

• 药材与资源 •

角蒿属药用植物叶绿体基因组比较及系统发育分析

马海霞¹, 张亚颖¹, 张玉婉², 益西卓玛¹, 梁焯森³, 李欣冉³, 郭云涛³, 封 烁^{1*}

1. 青海大学 省部共建三江源生态与高原农牧业国家重点实验室, 青海 西宁 810016

2. 青海大学农牧学院, 青海 西宁 810016

3. 青海大学医学部, 青海 西宁 810016

摘要: 目的 研究角蒿属 *Incarvillea* Juss. 药用植物叶绿体基因组特征、变异程度及系统发育关系, 为角蒿属物种分子鉴定、系统发育及药用植物资源保护等研究提供参考。方法 以 5 种角蒿属药用植物为研究对象, 通过高通量测序获取密生波罗花叶绿体基因组, 并结合同属 4 种药用植物的叶绿体基因组序列, 利用生物信息学方法比较 5 种角蒿属药用植物叶绿体基因组之间的结构特征及变异程度, 同时以凌霄为外类群分析了角蒿属的系统发育关系。结果 5 种角蒿属药用植物的叶绿体基因组为典型的四分体结构, 均包含一个大单拷贝区、一个小单拷贝区及 2 个反向重复区。密生波罗花 *Incarvillea compacta* 的叶绿体基因组序列长度为 150 144 bp, GC 含量为 40%, 共注释到基因 119 个, 其中 76 个蛋白质编码基因、35 个转运 RNA (transfer RNA, tRNA) 基因、8 个核糖体 RNA (ribosomal RNA, rRNA) 基因。重复序列分析发现两头毛 *I. arguta* 叶绿体基因组重复序列数量最少, 且只有两头毛同时含有 4 种类型的重复序列, 其余 4 种药用植物的重复序列总数及重复类型相同。密生波罗花有 44 个简单重复序列 (simple sequence repeats, SSRs), 且以单核苷酸 (A/T) 重复为主。叶绿体基因组比较分析发现角蒿属内 5 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, 黄波罗花 *I. lutea* 与其他 4 种药用植物相比, 具有较高的序列变异, 而两头毛在进化上最为保守, 同时发现藏波罗花 *I. younghusbandii* 和密生波罗花的序列相似性较高; 共线性结果表明, 角蒿 *I. sinensis*、密生波罗花及藏波罗花中存在程度不一的基因重排。选择压力分析结果表明在角蒿属 5 种药用植物中 *rps2*、*rps7*、*psbN* 蛋白编码基因均受到正选择。系统发育分析结果显示, 密生波罗花与藏波罗花亲缘关系最近, 与两头毛的亲缘关系最远。结论 5 种角蒿属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。其中, 两头毛的变异程度最低, 进化上最保守。就亲缘关系而言, 密生波罗花和藏波罗花的亲缘关系最近。基于叶绿体基因组构建的系统发育树在一定程度上可以揭示角蒿属物种的系统发育关系。

关键词: 角蒿属; 密生波罗花; 两头毛; 黄波罗花; 藏波罗花; 角蒿; 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(2024)17-5972-10

DOI: 10.7501/j.issn.0253-2670.2024.17.023

Chloroplast genome comparison and phylogenetic analysis of five medicinal plants of *Incarvillea*MA Haixia¹, ZHANG Yaying¹, ZHANG Yuwan², YIXiZhuoma¹, LIANG Zhuosen³, LI Xinran³, GUO Yuntao³, FENG Shuo¹

1. Qinghai University, State Key Laboratory of Plateau Ecology and Agriculture, Xining 810016, China

2. College of Agriculture and Animal Husbandry, Qinghai University, Xining 810016, China

3. College of Medicine, Qinghai University, Xining 810016, China

Abstract: Objective To understand the chloroplast genome characteristics, degree of variation and phylogenetic relationship of medicinal plants of *Incarvillea*. This study will provide a reference for molecular identification, phylogeny and conservation resources of *Incarvillea*. **Methods** Chloroplast genomes of *Incarvillea compacta* were obtained by highthroughput sequencing, and combined with chloroplast genome sequences from four species of *Incarvillea* medicinal plants. Structural characteristics and variation degree of chloroplast genomes of five medicinal plants of *Incarvillea* were compared by bioinformatics method. And the phylogenetic relationship of *Incarvillea* was analyzed

收稿日期: 2024-01-02

基金项目: 青海省“帅才科学家”负责制项目 (2024-SF-102); 青海大学农牧学院大学生科研训练项目 (NKX202304)

作者简介: 马海霞, 硕士研究生, 研究方向为植物生态学。E-mail: 18997063244@163.com

*通信作者: 封 烁, 副教授, 从事适应性进化研究。E-mail: fengshuo8894@126.com

with *Campsis grandiflora* as an outgroup. **Results** The chloroplast genomes of five species of *Incarvillea* had a double stranded ring structure, including a long single copy region, a short single copy region and two reverse repeats. The chloroplast genome sequence length of *I. compacta* was 150, 154 bp with a GC content of 40%. A total of 119 genes were annotated, including 76 protein-coding genes, 35 tRNA genes, and eight rRNA genes. Sequence repeat analysis showed that only *I. arguta* contained four repeat types, and the repeat sequence number and repeat types of the other four medicinal plants were the same. There were 44 SSRs in *I. compacta*, and SSRs were mainly single nucleotide (A/T). Chloroplast genome comparative analysis found that the degree of variation the non-coding region of chloroplast genome revealed a greater degree than the gene coding region of five *Incarvillea* medicinal species. Compared to the other four medicinal species, *I. lutea* had higher sequence variation, but *I. arguta* is the most evolutionarily conservative. It was also found to have high sequence similarity between *I. younghusbandii* and *I. compacta*. Collinearity analysis showed that gene rearrangement was found to varying degrees in *I. sinensis*, *I. compacta* and *I. younghusbandii*. Results of selection pressure analysis showed that *rps2*, *rps7* and *psbN* genes were positively selected in five medicinal plants of *Incarvillea*. Phylogenetic analysis revealed that *I. compacta* and *I. younghusbandii* were the closest in the genetic relationship, and the furthest relation between *I. compacta* and *I. arguta*. **Conclusion** The chloroplast genome structure of the five species of *Incarvillea* is conservative, but the degree of variation is different. Among them, *I. arguta* has the lowest degree of variation and is the most evolutionarily conservative. In terms of kinship, the relatives of *I. compacta* and *I. younghusbandii* are the closest. The chloroplast genome-based phylogenetic tree can reveal the phylogenetic relationship of *Incarvillea* species to a certain extent.

Key words: *Incarvillea* Juss.; *Incarvillea compacta* Maxim.; *Incarvillea arguta* (Royle) Royle.; *Incarvillea lutea* Bur. et Franch.; *Incarvillea younghusbandii* Sprague; *Incarvillea sinensis* Lam.; chloroplast genome; variation; evolution; phylogeny

叶绿体作为光合作用的细胞器，同时还兼具氨基酸、核苷酸、脂肪酸、植物激素、维生素等的生物合成功能，拥有独立完整的基因组，在陆地植物和藻类的生理、发育中发挥着重要作用^[1]。叶绿体基因组在被子植物中遵循母系遗传，序列较为保守，便于组装，同时携带大量自我调节的编码基因，是半自主性细胞器。相较于线粒体基因组和核基因组，叶绿体基因组在结构、基因数量和基因组成上更为保守，进化速率也相对适中^[2-3]。基于以上优势，叶绿体基因组在揭示物种起源、进化演变以及不同物种之间的亲缘关系等方面具有重要价值^[4]。近年来，叶绿体基因组已应用于羌活属^[5]、柴胡属^[6]、甘草属^[7]、芍药属^[8]及龙胆属^[9]等药用植物的分子标记和系统发育等研究，这推动了药用植物的分子鉴定和系统进化等研究进展，进一步促进了本草类基因组学的发展^[10]。

角蒿属两头毛亚属的两头毛 *Incarvillea arguta* (Royle) Royle.、波罗花亚属的密生波罗花 *I. compacta* Maxim.、黄波罗花 *I. lutea* Bur. et Franch.、藏波罗花 *I. younghusbandii* Sprague，以及角蒿亚属的角蒿 *I. sinensis* Lam.均为紫葳科一年或多年生的草本药用植物，多分布在喜马拉雅山和东亚等地区。角蒿属植物在全世界约有 15 种，分布在我国云南西北部、四川西部、西藏及青海等地的就有 11 种^[11]。角蒿属植物作为传统的药用植物，多被用于治疗各种临床疾病。邹琼宇等^[12]提出，角蒿属植物中的单萜生物碱作为其特征化合物，具有较强的镇痛作用。

除此之外，它们还有抗炎、抗贫血、抗癌等各种药理活性。而两头毛是各民族较为常用的中药，有抗炎、抗菌、防止胆石症等疗效^[13]。密生波罗花作为常见藏药，其根、花、种子均可入药^[14]，有理气止痛、平肝潜阳、清热除湿的功效，临床上可用于治疗胃痛、黄疸、消化不良、耳脓、月经不调、高血压、肺出血等症状。宋超等^[14]运用硅胶柱色谱法从密生波罗花中分离鉴定了 1 组以神经酰胺为主的混合物，发现密生波罗花神经酰胺具有显著的抗炎活性。同时，李伟博等^[15]发现密生波罗花中营养成分丰富且氨基酸种类齐全，具有很好的抗氧化活性，可作为天然抗氧化剂用于保健品的开发。黄波罗花入药部位为根部，具有滋补功效，同时它还有抗炎镇痛之疗效^[16]。藏波罗花与同亚属的密生波罗花、黄波罗花一道作为传统藏药材“欧曲”，可用于治疗消化不良、黄疸、高血压、肺出血等临床症状^[17]。

目前，关于角蒿属植物的研究主要集中于化学成分^[12-15]、药理活性^[12-13, 16-17]等方面，从比较基因组学角度阐述角蒿属药用植物叶绿体基因组的研究较少。同时，随着对角蒿属药用植物功效的深入研究，陆续挖掘出其不同的药用价值，但由于角蒿属植物形态和功效的相似以及种间变异错综复杂的原因，导致本草记载中常出现同名异物或同物异名的情况。因此对角蒿属药用植物的叶绿体基因组研究可作为本草基源考证的一个方法，为角蒿属物种分子鉴定、系统发育以及资源保护等研究提供一定的参考。

1 材料与仪器

1.1 材料

本研究的密生波罗花采自青海省玉树藏族自治州 (32°59'N, 97°13'E), 由青海大学益西卓玛讲师鉴定为密生波罗花 *I. compacta* Maxim. 的叶。采集长势良好、无病虫害的新鲜叶片的样品在液氮中进行冷冻, 随后置于-80 °C 冰箱保存备用。

1.2 仪器

MDF-U54V 型立式医用低温箱 (松下电器, 日本); DK-8D 型三孔电热恒温水槽 (上海瑞稳仪器设备厂); 5430/R 型小型台式高速冷冻离心机 (艾本德公司, 德国); Illumina NovaSeq 6000 (Illumina 公司, 美国)。

2 方法

2.1 DNA 提取及测序

用改良的 CTAB 法从密生波罗花新鲜叶片中分离出总 DNA^[18]。使用 NEBNext® Ultra™ DNA 文库制备试剂盒生成测序文库^[19]。文库质量合格后, 在 Illumina NovaSeq 6000 平台进行测序, 获得原始数据。使用 Fastp v0.23.0^[20] 软件对低质量序列进行过滤, 获得的 clean reads 用于后续拼接。藏波罗花 *I. younghusbandii* Sprague 叶绿体全基因组由课题组前期组装、注释得到^[21]。其余 3 种角蒿属植物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。

表 1 角蒿属和其他属物种的叶绿体基因组 GenBank 登记号

Table 1 Chloroplast GenBank accession numbers of species in <i>Incarvillea</i> and other genera	
物种	GenBank 登记号
两头毛 <i>I. arguta</i>	MG763885.1
角蒿 <i>I. sinensis</i>	MT937254.1
黄波罗花 <i>I. lutea</i>	MG674921.1
凌霄 <i>C. grandiflora</i>	NC_061042.1
羽叶照夜白 <i>Nyctocalos pinnatum Steenis</i>	NC_061699.1
齿叶猫尾木 <i>Dolichandrone spathacea</i> (L. f.) K. Schum.	NC_054307.1
葫芦树 <i>Crescentia cujete</i> L.	KT182634.2
火焰树 <i>Spathodea campanulata</i> P. Beauv	NC_049000.1
梓树 <i>Catalpa ovata</i> G. Don	NC_047416.1
黄金树 <i>Catalpa speciosa</i> Teas	MT319818.1
美国梓树 <i>Catalpa bignonioides</i> Walter	MW148805.1
紫叶美国梓树 <i>Catalpa purpurea</i> Griseb	MW148798.1
硬骨凌霄 <i>Tecomaria capensis</i> Spach	NC_037462.1

2.2 叶绿体基因组组装、注释及物理图谱绘制

利用 GetOrganelle v1.6.4^[22] 软件拼接密生波罗花叶绿体基因组, 用 Circlator v2.0.1^[23] 软件检测组装的正确性。采用 CPGAVAS2^[24] 软件进行注释, 进行手工调整后。密生波罗花的叶绿体基因组圈图采用 Chloroplot^[25] 在线软件 (<https://irscope.shinyapps.io/Chloroplot/>) 绘制。密生波罗花的叶绿体基因组序列和基因注释信息已上传至 NCBI 数据库, 登录号为 OR603942。

2.3 重复序列分析

利用 REPuter^[26] 软件分别识别 5 种角蒿属药用植物叶绿体基因组包含的正向、反向、回文和互补的重复序列, 使用软件默认参数。利用 MISA^[27] 软件 (<https://webblast.ipk-gatersleben.de/misa/>) 对 5 种角蒿属药用植物的叶绿体基因组进行 SSRs 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的最少重复次数分别为 10、5、4、3、3、3^[27]。

2.4 叶绿体基因组比较分析

反向重复区域 (inverted repeat, IR) 在叶绿体基因中具有高度保守性, IR 边界的扩张和收缩被认为是被子植物叶绿体全基因组大小变化的主要机制^[28]。使用 IRscope^[29] 软件 (<https://irscope.shinyapps.io/irapp/>) 对 5 种角蒿属药用植物叶绿体全基因组 IR 区的扩张与收缩进行可视化。

利用 mVISTA^[30] 软件 (<https://genome.lbl.gov/vista/mvista/submit.shtml>) 对 5 种角蒿属药用植物叶绿体基因组进行全局比对分析, 选用全局比对模式 (Shuffle-LAGAN), 参考背景选择角蒿, 目的是对 5 种角蒿属叶绿体基因组的整体差异进行可视化。

基因共线性分析可比较物种之间同源基因的位置, 揭示比较物种之间的基因结构以及基因顺序的异同。本研究使用 Mauve^[31] 软件对角蒿属内 5 种药用植物叶绿体基因组的保守区域、局部共线性和基因组重排倒置进行鉴定^[32]。

2.5 选择压力分析

选择压力是指外界作用在某个生物进化过程中的压力, 使得该生物能更好地适应环境^[33]。使用 KaKs_calculator v3.0^[34] 软件对角蒿属 5 种药用植物叶绿体基因组所共有的蛋白编码基因进行选择压力分析, 即通过两两比较计算出各个蛋白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)。一般认为, $K_a/K_s > 1$ 时该基因受到自然选择压力, 即正选择; $K_a/K_s = 1$ 时, 该基因受到中性选择; $K_a/K_s < 1$ 时, 则认为该基因受到纯化选择^[35]。

2.6 系统发育分析

使用 MAFFT v7.388^[36] 软件进行序列比对矫正, 利用 IQ-TREE v2.2.5^[37] 软件的 ModelFinder 筛选最优模型, 自展值设置为 1 000 构建最大似然 (maximum likelihood, ML) 树。

3 结果与分析

3.1 角蒿属叶绿体基因组结构的基本特征

角蒿属内 5 种药用植物的叶绿体基因组均为典型的环状四分体结构, 均包含 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)。5 种角蒿属药用植物中叶绿体基因组

总长度最短的是密生波罗花 (150 154 bp), 最长的为角蒿 (162 086 bp) (表 2)。5 种药用植物叶绿体基因组中基因总数为 117~126 个 (表 2), 其中蛋白质编码基因数量为 71~81 个, tRNA 基因数量为 35~38 个, rRNA 基因数量均为 8 个。

密生波罗花叶绿体基因组全长 150 154 bp (图 1), LSC 长 81 527 bp, SSC 长为 21 925 bp, IR 长为 23 351 bp。密生波罗花叶绿体基因组 GC 含量为 40%, 注释得到基因共 119 个, 其中包括 76 个蛋白编码基因 (表 2)、35 个 tRNA 基因和 8 个 rRNA 基因。密生波罗花叶绿体基因组内含子分析发现 (表 3), 大部分基因不含内含子, 1 个基因 (*ycf3*) 含有 2 个内含子。根据注释结果, 将不同的蛋白编码基因功能大致可以分为 4 类, 包括参与光合作用的基因 (42 个)、参与自我复制的基因 (25 个)、其他基因以及一些未知功能的基因 (表 3)。

表 2 5 种角蒿属叶绿体基因组序列的基本特征信息

Table 2 Basic information of chloroplast genomes of five plants of *Incarvillea*

物种	两头毛	黄波罗花	角蒿	藏波罗花	密生波罗花
叶绿体基因组大小/bp	159 135	159 032	162 086	159 323	150 154
LSC 长度/bp	87 192	86 538	82 632	80 197	81 527
SSC 长度/bp	19 023	18 926	8 666	9 030	21 925
IR 长度/bp	26 460	26 784	35 394	35 048	23 351
GC 含量/%	39	40	39	40	40
基因总数/个	117	121	126	120	119
蛋白编码基因数/个	71	77	81	77	76
tRNA 基因数/个	38	36	37	35	35
rRNA 基因数/个	8	8	8	8	8

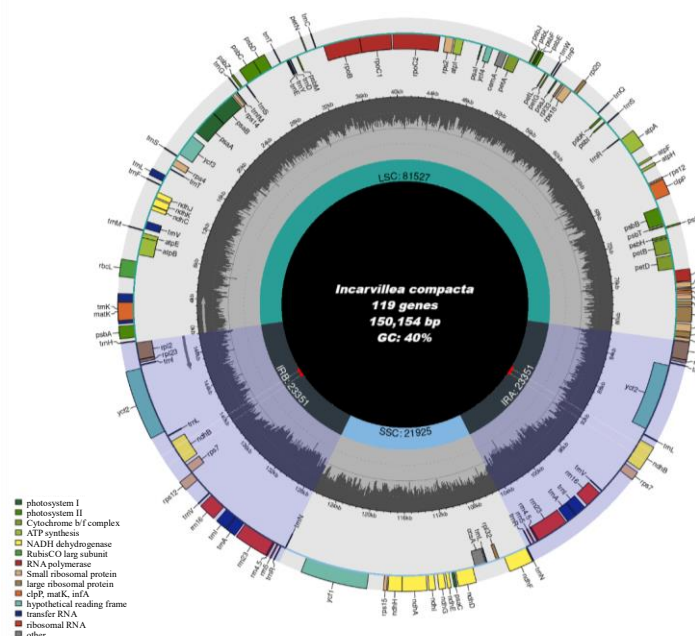


图 1 密生波罗花叶绿体基因组图谱

Fig. 1 Gene map of *I. compacta* chloroplast genome

表 3 密生波罗花叶绿体基因组注释信息

Table 3 Annotation information of cp genome in *I. compacta*

基因类别	基因分组	基因名称	数量
光合作用基因	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>	42
	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>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> , <i>ycf3</i> **	
	Subunits of NADH dehydrogenase	<i>ndhA</i> *, <i>ndhB</i> ×2*, <i>ndhC</i> , <i>ndhE</i> , <i>ndhH</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 photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	
	Subunit of rubisco	<i>rbcL</i>	
自我复制基因	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> *, <i>rpl2</i> ×2, <i>rpl20</i> , <i>rpl22</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	25
	DNA dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *	
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> ×2, <i>rps8</i> , <i>rps11</i> , <i>rps12</i> × 2*, <i>rps14</i> , <i>rps15</i> , <i>rps18</i> , <i>rps19</i>	
其他基因	c-type cytochrom synthesis	<i>ccsA</i>	5
	Envelop membrane protein	<i>cemA</i>	
	Maturase	<i>matK</i>	
	Protease	<i>clpP</i> *	
	Translational initiation factor	<i>infA</i>	
未知功能基因	Conserved open reading frames	<i>ycf1</i> , <i>ycf2</i> ×2, <i>ycf4</i>	4
总计			76

*含有 1 个内含子; **含有 2 个内含子。

*means one intron; **means two introns.

3.2 重复序列分析

叶绿体基因组的重复序列包括散在重复序列和串联重复序列^[38]。对 5 种角蒿属药用植物叶绿体基因组进行散在重复序列分析, 结果显示 5 种角蒿属药用植物中只有两头毛同时包含互补重复、反向重复, 同时两头毛相较于其他 4 种角蒿属植物而言, 它的重复序列数量最少 (图 2), 这表明两头毛在进化过程中可能更为保守。除两头毛外, 其余 4 种角蒿属植物的长重复序列总数相等, 同时构成的重复序列类型也相同, 即只有正向重复和回文重复这 2 种。其中, 黄波罗花所含有的正向重复序列数量最多, 有 36 个, 推测相较于其他 4 种角蒿属植物, 黄波罗花的叶绿体基因组更容易发生序列变异。

简单重复序列 (simple sequence repeats, SSRs) 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^[39]。SSRs 分析表明, 5 种药用植物的 SSRs 数量在 42~51 个 (表 4), 藏波罗花中 SSRs 数量最少为 42 个, 角蒿中 SSRs 数量最多有 51 个。5 种植物的 SSRs 类型多集中在单核苷酸重复, 其中 A/T 重复最多, 且多偏向于 T。二核苷酸重复只在两头毛和角蒿中出现。三核苷酸重复多集中在 AAG 和 TCT。

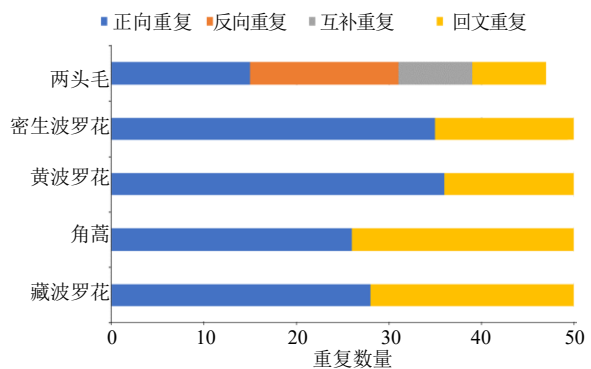


图 2 5 种角蒿属叶绿体基因组的重复序列

Fig. 2 Repeat sequences in chloroplast genomes of five plants of *Incarvillea*

3.3 角蒿属叶绿体基因组比较分析

为了解角蒿属 5 种药用植物间叶绿体基因组之间的差异, 将密生波罗花叶绿体基因组的 LSC-IRb (JLB)、SSC-IRb (JSB)、SSC-IRa (JSA)、LSC-IRa (JLA) 边界与其余植物进行比较。结果表明, 密生波罗花的 *rps19* 基因距 JLB 边界 56 bp, 藏波罗花的 *rps19* 基因距离 JLA 边界 40 bp, 而角蒿中 *rps19* 基因缺失。

表 4 5 种角蒿属叶绿体基因组的 SSRs 统计

Table 4 Statistical analysis of SSRs in chloroplast genomes of five plants of *Incarvillea*

SSRs 类型	两头毛	黄波罗花	角蒿	藏波罗花	密生波罗花
单核苷酸/个	26	27	30	24	25
二核苷酸/个	2	0	7	0	0
三核苷酸/个	10	3	3	4	4
四核苷酸/个	11	4	9	3	4
五核苷酸/个	1	4	0	3	3
六核苷酸/个	0	8	2	8	8
总计/个	50	46	51	42	44

两头毛的 JLB 边界在 *rps19* 和 *rpl22* 之间, *rps19* 基因距离 JLB 边界 3 bp; 黄波罗花的 JLB 边界在 *rpl22* 基因上 (图 3)。角蒿属内 5 种药用植物在基因组的 JLB、JSB、JSA、JLA 边界上基因的收缩和扩张存在明显差异。如 *rpl2* 基因仅在两头毛和黄波罗花中发生了基因收缩, 但 *rpl22* 基因在这两个物种中发生了基因扩张。在黄波罗花、藏波罗花和密生波罗花的 SSC 区存在 *ndhF* 基因缺失, 在两头毛 LSC 区存在 *trnH* 基因缺失。

同时, *rpl2* 基因均位于 5 种植物叶绿体基因组的 IR 区, *rpl22* 基因大多位于 LSC 区, 但在黄波罗花中, 该基因跨于 LSC 区和 IRb 区。在两头毛和黄波罗花叶绿体基因组中, *rps19* 基因位于 IR 区, *ycf1* 基因距离 JSB 边界 1 bp。同时, *ycf1* 基因在两头毛和黄波罗花中存在基因扩张现象, 但在其余 3 个物种中 *ycf1* 基因缺失。在密生波罗花和藏波罗花叶绿体基因组中, *psbA*、*trnH* 基因均在 LSC 区。

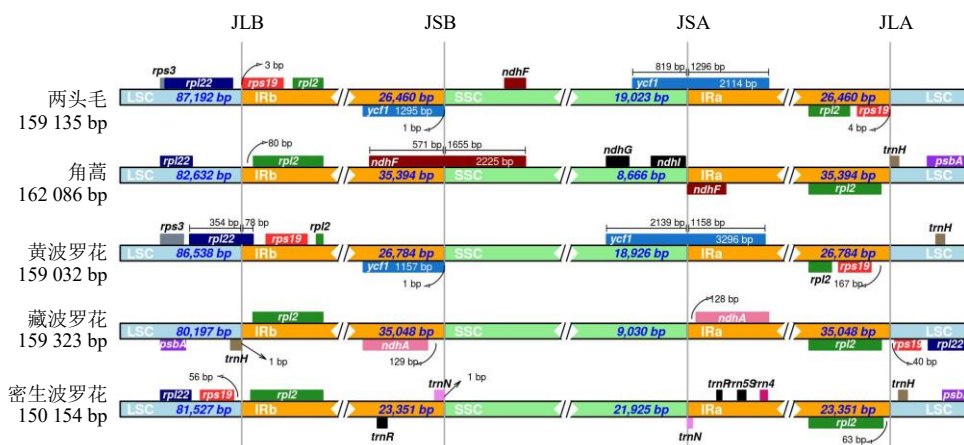


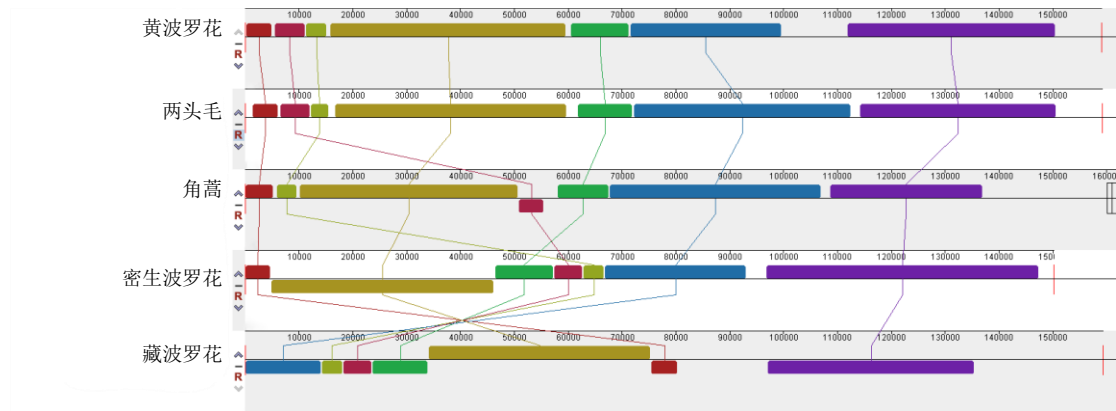
图 3 5 种角蒿属叶绿体基因组的边界分析

Fig. 3 Chloroplast genome boundary analysis of five plants of *Incarvillea*

以角蒿叶绿体基因组作为参考, 对角蒿属内另外 4 种植物叶绿体基因组序列的同源性进行了比较研究。结果显示, 角蒿属叶绿体基因组中非编码区存在明显差异, 变异程度高, 而在基因编码区差异较小, 具有较高的保守性。在叶绿体 4 个部分中, LSC 区的变异程度最高, 而 IRa 区的变异程度最低, 具有较高的保守性。此外, 角蒿属内不同植物的叶绿体基因组中 *atpB*、*atpF*、*atpI*、*psaA*、*rps7*、*rps16*、*rps12*、*ycf1*、*ndhF*、*ndhD*、*ycf2* 等基因的编码区存在显著差异; *atpH*、

petN、*psbM*、*rbcL*、*pasC* 等基因的编码区差异不明显。与其他同属植物相比, 黄波罗花具有较高的序列变异, 且藏波罗花和密生波罗花的序列相似性高。

通过对角蒿属内 5 种药用植物的叶绿体基因组进行共线性分析, 可进一步揭示角蒿属不同物种叶绿体基因组之间的差异。结果表明, 有 7 个局部共线块 (locally collinear block, LCB) 是这 5 种角蒿属药用植物所共有的 (图 4)。同时发现, 除了两头毛和黄波罗花外, 其余 3 个物种的叶绿体基因组均



相同的色块表示不同叶绿体基因组之间的同源区段，并用线条连接。

The same color blocks represented homologous segments between different chloroplast genomes and were connected by lines.

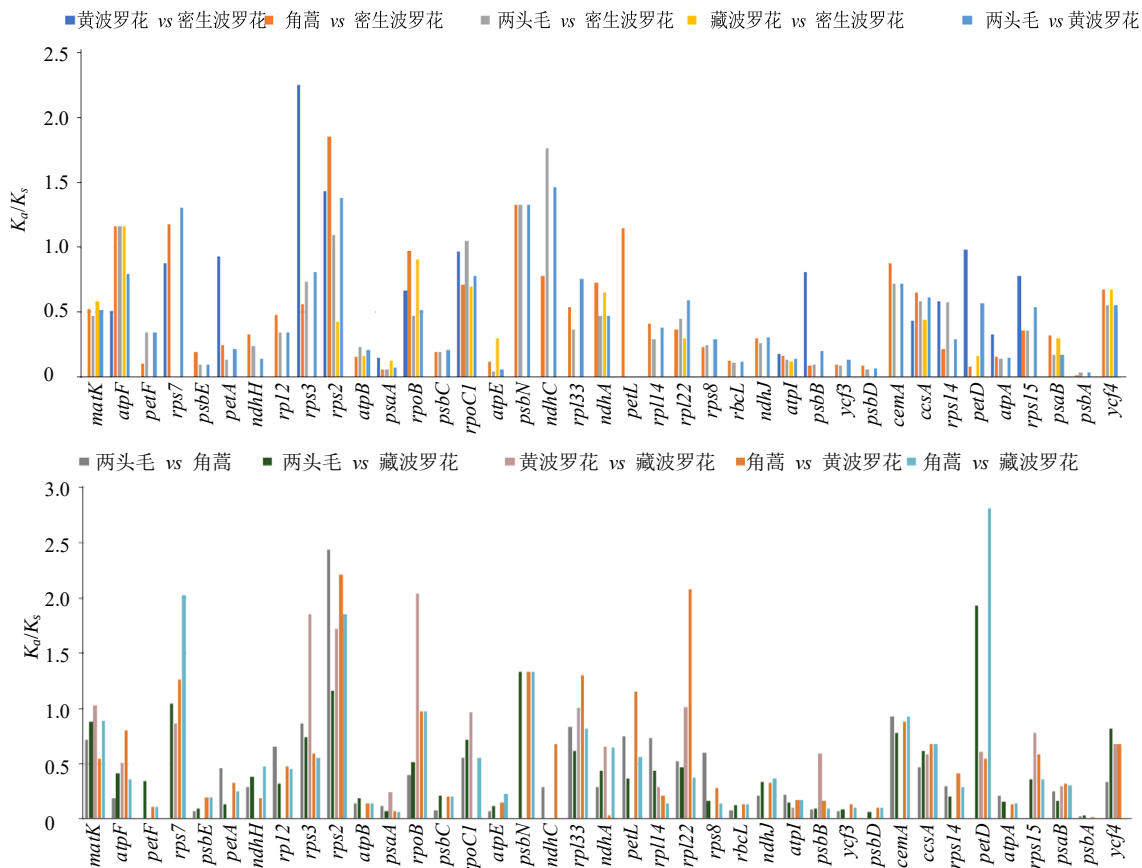
图 4 5 种角蒿属叶绿体基因组共线性分析

Fig. 4 Collinearity analysis of chloroplast genomes of five plants of *Incarvillea*

存在程度不一的基因重排。基因组重排除了可以增加遗传多样性，在促进生物的适应性方面也有积极作用。基因组重排通过改变生物体的基因表达模式和功能，从而导致新的适应性特征。

3.4 选择压力分析

为了研究角蒿属在叶绿体基因组中的进化特征，对密生波罗花叶绿体基因组及其 4 个近缘物种的 55 个共有蛋白编码基因进行了 K_a/K_s 计算（图 5）。图中



图中没有数值的基因 表示其中 K_a 或 K_s 值为 0。

A gene with no value in the figure indicates that the K_a or K_s value is 0.

图 5 角蒿属内部部分共有蛋白编码基因的 K_a/K_s 值

Fig. 5 K_a/K_s values of some common protein-coding genes in *Incarvillea*

没有数值的基因表示其中 K_a/K_s 值为 0。大部分与光系统、遗传系统相关的基因,如 *petA*、*psbA*、*atpA*、*atpB*、*atpE*、*atpI*、*rbcL*、*psbE*、*psbB*、*psbC*、*psbD*、*psbF*、*ndhH*、*ndhJ* 等基因的 K_a/K_s 值均小于 1,表明这些关键基因均受到纯化选择。上述基因在叶绿体基因组行使功能的过程中具有非常重要的作用,因此在进化上比较保守。同时,基因 *rps2* 除了在藏波罗花和密生波罗花的 K_a/K_s 比值小于 1 外,在其余物种对中 K_a/K_s 比值均大于 1,其中在两头毛和角蒿中的 K_a/K_s 值达到了 2.4377。这表明该基因对藏波罗花和密生波罗花的物种分化影响不大,但在促进两头毛和角蒿分化的过程中发挥重要作用。

此外, *rps3*、*rps7*、*psbN*、*petD*、*atpF* 等基因在 5 个物种间的 K_a/K_s 值较高,表明这些基因受到正选择。

3.5 基于叶绿体基因组的系统发育分析

基于 5 种角蒿属药用植物和其他同科物种的叶绿体基因组构建系统发育树(图 6)。结果显示,角蒿属的 5 个物种聚为一个大的分支。波罗花亚属内的 3 个物种(密生波罗花、藏波罗花和黄波罗花)聚为一支,其中,密生波罗花和藏波罗花亲缘关系最近,其次是黄波罗花,密生波罗花与两头毛的亲缘关系最远。同时发现,波罗花亚属和角蒿亚属亲缘关系最近,其次是两头毛亚属。

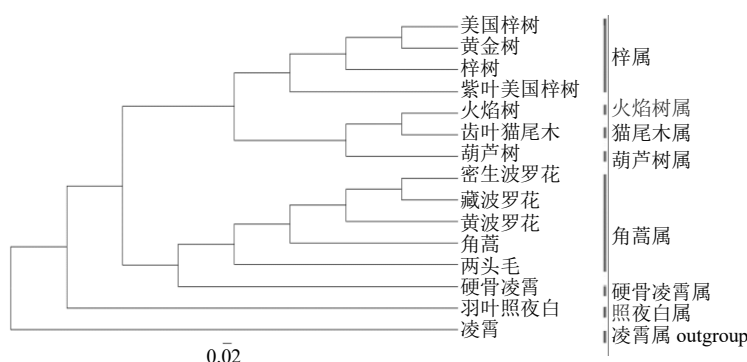


图 6 基于叶绿体基因组序列构建的最大似然 (ML) 系统发育树

Fig. 6 Maximum likelihood (ML) tree constructed based on complete chloroplast genomes

4 讨论

角蒿属植物在全世界约有 15 种,共 4 个亚属,其中分布在我国的角蒿属就有 3 个亚属:角蒿亚属、两头毛亚属和波罗花亚属^[11-12]。它们多集中分布于青藏高原和西南地区,而角蒿属在我国的这种集中的多样性爆发被认为是和早中新世发生青藏高原的隆起有关^[11]。

本研究对角蒿属内 5 种药用植物的叶绿体基因组进行比较分析,5 个物种的叶绿体基因组均为双链环状结构,由 4 部分组成,即 SSC、LSC、IRa 和 IRb。角蒿属内 5 种药用植物的叶绿体基因组长度在 150~162 kb,其中 LSC 长度在 80~87 kb,SSC 长度在 8~21 kb,而 2 个相同的反向重复序列 IRa 和 IRb 的长度范围在 23~35 kb,由此可见角蒿属植物的叶绿体基因组大小、基因组成等方面高度保守,其中造成角蒿属内药用植物叶绿体基因组大小存在差异的主要原因是 SSC 长度不一。对密生波罗花叶绿体基因组功能注释的 76 个基因主要分为 4 大类,与大部分物种的划分一致^[22, 24, 35]。

重复序列在植物的叶绿体基因组中发挥着重要的进化作用。有研究发现它能促进叶绿体基因组发生基因重排,不仅能增加物种的遗传多样性,并且能提高物种的适应性^[38]。角蒿属内 5 种药用植物中,只有两头毛叶绿体基因组中的重复序列类型最多、数量最少,推测该物种与其他同属物种相比,发生基因重排的可能性更小,在进化上可能更为保守。同时 SSRs 分析表明,5 种植物的 SSRs 类型多集中在单核苷酸重复,其中 A/T 重复最多,且多偏向于 T,该现象在桃属^[1]、山楂属^[40]等中也有报道,进一步印证在被子植物中的 SSRs 主要是由 polyA 和 polyT 组成。

角蒿属内 5 种药用植物叶绿体基因组的变异程度存在差异,两头毛在进化上最为保守,而黄波罗花叶绿体基因组序列变异程度最大。叶绿体全基因组变异分析及 IR 区的收缩或扩张分析都表明藏波罗花和密生波罗花叶绿体基因组的序列相似性较高,并且两者的边界基因类型差异不大,可能是因为亲缘关系更近的物种在边界基因类型方面表现出

较高的相似性,这与王孟豪等^[41]的研究结果相似。共线性分析表明两头毛和黄波罗花的叶绿体基因组未发生基因重排,其余 3 种植物叶绿体基因组中基因的排列顺序差异较大。同时比较波罗花亚属内的 3 个物种,发现它们种间基因的排列顺序差异较大,推断这可能与它们分布的不同海拔范围有关。黄波罗花主要分布在海拔 2 000~3 350 m 高山草坡或混交林下;藏波罗花生长在海拔 4 000~5 000 m 的高山沙质草甸及山坡砾石垫状灌丛中;密生波罗花分布的海拔范围在 2 600~4 100 m。与密生波罗花、黄波罗花相比,藏波罗花的生境更干、更寒冷,推测藏波罗花更能适应极端环境,该观点与张亚颖等^[21]的观点一致。而造成波罗花亚属内的 3 种植物叶绿体基因组中基因排列顺序的差异可能是与青藏高原发生隆起后,角蒿属内各物种为适应各自环境而发生适应性变异有关^[42]。

选择压力分析结果表明密生波罗花与 4 个近缘物种的叶绿体基因组中的蛋白编码基因受到的选择不同。大部分与光合作用、ATP 合酶相关的关键基因。例如, *rps2* 基因在棉属^[43]的叶绿体基因组中检测发现也受到了强烈正选择,在植物叶绿体核糖体合成过程中有重要作用,而 *rps2* 基因在拟南芥^[44]、东乡野生稻^[45]中作为抗性基因被广泛研究。而一些与光系统、遗传系统相关的基因,如 *atpA*、*atpB*、*atpE*、*atpI* 均受到纯化选择,该研究结果与刘巧珍等^[34]的结果一致。纯化选择是最普遍的自然选择机制,大部分叶绿体基因的纯化选择可能是维持其适应性状的进化结果^[46]。

系统发育分析的结果表明密生波罗花和藏波罗花的亲缘关系最近,与两头毛的亲缘关系最远。在形态学方面,密生波罗花和藏波罗花都是矮小的宿根草本植物,花萼均为钟状,花色相近,其叶为回羽状复叶,聚生于茎基部。根均为肉质,粗壮。蒴果都是具有明显特征的长披针形,且都具四棱^[11, 16]。陈绍田等^[47]对角蒿属 14 个种的花粉进行扫描电镜观察,发现密生波罗花和藏波罗花的花粉形态、大小及表面纹饰都极为相似,且两者的萌发孔均有 8 沟;在遗传学方面,前期本课题组^[21]基于叶绿体基因组序列估计藏波罗花与同属物种分化时间的研究表明,藏波罗花和密生波罗花的亲缘关系最近,其分化时间可以追溯到 4.66 百万年前。本研究为之后角蒿属内植物的亲缘关系确定、遗传多样性研究提供了理论依据,同时为该属物种鉴定及进一步探索个体水

平的遗传差异分析提供理论基础。

本研究通过对密生波罗花叶绿体基因组进行测序、组装及注释,并结合同属 4 个物种的叶绿体基因组进行比较分析,阐述了其变异程度及系统发育关系。结果显示,密生波罗花叶绿体基因组表型出典型的环状四分体结构,总长度为 150 154 bp, GC 含量为 40%,注释得到了 119 个基因。比较分析结果表明,5 种角蒿属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,其中波罗花亚属的变异程度比其他两个亚属的变异程度高,而两头毛在进化上最为保守;密生波罗花与藏波罗花的亲缘关系最近。以期对角蒿属物种分子鉴定、系统发育以及资源保护等研究提供一定的参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Du Z Y, Lu K, Zhang K, *et al.* The chloroplast genome of *Amygdalus* L. (Rosaceae) reveals the phylogenetic relationship and divergence time [J]. *BMC Genomics*, 2021, 22(1): 645.
- [2] 张悦. 红豆杉科叶绿体基因组结构多样化研究 [D]. 北京: 北京林业大学, 2019.
- [3] 舒军霞, 杨林, 周涛, 等. 刺柏属 4 种药用植物叶绿体基因组密码子偏好性分析 [J]. *中草药*, 2022, 53(23): 7507-7515.
- [4] 姜汶君, 郭梦月, 庞晓慧. 叶绿体基因组在药用植物鉴定及系统进化研究中的应用 [J]. *世界中医药*, 2020, 15(5): 702-708.
- [5] 杨娇. 羌活属叶绿体基因组进化及物种界定研究 [D]. 西安: 西北大学, 2018.
- [6] 张明英, 张雨曲, 李依民, 等. 北柴胡、紫花阔叶柴胡叶绿体全基因组解析及柴胡属药用植物叶绿体基因组比较与系统发育分析 [J]. *药学学报*, 2021, 56(02): 618-629.
- [7] 代江鹏, 蔡一鸣, 刘巧珍, 等. 甘草属 7 种植物叶绿体基因组密码子使用偏好性分析 [J]. *中草药*, 2023, 54(9): 2907-2916.
- [8] 张明英, 王西芳, 高静, 等. 美丽芍药叶绿体全基因组解析及系统发育分析 [J]. *药学学报*, 2020, 55(1): 168-176.
- [9] 董博然, 赵志礼, 倪梁红, 等. 基于叶绿体基因组的长梗秦艽、全萼秦艽物种 DNA 分子标记与鉴定 [J]. *药学学报*, 2021, 56(09): 2584-2591.
- [10] 张明英, 李依民, 程文萍, 等. 基于通用 DNA 条形码序列的黄精属药用植物分子鉴定 [J]. *中草药*, 2023, 54(1): 235-244.
- [11] 朱勇. 西藏藏波罗花的种群遗传多样性及角蒿属植物花色相关基因的分子进化研究 [D]. 上海: 复旦大学, 2010.
- [12] 邹琼宇, 陈德力, 黄园园, 等. 角蒿属植物化学成分及药理活性研究进展 [J]. *中草药*, 2016, 47(3): 499-511.

- [13] 左爱学, 杨艳, 朱萍, 等. 两头毛的化学成分和药理作用研究进展 [J]. 现代药物与临床, 2014, 29(1): 105-109.
- [14] 宋超, 管雪婷, 袁巧云, 等. 密生波罗花中神经酰胺类成分的抗炎作用 [J]. 现代食品科技, 2022, 38(11): 46-54.
- [15] 李伟博, 韩佳欣, 何洋等. 密生波罗花中营养成分及抗氧化活性 [J]. 食品工业, 2021, 42 (07): 342-346.
- [16] 彭晓阳. 黄波罗花化学成分及其生物活性研究 [D]. 上海: 第二军医大学, 2010.
- [17] 何青秀, 仁真旺甲, 王艺等. 藏药“欧曲”的品种考证、化学成分及药理作用研究进展 [J]. 成都中医药大学学报, 2022, 45 (01): 84-91.
- [18] 闫苗苗, 魏光成, 潘效红, 等. 一种适用于动物与植物总 DNA 提取的方法:改良 CTAB 法 [J]. 安徽农业科学, 2008, 36(20): 8488-8558.
- [19] Tvedte E S, Michalski J, Cheng S J, *et al.* Evaluation of a high-throughput, cost-effective Illumina library preparation kit [J]. *Sci Rep*, 2021, 11(1): 15925.
- [20] 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.
- [21] 张亚颖, 焦琬尧, 焦文睿, 等. 藏波罗花叶绿体基因组分析 [J]. 生物工程学报, 2023, 39(7): 2954-2964.
- [22] 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.
- [23] 尚明越, 王嘉乐, 周莹, 等. 濒危紫皮石斛叶绿体基因组结构及系统发育分析 [J]. 中草药, 2023, 54(19): 6424-6433.
- [24] 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.
- [25] Zheng S Y, Poczar P, Hyvönen J, *et al.* Chloroplot: An online program for the versatile plotting of organelle genomes [J]. *Front Genet*, 2020, 11: 576124.
- [26] 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.
- [27] Beier S, Thiel T, Münch T, *et al.* MISA-web: A web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [28] Wang W B, Yu H, Wang J H, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J]. *International Journal Of Molecular Sciences*, 2017, 18(11): 2288.
- [29] Amirouyefi A, Hyvönen J, Poczar P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [30] 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.
- [31] Darling A C E, Mau B, Blattner F R, Perna N T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements [J]. *Genome Research*, 2004, 14(7): 1394-1403.
- [32] 曾淑云. 地黄属植物叶绿体基因组研究 [D]. 西安: 西北大学, 2017.
- [33] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 2009, 25(11): 1451-1452.
- [34] 刘巧珍, 代江鹏, 朱鹏坚, 等. 栽培奇楠沉香叶绿体基因组特征及系统发育分析 [J]. 中国中药杂志, 2023, 48(20): 5531-5539.
- [35] 宋炎峰. 樱亚属 (subg. *Cerasus*) 叶绿体基因组分析及其系统发育研究 [D]. 南京: 南京林业大学, 2022.
- [36] 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.
- [37] 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.
- [38] 张玲. 中国霸王属(蕁藜科)的系统发育及霸王遗传多样性研究 [D]. 杨凌: 西北农林科技大学, 2022.
- [39] 尚明越, 王嘉乐, 周莹, 等. 濒危紫皮石斛叶绿体基因组结构及系统发育分析 [J]. 中草药, 2023, 54(19): 6424-6433.
- [40] Wu L W, Cui Y X, Wang Q, *et al.* Identification and phylogenetic analysis of five *Crataegus* species (Rosaceae) based on complete chloroplast genomes [J]. *Planta*, 2021, 254(1): 14.
- [41] 王孟豪, 冉航, 刘艳艳, 等. 槲木属及其近缘类群叶绿体基因组比较与系统发育分析 [J]. 植物科学学报, 2023, 41(2): 149-158.
- [42] Rana S K, Luo D, Rana H K, *et al.* Geo-climatic factors influence the population genetic connectivity of *Incarvillea arguta* (Bignoniaceae) in the Himalaya Hengduan Mountains biodiversity hotspot [J]. *J Systemat Evolution*, 2021, 59(1): 151-168.
- [43] 王宁. 棉属叶绿体基因组核苷酸进化和遗传关系研究 [D]. 西安: 西北大学, 2021.
- [44] Mindrinos M, Katagiri F, Yu G L, *et al.* The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats [J]. *Cell*, 1994, 78(6): 1089-1099.
- [45] 崔玲玲, 李磊, 韩小霞, 等. 东乡野生稻 *RPS2* 类候选抗病基因的克隆与分析 [J]. 湖南农业大学学报: 自然科学版, 2009, 35(2): 111-115.
- [46] Ren T, Li Z X, Xie D F, *et al.* Plastomes of eight *Ligusticum* species: Characterization, genome evolution, and phylogenetic relationships [J]. *BMC Plant Biol*, 2020, 20(1): 519.
- [47] 陈绍田, 管开云, 周浙昆, 等. 角蒿属(紫葳科)的花粉形态 [J]. 云南植物研究, 2003, 25(4): 457-464.