

## 基于简化基因组测序分析云南小桐子遗传多样性

郑周宜<sup>1</sup>, 刘雨婷<sup>1</sup>, 陈 春<sup>1</sup>, 王莎莎<sup>1,3,4\*</sup>, 唐 唯<sup>1,2,3\*</sup>

1. 云南师范大学生命科学学院, 云南昆明 650500; 2. 云南师范大学马铃薯科学研究院, 云南昆明 650500; 3. 生物能源持续开发利用教育部工程研究中心, 云南昆明 650500; 4. 云南省生物质能与环境生物技术重点实验室, 云南昆明 650500

**摘要:** 为深入了解云南省热带能源植物小桐子的遗传多样性和亲缘关系, 本研究以云南省 9 个种源地的小桐子群体为材料展开联合分析。为了研究群体多态性, 以 MISA 程序开发了 8 对 SSR 分子标记, PCR 扩增后选取多态性较好的 4 对引物以 9 个居群 95 个材料 DNA 进行了扩增, 发现 JC01 引物多态性最好, 达 0.711。聚类分析表明保山地区和西双版纳地区可能作为云南省小桐子的亲本或起源, 各居群间亲缘关系和地理位置有一定相关性。本研究对各采集地种子和幼苗时期生物学性状的综合考察发现地下部分鲜重 (UFW) 性状群体变异系数最大, 分离明显。选择 UFW 极端差异的 2 个居群 (版纳勐海和楚雄元谋) 以 BSA 分离法进行 RAD 简化基因组测序, 基因组组装分别得到 2783 和 3033 条 contigs, 利用 GenomeScope 对基因组组装数据分析发现 2 份材料均为二倍体, GWAS 分析在全基因组水平挖掘到有效 SNPs 22 756 个, 其中 UFW 高频 SNPs 集中于 9 号染色体, 有 3 个候选蛋白与 UFW 性状连锁。研究结果表明云南省不同地区小桐子居群具有较高的遗传多样性, 初步明确了云南本地小桐子种质的独特性。

**关键词:** 小桐子; 遗传多样性; 分子标记; 全基因组关联分析; 单核苷酸多态性

中图分类号: S565.9 文献标识码: A

## RAD-based Genomic Sequencing Analysis of *Jatropha curcas* L. in Yunnan to Reveal Its Genetic Diversity

ZHENG Zhouyi<sup>1</sup>, LIU Yuting<sup>1</sup>, CHEN Chun<sup>1</sup>, WANG Shasha<sup>1,3,4\*</sup>, TANG Wei<sup>1,2,3\*</sup>

1. School of Life Sciences, Yunnan Normal University, Kunming, Yunnan 650500, China; 2. Joint Academy of Potato Science, Yunnan Normal University, Kunming, Yunnan 650500, China; 3. Engineering Research Center of Sustainable Development and Utilization of Biomass Energy, Ministry of Education, Kunming, Yunnan 650500, China; 4. Key Lab of Yunnan Province for Biomass Energy and Environmental Biotechnology, Kunming, Yunnan 650500, China

**Abstract:** To gain insight into the genetic diversity and relationship of the tropic bioenergy plant physic nut (*Jatropha curcas* L.) in Yunnan, China, joint analysis of agronomic and molecular characters was performed in the present study, using different accessions collected from nine sources / geographical provenances in Yunnan. To investigate the population polymorphism, eight pairs of SSR molecular markers were developed by MISA program, and four of them with higher polymorphism were selected for PCR amplification using 95 DNA samples collected from nine populations. The results indicated that JC01 exhibited the best polymorphism (0.711). UPGMA clustering analysis suggested that Baoshan and Xishuangbanna might serve as the parents or origins of physic nut in Yunnan, which was partially related to the geographical origin of populations. Comprehensive investigations revealed among all examined biological traits of seeds or seedlings, the underground fresh weight (UFW) exhibited the highest coefficient of variation, with obvious separation among populations. Two individual seedlings with extreme variance in UFW (BNMH and CXYM) were selected for RAD-based genomic sequencing, while 2783 and 3033 contigs were assembled. Analysis of genome assembly data using GenomeScope revealed that both materials were diploid, and a total 22 756 valid SNPs were detected at the whole genome level using genome-wide association study (GWAS). High frequency-SNPs associated with UFW pref-

收稿日期 2022-06-07; 修回日期 2022-07-06

基金项目 国家自然科学基金项目 (No. 31860062, No. 31460182)。

作者简介 郑周宜 (2001—), 女, 本科生, 研究方向: 小桐子遗传育种。\*通信作者 (Corresponding author): 王莎莎 (WANG Shasha), E-mail: sasha\_wng@163.com; 唐 唯 (TANG Wei), E-mail: 4311@ynnu.edu.cn。

erably located on LG9, and three candidate proteins were linked to the UFW trait. The results indicated that the physic nut germplasm collected from different regions of Yunnan exhibited a high level of genetic diversity and uniqueness, which should facilitate further studies on its molecular breeding.

**Keywords:** physic nut (*Jatropha curcas* L.); genetic diversity; molecular markers; GWAS; SNP

**DOI:** 10.3969/j.issn.1000-2561.2023.03.008

小桐子 (*Jatropha curcas* L.) 属大戟科麻疯树属的多年生木本油料植物, 原产美洲, 现广泛分布于全世界热带及亚热带地区, 在我国尤以云南分布面积最广、资源数量最大<sup>[1-2]</sup>。小桐子具有生物产量大、适应范围广、耐贫瘠等特性, 是维持生态<sup>[3]</sup>、医药开发、生物防治及生产有机肥料的理想资源<sup>[4-5]</sup>。由于其种子含油率高达 60% 以上, 也被认为是具有巨大开发潜力的多用途能源植物树种<sup>[6]</sup>。尽管近年来国内外有关小桐子的研究报告数量大增, 但大多集中在药理毒理活性及新药开发<sup>[4-5]</sup>、生理适应机制<sup>[7-8]</sup>、生物柴油产业化<sup>[6]</sup>、经济用途开发<sup>[3]</sup>等方面, 关于其种质资源的遗传多样性研究起步相对较晚。另一方面, 种质资源不清、品种良莠不齐、基因资源库保存不完善等问题已成为限制国内小桐子产业开发的重要瓶颈<sup>[9]</sup>。

迄今为止, 来自中国、巴西、印度、墨西哥等多国的研究人员已先后采用 RAPD、AFLP 等多种分子标记对国内外小桐子种质进行了遗传多样性的鉴定和评价<sup>[10-11]</sup>, 发现我国范围内收集的小桐子种质资源遗传背景虽普遍低于它国 (南美、东南亚地区等)<sup>[12-13]</sup>, 但云南地区的小桐子居群具有较为丰富的遗传多样性, 且其居群内的遗传分化高于居群间<sup>[14-16]</sup>。上述研究在一定程度上印证了向振勇等<sup>[17]</sup>提出的云南小桐子种质来源较为复杂的推测。

基于上述溯源, 为了进一步了解近年云南地区小桐子种群结构特点, 本研究首先从种子和幼苗性状特点入手, 结合简化基因组测序和分子标记技术解析云南各地区小桐子的遗传多样性特点; 利用 GWAS 分析方法, 挖掘地下部分鲜重这一重要性状差异基因及地区独特性基因, 预期研究结果可为小桐子种质资源的保护利用、遗传改良以及杂交育种亲本选配提供重要遗传基础。

## 1 材料与方法

### 1.1 材料

供试材料来自云南省不同地区的 9 个居群, 种源地信息见表 1。每个地区根据植株分布情况

随机选取 50 个具有代表性的样株, 每个样株至少间隔 5 m 以上, 每个样株分别取 3 颗种子, 编号整理。

### 1.2 方法

1.2.1 幼苗培养 在各种源地的种子材料中挑选颗粒饱满、大小均一的种子 50 粒, 清洗净后以 1% CuSO<sub>4</sub> 溶液浸泡消毒 20 min, 经蒸馏水充分漂洗、浸泡、吸涨后于 26℃ 萌发。将萌发 14 d 后的种子移栽至装有等量营养土的花盆中, 置于人工气候箱中继续培养 14 d, 用于幼苗生物学性状的测定。

1.2.2 性状测定 (1) 百粒重及种子形状测定。在各种源地随机选取种子, 取 3 个重复, 每个重复 30 粒, 以电子天平分别称重 (精确至 0.01 g), 计算种子百粒重。以游标卡尺测定长径、宽径及侧径 (精确至 0.01 mm)。

(2) 生长指标及生物量测定。每个种源随机抽样 10 株苗, 于移栽后 14 d 进行株高测定 (精确至 0.01 cm)。生物量测定采用全称重法。样株在流动自来水下充分洗净、吸干表面水分, 分别对地上部分、地下部分进行鲜重称量。

(3) 数据处理。以 SPSS 软件 (v26.0) 进行数据统计、方差分析、多重比较 (Duncan's 法或非参数检验, 以不同小写英文字母表示显著差异,  $P < 0.05$ ) 及相关性分析。

利用变异系数 (CV) 分析种子性状变异性, 计算公式为:  $CV = (S/X) \times 100\%$  ( $S$  为标准差,  $X$  为平均值)。计算根冠比 = 地下部分 (鲜重) / 地上部分 (鲜重)  $\times 100\%$ 。

1.2.3 RAD-seq 基因组测序及分子标记开发 (1) 基因组 DNA 的提取。综合考虑幼苗性状 (鲜重、根冠比) 及平均变异系数, 取差异显著的 2 个居群材料 (各 10 株) 为 BSA 分离群体, 其中版纳勐海 BNMH 为 BSA1, 楚雄元谋 CXYM 为 BSA2, 采用 CTAB 法提取基因组 DNA (以第 2 片真叶为材料), 经琼脂糖凝胶电泳 (0.8%~1.0%) 和超微量分光光度计检测质量、纯度和完整性。

(2) 基因组测序及分子标记挖掘。构建插入

片段 350 bp 的文库, 测序在 ILLUMINA Hiseq 2000 平台进行 (Novogenes Inc.), 测序读长为两端 150 bp, 测序序列用 GenomeScope 2.0<sup>[18]</sup>进行 K-mer 分析确定染色体倍性。SNP 标记挖掘由 CLC Genomics Workbench 12.0.1 (QIAGEN, Inc.) 完成, 工具为 Basic Variation Detection (参数默认)。

(3) SSR 引物设计。取上述 2 个居群的小桐子材料进行 RAD-seq 基因组测序 (5×覆盖度), 以测序基因组为基础, 利用基于 Perl 的 MISA 程序挖掘 SSR 引物<sup>[19]</sup>[(X)n; (XX)n-2; (XXX)n-3; (XXXX)n-4; (XXXXX)n-6; (XXXXXX)n-7, n=10], 挖掘到的 SSR 标记以软件 primer3 (<https://primer3.org/>) 设计引物。引物位于 SSR 标记 2 个侧翼, 长度在 20~22 bp 之间, 正向引物的 5'端加上荧光标记 FAM, PCR 目标产物预期大小在 50~150 bp 范围内。引物用 QC filter 进行筛选, 去掉纯合位点, 引物由昆明擎科生物科技有限公司合成。

(4) 分子标记的电泳检测 (琼脂糖凝胶、毛细管电泳) 及遗传多样性分析。对检测和纯化好的不同种群小桐子基因组 DNA 进行筛选, 配制反应体系后进行 PCR 扩增, Tm 以引物合成报告单为准。对琼脂糖凝胶电泳初步确定扩增效果理想的引物进行毛细管电泳分析 (ABI 3500), 用 GeneMapper 5.0 (<https://tools.thermofisher.com/content/sfs/manuals/4476603A.pdf>) 读取条带, 用 MEGA 5.0<sup>[20]</sup>构建 UPGMA 系统发育树, 用 PopGene 1.31 (<https://sites.ualberta.ca/~fyeh/popgene.pdf>) 进行遗传多样性计算。

(5) 基于重要性状差异的 GWAS 分析。以地下部分鲜重统计数据为性状表型, 以全基因组 SNPs 为基因型, 使用 R 包 GAPIT 混合线性模型 (MLM) 进行 GWAS 分析 (全基因组关联分析, Genome Wide Association Study), 以 CMplot 包进行 GWAS 结果输出, 分别生成 SNPs 密度分布图和柱状曼哈顿图。

(6) 基因组注释及比较基因组分析。对上述重测序 2 份材料的 clean data 进行全基因组比对 (CLC Genomics Workbench 12.0.1) 得到共线性结果。对 2 个基因组进行 de novo 组装 (SOAPdenovo2<sup>[21]</sup>), 得到的组装序列用 Augustus (<http://bioinf.uni-greifswald.de/augustus/>) 进行 CDS 预测。对预测到的基因以 NCBI 中 protein databank 和 swiss-prot protein databank 进行同源注释和比较

分析, 找出同时具有以下 2 个特征的特异蛋白: ①仅在上述 2 份材料中存在的; ②与地下部分鲜重性状紧密连锁。

## 2 结果与分析

### 2.1 云南地区不同种源小桐子种子性状及其变异分析

百粒重测量结果 (表 1) 显示其变幅在 46.06~80.40 g, 不同种源的差异可达显著差异水平 ( $P < 0.05$ )。长径、宽径、侧径变幅依次为 13.00~19.50、7.50~12.00、5.20~10.10 mm, 最大平均值均为双江大文乡种源, 最小平均值为丽江华坪种源。对种子性状的变异性分析发现, 供试种子变异系数差别较为明显, 其中百粒重变异系数最大 (13.93%), 长径、宽径与侧径的变异系数依次为 3.04%、5.34%和 6.20%。

### 2.2 云南地区不同种源小桐子幼苗性状及其变异分析

从表 1 可以看出, 株高性状表现最好的是临沧耿马种源, 平均苗高为 11.46 cm, 版纳勐海种源具有相对最高的生物量积累 (全株鲜重 6.60 g), 丽江华坪、楚雄禄丰和楚雄元谋种源在苗高和生物量积累上均表现最差, 尽管楚雄元谋种源的根冠比表现最佳 (0.25)。方差分析表明不同种源的苗高、地上部分鲜重、全株鲜重及根冠比均存在显著差异。不同种源地小桐子表型性状变异程度依次为: 地上部分鲜重 > 全株鲜重 > 地下部分鲜重 > 根冠比 > 株高 > 百粒重 > 种子侧径 > 种子宽径 > 种子长径 (表 1), 尤以生物量积累的性状变异系数相对较高, 暗示这些表型性状存在遗传不稳定性, 易受环境压力的选择。

### 2.3 RAD-seq 基因组测序及结果分析

测序结果显示 BNMH (BSA1) 及 CXYM (BSA2) 的染色体为 2 倍体 (图 1A), 得到具有 EcoR I 酶切位点的 Reads 数分别为 64 734 412 条和 60 187 766 条。平均 Q<sub>20</sub> 为 98.6%, 平均 Q<sub>30</sub> 为 93.9%, 对 reads 进行有参组装 (参考基因组 RJC1) 后分别得到 2783 和 3033 条高质量的 contigs (表 2)。

### 2.4 SSR 引物开发及遗传多样性分析

利用 MISA 随机开发出 8 对 SSR 引物, PCR 扩增后利用琼脂糖凝胶电泳确认有效扩增引物为 4 对 (表 3); 以 4 对引物、9 个居群共 95 个 DNA

表 1 不同地理种源小桐子种子性状及变异比较  
Tab. 1 Comparison of variations among physic nut seeds from various origins geography

种源编号 Source number	产地 Place of origin	百粒重 100-seed weight/g	长径 Length to diameter/mm	宽径 Radial width/mm	侧径 Lateral diameter/mm	地下鲜重 Under-ground fresh weight/g	地上鲜重 Above-ground fresh weight/g	全株鲜重 Fresh weight of whole plant/g	根冠比 Root/shoot ratio	株高 Plant height/cm
LCGM	临沧耿马	61.79±3.96 <sup>ab</sup>	16.75±0.66 <sup>b</sup>	9.37±0.40 <sup>c</sup>	7.31±0.32 <sup>bc</sup>	0.85±0.25 <sup>a</sup>	3.46±0.30 <sup>ab</sup>	4.32±0.49 <sup>ab</sup>	0.24±0.06 <sup>a</sup>	11.46±1.14 <sup>a</sup>
PRJG	普洱景谷*	79.82±0.51 <sup>a</sup>	17.14±1.04 <sup>ab</sup>	10.60±0.87 <sup>ab</sup>	7.81±0.76 <sup>a</sup>					
HHYY	红河元阳	51.43±0.44 <sup>b</sup>	16.62±0.42 <sup>bc</sup>	10.35±0.40 <sup>b</sup>	7.34±0.41 <sup>bc</sup>	0.60±0.35 <sup>a</sup>	3.89±1.32 <sup>ab</sup>	4.49±1.63 <sup>ab</sup>	0.15±0.05 <sup>b</sup>	7.59±0.97 <sup>b</sup>
BSWY	保山瓦窑	65.49±1.07 <sup>ab</sup>	16.78±0.83 <sup>b</sup>	10.68±0.59 <sup>ab</sup>	7.32±0.48 <sup>bc</sup>	0.66±0.21 <sup>a</sup>	3.51±1.04 <sup>ab</sup>	4.17±1.17 <sup>ab</sup>	0.19±0.05 <sup>ab</sup>	8.41±0.38 <sup>ab</sup>
BNMH	版纳勐海	57.02±2.60 <sup>ab</sup>	16.55±0.77 <sup>bc</sup>	10.65±0.61 <sup>ab</sup>	7.49±0.63 <sup>b</sup>	1.02±0.79 <sup>a</sup>	5.58±3.17 <sup>ab</sup>	6.60±3.90 <sup>ab</sup>	0.17±0.06 <sup>b</sup>	8.42±1.57 <sup>ab</sup>
LCSJ	临沧双江	69.50±0.69 <sup>ab</sup>	16.69±0.57 <sup>b</sup>	10.77±0.40 <sup>ab</sup>	7.39±0.37 <sup>bc</sup>	0.80±0.25 <sup>a</sup>	5.15±0.41 <sup>a</sup>	5.96±0.56 <sup>a</sup>	0.16±0.04 <sup>b</sup>	7.40±0.40 <sup>b</sup>
CXLF	楚雄禄丰	54.27±1.66 <sup>ab</sup>	15.91±1.07 <sup>c</sup>	10.09±0.85 <sup>b</sup>	7.23±0.75 <sup>bc</sup>	0.77±0.18 <sup>a</sup>	3.18±0.41 <sup>b</sup>	3.96±0.48 <sup>b</sup>	0.25±0.06 <sup>a</sup>	8.31±1.57 <sup>ab</sup>
LJHP	丽江华坪	60.00±0.37 <sup>ab</sup>	16.35±0.68 <sup>bc</sup>	9.34±0.55 <sup>c</sup>	7.01±0.49 <sup>c</sup>	0.56±0.20 <sup>ab</sup>	2.74±0.92 <sup>b</sup>	3.30±0.98 <sup>b</sup>	0.22±0.12 <sup>ab</sup>	7.03±0.55 <sup>b</sup>
CXYM	楚雄元谋	53.29±6.77 <sup>ab</sup>	16.21±1.03 <sup>bc</sup>	10.29±0.69 <sup>b</sup>	8.22±0.66 <sup>a</sup>	0.52±0.22 <sup>b</sup>	2.85±1.46 <sup>b</sup>	3.38±1.63 <sup>b</sup>	0.22±0.12 <sup>ab</sup>	6.79±1.09 <sup>b</sup>
SJDWX	双江大文乡	63.76±5.38 <sup>ab</sup>	17.75±0.99 <sup>a</sup>	10.86±0.51 <sup>a</sup>	8.49±0.59 <sup>a</sup>	0.71±0.26 <sup>a</sup>	3.62±0.22 <sup>ab</sup>	4.33±0.31 <sup>ab</sup>	0.20±0.08 <sup>ab</sup>	8.68±0.25 <sup>ab</sup>
平均值		61.64	16.68	10.30	7.56	0.72	3.78	4.50	0.20	8.23
标准差		8.59	0.51	0.55	0.47	0.16	0.98	1.10	0.04	1.38
变异系数/%		13.93	3.04	5.34	6.20	21.90	25.87	24.49	18.41	16.81

注：\*表示该种源地种子经多次萌发均引起幼苗腐烂、未获得健康植株用于后续实验，推测是种子携带真菌所致；同列数据后不同小写字母表示处理间差异显著 ( $P<0.05$ )。

Note: \* means the seeds of this provenance after multiple germination caused seedling decay and did not obtain healthy plants for subsequent experiments, which was speculated to be caused by seed carrying fungus; Different lowercase letters after the same column of data indicate significant difference ( $P<0.05$ ).

样本进行毛细管电泳检测，结果发现引物多态性扩增条带最多为 6 条，最少为 2 条，各居群间利用 4 个毛细管电泳引物扩增遗传多样性水平较高（表 4）。以每个地区所有个体多态性条带为居群进行 UPGMA 聚类分析，结果表明保山地区和西双版纳地区可能作为云南省小桐子的亲本或起源，各居群间亲缘关系和地理位置有一定相关性（图 1B）。

### 2.5 基于地下部分鲜重差异的 GWAS 分析

基于在种子及幼苗性状调查分析中发现，地

下部分鲜重（underground fresh weight, UFW）为 9 个居群间差异最大，因此本研究针对 UFW，首先在全基因组水平挖掘到相关 InDel 标记 8182 个，SNPs 标记 82 155 个。因小桐子为杂合二倍体，因此经纯合和等位基因频率（ $\geq 50\%$ ）过滤后，共挖掘到有效 SNPs 22 756 个。GWAS 分析进一步发现，SNPs 分布于 11 对染色体（图 2A），UFW 高频 SNPs 集中于 9 号染色体（图 2B）。

### 2.6 注释及比较基因组分析

本研究通过对基因组测序材料 BNMH（版纳

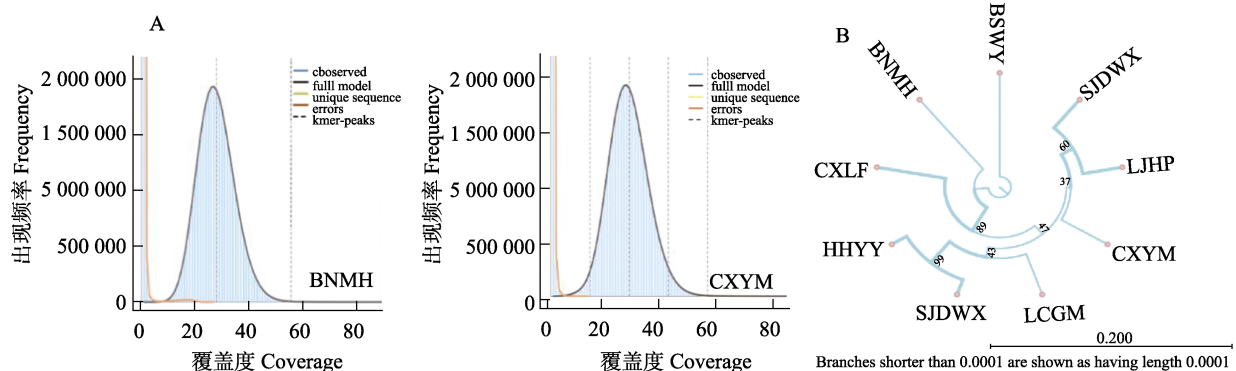


图 1 染色体倍性、基因组重复序列 (A) 与 UPGMA 分析 (B)

Fig. 1 Genome scope analysis of ploidy and genome repeat length (A) and UPGMA cluster analysis of 9 populations of physic nut (B)

表 2 小桐子基因组的组装特征

Tab. 2 Summary of genome assemble features among three cultivars of *J. curcas* L.

品种 Cultivar	登录号 Accession No.	重叠群数量 No. of contigs	G+C 含量 G+C content/%
BNMH	-	2783	35.08
CXYM	-	3033	34.87
RJC1	GCA_014843425.1	3383	33.65

表 3 基于 SSR 标记的 PCR 引物序列信息

Tab. 3 Detailed information of primers used for PCR analysis

引物名称 Primer name	引物序列 (5'-3') Primer sequence (5'-3')	扩增片段范围 Amplified fragment range/bp
JacCH1F	GACAAGATACACTTGGACTAGAAAA	98~149
JacCH1R	CCCCTTGAACATATTTGAT	
JacCH2F	CACTTAAAAGCCGAGTACTC	110~178
JacCH2R	AAAATAGAAAATGTATATATATACG	
JacGE1F	ATGCATAAAAAACATTAATTAT	118~144
JacGE1R	TCCTACATCCTCATAGTGAC	
JacGE2F	TTCGTATATCCAACAACATC	130~188
JacGE2R	GATTTTGATTCTTTTGCTGG	

表 4 SSR 引物多态性指数

Tab. 4 Polymorphism of SSR primers

SSR 位点 SSR locus	杂合度 <i>He</i>	多态性指数 PIC
JC01*	0.628	0.711
JC07	0.104	0.255
JC11	0.002	0.105
JC17*	0.816	0.639
JC33*	0.712	0.592
JC41	0.290	0.132
JCB2	0.158	0.041
JCD7*	0.784	0.481

注: \*为毛细管电泳所用引物。

Note: \* primers selected to capillary electrophoresis.

勐海) 和 CXYM (楚雄元谋) 进行注释和功能预测, 分别得到 27 133 和 26 282 个候选功能蛋白。Blast2GO 分析对预测蛋白进行 GO 功能注释, 发

现在生物学途径 (biological process) 中, 获得注释条目最多的是蛋白修饰、跨膜运输和转录调节; 在分子功能 (molecular function) 和细胞组件 (cellular component) 中分别为催化活性、结合功能、膜整合组分和胞质 (图 3)。

为了初步明确云南本地小桐子的独特性, 进一步比较分析发现, 仅在 BNMH 和 CXYM 两份测序材料中存在的候选蛋白共 40 个, 其中未注释的功能蛋白有 16 个, 和 UFW 性状连锁的有 3 个 (表 5)。

### 3 讨论

小桐子 (*J. curcas* L.) 是一种在全世界 (亚) 热带地区广泛种植的多用途油料作物, 其作为一种新兴能源植物的开发潜力而引起了全世界的广

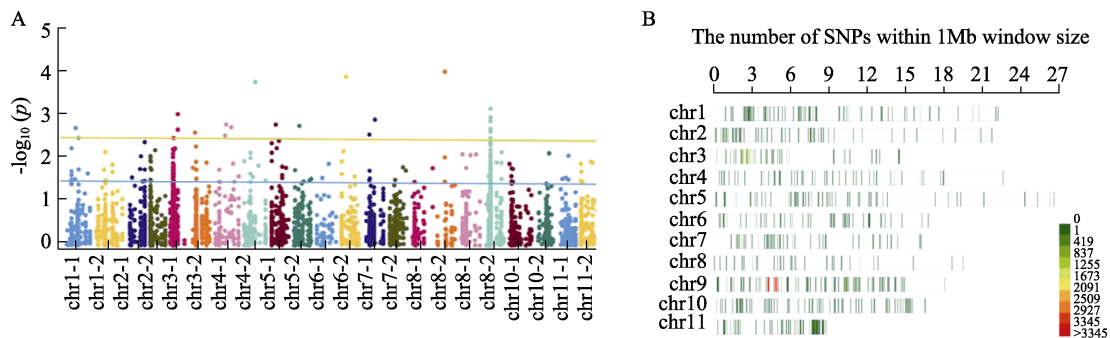


图 2 小桐子基因组测序中 UFW 性状的柱状曼哈顿图 (A) 及染色体内 SNP 密度分布 (B)

Fig. 2 Rectangular-Manhattan UFW trait plot (A) and SNPs density within 11 chromosomes (B) of *J. curcas* L.

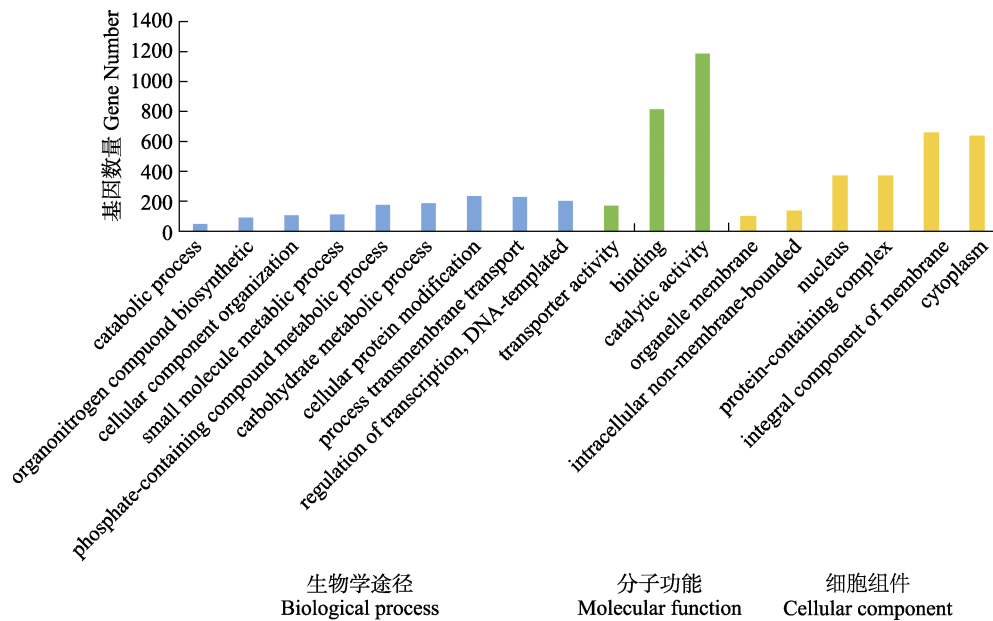


图 3 预测蛋白的 Blast2GO 分析  
Fig. 3 Blast2GO analysis of predicted proteins

表 5 BNMH 和 CXYM 中独特的预测蛋白信息  
Tab. 5 List of predicted unique proteins in BNMH and CXYM

蛋白序列号 Protein ID	目标蛋白数 HSPs	P	功能描述 Description
g12634.t1	4	0.42	Alpha-N-acetylgalactosaminidase NagBb from Bifidobacterium bifidum
g2785.t1	6	1.82	Chain a2, P10 [ <i>Paramecium bursaria</i> Chlorella virus 1]
g3265.t1	1	0.04	Chain B4, Ribosomal Protein L34 [ <i>Spinacia oleracea</i> ]
g6700.t1	84	0.00	Chain LC, Ribosomal protein L3 [ <i>Triticum aestivum</i> ]
g2971.t1	3	4.42×10 <sup>-4</sup>	Cryo-EM structure of a late human pre-40S ribosomal subunit - State C [ <i>Homo sapiens</i> ]
g7261.t1	1	7.46×10 <sup>-16</sup>	Crystal structure of Arabidopsis thaliana BIC2-CRY2 complex
g83.t1	9	0.21	Crystal structure of carbohydrate oxidase from Microdochium nivale
g7352.t1	2	5.29	Crystal Structure of Colicin E9 In Complex with Its Immunity Protein Im9 [ <i>Escherichia coli</i> ]
g983.t1	1	3.63	Crystal structure of diaminopimelate decarboxylase [ <i>Escherichia coli</i> ]
g13025.t1	2	4.40×10 <sup>-10</sup>	Crystal structure of human CPSF30 in complex with hFip1 [ <i>Homo sapiens</i> ]
g5788.t1	1	1.93×10 <sup>-4</sup>	Crystal structure of iron transporter VIT1 with zinc ions [ <i>Eucalyptus grandis</i> ]
g2802.t1	21	3.05	Crystal structure of P4397 complexed with c-di-GMP [ <i>Pseudomonas putida</i> KT2440]
g651.t1	7	1.11	Crystal Structure of Pyruvate Kinase from toxoplasma gondii
g726.t1	2	7.73	Crystal structure of tetR-family transcriptional regulator [ <i>Streptomyces coelicolor</i> A3(2)]
g3008.t1	1	8.40	Crystal structure of TLE1 N-terminal Q-domain residues 1-156 [ <i>Homo sapiens</i> ]
g5733.t1	3	1.71×10 <sup>-7</sup>	Crystal structure of UCH37-NFRKB Inhibited Deubiquitylating Complex [ <i>Homo sapiens</i> ]
g4966.t1	1	6.11	DHD127 [synthetic construct]
g615.t1	2	4.51	Glutamate-1-semialdehyde 2,1-Aminomutase from Thermosynechococcus elongatus [ <i>Synechococcus elongatus</i> ]
g10682.t1	49	1.71×10 <sup>-22</sup>	he catalytic core of Rad2 in complex with DNA substrate (complex II) [ <i>Saccharomyces cerevisiae</i> S288C]
g12671.t1	1	5.20	Solution structure of fibronectin type III domain of mouse hypothetical protein [ <i>Mus musculus</i> ]
g2630.t1	2	3.24	Structure of CBM32-3 from a family 31 glycoside hydrolase from Clostridium perfringens
g13597.t1	3	3.29×10 <sup>-55</sup>	Structure of the 70S chloroplast ribosome from spinach [ <i>Spinacia oleracea</i> ]
g10747.t1	95	9.90×10 <sup>-64</sup>	Structure of the photosynthetic complex I from <i>Thermosynechococcus elongatus</i>
g7098.t1	2	0.05	Tim8-Tim13 complex [ <i>Saccharomyces cerevisiae</i> ] complex [ <i>Saccharomyces cerevisiae</i> ]

续表 5 BNMH 和 CXYM 中独特的预测蛋白信息  
Tab. 5 List of predicted unique proteins in BNMH and CXYM (continued)

蛋白序列号 Protein ID	目标蛋白数 HSPs	<i>P</i>	功能描述 Description
g413.t1	0		Hypothetical protein
g611.t1	0		Hypothetical protein
g687.t1	0		Hypothetical protein
g770.t1	0		Hypothetical protein
g956.t1	0		Hypothetical protein
g2668.t1	0		Hypothetical protein
g2815.t1	0		Hypothetical protein
g3010.t1*	0		Hypothetical protein
g3274.t1	0		Hypothetical protein
g7384.t1*	0		Hypothetical protein
g7385.t1*	0		Hypothetical protein
g12632.t1	0		Hypothetical protein
g12720.t1	0		Hypothetical protein
g12768.t1	0		Hypothetical protein
g13142.t1	0		Hypothetical protein
g13276.t1	0		Hypothetical protein

注: \*为地下鲜重连锁的候选蛋白。

Note: \* was protein linked UFW.

泛关注,然而,现有的育种种质在遗传多样性上的匮乏成为小桐子完全商业化的主要阻碍,而种质资源的收集和评估对于培育高产、优质和高适应性的小桐子新品种至关重要。迄今为止在研究多样性的各种分子标记中,基于 PCR 的标记如 RAPD (随机扩增多态性 DNA)、ISSR (简单序列间重复)、SSR (简单序列重复)、EST-SSR 和 AFLP (扩增片段长度多态性) 均已被用于探究小桐子的遗传多样性和系统发生。PECINA-QUINTERO 等<sup>[22]</sup>借助 AFLP 技术对来自墨西哥中部和东南部 9 个州的 175 份种质 (包括有毒性、无毒性、蛋白质含量和种油含量对比大的基因型) 展开全面的遗传多样性分析,发现墨西哥小桐子具有很高的遗传多样性 (明显高于 GRATIVOL 等<sup>[23]</sup>对巴西来源小桐子的相关报道),并伴随着不同州的种群结构。SUDHEER 等<sup>[24]</sup>使用 RAPD 和 AFLP 技术对 42 个印度小桐子种质的分子多样性分析,指出其多态性的平均百分比为 26.47,显著低于墨西哥 (33.18)。这些结果都支持了目前的主流观点,墨西哥和中美洲可能是麻疯树起源和多样化中心<sup>[23-24]</sup>,而非洲和亚洲的小桐子种质存在低的遗传变异性<sup>[11, 25-26]</sup>。WEN 等<sup>[13]</sup>早期利用 36 个 EST-SSR 和 20 个 G-SSR 来自南美洲、中国和

印度尼西亚的 45 个种质中检测到 183 个多态性等位基因,平均遗传多样性指数为 0.5572,发现中国云南的种质依然显示出比其他地理区域 (如格林纳达和中国海南) 更高水平的遗传变异,可作为丰富小桐子遗传背景的重要种质资源。在本研究中,小桐子种子采集自云南省的不同地域,并首先用于多项形态学或农艺性状分析,包括种子径围、百粒重、幼苗株高和生物量积累等。这些性状对作物育种改良研究很重要,但应用于小桐子遗传多样性的系统研究还很少见到报道。本研究还发现小桐子种群内存在显著性状差异,尤以生物量相关性状的变异系数最大 (21.90%~25.87%),这与 PESTANA-CALDAS 等<sup>[27]</sup>和 LAVIOLA 等<sup>[28]</sup>的研究相似,均认为单株植物种子产量 (包括种子百粒重)、植株高度对小桐子遗传变异性贡献最小,而生长有关的特征则表现出最大的表型变异,这可能和地区适应性和驯化有密切关系。

有研究表明在小桐子中,分子标记和性状数据连锁不紧密<sup>[28-29]</sup>。经典的遗传定位依靠基因型和表型的紧密连锁<sup>[30]</sup>,结合基因组测序的 BSA 分析法扩大了群体内一致性数据和群体间差异性数据<sup>[31]</sup>,本研究选择地下鲜重这一极端差异性状的

2个分离群体, BNMH(版纳勐海)和CX YM(楚雄元谋)以BSA分离法进行全基因组测序, 分别组装获得2783和3033条contigs并通过GWAS对重要SNP标记的基因组区域进行高精度的关联和标记<sup>[32]</sup>, 在9号染色体上确定了5个UFW高频SNPs的峰值区域, 有3个候选蛋白与UFW性状连锁。从基因组测序和比较分析来看, 云南省小桐子具有较高的单核苷酸多态性, 具有特有蛋白40个(含未注释功能蛋白16个), 显示了较好的多样性、独特性和地区适应性。此外, 地下鲜重主要涉及根系生长和发育, 这一性状的差异也明显和地区驯化有关。

本研究利用MISA开发并随机选出8对SSR引物(有效扩增引物4对), 对9个居群共95个DNA样本进行毛细管电泳检测, 结果发现引物多态性条带最多为6个, 最少为2个, 引物位点的多态信息含量指数(PIC)变幅为0.041~0.711(平均为0.370), 高于VÁSQUEZ-MAYORGA等<sup>[33]</sup>(PIC=0.274±0.165)和KUMAR等<sup>[34]</sup>(PIC=0.25±0.16)的报道, 可作为小桐子种质亲缘关系鉴定、遗传变异检测等分子生物学研究的重要可用资源。此外, 群体结构研究虽表现为中度多态性<sup>[35]</sup>, 但本研究发现云南小桐子的杂合度(平均 $H_e=0.437$ )与目前世界范围内的小桐子报道相比十分可观, 明显高于WEN等<sup>[13]</sup>对南美洲和云南种质资源的报道( $H_e$ 分别为0.33、0.3473)。多数研究认为小桐子起源于中美洲, 但传入我国时间和路线不详<sup>[36]</sup>。本研究基于UPGMA聚类分析发现各居群间亲缘关系和地理位置有一定相关性, 各居群间遗传多样性较高, 这与WEN等<sup>[13]</sup>、RAFII等<sup>[37]</sup>的结论相一致, 暗示小桐子可能由多点多分化种传入云南, 传入后在各生长地之间基因交流较少, 变异主要为独立群体演化事件, 其主要结果为充分了解小桐子在中国的起源中心的遗传多样性和种群结构提供参考, 也为其作为可持续能源作物的复兴奠定新的理论基础。在后续研究中应进一步结合胞质基因组多态性特点及相关历史考据文献对小桐子在云南的传入、传播及分布特点进行更加深入的探讨。此外, 小桐子作为云南省的一个主要的外来物种, 在将来的育种工作中有必要加大对云南省(尤其是边境地区)的小桐子种质资源的勘探与发掘, 结合InDel等大片段标记<sup>[34, 38]</sup>, 开发适用于种质资源检测的各类有效标记, 进一步通过分子设计育种加速小桐子的种植

资源改良。

## 4 结论

本研究采集了云南省10个小桐子聚居地种子, 调查了9个重要农艺性状, 利用地下鲜重差异显著的2个BSA分离群体, 通过RAD-seq测序结合GWAS分析挖掘到SNPs 22 756个, 分析发现高频SNPs集中在9号染色体, 进一步定位到来自于版纳和楚雄群体的特有候选蛋白40个, 筛选了3个UFW候选蛋白。此外, 研究基于4对SSR标记对9个地区小桐子种质资源进行了群体结构分析, 得到了较好的多态性和推测了可能的起源中心。

## 参考文献

- [1] MONTES J M, MELCHINGER A E. Domestication and breeding of *Jatropha curcas* L.[J]. Trends in Plant Science, 2016, 21(12): 1045-1057.
- [2] 董敏, 何俊, 许彦红, 罗明灿. 能源树种麻风树产业发展回顾和分析[J]. 林业资源管理, 2017(1): 12-18.  
DONG M, HE J, XU Y H, LUO M C. Review and analysis on *jatropha* industry[J]. Forest Resources Management, 2017(1): 12-18. (in Chinese)
- [3] WU D M, FENG J Y, CHU S S, JACOBS D F, TONG X, ZHAO Q, CHEN X Y, ZENG S C. Integrated application of sewage sludge, earthworms and *Jatropha curcas* on abandoned rare-earth mine land soil[J]. Chemosphere, 2019, 214: 47-54.
- [4] ZHAO Y N, WANG Y B, WANG H F, WU Y M, MAKKAR H P, LIU J X. Nutritional value of detoxified *Jatropha curcas* seed cake protein isolates using rats as an animal model[J]. Animal Nutrition, 2018, 4(4): 429-434.
- [5] MAZUMDAR P, SINGH P, BABU S, SIVA R, HARIKRISHNA J A. An update on biological advancement of *Jatropha curcas* L.: new insight and challenges[J]. Renewable & Sustainable Energy Reviews, 2018, 91: 903-917.
- [6] THAPA S, INDRAWAN N, BHOI P R. An overview on fuel properties and prospects of *Jatropha* biodiesel as fuel for engines[J]. Environmental Technology & Innovation, 2018, 9: 210-219.
- [7] WANG H B, ZOU Z R, WANG S S, GONG M. Global analysis of transcriptome responses and gene expression profiles to cold stress of *Jatropha curcas* L.[J]. PLoS One, 2013, 8(12): e82817.
- [8] LIU H, WANG F F, PENG X J, HUANG J H, SHEN S H. Global phosphoproteomic analysis reveals the defense and

- response mechanisms of *Jatropha curcas* seedling under chilling stress[J]. International Journal of Molecular Sciences, 2019, 20(1): 208.
- [9] 王辉, 杜官本. 我国小桐子产业现状及发展趋势探析[J]. 中国粮油学报, 2020, 35(6): 190-195.  
WANG H, DU G B. Analysis on present industry situation and development trend of *Jatropha curcas* in China[J]. Journal of the Chinese Cereals and Oils Association, 2020, 35(6): 190-195. (in Chinese)
- [10] 欧文军, 王文泉, 李开绵. 120 份小桐子种质的分子遗传多样性分析[J]. 热带作物学报, 2009, 30(3): 284-292.  
OU W J, WANG W Q, LI K M. Molecular genetic diversity analysis of 120 accessions of *Jatropha curcas* L. germplasm[J]. Chinese Journal of Tropical Crops, 2009, 30(3): 284-292. (in Chinese)
- [11] TATIKONDA, WANI S P, KANNAN S, BEERELLI N, SREEDEVI T K, HOISINGTON D A, DEVI P, VARSHNEY R K. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant[J]. Plant Science, 2009, 176(4): 505-513.
- [12] ZHANG Z Y, GUO X L, LIU B Y, TANG L, CHEN F. Genetic diversity and genetic relationship of *Jatropha curcas* between China and Southeast Asian revealed by amplified fragment length polymorphisms[J]. African Journal of Biotechnology, 2011, 10(15): 2825-2832.
- [13] WEN M F, WANG H Y, XIA Z Q, ZOU M L, LU C, WANG W Q. Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L[J]. BMC Research Notes, 2010, 3(1): 42.
- [14] 王晓. 川滇地区麻疯树种质资源遗传多样性的 AFLP 分析[D]. 雅安: 四川农业大学, 2010.  
WANG X. Analysis on genetic diversity of *Jatropha curcas* Linn in Sichuan-Yunnan region, by AFLP technology[D]. Ya'an: Sichuan Agriculture University, 2010. (in Chinese)
- [15] 何玮, 郭亮, 王岚, 杨威, 唐琳, 陈放. 麻疯树种质资源遗传多样性的 ISSR 分析[J]. 应用与环境生物学报, 2007, 13(4): 466-470.  
HE W, GUO L, WANG L, YANG W, TANG L, CHEN F. ISSR analysis of genetic diversity of *Jatropha curcas* L.[J]. Chinese Journal of Applied and Environmental Biology, 2007, 13(4): 466-470. (in Chinese)
- [16] 郭国业, 徐莺, 唐琳, 陈放, 韩学琴. 不同地理种源麻疯树表型变异研究[J]. 四川农业大学学报, 2020, 38(2): 143-151, 160.  
GUO G Y, XU Y, TANG L, CHEN F, HAN X Q. Phenotypic variation analysis of *Jatropha curcas* from different geographical provenances[J]. Journal of Sichuan Agricultural University, 2020, 38(2): 143-151, 160. (in Chinese)
- [17] 向振勇, 宋松泉, 王桂娟, 陈茂盛, 杨成源, 龙春林. 云南山南不同种源地小桐子遗传多样性的 ISSR 分析[J]. 云南植物研究, 2007, 29(6): 619-624.  
XIANG Z Y, SONG S Q, WANG G J, CHEN M S, YANG C Y, LONG C L. Genetic diversity of *Jatropha curcas* (Euphorbiaceae) collected from southern Yunnan, detected by inter-simple sequence repeat (ISSR) [J]. Acta Botanica Yunnanica, 2007, 29(6): 619-624. (in Chinese)
- [18] VURTURE G W, SEDLAZECK F J, NATTESTAD M, UNDERWOOD C J, HAN F, GURTOWSKI J, SCHATZ M C. GenomeScope: fast reference-free genome profiling from short reads[J]. Bioinformatics, 2017, 33(14): 2202-2204.
- [19] BEIER S, THIEL T, MÜNCH T, UWE S, MASCHER M. MISA-web: a web server for microsatellite prediction[J]. Bioinformatics, 2017, 33(16): 2583-2585.
- [20] TAMURA K, PETERSON D, PETERSON N, STECHER G, NEI M, KUMAR S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molecular Biology & Evolution, 2011, 28(10): 2731-2739.
- [21] LUO R B, LIU B H, XIE Y L, LI Z Y, HUANG W H, YUAN J Y, HE G Z, CHEN Y X, PAN Q, LIU Y J, TANG J B, WU G X, ZHANG H, SHI Y J, LIU Y, YU C, WANG B, LU Y, HAN C L, CHEUNG D W, YIU S M, PENG S L, XIAOQIAN Z, LIU G M, LIAO X K, LI Y R, YANG H M, WANG J, LAM T W, WANG J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler[J]. Giga Science, 2012, 1(1): 18.
- [22] PECINA-QUINTERO V, ANAYA-LÓPEZ J L, ZAMARRIPA-COLMENERO A, NÚÑEZ-COLÍN C A, MONTES-GARCÍA N, SOLÍS-BONILLA J L, JIMÉNEZ-BECERRIL M F. Genetic structure of *Jatropha curcas* L. in Mexico and probable centre of origin[J]. Biomass and Bioenergy, 2014, 60: 147-155.
- [23] GRATIVOL C, DA FONSECA LIRA-MEDEIROS C, HEMERLY A S, FERREIRA P C G. High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions[J]. Molecular Biology Reports, 2011, 38(7): 4245-4256.
- [24] SUDHEER PAMIDIMARRI D V N, REDDY M P. Phylogeography and molecular diversity analysis of *Jatropha curcas* L. and the dispersal route revealed by RAPD, AFLP and nrDNA-ITS analysis[J]. Molecular Biology Reports, 2014, 41(5): 3225-3234.
- [25] BASHA S D, FRANCIS G, MAKKAR H P S, BECKER K, SUJATHA M. A comparative study of biochemical traits and molecular markers for assessment of genetic relationships

- between *Jatropha curcas* L. germplasm from different countries[J]. *Plant Science*, 2009, 176(6): 812-823.
- [26] BASHA S D, SUJATHA M. Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers[J]. *Euphytica*, 2007, 156(3): 375-386.
- [27] PESTANA-CALDAS C N, SILVA S A, MACHADO E L, DE SOUZA D R, CERQUEIRA-PEREIRA E C, SILVA M S. Genetic divergence through joint analysis of morpho- agronomic and molecular characters in accessions of *Jatropha curcas*[J]. *Genetics and Molecular Research*, 2016, 15(4).
- [28] LAVIOLA B G, ALVES A A, ROCHA R B, DRUMOND M A. *Jatropha*, challenges for a new energy crop[M]. New York: Springer, 2012: 71-94.
- [29] ALVES A A, BHERING L L, ROSADO T B, LAVIOLA B G, FORMIGHIERI E F, CRUZ C D. Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of the Brazilian physic nut germplasm bank[J]. *Genetics & Molecular Biology*, 2013, 36(3): 371-381.
- [30] WING R A, ZHANG H B, TANKSLEY S D. Map-based cloning in crop plants. Tomato as a model system: I. Genetic and physical mapping of jointless[J]. *Molecular & General Genetics MGG*, 1994, 242: 681-688.
- [31] TAKAGI H, ABE A, YOSHIDA K, KOSUGI S, NATSUME S, MITSUOKA C, UEMURA A, UTSUSHI H, TAMIRU M, TAKUNO S, INNAN H, CANO LM, KAMOUN S, TERAUCHI R. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations[J]. *Plant Journal*, 2013, 74(1): 174-183.
- [32] TIBBS CORTES L, ZHANG Z W, YU J M. Status and prospects of genome-wide association studies in plants[J]. *The Plant Genome*, 2021, 14(1): e20077.
- [33] VÁSQUEZ-MAYORGA M, FUCHS E J, HERNÁNDEZ E, HERRERA F, BARBOZA N M. Molecular characterization and genetic diversity of *Jatropha curcas* L. in Costa Rica[J]. *Peer J*. 2016, 5: e2931.
- [34] KUMAR Y H, RANJAN A, ASIF M H, MANTRI S, SAWANT S V, TULI R. EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera[J]. *Tree Genetics & Genomes*, 2011, 7(1): 207-219.
- [35] BOTSTEIN D, WHITE R L, SKOLNICK M, DAVIS R W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms[J]. *American Journal of Human Genetics*, 1980, 32(3): 314-331.
- [36] 栗宏林. 中国小桐子不同地理种源差异比较与种源亲缘关系研究[D]. 北京: 北京林业大学, 2008.
- JIA H L. Studies on genetic relationship of provenance and comparative analysis of provenance difference of *Jatropha curcas* Linn[D]. Beijing: Beijing Forestry University, 2008. (in Chinese)
- [37] RAFII M Y, SHABANIMOFRAD M, EDAROYATI M P, LATIF M A. Analysis of the genetic diversity of physic nut, *Jatropha curcas* L. accessions using RAPD markers[J]. *Molecular Biology Reports*, 2012, 39(6): 6505-6511.
- [38] BRESSAN E A, SEBBENN A M, FERREIRA R R, LEE T S G, FIGUEIRA A. *Jatropha curcas* L. (Euphorbiaceae) exhibits a mixed mating system, high correlated mating and apomixis[J]. *Tree Genetics & Genomes*, 2013, 9(4): 1089-1097.