

• 药材与资源 •

芍药属 4 种药用植物叶绿体基因组比较及系统发育分析

范丙友^{1*}, 陈 盈¹, 郝明扬¹, 高 凯², 刘改秀², 王萌瑶¹, 吴 疆^{3*}

1. 河南科技大学农学院, 河南 洛阳 471023

2. 洛阳市农林科学院, 河南省牡丹基因开发工程技术中心, 中原学者工作站, 河南 洛阳 471023

3. 榆林学院信息工程学院, 陕西 榆林 719000

摘要: 目的 研究芍药属 4 种药用植物叶绿体基因组特征、变异程度及芍药属和毛茛科植物的系统发育关系, 为芍药属药用植物的分类地位和系统发育提供参考。方法 应用二代高通量测序技术测定芍药 *Paeonia lactiflora* 叶绿体基因组, 利用比较基因组学方法分析了芍药与川赤芍 *P. Veitchii*、牡丹 *P. suffruticosa* 及杨山牡丹 *P. ostii* 等芍药属药用植物叶绿体基因组之间的结构特征及变异程度, 以山地虎耳草 *Saxifraga sinomontana* 等植物为外类群, 分析了芍药属植物与毛茛科植物的系统发育关系。结果 芍药属 4 种药用植物的叶绿体基因组均为典型的四分结构, 包含 1 个大单拷贝区 (large single copy, LSC)、1 个小单拷贝区 (small single copy, SSC) 和 2 个反向重复区 (inverted repeats, IRa 和 IRb)。芍药叶绿体基因组序列长度为 152 731 bp, GC 含量为 38.43%, 共注释到基因 126 个, 其中 82 个蛋白质编码基因、36 个 tRNA 基因和 8 个核糖体 rRNA 基因。重复序列分析发现芍药叶绿体基因组重复序列数量最多为 48 个, 杨山牡丹最少为 39 个; 只有芍药同时含有 4 种类型的重复序列。芍药含有最多的 SSR 序列达 52 个, 4 种植物 SSR 序列中单核苷酸重复占比达 81%~86%, 绝大多数为 A/T 重复。密码子偏好性分析表明芍药属 4 种药用植物中氨基酸出现频率均以亮氨酸最高, 半胱氨酸最低; 31 种密码子具偏好性, 且密码子偏好性与密码子第 3 位碱基具显著相关性。LSC/IRb、IRb/SSC、SSC/IRa 和 IRa/LSC 边界附近的基因类型相同, 相对保守, 但牡丹组和芍药组之间仍具有差异。选择压力分析结果表明, 绝大多数基因 K_a/K_s 值均小于 1, 受纯化选择, *matK*、*ndhB* 和 *rpoA* 受到正选择。叶绿体基因组比较分析发现芍药属内 4 种药用植物的叶绿体基因组中非编码区比基因编码区的变异程度大, 筛选出了 6 个基因间区和 2 个基因编码区的高变异区。系统发育分析结果显示, 芍药与美丽芍药亲缘关系最近, 所有芍药属植物被聚为一个分支, 外类群把芍药属和毛茛科植物分隔开来, 芍药属植物与毛茛科植物亲缘关系较远。结论 芍药属 4 种植物在叶绿体基因组结构特征方面较为保守, 但其变异程度存在一定差异; 筛选出的 8 个高变异区可为芍药属药用植物的条形码开发、物种鉴定及药用植物质量控制提供候选片段; 系统发育分析表明芍药属植物与外类群植物亲缘关系较近, 而与毛茛科植物亲缘关系较远, 为芍药属植物系统进化、芍药属药用植物的物种鉴定、质量控制及保护开发等研究奠定基础。

关键词: 芍药属; 药用植物; 芍药; 川赤芍; 牡丹; 杨山牡丹; 叶绿体基因组; 变异; 进化; 系统发育

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)13-4765-14

DOI: 10.7501/j.issn.0253-2670.2025.13.020

Chloroplast genomes comparison and phylogenetic analysis of four medicinal plants in *Paeonia*

FAN Bingyou¹, CHEN Ying¹, HAO Mingyang¹, GAO Kai², LIU Gaixiu², WANG Mengyao¹, WU Jiang³

1. College of Agriculture, Henan University of Science and Technology, Luoyang 471023, China

2. Peony Gene Development Engineering and Technology Center of Henan province, Central Plains Scholar Workstation, Luoyang Academy of Agriculture and Forestry Sciences, Luoyang 471023, China

3. School of Information Engineering, Yulin University, Yulin 719000, China

Abstract: Objective To understand the chloroplast genome characteristics, degree of variation of four medicinal plants in *Paeonia* and the phylogenetic relationship between *Paeonia* and Ranunculaceae. This study will provide a reference for classification status and

收稿日期: 2025-02-02

基金项目: 国家自然科学基金资助项目 (U1204323)

*通信作者: 范丙友 (1974—), 男, 教授, 博士, 主要从事植物分子生物学研究。E-mail: fanbingyou2005@163.com

吴 疆 (1981—), 男, 教授, 博士, 主要从事生物信息学研究。E-mail: wujiang@yulinu.edu.cn

phylogeny of medicinal plants in *Paeonia*. **Methods** High-throughput next-generation sequencing technology was applied to sequence and assemble chloroplast genome of *Paeonia lactiflora*. Comparative genomics methods were used to analyze the structural characteristics and degree of variation among the chloroplast genomes of four medicinal plants in *Paeonia*, including *P. lactiflora*, *P. veitchii*, *P. suffruticosa*, and *P. ostii*. The phylogenetic relationship between *Paeonia* and Ranunculaceae was analyzed, with *Saxifraga sinomontana* and other related plants as outgroups. **Results** The chloroplast genomes of four medicinal plants in *Paeonia* were all typical tetrad structures, including one large single copy region (LSC), one small single copy region (SSC), and two reverse repeat regions (IRa and IRb). The length of chloroplast genome of *P. lactiflora* was 152 731 bp, with GC content of 38.43%. A total of 126 genes were annotated, including 82 protein coding genes, 36 tRNA genes, and eight ribosomal rRNA genes. Repetitive sequence analysis revealed that the chloroplast genome of *P. lactiflora* had the highest number of repetitive sequences at 48, while *P. ostii* lowest at 39. Only *P. lactiflora* contains four types of repeat sequences simultaneously. *P. lactiflora* contained the largest number of SSR sequences with 52. Single nucleotide repeats accounted for 81%—86% of the SSR in the four medicinal plants chloroplast genome, with the vast majority being A/T repeats. Codon preference analysis showed that among the four medicinal plants of *Paeonia*, the highest frequency of amino acid was leucine, while lowest was cysteine. Thirty-one codons possessed preference, and codon preference was significantly correlated with the third base of the codon. The gene types near the boundaries of LSC/IRb, IRb/SSC, SSC/IRa and IRa/LSC were the same and relatively conserved, but there were still little differences between sect *Moutan* and *Paeonia*. The selection pressure analysis showed that the K_a/K_s values of the majority of genes were less than one. Only *matK*, *ndhB*, and *rpoA* were positively selected. Chloroplast genome comparative analysis revealed that the non-coding regions exhibited greater variability than the gene-coding regions. A total of six intergenic regions and two gene coding regions with high variability were screened out. The phylogenetic analysis showed that the relationship between *P. lactiflora* and *P. maire* was closest. All species in *Paeonia* were clustered into one branch, while outgroups can separate *Paeonia* from Ranunculaceae. The relationship between *Paeonia* and Ranunculaceae was relatively distant. **Conclusion** The chloroplast genome structural characteristics of the four medicinal plants in *Paeonia* are relatively conservative, with certain differences on chloroplast genome variation degree. The eight highly variable regions identified can be served as the candidate fragments for barcode development, species identification and quality control of medicinal plants. Phylogenetic analysis shows that species in *Paeonia* are more closely related to the outgroup plants, but far related to species in Ranunculaceae. This lays the foundation for phylogeny of *Paeonia* and research on species identification, quality control, protection and development of medicinal plants resources in *Paeonia*.

Key words: *Paeonia* L.; medicinal plants; *Paeonia lactiflora* Pall.; *Paeonia veitchii* Lynch; *Paeonia suffruticosa* Andr.; *Paeonia ostii* T. Hong & J. X. Zhang; chloroplast genome; variation; evolution; phylogeny

芍药属的牡丹 *Paeonia suffruticosa* Andr.和芍药 *P. lactiflora* Pall.均为中国传统名花，二者被尊称为“花王”和“花相”。除观赏和文化价值外，二者还具有重要的药用价值，其中牡丹根可被加工为丹皮，芍药根可被加工为白芍和赤芍^[1]。丹皮在东汉时期的《神农本草经》中首次被记载，丹皮具有清热凉血、活血化瘀的功效，是临床常用的清热凉血药，具有抗炎、调节免疫功能、保护心血管等作用^[2]。丹皮的产地广泛，包括安徽铜陵和亳州、山东菏泽、重庆垫江、四川中江等地，其中安徽铜陵凤凰山所产丹皮质量上乘，其原料来自该地区的牡丹主栽品种‘凤丹’及其野生种杨山牡丹 *P. ostii* T. Hong & J. X. Zhang^[3]。综合考虑产量、质量及药用历史和习惯等因素，杨山牡丹是生产高品质丹皮的理想原料。因此，杨山牡丹也应被纳入丹皮的基原植物范畴^[1]。

战国时期的《五十二病方》是记载芍药入药最

早的方书^[4]，但未将其区分为白芍或赤芍。梁朝陶弘景《本草经集注》首次提出了赤芍和白芍之分^[5]。白芍和赤芍在植物来源、加工方式和功效主治方面存在差别。其中，白芍是将芍药根去除头尾和细根后，置于沸水中水煮，再除去外皮切薄片入药；赤芍则是将芍药或川赤芍 *P. veitchii* Lynch 的干燥根采挖后，洗净切片、晒干即可入药^[6]。白芍具有养血调经、敛阴止汗、柔肝止痛、平抑肝阳的功效，常用于血虚萎黄、月经不调、自汗、盗汗、胁痛、腹痛、四肢挛痛、头痛眩晕等症^[7]。赤芍具清热凉血、散瘀止痛等功效，用于热入营血、温毒发斑、吐血衄血、目赤肿痛、肝郁胁痛、经闭痛经、癥腹痛、跌扑损伤、痈肿疮疡等^[8]。

芍药属植物具有一些原始的性状如离生心皮等，与毛茛科植物相似，最早被划入毛茛科^[9]；但芍药属植物花部结构如花托、花瓣数量和花蕊结构方面比典型的毛茛科植物更加复杂。基于最新的分

子证据, 众多学者支持将芍药属从毛茛科独立, 单独成科为芍药科^[10-11]。但《中国药典》2020 年版依然将丹皮、白芍和赤芍等中药材的基原植物牡丹、芍药和川赤芍等药用植物划归为毛茛科^[8]。芍药属植物分类地位的不统一对于芍药属药用植物的药材鉴定、质量把控、药理研究和临床应用都将产生潜在的深远负面影响。

随着测序技术的不断进步和成熟, 基因组学越来越多地被应用于系统分类研究。比较基因组学通过对不同物种基因序列的比较, 可为物种分类提供最直接的证据。已测定的栽培牡丹‘洛神晓春’和杨山牡丹核基因组高达 12.25 Gb^[12]和 12.28 Gb^[13]。由于核基因组测序成本高、组装难度大, 目前还不适宜应用细胞核全基因组序列来开展芍药属植物的系统发育研究。叶绿体基因组大小通常在 120~160 kb, 相较于庞大复杂的核基因组更易于测序和分析。叶绿体基因组编码的约 100~130 个基因中, 与光合作用紧密相关的基因序列和功能在漫长进化中保持相对稳定, 这使得芍药属植物间的叶绿体基因组存在可比对的保守区域^[14]。基于这些优势, 对芍药叶绿体基因组进行特征分析, 并与芍药属其他 3 种药用植物叶绿体基因组进行比较分析, 可为芍药属药用植物的分类地位提供科学依据, 也可芍药属植物的系统进化关系研究及芍药属药用资源植物的种质鉴定提供科学依据。

1 材料

芍药样本采自洛阳市农林科学院国家牡丹和芍药种质资源库 (N34°38'20", E112°28'51"), 经洛阳市农林科学院刘改秀研究员鉴定为芍药 *P. lactiflora* Pall. (标本号 LKLPB202402)。采集长势良好的芍药新鲜叶片, 用无菌水冲洗数次, 用液氮速冻后带回实验室内置于 -80 °C 冰箱保存备用。

芍药叶绿体基因组序列和注释信息已提交到 NCBI 数据库, GenBank 登录号为 OP205316。其余 3 种芍药属药用植物牡丹、杨山牡丹和川赤芍的叶绿体全基因组均从 NCBI 数据库检索下载, 具体信息见表 1。

2 方法

2.1 DNA 提取及测序

应用植物 DNA 提取试剂盒 (TaKaRa MiniBEST Plant Genomic DNA Extraction Kit) 提取芍药叶片总 DNA。1% 琼脂糖凝胶电泳检测 DNA 质量, 运用

表 1 芍药属 4 种药用植物及叶绿体基因组 GenBank 登录号

Table 1 Four medicinal plants of *Paeonia* and their chloroplast genome GenBank accession numbers

物种	拉丁名	GenBank 登录号
芍药	<i>P. lactiflora</i> Pall.	OP205316
牡丹	<i>P. suffruticosa</i> Andr.	MK701995
杨山牡丹	<i>P. ostii</i> T. Hong & J. X. Zhang	OP324592
川赤芍	<i>P. veitchii</i> Lynch	NC_032401

NanoDrop™ 核酸测定仪 (NanoDrop Technologies, 美国) 测定 DNA 浓度。检测合格后的总 DNA 利用 MGISEQ-200 平台 (菲沙基因公司) 进行建库测序。运用 fastp^[15] 软件对原始测序数据进行质控。

2.2 叶绿体基因组组装、注释及物理图谱绘制

使用软件 Getorganelle v1.7.6.1^[16] 对芍药叶绿体基因组进行组装。运用 Circlator v.2.0.1^[17] 检验已组装的叶绿体基因组是否含有缺口。运用在线注释工具 CPGAVAS2^[18] 对叶绿体基因组进行注释, 手工调整矫正获得完整准确的注释信息。利用在线工具 OGDRAW^[19] 绘制芍药叶绿体基因组圈图。

2.3 重复序列分析

使用在线工具 REPuter^[20] 对芍药叶绿体基因组包含的正向、反向、回文和互补等长重复序列进行分析。应用在线工具 MISA^[21] 进行 SSR 分析, 其参数设置如下: 单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸对应的最少重复次数分别为 10、6、4、3、3、3^[22]。

2.4 密码子使用偏好性分析

使用软件 Codon W v1.4.2 对叶绿体基因组 CDS 区密码子使用偏好性进行分析, 统计氨基酸出现的次数、频率和同义密码子的相对使用度 (relative synonymous codon usage, RSCU), 分析结果用 CPStools^[23] 可视化。

2.5 IR 边界收缩和扩张分析

通过在线工具 IRscope^[24] 对芍药属 4 种药用植物叶绿体基因组的 IR 边界进行比较分析。

2.6 选择压力分析

应用 KaKs_calculator v3.0 软件^[25] 对芍药属 4 种药用植物叶绿体基因组所共有的蛋白质编码基因进行选择压力分析。通过两两比较, 精准计算出各个蛋白质编码基因的同义突变频率 (K_s)、非同义突变频率 (K_a) 以及非同义突变频率与同义突变频率的比值 (K_a/K_s)。其中, K_a/K_s 的值具有重要的指示意义, 当 $K_a/K_s > 1$ 时, 意味着该基因受到自然选择压力, 即正

选择, 其可能正朝着更有利于生物适应性的方向进化; 当 $K_a/K_s=1$ 时, 表示该基因受到中性选择, 处于一种相对稳定的进化状态; 当 $K_a/K_s<1$ 时, 该基因受到纯化选择, 也就是负选择, 这表明基因可能会在进化过程中维持其原有的功能, 去除有害突变。

2.7 叶绿体基因组序列结构变异分析

以芍药叶绿体基因组 (OP205316) 为参考基因组, 应用在线工具 mVISTA^[26] 进行芍药、川赤芍、牡丹和杨山牡丹叶绿体基因组的结构变异分析; 利

用 DnaSPv6 计算 4 种植物叶绿体基因组的核苷酸多样性 (π)。

2.8 系统发育分析

下载芍药属 *Paeonia* L.、毛茛科 (*Ranunculaceae*) 及和芍药科亲缘较近的山地虎耳草 *Saxifraga sinomontana* J. T. Pan & Gornall、虎耳草 *S. stolonifera* Curtis、月季 *Rosa chinensis* Jacq 及葡萄 *Vitis vinifera* L. 的叶绿体基因组序列 (表 2); 使用软件 PhyloSuite v1.2.2^[27] 提取叶绿体基因组共

表 2 物种及叶绿体基因组序列信息

Table 2 Species and chloroplast genome sequences information

物种	拉丁名	目	科	属	登录号	分组
四川牡丹	<i>P. decomposita</i>	虎耳草目	芍药科	芍药属	NC_039425	牡丹组革质花盘亚组
矮牡丹	<i>P. jishanensis</i>	虎耳草目	芍药科	芍药属	NC_050330	牡丹组革质花盘亚组
卵叶牡丹	<i>P. qiui</i>	虎耳草目	芍药科	芍药属	NC_050333	牡丹组革质花盘亚组
牡丹	<i>P. suffruticosa</i>	虎耳草目	芍药科	芍药属	MK701995	牡丹组革质花盘亚组
杨山牡丹	<i>P. ostii</i>	虎耳草目	芍药科	芍药属	OP324592	牡丹组革质花盘亚组
紫斑牡丹	<i>P. rockii</i>	虎耳草目	芍药科	芍药属	OK235337	牡丹组革质花盘亚组
紫牡丹	<i>P. delavayi</i>	虎耳草目	芍药科	芍药属	NC_035718	牡丹组革质花盘亚组
狭叶牡丹	<i>P. potaninii</i>	虎耳草目	芍药科	芍药属	NC_050332	牡丹组肉质花盘亚组
大花黄牡丹	<i>P. ludlowii</i>	虎耳草目	芍药科	芍药属	NC_035623	牡丹组肉质花盘亚组
黄牡丹	<i>P. lutea</i>	虎耳草目	芍药科	芍药属	NC_050331	牡丹组肉质花盘亚组
布朗芍药	<i>P. brownii</i>	虎耳草目	芍药科	芍药属	NC_037880	北美芍药组
芍药	<i>P. lactiflora</i>	虎耳草目	芍药科	芍药属	OP205316	芍药组
美丽芍药	<i>P. mairei</i>	虎耳草目	芍药科	芍药属	MZ617462	芍药组
窄叶芍药	<i>P. anomala</i>	虎耳草目	芍药科	芍药属	MT210549	芍药组
川赤芍	<i>P. veitchii</i>	虎耳草目	芍药科	芍药属	NC_032401	芍药组
块根芍药	<i>P. intermedia</i>	虎耳草目	芍药科	芍药属	MT210547	芍药组
草芍药	<i>P. obovata</i>	虎耳草目	芍药科	芍药属	NC_026076	芍药组
多花芍药	<i>P. emodi</i>	虎耳草目	芍药科	芍药属	MT210548	芍药组
山地虎耳草	<i>S. sinomontana</i>	虎耳草目	虎耳草科	虎耳草属	MN104589	外类群
虎耳草	<i>S. stolonifera</i>	虎耳草目	虎耳草科	虎耳草属	NC_037882	外类群
月季	<i>Rosa chinensis</i>	蔷薇目	蔷薇科	蔷薇属	NC_065867	外类群
葡萄	<i>Vitis vinifera</i>	葡萄目	葡萄科	葡萄属	NC_007957	外类群
直距耬斗菜	<i>Aquilegia rockii</i>	毛茛目	毛茛科	耬斗菜属	NC_046738	唐松草亚科
距瓣尾囊草	<i>Urophysa rockii</i>	毛茛目	毛茛科	尾囊草属	NC_039742	唐松草亚科
东北扁果草	<i>Isopyrum manshuricum</i>	毛茛目	毛茛科	北扁果草属	NC_041541	唐松草亚科
丝叶唐松草	<i>Thalictrum</i>	毛茛目	毛茛科	唐松草属	MT922012	唐松草亚科
鸡爪草	<i>Calathodes oxycarpa</i>	毛茛目	毛茛科	鸡爪草属	NC_041475	金莲花亚科
矮金莲花	<i>Trollius farreri</i>	毛茛目	毛茛科	金莲花属	NC_050872	金莲花亚科
兴安升麻	<i>Actaea dahurica</i>	毛茛目	毛茛科	类叶升麻属	NC_041533	金莲花亚组
铁破锣	<i>Beesia calthifolia</i>	毛茛目	毛茛科	铁破锣属	MK253467	金莲花亚科
黑种草	<i>Nigella damascena</i>	毛茛目	毛茛科	黑种草属	MN648403	金莲花亚科
厚叶美花草	<i>Callianthemum</i>	毛茛目	毛茛科	美花草属	MK253466	毛茛亚科
水毛茛	<i>Ranunculus bungei</i>	毛茛目	毛茛科	毛茛属	MK253468	毛茛亚科
驴蹄草	<i>Caltha palustris</i>	毛茛目	毛茛科	驴蹄草属	MK253465	金莲花亚科
飞燕草	<i>Consolida ajacis</i>	毛茛目	毛茛科	飞燕草属	MK569484	金莲花亚科
还亮草	<i>Delphinium</i>	毛茛目	毛茛科	翠雀属	MK253461	金莲花亚科
北乌头	<i>Aconitum kusnezoffii</i>	毛茛目	毛茛科	乌头属	MK253471	金莲花亚科

有的编码蛋白质基因；使用软件 MAFFT v7.3.13^[28]进行多重序列比对；使用软件 IQ-TREE v1.6.8^[29]基于最大似然值（ML）法构建系统发育树，预测模型为 GTR+F+I+G4，自展值为 1 000；使用在线工具 iTOL^[30]对发育树进行可视化。

3 结果与分析

3.1 芍药叶绿体基因组结构特征

组装后的芍药叶绿体基因组为环状双链分子，具有典型的四分体结构，包含 1 个大单拷贝区（large single copy, LSC）、1 个小单拷贝区（small single

copy, SSC）和 2 个反向重复区（inverted repeats, IRa 和 IRb）。芍药叶绿体基因组总长度为 152 731 bp，其中 LSC 区、SSC 区和 2 个反向互补的 IRs 区长度分别为 84 402、16 969、25 680 bp（图 1）。总 GC 含量为 38.43%，SSC 区与 LSC 区、IRs 区的 GC 含量存在一定的差异，SSC 区为 32.72%，LSC 区和 IRs 区为 36.73%。芍药属 4 种药用植物中，杨山牡丹叶绿体基因组最长为 152 768 bp，牡丹最短为 152 608 bp。芍药叶绿体基因组比杨山牡丹小 37 bp，比川赤芍和牡丹分别大 49 bp 和 123 bp（表 3）。

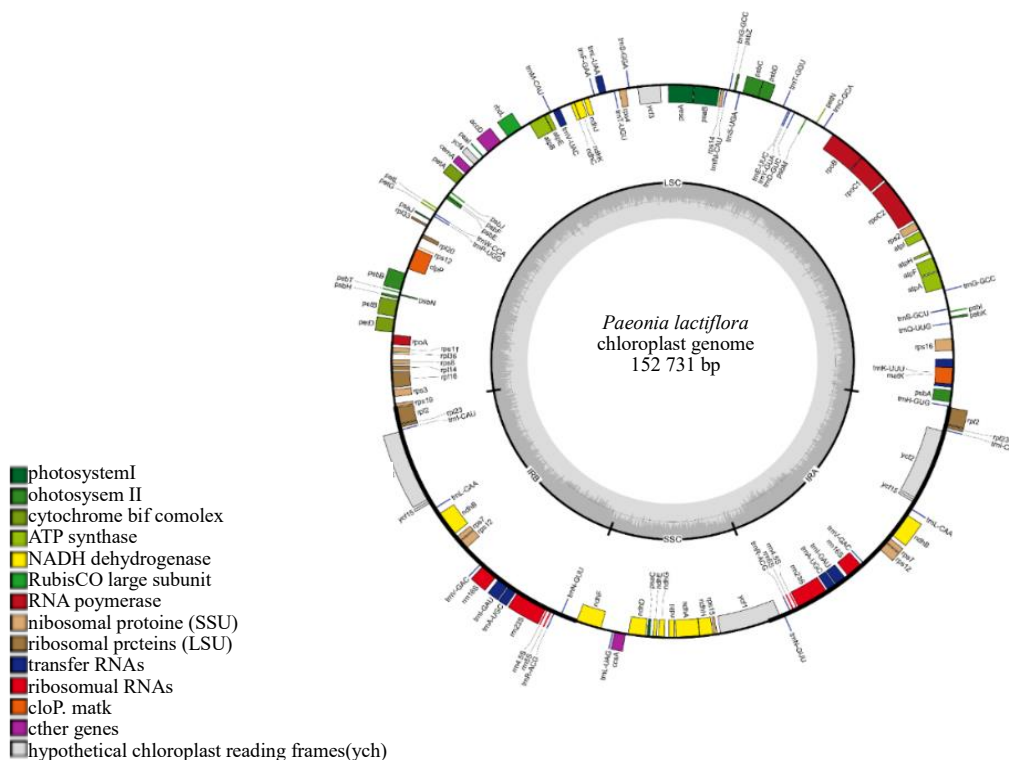


图 1 芍药叶绿体基因组图谱

Fig. 1 Gene map of *P. lactiflora* chloroplast genome

表 3 4 种芍药属植物叶绿体基因组序列的基本特征信息
Table 3 Basic characteristics among four medicinal plants chloroplast genomes of *Paeonia*

基因组特征	芍药	川赤芍	牡丹	杨山牡丹
基因组大小/bp	152 731	152 682	152 608	152 768
总GC量/%	38.43	38.42	38.36	38.34
LSC长度/bp	84 403	84 399	84 270	84 396
SSC长度/bp	16 970	16 980	17 049	17 079
IR长度/bp	25 680	25 653	25 646	25 648

3.2 芍药叶绿体基因组功能及分类

叶绿体注释结果表明，芍药叶绿体基因组共有 126 个基因，其功能大致可以分为 4 类，包括参与光合作用的基因 44 个、参与自我复制的基因 70 个、

其他基因 5 个，未知功能基因 7 个。编码 82 个蛋白质基因中 *ycf2*、*ycf15*、*rps7*、*rps12*、*rpl23*、*rpl2*、*ndhB* 有 2 个拷贝；36 个 tRNA 基因中 *trnA*-UGG、*trnG*-GCC、*trnI*-CAU、*trnI*-GAU、*trnL*-CAA、*trnN*-GUU、*trnR*-ACG、*trnV*-GAC 有 2 个拷贝；8 个 rRNA 基因中 *rrn16S*、*rrn23S*、*rrn4.5S* 和 *rrn5S* 均有 2 个拷贝（表 4）。在蛋白质编码基因中共发现 16 个含有内含子（Intron）的基因，其中 *clpP* 和 *ycf3* 含有 2 个内含子，其余 14 个基因 *atpF*、*ndhA*、*ndhB*、*petB*、*petD*、*rpl16*、*rpl2*、*rpoC1*、*rps16*、*trnA*-UGC、*trnI*-GAU、*trnK*-UUU、*trnL*-UAA 和 *trnV*-UAC 则仅有 1 个内含子（表 5）。

表 4 芍药叶绿体基因组注释信息

Table 4 Annotation information of chloroplast genome in *P. lactiflora*

基因类型	基因分组	基因名称	数量
光合作用	subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	44
	subunits of photosystem II	<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>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	
	subunits of NADH dehydrogenase	<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>	
	subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> [*] , <i>petD</i> [*] , <i>petG</i> , <i>petL</i> , <i>petN</i>	
	subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> [*] , <i>atpH</i> , <i>atpI</i>	
	large subunit of rubisco	<i>rbcL</i>	
自我复制基因	proteins of large ribosomal subunit	<i>rpl14</i> , <i>rpl16</i> [*] , <i>rpl2</i> ^{*(2)} , <i>rpl20</i> , <i>rpl23</i> (2), <i>rpl33</i> , <i>rpl36</i>	70
	proteins of small ribosomal subunit	<i>rps11</i> , <i>rps12</i> ^{**} (2), <i>rps14</i> , <i>rps15</i> , <i>rps16</i> [*] , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (2), <i>rps8</i>	
	subunits of RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> [*] , <i>rpoC2</i>	
	ribosomal RNAs	<i>rrn16S</i> (2), <i>rrn23S</i> (2), <i>rrn4.5S</i> (2), <i>rrn5S</i> (2)	
	transfer RNAs	<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(2), <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, <i>trnN</i> -GUU(2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG(2), <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, <i>trnY</i> -CAU	
其他基因	maturase	<i>matK</i>	5
	protease	<i>clpP</i> ^{**}	
	envelope membrane protein	<i>cemA</i>	
	acetyl-CoA carboxylase	<i>accD</i>	
	c-type cytochrome synthesis gene	<i>ccsA</i>	
未知功能基因	conserved hypothetical chloroplast ORF	<i>ycf1</i> , <i>ycf15</i> (2), <i>ycf2</i> (2), <i>ycf3</i> ^{**} , <i>ycf4</i>	7
总计			126

*含有 1 个内含子; **含有 2 个内含子; (2): 双拷贝基因。
*means one intron; **means two introns; (2) means double-copy genes.

表 5 芍药叶绿体基因组基因外显子和内含子信息

Table 5 Information on exons and introns of genes in chloroplast genome of *P. lactiflora*

基因	位置	外显子I/bp	内含子I/bp	外显子II/bp	内含子II/bp	外显子III/bp
<i>atpF</i>	LSC	160	691	410		
<i>clpP</i>	LSC	71	678	291	658	226
<i>ndhA</i>	SSC	544	1 015	539		
<i>ndhB</i>	IR	775	682	758		
<i>petB</i>	LSC	6	767	648		
<i>petD</i>	LSC	9	696	474		
<i>rpl16</i>	LSC	9	1 002	402		
<i>rpl2</i>	IR	397	665	434		
<i>rpoC1</i>	LSC	430	700	1 622		
<i>rps16</i>	LSC	46	812	233		
<i>trnA</i> -UGC	IR	40	715	35		
<i>trnI</i> -GAU	IR	42	933	35		
<i>trnK</i> -UUU	LSC	37	2 462	25		
<i>trnL</i> -UAA	LSC	37	524	50		
<i>trnV</i> -UAC	LSC	39	573	37		
<i>ycf3</i>	LSC	124	735	230	749	153

3.3 重复序列分析

在芍药属 4 种药用植物叶绿体基因组中, 芍药含有的长重复序列数量最多, 为 48 个, 川赤芍和牡丹次之, 均为 43 个, 而杨山牡丹长重复序列数量最少, 为 39 个。芍药属 4 种药用植物叶绿体基因组均含有正向重复和回文重复, 正向重复总数和回文重复分别介于 19~23 和 20~22。在正向重复中, 30~39 bp 的重复单元数量最多, 介于 11~14; 仅芍药含有 60~69 bp 的正向重复单元; 在回文重复中, 30~39 bp 的重复单元数量最多, 介于 10~13; 仅芍药和川赤芍含有 60~69 bp 的回文重复单元, 数量均为 2。在芍药属 4 种药用植物叶绿体基因组中, 只有芍药和川赤芍含有反向重复, 且反向重复的重复单元碱基长度仅为 30~39 bp 一类, 数目分别为 2 个和 1; 只有芍药含互补重复, 且互补重复单元碱基长度仅为 30~39 bp 一类, 数目仅为 2 个。表明只有芍药同时包含正向重复、反向重复、互补重复和回文重复, 而川赤芍缺少互补重复, 牡丹和杨山牡丹缺少反向重复和互补重复。

3.4 简单重复序列分析

简单重复序列 (simple sequence repeats, SSR) 是一段串联的短小 DNA 序列, 大小为 1~6 bp, 在叶绿体基因组中广泛分布。SSR 分析表明芍药和牡丹中 SSR 数量最多, 均为 52 个, 川赤芍次之为 51 个, 杨山牡丹中 SSR 数量最少为 48 个。

4 种植物的 SSR 类型中单核苷酸重复占比 81%~86%, 单核苷酸重复中绝大多数为 A/T 重复, 牡丹、芍药和杨山牡丹单核苷酸 C/G 重复中既含 C 重复, 也含 G 重复, 但川赤芍中仅含 G 重复; 芍药属 4 种药用植物中二核苷酸重复中均仅含有 AT/TA 重复, 不含 CG/GC 重复; 仅芍药叶绿体基因组中含有三核苷酸重复, 为 AAT。4 种芍药属叶绿体基因组中均缺少四核苷酸、五核苷酸和六核苷酸的简单重复序列。

3.5 密码子偏好性分析

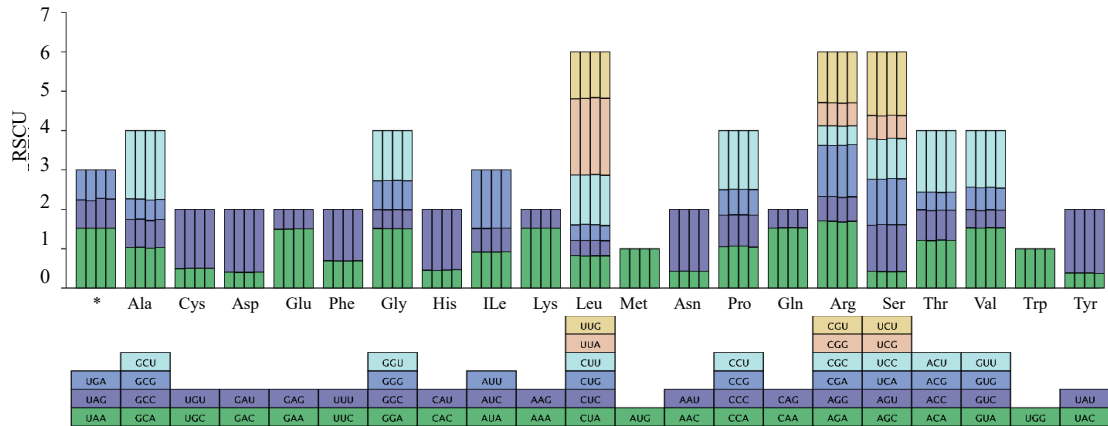
密码子使用频次统计表明, 芍药、川赤芍、牡丹和杨山牡丹的叶绿体基因组均编码 64 种密码子, 密码子总使用频次分别达 22 170、22 383、22 599 和 22 327 次, 其中 61 种密码子编码 20 种氨基酸, 其余 3 种密码子为终止密码子。在编码的氨基酸中, 亮氨酸 (Leu) 出现频率最高为 10.48%~10.49%, 异亮氨酸 (Ile) 次之, 为 8.64%~8.70%, 半胱氨酸 (Cys) 使用频率最低为 1.11%~1.12% (表 6)。

除编码甲硫氨酸的 AUG 和编码色氨酸的 UGG 外, 芍药属 4 种药用植物对其他 62 种密码子都具有偏好性且偏好性基本一致 (图 2)。其中 UUA 等 31 个密码子的 RSCU 大于 1, 以上偏好性密码子以 A 结尾的有 13 个、以 U 结尾的有 16 个、以 G 和 C 结尾均只有 1 个, 绝大多数以 A/U 结尾, 说明密码子偏好 A 或 U 结尾。而 AUA 等 31 个密码子对密

表 6 4 种芍药属植物叶绿体基因组中氨基酸出现的次数及占比

Table 6 Number and proportion of amino acid occurrence in four medicinal plants chloroplast genomes of *Paeonia*

芍药			川赤芍			牡丹			杨山牡丹		
氨基酸	次数	占比/%	氨基酸	次数	占比/%	氨基酸	次数	占比/%	氨基酸	次数	占比/%
Leu	2 325	10.49	Leu	2 355	10.52	Leu	2 370	10.49	Leu	2 339	10.48
Ile	1 915	8.64	Ile	1 937	8.65	Ile	1 967	8.70	Ile	1 937	8.68
Ser	1 648	7.43	Ser	1 670	7.46	Ser	1 700	7.52	Ser	1 657	7.42
Gly	1 581	7.13	Gly	1 592	7.11	Gly	1 609	7.12	Gly	1 587	7.11
Arg	1 305	5.89	Arg	1 326	5.92	Arg	1 335	5.90	Arg	1 333	5.97
Ala	1 284	5.79	Ala	1 283	5.73	Ala	1 293	5.72	Ala	1 289	5.77
Phe	1 254	5.66	Phe	1 268	5.67	Phe	1 273	5.63	Phe	1 260	5.64
Val	1 230	5.55	Val	1 236	5.52	Val	1 249	5.53	Val	1 237	5.54
Glu	1 153	5.20	Lys	1 168	5.22	Glu	1 169	5.17	Lys	1 164	5.21
Lys	1 147	5.17	Thr	1 160	5.18	Lys	1 163	5.15	Glu	1 159	5.19
Thr	1 147	5.17	Glu	1 159	5.18	Thr	1 162	5.14	Thr	1 146	5.13
Asn	1 041	4.70	Asn	1 051	4.70	Asn	1 063	4.70	Asn	1 047	4.69
Pro	952	4.29	Pro	962	4.30	Pro	979	4.33	Pro	961	4.30
Asp	894	4.03	Asp	898	4.01	Asp	909	4.02	Asp	901	4.04
Tyr	812	3.66	Tyr	816	3.65	Tyr	836	3.70	Tyr	820	3.67
Gln	788	3.55	Gln	797	3.56	Gln	801	3.54	Gln	793	3.55
His	529	2.39	His	531	2.37	Met	533	2.36	His	528	2.37
Met	522	2.36	Met	529	2.36	His	532	2.35	Met	524	2.35
Trp	394	1.78	Trp	396	1.77	Trp	403	1.78	Trp	395	1.77
Cys	249	1.12	Cys	249	1.11	Cys	253	1.12	Cys	250	1.12



自左至右分别代表芍药、川赤芍、牡丹和杨山牡丹。

From left to right: *P. lactiflora*, *P. veitchii*; *P. suffruticosa*, *P. ostii*, respectively.

图 2 芍药属 4 种药用植物叶绿体基因组相对同义密码子使用度

Fig.2 Relative synonymous codon usage in four medicinal plants chloroplast genome of *Paeonia*

码子的使用频次较少表现弱偏好性，密码子偏好性与碱基组成特别是密码子第 3 位碱基有显著相关性。

3.6 IR 边界收缩和扩张分析

芍药属 4 种药用植物叶绿体基因组的 IR 区均与 LSC 区和 SSC 区存在边界，即 JLB (LSC/IRb)、JSB (IRb/SSC)、JSA (SSC/IRa) 和 JLA (IRa/LSC)。4 个边界附近的基因类型相同，相对保守，但其边界仍具有一定差异。芍药属 4 种药用植物 JLB 边界均位于 *rps19* 基因内部，其中芍药和川赤芍 *rps19* 基因均向 IRb 区扩张 8 bp，牡丹和杨山牡丹 *rps19* 基因向 IRb 区扩张 2 bp；在 JSB 边界，川赤芍和杨山牡丹的 *ycf1* 基因向 SSC 区分别扩张了 3 bp 和 1 bp，而芍药和牡丹缺乏 *ycf1* 基因；牡丹和杨山牡丹的 *ndhF* 基因向 IRb 区扩张了 2 bp，而芍药和川赤芍的 *ndhF* 基因全长均位于 SSC 区；JSA 边界均位于 *ycf1* 基因内部，但芍药属 4 种药用植物 *ycf1* 基因长度及

其在 SSC 区和 IRa 区的长度略有不同；JLA 边界均位于 *trnH* 基因内部，芍药、川赤芍和杨山牡丹的 *trnH* 基因均向 IRa 扩张了 3 bp，而牡丹向 IRa 扩张了 4 bp（图 3）。

3.7 选择压力分析

选择压力是指环境或其他因素对生物个体或种群施加的一种压力，导致具有不同特征的个体在生存和繁殖上产生差异。它是推动生物进化的重要动力之一。本研究对芍药属 4 种药用植物叶绿体基因组中蛋白编码基因进行了选择压力分析。如图 4 显示，绝大多数基因两两相比的 K_a/K_s 值均小于 1，表明这些基因受到纯化选择，推测它们在叶绿体基因组行使功能的过程中具有非常重要的作用，因此非常保守。牡丹 vs 芍药、杨山牡丹 vs 芍药的 *matK* 和 *ndhB* 的 K_a/K_s 均大于 1，杨山牡丹 vs 芍药、牡丹 vs 杨山牡丹 *rpoA* 的 K_a/K_s 达 4.27 和 2.22，表明

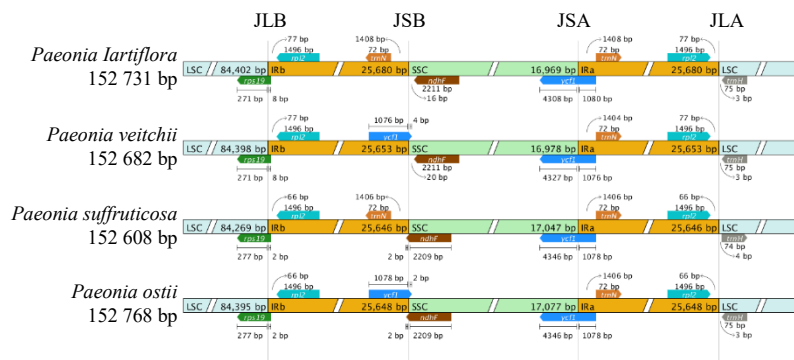


图 3 芍药属 4 种药用植物叶绿体基因组 LSC、IRs 和 SSC 边界区域比较

Fig. 3 Comparison of LSC, IRs and SSC border regions of chloroplast genome in four medicinal plants of *Paeonia*



图 4 4 种芍药属药用植物部分共有蛋白编码基因的 K_a/K_s 值

Fig. 4 K_a/K_s value of some common protein-coding genes of four medicinal plants in *Paeonia*

matK、*ndhB* 和 *rpoA* 受到正向选择，推测这些基因在牡丹、芍药和杨山牡丹驯化过程中发挥重要作用。

3.8 叶绿体基因组序列结构变异分析

为评估芍药属 4 种药用植物叶绿体基因组序列的差异程度，本研究以芍药 (OP205316) 为参考基因组，采用 mVISTA 在线工具，分别与川赤芍、牡丹和杨山牡丹的叶绿体基因组全长序列进行比对分析 (图 5)。由图 5 可见，与参考基因组相比，3 条叶绿体基因序列中非编码区变异高于蛋白编码区域；LSC 区、SSC 区变异明显大于 IR 区。整体来看，川赤芍与芍药之间的变异相对较小，牡丹、杨山牡丹与芍药之间的变异则更为明显，而牡丹与杨山牡丹存在很多相似的变异，这与芍药和川赤芍隶属于芍药组，牡丹和杨山牡丹隶属于牡丹组的分类地位一致。

由图 6 可见，在芍药属 4 种药用植物叶绿体全基因组的核酸多态性分析结果中，高多态性区间主要集中在 LSC 和 SSC 区间内，而 IR 区相对保守，多态性则较低，这与结构变异分析的结果一致。其中 *trnH-psbA*、*rpoC1* (exon2) - *rpoC1* (exon1)、*accD-psaI*、*petA-psbJ*、*rpl33-rpl20*、*rps3-rps19* 共 6 个区间及 *ndhF* 和 *ycf1* 共 2 个基因编码区的多态性

较高，可作为这 4 种药用植物条形码开发的候选片段及分子鉴定的候选基因。

3.9 芍药科和毛茛科植物的亲缘关系分析

以虎耳草科 (2 种)、蔷薇科 (1 种) 和葡萄科 (1 种) 作为外类群，基于叶绿体基因组共有基因序列的系统进化分析表明，所有芍药科植物被聚为一个分支，芍药科可被分为牡丹组、北美芍药组和芍药组，且牡丹组又被分为肉质花盘亚组和革质花盘亚组 (图 7)，符合中国植物志分类关系和 APG 分类关系。外类群把芍药科和毛茛科植物分隔开来，毛茛科基于叶绿体基因组 CDS 序列的系统进化分析结果偏离中国植物志分类关系，但隶属于唐松草亚科的 4 种植物直距耧斗菜 *A. rockii* Munz、距瓣尾囊草 *U. rockii* O. E. Ulbr.、东北扁果草 *I. manshuricum* Kom. 和丝叶唐松草 *T. foeniculaceum* Bunge 形成 1 个分支，绝大多数隶属于金莲花属的植物形成另外 1 个分支，仅鸡爪草 *C. oxycarpa* Sprague 和矮金莲花 *T. farreri* Stapf 介于唐松草亚科和金莲花亚科所在的分支之间；另外，另外隶属于毛茛亚科的厚叶美花草 *C. alata* Freyn 和水毛茛 *R. bungei* Steud. 与其他金莲花科植物形成 1 个分支。

芍药科和外类群虎耳草科植物聚类在一起，符合 APG 分类系统二者同属于虎耳草目。虎耳草目和蔷薇

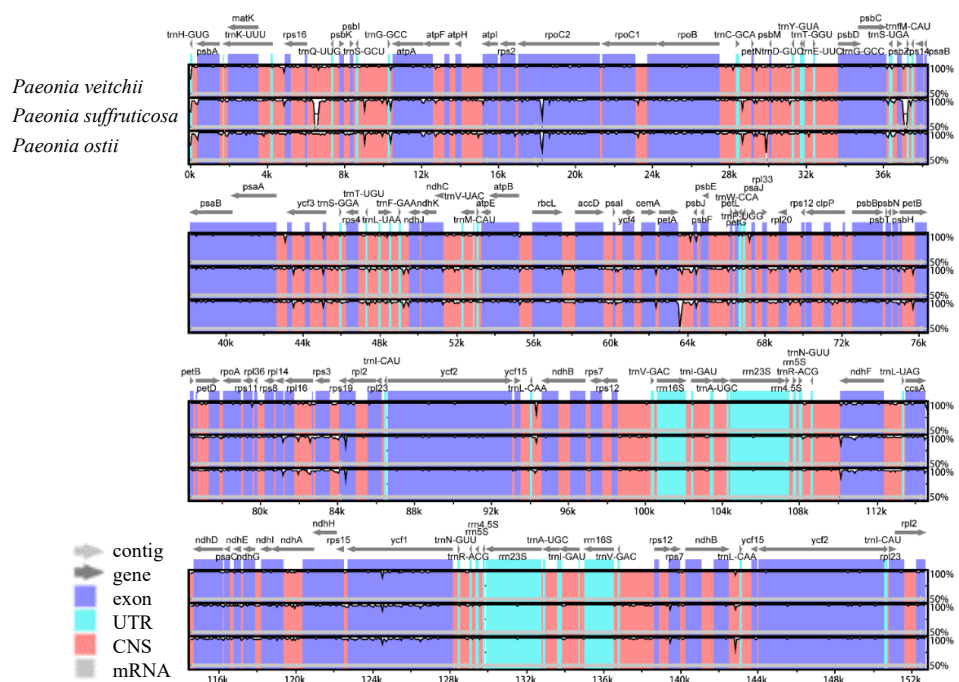


图 5 4 种芍药属植物叶绿体基因组全局比对分析

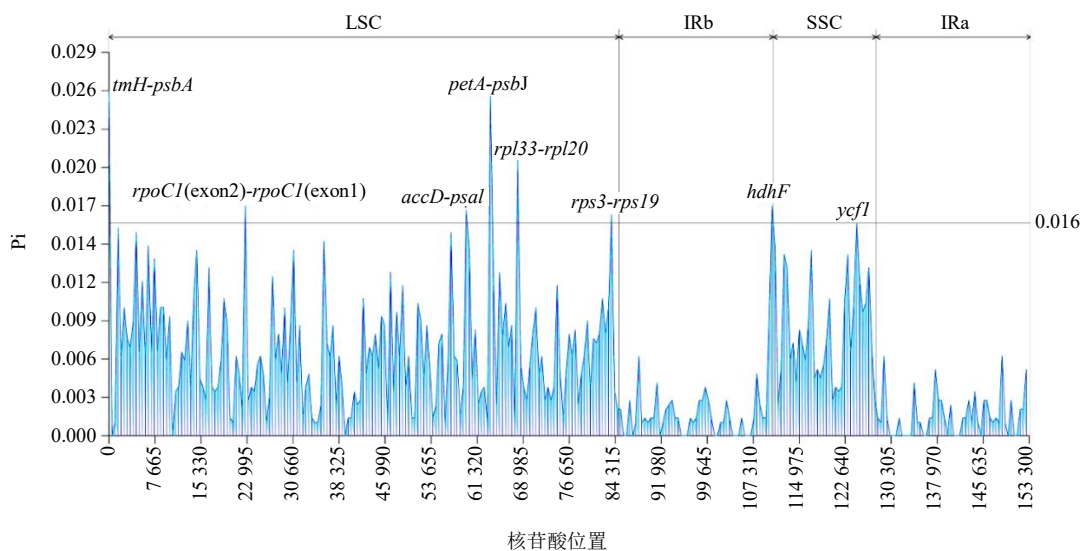
Fig. 5 Global alignment analysis of chloroplast genome of four medicinal plants in *Paeonia*

图 6 芍药属 4 种药用植物叶绿体基因组核酸多态性分析

Fig. 6 Nucleic acid polymorphism analysis of chloroplast genome of four medicinal plants in *Paeonia*

科的月季聚类在一起符合 APG 分类系统二者同属于蔷薇系。虎耳草目和月季组成的蔷薇系与葡萄目的葡萄聚类在一起，符合 APG 分类系统它们同属于被子植物门、真子叶植物纲下的超蔷薇类。

4 讨论

中国是芍药属植物的物种起源中心之一，在芍药属植物的分类、分布及演化等方面占据关键地

位。全球芍药属植物共有 34 种，被划分为牡丹组、芍药组和北美芍药组。不同组在植物形态特征和地理分布上存在差异，其中牡丹组 9 种皆为中国特有，芍药组 8 种在中国分布^[31]。芍药属植物含多种有效成分，这些成分在抗炎、免疫调节、保护心血管等方面发挥作用，被用于多种疾病的治疗和预防^[32]。

本研究完成了芍药叶绿体基因组测序并对芍

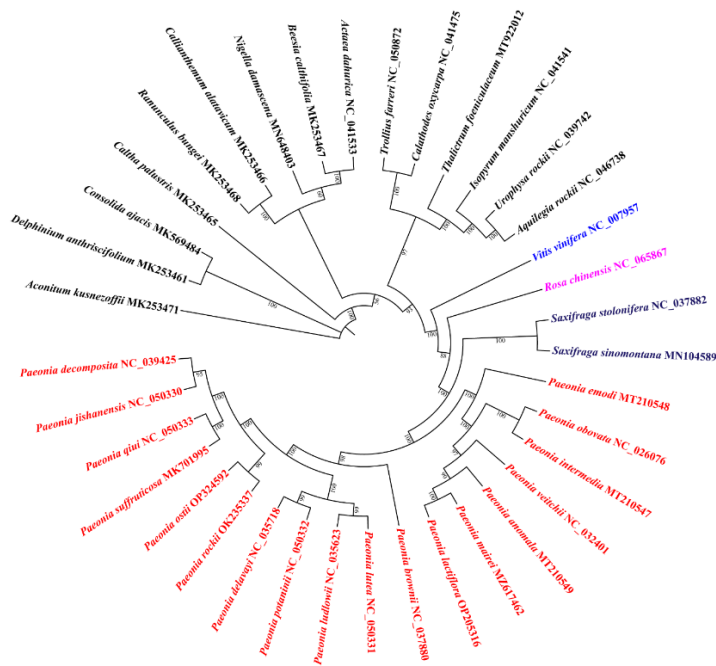


Fig. 7 Construction of ML phylogenetic tree among *Paeonia* and Ranunculaceae based on common CDS

药属 4 种药用植物的叶绿体基因组进行了比较分析。芍药叶绿体基因组和芍药属其他物种一样为双链环状四分结构^[10], 序列长度为 152 731 bp, 由 SSC、LSC、IRa 和 IRb 共 4 部分组成, GC 含量为 38.43%, 共注释到基因 126 个, 其中 82 个蛋白质编码基因、36 个 tRNA 基因和 8 个核糖体 rRNA 基因。芍药、川赤芍、牡丹和杨山牡丹的叶绿体基因组全长 152 608~152 768 bp, 其中 LSC 长度 84 270~84 403 bp, SSC 长度 16 970~17 079 bp, 而 2 个相同的反向重复序列 IRa 和 IRb 长度 25 646~25 680 bp。由此可见芍药属植物的叶绿体基因组在大小和基因组成等方面高度保守^[33], SSC 长度不一是芍药属药用植物叶绿体基因组大小存在差异的重要原因。

重复序列在植物的叶绿体基因组中可维持叶绿体基因组的稳定性^[34]。芍药的重重复序列数多达 48 个, 且仅芍药叶绿体基因组中含有正向、回文、反向及互补 4 种重复序列。SSR 分析结果表明, 芍药属 4 种药用植物的 SSR 类型多为单核苷酸重复, 其中 A/T 重复数最多, 且多偏向于 T, 这种现象普遍存在于芍药属其他物种叶绿体基因组中^[10]。研究报告叶绿体 SSR 分子标记可用于评估牡丹栽培种的多样性水平和群体结构, 从叶绿体基因组角度理解牡丹的种质特性^[35], 这为后续芍药属药用植物 SSR 分子标记的开发和应用奠定了理论基础。

密码子分析结果表明,芍药属 4 种药用植物编码的氨基酸中,亮氨酸出现频率均最高,异亮氨酸次之,半胱氨酸使用频率最低。有 31 种密码子的 RSCU 值大于 1,其中 UUA、GCU 和 AGA 的 RSCU 值排名前 3,研究报道这 3 种密码子为芍药属叶绿体基因组种的优化密码子^[36]。

IR 边界收缩和扩张会改变基因剂量,影响基因表达水平,还可能影响基因结构和功能,进而对植物的光合作用、生长发育、繁殖及环境适应等方面产生影响。IR 区边界分析发现牡丹和杨山牡丹之间、芍药和川赤芍之间在 IR 区边界特征上更为相似,这与它们分别隶属于牡丹组和芍药组的分类地位一致^[31]。

选择压力分析结果表明芍药属 4 种药用物种叶绿体基因组中的蛋白编码基因受到的选择压力存在差别。绝大多数基因非常保守, 受到纯化选择, 推测它们在叶绿体基因组行使功能的过程中具有非常重要的作用^[37]; 而仅有 *matK*、*ndhB* 和 *rpoA* 的 K_a/K_s 大于 1, 受正向选择。受正向选择的 3 个基因 *matK*、*ndhB* 和 *rpoA* 分别在虎耳草科金腰属 *Chrysosplenium* L.^[38]、葛 *Pueraria montana* var. *lobata* (Willd.) Sanjappa & Pradeep^[39] 和姜黄属 *Curcuma* L.^[40] 已有报道。*matK* 基因编码的成熟酶 K 参与叶绿体基因转录后的加工, 可精准剪接 RNA, 保证叶绿体基因正常表达。当环境改变, 正向选择

推动 *matK* 基因改变, 让植物调整光合作用相关蛋白的表达, 增强对环境的适应能力。*ndhB* 基因参与叶绿体呼吸链中的 NADH 脱氢酶复合体的组成。正向选择促使 *ndhB* 基因发生有利于提高其功能的变异, 优化 NADH 脱氢酶复合体的结构和活性, 增强电子传递效率, 提升叶绿体对光能的捕获、转化及利用效率。*rpoA* 基因编码 RNA 聚合酶的 α 亚基, 对转录起始和转录复合物组装起关键作用。正向选择使 *rpoA* 基因积累有益突变, 优化 RNA 聚合酶结构与功能, 提升转录效率。正向选择还让 *rpoA* 基因更精准地调控基因表达, 为叶片行使功能奠定基础。牡丹、芍药和杨山牡丹较川赤芍驯化程度高, 推测 *matK*、*ndhB* 和 *rpoA* 在牡丹、芍药和杨山牡丹的分化和驯化过程中发挥了重要作用。

叶绿体全基因组变异分析结果表明芍药组的芍药和川赤芍、牡丹组的牡丹和杨山牡丹叶绿体基因组之间的序列相似性更高, 这表明亲缘关系更近物种的叶绿体基因组更加保守^[35]。核酸多态性分析结果表明, 高多态性区间主要集中在 LSC 和 SSC 区间内, 而 IR 区相对保守, 多态性则较低^[41]。其中 *trnH-psbA*、*rpoC1* (exon2) -*rpoC1* (exon1)、*accD-psaI*、*petA-psbJ*、*rpl33-rpl20*、*rps3-rps19* 共 6 个基因间区及 *ndhF* 和 *ycf1* 共 2 个基因编码区的多态性较高, 它们可作为芍药属 4 种药用植物条形码开发的候选片段^[42], 为芍药属药用植物的分子鉴定奠定理论基础。

毛茛科是被子植物的原始类群之一, 其系统关系和科下演化关系一直备受争议^[43]。芍药科 (Paeoniaceae) 植物世界上仅一属 (芍药属 *Paeonia* L.), 分布于亚洲、欧洲、北美洲及非洲温带地区。恩格勒系统、边沁系统和虎克系统等世界上应用最广的分类系统最初都把芍药科置于毛茛科中^[9]。1908 年英国学者 Worsdell 发现芍药属植物的雄蕊群离心发育与毛茛科不同, 将其独立为芍药科, 但仍放在毛茛目中。越来越多的形态学、解剖学、孢粉学及分子证据表明, 芍药属是区别于毛茛科其他植物的独立类群^[10-11]。至 20 世纪 70 年代, 芍药属独立成科, 其分类地位在国际上获得一致支持, 被置于虎耳草目下^[44]。《中国药典》2020 年版将牡丹、芍药、川赤芍等芍药属药用植物归类于毛茛科下^[8]。芍药属药用植物的分类地位与国际上芍药属植物的分类地位的不一对于丹皮、白芍和赤芍等中药材的鉴定和质量把控、药理研究和临床应用都将产

生不利影响。本研究以虎耳草科的山地虎耳草和虎耳草、蔷薇科的月季和葡萄科的葡萄作为外类群, 基于叶绿体基因组共有基因序列进行系统进化分析, 结果表明芍药和美丽芍药的亲缘关系最近, 所有芍药科植物被聚为一个分支, 且芍药科形成牡丹组、北美芍药组和芍药组 3 个亚支, 牡丹组又形成肉质花盘亚组和革质花盘亚组 2 个亚亚支, 这与基于形态学的芍药属植物分类一致^[31], 表明叶绿体基因组可以作为芍药属植物分类的分子依据。外类群把芍药属和毛茛科植物分隔开来, 表明芍药属和虎耳草科、蔷薇科和葡萄科的亲缘根系更近, 而与毛茛科的亲缘关系更远^[45]。建议《中国药典》下一轮修订时召开学术研讨会, 讨论将入药的牡丹、芍药、川赤芍和杨山牡丹的分类地位与国际接轨。

本研究通过芍药叶绿体基因组的测序、组装及注释, 并结合同属 3 个药用植物的叶绿体基因组比较分析, 阐述了其变异程度及系统发育关系。叶绿体基因组比较分析结果表明, 芍药属 4 种药用植物在叶绿体基因组结构特征方面较为保守, 但其变异程度存在差异, 芍药与美丽芍药亲缘关系最近, 芍药属被聚为一个分支, 外类群把芍药属和毛茛科植物分隔开来, 芍药属植物与毛茛科的亲缘关系较远; 鉴定出的 6 个基因间区和 2 个基因编码区可作为这 4 种药用植物条形码开发的候选片段。研究结果为芍药属药用植物的分子鉴定、系统发育及资源保护等研究提供了一定参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 李嘉珏. 中国牡丹种质资源研究与利用 [M]. 郑州: 中原农民出版社, 2023: 35.
- [2] 张树蓉, 赵宏苏, 佟沫儒, 等. 牡丹皮化学成分、药理作用及其质量标志物 (Q-Marker) 的预测分析 [J]. 中草药, 2022, 53(16): 5215-5224.
- [3] 严安定, 宋帅, 孙银华, 等. 凤丹皮中丹皮酚类和芍药苷类成分含量的经时变化特征分析 [J]. 中药材, 2019, 42(3): 541-544.
- [4] 马王堆汉墓帛书整理小组. 五十二病方 [M]. 北京: 文物出版社, 1979: 47.
- [5] 南朝·梁·陶弘景编. 尚志钧尚元胜辑校. 本草经集注: 辑校本 [M]. 北京: 人民卫生出版社, 1994: 267.
- [6] 杨颜芳, 王晶娟, 张贵君, 等. 白芍与赤芍及其醇提物的红外光谱法整体结构解析及鉴定 [J]. 中草药, 2016, 47(19): 3508-3512.
- [7] 李静, 曹兰秀, 郑文. 白芍常用药对配伍探析 [J]. 中

- 医学报, 2024, 39(6): 1214-1219.
- [8] 中国药典 [S]. 2005 年版增补本. 2009.
- [9] 郭先锋. 中国芍药分类学研究进展 [J]. 北京林业大学学报, 2002, 24(3): 99-102.
- [10] Wu L W, Nie L P, Wang Q, *et al.* Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae [J]. *Sci Rep*, 2021, 11(1): 14643.
- [11] Zhou S L, Xu C, Liu J, *et al.* Out of the pan-Himalaya: Evolutionary history of the Paeoniaceae revealed by phylogenomics [J]. *J Syst Evol*, 2021, 59(6): 1170-1182.
- [12] Lv S Z, Cheng S, Wang Z Y, *et al.* Draft genome of the famous ornamental plant *Paeonia suffruticosa* [J]. *Ecol Evol*, 2020, 10(11): 4518-4530.
- [13] Yuan J H, Jiang S J, Jian J B, *et al.* Genomic basis of the giga-chromosomes and giga-genome of tree peony *Paeonia ostii* [J]. *Nat Commun*, 2022, 13(1): 7328.
- [14] Guo L L, Guo S, Xu J, *et al.* Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony [J]. *Ind Crops Prod*, 2020, 153: 112567.
- [15] Chen S F, Zhou Y Q, Chen Y R, *et al.* Fastp: An ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): i884-i890.
- [16] 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.
- [17] Hunt M, Silva N D, Otto T D, *et al.* Circlator: Automated circularization of genome assemblies using long sequencing reads [J]. *Genome Biol*, 2015, 16: 294.
- [18] 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.
- [19] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47(W1): W59-W64.
- [20] 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.
- [21] Beier S, Thiel T, Münch T, *et al.* MISA-web: A web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [22] Zhu Y Y, Zhang X W, Yan S F, *et al.* SSR identification and phylogenetic analysis in four plant species based on complete chloroplast genome sequences [J]. *Plasmid*, 2023, 125: 102670.
- [23] Huang L J, Yu H X, Wang Z, *et al.* CPStools: A package for analyzing chloroplast genome sequences [J]. *iMetaOmics*, 2024, 1(2): e25.
- [24] Amiryousefi A, Hyvönen J, Poczei P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [25] Zhang Z. KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences [J]. *Genomics Proteomics Bioinformatics*, 2022, 20(3): 536-540.
- [26] Poliakov A, Foong J, Brudno M, *et al.* GenomeVISTA: An integrated software package for whole-genome alignment and visualization [J]. *Bioinformatics*, 2014, 30(18): 2654-2655.
- [27] Zhang D, Gao F L, Jakovlić I, *et al.* PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. *Mol Ecol Resour*, 2020, 20(1): 348-355.
- [28] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.
- [29] Nguyen L T, Schmidt H A, von Haeseler A, *et al.* IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies [J]. *Mol Biol Evol*, 2015, 32(1): 268-274.
- [30] Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation [J]. *Nucleic Acids Res*, 2021, 49(W1): W293-W296.
- [31] 洪德元. 世界牡丹和芍药研究: 兼论分类学的基本原理 [M]. 北京: 科学出版社, 2024: 56.
- [32] 薛璇. 典型芍药科药材中物质成分、赋存特征及其环境健康研究 [D]. 合肥: 中国科学技术大学, 2022.
- [33] 周晓君, 张凯, 彭正锋, 等. 矮牡丹与芍药属其他 5 种叶绿体基因组特征的比较 [J]. 林业科学, 2020, 56(4): 82-88.
- [34] Wu Y, Liu F, Yang D G, *et al.* Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny [J]. *Front Plant Sci*, 2018, 9: 376.
- [35] Guo Q, Xue X, Wang D D, *et al.* Genetic diversity and population genetic structure of *Paeonia suffruticosa* by chloroplast DNA simple sequence repeats (cpSSRs) [J]. *Hortic Plant J*, 2025, 11(1): 367-376.
- [36] Chen Q H, Chen L, Teixeira da Silva J A, *et al.* The plastome reveals new insights into the evolutionary and domestication history of peonies in East Asia [J]. *BMC Plant Biol*, 2023, 23(1): 243.
- [37] 马海霞, 张亚颖, 张玉婉, 等. 角蒿属药用植物叶绿体基因组比较及系统发育分析 [J]. 中草药, 2024, 55(17):

- 5972-5981.
- [38] Wu Z H, Liao R, Yang T G, *et al.* Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae) [J]. *BMC Genomics*, 2020, 21(1): 621.
- [39] Zhou Y, Shang X H, Xiao L, *et al.* Comparative plastomes of *Pueraria montana* var. *Lobata* (Leguminosae: Phaseoleae) and closely related taxa: Insights into phylogenomic implications and evolutionary divergence [J]. *BMC Genomics*, 2023, 24(1): 299.
- [40] Gui L J, Jiang S F, Xie D F, *et al.* Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution [J]. *Gene*, 2020, 732: 144355.
- [41] Cai H B, Xu R, Tian P, *et al.* Complete chloroplast genomes and the phylogenetic analysis of three native species of Paeoniaceae from the Sino-Himalayan flora subKingdom [J]. *Int J Mol Sci*, 2023, 25(1): 257.
- [42] Yang X, Yu X L, Zhang X Y, *et al.* Development of mini-barcode based on chloroplast genome and its application in metabarcoding molecular identification of Chinese medicinal material *Radix Paeoniae Rubra* (Chishao) [J]. *Front Plant Sci*, 2022, 13: 819822.
- [43] Emadzade K, Lehnebach C, Lockhart P, *et al.* A molecular phylogeny, morphology and classification of genera of Ranunculeae (Ranunculaceae) [J]. *Taxon*, 2010, 59(3): 809-828.
- [44] Group T A P. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV [J]. *Bot J Linn Soc*, 2016, 181(1): 1-20.
- [45] 穆赢通, 樊东昌, 吕丽娟, 等. 毛茛科和芍药科叶绿体基因组密码子特征和系统发育比较 [J]. *植物研究*, 2022, 42(6): 964-975.

[责任编辑 时圣明]