

## • 药材与资源 •

## 药用植物蒺藜叶绿体基因组特征与系统进化分析

胡赛文<sup>1</sup>, 丁怡宁<sup>1</sup>, 毕光耀<sup>1</sup>, 李翠翠<sup>1</sup>, 苏 春<sup>2</sup>, 窦利军<sup>3</sup>, 李贺敏<sup>1</sup>, 夏 至<sup>1\*</sup>

1. 河南农业大学农学院, 河南 郑州 450002

2. 西北农林科技大学生命科学院, 陕西 杨凌 712100

3. 河南省信阳生态环境监测中心, 河南 信阳 464000

**摘要:** 目的 以药用植物蒺藜 *Polygonum aviculare* 为材料, 利用高通量技术测定基因组 DNA 序列, 对叶绿体基因组进行组装和序列分析, 为进一步开展药用植物蒺藜的群体遗传学和遗传多样性研究奠定基础。方法 利用华大 MGISEQ-2000PE150 测序平台, 双末端测序策略对其全基因组 DNA 建库测序, 测定蒺藜的基因组 DNA 序列, 用 NOVO Plasty 组装叶绿体基因组, 采用最大似然法 (maximum likelihood, ML) 构建系统进化树。结果 蒺藜叶绿体基因组全长为 163 461 bp, GC 值 37.5%, 具 1 个典型的四分区域结构, 包括 1 个大单拷贝区 (large single copy region, LSC)、1 对反向重复区 (inverted repeats, IR) 和 1 个短单拷贝区 (small single copy, SSC), 序列长度分别为 88 023、31 066、13 306 bp。蒺藜的叶绿体基因组共有 130 个基因, 其中编码蛋白基因、rRNA 基因与 tRNA 基因的数量分别为 83、8、37 个。系统进化分析表明, 蒺藜与木蓼属的额河木蓼构成一单系分支, 具有 100% 支持率。结论 蒺藜隶属于蓼科的蓼属, 与木蓼属的额河木蓼亲缘关系较近。蒺藜叶绿体基因组的组装、序列特征、间隔区的筛选及系统进化分析为后续开展群体遗传学和遗传多样性研究奠定基础。

**关键词:** 蒺藜; 蓼属; 叶绿体全基因组; 组装; 系统发育

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2022)09-2776-10

DOI: 10.7501/j.issn.0253-2670.2022.09.021

## Characterization and phylogenetic analysis of complete chloroplast genome of medicinal plant *Polygonum aviculare*

HU Sai-wen<sup>1</sup>, DING Yi-ning<sup>1</sup>, BI Guang-yao<sup>1</sup>, LI Cui-cui<sup>1</sup>, SU Chun<sup>2</sup>, DOU Li-jun<sup>3</sup>, LI He-min<sup>1</sup>, XIA Zhi<sup>1</sup>

1. College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China

2. College of Life Sciences, Northwest A&F University, Yangling 712100, China

3. Xinyang Environmental Monitoring Center of Henan Province, Xinyang 464000, China

**Abstract: Objective** The complete chloroplast genome of medicinal plant *Polygonum aviculare* sequenced by high-throughput technologies was assembled for the sequence analysis to provide evidence for its population genetics and diversity studies. **Methods** DNA library was constructed for *P. aviculare* with the paired-end strategy, and MGISEQ-2000PE150 was used to sequence DNA of *P. aviculare* in Beijing Genomics Institute (China). The complete chloroplast genome was assembled using NOVO Plasty software, and sequence analysis was performed based on gene annotation results. Phylogenetic analyses were performed using Maximum-Likelihood (ML) methods. **Results** The complete chloroplast genome of *P. aviculare* was 163 461 bp in length with a GC content of 37.5%. The chloroplast genome exhibited a typical quadripartite structure, including a large single copy region (LSC), a pair of inverted repeats (IR), and a small single copy (SSC), and the sequence lengths were 88 023, 31 066 and 13 306 bp. The chloroplast genome harbored 130 genes, including 83 protein-coding genes, eight rRNA genes, and 37 tRNA genes. Phylogenetic analyses result indicated that *P. aviculare* was sister to *Atraphaxis irtyschensis* with bootstrap value of 100%. **Conclusion** Our result verified that *P. aviculare* belonged to the Polygonaceae, and it was closely related to *A. irtyschensis*. Sequence assembly, sequence features, intergenic region screening and phylogenetic analysis of medicinal plant *P. aviculare* chloroplast genome provide a basis for its future studies on both population genetics and genetic diversity.

**Key words:** *Polygonum aviculare* L.; *Polygonum* L.; chloroplast genome; assembly; phylogenetic analysis

收稿日期: 2021-10-09

基金项目: 国家自然科学基金面上项目 (31770370); 河南省高等学校重点科研项目计划 (18A360006)

作者简介: 胡赛文 (1992—), 女, 硕士, 研究方向为中药资源的分子鉴定。

\*通信作者: 夏至, 教授, 主要从事中药资源的分子鉴定及中药资源可持续利用研究。E-mail: xiazhiemail@126.com

药用植物蒺藜 *Polygonum aviculare* L. 隶属于蓼科 (Polygonaceae) 蓼属 *Polygonum* L., 为一年生或多年生草本, 主要分布在我国河南、浙江、山东、河北、吉林等地<sup>[1-2]</sup>。蒺藜始载于《神农本草经》, 全草皆可入药<sup>[3]</sup>, 临床多用于治疗泌尿系统感染、肾炎、细菌性痢疾、非胰岛素依赖性糖尿病、结石等疾病<sup>[4-6]</sup>。《中国药典》2020 年版规定蒺藜有效成分为杨梅苷(含量不得少于 0.030%)<sup>[3]</sup>。此外, 中药材蒺藜还含有丰富的化学成分, 主要包括黄酮类、酚酸类、苯丙素类等<sup>[7]</sup>。现代药理学研究表明蒺藜具有利尿、降压、抑菌、抗氧化<sup>[8]</sup>和保肝等诸多的药理作用, 临床应用十分广泛。目前, 有关药用植物蒺藜的研究主要集中在化学成分分析和药理学研究等方面, 该植物叶绿体基因组组装、基因组特征和系统进化分析等未见报道。

叶绿体不仅是绿色植物光合作用的重要场所、也是绿色植物最重要的细胞器之一。它利用太阳能和二氧化碳转换成化学能与氧气, 维持地球生态平衡<sup>[9]</sup>、参与环境信号的响应, 在逆境响应中也起着重要作用<sup>[10]</sup>。叶绿体具有一整套用于光合作用、能量代谢、蛋白质合成及氮、硫同化相关的基因, 分布在大小为 120~180 kb 的环状基因组上, 具有结构保守, 母性遗传等特点<sup>[11-12]</sup>。叶绿体基因组一般为闭环双链 DNA 结构, 陆生植物的叶绿体基因组结构通常由 1 个大单拷贝区域 (large single copy, LSC)、1 个短单拷贝区域 (small single copy, SSC) 和 2 个反向重复区域 (inverted repeat, IR) 组成<sup>[13]</sup>。叶绿体基因组拥有相对独立的基因组和遗传序列, 并且不像核基因组一般有着复杂的重复序列, 其基因序列保守, 间隔区变异位点丰富, 适宜的进化速率<sup>[14]</sup>能够为植物不同等级的亲缘关系, 系统进化关系及遗传多样性研究提供较为可靠的信息<sup>[15]</sup>。

近年来, 随着测序技术的不断改进, 测序平台的不断升级, 一系列组装和注释软件如 plasmidSPAdes<sup>[16]</sup>、NOVOPlasty<sup>[17]</sup>、GetOrganelle<sup>[18]</sup>等的开发与更新, 多种重要的药用植物叶绿体基因组已完成测序和分析, 如丹参 *Salvia miltiorrhiza* Bunge<sup>[19]</sup>、人参 *Panax ginseng* C. A. Mey.<sup>[20]</sup>、三七 *P. notoginseng* (Burkill) F. H. Chen ex C. Chow & W. G. Huang<sup>[21]</sup>、粗茎秦艽 *Gentiana crassicaulis* Duthie ex Burk.<sup>[22]</sup>、曼陀罗 *Datura stramonium* L.<sup>[23]</sup>、五味子 *Schisandra chinensis* (Turcz.) Baill.<sup>[24]</sup>、铁皮石斛 *Dendrobium officinale* Kimura et Migo<sup>[25]</sup>和三叶崖爬藤

*Tetrastigma hemsleyanum* Diels & Gilg<sup>[26]</sup>等。本研究以药用植物蒺藜为材料, 利用高通量测序方法测定蒺藜基因组 DNA 序列, 并对该植物的叶绿体基因组进行组装和注释。进而分析蒺藜叶绿体基因组序列特征, IR 边界特征, 间隔区信息位点的特征, 并对蒺藜及其近缘 19 种植物的叶绿体基因组序列进行系统发育分析, 验证其在科级系统发育中的位置, 为药用植物蒺藜的种质资源的鉴定、开发和利用提供一定的理论依据。

## 1 材料与方法

### 1.1 材料

蒺藜新鲜叶片采集于河南省新乡市原阳县小吴庄村河南农业大学科教园区 (35°6'32"N, 113°56'34"E), 由河南农业大学生命科学院李家美教授和农学院中药材系夏至教授鉴定为蓼科植物蒺藜 *P. aviculare* L., 凭证标本保存于河南农业大学标本馆, 凭证标本号为 XZ-2020-11。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 冰箱备用。蒺藜及其近缘物种叶绿体基因组序列来源于 NCBI 数据库, 实验材料详细信息见表 1。

表 1 植物样品来源

Table 1 Plant samples used in this study

物种	来源	GenBank 登录号
<i>Afrobrunni chiaerecta</i> (Asch.) Hutch.	NCBI	MH286316
额河木蓼 <i>Atraphaxis irtyschensis</i> Yang et Han	NCBI	MG878984
吉木乃沙拐枣 <i>Calligonum jeminaicum</i> Z. M. Mao	NCBI	MN202608
泡果沙拐枣 <i>C. junceum</i> Flour	NCBI	MK854997
金荞麦 <i>Fagopyrum dibotrys</i> (D. Don) Hara	NCBI	MF491390
荞麦 <i>Fagopyrum esculentum</i> Moench.	NCBI	NC 010776
何首乌 <i>Fallopia multiflora</i> Thunb.	NCBI	MK330002
<i>Fallopia sachalinensis</i> (F. Schmidt) Ronse	NCBI	NC 047446
<i>Muehlenbeckia australis</i> (Forst. f.) Caused	NCBI	MG604297
山蓼 <i>Oxyria digyna</i> (L.) Hill	NCBI	MN564931
中华山蓼 <i>O. sinensis</i> Hemsl.	NCBI	KX774248
蒺藜 <i>Polygonum aviculare</i> Linn. var. <i>vegetum</i> Ledeb.	河南 原阳	MW044669
虎杖 <i>Reynoutria japonica</i> Houtt.	NCBI	MT301955
波叶大黄 <i>Rheum franzenbachii</i> Munt.	NCBI	MN564923
掌叶大黄 <i>Rheum palmatum</i> L.	NCBI	NC 027728
小大黄 <i>Rheum mpumilum</i> L.	NCBI	MN564927
皱叶酸模 <i>Rumex crispus</i> L.	NCBI	MN055629
羊蹄 <i>Rumex japonicus</i> Houtt.	NCBI	MN720269
<i>Symmeria paniculate</i> Benth.	NCBI	MH286353
蓝花丹 <i>Plumbago auriculata</i> Lam.	NCBI	MH286308

## 1.2 DNA 的提取和高通量测序

采用北京天根生化植物 DNA 提取试剂盒 (Tiangen Biotech Co., 中国) 提取样品篇蓄新鲜叶片的总 DNA, 利用 1% 的琼脂糖凝胶电泳检测 DNA 完整性。样品送至华大生物科技公司 (北京) 后, 使用 NanoDrop2000 微量分光光度计 (Thermo Scientific, 美国) 检测总 DNA 的纯度和浓度。MGISEQ-2000 PE150 测序平台测序, 测序完成后, 利用华大自主开发的过滤软件 SOAPnuke 过滤参数, 过滤步骤为 (1) 过滤接头: 测序 read 匹配上 adapter 序列的 25% 或者以上则删除整条 read; (2) 过滤低质量数据: 如果测序 read 中质量值低于 20 的碱基占整条 read 的 30% 或者以上则删除整条 read; (3) 去 N: 如果测序 read 中 N 含量占整条 read 的 1% 或者以上, 则删除整条 read。 (4) 获得 Clean reads。数据以 FASTQ 格式储存, 用于后续的拼接和注释。

## 1.3 叶绿体基因组的拼接和注释

叶绿体基因组的拼接采用 NOVOPlasty<sup>[17]</sup> 程序, 插入片段大小设为 150 bp。过滤后的 reads 用 Geneious 11.0.3<sup>[27]</sup> 拼接软件组装成重叠群, 并对组装中的简并碱基, 进行人工修正。利用 Geneq-Annotation of Organellar (<https://chlorobox.mpimp.golm.mpg.de/geseq.html>), 结合 NCBI 上已报道的何首乌 (GenBank 登录号: MK330002) 注释结果对篇蓄叶绿体全基因组进行基因注释, 参数为默认值, 最后进行手动调整。tRNA 用 ANAGORNV1.2.38 (<https://chlorobox.mpimp.golm.mpg.de/geseq.html>) 预测。注释完成后, 使用 Geneious11.0.3<sup>[27]</sup> 生成 GenBank 格式的文件, 并提交到 GeneBank (<https://www.ncbi.nlm.nih.gov/genbank/>), 登录号为 MW044669。利用在线工具 OGDRAW-DRAW Organelle Genome Maps (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 绘制叶绿体结构图。

## 1.4 IR 边界的收缩和扩张分析

IR 区域在叶绿体基因中具有高度保守性, 但是 IR/SC 边界区域仍存在细微的变化, IR 边界的膨胀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<sup>[28]</sup>。在植物进化过程中, IR/SC 边界不同程度的扩张和收缩导致了边界和基因组长度的多样性<sup>[29]</sup>。本研究使用 Geneious 11.0.3<sup>[27]</sup> 软件获得蓼科 7 个属 7 种植物叶绿体基因组的 IRA/IRB、LSC 和 SSC 和边界基因的序列长度, 进行比较分

析, 探讨蓼科植物叶绿体基因组 IR 边界的收缩和扩张特征。使用 Adobe illustrator 软件绘制 7 种蓼科植物叶绿体基因组 IR 边界对比图。

## 1.5 叶绿体基因组基因间隔区信息位点分析

相比叶绿体基因编码区, 叶绿体基因间隔区在近缘物种间往往具有更高的变异位点, 通常被用来构建属间、属内种间物种系统进化发育关系。本研究基于蓼科已注释的 11 个属 17 种植物的叶绿体基因组序列特征, 利用 phylosuite v1.2.1<sup>[30]</sup> 提取 17 个物种叶绿体基因组 32 个共有的间隔区, 利用 MAFFT<sup>[31]</sup> 进行多重比对。统计这些间隔区的信息位点百分率, 为下一步构建蓼科属级以下物种系统进化关系提供分子标记。

## 1.6 系统发育分析

从 NCBI 数据库选取已公开发布的 19 个蓼科物种叶绿体基因组数据 (表 1), 同时以近缘的白花丹科 (Plumbaginaceae) 蓝花丹 *Plumbago auriculata* L. 为外类群, 利用 phylosuite v1.2.1<sup>[30]</sup> 软件基于 MAFFT<sup>[31]</sup> 进行多重比对。系统发育分析采用最大似然法 (maximum likelihood, ML), 利用 CIPRES Science Gateway 服务器 (<http://www.phylo.org/>) 中 RaxML-HPC2 7.6.3 软件 (Stamatakis 2006) 构建系统发育树。利用 Bootstrap (BS) (1000 次重复) 检验各分支的支持率。系统发育树导出后利用 FigTree version1.4.2 查看。

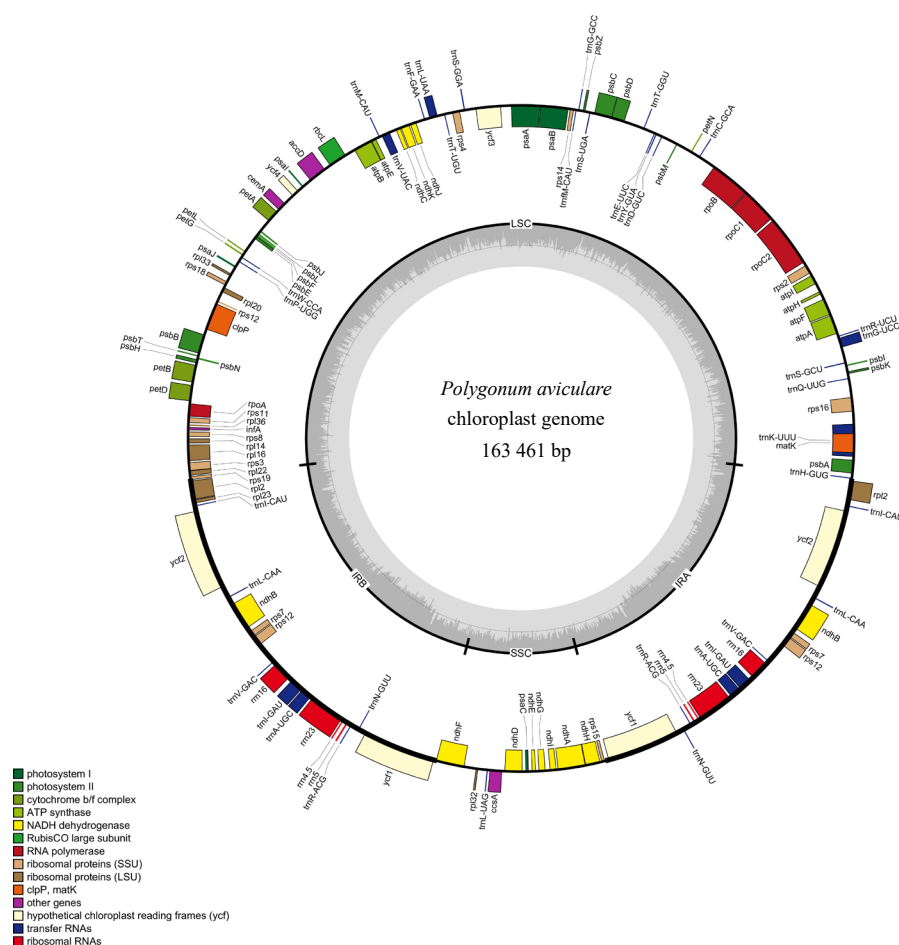
## 2 结果与分析

### 2.1 篇蓄叶绿体基因组的结构

测序结果去除接头和低质量的数据后, 共获得 2.32 Gb 原始数据, 原始 reads 为 15 471 394 条, Q20 达 97.86%。组装和注释后得到篇蓄的叶绿体基因组。结果表明, 篇蓄为共价闭合的双链环状分子 (图 1), 全长为 163 461 bp, GC 值 35.6%, 是一个典型的 4 分区域结构, 包括 1 个 LSC、1 对 IR 和 1 个 SSC, 它们的长度分别为 88 023、31 066、13 306 bp。IR、LSC 和 SSC 区域的 GC 值存在一定的差异, 其中, IR 区域的 GC 值最高 (41.3%), LSC (35.5%) 和 SSC (32.5%) 均较低, 详细信息见表 2。

### 2.2 篇蓄叶绿体基因组的组成和特点

篇蓄叶绿体基因组共包括 128 个基因, 非重复基因 113 个, 其中 83 个编码蛋白基因、8 个 rRNA 基因与 37 个 tRNA 基因。LSC 区包含的基因最多, 包括 60 个蛋白编码基因和 22 个 tRNA 基因; SSC 区包括 1 个 tRNA 基因 (*trnL-UAG*) 和 11 个蛋白



LSC 和 SSC: 大单拷贝区域、小单拷贝区域; IRA 和 IRB: 2 个反向重复区域; 内圈深色部分: GC 含量

LSC and SSC: large single copy and small single copy; IRA and IRB: two inverted repeats; the darker gray area in inner circle: GC content

图 1 蒺藜叶绿体基因组

Fig. 1 Map of *P. aviculare* chloroplast genome

表 2 蒺藜叶绿体基因组碱基组成及特征

Table 2 Base composition and characteristic of chloroplast genome in *P. aviculare*

区域	A/%	T/%	C/%	G/%	GC/%	蛋白编码基因数	tRNA 基因数	rRNA 基因数	基因总数	长度/bp
LSC	31.6	32.9	18.3	17.3	35.5	60	22	0	82	88 023
SSC	35.6	32.0	17.0	15.4	32.5	11	1	0	12	13 306
IRA/IRB	30.1	28.7	19.7	21.6	41.3	12	14	8	34	31 066
总计	31.1	31.5	19.1	18.4	37.5	83	37	8	128	163 461

因 *rps19*、*ndhF* 基因分别横跨 LSC/IRb、SSC/IRb 边界, 故未统计在表 2 内

the *rps19* and *ndhF* genes across the boundaries of LSC/IRb and SSC/IRb, and they were not listed in table 2

编码基因; 所有的 rRNA 基因、14 个 tRNA 基因和 12 个蛋白编码基因在 IR 区(表 2)。其中, 蛋白质编码基因中与自我复制相关基因除 rRNA 基因和 tRNA 基因外(表 3), 还包括 14 个核糖体小亚基基因、10 个核糖体大亚基基因和 4 个 RNA 聚合酶亚基基因; 光合作用相关的基因有 46 个, 包括 12 个 NADH 脱氢酶基因、5 个光合系统 I 基因、15 个光合系统 II 基因、6 个细胞色素复合体编码基因、6 个 ATP 合酶基因、1 个二磷酸核酮糖羧化酶大亚

基基因和 1 个依赖 ATP 的蛋白酶单元 *p* 基因; 此外还有 5 个其他功能基因及 6 个未知功能基因。在 tRNA 中 *trnI-CAU*、*trnL-CAA*、*trnV-GAC*、*trnI-GAU*、*trnA-UGC*、*trnR-ACG* 和 *trnN-GUU* 各有 2 个拷贝; 4 个核糖体 RNA 均有 2 个拷贝, 分别位于反向重复区 IRA 和 IRB。核糖体蛋白大小亚基编码的基因中, *rps7*、*rps12*、和 *rpl2* 这 3 个基因均有 2 个拷贝, 其余为 1 个拷贝。NADH 脱氢酶亚基中的 *ndhB* 基因及未知功能蛋白基因 *ycf2* 和 *ycf1* 的

表 3 蒺藜叶绿体基因组编码的基因

Table 3 Encoded genes present in chloroplast genome of *P. aviculare*

基因分组	基因名称	数量
tRNA	<i>trnH-GUG</i> 、 <i>trnK-UUU</i> <sup>a</sup> 、 <i>trnQ-UUG</i> 、 <i>trnS-GCU</i> 、 <i>trnG-UCC</i> <sup>a</sup> 、 <i>trnR-UCU</i> 、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnY-GUA</i> 、 <i>trnE-UUC</i> 、 <i>trnT-GGU</i> 、 <i>trnS-UGA</i> 、 <i>trnG-GCC</i> 、 <i>trnM-CAU</i> 、 <i>trnS-GGA</i> 、 <i>trnT-UGU</i> 、 <i>trnL-UAA</i> <sup>a</sup> 、 <i>trnF-GAA</i> 、 <i>trnV-UAC</i> <sup>a</sup> 、 <i>trnM-CAU</i> 、 <i>trnW-CCA</i> 、 <i>trnP-UGG</i> 、 <i>trnI-CAU</i> <sup>c</sup> 、 <i>trnL-CAA</i> <sup>c</sup> 、 <i>trnV-GAC</i> <sup>c</sup> 、 <i>trnI-GAU</i> <sup>ac</sup> 、 <i>trnA-UGC</i> <sup>ac</sup> 、 <i>trnR-ACG</i> <sup>c</sup> 、 <i>trnN-GUU</i> <sup>c</sup> 、 <i>trnL-UAG</i>	37
rRNA	<i>rrn16</i> <sup>c</sup> 、 <i>rrn23</i> <sup>c</sup> 、 <i>rrn4.5</i> <sup>c</sup> 、 <i>rrn5</i> <sup>c</sup>	8
核糖体蛋白小亚基	<i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> <sup>c</sup> 、 <i>rps8</i> 、 <i>rps11</i> 、 <i>rps12</i> <sup>ac</sup> 、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps16</i> <sup>a</sup> 、 <i>rps18</i> 、 <i>rps19</i>	14
核糖体蛋白大亚基	<i>rpl2</i> <sup>ac</sup> 、 <i>rpl14</i> 、 <i>rpl16</i> <sup>a</sup> 、 <i>rpl20</i> 、 <i>rpl22</i> 、 <i>rpl23</i> 、 <i>rpl32</i> 、 <i>rpl33</i> 、 <i>rpl36</i>	10
RNA 聚合酶	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> <sup>a</sup> 、 <i>rpoC2</i>	4
NADH 脱氢酶亚基	<i>ndhA</i> <sup>a</sup> 、 <i>ndhB</i> <sup>ac</sup> 、 <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>	12
光系统 I 亚基	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaI</i> 、 <i>psaJ</i>	5
光系统 II 亚基	<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>	15
细胞色素 b/f 复合物亚基	<i>petA</i> 、 <i>petB</i> 、 <i>petD</i> 、 <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>	6
ATP 合成酶亚基	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> <sup>a</sup> 、 <i>atpH</i> 、 <i>atpI</i>	6
Rubisco 大亚基	<i>rbcL</i>	1
蛋白酶	<i>clpP</i> <sup>b</sup>	1
成熟酶	<i>matK</i>	1
包裹膜蛋白	<i>cemA</i>	1
乙酰 CoA 羧化酶亚基	<i>accD</i>	1
细胞色素 C 合成酶	<i>ccsA</i>	1
翻译起始因子	<i>infA</i>	1
未知功能蛋白	<i>Ycf1</i> <sup>c</sup> 、 <i>ycf2</i> <sup>c</sup> 、 <i>ycf3</i> <sup>b</sup> 、 <i>ycf4</i>	6
总计		130

a 和 b 分别表示含有 1 个和 2 个内含子；c 表示含有 2 个拷贝基因

a represent one intron, b represent two introns; c represents two copies of the gene

拷贝数均为 2。

内含子在基因表达调控中发挥重要作用，蒺藜叶绿体基因组中有 18 个基因有内含子。其中，*trnK-UUU*、*rps16*、*trnG-UCC*、*atpF*、*trnL-UAA*、*trnV-UAC*、*petB*、*petD*、*rpl16*、*rpl2*、*ndhB*、*trnI-GAU*、*trnA-UGC*、*ndhA* 和 *rpoC1* 各有 1 个内含子，而 *rps12*、*clpP*、*ycf3* 具 2 个内含子。*matK* 基因位于 *trnK-UUU* 基因内，整个编码区为 *trnK-UUU* 内含子的一部分，存在序列共用现象；*psbD* 基因的 3'端与 *psbC* 基因的 5'端，*atpE* 基因的 3'端与 *atpB* 基因的 5'端、*rps3* 的 3'端与 *rpl22* 的 5'端重叠。

### 2.3 1R 边界的特征

蓼科 7 个属 7 个物种叶绿体全基因组的 IR-LSC 和 IR-SSC 边界比较表明，蓼科各属物种边界具有高度的保守性，其 LSC/IRb 边界 (JLB)，SSC/IRb 边界 (JSB)、SSC/IRa 边界 (JSA) 和 LSC/IRa 边界 (JLA) 的侧翼基因完全相同，但扩张程度存在一定的差异。蓼科 7 个物种在 JLB 边界全部位于 *rps19* 基因内，中华山蓼、掌叶大黄和何首乌中 LSC 和 IRb 区向 *rps19* 基因扩张的程度相同，均为 171、108 bp。泡果沙拐枣、荞麦、蒺藜和羊蹄的 LSC 和 IRb 区向 *rps19*

的扩张程度变异较小，分别为 174 bp 和 108 bp、176 bp 和 103 bp、172 bp 和 107 bp、169 bp 和 110 bp。蒺藜的 JLB 边界扩张范围，位于 7 个种之内。JSB 边界扩张范围显示，中华山蓼、掌叶大黄、泡果沙拐枣、蒺藜与羊蹄的边界非常相似，均位于 *ndhF* 基因内部，*ndhF* 基因在 IRb 和 SSC 区的扩张变异范围，分别为 95 bp 和 2134 bp，95 bp 和 2140 bp，19 bp 和 2225 bp，64 bp 和 2186 bp，62 bp 和 2182 bp。蒺藜的 JLB 边界扩张范围，位于 7 个种之内。但何首乌和荞麦的 *ndhF* 基因已全部扩张到 SSC 区内部，距 JSB 边界分别为 53 bp 和 196 bp。JSA 边界的扩张范围显示，其侧翼基因完全相同，均为 *rps15* 基因（位于 SSC 区）和 *ycf1* 基因（位于 IRa 区）。其中 7 个物种的 *rps15* 基因和 *ycf1* 基因距 JSA 边界的碱基长度范围变异较大，分别 9~267 bp 和 0~282 bp。蒺藜的 JSA 边界扩张范围，位于 7 个种之内。JLA 边界扩张范围显示，其侧翼基因完全相同，均为 *rpl2*（位于 IRa 区）和 *trnH*（位于 LSC 区）2 个基因。其中 7 个物种的 *rpl2* 和 *trnH* 2 个基因距 JLA 边界碱基长度范围变异不大，分别 163~207 bp 和 2~63 bp。蒺藜的 JLA 边界扩张范围，位于 7 个种之内。蒺藜及 6 个蓼科植物的叶绿体全基因组边界图见图 2。



图 2 篇蓄及 6 个蓼科植物的叶绿体全基因组边界图

Fig. 2 Border regions of complete chloroplast genomes of *P. aviculare* and six Polygonaceae plants

## 2.4 基因组特征比较分析

由于虎杖和额河木蓼的叶绿体基因组注释不完全，除了叶绿体基因组全长比较分析后，IR 区、SSC 区和 LSC 区序列长度未做分析。叶绿体基因组特征比较分析见表 4。篇蓄及蓼科 18 种近缘植物叶

绿体基因组的序列长度范围介于 158 851~170 974 bp，其中皱叶酸模的叶绿体全基因组的长度最短（158 851 bp），*Afrobrunnichia erecta* 的叶绿体基因组最长（170 974 bp）。篇蓄的叶绿体基因组长度为 163 461 bp，介于蓼科 19 个物种的叶绿体基因组长度

表 4 19 种植物叶绿体基因组的特征

Table 4 Chloroplast genome characteristics of 19 plants

物种	叶绿体基因组长度/bp	GC 值/%	LSC 长度/bp	SSC 长度/bp	IR 长度/bp
<i>Afrobrunnichia erecta</i>	170 974	37.1	88 058	13 654	34 631
额河木蓼	164 192	37.5	—	—	—
吉木乃沙拐枣	162 535	37.5	88 160	13 319	30 528
泡果沙拐枣	162 459	37.5	88 107	13 416	30 468
金荞麦	159 919	37.9	85 134	13 309	30 738
荞麦	159 599	38.0	84 888	13 875	30 412
何首乌	163 773	37.5	88 340	13 881	30 776
<i>Fallopia sachalinensis</i>	163 485	37.5	87 703	13 566	31 108
<i>Muehlenbeckia australis</i>	163 484	37.4	88 167	13 485	30 916
山蓼	160 698	37.5	85 748	13 172	30 889
中华山蓼	160 404	37.5	85 501	13 133	30 885
篇蓄	163 461	37.5	88 023	13 306	31 066
虎杖	163 410	37.5	—	—	—
波叶大黄	161 688	37.4	86 946	12 784	30 979
掌叶大黄	161 541	37.3	86 518	13 111	30 956
小大黄	161 749	37.3	86 997	12 812	30 970
皱叶酸模	158 851	37.6	85 039	12 657	30 399
羊蹄	159 292	37.5	85 028	13 006	30 629
<i>Symmeria paniculata</i>	162 501	38.3	86 920	13 435	31 073

“—” 代表 GenBank 注释缺失，不完全

“—”represent GenBank annotation missing

范围之内。篇蓄及蓼科 18 种近缘植物叶绿体基因组 GC 含量的范围为 37.1%~38.3%，其中 *Afrobrunnichia erecta* 的 GC 含量最低 (37.1%)，*Symmeria paniculata* 的 GC 含量最高 (38.3%)。篇蓄叶绿体基因组的 GC 含量为 37.5%，介于蓼科 19 个物种叶绿体基因组 GC 含量之内。利用 Geneious 11.0.3 分别得到篇蓄及蓼科 18 种近缘植物的 IR 区、LSC 区和 SSC 区序列，结果表明，IR 区的序列长度范围介于 30 399~34 631 bp，皱叶酸模的 IR 区最短 (30 399 bp)，*Afrobrunnichia erecta* 的 IR 区最长 (34 631 bp)。篇蓄 IR 区长度为 31 066 bp，介于蓼科 19 个物种 IR 区长度范围之内。LSC 区的序列长度范围介于 84 888~88 340 bp，其中荞麦 LSC 区的序列长度最短 (84 888 bp)，何首乌 LSC 区的序列长度最长 (88 340 bp)。篇蓄的 LSC 区的序列长度为 88 023 bp，介于蓼科 19 个物种 LSC 区长度范围之内。SSC 区的序列长度范围介

于 12 657~13 881 bp。其中，皱叶酸模的 SSC 最短 (12 657 bp)，何首乌最长为 (13 881 bp)。篇蓄的 SSC 区的序列长度为 13 306 bp，介于蓼科 19 个物种 SSC 区长度范围之内。

## 2.5 叶绿体基因组基因间隔区信息位点分析

基于蓼科 11 个属 17 种植物 (虎杖和 *Atraphaxis irtyschensis* 未统计在内) 的 32 个叶绿体基因组间隔区信息序列特征统计表明 (表 5)，在 32 个叶绿体基因间隔区中，变异位点百分率变化范围为 2.2%~23.4%，最高的为 psbK-psbI 基因间隔区，其变异位点百分率为 23.4%。超过 20% 有 8 个，分别为 ndhE-ndhG、psaJ-rpl33、psbK-psbI、rpl33-rps18、rps16-trnQ-UUG、trnE-UUC-trnT-GGU、trnF-GAA-ndhJ、trnT-GGU-psbD。这些变异位点百分率较高的叶绿体基因间隔区，能提供足够多的信息位点，为蓼科属间和种间物种进化关系及分子鉴定提供较高的分辨率。

表 5 蓼科植物 17 个物种的 32 个叶绿体基因间隔区矩阵位点信息

Table 5 Information of 32 chloroplast intergenic region in 17 species of Polygonaceae

基因间隔区	信息位点/%	非信息位点/%	一致位点/%	基因间隔区	信息位点/%	非信息位点/%	一致位点/%
accD-psaI	18.6	10.7	70.7	rps7-trnV-GAC	4.3	2.0	93.7
atpF-atpH	15.0	7.9	77.1	rps16-trnQ-UUG	21.1	12.8	66.1
atpH-atpI	18.6	11.5	69.9	rrn4.5-rrn5	13.5	1.9	84.6
clpP-psbB	16.1	7.3	76.6	rrn5-trnR-ACG	8.2	2.8	89.0
ndhB-rps7	2.2	2.0	95.8	trnC-GCA-petN	17.0	11.0	72.0
ndhC-trnV-UAC	12.4	10.3	77.3	trnD-GUC-trnY-GUA	19.2	7.6	73.2
ndhE-ndhG	20.7	20.4	58.9	trnE-UUC-trnT-GGU	20.8	13.2	66.0
petA-psbJ	16.2	8.9	74.9	trnF-GAA-ndhJ	20.1	9.3	70.6
petB-petD	9.9	8.8	81.3	trnL-CAA-ndhB	6.9	2.8	90.3
petN-psbM	19.3	9.2	71.5	trnL-UAA-trnF-GAA	19.5	6.8	73.7
psaJ-rpl33	23.0	13.9	63.1	trnN-GUU-ycf1	7.9	4.8	87.3
psbE-petL	19.0	8.9	72.1	trnP-UGG-psaJ	15.4	14.9	69.7
psbK-psbI	23.4	9.0	67.6	trnQ-UUG-psbK	13.0	9.9	77.1
psbM-trnD-GUC	17.1	12.4	70.5	trnR-ACG-trnN-GUU	6.0	3.1	90.9
rpl33-rps18	20.2	16.5	63.3	trnT-GGU-psbD	20.9	11.3	67.8
rpoB-trnC-GCA	19.3	12.1	68.6	ycf2-trnL-CAA	3.7	2.6	93.7

## 2.6 系统发育分析

叶绿体基因组序列用进行多重比对后，矩阵长 221 423 bp，其中信息位点 20 235 bp (9.1%)，非信息位点 22 977 bp (10.4%)，一致位点 178 211 bp (80.5%)，利用最大似然法和贝叶斯法构建篇蓄及 18 种蓼科植物叶绿体全基因组的系统发育树 (图 3)。系统发育树结果显示，蓼科植物除国外分布的 2 个属 *Afrobrunnichia erecta* 和 *Symmeria paniculata*，位于系统树基部位置。其余的取样类群可以分为 3 个主要分支，分支 A 仅

包括荞麦属 2 个种构成单系分支，具有 100% 支持率。分支 B 包括 3 个属，大黄属、山蓼属、酸模属构成单系分支具有 100%，分支 B 内这 3 个属的单系性都得到 100% 支持率。分支 C 包括 6 个属，虎杖属、木蓼属、沙拐枣属、*Muehlenbeckia australis*、何首乌属、蓼属构成单系分支具有 74% 支持率。分支 C 内，沙拐枣属的单系性都得到 100% 支持率。篇蓄与额河木蓼聚在一支具有 100% 支持率。何首乌属与虎杖属聚在一支，具有 100% 支持率。



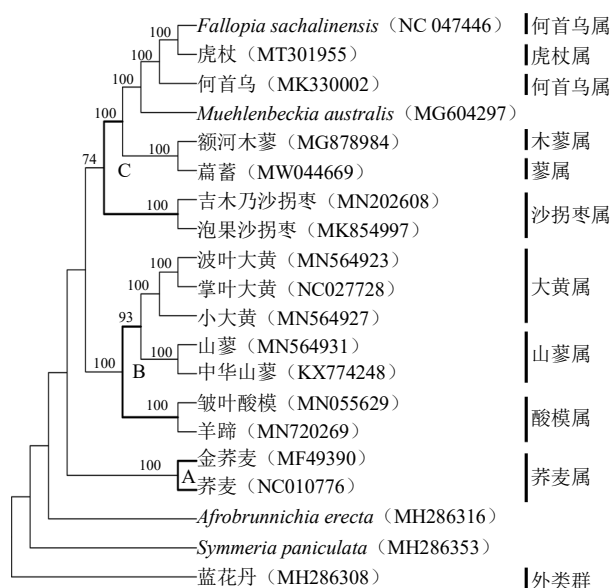


图 3 基于叶绿体全基因组构建的系统发育树

Fig. 3 Phylogenetic tree based on complete chloroplast genome

### 3 讨论

篇蓄作为我国传统的药材之一，具有重要的药用和经济价值。本研究完成了药用植物篇蓄叶绿体基因组的测序、组装与注释。对比篇蓄与蓼科其他 19 个物种的叶绿体基因组分析结果表明，篇蓄叶绿体基因组具有典型的 4 分区域结构，包括 1 个 LSC 区、1 对 IR 区和 1 个 SSC 区，其叶绿体基因组长度位于蓼科物种叶绿体基因组长度范围之内。被子植物中除部分列当科和兰科寄生植物的叶绿体基因组较短外，如 *Epifagus virginiana*<sup>[32]</sup>（叶绿体基因组全长 70 028 bp），天麻<sup>[33]</sup>（叶绿体基因组全长 35 304 bp）等，常见的叶绿体基因组大小为 120~180 kb，共编码 100~200 个基因，包括 70~80 个蛋白质编码基因，30~32 个 tRNA 基因，4 个 rRNA 基因<sup>[34]</sup>。此外，除部分列当科和兰科寄生植物的叶绿体基因组部分基因发生丢失，具有较快的进化速率外，大多数植物的叶绿体基因组结构和序列相当保守，重复片段较少<sup>[33]</sup>。篇蓄与蓼科其他 6 个属植物的叶绿体基因组 IR 边界均具有高度的保守性，其 LSC/IRb 边界(JLB)，SSC/IRb 边界(JSB)、SSC/IRa 边界(JSA)和 LSC/IRa 边界(JLA)的侧翼基因完全相同。且篇蓄的叶绿体基因组各边界的扩张范围，均位于蓼科物种叶绿体基因组 IR 边界的扩张范围之内。这表明在蓼科，篇蓄与其他物种的叶绿体基因组均具有较高保守性，适合用来解决属级以上分类等级的系统发育关系。

相对于叶绿体基因组编码区基因具有较高的保守性，叶绿体基因组的基因间隔区往往具有丰富的变异位点，本研究比较分析了蓼科 17 个物种的叶绿体基因组中 32 个基因间隔区，发现间隔区序列提供变异位点超过 20%以上共计有 8 个，分别是 *ndhE-ndhG*、*psaJ-rpl33*、*psbK-psbI*、*rpl33-rps18*、*rps16-trnQ-UUG*、*trnE-UUC-trnT-GGU*、*trnF-GAA-ndhJ*、*trnT-GGU-psbD*。这些基因间隔区在蓼科的属间和种间提供了丰富的信息位点。由于叶绿体基因组在大多数被子植物中为母系遗传，重组率低，核苷酸置换率适中<sup>[35]</sup>，进一步结合双亲遗传的核基因片段联合分析，为蓼科植物属下大范围的物种鉴定，杂交起源，多倍体物种的形成和系统进化分析提供可靠的分子标记片段。

为进一步界定药用植物篇蓄在蓼科的系统位置，基于蓼科 19 个物种叶绿体基因组全长构建的系统发育树结果表明，除国外分布的 2 个种 *Symmeria paniculata* 和 *Afrobrunni chiaerecta*，其余的 17 物种分别聚在 3 个不同分支内。篇蓄位于分支 C 内，与木蓼属的额河木蓼亲缘关系最近，二者构成一单系分支具有 100%的支持率，形态特征也支持这 2 个种所在蓼属和木蓼属亲缘关系较近，二者均具有茎直立，花被片五枚，柱头片状，瘦果具翅等特征<sup>[1]</sup>。该分支内，虎杖属与何首乌属亲缘关系较近，二者构成一单系分支，具有 100%支持率。分支 B 主要包括酸模族的 3 个属大黄属、酸模属和



山蓼属，叶绿体全基因组的系统发育树支持酸模族的单系性，具有 100% 的支持率，形态学特征显示该族具有花被片 4~6 枚，雄蕊 6~9 枚，瘦果具 3 棱等特征<sup>[1]</sup>。大黄属，酸模属和山蓼属的各自单系性也得到 100% 的支持率。分支 A 主要由荞麦属的 2 个种构成，也组成一单系群，具有 100% 的支持率。本研究基于叶绿体基因组数据表明，篇蓄隶属于蓼科的蓼属，与木蓼属的额河木蓼亲缘关系较近。

近年来，随着新一代测序技术和生物信息学的快速发展，越来越多的药用植物物种的叶绿体基因组被报道。叶绿体基因组作为超级条形码或基于叶绿体基因测序分析筛选 DNA 片段显示出巨大的物种识别潜力，特别是在物种亲缘关系较近的物种间<sup>[36]</sup>。本研究首次报道了蓼科中药材篇蓄的叶绿体全基因组，综合分析蓼科药用植物如大黄、何首乌和虎杖等的叶绿体基因组序列，结构和特征，筛选一批叶绿体基因组的间隔区。这为蓼科药用植物的分子鉴定，种质资源保护，系统进化关系及遗传多样性研究奠定基础。

**利益冲突** 所有作者均声明不存在利益冲突

## 参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志 (第 17 卷) [M]. 北京: 科学出版社, 1994: 126.
- [2] 国家中医药管理局《中华本草》编委会. 中华本草-2 [M]. 上海: 上海科学技术出版社, 1999: 639.
- [3] 中国药典 [S]. 一部. 2020: 348.
- [4] 王桂芝, 胡海涛, 董鹏达, 等. 中药篇蓄的研究现状 [J]. 黑龙江医药, 2010, 23(4): 614-616.
- [5] 赵爱华, 赵勤实, 林中文, 等. 篇蓄的化学成分研究 [J]. 天然产物研究与开发, 2002, 14(5): 29-32.
- [6] 汤迎爽, 宋红儒, 杨丽甲. 篇蓄的研究进展 [J]. 时珍国医国药, 2004, 15(1): 54.
- [7] 李曼曼, 刘增辉, 王海燕, 等. 篇蓄抑菌活性及化学成分研究 [J]. 天然产物研究与开发, 2014, 26(4): 526-530.
- [8] 郑泽生, 杨爱梅, 张富禄, 等. 篇蓄草提取物的抗氧化和抑菌活性研究 [J]. 中医药学报, 2019, 47(5): 55-58.
- [9] Daniell H, Lin C S, Yu M, *et al.* Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 134.
- [10] Pogson B J, Ganguly D, Albrecht-Borth V. Insights into chloroplast biogenesis and development [J]. *Biochim Biophys Acta BBA-Bioenerg*, 2015, 1847(9): 1017-1024.
- [11] Twyford A D, Ness R W. Strategies for complete plastid genome sequencing [J]. *Mol Ecol Resour*, 2017, 17(5): 858-868.
- [12] Yu X Y, Zuo L H, Lu D D, *et al.* Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis [J]. *Gene*, 2019, 689: 141-151.
- [13] Zhang F J, Wang T, Shu X C, *et al.* Complete chloroplast genomes and comparative analyses of *L. chinensis*, *L. anhuiensis*, and *L. Aurea* (Amaryllidaceae) [J]. *Int J Mol Sci*, 2020, 21(16): 5729.
- [14] 邢少辰, CLARKE JIHONG LIU. 叶绿体基因组研究进展 [J]. 生物化学与生物物理进展, 2008, 35(1): 21-28.
- [15] 张靖雯, 姜在民, 蔡靖. 紫丁香与羽叶丁香叶绿体 DNA 提取方法研究 [J]. 西北林学院学报, 2018, 33(4): 95-99.
- [16] Antipov D, Hartwick N, Shen M, *et al.* plasmidSPAdes: Assembling plasmids from whole genome sequencing data [J]. *Bioinformatics*, 2016, 32(22): 3380-3387.
- [17] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data [J]. *Nucleic Acids Res*, 2017, 45(4): e18.
- [18] 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.
- [19] Qian J, Song J Y, Gao H H, *et al.* The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza* [J]. *PLoS One*, 2013, 8(2): e57607.
- [20] Zhao Y B, Yin J L, Guo H Y, *et al.* The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng* [J]. *Front Plant Sci*, 2015, 5: 696.
- [21] 宋菊, 龙月红, 林丽梅, 等. 五加科植物叶绿体基因组结构与进化分析 [J]. 中草药, 2017, 48(24): 5070-5075.
- [22] 董博然, 赵志礼, 倪梁红, 等. 龙胆科叶绿体基因组结构特征及其药用植物鉴定意义 [J]. 中草药, 2020, 51(6): 1641-1649.
- [23] Yang Y, Dang Y Y, Li Q, *et al.* Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: Organizations and implications for genetic engineering [J]. *PLoS One*, 2014, 9(11): e110656.
- [24] 郭豪杰, 刘久石, 罗丽, 等. 五味子叶绿体基因组结构解析与比较分析 [J]. 中国科学: 生命科学, 2017, 47(7): 728-739.
- [25] Biswal D, Konhar R, Debnath M, *et al.* Chloroplast genome sequence annotation of *Dendrobium nobile* (Asparagales: Orchidaceae), an endangered medicinal orchid from northeast India [J]. *PLoS Curr*, 2017, 19: 1-8.

- [26] 蒋明, 王军峰, 应梦豪, 等. 三叶崖爬藤叶绿体基因组的组装与序列分析 [J]. 中草药, 2020, 51(2): 461-468.
- [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(12): 1647-1649.
- [28] Wang W B, Yu H, Wang J H, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J]. *Int J Mol Sci*, 2017, 18(11): 2288.
- [29] Dong W L, Wang R N, Zhang N Y, *et al.* Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution [J]. *Int J Mol Sci*, 2018, 19(3): 716.
- [30] 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.
- [31] 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.
- [32] Wolfe K H, Morden C W, Palmer J D. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant [J]. *PNAS*, 1992, 89(22): 10648-10652.
- [33] 马潇. 兰科质体基因组系统学研究 [D]. 北京: 中国科学院大学, 2018.
- [34] 李岩, 吕光辉, 张雪妮, 等. 十字花科植物叶绿体基因组结构及变异分析 [J]. 西北植物学报, 2017, 37(6): 1090-1101.
- [35] Drouin G, Daoud H, Xia J N. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants [J]. *Mol Phylogenet Evol*, 2008, 49(3): 827-831.
- [36] 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.

[责任编辑 时圣明]