

基于染色体水平栀子全基因组 SSR 特征分析

王轶钥^{1,2}, 蒋丽娟¹, 李培旺², 陈景震², 陈韵竹², 刘强¹, 周宵^{1,2*}, 杨艳^{2,3*}

1. 中南林业科技大学生命与环境科学学院, 湖南长沙 410002

2. 湖南省林业科学院 木本油料资源利用全国重点实验室, 湖南长沙 410004

3. 南方木本油料利用科学国家林草局重点实验室, 湖南长沙 410004

摘要:目的 通过全基因组数据开发栀子 *Gardenia jasminoides* 全基因组 SSR 分子标记。方法 基于栀子全基因组序列, 利用 MISA 软件搜索 SSR 位点, 从染色体长度、SSR 数量、不同重复类型、不同重复频次以及不同核苷酸类型的特征进行了分析。结果 栀子基因组共鉴定到 117 037 个 SSR, 长度在 12~817 bp, 平均长度为 24.68 bp。栀子基因组全长为 534.9 Mb, 最长的为 Chr09, 长达 110.58 Mb; 最短的为 Chr10, 仅有 30.74 Mb; SSR 出现频率为 444.23 bp/个。完全型核苷酸共计 102 001 个, 以二核苷酸数量最多, 六核苷酸最少, 其中长度大于 15 bp 的 SSR 共有 64 283 个, 占比为 67.61%; 混合型核苷酸 14 464 个, 占比 12.36%。各染色体组上 SSR 位点的频数密度在 157~280 个/Mb, 且 SSR 位点数量和染色体组长度呈高度相关。SSR 在栀子的 11 组染色体中以 Chr09 最多为 31 012 个, Chr10 最少为 5 360 个。不同 SSR 类型重复频次范围在 4~81, 重复频次主要集中在 4~20, 占比高达 96.30%。完全型核苷酸不同重复类型种类和不同重复基序类型分别有 2 429、1 112 种, 其优势重复单元依次为 T、AT、AAT、AAAT、AAAAG、AAAAAT, 栀子 SSR 序列在一定程度上偏好于 A、T 所构成的碱基。结论 丰富了栀子的 SSR 分子标记数据库, 可为后续栀子的核心种质构建、遗传多样性分析、品种鉴定及分子标记辅助育种等研究奠定基础。

关键词: 栀子; 全基因组; SSR 分子标记; SSR 特征分析; 栀子苷; 藏红花素

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)07-2495-10

DOI: 10.7501/j.issn.0253-2670.2025.07.023

Genome-wide SSR characterisation of *Gardenia jasminoides* based on chromosome level

WANG Yiyue^{1,2}, JIANG Lijuan¹, LI Peiwang², CHEN Jingzhen², CHEN Yunzhu², LIU Qiang¹, ZHOU Xiao^{1,2}, YANG Yan^{2,3}

1. College of Life and Environmental Sciences, Central South University of Forestry & Technology, Changsha 410002, China

2. State Key Laboratory of Woody Oil Resource Utilization, Hunan Academy of Forestry, Changsha 410004, China

3. Key Laboratory of Woody Oil Utilization Science in South National Forestry and Grassland Administration, Changsha 410004, China

Abstract: Objective To develop genome-wide SSR molecular markers for *Gardenia jasminoides* through genome data.

Methods Based on the whole genome sequence of *G. jasminoides*, searching for SSR loci by using MISA software, and analysing the characteristics in terms of chromosome length, number of SSRs, different repetition types, different repetition frequencies and different nucleotide types. **Results** A total of 117 037 SSRs were identified in *G. jasminoides* genome, the SSRs length was between 12—817 bp, the average length was 24.68 bp. The full length of *G. jasminoides* genome was 534.9 Mb, the longest one was Chr09, which was 110.58 Mb, the shortest one was Chr10, only 30.74 Mb, and the frequency of SSRs appearance was 444.23 bp each. There were a total of 102 001 complete nucleotides, with the highest number of

收稿日期: 2024-10-05

基金项目: 湖南省重大科技创新平台项目 (2024PT1001); 长沙市自然科学基金资助项目 (kq2208098, 2021JJ31143, kq2208099); 膜受体木本油脂功能成分筛选及大健康产品定向转化 (kq2301004); 郴州国家可持续发展议程创新示范区建设专项 (2022sfq53); 湖南省林业科学院国家光皮树、山苍子良种基地项目

作者简介: 王轶钥 (2003—), 女, 学士, 研究方向为植物学。E-mail: 2874686259@qq.com

*通信作者: 杨艳 (1980—), 女, 博士, 研究员, 研究方向为森林培育。E-mail: yangyanzupci@126.com

周宵 (1994—), 男, 博士, 研究方向为植物学。E-mail: zx924647234@163.com

dinucleotides and the lowest number of hexanucleotides. There were 64 283 SSR's length larger than 15 bp, accounting for 67.61% of the total. There were 14 464 mixed nucleotides, accounting for 12.36% of the total SSRs. The frequency density of SSR locis on each chromosome group was in the range of 157—280 Mb, and the number of SSR locis and the length of chromosome group were highly correlated. Among the 11 chromosomes of *G. jasminoides*, the highest number of SSRs in Chr09 was 31 012, and the lowest number of SSRs in Chr10 was 5 360. Different SSR types repeat frequency ranged between 4—81, repeat frequency mainly concentrated in 4—20, accounting for as high as 96.30%. There were 2 429 different repeat types and 1 112 different repeat motifs in complete nucleotides, and their dominant repeat units were T, AT, AAT, AAAT, AAAAG, AAAAAT in order. The SSR sequences of *G. jasminoides* preferred to some extent the bases composed of A and T.

Conclusion This study enriched the SSR molecular markers database of *G. jasminoides*, which can lay the foundation for the subsequent studies of core germplasm construction, genetic diversity analysis, variety identification and molecular marker-assisted breeding of *G. jasminoides*.

Key words: *Gardenia jasminoides* J. Ellis; whole genome; SSR molecular markers; SSR characterisation; geniposide; crocin

梔子 *Gardenia jasminoides* J. Ellis 是茜草科梔子属的灌木,在《本草纲目》中被称为“卮子”,因其果实形似商周酒器“卮”而得名,又名黄梔子、山梔子等。梔子花形大且花色洁白、美丽芳香,广泛种植于盆栽、庭院以供观赏。梔子原产于中国,主要分布于我国长江以南一带,日本、朝鲜、越南等地也均有分布^[1]。因梔子中含有梔子苷、藏红花素、京尼平苷等活性成分,是重要中药材之一,可用于临床治疗,也是我国卫生部于 2002 年颁布的第一批药食两用资源^[1],其性寒味苦,无毒;常以干燥成熟果实入药,具有消炎利湿、清热解毒等功效,还可用于消肿止痛、湿热黄疸、肝炎肾炎水肿、血热吐衄、火毒疮疡等病症^[2-3]。同时梔子黄色素、蓝色素是常用的天然染色剂,也是国际上流行的天然食品添加剂,且需求量逐年递增。随着人们对它认知度的提高,市面上已经出现很多以梔子花、果为原料的护肤品、洗漱用品、香水等,并且梔子油作为一种对人体有益的高端食用油具有广阔的市场^[4]。而目前对于定向型品种如高产花、高产果、高含油率、高梔子苷、高藏红花素等优良品种的开发少,借助 SSR 分子标记构建核心种质库,实现品种选育、多个优良性状聚合等满足市场需求^[5]。

目前,关于梔子的研究多集中在内含物及其药理作用,而在分子育种领域的研究不够深入,部分学者基于目标起始密码子多态性标记 (start codon targeted polymorphism, SCoT)^[6]、简单重复间序列标记 (inter-simple sequence repeats, ISSR)、相关序列扩增多态性标记 (Sequence—related amplified polymorphism, SRAP) 分子标记^[7],探究梔子不同居群间的遗传差异。Deng^[8]和刘春雷等^[10]利用 SSR

研究了部分地区的遗传多样性,表明梔子具有较高的遗传多样性水平,但收集的资源数量不够多且覆盖区域不够大。目前已有基于梔子 SSR 分子标记的遗传多样性研究,但对其全基因组 SSR 特征并没有进行深度分析。本研究基于梔子全基因组序列检索了其 SSR,深度剖析了全基因组 SSR 特征,从染色体长度、SSR 数量、不同重复类型、不同重复频次以及不同核苷酸类型的特征进行了分析,并讨论了梔子与药用植物、油料植物以及灌木植物的 SSR 特征的异同,丰富了梔子的 SSR 分子标记数据库,以期在后续的群体遗传多样性、品种鉴定及分子育种奠定基础。

1 材料与仪器

本研究使用的数据为梔子全基因组序列,由中国医学科学院北京协和医学院上传。基因组测序材料经中南林业科技大学杨艳研究员鉴定为梔子 *G. jasminoides* J. Ellis 叶片,采自重庆市南川区;Oxford Nanopore 为全基因组测序仪器。

2 方法

2.1 梔子基因组序列获取

梔子全基因组 (含 11 组染色体) 的序列数据 (GenBank: GCA_013103745.1) 从 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中下载获取。梔子全基因组 11 组染色体编号为 Chr01 (CM023095.1)、Chr02 (CM023096.1)、Chr03 (CM023097.1)、Chr04 (CM023098.1)、Chr05 (CM023099.1)、Chr06 (CM023100.1)、Chr07 (CM023101.1)、Chr08 (CM023102.1)、Chr09 (CM023103.1)、Chr10 (CM023104.1)、Chr11 (CM023105.1)。

2.2 SSR 位点检索

利用 MISA (1.0) 软件对梔子全基因组序列进

行 SSR 位点搜索, 参数标准设为由一到六核苷酸重复次数依次为 15、6、5、4、4、4; 距离在 100 bp 以上的视为一个 SSR 位点。梔子全基因组大小 534.9 Mb, 核型为 $2n=2x=22^{[11]}$ 。

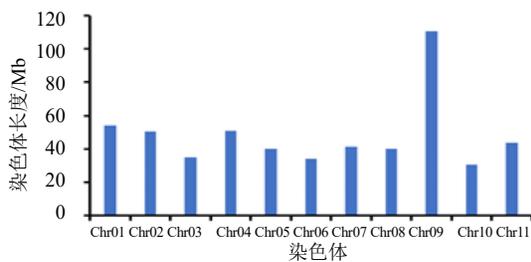
2.3 数据分析

文中的数据分析和图表制作均使用 Microsoft 2024 Excel 和 Origin 2024 软件完成。

3 结果与分析

3.1 梔子染色体长度、SSR 数量及频度密度分布特征

3.1.1 梔子染色体长度、数量 利用 MISA 软件对梔子全基因组进行 SSR 搜索, 共检索到 117 037 个 SSR, 长度在 12~817 bp, 平均长度为 24.68 bp。梔子基因组全长为 534.9 Mb, 最长的为 Chr09, 长达 110.58 Mb; 最短的为 Chr10, 仅有 30.74 Mb, SSR 出现频率为 444.23 bp/个 (图 1)。



Chr01~Chr11 代表染色体组编号 1~11。
Chr10 to Chr11 stands for chromosome numbers 1 to 11.

图 1 梔子染色体长度

Fig. 1 Chromosome length of *G. jasminoides*

3.1.2 梔子 SSR 分布频度密度 对检测到的 SSR 进行分类, 可分为 2 大类: (1) 完整型核苷酸: 即一至六核苷酸; (2) 混合型核苷酸: 不完全型 (c)、复合型 (c*)。检索到的 SSR 在染色体上共有 116 465 个, 占总数的 99.51% (一至六核苷酸共有 102 001 个, 占比 87.15%; 混合型核苷酸 14 464 个, 占比 12.36%)。本研究主要对染色体上完全型核苷酸进行特征分析, 11 组染色体上完全型核苷酸共计 102 001 个, 其 SSR 位点分布较为稳定。频度密度在 157~280 个/Mb, 平均每 1 Mb 中含有 179 个 SSR 位点; 其中 Chr09 最大 (280 个/Mb), 其次是 Chr11 (185 个/Mb) 和 Chr03 (176 个/Mb), Chr06 最小 (157 个/Mb) (图 2)。

3.1.3 梔子染色体长度与 SSR 位点数线性关系 从一到六核苷酸分别占各染色体的完全型 SSR 数量的百分比发现, Chr05 中的二核苷酸最高, 占比 46.26%; Chr08 上的六核苷酸最低, 占比为 1.06%,

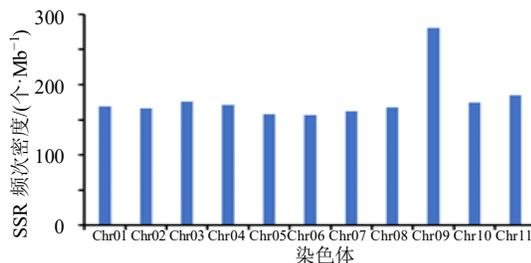


图 2 梔子全基因 SSR 在染色体组上的分布频度密度

Fig. 2 Distribution frequency density of whole gene SSR in *G. jasminoides* on chromosome

表明二核苷酸为梔子的优势种类。通过梔子全基因组 SSR 位点数量和染色体组长度的线性关系分析发现二者呈高度相关, 得到方程 $Y=0.002\ 9\ X+20.976$ ($R^2=0.970\ 2$), 即随着染色体组长度增大, SSR 的位点数随之增加 (图 3)。

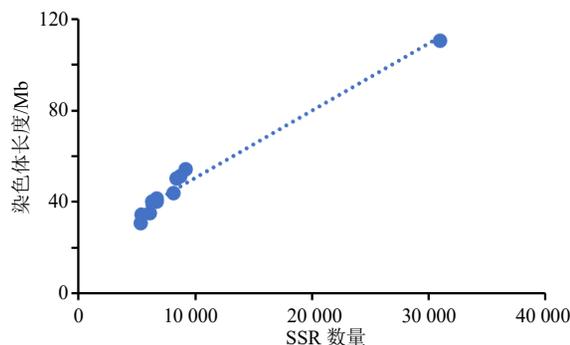


图 3 梔子染色体长度与 SSR 位点数线性方程

Fig. 3 Linear equation between chromosome length and SSR loci of *G. jasminoides*

3.2 梔子染色体不同 SSR 类型分布特征

3.2.1 不同核苷酸类型 SSR 在染色体上的分布 102 001 个 SSR 在梔子的 11 组染色体上均有分布, 平均每组染色体含有 9 272.81 个 SSR。其中, Chr09 中含有的 SSR 数量最多, 共 31 012 个, 占完全型核苷酸的 30.40%; 其次是 Chr01 和 Chr04, 分别含有 9 177, 8 740 个 SSR, 分别占比为 9.00%, 8.57%; SSR 数量最少的为 Chr06 和 Chr10, 各含有 5 372、5 360 个 SSR, 分别占完全型核苷酸的 5.27% 和 5.25% (图 4)。染色体不同 SSR 种类也有较明显差异, 从 Chr01 到 Chr11, 不同种类数分别为 817、776、675、835、671、603、713、692、1486、604、783, 最多的为 Chr09, 最少的为 Chr06 (图 5)。

3.2.2 不同 SSR 类型在染色体上的重复频次 从染色体水平来看, 不同 SSR 类型重复频次范围在 4~81, 重复次数在 4~5、6~10、11~15、16~20、

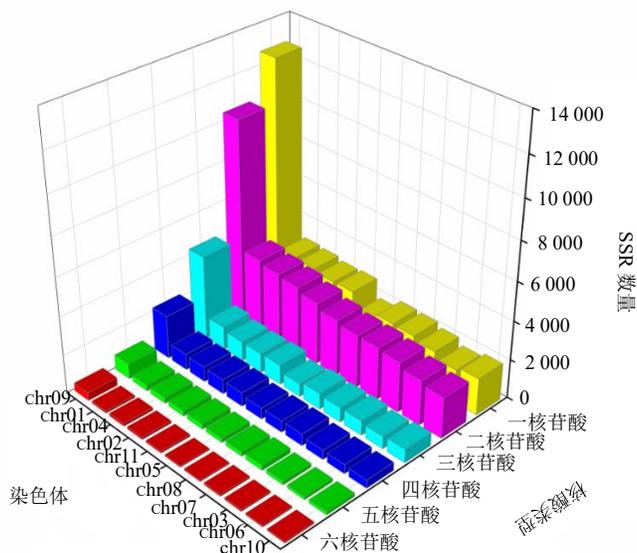


图 4 栀子全基因组不同核苷酸类型 SSR 在染色体上的分布

Fig. 4 Distribution of different nucleotide types of SSRs on chromosomes in whole genome of *G. jasminoides*

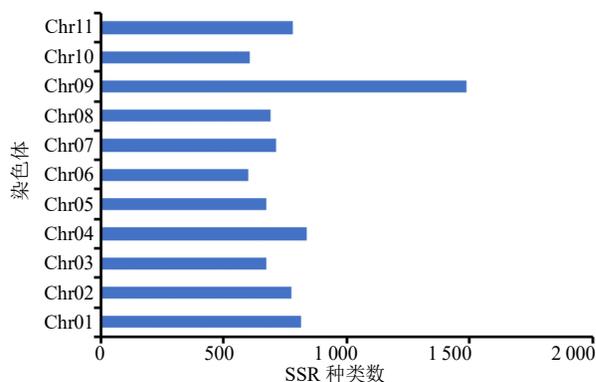


图 5 栀子染色体上 SSR 种类数

Fig. 5 Number of SSR types on *G. jasminoides* chromosome

21~25、26~30、31~35、36~40、41~45、46~50、51~81 间均有分布，分别有 18 594、38 388、17 599、23 646、3 330、328、72、28、9、5、2 个，重复次数分布在 6~10 的 SSR 位点数最多，共计 38 388 个；重复次数在 51~81 的 SSR 仅有 2 个，分别为一、二核苷酸。Chr01 到 Chr11 分别有 9 177、8 372、6 141、8 740、6 305、5 372、6 691、6 727、31 012、5 360、8 104 个 SSR 位点，Chr09 是 SSR 位点数最多的，主要分布在 6~10，占该染色体组的 34.06%。11 组染色体重复频次主要集中在 4~20，占比高达 96.30%，其中以重复频次 6~10 次最多，占比为 37.63%（图 6）。

3.2.3 不同 SSR 类型在染色体上的重复频次 从染色体水平来看，不同 SSR 类型重复频次范围在

4~81，重复次数在 4~5、6~10、11~15、16~20、21~25、26~30、31~35、36~40、41~45、46~50、51~81 均有分布，分别有 18 594、38 388、17 599、23 646、3 330、328、72、28、9、5、2 个，重复次数分布在 6~10 的 SSR 位点数最多，共计 38 388 个；重复次数在 51~81 的 SSR 仅有 2 个，分别为一、二核苷酸。Chr01 到 Chr11 分别有 9 177、8 372、6 141、8 740、6 305、5 372、6 691、6 727、31 012、5 360、8 104 个 SSR 位点，Chr09 是 SSR 位点数最多的，主要分布在 6~10，占该染色体组的 34.06%。11 组染色体重复频次主要集中在 4~20，占比高达 96.30%，其中以重复频次 6~10 次最多，占比为 37.63%（图 6）。

3.3 栀子全基因组 SSR 不同核苷酸类型特征分析

3.3.1 不同核苷酸类型长度分布 SSR 长度范围在 12~817 bp，不同核苷酸 SSR 长度存在极显著差异，SSR 长度 ≤ 15 bp 的有 37 718 个，均由一、二、三核苷酸构成，分别有 11 255、19 850、6 613 个，占总比的 11.03%、19.46%、6.48%；SSR 长度 > 15 bp 的有 64 283 个，并且在每一种类型核苷酸中均有出现，一至六核苷酸分别有 24 845、20 071、6 349、8 503、3 170、1 345，占总比的 24.36%、19.68%、6.22%、8.34%、3.11%、1.32%（图 7）。

3.3.2 完全型 SSR 不同核苷酸和重复类型的数量 栀子完全型核苷酸共计 102 001 个，从一至六核苷酸分别有 36 100、39 921、12 962、8 503、3 170、1 345 个，数量最多的是二核苷酸，占完全型核苷酸

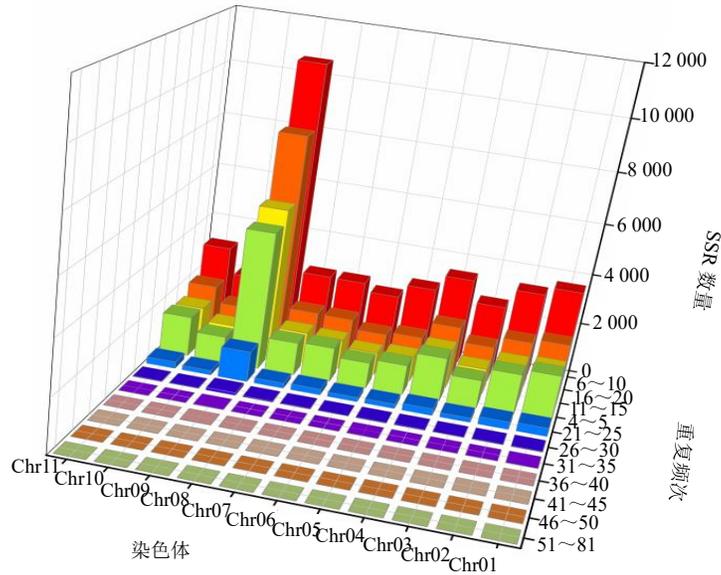


图 6 梔子不同染色体 SSR 重复频次分布

Fig. 6 Frequency distribution of SSR repeats on different chromosomes of *G. jasminoides*

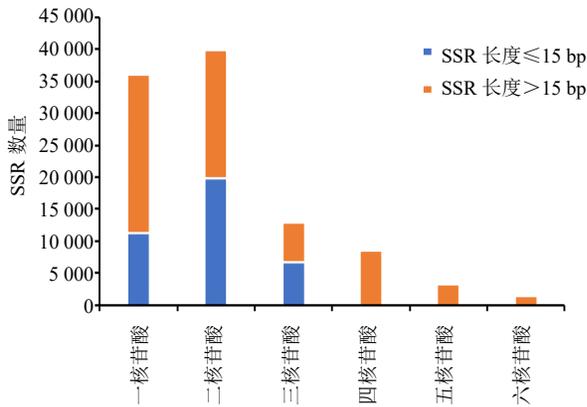


图 7 梔子全基因组 SSR 长度分布

Fig. 7 SSR length distribution of *G. jasminoides* whole genome

的 39.14%，最少的是六核苷酸，占比 1.32%；其不同重复类型种类共有 2 429 种，一至六核苷酸分别有 83、253、446、508、547、592 种不同重复类型，重复基序类型数共计 1 112 种，分别有 4、12、60、198、364、474 种，其中六核苷酸最多，但其仅占对应核苷酸重复基序理论数值 11.57%（表 1）。

3.3.3 不同重复类型 SSR 相关性分析 对梔子全基因组不同重复类型的 SSR 数量进行相关性分析发现，不同重复类型 SSR 位点间均呈显著相关关系，其中相关系数最大的为一核苷酸与三核苷酸（0.999），最小的为一核苷酸与五核苷酸，三核苷酸与五核苷酸（0.984）；在 SSR 数量上，六核苷酸和其他核苷酸间相关系数较低（0.988~0.991）（表 2）。

表 1 梔子完全型 SSR 不同核苷酸和重复类型的数量

Table 1 Number of different nucleotides and repeat types of complete SSRs in *G. jasminoides*

核苷酸重复类型	核苷酸总数	占完全型比例/%	重复类型数	重复基序类型数	重复基序理论数	占重复基序理论值比例/%
一核苷酸	36 100	35.39	83	4	4	100.00
二核苷酸	39 921	39.14	253	12	16	75.00
三核苷酸	12 962	12.71	446	60	64	93.75
四核苷酸	8 503	8.34	508	198	256	77.34
五核苷酸	3 170	3.11	547	364	1 024	35.55
六核苷酸	1 345	1.32	592	474	4 096	11.57
总计	102 001	100	2 429	1 112	5 460	20.37

3.3.4 不同核苷酸类型重复频次分布 梔子完全型 SSR 不同核苷酸类型的重复频次分布在 4~81，一核苷酸中有 91.19% 的 SSR 分布在 11~20 次，共

有 32 918 个；二核苷酸有 31 433 个 SSR 分布在 6~10 次，占二核苷酸的 78.74%；三核苷酸中各有 6 613、5 930 个 SSR 分布在 4~5、6~10；四、五、

表 2 栀子不同重复类型 SSR 相关性分析

Table 2 SSR correlation analysis of different replication types in *G. jasminoides*

重复单元	一核苷酸	二核苷酸	三核苷酸	四核苷酸	五核苷酸	六核苷酸
一核苷酸	1.000					
二核苷酸	0.993**	1.000				
三核苷酸	0.999**	0.995**	1.000			
四核苷酸	0.991**	0.998**	0.992**	1.000		
五核苷酸	0.984**	0.993**	0.984**	0.996**	1.000	
六核苷酸	0.991**	0.991**	0.990**	0.990**	0.988**	1.000

**在 0.01 级别 (双尾), 相关性显著; 不考虑碱基互补。

**Correlations are significant at 0.01 level (two-tailed). Base complementation is not considered.

六核苷酸重复频次主要分布在 4~5, 分别占比为 90.49%、94.98%、94.87%。SSR 重复频次大于 10 的主要为一核苷酸与二核苷酸, 共有 44 588 个, 占重复频次大于 10 的 SSR 的 99.04% (图 8)。

3.3.5 一至六核苷酸特征分析 一核苷酸共有 36 100 个, 其数量在完全型核苷酸中仅次于二核苷酸, 长

度范围在 15~81 bp, 重复频次在 15~81, 不同重复类型共 83 种, 相对应的不同重复基序类型有 4 种 (图 9)。一核苷酸长度主要分布在 11~20 bp, 占总数的 91.19%。其中, (T) 15 的长度最长, 为 84 360 bp, 数量最多, 共 5 624 个; (A) 81 重复频次最高, 为 81 次; 重复序列的重复频次有 99.09%集中在

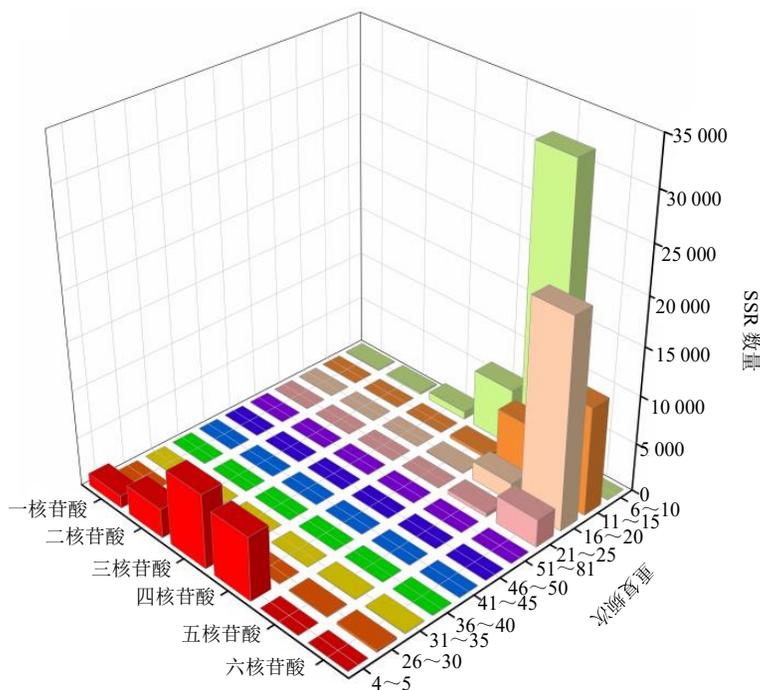


图 8 栀子不同基序类型的 SSR 重复频率分布

Fig. 8 Frequency distribution of SSR repeats of different motif types in *G. jasminoides*

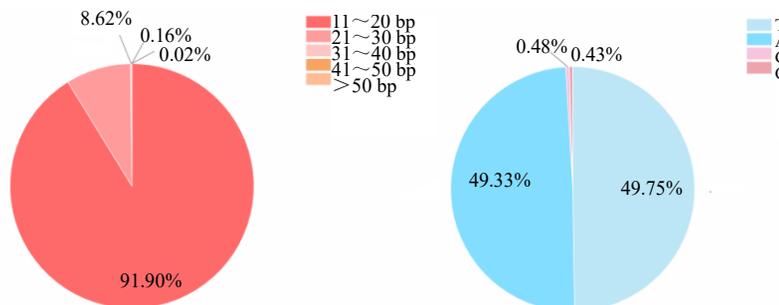


图 9 一核苷酸 SSR 位点中长度分布 (左) 和不同类型重复单元的占比 (右)

Fig. 9 Percentage (right) and length distribution (left) of different types of repeat units in mononucleotide SSR loci

T 和 A。一核苷酸重复基序中数量最多的是 T，占比为 49.75%，其次是 A、G、C，分别有 17 809、173、157 个，各占比 49.33%、0.48%、0.43%（图 9）。二核苷酸 SSR 数量共有 39 921 个，是完全型 SSR 中数量最多的，占完全型核苷酸的 39.14%。二核苷酸长度在 12~108 bp，主要分布在 12、14、16 bp，各有 12 559、7 291、5 137 个 SSR，分别占二核苷酸总数的 31.46%、18.26%、12.87%；其重复频次在 6~54，二核苷酸 SSR

重复频次主要集中在 6~10，共有 31 433 次，占二核苷酸总数的 78.74%。二核苷酸不同重复类型有 253 个，不同重复基序有 12 种，重复基序中数量最多的是 AT，共有 8 203 个，占二核苷酸总数的 20.55%，占完全型 SSR 的 8.04%；其次是 TA，共 5 180 个，占二核苷酸总数的 12.98%，占完全型 SSR 的 5.08%，重复基序数量最少的是 CG，仅 21 个，占二核苷酸总数的 0.05%（图 10）。

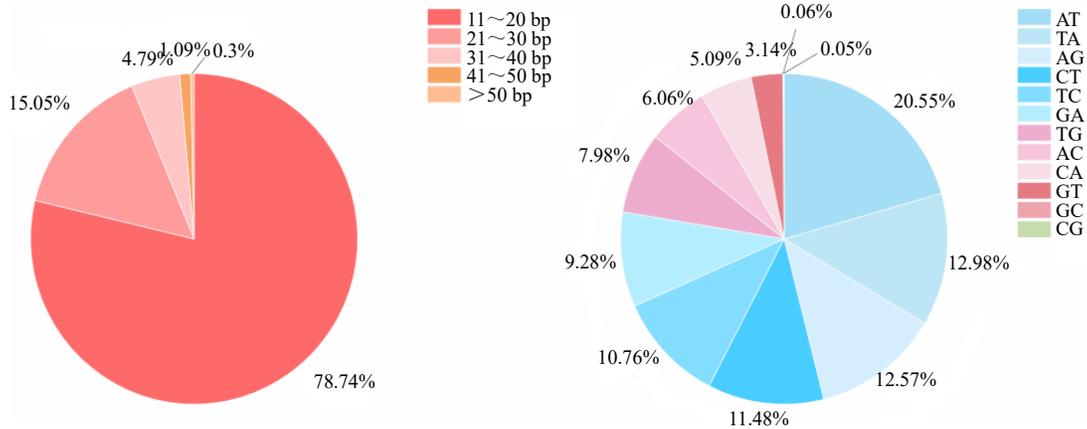


图 10 二核苷酸 SSR 位点中长度分布（左）和不同类型重复单元的占比（右）

Fig. 10 Percentage (right) and length distribution (left) of different types of repeat units in dinucleotide SSR locis

三核苷酸 SSR 共有 12 962 个，占完全型核苷酸的 12.71%，长度范围在 15~111 bp，重复频次在 5~37 次。三核苷酸长度主要分布在 15 bp 和 18 bp，分别占比 51.02%、24.31%。三核苷酸重复频次数量最多在 0~5 次，共计 6 613 个 SSR；不同重复基序类型中以 AAT、TTA 重复频次总和数量最多，各 280 次；（AAG）37 长度最长，为 111 bp。三核苷酸不同重复类型数 446 种，不同重复基序类型数 60 种。不同重

复频次数量最高的 SSR 基序是 AAG，重复频次 37 次，其次是 AAT，为 29 次。三核苷酸以 AAT、TTC、AAG 最多，分别有 1 306、1 151、1 039 个，占三核苷酸总数的 10.08%、8.88%、8.02%（图 11）。

四核苷酸 SSR 数量共有 8 503 个，占比为 8.34%，长度范围在 16~64 bp，重复频次在 4~16。其长度主要分布在 16 bp 和 20 bp，分别占比 67.53%、22.96%。四核苷酸重复频次数量最多的在 4~5 次，共 7 694 个

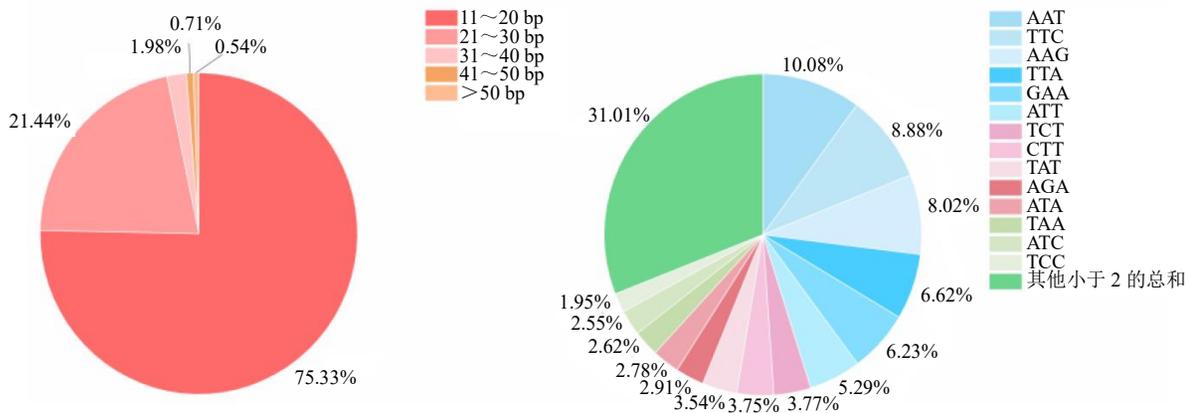


图 11 三核苷酸 SSR 位点中长度分布（左）和不同类型重复单元的占比（右）

Fig. 11 Percentage (right) and length distribution (left) of different types of repeat units in trinucleotide SSR locis

SSR；四核苷酸 SSR 有 508 种不同重复类型，198 种不同重复基序类型，其中以 TTCT、TTTC 重复频次总和数最高，不同重复频次计数均为 66 次；(TTCT)

16 的长度最长，为 64 bp。四核苷酸以 AAAT、TTTA、ATTT 的 SSR 数量最多，分别有 952 (11.20%)、928 (10.91%)、478 (5.62%) 个 SSR (图 12)。

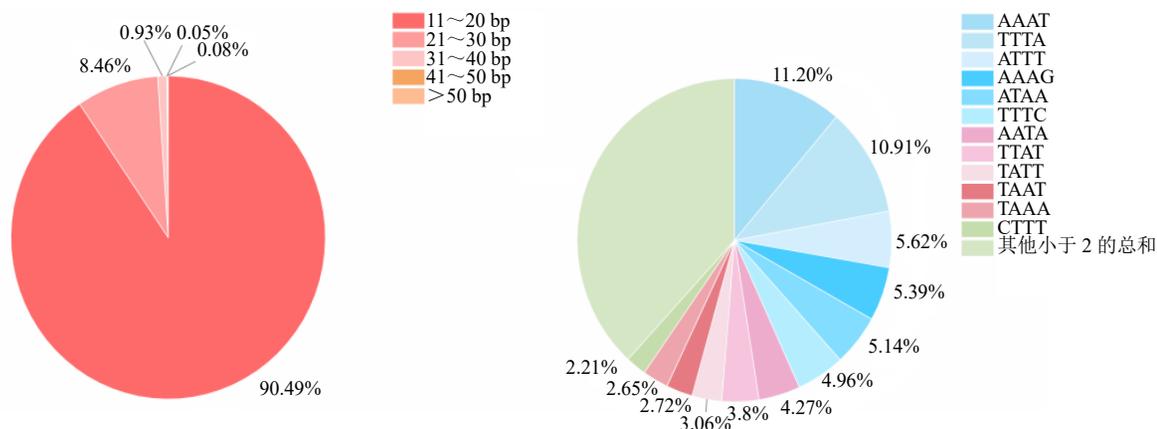


图 12 四核苷酸 SSR 位点中长度分布 (左) 和不同类型重复单元的占比 (右)

Fig. 12 Percentage (right) and length distribution (left) of different types of repeat units in tetranucleotide SSR locis

五核苷酸数量共 3 170 个，长度范围在 20~50 bp，意味着重复频次在 4~10。五核苷酸长度主要分布在 20 bp 以下，占比为 76.15%。其中以 (AAGAA) 10 重复频次最高以及长度最长。五核苷酸 SSR 不同重复类型共有 547 种，且有 364 种不同种类的重复基序。以重复基序 AAAAG、AAAAT、TTTTTC 的 SSR 数量最多，分别有 249

(7.85%)、197 (6.21%)、152 (4.79%) 个 SSR。不同重复类型的 SSR 数量大于 100 的仅有 4 个，且 SSR 数量小于 25 个的占的 80.88%。同一基序不同重复频次计数总和最多的 SSR 是 AAGAA，有 32 个，其次是 TTTTG、AAAAC、TTTTA、AAAAG、ATCAC，不同重复频次计数总和分别为 31、24、24、23、23 (图 13)。

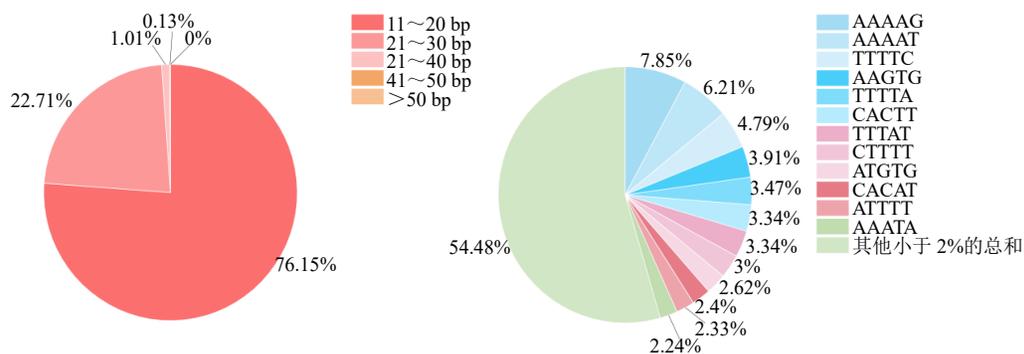


图 13 五核苷酸 SSR 位点中长度分布 (左) 和不同类型重复单元的占比 (右)

Fig. 13 Percentage (right) and length distribution (left) of different types of repeat units in pentanucleotide SSR locis

六核苷酸 SSR 共有 1 345 个，是完全型核苷酸中数量最少的，仅占 1.32%，长度范围在 24~66 bp，重复频次在 4~11 次。六核苷酸长度主要分布在 24 bp，分别占比 78.29%。六核苷酸重复频次数量最多在 0~5 次，共计 1 276 个 SSR；以重复基序 AAAAAT、AAAAAG 数量最多，分别有 62、60 个 SSR，分别占六核苷酸总数的 4.61%、4.46%。不同重复基序类型中

以(TTATCA)11 长度最长且重复频次最多，长 66 bp，重复频次 11 次。六核苷酸不同重复类型数 592 种，不同重复基序类型数 474 种。同一基序不同重复频次计数总和最多的 SSR 是 ATCATT、TAATGA，各有 41、39 个。在不同重复序列中数量最多的为 ATCATT、TAATGA，各 6 个，其次为 TTATCA、AAAAAG、TATCAT，分别有 5、4、4 个 (图 14)。

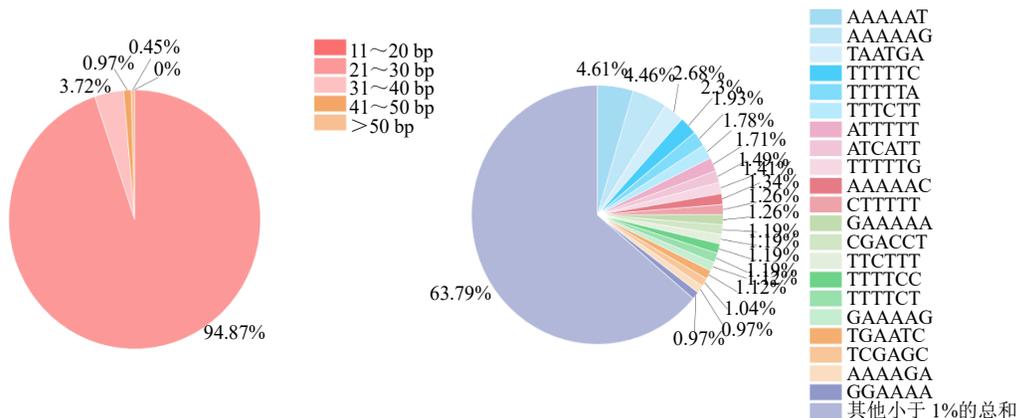


图 14 六核苷酸 SSR 位点中长度分布 (左) 和不同类型重复单元的占比 (右)

Fig. 14 Percentage (right) and length distribution (left) of different types of repeat units in hexanucleotide SSR loci

4 讨论

栀子是我国传统中药材之一，且栽培适应性强，是优良的园林绿化及室内盆栽植物。栀子花、果富含萜类化合物，具有抗炎、抗癌等作用，目前对栀子的研究主要集中在内含物提取及药理活性，国内外对其分子标记的研究进展相对缓慢^[12]。本研究利用 MISA 软件对栀子全基因组进行搜索并分析 SSR 特征，检测到 SSR 位点数共 117 037 个，完全型核苷酸有 102 001 个，多于药用植物丹参^[13]、杜鹃花^[14]、大叶铁线莲^[15]、无患子^[16]等，少于同为药食同源植物的薏苡^[17]、葛根^[18]，药用植物裸花紫珠^[19]、黄麻^[20]、闽楠^[21]以及同为灌木的怒江红山茶^[22]等。栀子 SSR 重复类型数量共 2 429 种，多于同为药用植物的半夏^[23]、丹参、薏苡、裸花紫珠，由此说明栀子全基因组 SSR 重复类型数量丰富。栀子全基因组 SSR 位点的不同重复基序类型共有 1 112 种，高于半夏、油茶^[24]、杜仲^[25]、丹参、文冠果^[26]等，表明栀子全基因组 SSR 位点丰富性高。

经统计栀子全基因组 SSR 数量丰富，不同核苷酸类型的 SSR 数量差异较为明显，栀子全基因组中二核苷酸 SSR 数量最多，占比为 34.3%，这与半夏、三七^[27]、杜鹃花、大叶铁线莲、葛根、怒江红山茶、光皮树^[28]一致，而在丹参中三核苷酸数量是最多的，占总 SSR 数量的 61.3%，在罂粟^[29]、薏苡、裸花紫珠全基因组中单核苷酸的 SSR 数量是最多的。从二核苷酸到六核苷酸，栀子 5 种完全型 SSR 数量随着重复基元的增多而减少。有研究表明 SSR 的低级重复单元越多则该物种进化水平较高，而高级重复单元越多则表明其变异频率低或进化时间短^[30]，

由此可判断栀子具有较高的物种进化水平。

栀子全基因组 SSR 不同重复基元中，从一至六核苷酸各优势重复基元依次为 T、AT、AAT、AAAT、AAAAG、AAAAAT，栀子 SSR 序列在一定程度上偏好于 A 与 T 所构成的碱基，这与同为中药材的三七、丹参、杜鹃花、黄麻结果一致，而小麦等植物中是以 C 和 G 为主要重复单元，在怒江红山茶中二核苷酸重复类型以 AG 基元占多数，三核苷酸重复类型中 AAG、ACC 和 ATC 占比较多，表明不同物种之间 SSR 重复基序普遍存在显著差异。碱基偏好性可以反映密码子的使用偏好情况，不同的密码子能够形成不同的氨基酸种类，使得不同植物有着不同的表型特征和生理表达。有研究表明，在植物中，二核苷酸基序(AT)_n重复类型最为普遍^[31-32]，导致这种现象的原因可能是富含 A/T 碱基的重复序列在 DNA 中比较容易解链^[33]。综上所述，本研究丰富了栀子的 SSR 分子标记库，可为栀子种质资源收集、核心种质构建、群体遗传多样性、品种鉴定、分子标记辅助育种及资源保护提供一定的研究基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志(第十九卷) [M]. 北京: 科学出版社, 1999: 332.
- [2] 赵梦亭, 朱如意, 应佳亮, 等. 栀子炮制历史沿革、临床应用及质量评价研究进展 [J]. 中华中医药杂志, 2021, 36(4): 2229-2237.
- [3] 中国药典 [S]. 一部. 2020: 263.
- [4] 孟祥乐, 李红伟, 李颜, 等. 栀子化学成分及其药理作用研究进展 [J]. 中国新药杂志, 2011, 20(11): 959-967.
- [5] 朱颖洁, 杨亚洁, 李群和, 等. 栀子果油的生物活性成

- 分及抗氧化活性评价 [J]. 现代食品科技, 2021, 37(2): 103-113.
- [6] 金锐, 张从合, 朱全贵, 等. 分子标记辅助选择在玉米抗病和抗虫育种上的应用 [J]. 安徽农业科学, 2021, 49(16): 10-15.
- [7] 潘媛, 陈大霞, 宋旭红, 等. 基于 SCoT 标记的栽培栀子种质资源遗传多样性研究 [J]. 中草药, 2018, 49(14): 3376-3381.
- [8] 姜武, 吴志刚, 陶正明, 等. 基于 ISSR 和 SRAP 标记的栀子种质遗传多样性研究 [J]. 中草药, 2019, 50(2): 510-516.
- [9] Deng S Y, Wang X R, Zhu P L, *et al.* Development of polymorphic microsatellite markers in the medicinal plant *Gardenia jasminoides* (Rubiaceae) [J]. *Biochem Syst Ecol*, 2015, 58: 149-155.
- [10] 刘春雷, 胡开治, 刘燕琴, 等. 栀子全基因组 SSR 标记开发及遗传多样性分析 [J]. 中国中医药信息杂志, 2022, 29(10): 110-115.
- [11] 杨兆起, 程尧楚, 田自强, 等. 栀子染色体核型带型研究 [J]. 中国中药杂志, 1995, 20(10): 588-590.
- [12] 邓绍勇, 李康琴, 朱培林, 等. 栀子遗传资源研究现状与展望 [J]. 农业科技与信息: 现代园林, 2015, 12(10): 756-760.
- [13] 徐海滨, 钱俊, 汪波, 等. 丹参基因组 SSR 位点的特征分析及引物开发 [J]. 世界科学技术—中医药现代化, 2013, 15(3): 367-370.
- [14] 龚记熠, 刘欢, 张习敏, 等. 杜鹃花属序列中微卫星位点特征分析和引物开发 [J]. 分子植物育种, 2018, 16(22): 7403-7406.
- [15] 宋琳琳, 王豆豆, 周晓君. 基于高通量测序的大叶铁线莲基因组微卫星特征分析 [J]. 分子植物育种, 2022, 20(4): 1087-1094.
- [16] 丁澜, 蒋丽娟, 李培旺, 等. 无患子全基因组 SSR 特征分析 [J/OL]. 分子植物育种, (2024-07-11) [2025-03-19]. <http://kns.cnki.net/kcms/detail/46.1068.s.20240708.1210.007.html>.
- [17] 杜磊, 蒙秋伊, 尚昆, 等. 薏苡基因组 SSR 标记开发与应用 [J]. 分子植物育种, 2022, 20(3): 887-894.
- [18] 周荣荣, 周骏辉, 南铁贵, 等. 葛根 SSR 特征及江西粉葛 DNA 身份证的构建 [J]. 中国中药杂志, 2019, 44(17): 3615-3621.
- [19] 黄梅, 李鑫, 王凯, 等. 裸花紫珠群体遗传结构及其有效成分与 SSR 标记的关联分析 [J]. 中草药, 2024, 55(23): 8162-8170.
- [20] 姚嘉瑜, 张立武, 赵捷, 等. 黄麻全基因组 SSR 鉴定与特征分析 [J]. 作物学报, 2019, 45(1): 10-17.
- [21] 熊远芳, 江淑珍, 连辉, 等. 闽楠基因组 SSR 的开发及其通用性分析 [J]. 分子植物育种, 2024, 22(23): 7764-7773.
- [22] 王大玮, 周凡, 沈兵琪, 等. 基于高通量测序的怒江红山茶微卫星位点特征分析 [J]. 植物研究, 2019, 39(1): 148-155.
- [23] 张君毅. 半夏遗传多样性分子标记及质量评价研究 [D]. 南京: 南京农业大学, 2007.
- [24] 代涵, 申铁, 石桃雄, 等. 油茶基因组 SSR 位点挖掘及遗传多样性分析 [J]. 作物杂志, 2024(3): 23-31.
- [25] 吴敏, 杜红岩, 乌云塔娜, 等. 杜仲基因组微卫星特征及 SSR 标记开发, 林业科学研究, 2015, 28(3): 387-393.
- [26] 乐琳琳, 杨晓明, 郁万文, 等. 基于文冠果 (*Xanthoceras sorbifolium* bunge.) 基因组 SSR 标记的开发及初步验证 [J]. 分子植物育种, 2023, 21(9): 2971-2979.
- [27] 揭应碧, 卢迎春, 宋婉玲, 等. 三七基因组 SSR 位点分析和多态性引物开发 [J]. 热带作物学报, 2018, 39(7): 1351-1358.
- [28] 周文才, 唐山, 何小三, 等. 光皮树基因组 SSR 特征分析 [J]. 分子植物育种, 2023, 21(5): 1618-1623.
- [29] 凌立贞, 杜毛毛, 张书东. 基于罂粟基因组的 SSR 位点鉴定及其特征分析 [J]. 六盘水师范学院学报, 2023, 35(6): 99-105.
- [30] 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.
- [31] Wang H, Varma Penmetsa R, Yuan M, *et al.* Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.) [J]. *BMC Plant Biol*, 2012, 12: 10.
- [32] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.
- [33] Gur-Arie R, Cohen C J, Eitan Y, *et al.* Simple sequence repeats in *Escherichia coli*: Abundance, distribution, composition, and polymorphism [J]. *Genome Res*, 2000, 10(1): 62-71.

[责任编辑 时圣明]