

美丽芍药叶绿体全基因组解析及系统发育分析

张明英, 王西芳, 高 静, 刘阿萍, 颜永刚, 杨新杰, 张 岗*

(陕西中医药大学药学院/陕西省秦岭中草药应用开发工程技术研究中心, 陕西 西安 712046)

摘要: 本研究利用 Illumina HiSeq X Ten 测序平台对芍药属药用植物美丽芍药进行了叶绿体全基因组测序, 通过生物信息学分析方法进行序列组装、注释和特征解析, 并将其与芍药属其他 11 种植物的叶绿体全基因组进行了比较基因组学和系统发育分析。结果显示, 美丽芍药叶绿体基因组全长 152 731 bp, GC 含量为 38.4%, 具有被子植物叶绿体基因组典型的四分体结构, 包括一个大单拷贝区 (large single copy, LSC)、一个小单拷贝区 (small single copy, SSC) 和一对反向重复区 (inverted repeat sequence, IRa/IRb), 长度分别为 84 402、16 969 和 25 680 bp; 共注释得到 136 个基因, 包括 90 个蛋白编码基因, 38 个 tRNA 基因和 8 个 rRNA 基因, 其中 7 个蛋白编码基因、7 个 tRNA 基因和 4 个 rRNA 基因分别在反向重复区发生了一次重复; 此外, 在美丽芍药叶绿体基因组中共检测出 28 个散在重复序列 (dispersed repeats)、10 个串联重复序列 (tandem repeats) 和 64 个简单重复序列 (simple sequence repeats, SSRs)。芍药属 12 个物种的叶绿体基因组在大小、基因组组成和排列顺序、GC 含量等方面高度保守; 同时, 非编码区 (包括基因间区和内含子) 序列的种间变异高于蛋白编码基因序列, LSC 区和 SSC 区序列变异高于 IR 区。系统发育分析结果以 100% 的支持率支持美丽芍药与芍药、草芍药和川赤芍共同构成一个单系分支, 并与芍药亲缘关系最近。本研究首次报道了美丽芍药叶绿体全基因组, 对其序列变异及结构特征进行了解析, 并基于叶绿体全基因组序列构建系统发育树, 明确了美丽芍药在芍药属内的系统发育位置, 研究结果将为美丽芍药的保护遗传学和资源开发利用等研究提供理论基础。

关键词: 芍药属; 美丽芍药; 叶绿体基因组; 高通量测序; 系统发育

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2020)01-0168-09

Complete chloroplast genome of *Paeonia mairai* H. Lév.: characterization and phylogeny

ZHANG Ming-ying, WANG Xi-fang, GAO Jing, LIU A-ping, YAN Yong-gang,
YANG Xin-jie, ZHANG Gang*

(College of Pharmacy and Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal
Medicine, Shaanxi University of Chinese Medicine, Xi'an 712046, China)

Abstract: The whole chloroplast genome of the medicinal plant *Paeonia mairai* H. Lév. was sequenced using the Illumina HiSeq X Ten platform and then assembled, annotated, and characterized by bioinformatic methods in this study. The complete chloroplast genome of *P. mairai* is 152 731 bp in length with the typical quadripartite structure, which consists of a large single copy-region (LSC, 84 402 bp), a small single copy-region (SSC, 16 969 bp), and a pair of inverted repeat regions (IRa and IRb, 25 680 bp), with an overall GC content of 38.4%. A total of 136 predicted genes, including 90 protein-coding genes, 38 tRNA genes and eight rRNA genes were identified.

收稿日期: 2019-08-19; 修回日期: 2019-09-10.

基金项目: 陕西省教育厅科学计划研究项目 (18JK0221); 陕西省自然科学基金基础研究计划 (2019JQ-876); 陕西中医药大学博士科研启动经费 (104080001); 陕西中医药大学“秦药”品质评价及资源开发学科创新团队项目 (2019-QN01); 陕西省高校青年杰出人才支持计划项目; 陕西中医药大学“思邈学者项目”。

*通讯作者 Tel: 86-29-38185165, E-mail: jay_gumling2003@aliyun.com

DOI: 10.16438/j.0513-4870.2019-0654

Among these, seven protein-coding genes, seven tRNA genes and four rRNA genes were found duplicated in the IR regions. In addition, 28 dispersed repeats, 10 tandem repeats, and 64 simple sequence repeats were detected within the whole chloroplast genome of *P. mairei*. Comparative analyses between 12 *Peaonia* species showed that the chloroplast genomes are highly conserved in length, gene content, gene order, and GC content. Meanwhile, the noncoding sequences (intergenic regions and introns) show a higher variation than the protein coding sequences, and sequences from the LSC region and SSC region are more variable than those from the IR regions. *P. mairei* was inferred forming in a distinct clade with *P. lactiflora*, *P. obovate*, and *P. anomala* subsp. *veitchii* with a 100% bootstrap value and is phylogenetically closest to *P. lactiflora*. These results may provide a basis for further genetic studies and the development and utilization of medicinal *P. mairei*.

Key words: *Paeonia*; *Paeonia mairei* H. Lév.; chloroplast genome; high throughput sequencing; phylogeny

叶绿体普遍存在于高等植物、藻类等自养生物细胞中,是光合作用的主要场所,同时执行着合成淀粉、脂肪酸、蛋白质及色素等功能^[1]。作为细胞内半自主性的细胞器,叶绿体拥有自身完整的一套遗传系统^[2],其遗传物质叶绿体基因组在裸子植物中一般为父系遗传,而在被子植物中以母系遗传为主^[3]。典型的被子植物叶绿体基因组为共价闭环状双链DNA分子,具有高度保守的四分体结构,包括一个大单拷贝区 (large single copy, LSC)、一个小单拷贝区 (small single copy, SSC) 和两个反向重复区 (inverted repeats, IRa/b),其大小一般为115~165 kb,编码约110~130个基因^[4]。

1986年,烟草 *Nicotiana tabacum* L.^[5] 和地钱 *Marchantia polymorpha* L.^[6] 叶绿体全基因组的成功测序和解析标志着叶绿体基因组研究的开始。新一代测序技术和生物信息学分析方法的快速发展,加之测序成本的不断降低,极大地推动了叶绿体基因组研究的发展,而“本草基因组学”^[7]的提出进一步使得药用植物叶绿体基因组研究成为热点。由于叶绿体基因组序列组成和结构相对保守,碱基变异速率适中,基因组相对较小,易于测序,且在细胞中常存在多个拷贝^[4],同时具有单亲遗传等特征,在植物分类与物种水平的系统发育等研究中具有明显优势,叶绿体系统发育基因组学分析已成为解决药用植物物种鉴定问题的有效途径^[8-10]。

美丽芍药 *Paeonia mairei* H. Lév. 为芍药科 *Paoniaceae* 芍药属 *Paeonia* 多年生草本植物,为中国特有物种,主要分布于云南东北部、贵州西部、四川中南部、湖北西南部、陕西南部 and 甘肃南部等地^[11]。其根可入药,具有行瘀活血、止痛之功效,在我国四川等一些地区,美丽芍药的根亦被作为中药材赤芍 (*Paoniae Radix Rubra*) 的来源之一^[12]。目前,芍药属芍药 *P. lactiflora* Pall.^[13]、川赤芍 *P. anomala* subsp. *Veitchii* (Lynch) D. Y. Hong & K. Y. Pan^[14]、凤丹 *P. ostii* T. Hong & J. X. Zhang^[15]、紫斑牡丹 *P. rockii* (S. G. Haw & Lauener) T. Hong & J.

J. Li ex D. Y. Hong^[16]、四川牡丹 *P. decomposita* Hand.-Mazz.^[17]、矮牡丹 *P. jishanensis* T. Hong & W. Z. Zhao^[18]、以及大花黄牡丹 *P. ludlowii* (Stern & G. Taylor) D. Y. Hong 和滇牡丹 *P. delavayi* Franch.^[19] 等多种植物的叶绿体基因组已经得到解析,并基于这些叶绿体基因组序列开展了不同阶元的系统发育研究。然而,美丽芍药的叶绿体基因组相关研究还未见报道, Stern^[20] 和 Hong^[21] 的分类系统都曾将芍药属下的物种划分为3个组,并将美丽芍药归于其中的芍药组 Sect. *Paeonia*, 但美丽芍药在芍药组内的系统发育位置还尚未解决。

本研究基于二代高通量测序技术和生物信息学分析方法,对美丽芍药开展叶绿体全基因组测序、组装、注释,及序列变异和结构特征解析,并将其与芍药属已报道的其他11种植物的叶绿体基因组序列进行比较基因组学和系统发育分析。旨在为美丽芍药的保护遗传学和资源开发利用、以及芍药属植物种间系统发育关系、物种鉴定等研究奠定基础。

材料与方法

实验材料 美丽芍药新鲜叶片材料采自陕西省宝鸡市太白县 (34°5'35"N, 107°26'35"E), 经陕西中医药大学标本馆王继涛高级实验师鉴定,凭证标本保存于陕西中医药大学药学院中药资源学与鉴定学科实验中心。

基因组DNA提取与测序 取新鲜幼嫩叶片,利用植物基因组DNA提取试剂盒 (TIANGEN, 北京) 提取总DNA,琼脂糖凝胶电泳和 Qubit 3.0 (Thermo Fisher Scientific, USA) 荧光定量仪检测DNA质量及浓度。检测合格后的总DNA被随机打断,构建长度约350 bp的插入片段文库,利用 Illumina HiSeq X Ten 平台进行双端测序,测序读长为150 bp。测序工作由华大基因完成。

叶绿体基因组序列组装及注释 测序完成后得到原始序列 (raw reads) 首先利用 NGS QC ToolKit^[22] 过滤去除接头及两端的低质量序列 (参数设置为 cut-off

value for percentage of read length = 80, cut-off value for PHRED quality score = 30), 得到高质量待分析序列 (即 clean reads)。在 Linux 系统下利用 GetOrganelle 程序包^[23]对 clean reads 进行组装, 序列延伸的种子 (seed) 为 GetOrganelle 默认 seed, 获得序列重叠群 (contigs), 再通过 Bandage 软件^[24]进行可视化编辑, 将 contigs 拼接成环并去除冗余序列, 初步完成叶绿体全基因组序列的从头组装 (*De novo assembly*)。之后, 再利用 Bowtie2 软件^[25]将原始序列重新映射 (mapping) 到组装完成的叶绿体基因组序列上, 检查验证 LSC/IRa、IRa/SSC、SSC/IRb 和 IR/LSC 四个节点及各 contigs 拼接处的正确性, 并修补有缺口 (gaps) 的位点, 最终获得美丽芍药完整的叶绿体全基因组序列。

以 NCBI 数据库中收录的同属植物芍药的叶绿体全基因组注释结果 (GenBank 登录号: NC_040983) 作为参考, 利用 PGA (Plastid Genome Annotator) 软件^[26]对组装得到的叶绿体全基因组序列进行注释, 在 Geneious 软件^[27]中对注释结果进行手动校正, 调整蛋白编码基因的起始、终止位置和基因内含子、外显子边界, 并利用 tRNAscan-SE 软件^[28]确定 tRNA 基因的边界。最后, 利用 Organellar Genome DRAW (OGDRAW) 软件^[29]绘制叶绿体全基因组图谱。组装并注释完成后的叶绿体全基因组序列提交至 NCBI 数据库, 获得登录号 MN061945。

叶绿体全基因组重复序列检测 叶绿体全基因组中的散在重复序列 (dispersed repeats) 利用 REpute 软件 (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) 进行检测, 参数设置为最小重复序列长度 = 30 bp, 重复序列间的相似性 >90%^[30]。串联重复序列 (tandem repeats) 利用 Tandem repeats finder 软件 (<https://tandem.bu.edu/trf/trf.html>) 进行检测, 参数设置选择默认值^[31]。简单重复序列 (simple sequence repeats, SSRs) 检测通过 MISA 软件^[32]完成, 参数设置为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, 三核苷酸和四核苷酸重复单元不少于 4 个, 五核苷酸和六核苷酸重复单元不少于 3 个, 且两个 SSRs 之间的最小距离为 100 bp。

芍药属种间叶绿体基因组比较分析 目前, NCBI 数据库共收录芍药属 11 种植物的叶绿体全基因组序列数据, 即草芍药 *P. obovata* Maxim. (GenBank 登录号: NC_026076)、芍药 (NC_040983)、川赤芍 (NC_032401)、紫斑牡丹 (MF488719)、大花黄牡丹 (NC_035623)、滇牡丹 (NC_035718)、凤丹 (NC_036834)、矮牡丹 (MG991935)、四川牡丹 (NC_039425)、北美芍药 *P. brownii* Douglas ex Hook. (MH191385) 和牡丹 *P. suffruticosa* Andr. (MH191384)。利用 mVISTA 软件^[33]基

于 LAGAN 模型对美丽芍药与这 11 种植物进行叶绿体全基因组比较分析。此外, 统计比较包括美丽芍药在内的所有 12 个物种的叶绿体全基因组长度、各个区域长度及其 GC 含量, 并利用 CORREL 函数分别计算全基因组长度变化与 LSC 区、SSC 区和 IR 区长度变化之间的相关性。

系统发育分析 下载 NCBI 数据库中收录的芍药属上述所有 11 种植物的叶绿体全基因组序列, 与本研究获得的美丽芍药叶绿体全基因组序列共同构建序列矩阵, 同时选择芍药属的近缘类群虎耳草目 Saxifragales 虎耳草科 Saxifragaceae 秦岭岩白菜 *Bergenia scopulosa* T. P. Wang 和毛茛目 Ranunculales 毛茛科 Ranunculaceae 黄连 *Coptis chinensis* Franch. 作为外类群, 利用 MAFFT 软件^[34]完成序列比对, 利用 RAxML^[35]以最大似然法 (Maximum Likelihood, ML) 进行系统发育关系重建分析, 核苷酸替换模型选择 GTR+G, 其他参数选择默认设置, 进行 1 000 次自展重复分析, 计算系统发育树各节点的靴带支持率 (Bootstrap values, BS)。

结果

1 叶绿体全基因组测序、组装及注释

测序共获得 80 175 282 条读长为 150 bp 的 raw reads, 质控除去接头和低质量序列之后得到 79 009 120 条 clean reads, 即 clean reads 占 raw reads 的比例为 98.55%。将 clean reads 组装、拼接后获得一条完整的环状叶绿体基因组序列。美丽芍药叶绿体全基因组序列总长度为 152 731 bp, 包括一个长度为 84 402 bp 的大单拷贝区、一个 16 969 bp 的小单拷贝区和一对长度分别为 25 680 bp 的反向重复区 (图 1)。基因组整体 GC 含量为 38.4%, SSC、LSC 和 IR 区的 GC 含量依次为 32.7%、36.7% 和 43.1%。

共注释得到 136 个基因, 包括 90 个蛋白编码基因 (*ycf68* 和 *rpl22* 为假基因)、38 个 tRNA 基因和 8 个 rRNA 基因。其中, 7 个蛋白编码基因、7 个 tRNA 基因和 4 个 rRNA 基因分别在反向重复区发生了一次重复 (表 1)。在这些基因中, 有 6 个 tRNA 基因 (*trnI-GAU*、*trnA-UGC*、*trnV-UAC*、*trnL-UAA*、*trnG-UCC*、*trnK-UUU*) 和 12 个蛋白编码基因 (*rpl2*、*rpl16*、*rpoC1*、*rps16*、*atpF*、*ndhB*、*ndhA*、*petD*、*petB*、*ycf3*、*clpP* 和 *rps12*) 包含内含子, 其中, 除 *ycf3*、*clpP* 和 *rps12* 各包含 2 个内含子外, 其余均只包含 1 个内含子 (表 2)。此外, 蛋白编码基因 *matK* 位于 *trnK-UUU* 基因的内含子区域之内, *rps12* 基因具反式剪切 (trans-spliced) 结构, 其 5' 端和 3' 端分别位于 LSC 区和 IR 区, 被分成 2 个独立的转录单元。

美丽芍药叶绿体全基因组的蛋白编码基因序列总

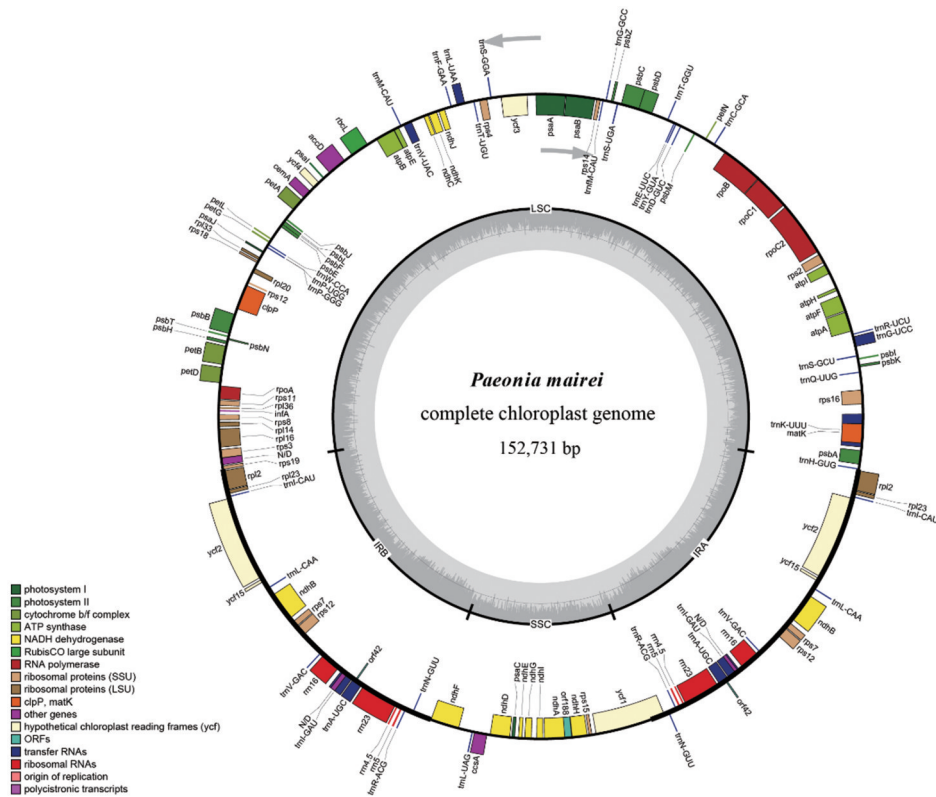


Figure 1 Chloroplast genome map of *P. mairii*. Genes inside and outside the circle are transcribed in a clockwise and counter clockwise direction, respectively. Genes are color-coded based on their functions. The grey area in the inner circle indicates the GC content of the chloroplast genome.

Table 1 Gene composition in the chloroplast genome of *P. mairii*. Bold font indicates genes duplicated in the IR regions; ^a represents pseudogenes; ^b and ^c indicate genes with one and two introns, respectively; ^d indicates the gene has 2 separate transcription units; ^e indicates the gene spans the large single copy region and the IR region

Category	Gene group	Gene name
Self-replication	Large subunit of ribosome	<i>rpl2^b</i> , <i>rpl23</i> , <i>rpl22^a</i> , <i>rpl16^b</i> , <i>rpl14</i> , <i>rpl36</i> , <i>rpl20</i> , <i>rpl33</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1^b</i> , <i>rpoC2</i>
	Small subunit of ribosome	<i>rps7</i> , <i>rps12^{c,d}</i> , <i>rps15</i> , <i>rps19^e</i> , <i>rps3</i> , <i>rps8</i> , <i>rps11</i> , <i>rps18</i> , <i>rps4</i> , <i>rps14</i> , <i>rps2</i> , <i>rps16^b</i>
	Ribosome RNA Genes	<i>rrn16</i> , <i>rrn23</i> , <i>rrn4.5</i> , <i>rrn5</i>
	Transfer RNA Genes	<i>trnI-CAU</i> , <i>trnL-CAA</i> , <i>trnV-GAC</i> , <i>trnI-GAU^b</i> , <i>trnA-UGC^c</i> , <i>trnR-ACG</i> , <i>trnN-GUU</i> , <i>trnL-UAG</i> , <i>trnP-GGG</i> , <i>trnP-UGG</i> , <i>trnW-CCA</i> , <i>trnM-CAU</i> , <i>trnV-UAC^b</i> , <i>trnF-GAA</i> , <i>trnL-UAA^b</i> , <i>trnT-UGU</i> , <i>trnS-GGA</i> , <i>trnJ-M-CAU</i> , <i>trnG-GCC</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnE-UUC</i> , <i>trnY-GUA</i> , <i>trnD-GUC</i> , <i>trnC-GCA</i> , <i>trnR-UCU</i> , <i>trnG-UCC^b</i> , <i>trnS-GCU</i> , <i>trnQ-UUG</i> , <i>trnK-UUU^b</i> , <i>trnH-GUG</i>
Photosynthesis	Subunits of ATP synthase	<i>atpB</i> , <i>atpE</i> , <i>atpI</i> , <i>atpH</i> , <i>atpF^b</i> , <i>atpA</i>
	Subunits of NADH-dehydrogenase	<i>ndhB^b</i> , <i>ndhH</i> , <i>ndhA^b</i> , <i>ndhI</i> , <i>ndhG</i> , <i>ndhE</i> , <i>ndhD</i> , <i>ndhF</i> , <i>ndhC</i> , <i>ndhK</i> , <i>ndhJ</i>
	Subunits of cytochrome b/f complex	<i>petD^b</i> , <i>petB^b</i> , <i>petG</i> , <i>petL</i> , <i>petA</i> , <i>petN</i>
	Subunits of photosystem I	<i>psaC</i> , <i>psaJ</i> , <i>psaI</i> , <i>psaA</i> , <i>psaB</i>
	Subunits of photosystem II	<i>psbH</i> , <i>psbN</i> , <i>psbT</i> , <i>psbB</i> , <i>psbE</i> , <i>psbF</i> , <i>psbL</i> , <i>psbJ</i> , <i>psbZ</i> , <i>psbC</i> , <i>psbD</i> , <i>psbM</i> , <i>psbI</i> , <i>psbK</i> , <i>psbA</i>
Other genes	Subunit of rubisco	<i>rbcL</i>
	proteins photosystem I assembly	<i>ycf4</i> , <i>ycf3^c</i>
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	c-Type cytochrome synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP^c</i>
	Translational initiation factor	<i>infA</i>
Maturase	<i>matK</i>	
Proteins of unknown function	<i>ycf2</i> , <i>ycf15</i> , <i>ycf68^a</i> , <i>ycf1^e</i>	

Table 2 List of genes with introns identified in the chloroplast genome of *P. mairei*

Gene	Location	Length/bp				
		Exon I	Intron I	Exon II	Intron II	Exon III
<i>rps16</i>	LSC	236	815	40		
<i>atpF</i>	LSC	411	691	159		
<i>rpoC1</i>	LSC	1 617	700	435		
<i>ycf3</i>	LSC	153	749	228	735	126
<i>rps12</i>	LSC-IR	114	–	232	535	26
<i>clpP</i>	LSC	228	658	292	675	71
<i>petB</i>	LSC	6	758	657		
<i>petD</i>	LSC	9	645	525		
<i>rpl16</i>	LSC	9	1 005	399		
<i>rpl2</i> (×2)	IR	435	668	393		
<i>ndhB</i> (×2)	IR	756	682	777		
<i>ndhA</i>	SSC	540	1 015	543		
<i>trnK-UUU</i>	LSC	35	2 452	37		
<i>trnG-UCC</i>	LSC	23	702	48		
<i>trnL-UAA</i>	LSC	37	524	50		
<i>trnV-UAC</i>	LSC	37	573	39		
<i>trnI-GAU</i> (×2)	IR	37	938	35		
<i>trnA-UGC</i> (×2)	IR	38	717	35		

长度为 78 273 bp, 约占叶绿体基因组总长度的 51.2%, 由 26 091 个密码子组成。其中, 编码亮氨酸 Leu 的密码子使用频率最高, 共 2 735 次, 约占总密码子使用率的 10.5%, 而编码半胱氨酸的密码子使用率最低, 仅 296 次, 占总密码子使用率的 1.1%。

2 重复序列检测

共检测到 28 个同时满足长度不小于 30 bp 且重复序列间相似度大于 90% 两个条件的散在重复序列, 包括 14 个回文重复 (palindromic repeats, P)、12 个正向重复 (forward repeats, F)、1 个互补重复 (complement repeats, C) 和 1 个反向重复 (reverse repeats, R), 其长度范围为 30~64 bp。检测到 10 条长度范围为 18~33 bp 的串联重复序列。串联重复序列在 4 个区域均有分布, 其中, LSC 区和 SSC 区各 2 条, 重复次数均为 2 次, 两个 IR 区各 3 条, 重复次数分别为 2、3 和 4 次。共检测到 64 个 SSRs, 包括 44 个单核苷酸、13 个双核苷酸和

7 个三核苷酸 SSRs, 四核苷酸、五核苷酸和六核苷酸 SSR 均未检测到。其中, 有 11 个 SSRs 分别分布于蛋白编码基因 *matK*、*rpoC2*、*atpB*、*cemA*、*rps3* 和 *ycf1* 的编码区内, 其余均分布于非编码区 (包括基因间区和内含子区)。

3 芍药属种间叶绿体基因组比较分析

美丽芍药与芍药属其他 11 种植物叶绿体全基因组序列间的比较分析结果见图 2。整体上, 12 个物种的叶绿体基因组基因构成和顺序高度一致, 同时, LSC 区和 SSC 区序列相比于 IR 区存在较多变异, 并且变异主要出现在非编码序列中。此外, 美丽芍药与其他 11 种植物叶绿体基因组序列在 *rps16-trnQ*、*psaA-ycf3*、*trnL-trnF*、*petA-psbJ*、*ndhF-trnL*、*rps15-ycf1* 等基因间区, 以及 *rpoC1*、*rpl16*、*ycf1* 基因的内含子区域存在明显差异。12 个物种叶绿体全基因组、LSC、SSC 和 IR 区的长度变化范围分别为 152 153~154 405 bp、84 365~86 057 bp、16 679~17 059 bp 和 24 863~25 745 bp, GC 含量变化范围分别为 38.3%~38.5%、36.6%~36.8%、32.6%~33.0% 和 43.1%~43.2% (表 3)。此外, CORREL 函数计算得到种间叶绿体全基因组长度变化与 LSC、SSC 和 IR 区长度变化之间的相关性系数分别为 0.657 4、0.360 7 和 0.371 8。

4 系统发育分析

系统发育分析结果显示 (图 3), 包括美丽芍药在内的芍药属 12 个物种以 100% 的支持率构成一个独立的单系分支, 与毛茛科、虎耳草科区分开来。同时, 这 12 个物种可以进一步划分为 3 个次级单系分支, 其中, 同来自芍药组的草芍药、川赤芍、芍药和美丽芍药 4 个物种构成一个单系的支持率为 100%, 并且美丽芍药和芍药的亲缘关系最近, 构成姐妹类群。来自牡丹组 sect. *Moutan* 的 7 个物种亦以 100% 的支持率聚为一支, 并与北美芍药构成姐妹关系。在整个系统发育树上, 除芍药组-北美芍药组 sect. *Onaepia*、凤丹-牡丹两

Table 3 Comparison of chloroplast sequences of *P. mairei* and other 11 species from *Paeonia*

Species	Whole genome		LSC		SSC		IR	
	Length/bp	G+C/%	Length/bp	G+C/%	Length/bp	G+C/%	Length/bp	G+C/%
<i>P. mairei</i>	152 731	38.4	84 402	36.7	16 969	32.7	25 680	43.1
<i>P. obovata</i>	152 698	38.4	84 387	36.7	17 027	32.7	25 642	43.2
<i>P. lactiflora</i>	152 747	38.4	84 412	36.7	17 033	32.7	25 651	43.1
<i>P. anomala</i> subsp. <i>veitchii</i>	152 682	38.4	84 398	36.7	16 978	32.7	25 653	43.1
<i>P. rockii</i>	152 821	38.3	84 477	36.6	17 048	32.6	25 648	43.1
<i>P. ludlowii</i>	152 687	38.4	84 426	36.7	16 983	32.7	25 639	43.1
<i>P. delavayi</i>	154 405	38.4	86 057	36.7	17 050	32.7	25 649	43.1
<i>P. ostii</i>	152 153	38.4	85 373	36.7	17 054	32.7	24 863	43.1
<i>P. jishanensis</i>	152 628	38.4	84 292	36.7	17 044	32.7	25 646	43.1
<i>P. decomposita</i>	152 601	38.4	84 265	36.7	17 036	32.6	25 650	43.1
<i>P. suffruticosa</i>	153 119	38.4	84 570	36.6	17 059	32.6	25 745	43.2
<i>P. brownii</i>	152 228	38.5	84 265	36.8	16 679	33.0	25 642	43.2

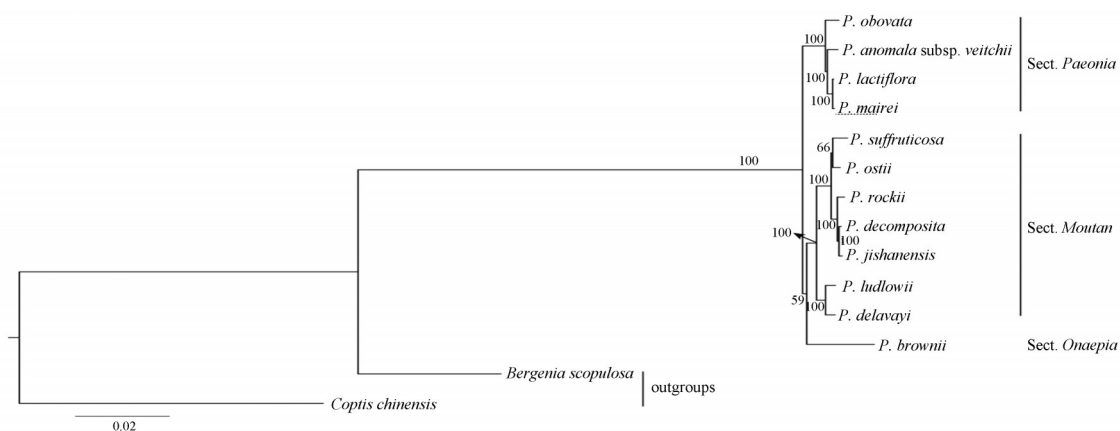


Figure 3 Maximum likelihood phylogenetic tree of 12 species from *Paeonia* inferred using a dataset of the whole chloroplast genome sequences. Numbers beside each node are bootstrap values (BS)

个节点的支持率分别为59%和66%之外,其余节点的支持率均为100%。

讨论

作为芍药属中国特有物种之一,美丽芍药曾被《中国物种红色名录》列为近危种^[36],化学成分研究表明,美丽芍药根部含芍药苷等多种芍药属植物所具有的特征化合物,并且具有较强的抗氧化活性,可以作为抗菌抗氧化药物新资源^[37],因而具有重要的保护、研究和开发利用价值。本研究基于二代高通量测序技术和生物信息学分析方法首次对美丽芍药叶绿体全基因组进行了分析研究。结果显示美丽芍药叶绿体基因组具有被子植物叶绿体基因组典型的四分体结构,并且其大小、结构、基因组成和顺序、CG含量等与芍药属其他已报道的植物高度相似^[13,15-19]。此外,分析发现,包括美丽芍药在内的芍药属12个物种叶绿体全基因组、LSC、SSC和IR区的长度差异范围分别为5~2 252 bp、0~1 792 bp、2~380 bp和0~882 bp,说明芍药属植物叶绿体基因组在进化过程中高度保守。种间叶绿体全基因组长度变化与LSC、SSC和IR区长度变化之间的相关性系数为LSC (0.657 4) > IR (0.371 8) > SSC (0.360 7),说明不同物种间叶绿体全基因组长度的变化主要受LSC区长度变化影响,而IR区和SSC区长度变化相对保守。

本文所研究的芍药属12种植物中,叶绿体基因组序列变异模式在不同区域表现出明显的差异(图2)。总体上,LSC区和SSC区序列变异高于IR区,IR区序列高度保守,这一结果与Guo等^[15]对芍药属5种植物叶绿体全基因组的比较分析结果相同,也与其他被子植物类群叶绿体基因组序列变异模式一致^[9,38-40]。同时,LSC区和SSC区所呈现出的高变异主要来自于非

编码序列(主要为基因间区序列,其次为部分基因的内含子序列),说明在芍药属中,非编码序列的进化速率快于蛋白编码基因序列,与多数被子植物叶绿体基因组序列进化式样一致^[38,40,41]。此外,美丽芍药与芍药属其他11种植物叶绿体基因组序列在*rps16-trnQ*、*psaA-ycf3*、*trnL-trnF*、*petA-psbJ*、*ndhF-trnL*、*rps15-ycf1*等基因间区,以及*rpoCl*、*rpl16*、*ycf1*基因的内含子区均存在明显差异,这些区域可以为芍药属物种鉴定特异性DNA条形码的筛选提供参考。

叶绿体基因组中的SSRs因含量丰富、多态性高,且兼具单亲遗传模式等优点而被作为分子标记广泛用于群体遗传学^[42]、谱系地理学^[43]等研究。本研究从美丽芍药叶绿体全基因组中共检测出44个单核苷酸、13个双核苷酸和7个三核苷酸SSRs。这些SSRs可以为美丽芍药的遗传多样性及保护遗传学等相关研究提供候选分子标记。在这些SSRs中,单核苷酸和双核苷酸的重复序列类型分别以A/T(约占单核苷酸SSRs总数的95.45%)和AT/AT(约占双核苷酸SSRs总数的92.30%)为主,三核苷酸SSRs的重复单元全部由A、T碱基组合(AAT、ATA、TAT、TTA、TAA)构成,这也与其他被子植物叶绿体基因组中的SSRs序列构成一致^[39,40,44],进一步支持了叶绿体SSRs主要是由短的polyA和polyT重复所构成,而较少含有C或G串联重复的观点^[45]。此外,这些SSRs主要分布在基因组中的LSC区(50个),其次为SSC区(10个),两个IR区分别仅检测到2个重复次数分别为10和12的A、T单核苷酸SSRs,因此,推测这些高A/T含量的SSRs可能是引起叶绿体基因组中CG含量偏低及各区域碱基含量偏差的原因之一(本研究中美丽芍药叶绿体基因组中平均GC含量为38.4%,LSC、SSC、和IR区的GC含量分别为36.7%、32.7%和43.1%)。

芍药属是一个分类与系统发育研究中较为复杂的类群^[46]。依据形态特征和分子系统学等研究结果, 属下物种被划分为3个组, 即牡丹组、芍药组和北美芍药组^[20,21], 但由于经历了复杂的网状进化及多倍化演化历程, 有限的DNA片段 (ITS、*matK*、*psbA-trnH*及*trnL-trnF*等) 未能提供足够的系统发育信息来有效解决其种间关系^[46-48]。本研究基于叶绿体全基因组序列构建的系统发育树将芍药属12个物种明确划分为3个独立的单系分支, 与Stern^[20]和Hong等^[21]划分的3个组分别对应, 并且各分支内部不同物种间的系统发育关系也得到较好解决 (除凤丹-牡丹之间姐妹关系的支持率为66%外, 其余所有种间关系的支持率均为100%)。说明叶绿体全基因组序列包含丰富的物种演化信息, 是解决芍药属种间系统发育关系的一个有效途径。

References

- [1] Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids [J]. *Annu Rev Plant Physiol Plant Mol Biol*, 2000, 51: 111-140.
- [2] McFadden GI. Primary and secondary endosymbiosis and the origin of plastids [J]. *J Phycol*, 2001, 37: 951-959.
- [3] Corriveau JL, Coleman AW. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species [J]. *Am J Bot*, 1988, 75: 1443-1458.
- [4] Jansen RK, Ruhlman TA. *Plastid Genomes of Seed Plants*, in *Genomics of Chloroplasts and Mitochondria* [M]. Dordrecht: Springer Netherlands, 2012.
- [5] Sugiura M, Shinozaki K, Zaita N, et al. Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genes [J]. *Plant Sci*, 1986, 44: 211-217.
- [6] Ohyama K, Fukuzawa H, Kohchi T, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA [J]. *Nature*, 1986, 322: 572-574.
- [7] Chen SL, Song JY. *Herbgenomics* [J]. *China J Chin Mater Med* (中国中药杂志), 2016, 41: 3881-3889.
- [8] Zhang N, Erickson DL, Ramachandran P, et al. An analysis of *Echinacea* chloroplast genomes: implications for future botanical identification [J]. *Sci Rep*, 2017, 7: 216.
- [9] Zhu S, Niu Z, Xue Q, et al. Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes [J]. *Acta Pharm Sin B*, 2018, 8: 969-980.
- [10] Fu CN, Wu CS, Ye LJ, et al. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide [J]. *Sci Rep*, 2019, 9: 2773.
- [11] Hong DY, Pan KY, Turland NJ. *Paeoniaceae in Flora of China* [M]. Beijing: Science Press/St. Louis: Missouri Botanical Garden Press, 2001.
- [12] Editorial Committee of Chinese Materia Medica. *Chinese Materia Medica* (中华本草) [M]. Shanghai: Shanghai Science and Technology Press, 1999.
- [13] Lee M, Park JH, Gil JS, et al. The complete chloroplast genome of *Paeonia lactiflora* Pall. (Paeoniaceae) [J]. *Mitochondr DNA Part B Resour*, 2019, 4: 2715-2716.
- [14] Zhang G, Sun J, Li YM, et al. The complete chloroplast genome of *Paeonia anomala* subsp. *Veitchii* [J]. *Mitochondr DNA Part B Resour*, 2016, 1: 191-192.
- [15] Guo S, Guo L, Zhao W, et al. Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii* [J]. *Molecules*, 2018, 23: 246.
- [16] Bai GQ, Guo HW, Zhao N, et al. The complete chloroplast genome of *Paeonia rockii* (Paeoniaceae), an endangered endemic species to China [J]. *Conserv Genet Resour*, 2018, 10: 453-456.
- [17] Chen YM, Zhou QJ, Sun LJ, et al. The chloroplast genome of *Paeonia decomposita* (Paeoniaceae), an endangered wild tree peony from Sichuan, China [J]. *Conserv Genet Resour*, 2019, 11: 59-61.
- [18] Zhou XJ, Song LL, Peng ZF, et al. The complete chloroplast genome sequence of *Paeonia jishanensis* (Paeoniaceae), a rare wild tree peony [J]. *Mitochondr DNA Part B Resour*, 2019, 4: 503-504.
- [19] Li HE, Guo QQ, Zheng WL. Characterization of the complete chloroplast genomes of two sister species of *Paeonia*: genome structure and evolution [J]. *Conserv Genet Resour*, 2018, 10: 209-212.
- [20] Stern FC. *A Study of the Genus Paeonia* [M]. London: The Royal Horticulture Society, 1946.
- [21] Hong DY. *Peonies of the World: Taxonomy and Phytogeography* [M]. Richmond: Royal Botanic Gardens, 2010.
- [22] Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data [J]. *PLoS One*, 2012, 7: e30619.
- [23] Jin JJ, Yu WB, Yang JB, et al. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data [J]. *BioRxiv*, 2018: 256479.
- [24] Wick RR, Schultz MB, Zobel J, et al. Bandage: interactive visualization of de novo genome assemblies [J]. *Bioinformatics*, 2015, 31: 3350-3352.
- [25] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2 [J]. *Nat Methods*, 2012, 9: 357-359.
- [26] Qu XJ, Moore MJ, Li DZ, et al. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes [J]. *Plant Methods*, 2019, 15: 50.
- [27] Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28: 1647-1649.
- [28] Lowe TM, Chan PP. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes [J]. *Nucleic*

- Acids Res, 2016, 44: W54-W 57 .
- [29] Greiner S, Lehwark P, Boc R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes [J]. Nucleic Acids Res, 2019, 47: W59-W64.
- [30] Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Res, 2001, 29: 4633-4642.
- [31] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. Nucleic Acids Res, 1999, 27: 573-580.
- [32] Beier S, Thiel T, Munch T, et al. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 2017, 33: 2583-2585.
- [33] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics [J]. Nucleic Acids Res, 2004, 32 (suppl_2): W273-W279 .
- [34] Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program [J]. Bioinformatics, 2010, 26: 1899-1900.
- [35] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies [J]. Bioinformatics, 2014, 30: 1312-1313.
- [36] Wang S, Xie Y. China Species Red List (中国物种红色名录) [M]. Beijing: Higher Education Press, 2004.
- [37] Shi Y. Study on Chemical Compositions of *Paeonia mairei* Levl. and Their Bioactivities (美丽芍药化学成分及其生物活性分析) [D]. Xi'an: Shaanxi University of Science and Technology, 2015.
- [38] Shen XF, Wu ML, Liao BS, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua* [J]. Molecules, 2017, 22: E1330.
- [39] Guo HJ, Liu JS, Luo L, et al. Complete chloroplast genome sequences of *Schisandra chinensis*: genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms [J]. Sci China C (中国科学: 生命科学), 2017, 47: 728-739.
- [40] Yang QQ, Jiang M, Wang LQ, et al. Complete chloroplast genome of *Allium chinense*: comparative genomic and phylogenetic analysis [J]. Acta Pharm Sin (药学报), 2019, 54: 173-181.
- [41] Zhang MY, Fritsch PW, Ma PF, et al. Plastid phylogenomics and adaptive evolution of *Gaultheria* series *Trichophyllae* (Ericaceae), a clade from sky islands of the Himalaya-Hengduan Mountains [J]. Mol Phylogenet Evol, 2017, 110: 7-18.
- [42] Du Q, Wang B, Wei Z, et al. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers [J]. J Hered, 2012, 103: 853-862.
- [43] Chmielewski M, Meyza K, Chybicki I, et al. Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland [J]. iForest, 2015, 8: 765-771.
- [44] Tian N, Han LM, Chen C, et al. The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species [J]. PLoS One, 2018, 13: e0192956.
- [45] Kuang DY, Wu H, Wang YL, et al. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics [J]. Genome, 2011, 54: 663-673.
- [46] Sun YL, Hong SK. Phylogenetic relationship and evolution analysis of the peony *Paeonia* species using multi-locus deoxyribonucleic acid (DNA) barcodes [J]. J Med Plants Res, 2012, 6: 5048-5058.
- [47] Sang T, Crawford DJ, Stuessy TF. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution [J]. Proc Natl Acad Sci U S A, 1995, 92: 6813-6817.
- [48] Sang T, Crawford D, Stuessy T. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae) [J]. Am J Bot, 1997, 84: 1120-1136.