

DOI:10.13275/j.cnki.lykxyj.2022.004.005

刀状黑黄檀基因组特征与种群遗传变异分析

刘宇^{1,2}, 郑勇奇¹, 李长红¹, 林富荣¹, 黄平^{1*}

(1. 林木遗传育种国家重点实验室, 中国林业科学研究院林业研究所, 国家林业和草原局森林培育重点实验室, 北京 100091;
2. 温州市资源植物创新利用重点实验室, 浙江省农业科学院浙江省亚热带作物研究所, 浙江温州 325005)

摘要: [目的] 本研究旨在分析刀状黑黄檀基因组基本特征, 发掘一套多态性好、重现性高的基因组 SSR 分子标记并用于评估其野生种群遗传变异程度。[方法] 利用二代测序 (NGS) 方法初步评估刀状黑黄檀基因组大小、杂合度和重复序列; 利用 MISA 对基因组上潜在的 SSR 位点进行检索, 解析 SSR 类型与特征; 利用 Primer Premier v 5.0 设计合成 300 对引物, 通过琼脂糖凝胶电泳、非变性聚丙烯酰胺凝胶电泳对潜在的 SSR 位点进行初筛与复筛, 最后采用荧光标记毛细管电泳进行 SSR 位点多态性与种群遗传多样性分析。[结果] 刀状黑黄檀基因组特征分析结果显示: 其基因组大小约为 706.92 Mb, 杂合度为 1.26%, 重复序列比率为 55.74%。通过筛选实验, 发掘了 27 个具有良好多态性的 SSR 标记, 并进行了验证。结果显示 27 个 SSR 位点共扩增得到 117 个等位基因, 多态性信息含量 (PIC) 介于 0.149 ~ 0.803。对 3 个野生种群遗传多样性分析结果显示: 种群水平的平均期望杂合度 (H_e) 为 0.504, 遗传分化系数 (F_{ST}) 为 0.034; AMOVA 分析显示: 种群间变异占比 3.46%, 种群内变异占比 96.54%, 这表明遗传变异主要来源于种群内。[结论] 刀状黑黄檀基因组属于高杂合、高重复的复杂基因组, 研究结果为制定基因组精细组装策略提供了参考依据; 发掘与验证的 27 个 SSR 标记具有较好的多态性与扩增稳定性; 3 个野生种群具有中等遗传多样性和较低的遗传分化, 本研究为开展刀状黑黄檀的种质资源收集保存与分析评价提供科学依据。

关键词: 刀状黑黄檀; 下一代测序技术; 基因组 SSR; 多态性; 遗传多样性

中图分类号: S722.3

文献标志码: A

文章编号: 1001-1498(2022)04-0044-10

刀状黑黄檀 (*Dalbergia cultrata* Graham ex Benth) 隶属于豆科 (Fabaceae) 黄檀属 (*Dalbergia*), 高大落叶乔木, 其主要分布区位于中国云南省南部以及相毗邻的中南半岛的热带与亚热带地区^[1]。刀状黑黄檀是珍贵的红木树种, 由于其心材呈黑褐色, 新切面具有酸香气味, 新切面紫、黑或栗褐, 常带紫或黑褐窄条纹, 在中国《红木》(GB/T 18107-2017) 国家标准中被归类于黑酸枝木系, 是制作高档家具、乐器以及工艺品的优质材料; 同时, 心材和根可入药, 具有较好的药用价值^[2]; 刀状黑黄檀根系发达, 对土壤要求不高, 且具有良好的抗病、抗虫等特性, 是困难立地生态修复的优良树种^[3]。以往研究主要关注了刀状黑黄

檀的野生资源分布以及气候变化对其影响、育苗方法、木材内含物化学成分等方面^[3-6]。刀状黑黄檀心材形成周期长导致了突出的供需矛盾, 也造成了野生资源的过度开发, 而经济社会高速发展过程中的土地利用转变以及气候变化的风险可能进一步挤压其原生栖息地范围, 致使野生资源数量进一步减少, 野生资源正在遭受遗传侵蚀的风险^[4]。尽管刀状黑黄檀早已被列入 IUCN (International Union for Conservation of Nature) 红色名录 (近危级, Near threatened, NT) 以及《中国高等植物受威胁物种名录》(易危级, Vulnerable, VU A2c)^[7], 但是关于该物种种群遗传保护研究尚未见报道, 也缺少相关背景基因组信息以及物种特异性的遗传标记。

收稿日期: 2022-02-22 修回日期: 2022-03-14

基金项目: 国家自然科学基金项目 (31761143002); 国家林业与草原种质资源库 (2005DKA21003)

作者简介: 刘宇, 博士, 助理研究员。主要研究方向: 林木遗传育种。Email: liuyu@caf.ac.cn

* 通讯作者: 黄平, 博士, 助理研究员。主要研究方向: 林木种质资源保护。Email: huangping@caf.ac.cn

分子标记是有效量化和科学评价种群遗传变异程度的分析工具,被广泛应用于遗传图谱构建、数量性状定位、遗传多样性与遗传结构评价等方面^[8]。SSR分子标记是群体遗传学研究中最常用的DNA分子标记之一,随着NGS测序技术的发展与成本降低,基于测序数据,通过生物信息学分析方法可快速、高效的发掘出大量的潜在SSR标记位点,进一步结合分子实验筛选验证则可对多态性基因组SSR标记进行鉴定,在林草遗传分析中已有诸多应用,如元宝枫(*Acer truncatum* Bunge)、象草(*Pennisetum purpureum* Schum.)和开心果(*Pistacia vera* L.)等^[9-11]。

为了掌握刀状黑黄檀的遗传信息,开发一套可用于遗传分析的适宜分子标记,本研究基于NGS测序技术对刀状黑黄檀基因组的基本特征进行了初步评估,通过实验筛选开发了一套重复性高、多态性好的SSR标记,并在其野生种群中进行了初步验证与评价分析。研究结果不仅可为刀状黑黄檀群体遗传学研究提供可靠分析工具,而且为

进一步深入开展种质资源收集保存、分析评价提供科学依据。

1 材料与方法

1.1 试验材料

基因组测序试验材料:2018年3月在云南省普洱市六顺乡(LSX)采集刀状黑黄檀成熟种子,采种母树凭证标本存放于中国林业科学研究院林业研究所。于当年在中国林科院科研温室播种育苗,育苗基质为草炭土与蛭石(3:1),常规水肥管理,选择长势良好且健康的植株,采集新鲜叶片,液氮保存,用于基因组DNA提取。

SSR分子标记筛选试验材料:选择云南省普洱市六顺乡(LSX)、江城(JC),西双版纳傣族自治州景洪(JH)等3个地点进行采样,采样地理环境信息见表1。每个采样点随机选择30个个体,共计90个样品,不同个体间距离大于50m。采集新鲜叶片,经硅胶完全干燥后,放置于-20℃冰箱保存,用于DNA提取。

表1 刀状黑黄檀采样点信息

Table 1 Sampling information of *D. cultrata*

序号 No.	群体 Populations	代码 Code	经度(E) Longitude	纬度(N) Latitude	海拔 Altitude/m	年平均温度 Annual mean temperature/℃	年平均降水量 Annual mean precipitation/mm
1	六顺乡 Liushunxiang	LSX	100°47'	22°40'	800~1 250	19.9	1 460.4
2	江城 Jiangcheng	JC	101°43'	22°33'	800~1 250	20.3	1 783.0
3	景洪 Jinghong	JH	100°31'	21°59'	600~1 350	20.2	1 450.3

1.2 DNA提取和测序

采用CTAB法^[12]从叶片样品中提取基因组DNA,分别用1.0%琼脂糖凝胶电泳和ScanDrop 100分光光度计(Molecular Device, California, United States)检测DNA浓度及质量。利用illumina HiSeq™ 2500测序平台(illumina, San Francisco, United States),基于标准规程构建一个插入片段大小为300 bp的文库。基于K-mer分析,估计基因组大小、杂合度以及重复序列比率^[13]。

参考Luo等的方法^[14],利用SOAPdenovo v 2.04软件进行基因组初步组装。构建Contig序列(K-mer=17)并进一步组装为Scaffold,使用GapCloser v 1.12-r6修补Scaffold内部缺口。以每10 kb作为1个窗口,进行非重叠滑窗统计,计算每个窗口的平均深度与GC含量,获得GC-depth

点图,进而估计基因组GC含量。

1.3 SSR位点检测与引物设计

采用MISA(<http://pgrc.ipk-gatersleben.de/misa/misa.html>)对初步组装Scaffold进行潜在SSR位点检索,挖掘候选SSR标记。参数设置为:二至六核苷酸基序重复单元最小重复次数值分别设置为6、5、4、4、4。利用Primer Premier v 5.0软件设计引物,引物参数分别为:引物长度18~22 bp,产物大小100~300 bp,退火温度55~62℃,最优值为60℃。随机挑选300对引物,由睿博兴科生物公司(北京)合成。

1.4 SSR引物筛选与验证

从六顺乡(LSX)、江城(JC)和景洪(JH)3个种群90个样品中随机挑选出4个样品对300对引物进行琼脂糖凝胶电泳初筛,选择扩增条

带清晰, 稳定性良好的引物进行复筛。复筛时在 90 个样品中随机挑选 8 个样品进行 8.00% 非变性聚丙烯酰胺凝胶电泳检测, 扩增产物条带明亮、清晰, 无背景噪音, 即复筛合格引物。然后, 利用荧光毛细管电泳检测候选 SSR 位点多态性, 最后利用多态性 SSR 位点分析 3 个种群 (90 个个体) 遗传变异程度, 初步评估 SSR 位点适用性以及种群遗传变异程度。

PCR 反应体系为: 基因组 DNA 50 ng, 2 × Taq MasterMix (Dye) (康为世纪, 北京) 12.5 μL, 正向引物 (10 μmol·L⁻¹) 0.5 μL, 反向引物 (10 μmol·L⁻¹) 0.5 μL, 用 ddH₂O 补至 25 μL。PCR 程序为: 96 °C 预变性 3 min, 然后 96 °C 变性 30 s, 60 °C 退火 30 s, 72 °C 延伸 45 s, 共 30 个循环, 最后 72 °C 延伸 7 min。

利用 GeneMaker v 2.2.0 对 SSR 位点进行基因分型, 并进行二次人工复核。

1.5 遗传多样性分析评价

采用 GenAIEx v 6.502^[15] 计算下列群体遗传参数, 包括观测等位基因数 (N_a), 有效等位基因数 (N_e), 观测杂合度 (H_o), 期望杂合度 (H_e), Shannon's 信息指数 (I), 种群间遗传分化系数 (F_{ST}), 哈温平衡偏离显著性 ($HWE (p)$) 等。采用 GeneALEX v 6.502^[15] 进行分子方差分析 (AMOVA) 以检测遗传变异来源。采用 CERVUS v 3.0.4^[16] 计算位点多态性信息含量 (PIC)。

2 结果与分析

2.1 基因组特征分析

基于 illumina HiSeq™ 2500 平台测序共生成 94.52 Gb 高质量数据, 覆盖度约为基因组的 137 倍。测序质量值 Q20 和 Q30 分别为 96.24% 和 90.23%, 表明测序质量良好, 可用于后续分析。使用 SOAPdenovo 软件进行组装和拼接, 选择 K-mer 为 17, 初步组装基因组, 共获得 463 258 个 Scaffold, Scaffold N50 长度约为 2 381 bp (表 2)。将所有高质量数据用于 K-mer 分析, K-mer 分布曲线结果显示: K-mer 期望深度 (主峰) 对应的测序深度为 126 ×, GenomeScope 估计基因组约为 706.92 Mb, 基因组杂合度为 1.26%, 重复序列比率为 55.74%, 平均 GC 含量为 34.11%。

2.2 SSR 标记类型

利用 MISA 进行 SSR 候选位点检索, 从

表 2 刀状黑黄檀基因组组装信息

Table 2 Information of the assembled genome sequences of *D. cultrata*

组装信息 Information of assembly	长序列片段 Scaffold	重叠群 Contig
拼接序列数量 Number of sequences	463 258	720 549
基因组大小 Total length/bp	777 543 915	776 970 961
平均长度 Average Length/bp	1 678.43	1 078.30
最大长度 Maximum Length/bp	96 195	96 195
N50长度 N50 Length/bp	2 381	1 576
N90长度 N90 Length/bp	692	514

463 258 个 Scaffold 中总共鉴定出 142 891 个 SSR 标记 (表 3), 其中, 二核苷酸 (Di-) 位点数量最多, 占总数的 61.21%, 主要重复类型为 AT/AT, 占 54.75%; 三核苷酸 (Tri-) 次之 (24.88%), 主要类型为 AAT/ATT, 占 54.23%; 四核苷酸 (Tetra-) 占比为 7.45%, 主要类型为 AAAT/ATTT, 占比为 68.11%; 五核苷酸 (Penta-) 占比为 1.11%, 主要类型为 AAAAT/ATTTT。

表 3 刀状黑黄檀的基因组 SSR 分布特征

Table 3 Characteristic and distribution of SSR in *D. cultrata* genome

SSR 重复基序 SSR repeat motif	SSR 数量 Number of SSR	频率 Frequency/%	主要类型 Main type
二核苷酸 Dinucleotide	87 464	61.21	AT/AT
三核苷酸 Trinucleotide	35 558	24.88	AAT/ATT
四核苷酸 Tetranucleotide	10 651	7.45	AAAT/ATTT
五核苷酸 Pentanucleotide	1 593	1.11	AAAAT/ATTTT
其它 Others	7 625	5.34	—
总计 Total	142 891	—	—

2.3 SSR 筛选与有效性验证

琼脂糖凝胶电泳初筛结果显示: 128 对引物具有清晰稳定的扩增条带, 占检测引物比例为 42.67%; 经过 8.00% 非变性聚丙烯酰胺凝胶复筛检测, 最终共筛选了 27 对扩增稳定, 条带清晰明亮, 多态性良好的引物, 约占检测引物总数的 9.00%, 引物详细信息见表 4。

利用筛选后具有多态性的 SSR 位点对 3 个种群的 90 个刀状黑黄檀个体进行 SSR 基因分型 (图 1)。检测结果显示: 在 27 个 SSR 位点上共检测到 117 个等位基因, 观测等位基因数 (N_a) 最少为 2 个, 最多为 10 个, 均值为 4.333; 有效等

表 4 刀状黑黄檀 27 个 SSR 位点的侧翼引物信息

Table 4 Characteristics of flanking primers of 27 SSR markers developed for *D. cultrata*

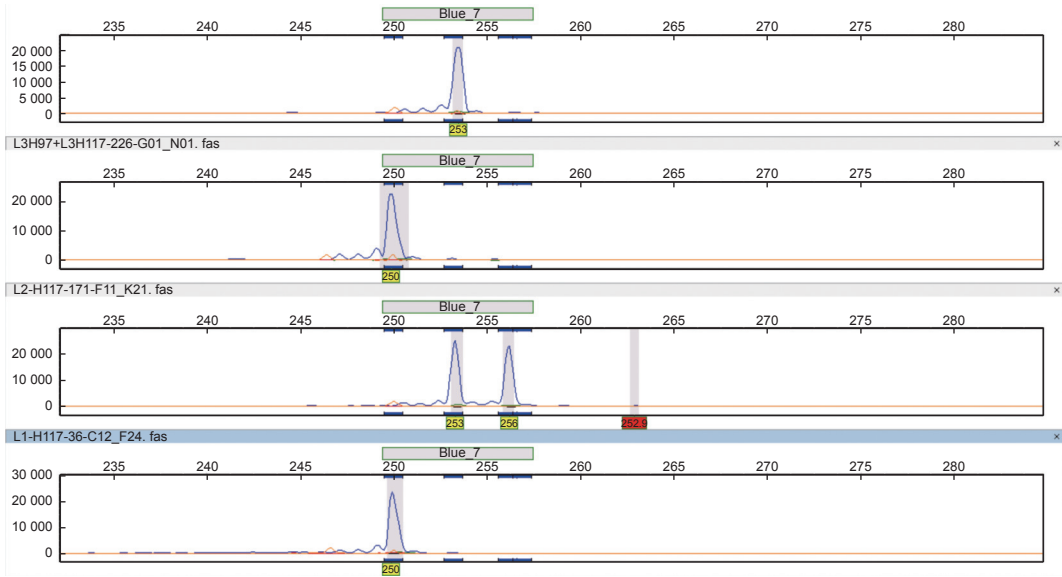
位点 Locus	重复单元 Repeat Motif	正向引物序列 Forward Primer	反向引物序列 Reverse Primer	产物大小 Product Size/bp	退火温度 Tm/(°C)
HHT5	(TTAT) ₄	GCAGCCAACCTATGTTCCAT	TAAGTGCTGAGTGATTGCCG	180	60
HHT10	(CCTGG) ₄	GGCCAGTTTGATAAACCCAT	AGTCAGGGCACTGAAGGAGA	195	60
HHT14	(CAC) ₇	GACCTCGACAAGCTCAAAGC	CAGCCCAATAAGTTGTCCGGT	254	60
HHT23	(AAAAT) ₄	GATCGGGCTCGTGGTTAATA	ATCTTCGGGCCATTTTCTTT	161	60
HHT45	(AGAAAA) ₄	TCCATCATCCGTCACAAAGA	AATGCTGTGCTTTCCAGGTC	276	60
HHT49	(AAT) ₁₂	CAGAAGCCAGAAACGTCGTA	GGGGCCGAAAGAAAAGATAG	186	60
HHT60	(CTAT) ₅	CCACTACTGCTCCTTCGCAT	ACAACGTACGTAGGCAACCC	160	60
HHT62	(AAC) ₇	CGCCCTTCTCCATATTCAA	GGAAATTGGTTGGCTTCGTA	277	60
HHT85	(TC) ₁₀	ACACGGCACACATTCTCGTA	CGCTCCATGAAGGAAGAAAG	234	60
HHT97	(GCA) ₁₁	CCACAACCAGCAACAACAAC	AAGTCCTTCATCATCCCGTG	223	60
HHT117	(ATT) ₇	ATAAGTATGTGCCGGCATGG	GAGATGGATGCATCGGTAAAA	253	60
HHT119	(AC) ₁₀	TCATGGTACTGTCATCCCA	CTCAAGGCTGACCAAAGGAG	255	60
HHT121	(AT) ₁₀	GCTTGAGGAGATGGTGAAGG	CCAATTTTCGGTAATGACAAGA	279	59
HHT124	(GA) ₁₀	GGAAATGCCACACAAGAGT	TGACACATGTGGGATATGGA	187	58
HHT126	(TG) ₁₀	CAGATGATCCAACCCAATCA	TACCAAACCCAAACCCAATC	242	60
HHT160	(CT) ₁₁	TCAAATCGCAAATGGCAATA	GGAATACGACGTCATGCAAA	188	60
HHT161	(AC) ₁₀	GACACAGGAAAAGGAAGATGG	GTGGAGCCGACAAGAAATGT	185	60
HHT211	(TTTTA) ₄	ATATCATTTCCACACAGCCA	CCAAAAACAACAAGCTGGAA	221	59
HHT216	(AAAAC) ₄	AGGGCTGCAAAAAGTGAAGA	GGTTGTAGCCGAGGAAAGAA	263	59
HHT219	(AATAAA) ₄	TGGTGAACTGGTCCATGAG	TTTGACCGCATGTTACCTGA	236	60
HHT223	(ATTAG) ₄	CCGCAAAATATATGACGGCT	GAGAGAGGCACATTTGACCG	203	61
HHT224	(AAAAT) ₄	AGCCCAATCCTTTTCTTCGT	ATTGCAATACATGGACCCGT	265	60
HHT226	(TTTTA) ₄	TGTCTCGCTCTCGCAGATAA	CAACCGCCAAAATTACATGA	271	59
HHT229	(ATTTT) ₄	TGGAGGGTGTGTTTGTGCTT	GCTGGAGAAGAGGTCATTGC	279	60
HHT230	(TTTGCA) ₄	CTTACGAGTGGGAATGGGAA	AAAAGAGATGGGGCGAAGAT	214	60
HHT247	(TTGAG) ₄	GAAGTAGGCAGGAAGCATCG	ATTAGCCGGTTGTGGTTTTG	199	60
HHT255	(CAAACA) ₅	TATGATGGCTTTTGGGGAGA	CCGGTGTGTTGTTGCTATG	273	60

位基因数 (N_e) 最少为 1.193, 最多为 5.757, 均值为 2.397。位点水平的观测杂合度 (H_o) 最小值为 0.111 (HHT226), 最大值为 0.833 (HHT119), 均值为 0.451; 期望杂合度 (H_e) 最小值为 0.162 (HHT226), 最大值为 0.826 (HHT121), 均值为 0.523。多态性信息含量 (PIC) 最小值为 0.149 (HHT226), 最大值为 0.803 (HHT121), 均值为 0.463, 其中, 11 个位点具有高多态性 ($PIC > 0.5$), 13 个位点具有中等多态性 ($0.25 < PIC \leq 0.5$), 仅有 HHT23、HHT224 和 HHT226 等 3 个位点多态性低于 0.25。Shannon's 信息指数 (I)

的平均值为 0.957, 其中, 最小值为 0.300, 最大值为 1.894。哈温平衡 (HWE) 检验结果显示: HHT5、HHT10 等 8 个位点符合哈温平衡 (表 5)。

2.4 遗传多样性分析评价

在 3 个野生种群中检验了 27 个 SSR 位点在群体遗传分析中有效性与适用性, 结果 (表 6) 显示: 在六顺乡 (LSX)、江城 (JC) 以及景洪 (JH) 种群中, 观测等位基因 (N_a) 数量分别为 3.407、3.444、3.741; 有效等位基因 (N_e) 数量分别为 2.194、2.225、2.453。在 3 个种群中, 具有中高度多态性的 SSR 位点分别占比 85.19% (LSX)、



注：横坐标为 SSR 扩增片段大小 (nt)；纵坐标为荧光信号检测强度 (flu)

Notes: X-axis indicates the size of SSR loci (nt) and Y-axis indicates intensity of fluorescence signal (flu)

图 1 部分 SSR 位点毛细管电泳检测结果

Fig. 1 Fluorescence capillary electrophoresis test results of part of SSR loci

88.89% (JC)、85.19% (JH)。种群水平的期望杂合度 (H_e) 分别为 0.491 (LSX)、0.493 (JC)、0.528 (JH)，均值为 0.504；观测杂合度 (H_o) 分别为 0.472 (LSX)、0.416 (JC)、0.461 (JH)，均值为 0.450，结果表明 3 个刀状黑黄檀野生种群具有中等的遗传多样性。

F_{ST} 值可用于评价种群间遗传分化程度，3 个种群之间的遗传分化系数 (F_{ST}) 为 0.002~0.115，平均值为 0.034 (表 5)。AMOVA 分析结果显示：刀状黑黄檀种群内遗传变异和种群间遗传变异分别为 96.54% 和 3.46%，这表明刀状黑黄檀遗传变异主要来源于种群内 (表 7)。

3 讨论

3.1 基因组特征与精细组装策略

自 Galbraith 等^[17] 利用流式细胞术测定植物细胞核 DNA 含量以来，流式细胞术成为了测定植物基因组大小的标准方法。随着高通量测序技术的发展，基于 K-mer 分析的生物信息学估计基因组特征方法的出现，对于探索未知基因组物种提供了极大帮助，已经广泛应用于植物基因组大小评估和特征分析^[18]，并且 K-mer 分析相较于流式细胞术分析法的预测结果可信度更高、信息更丰富。以往普遍认为黄檀属植物的基因组可能比较大，在 DNA-C 值数据库中黄檀属植物基因组 (1C) 大小范围

为 1 430~1 928 Mb，平均值为 1 729 Mb^[19]。然而，在本研究中，基于 K-mer 分析结果估计刀状黑黄檀的基因组大小约为 706.92 Mb，基因组杂合度为 1.26%，重复序列比率为 55.74%。这一结果表明，其基因组属于高杂合、高重复的复杂基因组，基因组大小略高于降香黄檀 (*Dalbergia odorifera* T. Chen) 的 653.45 Mb^[20]，这可能是由于物种差异以及较高的基因组杂合度等因素共同导致的^[21]。刀状黑黄檀的基因组高杂合性可能与该物种复杂的分化过程有关，Vatanparast 等^[22] 基于分子证据认为，黄檀属起源于南美洲，但其只用了 ITS 片段，结果缺乏有力支撑；崔菲等^[23] 利用核基因和叶绿体片段构建了系统发育树，但基部类群也并未明确，暂未确定黄檀属的明确起源地区与分化趋势。然而，从基因组水平来解读物种系统演化的过程将是未来植物分类与系统进化研究发展的趋势。因此，对于刀状黑黄檀基因组特征初步分析和评价有利于制定合适的基因组精细组装策略，选用最新的三代 Pacbio 测序平台 HiFi 模式测序可能对于解决高杂合度在基因组组装时所产生的不利影响具有较大帮助，结合 Hi-C 技术和三代测序平台的全长转录组分析，将有助于基因组染色体水平的组装并获得更加精确的基因组注释信息。

3.2 基因组 SSR 标记特征

SSR 分子标记是保护遗传学中最常用的

表5 27个SSR位点多态性信息

Table 5 Characterization of locus-level polymorphic information at 27 SSR loci

位点 Locus	观测等位基因 N_a	有效等位基因 N_e	观测杂合度 H_o	期望杂合度 H_e	多态性信息含量 PIC	群体间遗传分化系数 F_{ST}	Shannon's信息指数 I	哈温平衡 HWE (p)
HHT5	3.000	2.032	0.367	0.508	0.390	0.115	0.743	0.058
HHT10	3.000	1.970	0.438	0.492	0.387	0.008	0.742	0.539
HHT14	4.000	2.627	0.477	0.619	0.559	0.056	1.116	0.000**
HHT23	3.000	1.282	0.170	0.220	0.198	0.031	0.400	0.012*
HHT45	3.000	2.272	0.533	0.560	0.489	0.004	0.936	0.290
HHT49	6.000	2.855	0.412	0.650	0.590	0.060	1.245	0.000**
HHT60	4.000	2.203	0.444	0.546	0.467	0.022	0.929	0.151
HHT62	2.000	1.979	0.602	0.495	0.372	0.021	0.688	0.042*
HHT85	9.000	4.014	0.719	0.751	0.716	0.027	1.697	0.000**
HHT97	6.000	4.442	0.722	0.775	0.742	0.025	1.618	0.000**
HHT117	3.000	1.795	0.233	0.443	0.397	0.033	0.780	0.000**
HHT119	6.000	3.311	0.833	0.698	0.651	0.045	1.382	0.000**
HHT121	10.000	5.757	0.573	0.826	0.803	0.082	1.894	0.000**
HHT124	6.000	2.558	0.344	0.609	0.562	0.030	1.166	0.000**
HHT126	8.000	1.933	0.202	0.483	0.467	0.034	1.124	0.000**
HHT160	8.000	1.913	0.467	0.477	0.438	0.024	1.001	0.999
HHR161	5.000	2.914	0.667	0.657	0.597	0.026	1.238	0.000**
HHT211	3.000	2.343	0.584	0.573	0.504	0.023	0.960	0.089
HHT216	3.000	2.047	0.534	0.511	0.411	0.022	0.796	0.222
HHT219	2.000	1.636	0.506	0.389	0.313	0.008	0.577	0.005**
HHT223	4.000	2.048	0.517	0.512	0.443	0.014	0.901	0.066
HHT224	2.000	1.290	0.146	0.225	0.200	0.003	0.385	0.001**
HHT226	2.000	1.193	0.111	0.162	0.149	0.051	0.300	0.003**
HHT229	2.000	1.600	0.295	0.375	0.305	0.103	0.562	0.047*
HHT230	4.000	2.684	0.409	0.627	0.550	0.023	1.068	0.000**
HHT247	2.000	1.468	0.398	0.319	0.268	0.002	0.499	0.020*
HHT255	4.000	2.556	0.477	0.609	0.542	0.011	1.089	0.005**
均值Mean	4.333	2.397	0.451	0.523	0.463	0.034	0.957	--

注:“**”表明显著性水平 ($*p < 0.05$, $**p < 0.01$)。

Notes:“**”Indicates significant levels ($*p < 0.05$, $**p < 0.01$).

DNA 标记之一,与 RFLP、AFLP 和 RAPD 相比,SSR 标记具有信息丰富、易于判定、稳定和共显性的特点^[24]。以往研究也发现,在遗传分析方面,基因组 SSR 相比于 EST-SSR 以及 SNP 具有更高的多态性^[25]。刀状黑黄檀基因组包含了大量的 SSR 分子标记位点,平均 5 442 bp 存在 1 个,且重复类型丰富,变异基序多样,以双、三核苷酸重复基序的占比较高。以往研究认为,重复单元长度变化与选择压力有关,重复单元长度越长,所受的

选择压力越大,拷贝数就越少,因此,基因组中长度较短的微卫星变异速率较快,而较长的重复单元变异速率较慢,所以,基因组中低级重复单元较多则表示该物种进化水平较高,而高级重复单元比例高的物种其进化时间短或变异频率较低^[26-27]。

3.3 多态性 SSR 验证与多样性评价

本研究基于 NGS 测序和初步组装数据发掘了刀状黑黄檀基因组 SSR 候选位点,并结合实验室筛选,验证了 27 个多态性较好,扩增稳定的

表 6 3 个刀状黑黄檀野生种群遗传多样性

Table 6 Genetic diversity of 3 wild populations of *D.cultrata* in China

群体 Population	观测等位基因 N_a	有效等位基因 N_e	Shannon's 信息指数 I	观测杂合度 H_o	期望杂合度 H_e
六顺乡 LSX	3.407	2.194	0.859	0.472	0.491
江城 JC	3.444	2.225	0.862	0.416	0.493
景洪 JH	3.741	2.453	0.949	0.461	0.528
均值 Mean	3.531	2.291	0.890	0.450	0.504

表 7 刀状黑黄檀种群间和种群内 AMOVA 分析

Table 7 AMOVA analysis among and within *D.cultrata* populations

变异来源 Source of variation	自由度 Degree of freedom	方差和 Sum of squares	变异组分 Variance component	占变异的比例 Proportion in total variation/%
种群间 Among population	2	47.522	23.761	3.46
种群内 Within population	177	1 263.417	14.317	96.54
总计 Total	179			100.00

SSR 位点。位点水平的观测等位基因数量 (N_a) 变异范围为 2~10 个, 均值为 4.333, 这也表明筛选验证的位点具有较好的等位基因变异, 其中, 超过 80% 以上的 SSR 位点具有中高度的多态性信息含量。研究发现, 27 个 SSR 位点中仅有 8 个位点符合哈温平衡, 类似的情况在生态位较为接近的降香黄檀种群中也有所体现, 同样具有一定比率的 SSR 位点不符合哈温平衡^[28]。符合哈温平衡需要满足随机交配且足够大的种群数量, 而在资源调查中发现, 刀状黑黄檀的野生种群均为零星分布, 且遭遇了严重的生境片段化, 几乎没有连续分布的刀状黑黄檀纯林^[4], 因此, 种群有效个体数量偏低, 非随机交配等因素可能是造成 SSR 标记位点偏离

哈温平衡的主要原因。

3.4 黄檀属重要树种遗传多样性比较

以 3 个野生种群为对象, 对刀状黑黄檀遗传多样性进行了初步分析与评价, 结果显示 3 个种群的期望杂合度 (H_e) 均值为 0.504, 明显低于异交植物的平均值 0.65^[29], 通过与其他黄檀属内植物遗传多样性比较发现 (表 8), 3 个刀状黑黄檀野生种群的期望杂合度明显高于海南分布的降香黄檀 ($H_e = 0.370$)^[28], 但低于泰国、老挝等地分布的交趾黄檀 (*Dalbergia cochinchinensis* Pierre) ($H_e = 0.550$ & $uH_e = 0.570$) 以及柬埔寨和越南等地分布的奥氏黄檀 (*Dalbergia oliveri* Gamble ex Prain) ($H_e = 0.730$)^[30-31] 和南美洲巴西热带地区分布的

表 8 黄檀属不同物种遗传多样性比较

Table 8 Comparative analysis of genetic diversity in *Dalbergia* genus

物种名 Species name	拉丁名 Scientific name	分布区域 Distribution	种群数量 No. of populations	样本数量 No. of individuals	标记类型 Type of marker	群体遗传参数 Genetic Parameters	参考文献 Reference
刀状黑黄檀	<i>Dalbergia cultrata</i>	中国云南等, 热带	3	90	SSR	$H_e = 0.504$	本研究
降香黄檀	<i>Dalbergia odorifera</i>	中国海南, 热带	7	42	SSR	$H_e = 0.370$	[28]
交趾黄檀	<i>Dalbergia cochinchinensis</i>	泰国、老挝、柬埔寨等, 热带	26	677	SSR	$H_e = 0.550$ $uH_e = 0.570$	[30] [31]
奥氏黄檀	<i>Dalbergia oliveri</i>	老挝、柬埔寨和越南等, 热带	23	616	SSR	$H_e = 0.730$	[30]
巴西黑黄檀	<i>Dalbergia nigra</i>	巴西, 热带	3 1	235 107	SSR	$H_e = 0.735$ $H_e = 0.682$	[32] [33]
黄檀	<i>Dalbergia hupeana</i> Hance	中国四川和陕西, 亚热带	3	90	SSR	$H_e = 0.659 \sim 0.740$, $uH_e = 0.670 \sim 0.752$	[34]
多体蕊黄檀	<i>Dalbergia polyadelpa</i> Prain	中国云南, 热带	1	29	SSR	$H_e = 0.512$, $uH_e = 0.522$	[34]
南岭黄檀	<i>Dalbergia balansae</i> Prain	中国云南, 热带	1	29	SSR	$H_e = 0.635$, $uH_e = 0.648$	[34]

巴西黑黄檀 (*Dalbergia nigra* (Vell.) Benth.) ($H_e = 0.735$ & $H_e = 0.682$) [32-33]。比较结果表明:中国野生分布的刀状黑黄檀种群遗传多样性处于较低的水平,这也暗示中国分布的刀状黑黄檀野生资源遭受了明显的遗传侵蚀,这可能是由于过去的过度砍伐与土地利用转变等因素造成的。在 IUCN 红色名录中,交趾黄檀 (VU, A1cd) 和降香黄檀 (VU, A1d) 濒危等级都高于刀状黑黄檀 (NT 级别),在中国受威胁高等植物名录中,刀状黑黄檀被列为 VU 级,从野生种群遗传多样性水平角度而言,本研究结果也支持将刀状黑黄檀的濒危等级提高到与交趾黄檀和降香黄檀相同的级别。尽管刀状黑黄檀已经被列入国家二级保护植物,但在野外调查研究中发现,土地利用转变导致的生境片段化也可能影响其野生种群的稳定性及限制种群间的基因交流与种群内的繁殖更新^[3]。因此,今后的工作应将全面深入开展遗传多样性评估、繁殖生物学及景观遗传学等方面的研究作为重点方向,用于制定科学的资源保护与种群恢复策略。

4 结论

本研究利用 NGS 技术对刀状黑黄檀基因组特征进行了分析,结果表明,其基因组大小约为 706.92 Mb,杂合度为 1.26%,属于高杂合的复杂基因组,提出了基于 PacBio 三代测序和 Hi-C 辅助组装的精细测序组装策略。通过实验发掘并验证了 27 个扩增条带清晰、多态性良好的 SSR 位点,并初步评估了 3 个野生刀状黑黄檀种群的遗传多样性,结果表明其具有中等水平的遗传多样性和较低的遗传分化程度。

参考文献:

- [1] 李世晋. 亚洲黄檀[M]. 北京: 科学出版社, 2017: 60-61.
- [2] 王磊, 刘久东, 耿云芬, 等. 黑黄檀的植物学特性及资源培育技术概况[J]. 林业调查规划, 2016, 41(4): 75-79.
- [3] Liu Y, Huang P, Lin F R, et al. Maxent modelling for predicting the potential distribution of a near threatened rosewood species (*Dalbergia cultrata* Graham ex Benth)[J]. Ecological Engineering, 2019, 141: 105612.
- [4] 张辉, 王立新, 付强, 等. 澜沧江中下游黑黄檀种群结构及保护对策[J]. 西南林学院学报, 2010, 30(1): 25-28.
- [5] 邱琼, 金美英, 杨德军, 等. 黑黄檀容器育苗基质筛选试验[J]. 湖南林业科技, 2015, 42(1): 19-22.
- [6] Donnelly D M X, O'Reilly J, Thompson J. Neoflavanoids of *Dalbergia cultrata*[J]. Photochemistry, 1972, 11(2): 823-826.
- [7] 覃海宁, 杨永, 董仕勇, 等. 中国高等植物受威胁物种名录[J]. 生物多样性, 2017, 25(7): 696-744.
- [8] Sun R X, Lin F R, Huang P, et al. Moderate genetic diversity and genetic differentiation in the relict tree *Liquidambar formosana* Hance revealed by genic simple sequence repeat markers[J]. Frontiers in Plant Science, 2016, 7: 1411.
- [9] Wang R K, Fan J S, Chang P, et al. Genome survey sequencing of *Acer truncatum* Bunge to identify genomic information, simple sequence repeat (SSR) markers and complete chloroplast genome[J]. Forests, 2019, 10(2): 87.
- [10] Wang C R, Yan H D, Li J, et al. Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum 'Zise') and identification of its SSR markers[J]. Molecular Breeding, 2018, 38(7): 22-31.
- [11] Motalebipour E Z, Kafkas S, Khodaeiaminjan M, et al. Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species[J]. BMC Genomics, 2016, 17(1): 998.
- [12] Doyle J J T, Doyle J L. Isolation of plant DNA from fresh tissue[J]. Focus, 1990, 12(1): 13-15.
- [13] Vurture G W, Sedlazeck F J, Nattestad M, et al. GenomeScope: Fast reference-free genome profiling from short reads[J]. Bioinformatics, 2017, 33(14): 2202-2204.
- [14] Luo R B, Liu B H, Xie Y L, et al. SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler[J]. GigaScience, 2012, 1(1): 11-16.
- [15] Peakall R, Smouse P E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update[J]. Bioinformatics, 2012, 28(19): 2537-2539.
- [16] Kalinowski S T, Taper M L, and Marshall T C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity[J]. Molecular Ecology, 2007, 16(5): 1099-1106.
- [17] Galbraith D W, Harkins K R, Maddox J M, et al. Rapid flow cytometric analysis of the cell cycle in intact plant tissues[J]. Science, 1983, 220(4601): 1049-1051.
- [18] 朱礼明, 黎梦娟, 张景波, 等. 西伯利亚白刺基因组信息初探[J]. 林业科学研究, 2020, 33(1): 144-151.
- [19] Leitch I J, Johnston E, Pellicer J, et al. Angiosperm DNA C-values database[DB]. 2019, Release 7.1, <https://cvalues.science.kew.org>.
- [20] Hong Z, Li J, Liu X J, et al. The chromosome-level draft genome of *Dalbergia odorifera*[J]. GigaScience, 2020, 9(8): 1-8.
- [21] Veleba A, Bureš P, Adamec L, et al. Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae[J]. New Phytologist, 2014, 203(1): 22-28.
- [22] Vatanparast M, Klitgård B B, Adema F A C B, et al. First mo-

- lecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography[J]. South African Journal of Botany, 2013, 89: 143-149.
- [23] 崔 菲. 黄檀属(豆科)的分子系统学研究[D]. 北京: 中国科学院大学, 2014: 52-68.
- [24] Song Q J, Shi J R, Singh S, *et al.* Development and mapping of microsatellite (SSR) markers in wheat[J]. Theor Appl Genet, 2005, 110(3): 550-560.
- [25] Lu X, Wang H X, Dai P, *et al.* Characterization of EST-SSR and genomic-SSR markers in the clam, *Meretrix meretrix*[J]. Conservation Genetics Resources, 2011, 3(4): 655-658.
- [26] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: Survey and analysis[J]. Genome Research, 2000, 10(7): 967-981.
- [27] Cardle L, Ramsay L, Milbourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants[J]. Genetics, 2000, 156(2): 847-854.
- [28] Liu F M, Hong Z, Xu D P, *et al.* Genetic diversity of the endangered *Dalbergia odorifera* revealed by SSR markers[J]. Forests, 2019, 10: 225.
- [29] Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants[J]. Molecular Ecology, 2004, 13(5): 1143-1155.
- [30] Hartvig I, So T, Changtragoon S, *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits[J]. Ecology and Evolution, 2018, 8(1): 530-545.
- [31] Hartvig I, So T, Changtragoon S, *et al.* Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): Recommendations for management and sustainable use[J]. Conservation Genetics, 2020, 21(4): 677-692.
- [32] Resende L C, Ribeiro R A, Lovato M B. Diversity and genetic connectivity among populations of a threatened tree (*Dalbergia nigra*) in a recently fragmented landscape of the Brazilian Atlantic Forest[J]. Genetica, 2011, 139(9): 1159-1168.
- [33] Leite F A B, Brandão R L, Buzatti R S D O, *et al.* Fine-scale genetic structure of the threatened rosewood *Dalbergia nigra* from the Atlantic Forest: Comparing saplings versus adults and small fragment versus continuous forest[J]. Tree Genetics and Genomes, 2014, 10(2): 307-316.
- [34] Li C H, Zheng Y Q, Liu Y, *et al.* Development of genomic SSR for the subtropical hardwood tree *Dalbergia hupeana* and assessment of their transferability to other related species[J]. Forests, 2021, 12(6): 804.

Genomic Characteristics and Population Genetic Variation of *Dalbergia cultrata* Graham ex Benth in China

LIU Yu^{1,2}, ZHENG Yong-qj¹, LI Chang-hong¹, LIN Fu-rong¹, HUANG Ping¹

(1. State Key Laboratory of Tree Genetics and Breeding; Research Institute of Forestry, Chinese Academy of Forestry; Laboratory of Forest Silviculture and Tree Cultivation, Beijing 100091, China; 2. Wenzhou Key Laboratory of Resource Plant Innovation and Utilization, Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou 325005, Zhejiang, China)

Abstract: [Objective] To analyze the characterization of *Dalbergia cultrata* genome, and develop a set of high polymorphic SSR molecular markers for assessment of genetic variation of wild population. [Method] Next Generation Sequencing (NGS) was performed to evaluate the characteristic of genome, and MISA was used to mine candidate genomic SSR loci from the assembled data. The flanking primers of candidate SSR loci were designed by Primer Premier v 5.0, and screened by polyacrylamide gel electrophoresis (PAGE). The polymorphic and population genetic analysis was performed by capillary electrophoresis (CE). [Result] The genome size of *D. cultrata* was about 706.92 Mb and the heterozygosity was about 1.26%. The rate of repetitive sequence was 55.74%. A total of 27 polymorphic SSR loci were screened and 117 alleles were amplified on 27 SSR loci. The value of the polymorphism information content (PIC) varied from 0.149 to 0.803. The population genetic analysis showed the mean expected heterozygosity (H_e) and the coefficient of genetic differentiation (F_{ST}) was 0.504 and 0.034, respectively. AMOVA analysis revealed the genetic variation within the populations (96.54%) was much higher than that among the populations (3.46%). [Conclusion] The genome of *D. cultrata* belongs to high heterozygosity and highly repetitive complex genome, which provides important basic data to make a fine assembly strategy. A set of novel genomic SSR markers shows good polymorphism, and stability. Moderate genetic diversity and low genetic differentiation of wild populations of *D. cultrata* are revealed by SSR markers. This study would facilitate to conserve and assess wild germplasms of *D. cultrata*.

Keywords: *Dalbergia cultrata*; Next Generation Sequencing; genomic SSR; polymorphism; genetic diversity.

(责任编辑:张研)