

DOI: 10. 11931/guihaia. gxzw202402031

不同来源广藿香叶绿体基因组研究

行冰楠, 梁莹莹, 吴文如*, 陆亚茹, 邹何元, 彭晓祺

(广州中医药大学 中药学院, 广州 510006)

摘要: 为了研究广藿香 (*Pogostemon cablin*) 叶绿体基因组的结构特征, 比较不同来源广藿香叶绿体基因组的差异, 该研究利用 DNBSeg 测序平台对广藿香全基因组进行测序, 使用 getOrganelle 组装完整的广藿香叶绿体基因组, 通过 CPDAS2 网站对该叶绿体基因组进行注释, 对不同来源的广藿香叶绿体基因组进行基本结构特征分析、IR/SC 的边界比较、基因组比较及共线性分析、简单重复序列及散在重复序列、多态性分析和同义密码子相对使用度分析。结果表明: (1) 20 个不同来源广藿香叶绿体基因组全长 152 461~152 510 bp, 注释得到 132 个基因, 其中包含 87 个 CDS, 37 个 tRNA 基因和 8 个 rRNA 基因。(2)mVISTA 比对发现 *atpF*、*atpF-atpH*、*rps16-trnQ_UUG*、*rpoB-trnC_GCA*、*accD*、*psaI-ycf4*、*petA-psbJ*、*rpl16*、*rps15-ycf1* 为高变区。(3) 核酸多样性大于 0.002 的位点位于 *trnM-CAU-atpB* 间隔区、*ycf4*、*rpl32*、*rpl32-trnL-UAC* 间隔区。(4) 检测到共有 64 种密码子编码 20 个氨基酸, 偏好性较强的密码子有 33 个, 其中以 A/U 结尾的密码子占大多数。(5) 检测出 74~76 个 SSR 及 15~18 个回文重复序列、12~17 个正向重复序列。(6) 经遗传距离分析与系统发育分析, 仅 GSY_MLXY 与其他栽培类型的亲缘关系较远。该研究获得了 20 个不同来源广藿香叶绿体基因组结构信息以及差异位点, 为分子标记的开发及优良种质筛选提供基础资料。

关键词: 广藿香, 叶绿体基因组, 重复序列, 多态性, 同义密码子相对使用度

中图分类号: **文献标识码:** A **文章编号:**

Analysis of chloroplast genome of *Pogostemon cablin* of different origins

XING Bingnan, LIANG Yingying, WU Wenru*, LU Yaru, ZOU Heyuan, PENG Xiaoqi

(School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine,

Guangzhou 510006, China)

Abstract: *Pogostemon cablin* (Blanco) Benth. possesses significant medical and industrial values, as it can be used for medicinal purposes as well as for essential oil extraction. However, the yield and quality of *P. cablin* can vary depending on the ecological environment and artificial cultivation measures employed in different production regions and origins. In order to study the structural characteristics and compare the differences of the chloroplast genome of *P. cablin* of different origins, this study used the DNBSeg sequencing platform to sequence the whole genome of *P. cablin*, used getOrganelle to assemble the complete chloroplast genome, annotated the chloroplast genome through the OGDRAW website, and analyzed the basic structural characteristics, IR/SC boundary comparison, genome comparison and collinearity analysis, simple repeat sequences and interspersed repeat sequences, polymorphism analysis and relative usage analysis of synonymous codons. The results were as follows: (1) The full length of the chloroplast genomes of 20 different origins of patchouli was 152 461~152 510 bp, and 132 genes were annotated, including 87 CDS, 37 tRNA genes and 8 rRNA genes. (2) The mVISTA comparison found that *atpF*, *atpF-atpH*, *rps16-trnQ_UUG*, *rpoB-trnC_GCA*, *accD*, *psaI-ycf4*, *petA-psbJ*, *rpl16*, and *rps15-ycf1* were hypervariable regions. (3) The sites with nucleic acid diversity greater than 0.002 were located in the *trnM-CAU-atpB* interval, *ycf4*, *rpl32*, and *rpl32-trnL-UAC* interval. (4) Analysis of the relative usage of synonymous codons detected a total of 64 codons encoding 20 amino acids, and there were 33 highly preferred codons, among which codons ending in A/U accounted for the majority. (5) 74-76 SSRs, 15-18 palindrome repeat sequences, and 12-17 forward repeat sequences were detected.

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(202210572232)。

第一作者: 行冰楠(1993—), 硕士研究生, 主要从事中药品种鉴定与质量标准研究, (Email) 526928734@qq.com。

***通信作者:** 吴文如, 教授, 硕士生导师, 研究方向为中药品种鉴定与质量标准研究, (Email) wuwenru@gzucm.edu.cn。

(6)After genetic distance analysis and phylogenetic analysis, only GSY_MLXY has a distant genetic relationship with other cultivated types. In this study, the genome structure and different sites identified of chloroplasts from 20 different sources of *P. cablin* were obtained, which provided basic data for the development of molecular markers and the selection of superior germplasm.

Key words: *Pogostemon cablin*, chloroplast genome, repetitive sequence, single nucleotide polymorphism, relative usage of synonymous codon

广藿香 (*Pogostemon cablin*) 为唇形科刺蕊草属植物, 以干燥地上部分入药, 其功效为芳香化浊, 开胃止呕, 发表解暑 (国家药典委员会, 2020), 在医药、食品和化妆品加工业等方面有广泛的应用, 具有较高的经济价值, 亦是新型冠状病毒肺炎防疫中的重要中药 (李浩等, 2021)。广藿香, 这种具有独特香气的植物, 可以根据其化学成分划分为两大类: 广藿香酮型和广藿香醇型。广州和肇庆地区所产出的广藿香, 因其独特的香气, 被称为“牌香”或“肇香”。而广东湛江地区的吴川、遂溪、雷州, 以及海南省的万宁等地生产的广藿香, 人们习惯地称之为“琼香”, 亦或“南香”。传统认为酮型广藿香作药用, 醇型广藿香多用于提取挥发油 (罗集鹏等, 2005), 但广藿香培植过程中存在引种混乱、种质不清等问题 (顾艳等, 2022), 影响其药材的质量。吴文如等 (2019) 通过广藿香及其混伪品藿香等 ITS 序列的差异, 实现了广藿香特异性 PCR 鉴别。He 等 (2014) 采用 *rbcL*、*psbA-trnH* 等序列作为 DNA 条形码对广藿香化学型进行了鉴定, 但未达到准确区分的效果。传统的 DNA 条形码短片段可以实现特定物种的种间鉴定, 但在同一物种的种内差异研究中表现不佳。为探究广藿香种质的差异, 筛选优良种质, 以保障临床用药的安全有效, 进行不同来源广藿香的叶绿体基因组研究。

叶绿体作为一种独特的细胞器, 不仅承载着细胞内的自主遗传信息, 更在近年来凭借其特性, 成为了揭示植物进化和系统发育关系的强大工具。其基因组的特点在于其结构简单、分子量小、拷贝数众多, 以及基因种类的保守性 (Dobrogojski et al., 2020)。目前叶绿体基因组研究已经应用于安息香属 (Song et al., 2022)、苍术属 (Wang et al., 2021) 和桃金娘目 (Zhang et al., 2021) 等多种药用植物的系统发育学和优良品种选育, 且相对于传统 DNA 条形码短片段, 叶绿体全基因组具备更为丰富的变异位点, 在种间水平上提供了更高的分辨率, 因此鉴定效率更高。这使得叶绿体基因组在分子标记、近缘物种辨识等领域中展现出了巨大的潜力和应用价值。

本研究借助二代测序技术, 获得 20 个不同来源的广藿香叶绿体全基因组数据, 对这些广藿香叶绿体基因组进行组装、注释并绘制其相关的基因组图谱, 进行其基本结构特征分析、IR/SC 的边界比较、基因组比较及共线性分析、简单重复序列及散在重复序列、多态性分析、同义密码子相对使用度分析、遗传距离分析和系统发育, 拟探讨以下问题: (1) 不同来源的广藿香叶绿体基因组序列有何特征; (2) 20 个不同来源广藿香之间有何种差异; (3) 20 个不同来源广藿香之间的亲缘关系如何。以期揭示不同来源广藿香叶绿体基因组的差异, 为其分子标记的开发和优良种质筛选提供基础资料。

1 材料与amp;方法

1.1 材料

样品来自广东广州、肇庆、阳江、云浮等地, 经广州中医药大学吴文如教授鉴定为唇形科刺蕊草属植物广藿香。样品分为两份, 一份扦插、移栽于广州中医药大学时珍山药圃, 一份于 -20 °C 冰箱保存。采集的 20 个广藿香叶绿体基因组序列及注释信息已上传至 NCBI 数据库, 实验样品采集信息和 GenBank 登录号见表 1。

表 1 广藿香样品信息表
Table 1 *Pogostemon cablin* sample information

编号 Number	采集地点 Collection location	经纬度 Latitude and longitude	GenBank 登录号 GenBank accession number
GZY	广东省广州市番禺区广州中医药大学药科楼 Pharmacy Building, Guangzhou University of Chinese Medicine, Panyu District, Guangzhou City, Guangdong Province	113°24'36" E, 23°04'01" N	PP383983
GY	广东省肇庆市高要区南药产业园 Southern Medicine Industrial Park, Gaoyao District, Zhaoqing City, Guangdong Province	112°15'56" E, 23°11'07" N	PP383989
ZQXY	广东省肇庆市端州区肇庆学院 Zhaoqing University, Duanzhou District, Zhaoqing City, Guangdong Province	112°30'06" E, 23°06'46" N	PP383998
YC	广东省阳江市阳春市马水镇马兰村 Malan Village, Mashui Town, Yangchun City, Yangjiang City, Guangdong Province	111°37'28" E, 22°07'38" N	PP383994
YC2	广东省阳江市阳春市马水镇马兰村 Malan Village, Mashui Town, Yangchun City, Yangjiang City, Guangdong Province	111°37'28" E, 22°07'38" N	PP383995
SZS	广东省广州市番禺区广州中医药大学时珍山药圃 ShiZhen Shan, Guangzhou University of Chinese Medicine, Panyu District, Guangzhou City, Guangdong Province	113°24'38" E, 23°03'51" N	PP383993
YF2	广东省云浮市罗定市 Luoding City, Yunfu City, Guangdong Province	111°11'47" E, 22°40'33" N	PP383996
ZQZC	广东省肇庆市高要区乐城镇银村 Yin Village, Lecheng Town, Gaoyao District, Zhaoqing City, Guangdong Province	112°20'35" E, 23°20'07" N	PP383999
GSY_SP	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP383986
GSY_GY	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP383984
GSY_YC	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP383987
GSY_YN	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP383988
GSY_HN	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP383985
GSY_MLXY	广东省广州市天河区广东食品药品职业学院 Guangdong Food and Drug Vocational College, Tianhe District, Guangzhou City, Guangdong Province	113°22'59" E, 23°12'38" N	PP375409
ZQSP	广东省肇庆市高要区河台镇梁村 Liang Village, Hetai Town, Gaoyao District, Zhaoqing City, Guangdong Province	112°13'38" E, 23°20'07" N	PP383997
SX	广东省广州市荔湾区广州采芝林药业有限公司 Cai Zhi Lin Pharmaceuticals, Liwan District, Guangzhou City, Guangdong Province	110°25'01" E, 21°37'72" N	PP383992
P1	广东省云浮市云城区前锋镇崖楼村 Yalou Village, Qianfeng Town, Yuncheng District, Yunfu City, Guangdong Province	112°08'45" E, 22°48'59" N	PP375407
P3	广东省云浮市云城区前锋镇崖楼村 Yalou Village, Qianfeng Town, Yuncheng District, Yunfu City, Guangdong Province	112°08'45" E, 22°48'59" N	PP383990
P7	广东省云浮市云城区前锋镇崖楼村 Yalou Village, Qianfeng Town, Yuncheng District, Yunfu City, Guangdong Province	112°08'45" E, 22°48'59" N	PP375408
P10	广东省云浮市云城区前锋镇崖楼村 Yalou Village, Qianfeng Town, Yuncheng District, Yunfu City, Guangdong Province	112°08'45" E, 22°48'59" N	PP383991

1.2 方法

1.2.1 基因组 DNA 提取及测序

采用 CTAB 法提取广藿香叶片总 DNA，使用 Qubit Fluorometer 核酸蛋白定量仪（美国赛默飞公司）和 1%琼脂糖凝胶电泳检测总 DNA 浓度和质量。总 DNA 经检测合格后，交由华大基因科技有限公司，使用 DNBSeg 测序平台双末端测序策略构建片段大小为 150 bp 的测序文库进行测序。

1.2.2 叶绿体基因组组装及注释

测序得到原始测序数据，使用 SOAPnuke 软件（Chen et al., 2018）对低质量数据进行过滤，去除带接头(adapter)的 reads；去除长度<150 bp 的 reads；去除 N(N 表示无法确定碱基信息)的比例 1.0%以上的 reads；去除 read 中 polyX (X 可为 A, T, C 或 G) 长度超过 50 bp 的 reads；去除低质量 reads(即质量值 $Q \leq 20$ 的碱基数占整条 read 长度的 30%以上的 reads)，将最后得到 clean data 用于后续分析。使用 getOrganelle (Jin et al., 2020) v1.7.7.0 进行组装，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 21, 45, 65, 85；拼接完成后用 Bandage v0.8.1 软件可视。使用 Genenious v9.0.2 检查并调整序列。以广藿香的叶绿体基因组 (NC_042796.1) 序列为参考，通过 CPGAVAS2 网站 (<http://47.96.249.172:16019/analyzer/home>) 进行基因组注释，使用在线工具 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 绘制叶绿体基因组图谱。

1.2.3 简单重复序列与散在重复序列

使用 MISA (Beier et al., 2017) 软件的 Perl 脚本探索简单重复序列 (simple sequence repeats, SSR) 在叶绿体全基因组序列上的分布，设置单核苷酸重复单元大于等于 10 个，二核苷酸重复单元大于等于 5 个，三核苷酸重复单元大于等于 4 个，四、五、六核苷酸重复单元大于等于 3 个。应用在线软件 REPuter (Kurtz et al., 2001) (<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)、互补重复序列 (C) 和回文重复序列 (P)；基因组中的散在重复序列识别参数设置：海明距离设置为 3，最小重复长度不小于 11 bp，其他为默认值，取 e-value 不超过 1×10^{-5} 的重复序列。

1.2.4 边界收缩与扩张

使用 JSHY-Cloud (<http://cloud.genepioneer.com:9929>) 的 CPJSdraw 边界绘制工具比较不同来源的广藿香叶绿体基因组与参考序列 NC_042796.1 反向重复区与单拷贝区的边界 (Raubeson et al., 2007)，可视化其结果。

1.2.5 叶绿体基因组比较与共线性分析

使用 mVISTA (Frazer et al., 2004) 在线软件 (<https://genome.lbl.gov/vista/mvista/submit.shtml>)，参数设定中选择 shuffle-LAGAN 模式对 20 个不同来源的广藿香叶绿体基因组进行比较分析，以石牌广藿香叶绿体基因组 (NC_042796.1) 作为参考。使用 mauve (Darling et al., 2004) (version 20150226) 对广藿香叶绿体基因组进行共线性分析。

1.2.6 多态性位点分析

使用 mafft (Katoh & Standley, 2013) v7.490 软件比对 20 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 (Rozas et al., 2017) v6.12.03 软件，计算 20 个广藿香叶绿体基因组的核苷酸多样性水平，设置窗口长度 (window length) 为 600，步长 (step size) 为 200。

1.2.7 同义密码子相对使用度 (relative synonymous codon usage, RSCU)

采用 CodonW v1.4.4 计算各基因的 RSCU (Dana & Tuller, 2014)，并利用在线工具 RSCU-热图 (https://www.bioinformatics.com.cn/plot_basic_cluster_heatmap_plot_024) 绘制广藿香叶绿体基因组的 RSCU 热图。

1.2.8 遗传距离分析

使用“1.2.6”项下经 mafft v7.490 对齐后的 fasta 文件，导入 MEGA-X 软件，形成 mega 文件，设置自展值 Bootstrap method 1000，选择 p-distance 模式，成对计算不同来源广藿香叶绿体基因组的遗传距离 (Kumar et al., 2018)。

1.2.9 系统发育分析

使用 mafft v7.490 对齐 20 个广藿香叶绿体基因组及外类群荆芥 *Nepeta cataria* 的叶绿体基因组序列 (MT663220.1)，生成的 fasta 文件使用 iqtree v2.2.0.3 软件，以 MFP 模式，自展值 1 000，建立 ML 树。

2 结果与分析

2.1 广藿香叶绿体基因组构成

广藿香叶绿体基因组为环状双链分子，且呈典型的四分体结构（图 1），该结构包含 1 个大单拷贝区（LSC），2 个反向重复区（IR）和 1 个小单拷贝区（SSC）。基因组长度为 152 461~152 510 bp，GC 含量约为 38.2%。20 个不同来源广藿香叶绿体基因组中除 GSY_MLXY 的 SSC 为 17 571 bp，其余均为 17 584 bp；GSY_MLXY 的 IRa 与 IRb 为 25 664 bp，其余均为 25 662 bp；GZY、P7、SX、SZS、YC、YF2、ZQXY、ZQZC、GSY_GY、GSY_YC 的 LSC 为 83 554 bp，GSY_SP、GSY_YN、GSY_HN、P1、P3、P10、YC2、ZQSP、GY 的 LSC 为 83 553 bp，GSY_MLXY 的 LSC 为 83 611 bp（表 2）。20 个不同来源广藿香的叶绿体基因组结构特征一致，说明其叶绿体基因组结构保守。经 DNAMAN 比对，P7、GZY、SX、SZS、YC、YF2、ZQXY、ZQZC、GSY_GY、GSY_YC 这 10 个序列完全相同，以下简称 P7 等；P1、P3、P10、GSY_SP、GSY_YN、GSY_HN、YC2、ZQSP、GY 这 9 个序列完全相同，以下简称 P1 等。

表 2 广藿香叶绿体基因组的基本组成

Table 2 Basic composition of the chloroplast genome of *Pogostemon cablin*

编号 Number	长度 Length (bp)				含量 Content (%)				
	LSC	SSC	IR	Total	A	T	G	C	GC
P7 等 P7, etc	83 554	17 584	25 662	152 462	30.53	31.22	18.84	19.40	38.24
P1 等 P1, etc	83 553	17 584	25 662	152 461	30.53	31.22	18.84	19.40	38.24
GSY_MLXY	83 611	17 571	25 664	152 510	30.54	31.21	18.85	19.40	38.25

2.2 广藿香叶绿体基因组注释与归类

共注释出 132 个基因，这些基因中包含 87 个 CDS，37 个 tRNA 基因和 8 个 rRNA 基因，其中多拷贝基因有 18 个，除 *trnM*-CAU 为 3 个拷贝，*ndhB*、*rpl2*、*rpl23*、*rps12*、*rps7*、*rrn16*、*rrn23*、*rrn4.5*、*rrn5*、*trnA*-UGC、*trnL*-CAA、*trnN*-GUU、*trnR*-ACG、*trnV*-GAC、*ycf15*、*ycf2*、*trnI* 均为 2 个拷贝；18 个基因含有内含子，除 *rps12*、*clpP*、*ycf3* 含 2 个内含子，*ndhA*、*ndhB*、*petB*、*petD*、*atpF*、*rpl16*、*rpl2*、*rps16*、*rpoC1*、*trnA*-UGC、*trnC*-ACA、*trnI*、*trnK*-UUU、*trnL*-UAA、*trnS*-CGA 均含 1 个内含子（表 3）。

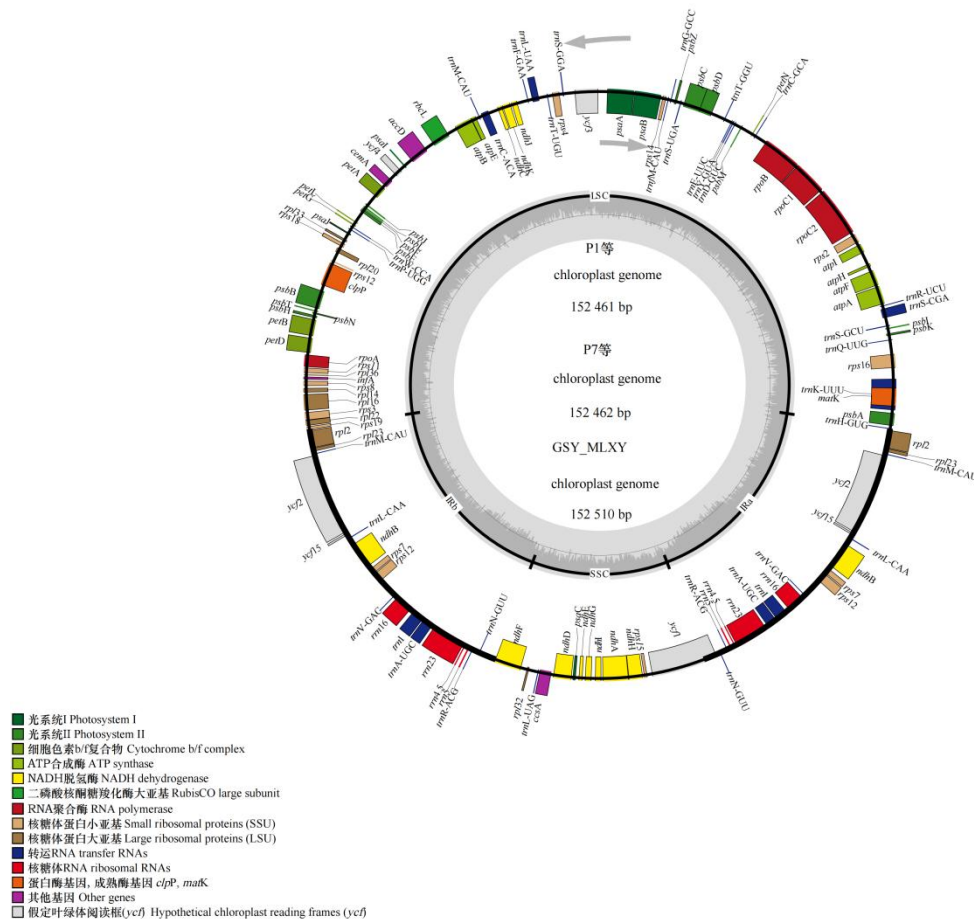


图 1 广藿香叶绿体基因组谱图

Fig. 1 Chloroplast genome profile of *Pogostemon cablin*

表 3 广藿香叶绿体基因组基因功能注释与分类

Table 3 Functional annotation and classification of chloroplast genome genes in *Pogostemon cablin*

基因功能 Gene fuction	基因分类 Gene category	基因名 Gene name	基因数 Gene number
光合作用 Photosynthesis	光系统 I 基因 Subunit of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>	5
	光系统 II 基因 Subunit of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
NADH 氧化还原酶基因 Subunit of NADH-dehydrogenase	细胞色素 b/f 复合体基因 Subunit of cytochrome b/f complx	<i>ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
	ATP 酶基因 Subunit of ATP synthase	<i>petA, petB*, petD*, petG, petL, petN</i>	6
	核酮糖二磷酸羧化酶大亚基基因 RubisCO large subunit	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
自我复制 Self-replication	核糖体蛋白大亚基基因 Large subunit of ribosome protein	<i>rbcl</i>	1
		<i>rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36</i>	11

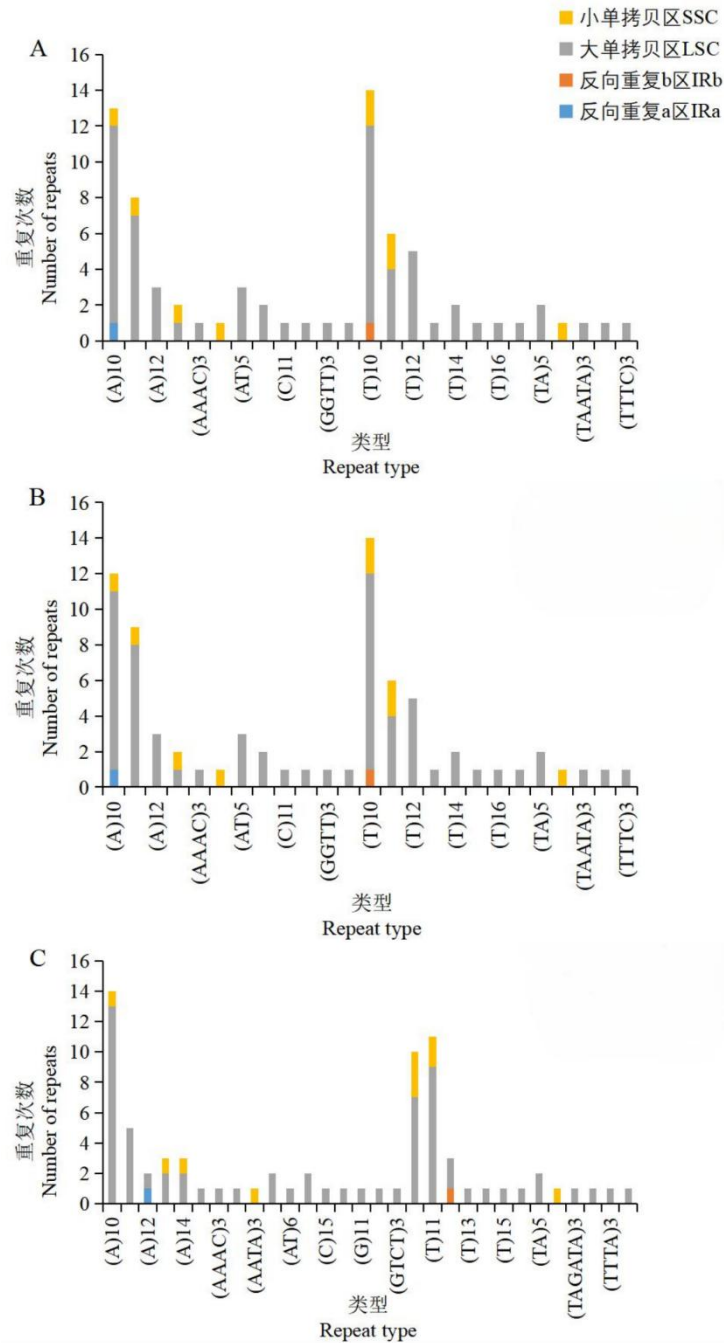
基因功能 Gene function	基因分类 Gene category	基因名 Gene name	基因数 Gene number
	核糖体蛋白小亚基基因 Small subunit of ribosome protein	<i>rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8</i>	14
	RNA 聚合酶基因 Subunit RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>	4
	核糖体 RNA 基因 Ribosomal RNAs	<i>rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)</i>	8
	转运 RNA 基因 Transfer RNAs	<i>trnA-UGC*(2), trnC-ACA*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnI*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU(3), trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-CGA*, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnW-CCA, trnY-GUA, trnY-M-CAU</i>	37
其他基因 Other genes	成熟酶 Maturase	<i>matK</i>	1
	蛋白酶 Protease	<i>clpP**</i>	1
	包膜蛋白 Envelope membrane protein	<i>cemA</i>	1
	乙酰辅酶 A 羧化酶基因 Subunit of acetyl-CoA-carboxylase	<i>accD</i>	1
	C 型细胞色素合成酶 C-type cytochrome synthase	<i>ccsA</i>	1
	翻译起始因子 Translation initiation factor	<i>infA</i>	1
未知功能基因 Unknow function gene	假定叶绿体阅读框(<i>ycf</i>) Hypothetical chloroplast reading frames (<i>ycf</i>)	<i>ycf1, ycf15(2), ycf2(2), ycf3**, ycf4</i>	7

注: *表示含有 1 个内含子, **表示含有 2 个内含子, ()中数字表示多拷贝基因的拷贝数。

Note: * indicates gene containing one intron; ** indicates gene containing two introns; numbers in () indicate gene with duplicate copy.

2.3 简单重复序列与散在重复序列

利用在线工具 MISA 检测广藿香叶绿体基因组中 SSR 的分布情况 (图 2), 共检测到 74~76 个 SSR, 以单核苷酸的重复为主, 占 76% 以上, 且以 T、A 为重复单元为主, 与 P1 等、P7 等不同, GSY_MLXY 叶绿体基因组在 *rps2* 和 *rpoC2* 间隔区处有 1 个以 (C) 15 的 C 重复单元 SSR, 并且在 *petB* 内含子处有 1 个以 (G) 11 的 G 重复单元的 SSR。各不同来源的广藿香叶绿体基因组的 SSR 中, 63~64 个分布于 LSC 区, 占比 84% 以上, IR 区仅有 2 个 SSR。P1 等、P7 等广藿香叶绿体基因组检测到 27 个散在重复序列, 15 个回文重复序列, 12 个正向重复序列, 而 GSY_MLXY 的叶绿体基因组检测到 35 个散在重复序列, 18 个回文重复, 17 个正向重复序列。另外, 广藿香叶绿体基因组均在 LSC 区, 有 1 个 205 bp 的回文重复序列, 此重复序列在叶绿体基因组拼接时解环时就被发现, 给广藿香叶绿体基因组的拼接与解环增加了难度。



A、B、C 分别为 P1 等 9 个广藿香叶绿体基因组、P7 等 10 个广藿香叶绿体基因组与 GSY_MLXY 广藿香叶绿体基因组的 SSR 类型与数量在不同区域的分布图。

A, B, and C were the distribution of SSR types and quantities in nine *Pogostemon cablin* chloroplast genomes such as P1, ten *P. cablin* chloroplast genomes such as P7 and GSY_MLXY *P. cablin* chloroplast genomes in different regions.

图 2 广藿香叶绿体基因组 SSR 类型与数量图及 SSR 分布区域图

Fig. 2 SSR type and quantity map and SSR distribution area of *Pogostemon cablin* chloroplast

2.4 边界收缩与扩张

IR/SC 区域边界的收缩和扩张是叶绿体变化的主要原因。研究比较了不同来源的广藿香叶绿体基因组与参考序列 NC_042796.1。LSC-IRb 的边界位于基因 *rps19* 上，IRb 区距边界最近的 *rpl2* 基因，与边界的距离均为 97 pb；IRb-SSC 的边界位于基因 *ndhF* 上，距离此边界最近的基因为 *trnN*，P1 等、P7 等与此边界的距离 1 422 bp，GSY_MLXY 与此边界的距离 1 420 bp；SSC-IRa 的边界

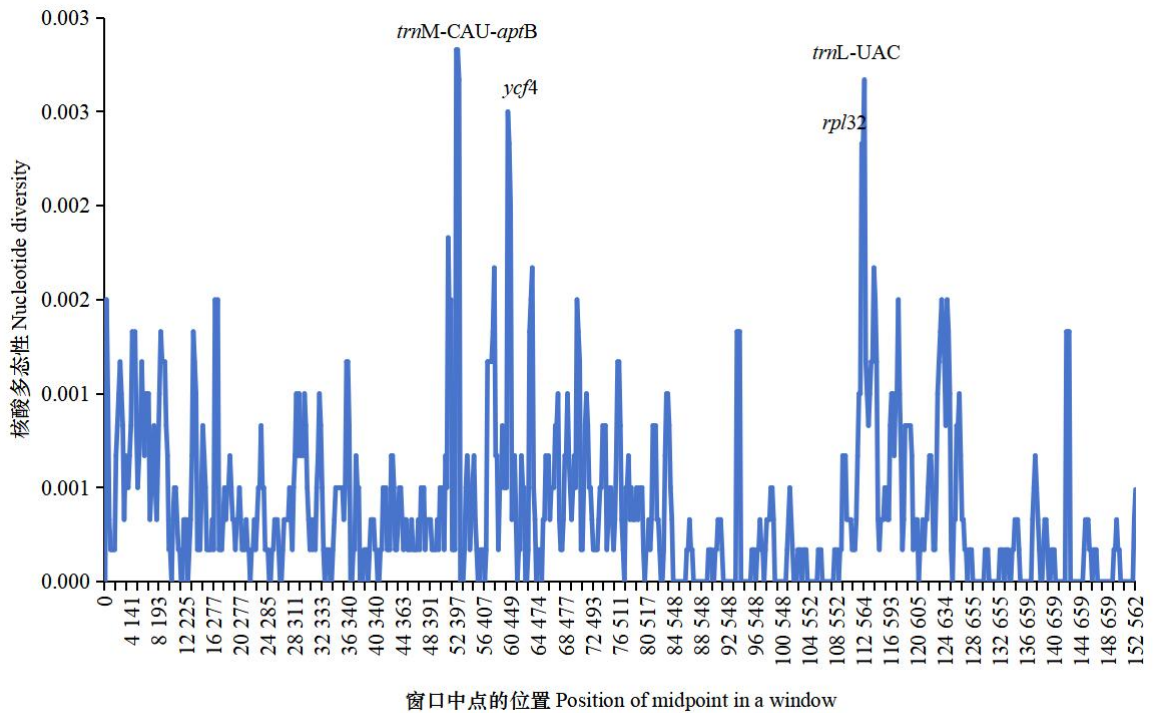


图 5 广藿香叶绿体基因组共线性图

Fig. 5 Collinear diagram of chloroplast genome of *Pogostemon cablin*

2.6 多态性位点分析

利用 DnaSP v6.12.03 软件计算 20 个广藿香叶绿体基因组的核苷酸多样性水平。共调查了 152 765 个位点，其中总共包括 760 个多态性位点。Pi 值范围为 0~0.002 83，平均值为 0.000 395 947，序列相似度高。编码区的可变位点多于非编码区，LSC 区、SSC 区的变异比 IR 区域的分歧更大。核酸多样性大于 0.002 的位点位于 *trnM-CAU-atpB* 间隔区、*ycf4*、*rpl32*、*trnL-UAC* 间隔区（图 6），或可用于广藿香栽培类型的鉴定。



窗口长度为 600 bp，步长为 200 bp

The window length is 600 bp, the step size is 200 bp.

图 6 20 个完整广藿香叶绿体基因组的滑动窗口分析

Fig. 6 Sliding window analysis of 20 complete *Pogostemon cablin* chloroplast genomes

2.7 同义密码子相对使用度分析

根据广藿香叶绿体基因组密码子 RSCU 值的统计结果显示，检测到共有 64 种密码子编码 20 个氨基酸（终止子不编码氨基酸），以 RSCU 大于 1.0 为标准，获得偏好性较强的密码子有 33 个，

其中以 A/U 结尾的密码子占大多数（图 7），表明广藿香叶绿体基因组主要偏好第三位碱基为 A 或 U 的密码子，这与已报道的大多数药用植物叶绿体基因组同义密码子相对使用度分析的结果（Morton, 1993; Wang et al., 2018; Wu et al., 2021; Wang et al., 2022; Guo et al., 2022）相似，然而这些密码子的 RSCU 均小于 2.0，说明基因组中不存在极强偏好性的密码子。

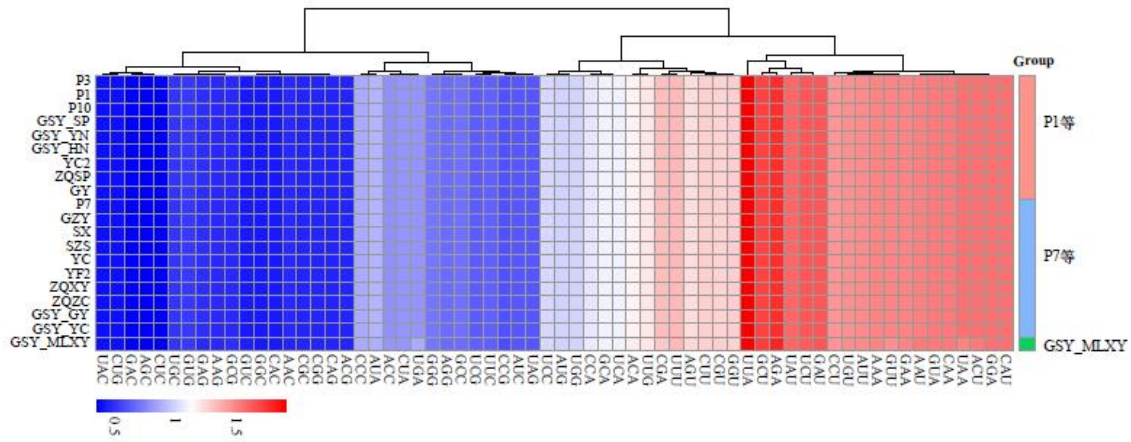


图 7 广藿香叶绿体基因组 RSCU 热图

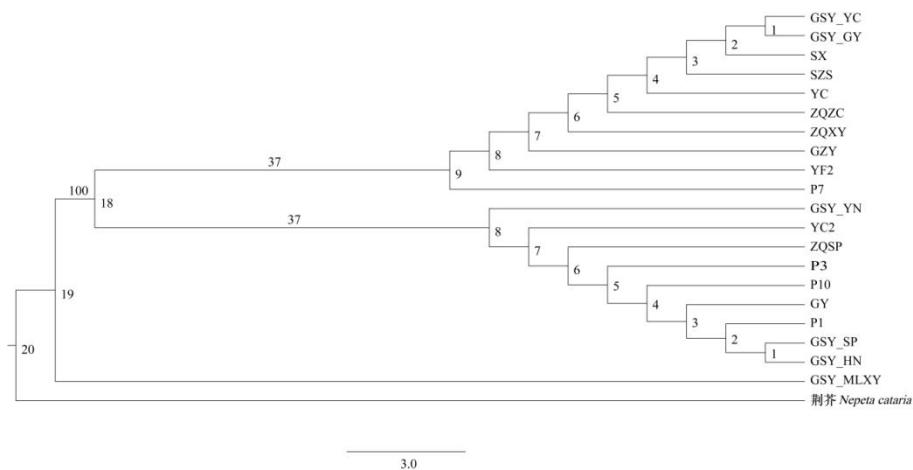
Fig. 7 Heatmap of chloroplast genome RSCU of *Pogostemon cablin*

2.8 遗传距离分析

使用 MEGA-X 计算 20 个广藿香叶绿体基因组的遗传距离，结果平均遗传距离为 0.000 244 142，仅 GSY_MLXY 与其他叶绿体基因组的成对遗传距离为 0.003 975，大于平均遗传距离，其余均为 0（表 4）。

2.9 系统发育分析

使用 iqtree v2.2.0.3，以唇形科植物荆芥的叶绿体基因组为外类群建立 ML 树。P7、YF2、GZY、ZQXY、ZQZC、YC、SX、GSY_GY、GSY_YC、P1、P3、P10、GSY_SP、GSY_YN、GSY_HN、YC2、ZQSP、GY 为一个支持率高的单系（BS=100%）说明这 19 种来源的广藿香亲缘关系较近，这 19 种广藿香之间的系统发育分支的支持率较低，难以区分。相比之下，它们与 GSY_MLXY 亲缘关系较远，与 GSY_MLXY 聚为一支，均来源于刺蕊草属，同时，这些分支同外类群荆芥属荆芥（*Nepeta cataria*）共同构成唇形科（图 8）。



节点数值表示自展支持率。

Numbers at nodes represent the bootstrap support value.

图 8 基于广藿香 20 条叶绿体全基因组构建的系统发育树 (ML)

Fig. 8 Phylogenetic tree based on twenty chloroplast genomes of *Pogostemon cablin* (ML)

表 4 20 条广藿香叶绿体基因组遗传距离

Table 4 Genetic distance of twenty chloroplast genomes of *Pogostemon cablin*

	GZY	GSY_GY	GSY_HN	GSY_MLXY	GSY_SP	GSY_YC	GSY_YN	GY	P1	P3	P7	P10	SX	SZS	YC	YC2	YF2	ZQSP	ZQXY	ZQZC	
GZY																					
GSY_GY	0																				
GSY_HN	0	0																			
GSY_MLXY	0.003975	0.003975	0.003975																		
GSY_SP	0	0	0	0.003975																	
GSY_YC	0	0	0	0.003975	0																
GSY_YN	0	0	0	0.003975	0	0															
GY	0	0	0	0.003975	0	0	0														
P1	0	0	0	0.003975	0	0	0	0													
P3	0	0	0	0.003975	0	0	0	0	0												
P7	0	0	0	0.003975	0	0	0	0	0	0											
P10	0	0	0	0.003975	0	0	0	0	0	0	0										
SX	0	0	0	0.003975	0	0	0	0	0	0	0	0									
SZS	0	0	0	0.003975	0	0	0	0	0	0	0	0	0								
YC	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0							
YC2	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0	0						
YF2	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0	0	0					
ZQSP	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0	0	0	0				
ZQXY	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
ZQZC	0	0	0	0.003975	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3 讨论

3.1 广藿香叶绿体基因组的结构特征

本研究组装并注释了 20 个不同来源广藿香的叶绿体基因组。总体来说，与大部分陆生植物的叶绿体基因组一样，广藿香的叶绿体基因组也呈现出典型的四分体结构，且不同来源广藿香叶绿体基因组共线性结果为单线型，未见重排或倒置，说明本研究中广藿香样品的叶绿体基因组结构高度保守。与以往广藿香叶绿体基因组研究相比，He 等（2016）首次报道完整的广藿香叶绿体基因组（NCBI 登录号：KX230834），为广东阳春广藿香的叶绿体基因组，但经 MUMmer v3.23 检查，其与广藿香参考基因组序列（NC_042796.1）共线性结果不佳，该序列 LSC 区的一部分序列连接在 IRa 区之后。而与 Zhang 等（2020）的研究相比，样品来源更加丰富且具有代表性。样品来源于广藿香不同的主产地，如云浮（YF2、P1、P3 等）、肇庆（GY、ZQXY、ZQSP）、湛江遂溪（SX）、阳春（YC、YC2）、海南（GSY_HN）等，此外还涉及到石牌广藿香的保种地广东食品药品职业学院（GSY_SP），以及境外来源地如越南（GSY_YN）、马来西亚（GSY_MLXY）等。这 20 个不同来源广藿香的叶绿体基因组全长 152 461~152 510 bp，GC 含量约为 38.2%，注释得到 132 个基因，这些基因中包含 87 个 CDS，37 个 tRNA 基因和 8 个 rRNA 基因，这也与大部分叶绿体基因组研究中的功能基因结果相似。

SSR 标记是基于 DNA 的有效分子标记，用于各种生物体的基因检测（Bei L et al., 2023）。本研究检测出 74~76 个 SSR 及 15~18 个回文重复序列、12~17 个正向重复序列，首次报道广藿香叶绿体基因组 LSC 区存在 1 个 205 bp 的回文重复片段，这些片段还需进一步深入验证，或许可用于广藿香物种的分子鉴定。

同义密码子的使用是控制基因表达的基本且保守的机制，密码子的最优性对个体表达具有重要性（Barrington et al., 2023）。本研究检测到共有 64 种密码子编码 20 个氨基酸，同义密码子相对使用度分析结果表明广藿香叶绿体基因组主要偏好第三位碱基为 A 或 U 的密码子，这与已报道的大多数药用植物相关分析的结果（Morton, 1993; Wang et al., 2018; Wu et al., 2021; Wang et al., 2022; Guo et al., 2022）相一致。

3.2 不同来源广藿香叶绿体基因组的差异

20 个不同来源广藿香叶绿体基因组中，GSY_MLXY 的基因组长度最大，为 152 510 bp，P7 等次之，P1 等长度最小；且 GSY_MLXY 的 SSC 为 17 571 bp，其余均为 17 584 bp；GSY_MLXY 的 IRa 与 IRb 为 25 664 bp，其余均为 25 662 bp；GSY_MLXY 的 LSC 为 83 611 bp，与 P1 等、P7 等 LSC 的 83 553- 83 554 bp 也有较大差别。IR/SC 边界仅 GSY_MLXY 与其他叶绿体基因组存在 2 bp 的收缩或扩增。发现 *rps16-trnQ*_UUG 等 9 个间区为高变区，核酸多样性大于 0.002 的位点位于 *trnM-CAU-atpB* 间隔区、*ycf4* 等 4 个位置，除 GSY_MLXY 外，P1 等和 P7 等叶绿体基因组序列相对比，仅 *ycf3-trnS_GGA* 间隔区存在差异位点。这些差异位点可以进一步结合核基因、线粒体基因组进行分析，获得广藿香物种鉴定有效的分子标记位点。

3.3 不同来源广藿香叶绿体基因组的亲缘关系

遗传距离分析与系统发育分析结果表明，仅 GSY_MLXY 与其他来源广藿香的亲缘关系较远，回顾前面的结构特征和差异分析，GSY_MLXY 的叶绿体基因组序列与其他来源广藿香存在明显差异，这或许是因为 GSY_MLXY 来源于马来西亚，与内陆广藿香存在地理隔离，其亲缘关系较远的原因可能与其生长环境、遗传背景以及物种演化等多个因素有关。首先，广藿香原产于马来西亚，并在东南亚国家、西印度群岛和巴拉主等地广泛种植。这些地区的气候、土壤等环境因素可能与内陆地区存在显著差异，导致广藿香在不同地理环境下的遗传变异和适应性演化。其次，广藿香作为一种植物，不同地区的广藿香可能存在着不同的遗传变异和基因交流情况，这些差异可能导致它们在遗传上的距离较远。最后，在漫长的演化过程中，广藿香可能经历了多种自然选

择和人工选择的压力，这些压力可能导致其在不同地区的种群发生分化，进而形成不同的亚种或变种，这些亚种或变种之间的亲缘关系可能较远。总而言之，马来西亚广藿香与内陆广藿香亲缘关系较远的原因可能是多方面的，包括生长环境、遗传背景以及物种演化等多个因素的综合作用。而内陆的广藿香大多为无性繁殖，这也可能是其遗传物质多样性退化的原因之一。这些发现可以为区分马来西亚来源的广藿香与其他来源广藿香提供参考依据。除 GSY_MLXY 外，P1 等和 P7 等亲缘关系较近，支持率较低，难以区分，可能是因为亲缘关系密切的种或变种之间在叶绿体特定基因上通常变异较小。

单一依靠叶绿体基因组反映物种完整的遗传变异情况较为有限（向坤莉等，2021）。而包括核心基因和非核心基因的泛基因组研究或许能够了解完整的物种遗传变异情况，更可以得到物种全面而准确的变异信息（SNP、Indel、CNV、PAV 等）(Jia et al., 2024)。因此，未来泛基因组将逐渐取代单一参考基因组，成为研究动植物进化、选择、基因功能和育种的“新标准”。本研究对于广藿香的叶绿体基因组研究也将会为后续泛基因组在广藿香中的应用提供相关数据基础。

参考文献

- BEIER S, THIEL T, MUNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction[J]. *Bioinformatics*, 33(16): 2583-2585.
- BARRINGTON CL, GALINDO G, KOCH AL, et al., 2023. Synonymous codon usage regulates translation initiation[J]. *Cell Rep*, 42(12):113413.
- BEI L, HE C, LIU J, et al., 2023. Genome-wide identification and characterization of microsatellite markers in Bactrian Camel[J]. *Genomics*, 115(6): 110726.
- CHEN Y, CHEN Y, SHI C, et al., 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data[J]. *Gigascience*, 7(1): 1-6.
- CHINESE PHARMACOPOEIA COMMISSION, 2020. Pharmacopoeia of the People's Republic of China[M]. Beijing: China Med Sci Press: 46-47. [国家药典委员会. 2020. 中华人民共和国药典一部 [M]. 一部. 北京: 中国医药科技出版社, 46-47.]
- DOBROGOJSKI J, ADAMIEC M, LUCIŃSKI R, 2020. The chloroplast genome: a review[J]. *Acta Physiologiae Plantarum*, 42(6): 155-167.
- DARLING AC, MAU B, BLATTNER FR, et al., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements[J]. *Genome Res*, 14(7): 1394-403.
- DANA A, TULLER T, 2014. The effect of tRNA levels on decoding times of mRNA codons[J]. *Nucleic Acids Res*, 42(14): 9171-81.
- FRAZER KA, PACHTER L, POLIAKOV A, et al., 2004. VISTA: computational tools for comparative genomics[J]. *Nucleic Acids Res*, 32(Web Server issue): W273-9.
- GU Y, MEI Y, XU SQ, et al., 2022. Research progress on germplasm resources and cultivation techniques of *Poaostemon cablin*[J]. *J Trop Crop*, 43(8): 1595-1603. [顾艳, 梅瑜, 徐世强, 等. 2022. 广藿香种质资源及栽培技术研究进展[J]. *热带作物学报*, 43(8): 1595-1603.]
- GUO S, LIAO X, CHEN S, et al., 2022. A Comparative Analysis of the Chloroplast Genomes of Four *Polygonum Medicinal Plants*[J]. *Front Genet*, 13: 764534.
- HE Y, WAN F, XIONG L, et al., 2014. Identification of two chemotypes of *Pogostemon cablin* (Blanco) Benth. through DNA barcodes[J]. *Z Naturforsch C J Biosci*, 69(5-6): 253-8.

- HE Y, XIAO H, DENG C, et al., 2016. The Complete Chloroplast Genome Sequences of the Medicinal Plant *Pogostemon cablin*[J]. *Int J Mol Sci*, 17(6): 820.
- JIN JJ, YU WB, YANG JB, et al., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes[J]. *Genome Biol*, 21(1): 241.
- JIA M, WANG J, CAO D, et al., 2024. The pan-plastome of *Hemerocallis citrina* reveals new insights into the genetic diversity and cultivation history of an economically important food plant[J]. *BMC Plant Biol*, 24(1): 44.
- KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. *Nucleic Acids Res*, 29(22): 4633-4642.
- KATOH K, STANDLEY DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Mol Biol Evol*, 30(4): 772-780.
- KUMAR S, STECHER G, LI M, et al., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms[J]. *Mol Biol Evol*, 35(6): 1547-1549.
- LI H, LIU W, HOU HX, et al., 2021. Patent compound Chinese medicine prescriptions in preventing and treating COVID-19[J]. *Centr S Pharm*, 19(11): 2426-2431. [李浩, 刘伟, 侯贺祥, 等. 2021. 中药防治新型冠状病毒肺炎专利复方用药规律分析[J]. *中南药学*, 19(11): 2426-2431.]
- LUO JP, FENG YF, HE B, et al., 2005. A study of the authenticity of *Pogostemon cablin*[J]. *Chin Med Mat*, 28(12): 1121-1125. [罗集鹏, 冯毅凡, 何冰, 等. 2005. 广藿香的道地性研究[J]. *中药材*, 28(12): 1121-1125.]
- MORTON BR, 1993. Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability[J]. *J Mol Evol*, 37(3): 273-80.
- RAUBESON LA, PEERY R, CHUMLEY TW, et al., 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*[J]. *BMC Genomics*, 8: 174.
- ROZAS J, FERRER-MATA A, SÁNCHEZ-DELBARRIO JC, et al., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets[J]. *Mol Biol Evol*, 34(12): 3299-3302.
- SONG Y, ZHAO W, XU J, et al., 2022. Chloroplast genome evolution and species identification of styrax (Styracaceae)[J]. *Biomed Res Int*, 2022:5364094.
- WANG Y, WANG S, LIU Y, et al., 2021. Chloroplast genome variation and phylogenetic relationships of *Atractylodes* species[J]. *BMC Genomics*, 22(1): 103.
- WANG S, YANG C, ZHAO X, et al., 2018. Complete chloroplast genome sequence of *Betula platyphylla*: gene organization, RNA editing, and comparative and phylogenetic analyses[J]. *BMC Genomics*, 19(1): 950.
- WU L, NIE L, WANG Q, et al., 2021. Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae[J]. *Sci Rep*, 11(1): 14643.
- WANG Z, CAI Q, WANG Y, et al., 2022. Comparative analysis of codon bias in the chloroplast genomes of Theaceae species[J]. *Front Genet*, 13: 824610.
- WU WR, AN X, LAI HL, et al., 2019. Identification of *Pogostemon cablin* (Blanco) Benth. with the counterfeit products by allele-specific PCR based on ITS2 sequence[J]. *Lishizhen Med Mat Med Res*, 30(4): 897-900. [吴文如, 安鑫, 来慧丽, 等. 2019. 基于 ITS2 序列位点特异性 PCR 鉴别广藿香及其混伪品[J]. *时珍国医国药*, 30(4): 897-900.]
- XIANG KL, HE WC, ZOU Y, et al., 2021. Application of pan-genome in genetic diversity and functional genomics[J]. *Guihaia*, 41(10): 1674-1682. [向坤莉, 贺文闯, 邹益, 等. 2021. 泛基因组研

究在遗传多样性和功能基因组学中的应用[J]. 广西植物, 41(10): 1674-1682.]

ZHANG XF, LANDIS JB, WANG HX, et al., 2021. Comparative analysis of chloroplast genomestructure and molecular dating in Myrtales[J]. BMC Plant Biol, 21(1): 219.

ZHANG CY, LIU TJ, MO XL, et al., 2020. Comparative analyses of the chloroplast genomes of patchouli plants and their relatives in *Pogostemon* (Lamiaceae)[J]. Plants (Basel), 9(11): 1497.