

苹婆叶绿体全基因组序列特征及苹婆属系统发育分析

梁莹莹, 吴文如*, 彭晓祺, 张雨薇, 陆亚茹, 邹何元

广州中医药大学中药学院, 广东 广州 510006

摘要: 目的 分析苹婆 *Sterculia monosperma* 的叶绿体基因组结构信息, 以及苹婆属的种间亲缘关系, 为苹婆属的系统发育和物种鉴定研究提供参考。方法 基于 DNBSEQ 测序平台, 对苹婆叶绿体基因组进行测序, 利用生物信息学手段进行组装、注释和结构分析, 并与从 NCBI 下载的假苹婆 *S. lanceolata*、西蜀苹婆 *S. lanceifolia*、小花苹婆 *S. micrantha*、粉苹婆 *S. euosma* 和短柄苹婆 *S. brevissima* 5 种苹婆属植物的叶绿体基因组进行多序列比对和系统发育等相关分析。结果 苹婆叶绿体基因组具有典型的四分体结构, 全长为 160 178 bp, 总 GC 含量为 37.0%, 编码 130 个基因, 得到 109 条 SSR 序列, 显示 30 条序列的 RSCU 大于 1。6 种苹婆属相关比对分析可知苹婆属叶绿体基因组序列具有一定相似性, 大单拷贝区 (large single-copy region, LSC) 存在 9 个不同程度差异片段, Pi 值显示有 4 个片段有高变异性, 系统发育树表明假苹婆和粉苹婆、西蜀苹婆和小花苹婆亲缘关系更近, 短柄苹婆和苹婆相对独立。结论 解析了苹婆的叶绿体基因组结构和特征, 探究了苹婆属植物种间亲缘关系, 为苹婆属物种鉴定、资源保护和开发及系统进化提供理论依据。

关键词: 苹婆; 假苹婆; 苹婆属; 叶绿体基因组; 系统发育

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)19-6721-09

DOI: 10.7501/j.issn.0253-2670.2024.19.024

Characteristics of chloroplast whole genome sequence of *Sterculia monosperma* and phylogenetic analysis of genus *Sterculia*

LIANG Yingying, WU Wenru, PENG Xiaoqi, ZHANG Yuwei, LU Yaru, ZOU Heyuan

School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

Abstract: Objective To provide reference for phylogeny and species identification *Sterculia* L. by studying the chloroplast genome structure information of *Sterculia monosperma* and genetic relationship between species of *Sterculia* L. **Methods** Based on the DNBSEQ sequencing platform, the chloroplast genome of *S. monosperma* was sequenced, and bioinformatics methods were used to assemble, annotate and structurally analyze it. The multiple sequence alignment and phylogenetic correlation were carried out between the chloroplast genomes of *S. monosperma* and five species of *Sterculia* L. from NCBI, including *S. lanceolata*, *S. lanceifolia*, *S. micrantha*, *S. euosma*, *S. brevissima*. **Results** The chloroplast genome of *S. monosperma* has a typical tetrad structure, with the total length of 160 178 bp, the total GC content of 37.0%, and encoding 130 genes. A total of 109 SSR sequences are obtained, showing that the RSCU of 30 sequences was greater than 1. The related comparison analysis of six species of *Sterculia* L. shows that the chloroplast genome sequences have certain similarities. There are nine fragments with varying degrees of difference in the large single-copy region (LSC) region. The Pi value shows that four fragments have high variability. The phylogenetic tree shows that *S. lanceolata* and *S. euosma*, *S. lanceifolia* and *S. micrantha* are more closely related, while *S. brevissima* and *S. monosperma* are relatively independent. **Conclusion** The structure and characteristics of the chloroplast genome of *S. monosperma* were analyzed, and the genetic relationship between species of *Sterculia* L. was explored, providing a theoretical basis for species identification, resource protection and development, and systematic evolution of *Sterculia* L..

Key words: *Sterculia monosperma* Vent.; *Sterculia lanceolata* Cav.; *Sterculia* L.; chloroplast genome; phylogeny

收稿日期: 2024-02-02

基金项目: 中央本级重大增减支项目“名贵中药资源可持续利用能力建设项目”(2060302); 广州中医药大学大学生创新创业训练计划项目(202310572342); 2020 年度广东省科技特派员项目(KTP20200137)

作者简介: 梁莹莹(2001—), 女, 硕士研究生, 研究方向为中药品种鉴定与质量标准研究。E-mail: 750451261@qq.com

*通信作者: 吴文如(1976—), 教授, 硕士生导师, 从事中药品种鉴定与质量标准研究。E-mail: wuwenru@gzucm.edu.cn

苹婆 *Sterculia monosperma* Vent. 来源于梧桐科植物苹婆属乔木，主要分布于广东、广西的南部、福建东南部、云南南部和台湾等地^[1]，多为人工栽培。苹婆以种子入药，具有消食和胃、明目壮阳、解毒杀虫等作用，可用于食积、胃翻呕吐^[2]，以及治疗痔疮、中耳炎和小肠疝气、血痢等^[3]。苹婆种子中含有的多酚类成分，能清除自由基，是制作抗氧化剂等保健产品的天然原料^[4]，亦可抑制肠道葡萄糖转运和吸收，调控糖尿病患者的血糖水平^[3]。

苹婆种子的别名较多，如凤眼果、九层皮、枇杷果、七姐果、罗晃子^[1]等，易导致混淆和临床误用^[5]。然而，苹婆植物名与同科属的假苹婆 *S. lanceolata* Cav.、西蜀苹婆 *S. lanceifolia* Roxb.、粉苹婆 *S. euosma* W. W. Sm. 等名称相似，且苹婆与不能入药的假苹婆种子外形十分相似，较难区分^[6-7]。苹婆属作为梧桐科的模式属^[1]，是梧桐科的重要组成部分。有研究者基于苹婆属种间在木材解剖结构的差异^[8-9]，将苹婆属分为 *Sterculia* A 和 *Sterculia* B 类群^[10]。但在分子系统发育研究中，苹婆属的相关研究较少，未能较好的揭示苹婆属植物的种间亲缘关系。

随着基因测序技术的发展，叶绿体基因组被广泛用于植物分子系统发育学、群体遗传学和保护遗传学研究^[11]，其体积小，多为单亲遗传，缺乏重组，因此，能稳定有效地扩展遗传信息，用于保护药用植物的种质资源。叶绿体基因组中发现的各类突变事件，为探究物种的进化历史提供了丰富信息，建立了植物之间的关系，对药用植物来说，这提供了有效的遗传标记来解决物种的鉴定问题^[12]。

目前，苹婆的研究多集中在成分分析、种质评价与病虫害等方面，尚未见苹婆的叶绿体基因组序列特征报道。基于此，本研究将通过探究苹婆叶绿体基因组结构及相关信息，并选取包括苹婆在内的 6 种苹婆属植物进行种间的系统发育分析，从分子的角度了解苹婆属植物种间的亲缘关系，为后续苹婆属的物种鉴定、资源开发与利用和遗传分类等研究提供理论基础。

1 材料与仪器

1.1 材料

植物样品采集：样品来源于广州中医药大学药王山（113°23'E，23°3'N），经广州中医药大学吴文如教授鉴定为梧桐科苹婆属植物苹婆 *S. monosperma* Vent.。将采集的新鲜嫩叶储存于中药鉴定学实验

室，4℃冰箱保存，用于后续提取总 DNA。苹婆新测序的叶绿体基因组组装和注释信息已上传 NCBI 网站（登录号 PP467378）。

其他苹婆属叶绿体基因组来源：从 NCBI 数据库（<https://www.ncbi.nlm.nih.gov/>）下载其他 5 种苹婆属的叶绿体基因组数据，用于后续分析。具体见表 1。

表 1 用于叶绿体基因组分析的苹婆属样品来源
Table 1 Sources of genus *Sterculia* L. for analysis of chloroplast genome

编号	物种名	Genbank 号
JPP	假苹婆 <i>S. lanceolata</i>	NC_054170.1
XSPP	西蜀苹婆 <i>S. lanceifolia</i>	MZ901905.1
XHPP	小花苹婆 <i>S. micrantha</i>	MZ901907.1
FPP	粉苹婆 <i>S. euosma</i>	ON881466.1
DBPP	短柄苹婆 <i>S. brevissima</i>	ON881824.1

1.2 仪器

HH-24 型数显恒温水浴锅（北京欧莱德科学仪器有限公司），B-500 型超微量分光光度计（上海元析仪器有限公司），JS-2012 型凝胶成像仪（上海培清科技有限公司），Power Pac Basic 型水平电泳仪（美国 Bio-Rad 公司）。

2 方法

2.1 总 DNA 提取和测序

采用改良的 CTAB 法^[14]提取苹婆的新鲜叶片总 DNA。使用 B-500 超微量分光光度计（上海元析仪器有限公司）和 1% 琼脂糖凝胶电泳检测总 DNA 的浓度和完整性。检测合格后委托深圳华大基因科技有限公司使用 DNBSEQ 测序平台进行二代测序，得到 3 Gb 的 clean data。

2.2 叶绿体基因组组装、注释

使用 getOrganelle v1.7.7.0^[15] 组装 clean data，参考数据库为 embplant_pt（陆生植物叶绿体），最大扩充循环数为 10，调用 K-mer 值为 65 105 127；拼接完成后用 Bandage v0.8.1 软件可视。以苹婆（MZ958831.1）为参考序列，通过 CPGAVAS2^[16] 网站（<http://47.96.249.172:16019/analyzer/home>）对苹婆、西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆进行基因组注释，Genenious R9.0.2 检查并调整注释结果，最后使用在线工具 OGDRAW^[17]（<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>）绘制叶绿体基因组图谱。

2.3 简单重复序列和散在重复序列

使用 MISA^[18] (<https://webblast.ipk-gatersleben.de/misa/>) 探索简单重复序列 (SSR) 在叶绿体全基因组序列上的分布, 设置参数: 单核苷酸重复单元不少于 10 个, 二核苷酸重复单元不少于 5 个, 三核苷酸重复单元不少于 4 个, 四、五、六核苷酸重复单元不少于 3 个, 设置 2 个 SSR 间最大长度为 0。应用 REPuter^[19] (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) 检测叶绿体基因组中的散在重复序列, 包括正向重复序列 (F)、反向重复序列 (R)、回文重复序列 (P) 和互补重复序列 (C), 设置参数: 海明距离设置为 3, 最小重复长度不少于 30 bp。

2.4 密码子使用分析

同义密码子使用度 (relative synonymous codon usage, RSCU) 是每个同义密码子用于编码特定氨基酸的相对频率^[20]。对苹婆叶绿体基因组中的编码基因序列进行筛选, 过滤重复序列及长度小于 300 bp 的序列, 保留以 ATG 为起始的序列。过滤后的序列采用 CodonW v1.4.4 (<http://codonw.sourceforge.net/>) 软件对苹婆和同义密码子相对使用度进行分析和统计, 使用 RStudio (2023.06.2 +561 版本) 绘图。

2.5 苹婆属叶绿体基因组比较

为评估苹婆属叶绿体基因组序列的差异程度, 本研究以新注释的苹婆作为参考基因组, 采用 mVISTA^[21] (<https://genome.lbl.gov/vista/mvista/submit.shtml>), 选择 Shuffle-LAGAN 模式, 对苹婆、假苹婆、西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆这 6 条叶绿体基因组全长序列进行比对分析。

2.6 苹婆属叶绿体基因组的边界收缩与扩张

IR 区域边界的收缩和扩张被认为是反复发生的过程, 其区域的收缩与扩张会导致叶绿体基因组的变化^[20]。为评估苹婆属植物叶绿体基因组各分区边界的差异, 本研究采用 CPJSdraw^[23]对苹婆、假苹婆、西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆叶绿体基因组的边界信息进行分析。

2.7 苹婆属叶绿体基因组多态性位点分析

单核苷酸多态性^[24] (single nucleotide polymorphism, SNP) 主要是指基因组水平上由单核苷酸变异引起的 DNA 序列多态性。使用 mafft^[25]v7.490 软件比对 8 条叶绿体基因组序列, 对齐后导出 fasta 格式文件。将所得 fasta 文件导入 DnaSP^[26]v6.12.03 软件, 计算 8 个苹婆属植物叶绿

体基因组的核苷酸多样性水平, 设置窗口长度为 600, 步长为 200。

2.8 系统发育分析

从 NCBI 上下载锦葵目梧桐科梭罗树属植物梭罗树 *Reevesia pubescens* (NC_063749.1) 和银叶树属植物长柄银叶树 *Heritiera angustata* (NC_037784.1) 的叶绿体基因组序列, 以这 2 种植物作为外类群, 加上 6 种苹婆属植物构建系统发育树。使用 mafft^[25]v7.490 软件进行多序列比较, 采用 tbtools v2.069 中的 trimAL^[27]工具自动修剪后, 通过 IQ-TREE^[28] v2.2.0.3 软件获得最优模型 K3Pu+F+G4, 构建最大似然 (maximum likelihood, ML) 系统发育树, 参数设置为 -m K3Pu+F+G4 -B 1000 -T AUTO。

3 结果与分析

3.1 苹婆及苹婆属叶绿体基因组结构与特征

苹婆及苹婆属叶绿体基因组全基因组谱图见图 1。6 种苹婆属叶绿体基因组序列呈现双链环状, 且分为 4 个区域, 即大单拷贝区 (LSC 区), 小单拷贝区 (SSC 区) 和 2 个反向重复区 (IR 区)。由表 2 可知, 这 6 种苹婆属植物序列长度不同, 序列全长为 160 178~161 172 bp, 除粉苹婆外, 总 GC 含量均为 37.0%, 其中苹婆的 LSC 区长度为 89 553 bp, GC 含量为 34.8%, SSC 区长度为 19 697 bp, GC 含量为 31.6%, IR 区的长度 25 464 bp, GC 含量为 43.0%。由表 3 可知, 6 种苹婆属叶绿体基因组序列编码了 130 个基因, 基因功能主要由化合作用和自我复制为主, 包括 37 个转运 RNA (tRNA)、8 个核糖体 RNA (rRNA), 除假苹婆缺少 *infA* 基因外, 其他苹婆属植物均有 85 个蛋白编码基因。

3.2 苹婆的简单重复序列与散在重复序列

由 REPuter 生成的结果可知, 总共生成 70 条散在重复序列, 其中含有 6 条互补重复序列 (C)、25 条正向重复序列 (F)、29 条回文重复序列 (P)、10 条反向重复序列 (R)。

通过 MISA 在线工具检测苹婆叶绿体基因组的简单重复 (SSR) 序列, 结果显示, 总共得到 109 条 SSR 序列, 主要分为 6 种 SSR, 包括 76 条单核苷酸 SSR、15 条二核苷酸 SSR、4 条三核苷酸 SSR、10 条四核苷酸 SSR、3 条五核苷酸 SSR、1 条六核苷酸 SSR。SSR 类型中单核苷酸 SSR 的 A 重复类型最多, 占有所有 SSR 序列的 37.61%。由图 2 可知, LSC 区的 SSR 序列最多, 占有所有 SSR 序列的 73.39%。

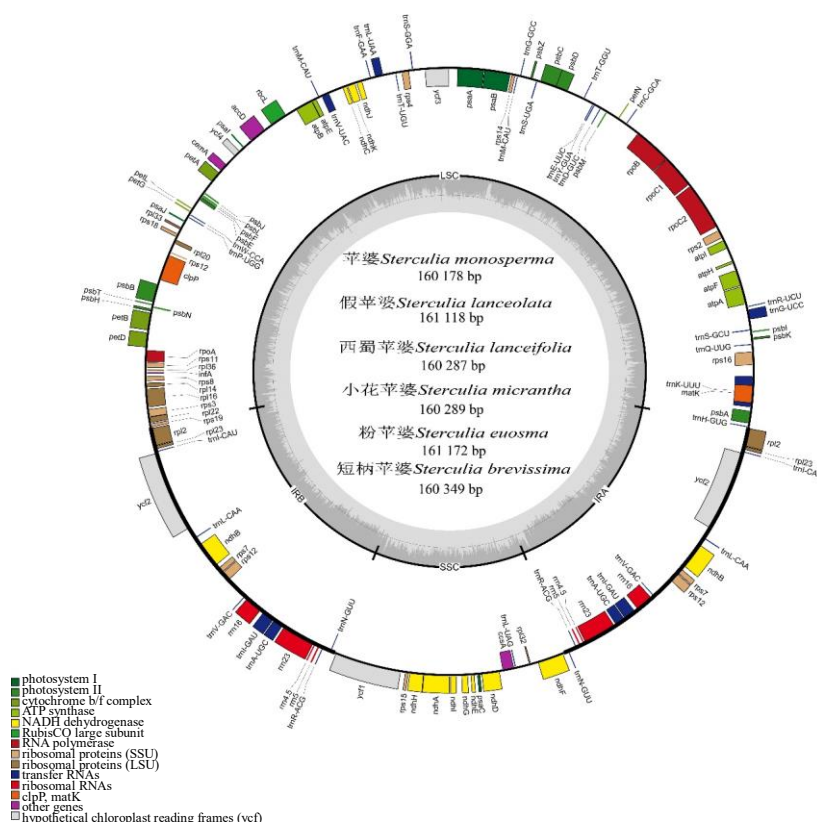


图 1 苹婆属 6 种植物叶绿体基因组谱图

Fig. 1 Chloroplast genome profile of six species of genus *Sterculia*

表 2 苹婆属 6 种植物叶绿体基因组的基本组成

Table 2 Basic composition of chloroplast genome of six species of genus *Sterculia* L.

项目	苹婆	假苹婆	西蜀苹婆	小花苹婆	粉苹婆	短柄苹婆
基因组长度/bp	160 178	161 118	160 287	160 289	161 172	160 349
LSC 区长度/bp	89 553	89 520	89 335	89 357	89 575	89 419
SSC 区去长度/bp	19 697	20 818	19 640	19 640	20 331	19 656
IR 区长度/bp	25 464	25 390	25 646	25 646	25 633	25 637
总基因数量	130	129	130	130	130	130
转运 RNA 基因数	37	37	37	37	37	37
核糖体 RNA 基因数	8	8	8	8	8	8
蛋白编码基因数	85	84	85	85	85	85
总 GC 含量/%	37.0	37.0	37.0	37.0	36.9	37.0
LSC 区 GC 含量/%	34.8	34.8	34.8	34.8	34.8	34.8
SSC 区 GC 含量/%	31.6	31.2	31.7	31.7	31.5	31.7
IR 区 GC 含量/%	43.0	43.1	42.7	42.7	42.8	42.8

3.3 苹婆的密码子使用分析

密码子使用对基因组进化具有重要作用，同义密码子使用的平等是自然选择、突变和遗传漂变造成的^[20]。通过对苹婆叶绿体基因组中的编码基因序列进行筛选，过滤得到 52 条 CDS 序列。图 3 是这 52 条序列的相对同义密码子使用度 (RSCU) 结果。当 RSCU>1 时，表示该密码子使用频率较高，偏好性较强；当 RSCU<1 时，表示该密码子使用频率较低，偏好性较弱；当 RSCU=

1 时，该密码子无偏好性^[29]。由图 3 可知，有 30 条序列的 RSCU 大于 1，使用频率较高，其中有 29 条序列是以 A 或 U 结尾，仅有 1 条序列以 G 结尾，说明苹婆叶绿体基因组的密码子偏好性多以 A 或 U 结尾。由 6 种密码子编码的精氨酸 (Arg)、亮氨酸 (Leu)、丝氨酸 (Ser) 是该叶绿体基因组中最丰富的氨基酸。

3.4 苹婆属叶绿体基因组多序列比对

mVISTA 结果 (图 4) 显示，6 种苹婆属植物叶

表 3 苹婆属 6 种植物叶绿体基因组基因组成
Table 3 Gene composition of chloroplast genome of six species of genus *Sterculia*

基因功能	基因分类	基因名
光合作用	光系统 I 基因	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaI</i> 、 <i>psaJ</i>
	光系统 II 基因	<i>psbA</i> 、 <i>psbB</i> 、 <i>psbC</i> 、 <i>psbD</i> 、 <i>psbE</i> 、 <i>psbF</i> 、 <i>psbH</i> 、 <i>psbI</i> 、 <i>psbJ</i> 、 <i>psbK</i> 、 <i>psbL</i> 、 <i>psbM</i> 、 <i>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>
	NADH 氧化还原酶基因	<i>ndhA</i> *、 <i>ndhB</i> *(2)、 <i>ndhC</i> 、 <i>ndhD</i> 、 <i>ndhE</i> 、 <i>ndhF</i> 、 <i>ndhG</i> 、 <i>ndhH</i> 、 <i>ndhI</i> 、 <i>ndhJ</i> 、 <i>ndhK</i>
	细胞色素 b/f 复合体基因	<i>petA</i> 、 <i>petB</i> *、 <i>petD</i> *、 <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>
	ATP 酶基因	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> *、 <i>atpH</i> 、 <i>atpI</i>
	核酮糖二磷酸羧化酶大亚基基因	<i>rbcL</i>
自我复制	核糖体蛋白大亚基基因	<i>rpl14</i> 、 <i>rpl16</i> *、 <i>rpl2</i> *(2)、 <i>rpl20</i> 、 <i>rpl22</i> 、 <i>rpl23</i> (2)、 <i>rpl32</i> 、 <i>rpl33</i> 、 <i>rpl36</i>
	核糖体蛋白小亚基基因	<i>rps11</i> 、 <i>rps12</i> ***(2)、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps16</i> *、 <i>rps18</i> 、 <i>rps19</i> 、 <i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> (2)、 <i>rps8</i>
	RNA 聚合酶基因	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> *、 <i>rpoC2</i>
	核糖体 RNA 基因	<i>rrn16</i> (2)、 <i>rrn23</i> (2)、 <i>rrn4.5</i> (2)、 <i>rrn5</i> (2)
	转运 RNA 基因	<i>trnA</i> -UGC*(2)、 <i>trnC</i> -GCA、 <i>trnD</i> -GUC、 <i>trnE</i> -UUC、 <i>trnF</i> -GAA、 <i>trnG</i> -GCC、 <i>trnG</i> -UCC*、 <i>trnH</i> -GUG、 <i>trnI</i> -CAU(2)、 <i>trnI</i> -GAU*(2)、 <i>trnK</i> -UUU*、 <i>trnL</i> -CAA(2)、 <i>trnL</i> -UAA*、 <i>trnL</i> -UAG、 <i>trnM</i> -CAU(2)、 <i>trnN</i> -GUU(2)、 <i>trnP</i> -UGG、 <i>trnQ</i> -UUG、 <i>trnR</i> -ACG(2)、 <i>trnR</i> -UCU、 <i>trnS</i> -GCU、 <i>trnS</i> -GGA、 <i>trnS</i> -UGA、 <i>trnT</i> -GGU、 <i>trnT</i> -UGU、 <i>trnV</i> -GAC(2)、 <i>trnV</i> -UAC*、 <i>trnW</i> -CCA、 <i>trnY</i> -GUA
其他基因	成熟酶	<i>matK</i>
	蛋白酶	<i>clpP</i> **
	包膜蛋白	<i>cemA</i>
	乙酰辅酶 A 羧化酶基因	<i>accD</i>
	C 型细胞色素合成酶	<i>ccsA</i>
	翻译起始因子	<i>infA</i>
未知功能基因	开放阅读框	<i>ycf1</i> 、 <i>ycf2</i> (2)、 <i>ycf3</i> **、 <i>ycf4</i>

*含有 1 个内含子，**含有 2 个内含子，括号中数字表示多拷贝基因的拷贝数。
* indicates gene containing one intron ; ** indicates gene containing two intron ; bracket indicates gene with duplicate copy.

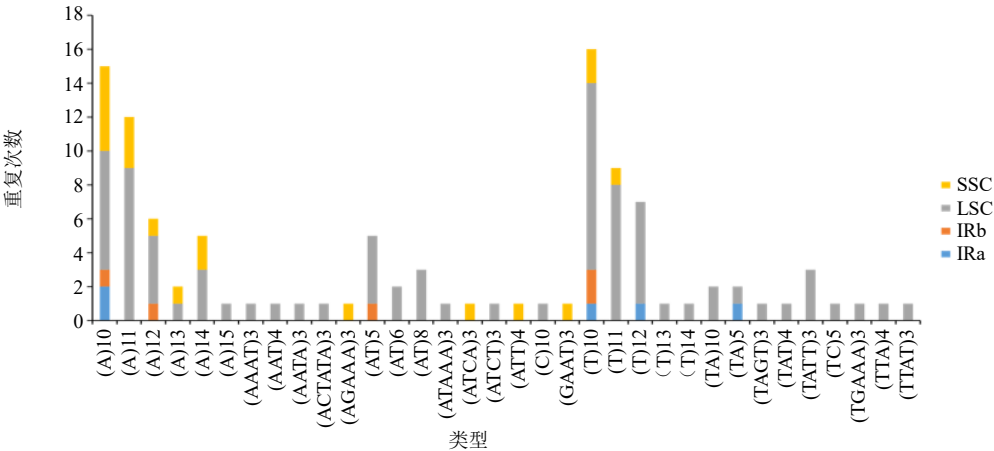


图 2 苹婆叶绿体基因组 SSR 类型及数量分布区域图
Fig. 2 Regional map of SSR type and quantity distribution in chloroplast genome of *S. monosperma*

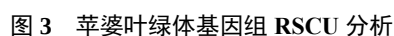


Fig. 3 RSCU analysis in chloroplast genome of *S. monosperma*

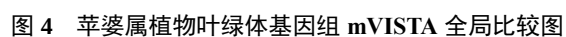


Fig. 4 mVISTA global comparison of genus *Sterculia* chloroplast genome

绿体基因组序列具有一定相似性,在 LSC 区中,6 条序列在 *matK-rps16*、*trnE-UUC-trnT-GGU*、*trnS-UGA-trnM-CAU*、*psaA-ycf3*、*rps4-ndhJ*、*ndhC-trnV-UAC*、*atpB-rbcL*、*psbE-petL*、*petD-rpoA* 的基因间区具有不同程度的差异,相似度低,在 SSC 和 IR 区中的 *trnN-GUU-ycf1*、*trnL-UAG-rpl32*、*ndhF-trnN-GUU*、*trnV-GAC-rps12* 基因间区有较低的变异,相比于 LSC 区,SSC 和 IR 区较为保守,整体序列的外显子区域(exon)比基因间区(CNS)的保守性更高。此外,假苹婆与其他苹婆属植物相比,差异性较大,在 *trnR-UCU-atpA*、*trnC-GCA-petN*、*psbC-psbZ*、*ycf3-trnS-GGA*、*trnN-GUU-ycf1* 基因间区具有较大变异。

3.5 苹婆属的边界收缩与扩张

6 种苹婆属植物叶绿体基因组的 IR 区均与

LSC 区、SSC 区存在 LSC/IRb、IRb/SSC、SSC/IRa、IRa/SSC 边界(图 5)。4 个边界附近的基因类型相同,相对保守,但各种之间仍具有差异。除苹婆和假苹婆的 LSC/IRb 边界距离 *rps19* 基因 42、21 bp 之外,其余苹婆属植物这一边界位于 *rps19* 基因内部,且均向 IRb 区扩张 6 bp。在 IRb/SSC 边界,西蜀苹婆和小花苹婆的 *ycf1* 基因向 IRb 区扩张了 1 bp,其余基因组 IRb/SSC 边界的大多位于 *trnN* 和 *ycf1* 之间。SSC/IRa 边界均位于 *ndhF* 和 *trnN* 之间,但基因距离边界长度有差异。IRa/SSC 边界均位于 *rpl2* 和 *trnH* 之间,除假苹婆 *trnH* 基因距离边界有 173 bp 外,其余苹婆属植物 *trnH* 基因与边界的距离在 2~9 bp。



图 5 苹婆属叶绿体基因组各分区边界比较图

Fig. 5 Comparison of boundaries of each partition of chloroplast genome of genus *Sterculia* L.

3.6 苹婆属多态性位点分析

单核苷酸多态性(SNP)是对确定遗传多样性和亲缘关系非常有利的标记。通过对 6 种苹婆属植物的核酸多态性分析可知(图 6),Pi 值在 0~0.03 133, 平均值是 0.001 919。总共获得 4 个高变异片段,即位于 LSC 区的 *trnL-UAA-trnF-GAA* 和 *psbB*、位于 IR 区的 *ycf1*、*rrn23*, 这些片段 Pi 值均大于 0.01, 可以成为苹婆属物种鉴定的候选片段。

3.7 苹婆属系统发育分析

以锦葵目梧桐科梭罗树 *Reevesia pubescens* Mast.和长柄银叶树 *Heritiera angustata* Pierre 作为外类群进行系统发育分析,结果显示(图 7),苹

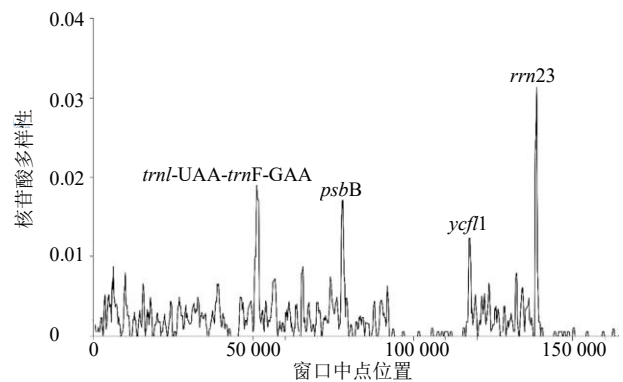


图 6 苹婆属植物叶绿体基因组的滑动窗口分析
Fig. 6 Sliding window analysis of chloroplast genomes of genus *Sterculia*



图 7 基于梧桐科 8 条叶绿体全基因组构建的系统发育树

Fig. 7 Phylogenetic tree constructed based on eight chloroplast whole genomes from Sterculiaceae

婆属为一个支持率高的单系，自展值均为 100%，与梭罗树和长柄银叶树区分开来。假苹婆和粉苹婆聚为一支，为分支 II，亲缘关系较近。苹婆相对独立，与短柄苹婆、西蜀苹婆和小花苹婆共同处于同一分支，为分支 I，西蜀苹婆与小花苹婆单独聚为一支，亲缘关系较近。

4 讨论

相比于核基因组，叶绿体基因组具有保守性，进化速率较慢，被广泛用于植物的系统发育学研究。基于叶绿体基因组中丰富的遗传信息，能筛选出有效的分子标记，有助于药用植物中的物种鉴定工作^[11]。

苹婆具有药食两用的价值，常用于消食和胃，治疗食积，且多酚类化合物具有抗氧化等作用，其药用功效仍有较大的挖掘空间^[2,4]。在物种鉴别和分子系统发育研究中，苹婆及苹婆属的研究均较少。苹婆与假苹婆外形相似，都属于高大乔木，有学者通过外观形态、显微特征^[7]和 DNA 条形码^[6]进行鉴别分析，本研究在相关基础上补充了苹婆叶绿体基因组信息。黎若竹等^[13]和 Wang 等^[30]在探究锦葵科及其亚科分类争议中涉及到苹婆的叶绿体基因组，但选择的苹婆属植物较少，且未全面对苹婆的叶绿体基因组进行分析，更多的是聚焦于科及亚科的系统发育关系。由于公共数据库中除假苹婆外，西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆植物数据均未被注释，本研究提供了苹婆新的测序数据并在全面分析苹婆的叶绿体基因组信息的同时，完善了西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆这 4 种苹婆属植物的叶绿体基因组注释信息，对苹婆、假苹婆、西蜀苹婆、小花苹婆、粉苹婆和短柄苹婆进行系统发育分析，从而能获得更清晰的苹婆属植物种间亲缘关系。

本研究中，新测序获得的苹婆叶绿体基因组具有典型的四分体结构，具有 LSC 区、SSC 区和 2 个 IR 区，与大多数被子植物的叶绿体基因组结构相似^[31]，体现出叶绿体基因组的保守性和遗传稳定

性。序列全长与黎若竹^[13]、Wang 等^[30]研究相比，长度稍短，为 160 178 bp。在基因注释方面，黎若竹等^[13]注释得到 114 个基因，而本研究总共注释得到 130 个功能基因，并且是首次对苹婆进行重复序列和散在重复序列分析，总共生成 70 条散在重复序列，获得 109 个 SSR 位点，以单核苷酸 SSR 的 A 重复类型最多，这些重复序列可用于苹婆的分子标记开发。此外，本研究是首次对苹婆的密码子进行分析，发现有 30 条序列的 RSCU 大于 1，使用频率较高，其中有 29 条序列是以 A 或 U 结尾，说明苹婆叶绿体基因组密码子偏好性多以 A 或 U 结尾。通过 mVISTA 进行苹婆属叶绿体基因组的多序列比对显示，6 种苹婆属具有一定相似性，LSC 区的差异较大，即在 *matK-rps16*、*trnE-UUC-trnT-GGU* 等 9 个基因间区具有不同程度的差异，Pi 分析得到 4 个高变异区域，其中 LSC 区的 *trnL-UAA-trnF-GAA* 位于 *rps4-ndhJ* 之间，与多序列比对结果有重合，提示这个片段很可能有是有利于开发苹婆属植物分子标记的位点。IR 边界收缩与扩张分析显示，苹婆属 IR 边界的基因种类相同，但有不同程度的收缩和扩张，这可能是导致本研究中 6 种苹婆属叶绿体基因组长度差异的原因。在系统发育分析中，黎若竹等^[13]只涉及到短柄苹婆、假苹婆和苹婆这 3 类苹婆属植物，Wang 等^[30]则只涉及到苹婆与假苹婆 2 个物种，相比之下，本研究涉及到的苹婆属物种更多，结果显示假苹婆和粉苹婆聚为一支，亲缘关系更近，短柄苹婆和苹婆相对独立，与西蜀苹婆和小花苹婆共同处于同一分支，一定程度上揭示了苹婆属植物的种间亲缘关系。

苹婆属作为梧桐科的模式属，在梧桐科的物种研究中占有重要地位。本研究得到的苹婆及苹婆属的叶绿体基因组相关信息和系统发育关系，可用于苹婆属的物种鉴定和遗传关系研究，同时将有助于梧桐科和锦葵科相关的争议探究。

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 (第四十九卷, 第二分册) [M]. 北京: 科学出版社, 1984: 122-123.
- [2] 张鹏, 花燕莹, 陈欣荣, 等. 苹婆果壳总酚酸提取工艺优化及抗氧化性研究 [J]. 中国野生植物资源, 2021, 40(3): 20-27.
- [3] 杨玉娇. 凤眼果壳酚类化合物的鉴定及抑制淀粉消化

- 酶和调控肠道葡萄糖吸收的机制研究 [D]. 广州: 华南理工大学, 2022.
- [4] 任惠, 周婧, 李一伟, 等. 苹婆种子营养及抗氧化活性 [J]. 植物科学学报, 2013, 31(2): 203-208.
- [5] 张敏. “罗晃子”名实考辨 [J]. 中国中药杂志, 1991(9): 518-519.
- [6] 彭晓祺, 吴文如, 来慧丽, 等. 凤眼果的生药学研究 [J]. 广西植物, 2024, 44(6): 1028-1041.
- [7] 苏雨苗, 陈诗敏, 杜勤, 等. 苹婆与假苹婆的比较鉴别研究 [J]. 广州中医药大学学报, 2019, 36(3): 414-419.
- [8] Rao K S, Rajput K S, Srinivas T. Comparative structure of vascular cambium and its derivatives in some species of *Sterculia* [J]. *IAWA J*, 1996, 17(3): 311-318.
- [9] 解新明, 张寿洲, 李勇, 等. 梧桐科分类学研究评述 [J]. 广西植物, 2003, 23(4): 311-317.
- [10] Chattaway M. The wood of the Sterculiaceae I. specialisation of the vertical wood parenchyma within the sub-family sterculieae [J]. *N Phytol*, 1932, 31(2): 119-132.
- [11] Cay S B, Cinar Y U, Kuralay S C, et al. Genome skimming approach reveals the gene arrangements in the chloroplast genomes of the highly endangered *Crocus* L. species: *Crocus istanbulensis* (B.Mathew) Rukšāns [J]. *PLoS One*, 2022, 17(6): e0269747.
- [12] Song Y, Zhao W J, Xu J, et al. Chloroplast genome evolution and species identification of *Styrax* (Styracaceae) [J]. *Biomed Res Int*, 2022, 2022: 5364094.
- [13] 黎若竹, 蔡杰, 杨俊波, 等. 利用叶绿体基因组数据解析锦葵科梧桐亚科的系统位置和属间关系 [J]. 广西植物, 2022, 42(1): 25-38.
- [14] Doyle J, Doyle J L, Doyle J, et al. A rapid DNA isolation procedure for small amounts of fresh leaf tissue [J]. *Phytochem Bull*, 1987, 19:11-15
- [15] Jin J J, Yu W B, Yang J B, et al. GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes [J]. *Genome Biol*, 2020, 21(1): 241.
- [16] Shi L C, Chen H M, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. *Nucleic Acids Res*, 2019, 47(W1): W65-W73.
- [17] Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes [J]. *Curr Genet*, 2007, 52(5/6): 267-274.
- [18] Beier S, Thiel T, Münch T, et al. MISA-web: A web server for microsatellite prediction[J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [19] Kurtz S, Choudhuri J V, Ohlebusch E, et al. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Res*, 2001, 29(22): 4633-4642.
- [20] Wu J, Zhang J, Guo X H, et al. Comprehensive analysis of complete chloroplast genome sequence of *Plantago asiatica* L. (Plantaginaceae) [J]. *Plant Signal Behav*, 2023, 18(1): 2163345.
- [21] Frazer K A, Pachter L, Poliakov A, et al. VISTA: Computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, 32(Web Server issue): W273-W279.
- [22] Darling A C, Mau B, Blattner F R, et al. Mauve: Multiple alignment of conserved genomic sequence with rearrangements [J]. *Genome Res*, 2004, 14(7): 1394-1403.
- [23] Li H E, Guo Q Q, Xu L, et al. CPJSDraw: Analysis and visualization of junction sites of chloroplast genomes [J]. *PeerJ*, 2023, 11: e15326.
- [24] 刘心雨, 姜丹, 杨楚楚, 等. 基于 ITS 条形码的滇鸡血藤及混伪品分子鉴定 [J]. 中国现代中药, 2024, 26(01): 10-17.
- [25] Katoh K, Misawa K, Kuma K, et al. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30(14): 3059-3066.
- [26] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J C, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. *Mol Biol Evol*, 2017, 34(12): 3299-3302.
- [27] Capella-Gutiérrez S, Silla-Martínez J M, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses [J]. *Bioinformatics*, 2009, 25(15): 1972-1973.
- [28] Minh B Q, Schmidt H A, Chernomor O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era [J]. *Mol Biol Evol*, 2020, 37(5): 1530-1534.
- [29] 陆奇丰, 骆文华, 黄至欢. 两种梧桐叶绿体基因组密码子使用偏性分析 [J]. 广西植物, 2020, 40(2): 173-183.
- [30] Wang J H, Moore M J, Wang H X, et al. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies [J]. *Gene*, 2021, 765: 145103.
- [31] 周军辉, 高晓敏, 李莎莎, 等. 细辛叶绿体基因组结构特征与系统进化分析 [J]. 中草药, 2023, 54(24): 8191-8199.

[责任编辑 时圣明]