

DOI: 10.11931/guihaia.gxzw202207030

张梦洁, 杨秀瑶, 尹拓, 等, 2023. 轮叶蒲桃叶绿体基因组特征及系统发育分析 [J]. 广西植物, 43(10): 1849–1860.

ZHANG MJ, YANG XY, YIN T, et al., 2023. Analysis of chloroplast genomic characteristics and phylogeny in *Syzygium grijsii* [J]. *Guihaia*, 43(10): 1849–1860.



## 轮叶蒲桃叶绿体基因组特征及系统发育分析

张梦洁<sup>1</sup>, 杨秀瑶<sup>1</sup>, 尹拓<sup>2</sup>, 韩沛辰<sup>2</sup>, 杜朝金<sup>2</sup>, 张汉尧<sup>1\*</sup>

( 1. 西南林业大学 林学院, 西南地区生物多样性保育国家林业与草原局重点实验室, 昆明 650224;

2. 西南林业大学 林学院, 云南省高校林木遗传改良与繁育重点实验室, 昆明 650224 )

**摘要:** 轮叶蒲桃 (*Syzygium grijsii*) 系桃金娘科 (Myrtaceae) 蒲桃属 (*Syzygium*) 常绿灌木, 其开发前景较好, 但其叶绿体基因组特征及系统发育关系尚未有相关报道。为弥补轮叶蒲桃基因组学方面的空缺, 该文对轮叶蒲桃的叶绿体基因组进行了系统的研究。运用 Illumina 高通量测序, 并在 GetOrganelle 平台进行完整组装, 同时利用组装好的数据分析轮叶蒲桃叶绿体基因组的结构特征和系统发育关系, 其中包括轮叶蒲桃叶绿体基因组结构、功能及特征、密码子偏好性分析、叶绿体基因组的比较分析和系统发育的分析。结果表明: (1) 轮叶蒲桃叶绿体基因组大小为 158 591 bp, 包含 129 个基因。其中, rRNA 基因 8 个, tRNA 基因 37 个, 蛋白编码基因 84 个。分析检测到 39 个重复序列和 84 个 SSR 位点。(2) 密码子偏好性分析发现轮叶蒲桃叶绿体基因组中末端存在对 A/U 的偏性, 使用最多的是编码亮氨酸的密码子。(3) 与近缘种比较, 轮叶蒲桃的边界长度保守, 边界处的基因种类与多个蒲桃属物种相似; 轮叶蒲桃叶绿体基因组在 LSC 和 SSC 区变异度较大, 有 45 处  $0.010 < P_i < 0.015$ , 核苷酸多样性水平高。(4) 系统发育分析表明, 轮叶蒲桃与蒲桃 (*S. jambos*)、滇边蒲桃 (*S. forrestii*) 和乌墨 (*S. cumini*) 亲缘关系最近, 其次是马六甲蒲桃 (*S. malaccense*)。因此, 该文结果表明轮叶蒲桃的叶绿体基因组结构保守, 特征明显, 与同属物种相比, 其序列具有相似性, 核苷酸具有多样性, 并且其与多种桃金娘科物种之间存在亲缘关系。该文对轮叶蒲桃叶绿体基因组进行基因组特征和系统发育分析, 为轮叶蒲桃的种质资源鉴定和开发利用奠定理论基础。

**关键词:** 轮叶蒲桃, 叶绿体基因组, 系统发育关系, 基因组特征, 比较分析

中图分类号: Q943 文献标识码: A 文章编号: 1000-3142(2023)10-1849-12

## Analysis of chloroplast genomic characteristics and phylogeny in *Syzygium grijsii*

ZHANG Mengjie<sup>1</sup>, YANG Xiuyao<sup>1</sup>, YIN Tuo<sup>2</sup>, HAN Peichen<sup>2</sup>,  
DU Chaojin<sup>2</sup>, ZHANG Hanyao<sup>1\*</sup>

( 1. Key Laboratory of Biodiversity Conservation of Southwest China, National Forestry and Grassland Administration, Forestry College of Southwest Forestry University, Kunming 650224, China; 2. Key Laboratory of Forest Genetic Improvement and Breeding, Forestry College of Southwest Forestry University, Kunming 650224, China )

收稿日期: 2022-12-04

基金项目: 国家自然科学基金项目 (32160556)。

第一作者: 张梦洁 (1999-), 硕士研究生, 研究方向为林木遗传育种, (E-mail)1481436191@qq.com。

\*通信作者: 张汉尧, 教授, 博士生导师, 主要从事林木和酵母遗传育种研究, (E-mail)hanyaoz@163.com。

**Abstract:** *Syzygium grijsii*, an evergreen shrub of Myrtaceae and *Syzygium*, has good prospects for exploitation, but its chloroplast genome characteristics and phylogenetic relationships have not been reported. To fill the gap in the genomics of *S. grijsii*, a systematic study of the chloroplast genome of *S. grijsii* was conducted in this paper. Illumina high-throughput sequencing was used to perform complete assembly in the GetOrganelle platform, and the assembled data were also used to analyze the structural features and phylogenetic relationships of the *S. grijsii* chloroplast genome, including the structure, function and characteristics of the *S. grijsii* chloroplast genome, codon preference analysis, comparative analysis of the chloroplast genome and phylogenetic analyses. The results were as follows: (1) The chloroplast genome of *S. grijsii* was 158 591 bp in size, which contained 129 genes. Among them, 8 were rRNA genes, 37 were tRNA genes and 84 were protein-coding genes. The analysis detected 39 repetitive sequences and 84 SSR loci. (2) Codon preference analysis revealed that there was a bias towards A/U at the ends of the *S. grijsii* chloroplast genome, and the most used codon was the one encoding leucine. (3) In comparison with closely related species, the boundary length of *S. grijsii* was conserved, and the gene species at the boundary were similar to several *Syzygium* species; the *S. grijsii* chloroplast genome had a high degree of variability in the LSC and SSC regions, with 45 sites of  $0.010 < P_i < 0.015$  and a high level of nucleotide diversity. (4) Phylogenetic analysis showed that *S. grijsii* was most closely related to *S. jambos*, *S. forrestii* and *S. cumini*, followed by *S. malaccense*. The study concluded that *S. grijsii* had a conserved chloroplast genome structure with distinct features, sequence similarity and nucleotide diversity compared to *Syzygium* species, and affinity between it and various Myrtaceae species. In this paper, the genomic characterization and phylogenetic analysis of the chloroplast genome of *S. grijsii* are conducted to lay the theoretical foundation for the identification and exploitation of the germplasm resources of *S. grijsii*.

**Key words:** *Syzygium grijsii*, chloroplast genome, phylogenetic relationship, genomic characteristics, comparative analysis

叶绿体(chloroplast, cp)是绿色植物进行光合作用的场所,也是绿色植物所特有的细胞器。叶绿体具有相对自主的遗传系统,有自己的遗传物质,能够进行半自主的复制(Xue et al., 2019)。在高等植物中,大多数植物的叶绿体基因组包括一个大单拷贝区(large single copy, LSC)、一个小单拷贝区(small single copy, SSC)和两个反向的重复区(inverted repeats, IRs)(Li et al., 2013)。叶绿体基因组按功能可分为参与叶绿体生物合成的基因、参与叶绿体基因表达的基因、参与光合作用的基因和一些功能未知的开放阅读框(孙志轩等,2022)。叶绿体拥有简单的基因组结构,它包含大量的遗传信息,基因组内碱基交换的速度处于中间水平,分子进化速度在编码区和非编码区的差别很大,所以叶绿体基因组也被广泛利用在系统发育推断和群体遗传学等领域(赵玉芬等,2022)。近年来,随着高通量测序技术越发成熟,近6 000个物种的叶绿体基因组成功被组装和注释。目前,叶绿体基因组分析已经涵盖了大部分被子植物,例如姜科(Zingiberaceae)(Liew et al., 2022)、兰科(Orchidaceae)(Xie et al., 2022)、石竹

属(*Dianthus*)(Lin et al., 2022)等,这些分析为研究物种变异、物种遗传和系统发育关系等方面做出了卓越的贡献。

桃金娘科(Myrtaceae)主要分布于热带和近热带地区,多为乔木或者灌木,约有150属5 600种,我国现有近100种。蒲桃属(*Syzygium*)是桃金娘科下的一个重要属,全世界约有500种,中国约有72种(陈介,2007)。因为蒲桃属植物外形美观,抗逆性强,所以大多用于城市绿化,但随着近年研究的深入,蒲桃属中的轮叶蒲桃、蒲桃和洋蒲桃等也被发掘出了药用价值。轮叶蒲桃根和叶可用于祛风散寒,活血化瘀,治疗跌打损伤、伤风感冒和风湿性头痛,具有巨大的开发利用价值(罗德志等,2020)。目前,在蒲桃属中,仅有8个物种通过高通量测序获得了完整的叶绿体基因组,学者们对蒲桃属的主要研究重点包括植物化学成分的分 离鉴定以及药物研发等方面。一些学者也探讨过桃金娘科的系统发育关系,但只停留于系统发育分析(王雪芹等,2021)。在蒲桃属中,仅有肖蒲桃的叶绿体基因组被详细分析过,但包括轮叶蒲桃在内的其余物种的叶绿体基因组特征和系统发育

相关分析等方面的研究成果则尚未见详细报告。

本文首次对轮叶蒲桃进行测序,并对测序结果进行组装分析。分析内容涵盖了基因组结构、功能及特征、重复序列、SSR 分析和系统发育分析等多方面,并且首次对轮叶蒲桃及其亲缘物种进行对比分析,对其物种间遗传变异做出了阐述。轮叶蒲桃开发前景较好,叶绿体基因组资源可以为其物种进化提供有效信息,但轮叶蒲桃叶绿体基因组特征和相关系统发育关系尚未有相关报道,影响了其进一步开发利用。本研究通过使用高通量测序技术对轮叶蒲桃的叶绿体基因组进行解码,并以已发表的密切相关物种肖蒲桃的叶绿体基因组序列作为参考进行拼接,得到完整的轮叶蒲桃叶绿体基因组,并在此基础上进行分析,解析轮叶蒲桃与其他桃金娘科植物的系统发育关系,为轮叶蒲桃在个体发育过程中存在的遗传和变异问题及其开发和系统研究提供理论基础。

## 1 材料与amp;方法

### 1.1 实验材料、测序和组装

轮叶蒲桃材料采集于福建龙海,用液氮冷冻处理后交与安诺优达生物技术有限公司,在 Illumina NovaSeq 6 000 平台上测序(中国浙江),测序共获得 4.8 G 数据量,测序读长为 PE 150。通过 GetOrganelle(v 1.6.2)进行数据组装(Jin et al., 2020)。组装后用在线软件 CPGAVAS 2 (<http://47.96.249.172:16019/analyzer/view>)进行注释(Liu et al., 2012),使用 Geneious(v 9.0.2)进行人工校正。测序产生的原始数据(Submission ID: SUB11257174; BioProject ID: PRJNA826323)和组装完成的叶绿体基因组信息(GenBank accession number: ON243764)均上传至 NCBI 数据库。

### 1.2 叶绿体基因组结构、功能分析

通过在线软件 CPGAVAS2 (<http://47.96.249.172:16019/analyzer/view>)注释轮叶蒲桃叶绿体基因组后,根据注释信息,用 Geneious(v 9.0.2)对轮叶蒲桃叶绿体基因组进行基因组组成和基因功能分析(穆雨农,2020)。通过在线软件 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>)绘制轮叶蒲桃叶绿体基因组的物理图谱(Lohse et al., 2009)。

### 1.3 重复序列和 SSR 位点分析

使用在线软件 REPuter(Kurtz et al., 2001) (<https://bibiserv.cebitec.uni-bielefeld>)对轮叶蒲桃叶绿体基因组重复序列进行分析,其中包括同向、反向、互补和回文重复。软件的参数设置如下:重复序列数目上限为 1 000,最小重复长度为 30,汉明距离为 3(表示一对重复序列的相似度不能小于 90%),编辑距离为默认参数。

通过在线软件 MISA(v 2.1)检测轮叶蒲桃叶绿体基因组的简单重复序列(Beier et al., 2017)。参数设置如下:单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的重复次数分别设置为 10、5、4、3、3、3,其余均为默认。得到的结果文件用 Krait(v1.3.3)软件进行手动修正(Du et al., 2018)。

### 1.4 密码子偏好性分析

通过 CodonW(v 1.4.2)软件对轮叶蒲桃叶绿体基因组进行密码子偏好性分析。为了减少误差,本研究选择的蛋白质编码区长度 $\geq 300$  bp,并且计算出了轮叶蒲桃叶绿体蛋白编码基因的相对同义密码子使用度 RSCU 值(relative synonymous codon usage),参数均为初始默认值。

### 1.5 基因组的比较分析

使用在线网站 JSHYCloud (<http://cloud.genepioneer.com:9929>)比较轮叶蒲桃与 4 个近缘物种,即白果蒲桃(*Syzygium album*)、马六甲蒲桃(*S. malaccense*)、香蒲桃(*S. odoratum*)和肖蒲桃(*S. acuminatissimum*)的 IR 边界(LSC/IRb、IRb/SSC、SSC/IRa 和 IRa/LSC)信息。

通过在线软件 mVISTA (<https://genome.lbl.gov/vista/index.shtml>)中的 Shuffle-LAGAN 模型比较轮叶蒲桃与以上 4 个近缘物种基因组序列差异(Frazer et al., 2004)。

通过 Geneious(v 9.0.2)对轮叶蒲桃及 4 个近缘种进行 MAFFT 比对(陈韵等,2017),再使用 DnaSP(DNA Sequence Polymorphism)进行核酸变异度分析(Librado & Rozas, 2009),滑动窗口长度为 600 sites,步长为 200 sites,其余所有参数均为默认,得出  $P_i$  值。

### 1.6 系统发育分析

为阐明轮叶蒲桃与桃金娘科其他物种的系统发育关系,从 GenBank 数据库(<https://www.ncbi.nlm.nih.gov/nucleotide>)中下载了桃金娘科 22 个

物种的叶绿体基因组序列,采用 Geneious(v 9.0.2) 软件(MAFFT 插件)比对 23 种叶绿体基因组全序列。使用 RAxML(v 8.2.12)构建最大似然进化树,最佳模型为 GTR+F+R2,bootstrap 值设置为 1 000,以推断各节点的支持率。

## 2 结果与分析

### 2.1 叶绿体基因组结构、功能及特征

轮叶蒲桃叶绿体基因组全长 158 591 bp(图 1),共编码 129 个基因,其中包括 8 个 rRNA 基因,37 个 tRNA 基因,84 个蛋白编码基因。轮叶蒲桃叶绿体基因组为环状。其中,大单拷贝区的长度为 88 046 bp,GC 含量为 34.8%;小单拷贝区长度为 18 409 bp,GC 含量为 30.9%;单个反向重复区的长度为 26 068 bp,GC 含量为 42.7%,明显高于 LSC 区和 SSC 区。在轮叶蒲桃叶绿体基因组中,总 GC 含量为 37.0%,低于 AT(63.0%)含量。由图 2 可知,白色、灰色和黑色区域为外显子,箭头表示基因的方向。轮叶蒲桃包含了 3 个外显子,其中两个外显子因在反向重复区而分别具有 2 个拷贝,5'端位于 LSC 区,3'端位于 IR 区。此外,轮叶蒲桃叶绿体基因组中有 22 个含有内含子的基因,其中在 19 个基因中发现 1 个内含子,另外 3 个基因中发现有两个内含子(表 1)。其中,*rps12* 作为一种反式剪切基因。

### 2.2 重复序列和 SSR 分析

在轮叶蒲桃叶绿体基因组中可以检测到 39 个大小为 30~45 bp 的重复序列,包含了 1 个反向重复,19 个回文重复及 19 个同向重复,未检测到互补重复(图 3)。这些重复序列分布广泛,有 18 个(包含 9 个同向重复,9 个回文重复)位于 *ycf2*、*psaB*、*psaA*、*trnS-GCU*、*trnS-UGA*、*ycf1* 基因上,9 个跨区域,8 个位于基因间区(IGS, intergenic spacer),4 个位于内含子区域,4 个最长的重复序列(45 bp)全部位于 *ycf2* 基因上。

在轮叶蒲桃叶绿体基因组中共检测到 84 个 SSR 位点,平均长度(SSR 总长度/SSR 总计数)为 11.34 bp。其中,单核苷酸 60 个,二核苷酸 2 个,三核苷酸 8 个,四核苷酸 13 个,五核苷酸 1 个。此外,轮叶蒲桃叶绿体基因组中大部分 SSR 位点位于 IGS 区域(64 个),少部分位于基因上(20 个)。叶绿体中两个二核苷酸(TA、TC),长度均为 10

bp,三核苷酸的长度均为 12 bp,四核苷酸重复长度仅有一个为 16 bp,其余均为 12 bp,唯一的五核苷酸长度为 15 bp。

### 2.3 密码子偏好性分析

对轮叶蒲桃叶绿体基因组的 41 个大于 300 bp 的蛋白质编码序列进行密码子偏好性分析(表 2)。轮叶蒲桃叶绿体基因密码子中,RSCU 值>1 的密码子有 30 个,RSCU 最高值为 1.98,为亮氨酸中的 UUA,最低值为 0.33,为亮氨酸中的 CUG。其中,有 29 个以 A/U 结尾,说明轮叶蒲桃叶绿体基因组密码子更偏向以 A/U 结尾。轮叶蒲桃叶绿体基因的密码子总数为 17 506 个,其中 1 846 个(10.54%)为编码亮氨酸的密码子,仅有 195 个(1.11%)密码子编码半胱氨酸,表明在轮叶蒲桃叶绿体基因组中使用最多的是亮氨酸,使用最少的是半胱氨酸。在所有密码子中,异亮氨酸中的 AUU 使用次数最多,次数为 786,使用次数最少的是半胱氨酸中的 UGC,次数为 44。

### 2.4 叶绿体基因组的比较分析

通过在线网站 JSHYCloud 比较分析了蒲桃属 5 个物种的 LSC、SSC 和 IR 区的边界位置。结果表明这 5 个物种的 IR 区长度较保守(26 068~26 472 bp)。结果如图 4 所示,5 个物种 LSC/IRb 边界在 *rps19* 基因内部,并且扩张程度均一致。轮叶蒲桃和马六甲蒲桃的 *ndhF* 基因紧挨着 IRb/SSC 边界,而白果蒲桃、香蒲桃和肖蒲桃的 *ndhF* 基因距边界为 53~58 bp。5 个物种的 SSC/IRa 边界都位于 *ycf1* 基因上,轮叶蒲桃和马六甲蒲桃的 *ycf1* 基因朝 SSC 区分别进入了 4 851 bp 和 4 857 bp,朝 IRa 区进入了 750 bp。白果蒲桃、香蒲桃和肖蒲桃均有 4 857 bp 位于 SSC 区、1 188 bp 位于 IRa 区。

图 5 结果显示(灰色箭头表示基因的方向,纵坐标表示从 50%到 100%的一致度;UTR 表示非翻译区,CNS 表示保守非编码序列),以白果蒲桃为参照序列,轮叶蒲桃的基因组与另外 4 种植物叶绿体基因组之间有所区别。在变异较大的 LSC 和 SSC 区,轮叶蒲桃的变异度较另外 4 种而言较大。5 个物种的 IR 区较为保守,LSC 和 SSC 区中 *psbM-psbD*、*psbZ-trnM(CAT)*、*trnM(CAT)-atpE*、*ycf4-psbE*、*ndhF-ccsA* 基因之间存在较大的序列差异。

使用 DnaSP 软件进行分析且经过 MAFFT 比对后的 5 种桃金娘科植物叶绿体基因组序列结果如图 6 所示,X 轴表示核苷酸的位置,Y 轴表示每

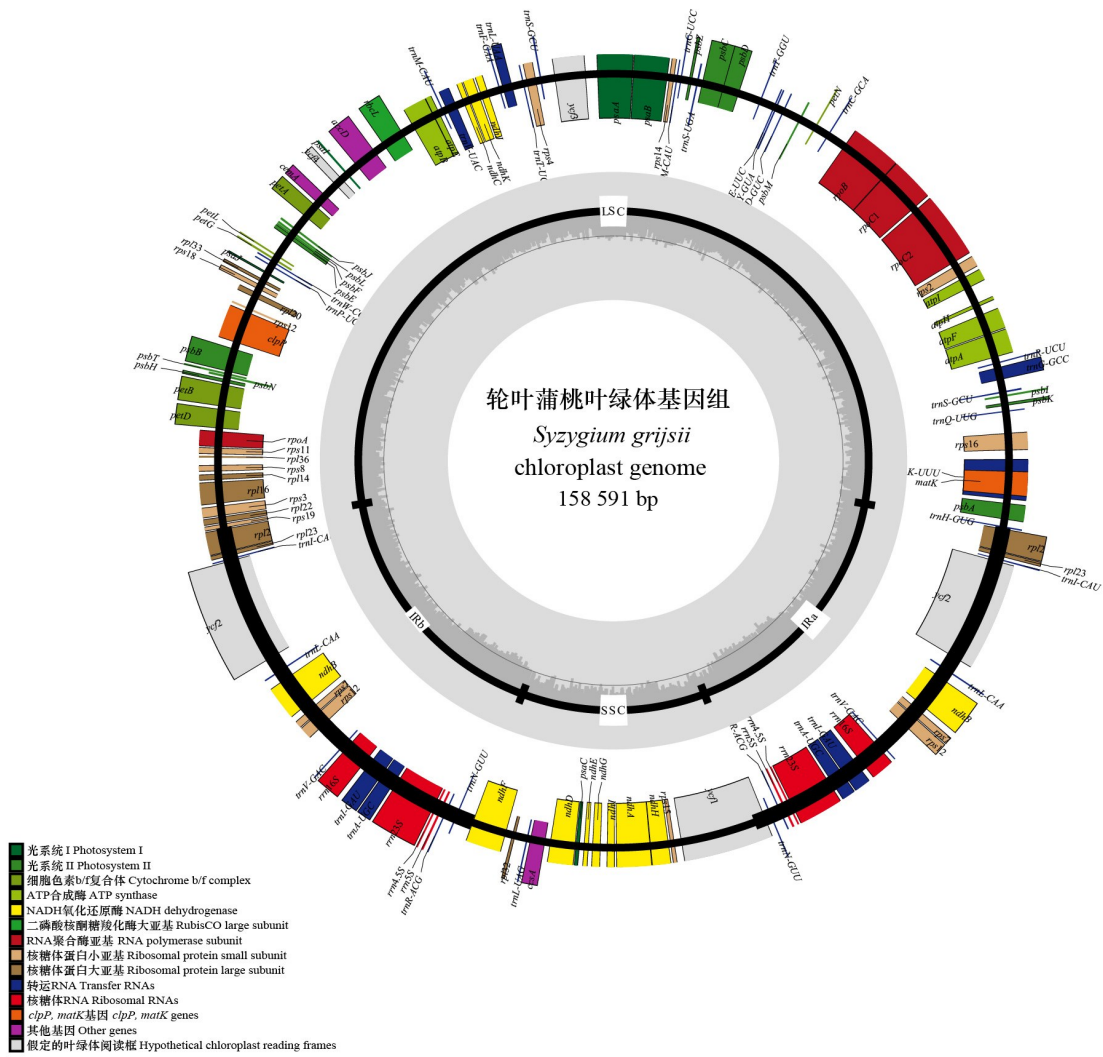


图 1 轮叶蒲桃叶绿体基因组图谱  
Fig. 1 Chloroplast genome map of *Syzygium grijsii*

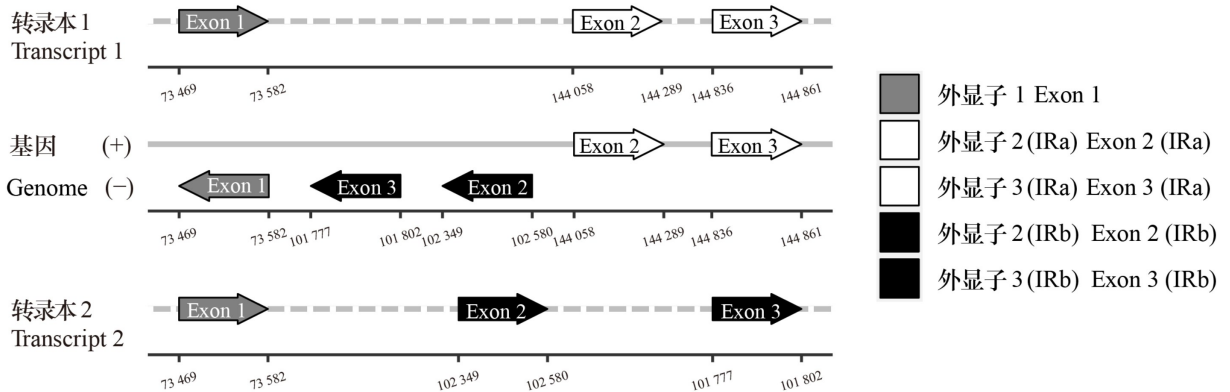


图 2 轮叶蒲桃叶绿体基因组的反式剪切基因组结构示意图  
Fig. 2 Schematic diagram of the *trans*-sheared genome structure in *Syzygium grijsii* chloroplast genome

表 1 轮叶蒲桃叶绿体基因组基因功能总结

Table 1 Summary of chloroplast genome gene functions in *Syzygium grijsii*

基因功能分类 Category for gene	基因分组 Group of gene	基因名称 Name of gene
参与光合作用的基因 Gene involving in photosynthesis	光合系统亚基 Subunit of photosystem	<i>psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	细胞色素 b/f 复合体亚基 Subunit of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>
	NADH-脱氢酶亚基 Subunit of NADH-dehydrogenase	<i>ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	ATP 合成酶亚基 Subunit of ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
自身复制相关基因 Self-replicating related gene	DNA 依赖型 RNA 聚合酶 DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	核糖体蛋白大亚基 Ribosomal protein large subunit	<i>rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36</i>
	核糖体蛋白小亚基 Ribosomal protein small subunit	<i>rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8</i>
	核糖体 RNA Ribosomal RNAs	<i>rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2)</i>
	转运 RNA Transfer RNAs	<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC*, trnG-UCC, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACC(2), trnR-UCU, trnS-GCU(2), trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnYm-CAU</i>
其他功能基因 Other functional gene	乙酰-CoA-羧化酶亚基 Subunit of acetyl-CoA carboxylase	<i>accD</i>
	C 类型细胞色素合成基因 C-type cytochrome synthesis gene	<i>ccsA</i>
	包膜蛋白基因 Envelop membrane protein	<i>cemA</i>
	蛋白酶 Protease	<i>clpP**</i>
	成熟酶 Maturase	<i>matK</i>
未知功能基因 Unknown functional gene	保守开放性阅读框 Conserved open reading frames	<i>yef1, yef2(2), yef3**, yef4</i>

注: \*表示有一个内含子的基因; \*\*表示有两个内含子的基因; 括号内的数字表示多拷贝基因的拷贝数量。

Note: \* indicates the gene with one intron; \*\* indicates the gene with two introns; the number in parentheses indicates number of copies of a multi-copy gene.

个窗口的核苷酸多样性, 可显著发现 IR 区的变异度比 LSC 和 SSC 区更低。  $P_i$  值最高点为 0.015, 位于 LSC 区的 *psbZ-psaB* 基因间区。图中有 2 处  $P_i = 0.015$  的热点, 45 处  $0.010 < P_i < 0.015$ , 可以看出核苷酸多样性水平较高。

## 2.5 系统发育分析

结果如图 7 所示, 分支上的数字为 1 000 次自展支持率。系统发育树中具有 20 个支持率为 100% 的拓扑结构节点, 遗传变异度为 0.01, 可信度比较高。蒲桃属、桉属 (*Eucalyptus*)、伞房桉属

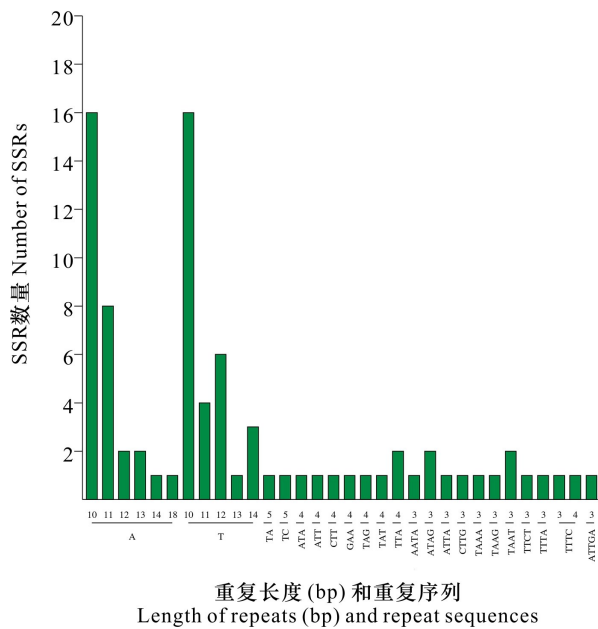


图3 轮叶蒲桃叶绿体基因组中的SSR类型和分布  
Fig. 3 Types and distribution of SSR in *Syzygium grijsii* chloroplast genome

(*Corymbia*) 和杯果木属 (*Angophora*) 亲缘关系较近。在蒲桃属中, 轮叶蒲桃与滇边蒲桃和乌墨亲缘关系最近, 其次是蒲桃。

### 3 讨论与结论

轮叶蒲桃叶绿体基因组全长 158 591 bp, 总 GC 含量为 37.0%, 与已发表的桃金娘科物种叶绿体基因组 (王雪芹等, 2021) 相似。轮叶蒲桃完整叶绿体基因组对 A/T 具有偏好性 (推测可能是因为 A/T 比 G/C 更容易解链) (左文明等, 2019; 赵祺等, 2021)。另外, 轮叶蒲桃完整叶绿体基因组中有 129 个基因被编码, 这些基因中含有蛋白编码基因 84 个和 tRNA 基因 37 个, 与邓演文等 (2021) 的研究结果一致, 表明轮叶蒲桃叶绿体基因组相对保守。可以根据这一性状推测出易变性存在于轮叶蒲桃叶绿体基因组序列, 并且易发生变异。

有研究者发现, 重复序列分析对于分析叶绿体的基因重排、碱基替换、系统发育和基因组进化有重要的作用 (Smith et al., 2002; Nie et al., 2012)。轮叶蒲桃叶绿体基因组中包含了 3 种重

复序列类型, 大部分位于基因上, 其中, 最大的四个重复序列均位于 *ycf2* 基因上。在轮叶蒲桃的叶绿体基因组中检测到 84 个 SSR 位点, 随着基序长度的增加, SSR 的频率降低, 具有多态性。SSR 多态性是由重复单元的伸长或缩短引起, 广泛用于种群遗传多样性或种群分类分析 (Xu et al., 2004; Zhao et al., 2019)。轮叶蒲桃的单核苷酸均为 A/T 型, 与蒲桃属的肖蒲桃一致 (邓演文等, 2021)。由前人研究可知, 从叶绿体基因组中鉴定出的重复序列和 SSR 可以作为未来木材生态学、进化和追踪研究的宝贵标记 (申响保等, 2021; 梁燕妮等, 2022)。

叶绿体基因对于密码子的使用普遍具有偏好性, 而通过密码子的偏好性能反映物种对外界环境的选择压力, 在叶绿体基因组的研究中发挥着重要作用 (Morton et al., 1999; Asaf et al., 2017)。轮叶蒲桃叶绿体基因组中编码亮氨酸的密码子 (1 846, 10.54%) 使用的频率最多, 编码半胱氨酸的密码子 (195, 1.11%) 使用的频率最少。分析表明, 轮叶蒲桃密码子具有很强的 A/U 碱基偏性, 与睡莲、木兰等 (季凯凯等, 2020; 黄祥等, 2022) 被子植物一致。这说明在被子植物间密码子偏好性具有相似性, 这种相似性的产生或许与自然选择、遗传变异或者其他方面有关系。

虽然蒲桃属植物叶绿体基因组的部分特征与其他被子植物基本一致 (王玲等, 2012), 但是轮叶蒲桃和蒲桃属其他物种之间亲缘关系的远近尚不明确。蒲桃属的 5 个物种边界处的基因种类比较保守。由分析可看出, 轮叶蒲桃与白果蒲桃的边界区域比较相似, 可能是具有较为接近的起源时间 (王一麾等, 2022)。IR 区的收缩和扩张广泛存在于 5 个蒲桃属植物中, 进行进一步的评估可以揭示轮叶蒲桃及相关类群的进化演变过程。

在全序列比对分析中, 参试的 5 个蒲桃属物种间具有较高的相似性, 同时也存在几个变异较大的区域, 这些区域的差异位点可以应用于种质资源的鉴定、DNA 条形码的开发以及后续其他研究 (Kim et al., 2004)。从核苷酸多样性分析可以看出, 变异位点的  $P_i$  值位于 0.00 067 ~ 0.015 之间, 有较高水平的核苷酸多样性, LSC 区和 SSC 区是轮叶蒲桃叶绿体基因组中会发生高频变异的区域, 这些区间可以作为进化分析的区域 (Park et al., 2019), 有助于研究轮叶蒲桃的种间关系。

表 2 轮叶蒲桃叶绿体基因密码子偏好性分析  
Table 2 Analysis of codon preference in *Syzygium griseum* chloroplast genes

氨基酸 Amino acid	占比 Ratio (%)	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	占比 Ratio (%)	密码子 Codon	数量 Number	RSCU		
亮氨酸 Leucine	10.54	<u>UUA</u>	608	1.98	苯丙氨酸 Phenylalanine	5.43	<u>UUU</u>	615	1.29		
		<u>UUG</u>	364	1.18			<u>UUC</u>	336	0.71		
		<u>CUU</u>	391	1.27	酪氨酸 Tyrosine	3.83	<u>UAU</u>	547	1.63		
		CUC	127	0.41			<u>UAC</u>	123	0.37		
		异亮氨酸 Isoleucine	9.03	CUA	253	0.82	终止密码子 Termination	0.23	<u>UAA</u>	26	1.90
				CUG	103	0.33			<u>UAG</u>	9	0.66
<u>AUU</u>	786			1.49	UGA	6			0.44		
AUC	305			0.58	组氨酸 Histidine	2.46	<u>CAU</u>	336	1.56		
AUA	490			0.93			<u>CAC</u>	95	0.44		
甲硫氨酸 Methionine	2.30	AUG	402	1.00	谷氨酰胺 Glutamine	3.78	<u>CAA</u>	515	1.56		
		缬氨酸 Valine	5.15	<u>GUU</u>			329	1.46	<u>CAG</u>	147	0.44
GUC	97			0.43	天冬酰胺 Asparagine	5.05	<u>AAU</u>	686	1.55		
<u>GUA</u>	347			1.54			<u>AAC</u>	198	0.45		
GUG	129			0.57	赖氨酸 Lysine	5.40	<u>AAA</u>	717	1.52		
丝氨酸 Serine	7.52	<u>UCU</u>	389	1.77			<u>AAG</u>	228	0.48		
		UCC	199	0.91	天冬氨酸 Aspartic acid	4.08	<u>GAU</u>	591	1.66		
		<u>UCA</u>	274	1.25			<u>GAC</u>	123	0.34		
		UCG	130	0.59	谷氨酸 Glutamic acid	5.51	<u>GAA</u>	730	1.51		
		<u>AGU</u>	249	1.13			<u>GAG</u>	234	0.49		
脯氨酸 Proline	4.00	AGC	76	0.35	半胱氨酸 Cysteine	1.11	<u>UGU</u>	151	1.55		
		<u>CCU</u>	279	1.59			<u>UGC</u>	44	0.45		
		CCC	134	0.77	色氨酸 Tryptophan	1.66	UGG	291	1.00		
		<u>CCA</u>	212	1.21			精氨酸 Arginine	6.18	<u>CGU</u>	211	1.17
		CCG	75	0.43					<u>CGC</u>	67	0.37
苏氨酸 Threonine	4.96	<u>ACU</u>	329	1.52	<u>CGA</u>	265	1.47				
		ACC	150	0.69	<u>CGG</u>	67	0.37				
		<u>ACA</u>	288	1.33	<u>AGA</u>	350	1.94				
		ACG	101	0.47	AGG	122	0.68				
丙氨酸 Alanine	5.19	<u>GCU</u>	403	1.77	甘氨酸 Glycine	6.57	<u>GGU</u>	369	1.28		
		GCC	152	0.67			<u>GGC</u>	105	0.36		
		<u>GCA</u>	250	1.10			<u>GGA</u>	500	1.74		
		GCG	104	0.46			GGG	177	0.62		

注：带下划线所对应的密码子表示被高频使用的密码子(RSCU>1)。

Note: The underlined codon indicates a high frequency codon (RSCU>1).

本研究从 NCBI 已公开的植物中选取了包括轮叶蒲桃在内的 23 种植物,通过系统进化分析确定了轮叶蒲桃的进化地位和亲缘关系。轮叶蒲桃、蒲桃、滇边蒲桃和乌墨构成了一个支持率高的单支,与马六甲蒲桃的关系比较密切,亲缘关系较近。在中国植物志分类系统中,轮叶蒲桃、滇边蒲

桃和乌墨属小花蒲桃组,这与本研究中构建的系统发育树一致,表明该文聚类分析结果具有可靠性。虽然蒲桃属物种组成丰富,但是目前所有的系统发育数据较少,轮叶蒲桃系统发育关系的研究弥补了部分空缺,但之后依然需要更加深入的分析。



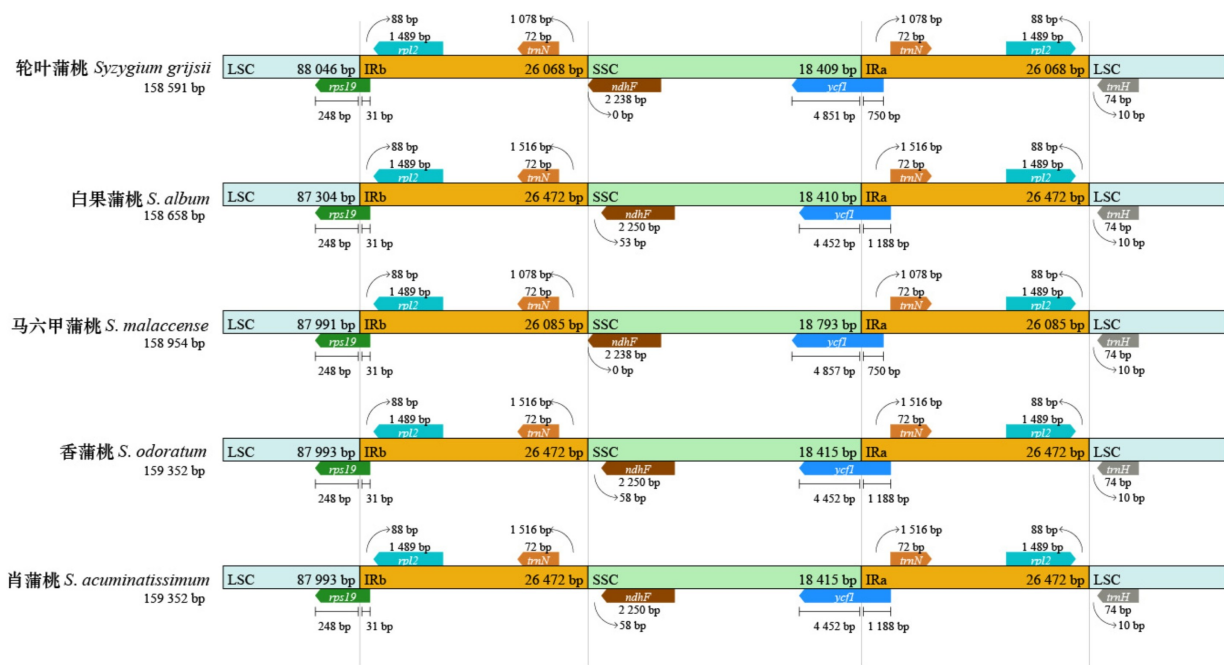


图 4 蒲桃属 5 种植物叶绿体基因组边界分析

Fig. 4 Analysis of chloroplast genome boundaries in five *Syzygium* species

通过高通量测序和组装轮叶蒲桃叶绿体样本,获得了轮叶蒲桃完整的叶绿体基因组。该叶绿体基因组的长度为 158 591 bp,具有典型的环状结构,含有 129 个功能基因。叶绿体基因组含有丰富的 SSR 位点(A/T 作为主要成分),并且以 A/U 结尾的密码子在轮叶蒲桃叶绿体基因组中更容易被使用。通过叶绿体基因组比较分析可以发现,轮叶蒲桃叶绿体基因组的 IR 边界保守,与蒲桃属其他物种间有较高的相似性,核苷酸水平高。系统发育树支持率比较高,物种间亲缘关系较近。轮叶蒲桃叶绿体基因组特征和系统发育研究对于了解蒲桃属的不同物种非常重要,有助于开发该属的植物资源,并为进一步研究提供理论基础。

#### 参考文献:

- ASAF S, WAQAS M, KHAN AL, et al., 2017. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species [J]. *Front Plant Sci*, 8: 304.
- BEIER S, THIEL T, MUNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 33(16): 2583-2585.
- CAVALIER-SMITH T, 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses [J]. *Current Biol*, 12(2): R62-R64.
- CHEN J, 2007. *Flora Reipublicae Popularis Sinicae*: Vol.13 [M]. Beijing: Science Press: 351-352. [陈介, 2007. 中国植物志: 第十三卷 [M]. 北京: 科学出版社: 351-352.]
- CHEN Y, 2017. Comparative analysis of structural organization and evolution of *Chrysanthemum* and *Ajania* chloroplast genomes [D]. Nanjing: Nanjing Agricultural University: 4-16. [陈韵, 2017. 菊属、亚菊属叶绿体 DNA 的基因组结构与进化分析 [D]. 南京: 南京农业大学: 4-16.]
- DENG YW, LIU XZ, ZHUO DL, et al., 2021. Structural characteristics and phylogenetic relationship of chloroplast genome of *Syzygium acuminatissimum* [J]. *Guangdong Agric Sci*, 48(6): 38-46. [邓演文, 刘晓洲, 卓定龙, 等, 2021. 肖蒲桃叶绿体基因组结构特征及系统发育关系 [J]. *广东农业科学*, 48(6): 38-46.]
- DU L, ZHANG C, LIU Q, et al., 2018. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design [J]. *Bioinformatics*, 34(4): 681-683.
- FRAZER KA, PACHTER L, POLIAKOV A, et al., 2004. VISTA: computational tools for comparative genomics [J]. *NuclAcids Res*, 32(2): W273-W279.
- HUANG X, CHU GM, ZHENG XK, et al., 2022. Analysis of codon usage bias and phylogenetic of chloroplast genome in *Nymphaea* [J]. *J Agric Sci Technol*, 24(4): 75-84. [黄祥, 楚光明, 郑新开, 等, 2022. 睡莲属叶绿体基因组密码子偏好性及系统发育分析 [J]. *中国农业科技导报*, 24(4): 75-84.]



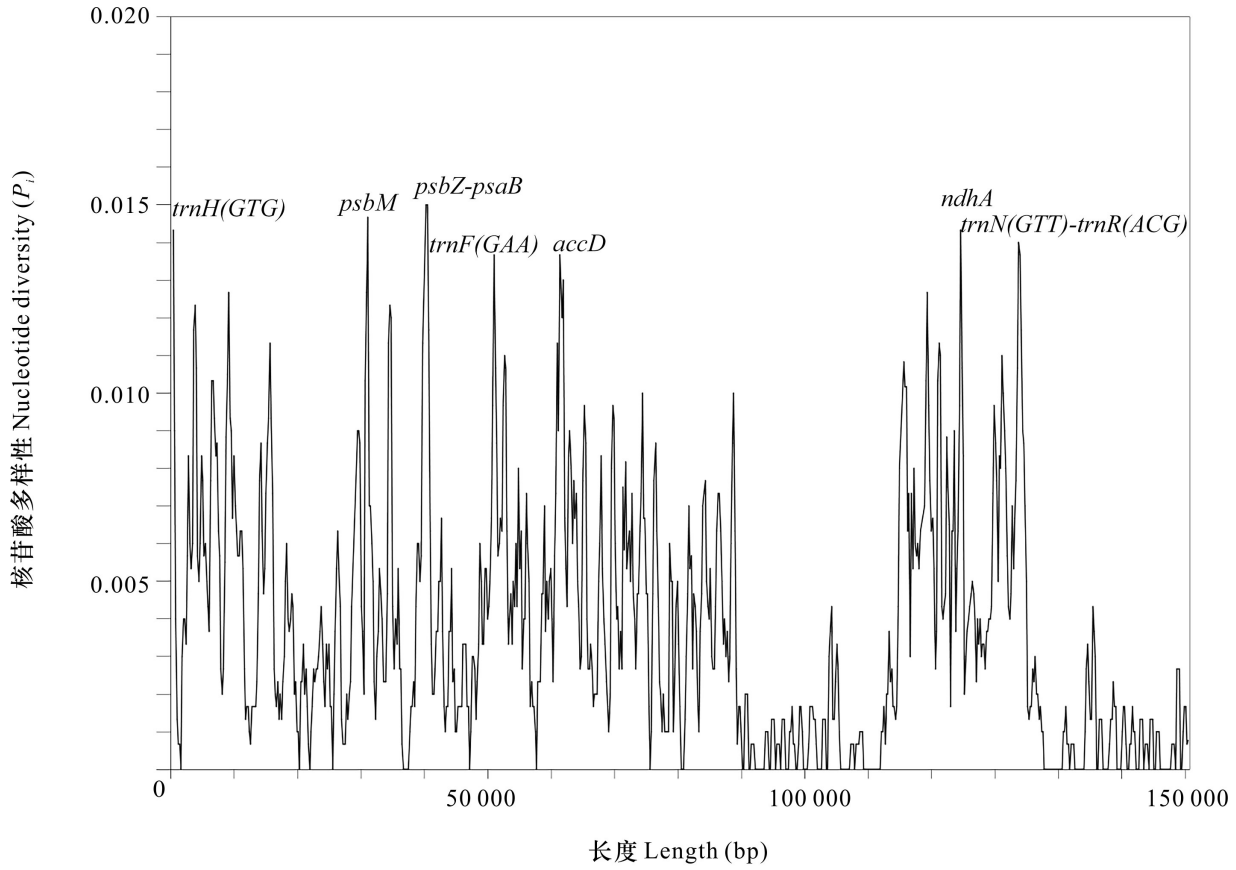


图 6 滑动窗口分析 5 个蒲桃属植物叶绿体基因组核苷酸多样性

Fig. 6 Sliding window analysis of chloroplast genomic nucleotide diversity in five *Syzygium* species

- species based on complete chloroplast genome sequences [J]. *Inter J Mol Sci*, 23(20): 12567.
- LIU C, SHI L, ZHU Y, et al., 2012. CPGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J]. *BMC Genom*, 13: 715.
- LOHSE M, DRECHSEL O, KAHLAU S, et al., 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucl Acids Res*, 41(W1): W575–W581.
- LUO DZ, FU RR, GAN YF, et al., 2020. Chemical constituents of *Syzygium grijsii* [J]. *J Chin Med Mat*, 43(5): 1130–1133. [罗德志, 付容容, 甘宇汾, 等, 2020. 轮叶蒲桃的化学成分研究 [J]. *中药材*, 43(5): 1130–1133.]
- MORTON BR, 1999. Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis* [J]. *Proc Natl Acad Sci*, 96(9): 5123–5128.
- MU YN, 2020. Phylogenetic and taxonomic research of *Hypolepis* in China based on the complete chloroplast genome [D]. Harbin: Harbin Normal University: 17–36. [穆雨农, 2020. 基于叶绿体全基因组对中国姬蕨属的系统分类学研究 [D]. 哈尔滨: 哈尔滨师范大学: 17–36.]
- NIE X, LV S, ZHANG Y, et al., 2012. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*) [J]. *PLoS ONE*, 7(5): e36869.
- PARK I, YANG S, KIM WJ, et al., 2019. Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species [J]. *Molecules*, 24(6): E1038.
- SHEN XB, ZHU YJ, XU GB, 2021. Distribution characteristics of SSR loci and development of molecular markers in *Taxus fuana* [J]. *J Cent S Univ For Technol*, 41(4): 139–147. [申响保, 朱妍洁, 徐刚标, 2021. 密叶红豆杉 SSR 位点分布特征及分子标记开发 [J]. *中南林业科技大学学报*, 41(4): 139–147.]
- SUN ZX, AO PX, BI YF, et al., 2022. Complete chloroplast genome sequence and characteristics analysis of *Medica gosativa* ‘Deqin’ [J]. *Acta Agr Sin*, 30(2): 320–328. [孙志轩, 敖平星, 毕玉芬, 2022. ‘德钦’紫花苜蓿叶绿体基因组序列及特征分析 [J]. *草地学报*, 30(2): 320–328.]
- WANG L, DONG WP, ZHOU SL, 2012. Structural mutations and reorganizations in chloroplast genomes of flowering plants

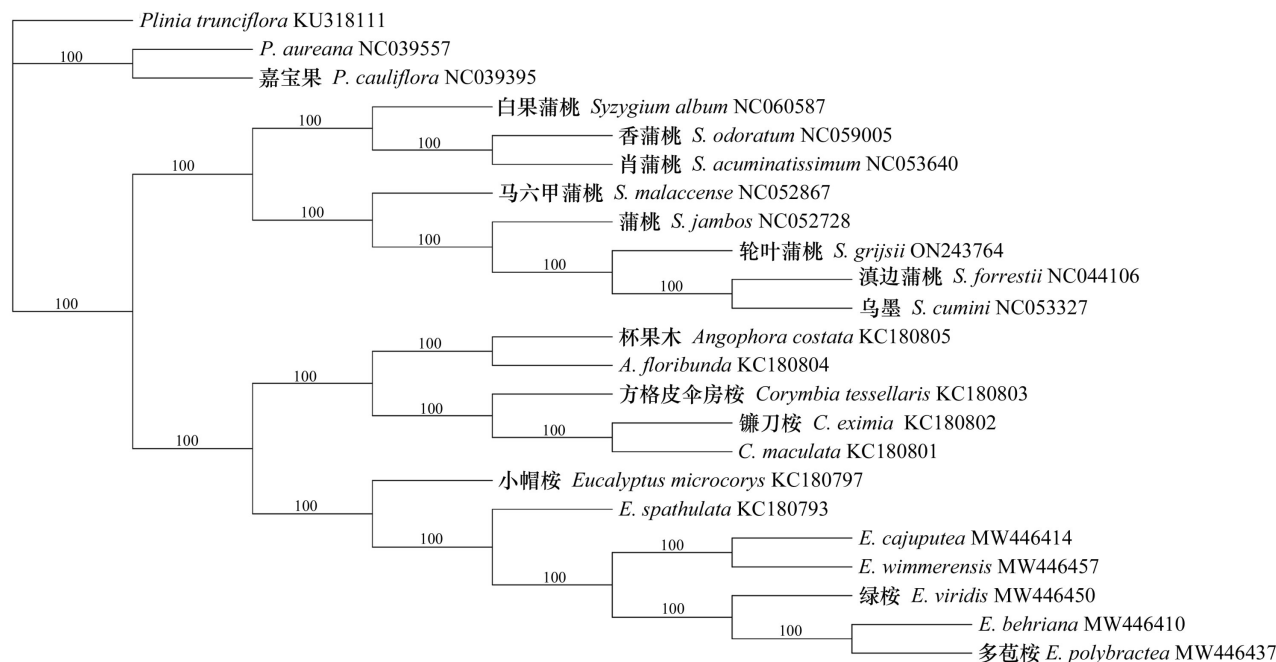


图 7 基于叶绿体基因组序列用最大似然法构建的 23 个桃金娘科物种的系统进化树

Fig. 7 Phylogenetic trees of 23 Myrtaceae species based on chloroplast genome sequences constructed by maximum likelihood (ML) method

[J]. Acta Bot Boreal-Occident Sin, 32(6): 1282–1288.  
[王玲, 董文攀, 周世良, 2012. 被子植物叶绿体基因组的结构变异研究进展 [J]. 西北植物学报, 32(6): 1282–1288.]

WANG XQ, SONG WW, XIAO JJ, et al., 2021. Phylogeny of Myrtales and related groups based on chloroplast genome [J]. Guihaia, 41(1): 68–80. [王雪芹, 宋卫武, 肖建加, 等. 基于叶绿体基因组探讨桃金娘目及其近缘类群的系统发育关系 [J]. 广西植物, 41(1): 68–80.]

WANG YH, XIE YF, ZHANG ZX, et al. The complete chloroplast genome of *Sloanea sinensis* and the systematic status of Elaeocarpaceae [J]. Guihaia, 42(1): 39–48. [王一麾, 谢宜飞, 张志翔, 等, 2022. 猴欢喜叶绿体全基因组及杜英科系统地位分析 [J]. 广西植物, 42(1): 39–48.]

XIE H, ZHANG L, ZHANG C, et al., 2022. Comparative analysis of the complete chloroplast genomes of six threatened subgenus *Gynopodium* (*Magnolia*) species [J]. BMC genomics, 23(1): 1–14.

XU XL, XU LA, HUAN MR, et al., 2004. Genetic diversity of microsatellites (SSRs) of natural populations of *Quercus variabilis* [J]. Hereditas (Beijing), 26(5): 683–688.

XUE S, SHI T, LUO W, et al., 2019. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina* [J]. Horticult Res, 6(1): 626–638.

ZHAO KK, LANDREIN S, BARRETT R, et al., 2019. Phylogeographic analysis and genetic structure of an endemic Sino-Japanese disjunctive genus *Diabelia* (Caprifoliaceae) [J]. Front Plant Sci, 10: 913.

ZHAO Q, YU JX, QIN YW, et al., 2021. Assembly and sequence analysis of the chloroplast genome of *Championella sarcorrhiza* based on high-throughput sequencing [J]. Chin Tradit Herb Drugs, 52(6): 1744–1750. [赵祺, 余佳兴, 秦宇雯, 等, 2021. 基于高通量测序的菜头肾叶绿体基因组的组装及序列分析 [J]. 中草药, 52(6): 1744–1750.]

ZHAO YF, 2022. Advances in the application of chloroplast genome in botanical research [J]. Biol Teach, 47(3): 83–85. [赵玉芬, 2022. 叶绿体基因组在植物学研究中的应用进展 [J]. 生物学教学, 47(3): 83–85.]

ZUO WM, ZENG Y, YANG CF, et al., 2019. High-throughput sequencing of the complete chloroplast genome of *Rheum tanguticum* and its application in species identification [J]. Chin Tradit Herb Drugs, 50(22): 5545–5553. [左文明, 曾阳, 杨春芳, 等, 2019. 基于高通量技术的唐古特大黄叶绿体全基因组测序及应用研究 [J]. 中草药, 50(22): 5545–5553.]

(责任编辑 周翠鸣)