

• 药材与资源 •

毛重楼叶绿体基因组序列特征及其系统发育分析

江媛^{1,2}, 杨青淑^{1,2}, 王婧^{1,2}, 杨成金³, 黄林芳², 杨燕⁴, 段宝忠^{1*}

1. 大理大学药学院, 云南 大理 671000
2. 中国医学科学院药用植物研究所, 北京 100193
3. 云南白药集团中药资源有限公司, 云南 昆明 650504
4. 云南省科学技术院, 云南 昆明 650051

摘要: **目的** 获取毛重楼 *Paris mairei* 叶绿体基因组信息特征并对其系统位置进行研究。**方法** 采用 Illumination 测序技术对毛重楼叶绿体基因组进行测序, 对其进行组装、注释和特征分析, 并构建最大似然 (maximum likelihood, ML) 系统发育树。**结果** 毛重楼叶绿体基因组大小为 163 918 bp, 总 GC 含量为 37.05%, 大单拷贝区 (large single-copy, LSC)、小单拷贝区 (small single-copy region, SSC) 分别为 84 173、13 054 bp, 反向互补重复区 (inverted repeats, IR) 大小为 33 296 bp, 共编码 113 个基因, 包括 79 个蛋白质编码基因, 30 个 tRNA 基因和 4 个 rRNA 基因。在系统进化树中, 重楼属 *Paris* L. 与藜芦科 (Melanthiaceae Batsch ex Borkh.) 关系较近, 与百合科 (Liliaceae Juss.) 关系较远。**结论** LSC 区和 SSC 区序列变异高于 IR 区; 相较于重楼属, 毛重楼与蚤休组的亲缘关系最近, 应将毛重楼归属为蚤休组; 相较于百合科, 毛重楼与藜芦科藜芦属亲缘关系更近, 应将蚤休组从百合科中独立, 归属为藜芦科。

关键词: 毛重楼; 叶绿体基因组; 重楼属; 系统发育; 藜芦科

中图分类号: R282 文献标志码: A 文章编号: 0253-2670(2021)13-4014-09

DOI: 10.7501/j.issn.0253-2670.2021.13.024

Complete chloroplast genome of *Paris mairei*: characterization and phylogeny

JIANG Yuan^{1, 2}, YANG Qing-shu^{1, 2}, WANG Jing^{1, 2}, YANG Cheng-jin³, HUANG Lin-fang², YANG Yan³, DUAN Bao-zhong¹

1. College of Pharmaceutical Science, Dali University, Dali 671000, China
2. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China
3. Yunnan Baiyao Group Traditional Chinese Medicine Resources Co., Ltd., Kunming 650504, China
4. Yunnan Provincial Academy of Science and Technology, Kunming 650051, China

Abstract: Objective To obtain the complete chloroplast genome information and to estimate phylogenetic relationships among *Paris mairei*. **Methods** The Illumina platform was used to sequence the complete chloroplast genome of *P. mairei*. The bioinformatics software were used to assemble, annotate and characterize the complete chloroplast genome of *P. mairei*. RAxML was used to reconstruct a maximum-likelihood (ML) phylogenetic tree. **Results** The complete chloroplast genome of *P. mairei* was 163 819 bp, containing inverted repeated of 33 296 bp each, a large single-copy region of 84 173 bp, and a small single-copy region of 13 054 bp. A total of 113 genes were identified including 79 protein-coding, 30 tRNAs, and 4 rRNAs. Phylogenetic analysis indicated that *Paris* presented a closer relationship with Melanthiaceae than Liliaceae. **Conclusion** The IR region variability was significantly inferior to LSC and SSC. Besides, phylogenetic analysis revealed that *P. mairei* was closely related to *Veratrum*.

Key words: *Paris mairei* H. Lévl.; chloroplast genome; *Paris* L.; phylogeny; Melanthiaceae

收稿日期: 2020-12-05

基金项目: 国家自然科学基金项目 (31860080); 云南省重大科技专项 (202002AA100007); 大理大学创新团队项目 (ZKLX2019318); 云南省院士专家工作站 (宋经元, 2021-1)

作者简介: 江媛 (1997—), 硕士研究生, 研究方向为中药资源与鉴定。E-mail: yjiang@yntcm.ac.cn

*通信作者: 段宝忠, 教授, 研究方向为中药资源与鉴定。Tel: (0872)2257411 E-mail: bzduan@126.com

重楼属 *Paris* L. 植物全球共有约 30 余种, 主要分布于我国西南地区, 有药用记载的有 11 种^[1]。其中滇重楼 *P. polyphylla* Smith var. *yunnanensis* (Franch.) Hand. -Mazz. 和华重楼 *P. polyphylla* Smith var. *chinensis* (Franch.) Hara, 是《中国药典》2020 年版收载的品种^[2]。由于长期的掠夺式采挖, 重楼属植物资源遭到了毁灭性的破坏, 多个物种已濒临枯竭^[3-4], 加强该属植物种质资源的保护和相关基础研究迫在眉睫。毛重楼 *P. mairei* H. Léveillé 为重楼属植物, 主要分布在云南、四川和贵州, 以根茎入药, 具有清热解毒、平喘止咳、消肿散瘀, 常用于治疗疔疮痈肿、咽喉肿痛、咳嗽、蛇虫咬伤、跌扑伤痛等病症^[5]。现代研究表明, 毛重楼含有螺甾烷型、呋甾烷型、五环三萜型等多种皂苷类化合物^[6-8], 具有抗肿瘤、抗病毒等重要药用价值^[9]。目前, 有关毛重楼的研究主要集中在栽培、遗传多样性、化学成分和药效学研究等方面^[5,8,10-11]。此外, Song 等^[12]虽报道了毛重楼的叶绿体全基因组, 但未对该植物的基因序列特征、功能与分类、密码子组成等进行研究, 且未与重楼属其他植物叶绿体基因组进行比较, 限制了对其遗传背景、种质资源保护及系统发育进化等方面的认识。鉴于此, 本研究对毛重楼、峨眉重楼 *P. polyphylla* var. *emeiensis* H. X. in, H. Zhang、滇重楼 3 个重楼属物种的新鲜叶片进行了测序, 并对毛重楼叶绿体基因组结构特征和系统进化进行了解析, 同时比较了上述 3 个物种的叶绿体基因组特征差异, 以期研究毛重楼优良品种选育、种质资源保护和植物系统进化发育奠定科学基础。

1 材料

样品的新鲜叶片采自云南省永平县龙门乡种质资源圃 (N99°53', E25°54'), 经大理大学段宝忠教授鉴定为毛重楼 *P. mairei* H. Léveillé (标本号 YN20200820A6)、峨眉重楼 *P. polyphylla* var. *emeiensis* H. X. in, H. Zhang (标本号 YN20191003B2) 和滇重楼 *P. polyphylla* Smith var. *yunnanensis* (Franch.) Hand. -Mazz. (标本号 YN20191002B4) 健康叶片装入取样袋后带回实验室, 用无菌水冲洗数次, 晾干后置于 -80 °C 冰箱保存备用。

2 方法

2.1 基因组 DNA 的提取和测序

取毛重楼、峨眉重楼和滇重楼的新鲜叶片, 采用植物基因组 DNA 试剂盒 (TIANGEN 公司, 北京) 提取总 DNA。采用琼脂糖凝胶电泳检测 DNA 质量。检测合格后的总 DNA 利用 Illumina HiSeq

X Ten 平台进行建库测序。原始序列采用 NGS QC ToolKit 进行质控。原始数据提交到 GenBank, 登录号 MW694831。

2.2 叶绿体基因组的组装与注释

以峨眉重楼叶绿体全基因组序列 (NC_050052) 为序列延伸种子, 在 Linux 系统中运行脚本。使用 NOVOPlasty (V.3.8.3) 对原始测序数据进行组装。采用 Circlator (V.2.0.1)^[13] 软件, 检测 reads 映射的覆盖度和各 contigs 的连接处, 检验组装的正确性。注释、编辑和校正采用 Geneious (V.9.1.4) 进行。叶绿体基因组环状示意图采用 OGDRAW 在线工具绘制^[14]。

2.3 重复序列和密码子使用分析

采用 CPGAVAS2 (<http://47.96.249.172:16019/analyser/home>) 进行重复序列和密码子使用分析^[15]。其中, MISA 用于发现微卫星序列分析^[16], 参数阈值设置为: 单、双、3、4、5、6, 核苷酸参数为 10、6、5、5、5、5; TRF 软件用于串联重复序列分析^[17], 重复单元大小设定 ≥ 7 ; VMATCH 软件用于分析散布重复序列^[18]; GC 含量和密码子使用偏好采用 EMBOSS (v6.3.1)^[19] 中的 Cusp 程序计算获得。

2.4 基因组比较分析

使用 IRscope 可视化工具, 比较毛重楼、滇重楼和峨眉重楼叶绿体基因组 4 个区域边界的差异。并采用 mVISTA 对其进行全基因组比对分析。

2.5 系统进化分析

为探讨毛重楼的系统位置, 本研究按照《中国植物志》百合科分类系统, 从 NCBI 下载了百合科下 22 条物种的叶绿体基因组序列, 包括广义重楼属 *Paris* 植物 15 条, 藜芦属 *Veratrum* 植物 5 条和贝母属 *Fritillaria* L. 植物 2 条。同时选择与百合科亲缘关系较近的菝葜科 (Smilacaceae) 菝葜 *Smilax china* L. 和白背牛尾菜 *S. nipponica* Miq. 作为外类群, 构建最大似然 (maximum likelihood, ML) 系统发育树。具体步骤: 采用 MAFFT (V.7.0)^[20-21] 进行全基因组比对, 然后使用 RAxML (V.8.2.11) 软件, 核苷酸替代模型设置为 GTR+I+G (Bootstrap 1000), 具体参数为: “raxmlHPC-PTHREADS-SSE3 -f a -N 1000 -m GTRGAMMA -x 551314260 -p 551314260 -o Smilax_china_NC_049022, Smilax_nipponica_NC_049024 -T 20”。

3 结果与分析

3.1 叶绿体基因组结构

利用 Illumina HiSeq X Ten 测序, 去除接头和

低质量的数据后,共得到 32 817 588 条长度约为 150 bp 的序列。利用 NOVOPlasty 进行组装,经注释后得到毛重楼的叶绿体基因组。结果表明,毛重楼叶绿体基因组总长 163 819 bp, G/C 含量为 37.05%, A/T 含量为 62.95%, 有明显的 AT 偏向性。基因组呈双链环状(图 1), 具有典型的 4 分区域结构, 包括 1 个大单拷贝区 (large single-copy, LSC)、1 对反向互补重复

区 (inverted repeats, IR) 和 1 个小单拷贝区 (small single-copy region, SSC), 其大小分别为 84 173、33 296、13 054 bp。此外, LSC、IR 和 SSC 区域的 GC 值存在一定的差异, 碱基组成结果见表 1。由表 1 可看出, IR 区的 G/C 含量最高 (39.73%), 其次是 SSC 区 (35.70%) 及 LSC 区 (32.16%), 其可能原因是 IR 区含有高 GC 含量的 rRNA 基因^[22]。

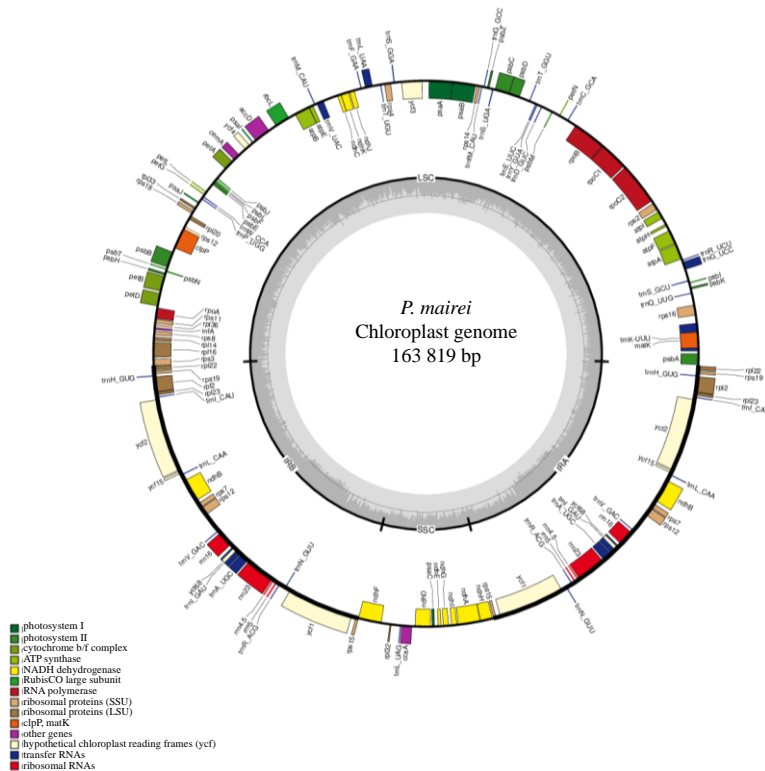


图 1 毛重楼叶绿体基因组图谱

Fig. 1 Gene map of chloroplast genome in *P. mairei*

表 1 毛重楼叶绿体基因组碱基组成

Table 1 Base composition of chloroplast genome in *P. mairei*

区域	T/%	A/%	C/%	G/%	GC/%	碱基长度/bp
IRa	28.60	31.68	18.83	20.89	39.73	33 296
IRb	31.68	28.60	20.89	18.83	39.73	33 296
LSC	32.70	31.60	18.19	17.51	32.16	84 173
SSC	32.07	35.77	16.70	15.46	35.70	13 054
总量	31.61	31.34	18.75	18.30	37.05	163 819

3.2 叶绿体基因组功能及分类

毛重楼叶绿体基因组共包含 113 个基因, 其中编码蛋白基因、rRNA 基因与 tRNA 基因的数量分别为 79、4、30 个。根据其功能可以把它们分为 4 大类: 分别是与光合作用有关的基因、复制有关的基因、未知功能的蛋白质基因, 以及成熟酶基因 (*matK*)、囊膜蛋白基因 (*cemA*) 等其他基因。在这些基因中, 有 10 个蛋白质编码基因 (*ndhB*、*rpl2*、*rpl22*、*rpl23*、*rps12*、

rps15、*rps19*、*rps7*、*ycf1*、*ycf2*)、7 个 tRNA 编码基因 (*tmA-UGC*、*tmH-GUG*、*tmL-CAU*、*tmL-CAA*、*tmN-GUU*、*tmR-ACG*、*tmV-GAC*) 和 4 个 rRNA 编码基因 (*rm16S*、*rm23S*、*rm4.5S*、*rm5S*) 位于 IR 区。对毛重楼叶绿体基因内含子进行统计分析, 发现共有 22 个基因含有内含子, 其中仅有 4 个蛋白质编码基因 (*ycf3*、*clpP*、*ycf1*、*ycf1*) 含有 2 个内含子, 其余 18 个基因只有 1 个内含子 (表 3)。

表 2 毛重楼叶绿体基因组基因列

Table 2 Genes list of chloroplast genome in *P. mairei*

基因类别	基因功能	基因名称
rRNA	核糖体 RNA 基因	<i>rrn16S</i> (×2), <i>rrn23S</i> (×2), <i>rrn5S</i> (×2), <i>rrn4.5S</i> (×2)
tRNA	转运 RNA 基因	<i>trnA-UGC</i> (×2), <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnI-M-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> (×2), <i>trnI-CAU</i> (×2), <i>trnI-GAU</i> (×2), <i>trnK-UUU</i> , <i>trnL-CAA</i> (×2), <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> (×2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> (×2), <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> (×2), <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
光合作用基因	ATP 合酶亚基	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
	亚光系统 II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> , <i>ycf3</i>
	NADH-脱氢酶亚基	<i>ndhA</i> , <i>ndhB</i> (×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	细胞色素复合物 b/f 亚基	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	亚光系统 II	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
复制基因	Rubisco 酶大亚基	<i>rbcL</i>
	Rubisco 酶小亚基	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> (×2), <i>rpl20</i> , <i>rpl22</i> (×2), <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	DNA 依赖核酸聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
其他基因	核糖体小亚基	<i>rps11</i> , <i>rps12</i> (×2), <i>rps14</i> , <i>rps15</i> (×2), <i>rps16</i> , <i>rps18</i> , <i>rps19</i> (×2), <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i>
	乙酰辅酶 A 羧化酶亚基	<i>accD</i>
	c 型细胞色素合成基因	<i>ccsA</i>
	包裹膜蛋白	<i>cemA</i>
	蛋白酶	<i>clpP</i>
	转录起始因子	<i>infA</i>
未知基因	成熟酶	<i>matK</i>
	保守开放性阅读框	<i>ycf1</i> (×2), <i>ycf2</i> (×2), <i>ycf4</i>

表 3 毛重楼叶绿体基因中内含子的位置和长度

Table 3 Location and length of chloroplast genes containing introns in *P. mairei*

基因	起点	终点	外显子 I 长度/bp	内含子 I 长度/bp	外显子 II 长度/bp	内含子 II 长度/bp	外显子 III 长度/bp
<i>trnK-UUU</i>	1391	4066	37	2604	35		
<i>rps16</i>	4823	5824	42	750	210		
<i>trnG-UCC</i>	9309	10 054	24	674	48		
<i>atpF</i>	11 997	13 342	145	791	410		
<i>rpoC1</i>	20 897	23 687	434	706	1651		
<i>ycf3</i>	42 447	44 438	124	729	230	750	159
<i>trnL-UAA</i>	47 241	47 938	35	613	50		
<i>trnV-UAC</i>	52 166	52 838	39	597	37		
<i>clpP</i>	70 663	72 663	71	780	291	636	223
<i>petB</i>	75 575	77 035	6	813	642		
<i>petD</i>	77 233	78 456	6	741	477		
<i>rpl2</i>	85 238	86 720	394	661	428		
<i>ndhB</i>	95 637	97 864	775	695	758		
<i>trnI-GAU</i>	103 478	104 489	42	935	35		
<i>trnA-UGC</i>	104 549	105 422	38	801	35		
<i>ycf1</i>	112 987	116 790	2782	57	726	30	209
<i>ndhA</i>	127 094	129 238	553	1062	530		
<i>ycf1</i>	131 203	135 006	2782	57	726	30	209
<i>trnA-UGC</i>	142 571	143 444	38	801	35		
<i>trnI-GAU</i>	143 504	144 515	42	935	35		
<i>ndhB</i>	150 129	152 356	775	695	758		
<i>rpl2</i>	161 273	162 755	394	661	428		

3.3 散在重复序列、串联重复序列及 SSR 分析

散在重复序列以 E 值 $<1 \times 10^{-4}$ 为阈值进行筛选, 结果显示重复单元以散在方式分布于基因组内, 包括回文重复序列 26 条、正向重复序列 23 条, 其长度均为 110~180 bp。串联重复序列按照总长度大于 20 bp, 且重复单元间相似性 $\geq 90\%$ 进行筛选, 结果显示共有 15 个串联重复序列。简单重复序列 (simple sequence repeat, SSRs) 分析表明, 共有 80 个 SSRs 位点。其中包括 60 个单核苷酸重复基序, 13 个二核苷酸重复基序, 3 个三核苷酸重复基序, 2 个五核苷酸重复基序和 2 个六核苷酸重复基序, 未发现四核苷酸重复基序。SSR 的类型以 A/T 为主, 共有 59 个; 其次为 AT/AT, 共有 12 个 (表 4)。

3.4 叶绿体基因组密码子使用分析

毛重楼叶绿体基因组密码子使用频次统计结果显示, 共有 41 893 个密码子 (图 2)。在这些密码子中, 异亮氨酸 (Leu) 为频次最高的氨基酸, 编码 4766 次, 占 10.21%; 密码子总数 < 2000 次的氨基酸有半胱氨酸 (Cys)、天冬氨酸 (Asp)、组氨酸 (His)、甲硫氨酸 (Met)、脯氨酸 (Pro)、谷氨酰胺 (Gln)、精氨酸 (Arg)、色氨酸 (Trp)、酪氨酸 (Tyr)、终止子 (TER), 其中编码频次最低的氨基酸为半胱氨酸 (Cys), 仅编码 603 次, 占总密码子的 1.29%。这与蒙古韭 *Allium mongolicum* Regel、金铁锁 *Psammosilene tunicoides* W. C. Wu &

表 4 毛重楼叶绿体基因组 SSRs 类型及数量

Table 4 Simple sequence repeats of chloroplast genome in *P. mairei*

核苷酸类型	SSR 重复序列	重复序列个数
单核苷酸	A	30
	C	1
	T	29
二核苷酸	AC	1
	AT	7
	TA	5
三核苷酸	AAT	2
	ATA	1
五核苷酸	CATTT	1
	GAAAT	1
六核苷酸	ATACTA	1
	AATGAC	1

C. Y. Wu 等陆生植物观察到的趋势一致^[23-24]。图 2 显示了毛重楼叶绿体基因组所有蛋白质编码基因的 20 个氨基酸和终止密码子的密码子个数。

3.5 IR 边界变化分析

研究显示, 毛重楼叶绿体基因组共存在 4 个边界, 分别是 IRB-SSC、IRB-LSC、IRA-SSC、IRA-LSC。所选的 3 个重楼属植物叶绿体基因组相对保守 (图 3), 3 个物种的 IRB-LSC 边界均位于 *rps3*, 且该基因在 IRB 区域长度均为 6 bp。IRB-SSC 的边界与 *rps15* 存在间隙, 其中该基因在毛重楼中扩张, 距 SSC 区仅 3 bp, 而峨眉重楼与滇重楼均为 21 bp。所有物种的 IRA-

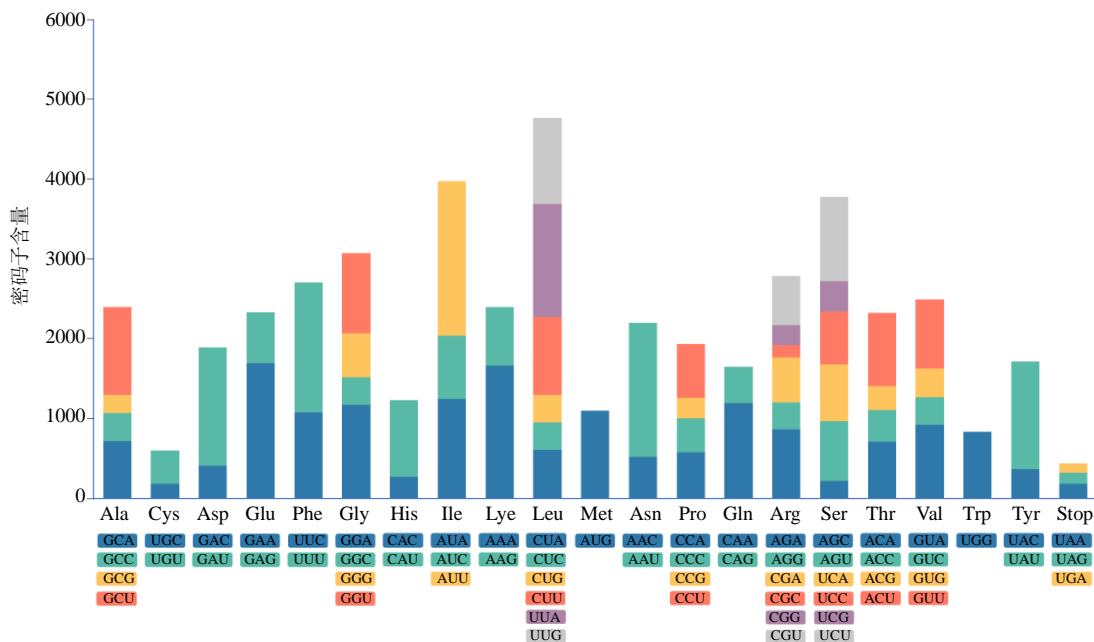


图 2 毛重楼蛋白质编码基因中 20 个氨基酸和终止密码子的密码子含量

Fig. 2 Codon content of 20 amino acid and stop codons in all protein-coding genes of chloroplast genome of *P. mairei*

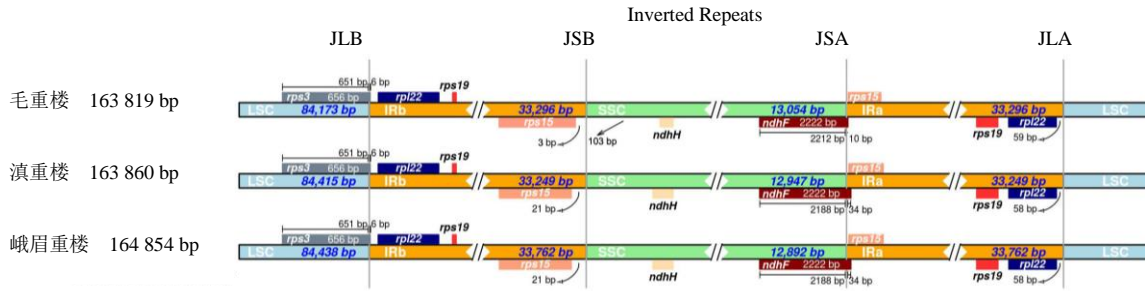


图3 重楼属3种植物叶绿体基因组的LSC、IRs和SSC边界区域的比较分析

Fig. 3 Comparison of LSC, IRs, and SSC border regions among three *Paris* species

SSC 边界均位于 *ndhF* 内, 该基因相对保守长度均为 2222 bp。此外, IRA-LSC 边界均与 *rpl22* 基因存在着 58~59 bp 的间隙。

3.6 基因组序列变异分析

尝试与 NCBI 中重楼属部分物种进行比较, 但发现 NCBI 数据库的多条重楼属物种 *ycf1* 和 *rps15* 基因缺失达 5300 bp, 尚不清楚是物种问题还是测序深度所造成, 因此, 采用了本课题组装拼接的序列进行比对, 以期更进一步理解他们之间的差异。全局比对采用在线基因组比对工具 mVISTA, 以项目组已发表的峨眉重楼叶绿体基因组 (NC_050052) 为参照, 对毛重楼和滇重楼 (NC_052909) 叶绿体基因组序列进行比对分析 (图 4)。结果显示, 3 条叶绿体基因组序列中非编码区变异高于保守的蛋白编码区域, LSC 区、IR 区变异明显大于 SSC 区, rRNA 基因高度保守,

几乎没有变异。由图可见, 变异较大的基因有 *atpF*、*trnL-UAA*、*cemA* 和 *accD*, 其他基因的保守程度非常高, 绝大多数的基因相似度都在 90% 以上。3 个重楼属植物的基因间区, 如 *trnT-UGU-trnL-UAA*、*rbcl-accD*、*trnL-CAU-ycf2*、*trnF-trnV*、*trnE-UUC-trnT-GGU*、*trnS-GCU-trnG-UCC*、*rps8-rps14* 等, 其基因间区变异均大于基因区。此外, 毛重楼变异相对较大, 如 *trnQ-UUG* 和 *ycf2* 基因, 以及 *psbK-psbL*、*atpH-atpL* 和 *petD-rpoA* 基因间区。这些位点为毛重楼的分子鉴定提供了新的位点资源。

3.7 毛重楼叶绿体基因组系统发育分析

为确定毛重楼的系统位置, 基于 24 条序列构建的 ML 系统进化树见图 5。从图 5 可见, 2 个外类群物种独立为一支; 毛重楼与蚤休属 *Euthyra* L. 的禄劝花叶重楼等聚成一支, 再与延龄草属 *Trillium* L.

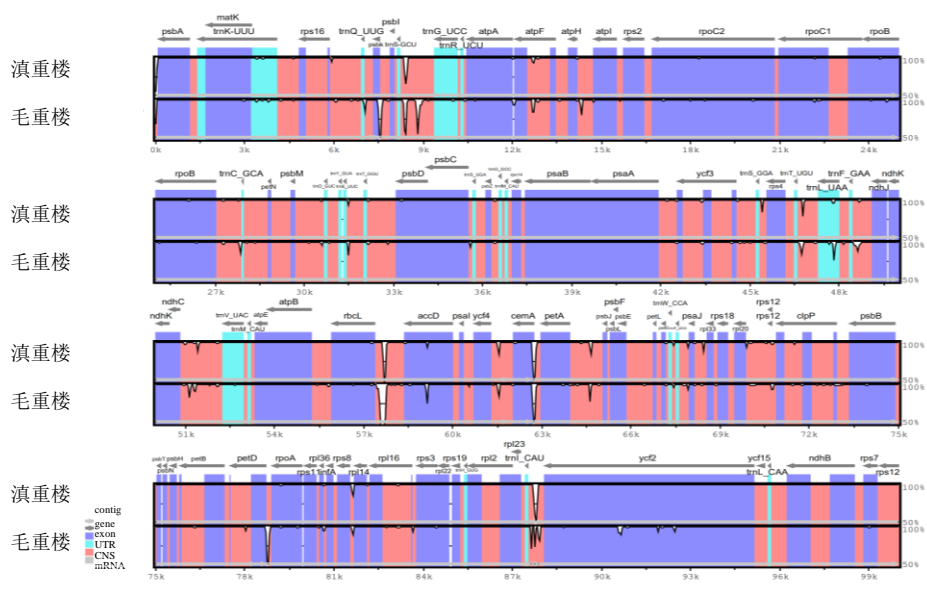


图4 毛重楼与相关重楼属植物叶绿体基因组全局比对图

Fig. 4 Global alignment of chloroplast genomes of three *Paris* species

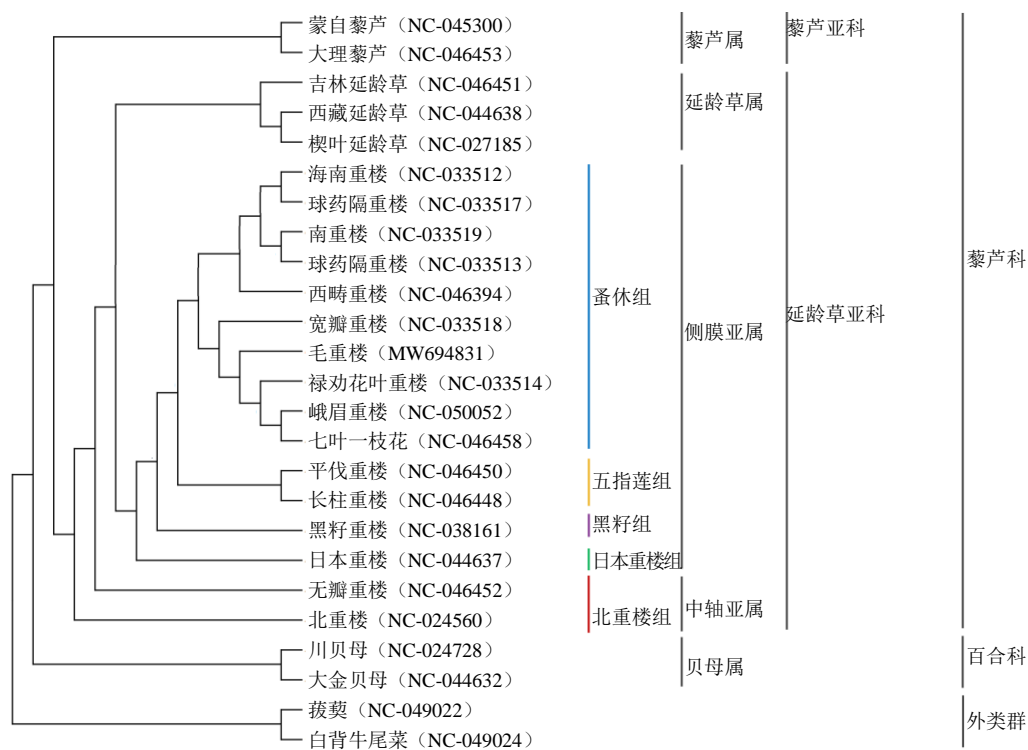


图 5 基于叶绿体全基因组序列构建 ML 系统进化树

Fig. 5 Phylogenetic analysis based on chloroplast genome sequences by maximum likelihood (ML) tree

聚成一支，之后与北重楼属 *Paris* L. 聚为一支，这 3 支在《被子植物 APG 分类法 IV》(APG IV) 分类系统均归属为延龄草亚科 (Subfam. Trillioideae A. Gray)，表明相较北重楼属而言，毛重楼与延龄草属的亲缘关系更近；延龄草亚科 (Subfam. Trillioideae) 与藜芦亚科 (Subfam. Melanthioideae Eaton) 聚为一支归属于藜芦科。值得一提的是，百合科贝母属与藜芦科分布在不同的进化支上，支持率为 100。在之前的研究中，学术界对重楼属植物的系统进化关系一直存在着较大争议^[25-27]。如李恒^[26]和 Dumortier^[28]提出重楼属与延龄草属二者构成一个单系类群，亲缘关系最近，应归属于延龄草科 (Trilliaceae)；而在《中国植物志》中重楼被归属于百合科 (Liliaceae)。本研究结果表明，相较藜芦科，毛重楼与百合科的关系较远，基于叶绿体基因组的系统进化分析，支持将毛重楼从百合科中独立，归属为藜芦科，这一结果与前人的研究结果一致^[29]。从属下等级来看，基于叶绿体基因组的 ML 树显示，广义重楼属被分为 5 个类群，分别是蚤休组 *Euthyra* Sect.、五指莲组 *Axialis* Sect.、黑籽组 *Thibeticae* Sect.、日本重楼 *Kinugasa* Sect.

和北重楼组 *Paris* Sect.，这一结果与前人的研究结果存在一定的分歧，如 Franchet 认为重楼属分为 *Euparis* Sect.^[30]和蚤休组 2 个组；而 Hara^[31]和 Takhtajian^[32]学者则认为重楼属应分为北重楼组、日本重楼组和侧膜组 3 个组^[31-32]。李恒^[26,33]则将其分为侧膜亚属和中轴亚属，而后再分成海南重楼组 *Dunniana* Sect.、蚤休组、禄劝花叶重楼组 *Marmoratae* Sect.、球药隔重楼组 *Fargesiana* Sect.、五指莲组、黑籽组、日本重楼组和北重楼组。鉴于重楼属植物分类的分歧，从叶绿体基因组系统发育的角度来看，重楼属植物应分为 5 个类群较为恰当。

4 讨论

本研究完成了毛重楼叶绿体基因组的测序、组装和注释，并对其结构、GC 含量等进行了分析。研究表明毛重楼具有典型的环状 DNA 分子，且具有保守的 4 分状结构，全长 163 819 bp，与本课题前期组装的同属植物滇重楼、峨眉重楼等叶绿体基因组大小相似^[34-35]，其 GC 含量为 36.5%，共编码 125 基因，其中 14 个蛋白编码基因含有内含子。此外，在该基因组中存在一些未知功能的基因，如

ycf1、*ycf2* 和 *ycf4*，到目前为止，功能尚未清楚的 *ycf15* 基因虽仍在大多数被子植物中存在，但在重楼属植物的叶绿体基因组中已完全消失。本研究从毛重楼叶绿体基因组中分别共检测到 80 个 SSR，共发现 15 条串联重复序列及 49 条散在重复序列。这些重复序列可为重楼属物种分子标记开发，物种或产品鉴定提供理论依据。

基于叶绿体基因组的 ML 进化树显示，广义重楼属^[26]被分为蚤休组、五指莲组、黑籽组、日本重楼组和北重楼组 5 个类群，其中毛重楼与蚤休组聚为一支，相较于北重楼组，毛重楼与蚤休组的亲缘关系最近，建议将毛重楼归属为蚤休组。此外，相较于百合科，毛重楼与藜芦科藜芦属亲缘关系更近，支持将蚤休属归于藜芦科的观点。本研究表明叶绿体基因组可在物种水平上有效解决重楼属的系统进化问题。毛重楼叶绿体基因组的组装和序列分析，为后续开展群体遗传学和遗传多样性研究奠定了科学基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 刘玉雨, 徐福荣, 范敏, 等. 重楼属植物在少数民族医药中的应用 [J]. 世界科学技术—中医药现代化, 2019, 21(3): 449-456.
- [2] 中国药典 [S]. 一部. 2020: 23.
- [3] 方海兰, 夏从龙, 段宝忠, 等. 基于 DNA 条形码的中药材种子种苗鉴定研究: 以重楼为例 [J]. 中药材, 2016, 39(5): 986-990.
- [4] 鄢连和, 姜程曦, 朱美晓. 畲族珍稀濒危特有药用植物资源现状分析 [J]. 中草药, 2014, 45(22): 3351-3355.
- [5] 黄贤校, 高文远, 谷克仁, 等. 毛重楼的化学成分研究 [J]. 中草药, 2009, 40(9): 1366-1369.
- [6] 刘杨. 南重楼和毛重楼的化学成分研究 [D]. 西安: 中国人民解放军空军军医大学, 2018.
- [7] 黄圆圆, 刘大会, 彭华胜, 等. 15 种重楼属植物中 8 种甾体皂苷的含量测定 [J]. 中国中药杂志, 2017, 42(18): 3443-3451.
- [8] 刘江. 滇产 7 种重楼属药用植物主要成分含量分析及止血、镇痛药效学研究 [D]. 大理: 大理大学, 2017.
- [9] 刘潇潇, 王磊, 隆颖, 等. 毛重楼的化学成分研究 [J]. 中国中药杂志, 2014, 39(16): 3107-3111.
- [10] 辛本华. 重楼属植物遗传多样性研究与品质评价分析 [D]. 雅安: 四川农业大学, 2012.
- [11] 张海珠, 赵飞亚, 陶爱恩, 等. 基于成分-活性整体相似性的重楼替代资源筛选 [J]. 中草药, 2018, 49(18): 4366-4373.
- [12] Song Y, Xu J, Zhang Y J, *et al.* Characterization of complete chloroplast genome of traditional Chinese medicinal plants *Paris Mairei* and its phylogenetic positions [J]. *Mitochondrial DNA B Resour*, 2019, 4(2): 3822-3823.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Beier S, Thiel T, Münch T, *et al.* MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [17] Benson G. Tandem repeats finder: A program to analyze DNA sequences [J]. *Nucleic Acids Res*, 1999, 27(2): 573-580.
- [18] 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.
- [19] Langmead B, Trapnell C, Pop M, *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome [J]. *Genome Biol*, 2009, 10(3): R25.
- [20] Katoh K, Rozewicki J, Yamada K D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization [J]. *Brief Bioinform*, 2019, 20(4): 1160-1166.
- [21] Kuraku S, Zmasek C M, Nishimura O, *et al.* Leaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity [J]. *Nucleic Acids Res*, 2013, 41(web server issue): W22-W28.
- [22] 吴茜, 姜梅, 陈海梅, 等. 旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组比较分析和系统发育研究 [J]. 药学报, 2020, 55(5): 1042-1049.
- [23] 王媛媛, 杨美青. 蒙古韭叶绿体基因组密码子使用偏好性分析 [J]. 分子植物育种, 2021, 19(4): 1084-1092.
- [24] 高亚芳, 刘莹莹, 杨从卫, 等. 金铁锁叶绿体基因组序列及其系统发育分析 [J]. 中草药, 2019, 50(22): 5532-5536.
- [25] Song Y, Wang S J, Ding Y M, *et al.* Chloroplast genomic resource of *Paris* for species discrimination [J]. *Sci Rep*, 2017, 7(1): 3427.
- [26] 李恒. 重楼属植物 [M]. 北京: 科学出版社, 1998: 269.

- [27] 赵志勇, 高文远, 黄贤校, 等. 重楼属植物分类学研究进展 [J]. 药物评价研究, 2010, 33(3): 244-246.
- [28] Dumortier B C. *Analyse Des Familles Des Plantes : Avec L'indication Des Principaux Generes Qui s'y Rattachent* [M]. *Toumay: Casterman*, 1829: 126.
- [29] 杨丽芳. 重楼属叶绿体系统发育基因组学与生物地理学研究 [D]. 昆明: 云南大学, 2019.
- [30] Franchet A. *Plantarum sinensium ecloge secunda* [J]. *J Bot*, 1898, 12(17/18): 253-264.
- [31] Hara H. Variation in *Paris polyphylla* Smith, with reference to other Asiatic species [J]. *J Fac Sci Univ Tokyo, Section 3*, 1969, 10(10): 141-180.
- [32] Takhtajan A. A revision of *Daiswa* (Trilliaceae) [J]. *Brittonia*, 1983, 35(3): 255-270.
- [33] 李恒. 重楼属的分类研究 [J]. 植物研究, 1986, 6(1): 109-144.
- [34] Fan M, Liu J, Wang J, *et al.* The first complete chloroplast genome sequence of *Paris polyphylla* var. *emeiensis*, a rare and endangered species [J]. *Mitochondrial DNA Part B*, 2020, 5(3): 2172-2173.
- [35] Guan X, Yang Q S, Wang S, *et al.* Chloroplast phylogenomic analysis provides insights into the evolution of *Paris liana* sp. nov [J]. *Mitochondrial DNA Part B*, 2021, 6(2): 346-348.

[责任编辑 时圣明]