCTTT	P4	(GAAT)5	5.000	2.220	0.576	0.424	0.550	1.097	0.512
jap17	F:GAAGAAGAGCGGCTACTGCA R:AAGAGTGAAGCAGCAGCAGT	P3	(CAC)5	5.000	3.723	0.766	0.234	0.731	1.417	0.685
jap18	F:AGGCTACATGATGATTGAGACGA R:TTGGCCCCACCTGATCAAAA	P3	(ACA)6	4.000	2.504	0.627	0.373	0.601	1.054	0.524
jap21	F:ACCCTCACCTTCACCCTCAT R:TGGGTATCGCATCGTCAAAA	P3	(CAT)5	4.000	1.419	0.308	0.692	0.295	0.624	0.281
jap44	F:ATGGAGGGCTAGTGAATGCG R:TCCACACTCGGTACACCAAA	P3	(GAT)5	4.000	1.959	0.511	0.489	0.490	0.882	0.431
jap51	F:ACGTGCAATGTTAATGGCACG R:TGCAGCACATTTGGCATTCT	P3	(ATT)6	4.000	2.575	0.641	0.359	0.612	1.121	0.555
jap53	F:GTCTGGGGAATGAGTTGCCA R:AGCCTTTGATCGAGTCAGGT	P3	(TTG)14	7.000	3.556	0.750	0.250	0.719	1.549	0.686
jap73	F:TCAAACCTCTACCTGGCTGC R:CTGTATGTGCCGACATGGT	P3	(GCT)6	6.000	2.642	0.649	0.351	0.622	1.328	0.594
yjap4	F:CAACTTCACCGGTGATCCA R:TCCTTCCAATTGAAACTAAACCT	P3	(TTG)5	4.000	2.483	0.623	0.377	0.597	1.119	0.552
yjap5	F:AAAGGCGTGGTGAAAGGTCT R:CACATATGGTTGCTGCAGCC	P3	(GAA)5	10.000	6.698	0.888	0.112	0.851	2.084	0.835
yjap6	F:TTCTAGGGCTTCGGCAACTG R:CCTTCCCTCGAAACTGCA	P3	(CTG)5	4.000	1.714	0.435	0.565	0.417	0.837	0.393
yjap9	F:TGGCAACTTTCCCCAAATTC R:TGAGCGAGTCATATAGGCGAG	P3	(TTC)6	4.000	2.361	0.601	0.399	0.576	0.980	0.485
yjap10	F:AACGGGAAAAGAGTGAGCGAG R:GTGTGTTTCCCTTTTCCGGC	P3	(ATT)5	3.000	1.834	0.475	0.525	0.455	0.803	0.410
yjap12	F:GCATAGACTCTGCTGGCCAA R:AGCAGTGGCAGATATGGCTG	P3	(TCA)5	5.000	2.852	0.678	0.323	0.649	1.305	0.614
yjap19	F:GGTGGATGAGCTTCTGAAGAGA R:CCAATGGCTTCTTCCCTGT	P3	(TTG)6	5.000	2.028	0.529	0.471	0.507	0.994	0.463
yjap24	F:GCTTTTGCAGCTCTACTTTGT R:TTTGCTGTAAAGACACGCGC	P3	(CAC)5	4.000	3.429	0.739	0.261	0.708	1.309	0.659
yjap36	F:GACACACACGGTTGCTTGAC R:AGAGTTTCTGCTGTGGTGGA	P3	(GCA)6	4.000	2.969	0.692	0.308	0.663	1.200	0.600
yjap39	F:TCACTACCTGTATCCGCTCT R:TGGAAGCAGAAAGGGCATCT	P3	(ATC)7	4.000	2.743	0.663	0.337	0.635	1.140	0.566
yjap47	F:GCTTCGCCTTTGGTTCCGTA R:CAACTCGCTCTCTCCCTTCC	P3	(ATG)7	4.000	1.419	0.308	0.692	0.295	0.624	0.281
yjap48	F:GGGCATTTTCTGGTGAGGA R:AGCATCCAAACCTGATTGT	P3	(GTT)5	4.000	2.642	0.649	0.351	0.622	1.105	0.553
yjap51	F:CGGAGAGAGATACTAGGCGC R:CGCGGCTGAATTTTAGAGCC	P2	(TA)6	7.000	3.692	0.761	0.239	0.729	1.595	0.701
yjap53	F:CATTCAATCCCCACCAGGT R:GGGACCGGTCGAAACCAATA	P2	(TC)10	5.000	2.034	0.533	0.468	0.508	1.056	0.483
yjap59	F:GGATCCAGCATGTCTGAACT R:TGCTTTGGCCCTTCTCATT	P2	(AT)8	4.000	2.215	0.573	0.428	0.549	1.013	0.500
yjap89	F:GAGTAAAGTGAGAGCGTGT R:AGTTTCCGAGAAGTGCCTGT	P4	(CTAT)5	4.000	3.310	0.728	0.272	0.698	1.283	0.645
yjap105	F:TGTTTAGCAACAAGGATGGT R:GCTGGCAGGGACAAAAAGTG	P2	(GA)7	6.000	4.033	0.788	0.212	0.752	1.547	0.714
yjap109	F:AACTTCTGTGACAGGGCAGG R:GCATATGCGGCAATGGGAAG	P2	(GT)6	5.000	2.504	0.627	0.373	0.601	1.188	0.559
yjap120	F:GCCAGAGCTCCGATAAGGTC R:TGCATCGCGTGTGTATGTGA	P2	(AT)6	8.000	4.966	0.833	0.167	0.799	1.848	0.779
yjap153	F:ACATGAGGGTTTGCCGATCC R:GGAGTATGGCCAGGTTTCGTC	P3	(GTG)6	5.000	2.852	0.678	0.323	0.649	1.253	0.596
yjap169	F:AGTTCGAGGGATTGTGCGTT R:TGTGGAATTCGTCAATGGCG	P3	(TTC)7	3.000	1.674	0.420	0.580	0.403	0.722	0.363
yjap171	F:TCAGAGAGGGGTAACAGCA R:TCGGCTTATACACTGCGAG	P3	(ACC)5	4.000	2.483	0.623	0.377	0.597	1.119	0.552

表 3(续)
Tab.3 (continued)

引物 Primer	引物序列(5'—3') Primer sequence (5'—3')	SSR类型 Type of SSR	重复基序 Repeat motif	N_a	N_e	H_e	H_o	H	I	PIC
yjap175	F:GTTTCAATGCGAGCTCGACG R:TCTCATTTCTCCCAAAATTCATCAGA	P3	(TTA)6	9.000	3.388	0.736	0.265	0.705	1.647	0.681
yjap183	F:GGGGCCTCATTTTCCTCAA R:TTTTCGGTGGTGGAGCAAT	P3	(CAC)5	5.000	3.236	0.721	0.279	0.691	1.342	0.644
均值 Mean				5.057	2.907	0.638	0.363	0.611	1.217	0.568

注： H_o ，观测杂合度； N_a ，观测等位基因数； N_e ，有效等位基因数； H_e ，期望杂合度； I ，Shannon信息指数； H ，Nei's遗传多样性指数；PIC，多态性信息含量。Notes: H_o , observation of heterozygosity; N_a , observed number of alleles; N_e , effective number of alleles; H_e , expected heterozygosity; I , Shannon information index; H , Nei's genetic diversity index; PIC, polymorphism information content.

干净读长分别占原始数据的 87.08% 和 92.53%，Q30 比例分别高达 93.2% 和 95.9%，表明碱基出错率低；GC 含量分别为 36.18% 和 36.75%，AT 含量略高于 GC 含量，上述结果表明样本数据量达到预期目标，测序数据可用于后续分析。通过 RAD-seq 技术开发应用 SSR 标记的研究充分体现了 RAD-seq 技术通量高、稳定性好、不受基因组序列限制等优势。因此，对于尚未获得足够基因组序列数据的物种，RAD-seq 技术为其 SSR 分子标记的开发提供了一种简单有效的途径。

在本研究中，利用 MISA 软件从金银花简化基因组中查找到的 46 999 个 SSR 位点可分为 P 型和 C 型。在 P 型 SSR 位点中，单核苷酸重复型出现的频率最高，占总位点数的 49.15%，六核苷酸重复型 SSR 位点比例最低，仅占 0.20%。单核苷酸重复单元占主导，可能与目前金银花高通量测序读长相对较短、大量多碱基重复序列未得到充分挖掘有很大关系^[23]。对于不同类型的金银花 SSR 重复单元，SSR 分布频率会逐渐随重复单元数量的增加而降低，该结论与宋立肖等^[24]的研究结果相似。研究发现，基因组大小与 SSR 数量或 SSR 密度之间呈负相关，金银花和高粱(*Sorghum bicolor*)等物种基因组 SSR 的分布特征也证实了这一观点^[25-26]。此外，该领域的一些学者发现在基因组中，中低级重复单元比例较多的物种的进化水平较高，而高级重复单元较多表明该物种进化时间段或变异频率较低^[27]，由此可以推测金银花的物种进化水平较高。但在大多数植物的 SSR 分子标记中，二核苷酸重复单元或三核苷酸重复单元占主导，例如在丁香(*Syringa oblata*)基因组的 SSR 位点中，二核苷酸重复型占优势^[28]，在油橄榄(*Olea europaea*)中三核苷酸重复型占优势^[29]，不同物种间的差异可能与基因组大小有关^[30]。

在金银花 SSR 基序中共发现 319 种重复类型，以 A/T、AG/CT、AT/TA、AAT/ATT、GAA/TTC、TAA/TAA、AAAT/ATTT、TAAT/ATTA、TAAA/TTTA、

TTTTC/GAAAA、TTTCT/AGAAA 和 AAAAT/ATTTT 基序最为丰富，说明 A/T 碱基在金银花基因组 SSR 位点中占绝对优势，此结果与在毛果杨(*Populus trichocarpa*)、石榴(*Punica granatum*)中的 SSR 序列研究结果相似^[21,31]。本研究所获得的干净读长及所检测到的 SSR 位点的基序中，均存在 A、T 碱基含量高于 G、C 碱基的现象，且在 SSR 位点的基序中，A、T 含量越高，该基序出现的频率越高。该现象的出现可能与 DNA 中 AT 碱基对比 GC 碱基对更易解链有关，也可能与 SSR 位点查找工具及参数设置有关^[32]。另外，C 型 SSR 位点占总位点数的 12.36%，远高于灵宝杜鹃(*Rhododendron henanense* subsp. *lingbaoense*) SSR 序列中 C 型 SSR 位点的比例(8.14%)^[33]，但目前还未见关于金银花基因组序列中 C 型 SSR 位点的研究报道。

研究发现二核苷酸基因座的广泛使用会导致所有等位基因的大小非常接近，难以区分，从而可能引起误读并产生混淆^[29,34]，三核苷酸 SSR 标记已被证明具有高度多态性，并且能够在大豆(*Glycine max*)和小麦(*Triticum aestivum*)等植物中稳定遗传^[35-36]。本研究所筛选的 35 对 SSR 标记中，三碱基及以上重复的标记占 29 对(82.86%)，这些具有更长重复基序的 SSR 标记的开发将为金银花品种的鉴定和遗传分析提供更多、更有效的信息。此外，本研究所获得的 35 对多态性引物中，有 3 对引物的等位基因数与有效等位基因数接近，其余的 32 对差异较大，说明所筛选的绝大多数 SSR 引物的等位基因分布不均，这可能与样本容量不足以致主效等位基因缺失有关。PIC 的大小能够反映 SSR 位点的变异程度，该值越接近 1，该群体杂合体的比率越大，多态性越高^[37]。本研究所筛选的引物中高多态性引物(PIC > 0.5)有 26 对，其中有 15 对同时具有较高的 PIC(> 0.5)和较多的等位基因(≥ 5)，表明这些引物具有较高的遗传变异性，在进行图谱构建及分子鉴定中可优先采用。

4 结 论

本研究首次通过 RAD-seq 测序数据进行金银花 SSR 分子标记的大规模开发,并对金银花不同类型的 SSR 位点的序列特征及分布规律进行分析归纳。利用随机选择的 12 个金银花样本,通过 SSR 标记结合毛细管电泳的自动检测法,筛选出 35 对稳定的多态性引物。其中,26 对高多态性引物中有 15 对同时具有较高的 PIC(> 0.5)和较多的等位基因(≥ 5),在后续的图谱构建、遗传多样性分析、QTL 定位等遗传学研究中可优先采用。

参 考 文 献

- [1] Cai A C, Liao H Y, Wang C C, et al. A comprehensive study of the aerial parts of *Lonicera japonica* Thunb. based on metabolite profiling coupled with PLS-DA[J]. *Phytochemical Analysis*, 2020, 31(6): 786–800.
- [2] 朱凤洁, 杨健, 袁媛, 等. 金银花种质资源化学指纹图谱及代谢物相似性分析[J]. 中国中药杂志, 2018, 43(12): 2575–2579.
Zhu F J, Yang J, Yuan Y, et al. Chemical fingerprinting and similarity analysis of *Lonicera japonica* resources[J]. *China Journal of Chinese Materia Medica*, 2018, 43(12): 2575–2579.
- [3] 蔡芷辰, 刘训红, 王程成, 等. 金银花分子生物学研究进展[J]. 中国中药杂志, 2020, 45(6): 1272–1278.
Cai Z C, Liu X H, Wang C C, et al. Research progress in molecular biology of *Lonicerae Japonicae* Flos[J]. *China Journal of Chinese Materia Medica*, 2020, 45(6): 1272–1278.
- [4] 张贵生, 王林冷, 方元平, 等. 基于 SSR 标记的大别山区金银花遗传多样性分析[J]. 分子植物育种, 2019, 17(19): 6543–6548.
Zhang G S, Wang L L, Fang Y P, et al. Genetic diversity analysis of *Lonicera japonica* in dabie mountains based on SSR markers[J]. *Molecular Plant Breeding*, 2019, 17(19): 6543–6548.
- [5] 徐石勇, 李欧静, 张舒玮, 等. 金银花 SSR 指纹图谱的构建及遗传多样性分析[J]. 天津农业科学, 2015, 21(5): 11–14, 18.
Xu S Y, Li O J, Zhang S W, et al. Construction of SSR-based molecular finger printing and analysis of genetic diversity for *Lonicera japonica* [J]. *Tianjin Agricultural Sciences*, 2015, 21(5): 11–14, 18.
- [6] 孙稚颖, 姚辉, 王振中, 等. 金银花种质资源遗传多样性的 ISSR 分析[J]. 世界科学技术-中医药现代化, 2013, 15(9): 1890–1895.
Sun Z Y, Yao H, Wang Z Z, et al. ISSR analysis on genetic diversity of germplasm resource of *Lonicerae Japonicae* Flos[J]. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*, 2013, 15(9): 1890–1895.
- [7] Heckenberger M, Vort J, Peleman J, et al. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties(II): genetic and technical sources of variation in AFLP data and comparison with SSR data[J]. *Molecular Breeding*, 2003, 12: 97–106.
- [8] 王久利, 朱明星, 徐明行, 等. 基于 RAD-seq 技术的异型花 SSR 信息分析[J]. 植物研究, 2017, 37(3): 447–452.
Wang J L, Zhu M X, Xu M X, et al. Analysis on SSR in *Sinoswertia tetraptera* base on RAD-seq[J]. *Bulletin of Botanical Research*, 2017, 37(3): 447–452.
- [9] 黄兴发, 尹跃, 赵建华, 等. 黑果枸杞基因组 SSR 标记开发及遗传多样性分析[J]. 西北农林科技大学学报(自然科学版), 2021, 49(1): 126–135.
Huang X F, Yin Y, Zhao J H, et al. Development of genomic SSR markers and genetic diversity analysis of *Lycium ruthenicum* Murr. [J]. *Journal of Northwest A&F university (Natural Science Edition)*, 2021, 49(1): 126–135.
- [10] Kumar A A, Deepak S, Archana V, et al. MSDB: a comprehensive, annotated database of microsatellites[J]. *Nucleic Acids Research*, 2020, 48(D1): D155–D159.
- [11] Andrews K R, Good J M, Miller M R, et al. Harnessing the power of RAD-seq for ecological and evolutionary genomics[J]. *Nature Review Genetics*, 2016, 17(2): 81–92.
- [12] Hou Y, Nowak M D, Mirré V, et al. RAD-seq data point to a northern origin of the arctic-alpine genus *Cassiope* (Ericaceae)[J]. *Molecular Phylogenetics and Evolution*, 2016, 95: 152–160.
- [13] Yuan Z, Fang Y, Zhang T, et al. The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology[J]. *Plant Biotechnology Journal*, 2018, 6(7): 1363–1374.
- [14] Miller M R, Dunham J P, Amores A, et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers[J]. *Genome Research*, 2007, 17(2): 240–248.
- [15] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads[J]. *Embnet Journal*, 2011, 17(1): 10–12.
- [16] Andreas U, Ioana C, Triinu K, et al. Primer3–new capabilities and interfaces[J]. *Nucleic Acids Research*, 2012, 40(15): e115.
- [17] Schuelke M. An economic method for the fluorescent labeling of PCR fragments[J]. *Nature Biotechnology*, 2000, 18: 233–234.
- [18] Hulce D, Li X, Snyder-Leiby T, et al. GeneMarker genotyping software: tools to increase the statistical power of DNA fragment analysis[J]. *Journal of Biomolecular Techniques*, 2011, 22(Suppl.): S35–S36.
- [19] Yeh F C, Yang R C, Boyle T, et al. POPGENE Version 1.32: Microsoft Windows-based freeware for populations genetic analysis[Z]. Edmonton: University of Alberta, 1999.
- [20] Marshall T C, Slate J, Kruuk L, et al. 1998. Statistical confidence for likelihood-based paternity inference in natural populations[J]. *Molecular Ecology*, 1998, 7(5): 639–655.
- [21] 洪文娟, 郝兆祥, 刘康佳, 等. 基于石榴全基因组序列的 SSR 标记开发及鉴定[J]. 北京林业大学学报, 2019, 41(8): 38–47.
Hong W J, Hao Z X, Liu K J, et al. Development and identification of SSR molecular markers based on whole genomic sequences of *Punica granatum* [J]. *Journal of Beijing Forestry*

- University, 2019, 41(8): 38–47.
- [22] Feng J Y, Zhao S, Li M, et al. Genome-wide genetic diversity detection and population structure analysis in sweet potato (*Ipomoea batatas*) using RAD-seq[J]. *Genomics*, 2020, 112(2): 1978–1987.
- [23] 刘瑞娟, 路兴旺, 窦全文. 基于简化基因组测序开发垂穗披碱草 (*Elymus nutans*) SSR 标记[J]. 分子植物育种, 2018, 16(6): 1888–1894.
- Liu R J, Lu X W, Dou Q W. Development of SSR markers in *Elymus nutans* based on reduced-representation genome sequencing[J]. *Molecular Plant Breeding*, 2018, 16(6): 1888–1894.
- [24] 宋立肖, 李国旗, 靳长青, 等. 大麻状罗布麻的全基因组分析和 SSR 标记开发[J]. 植物遗传资源学报, 2019, 20(5): 1309–1316.
- Song L X, Li G Q, Jin C Q, et al. Whole genome sequencing and development of SSR markers in *Apocynum cannabinum*[J]. *Journal of Plant Genetic Resources*, 2019, 20(5): 1309–1316.
- [25] Morgante M, Hanafey M, Powell W, et al. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes[J]. *Nature Genetics*, 2002, 30(2): 194–200.
- [26] Huo N, Lazo G, Vogel J P, et al. The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences[J]. *Functional Integrative Genomics*, 2008, 8(2): 135–147.
- [27] Tóth G, Gáspári Z, Jurka J, et al. Microsatellites in different eukaryotic genomes: survey and analysis[J]. *Genome Research*, 2000, 10(7): 967.
- [28] Yang Y Y, He R Q, Zheng J, et al. Development of EST-SSR markers and association mapping with floral traits in *Syringa oblata*[J]. *BMC Plant Biology*, 2020, 20(1): 436.
- [29] Li D Y, Long C, Pang X M, et al. The newly developed genomic-SSR markers uncover the genetic characteristics and relationships of olive accessions[J]. *PeerJ*, 2020, 8: 341–345.
- [30] Haydar K, Ying L, Wieland M. Survey of simple sequence repeats in completed fungal genomes[J]. *Molecular Biology & Evolution*, 2005, 22: 639–649.
- [31] Tuskan G A, Gunter L E, Yang Z K, et al. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*[J]. *Canadian Journal of Forest Research*, 2004, 34: 85–93.
- [32] 阎毛毛, 戴晓港, 李淑娴, 等. 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, 30(1): 103–109.
- Yan M M, Dai X G, Li S X, et al. Sequence analysis and comparison of EST-SSRs in pine, poplar and eucalyptus[J]. *Genomics and Applied Biology*, 2011, 30(1): 103–109.
- [33] 周晓君, 王海亮, 李方玲, 等. 基于 RAD-seq 技术开发灵宝杜鹃多态性 SSR 标记[J]. 农业生物技术学报, 2019, 27(1): 55–62.
- Zhou X J, Wang H L, Li F L, et al. Development of polymorphic SSR markers in *Rhododendron henanense* subsp. *lingbaoense* based on RAD-seq[J]. *Journal of Agricultural Biotechnology*, 2019, 27(1): 55–62.
- [34] Beghè D, Molano J F G, Fabbri A, et al. Olive biodiversity in Colombia, a molecular study of local germplasm[J]. *Scientia Horticulturae*, 2015, 189: 122–131.
- [35] Cregan P B, Jarvik T, Bush A L, et al. An integrated genetic linkage map of the soybean genome[J]. *Crop Science*, 1999, 39: 1464–1490.
- [36] Song Q J, Fickus E W, Cregan P B, et al. Characterization of trinucleotide SSR motifs in wheat[J]. *Theoretical and Applied Genetics*, 2002, 104: 286–293.
- [37] 金玲, 刘明国, 董胜君, 等. 97 个山杏无性系的遗传多样性及 SSR 指纹图谱[J]. *林业科学*, 2018, 54(7): 51–61.
- Jin L, Liu M G, Dong S J, et al. Genetic diversity and fingerprints of 97 *Armeniaca sibirica* clones based on SSR markers[J]. *Scientia Silvae Sinicae*, 2018, 54(7): 51–61.

(责任编辑 范娟 赵田芸
责任编委 林金星)