

• 农业生物技术 •

长爪柃[木衣]叶绿体基因组特征系统发育及密码子偏好性分析

李连星¹, 彭劲谕¹, 王大玮^{1,2}, 段安安¹

1 西南林业大学 云南省高校林木遗传改良与繁育重点实验室, 云南 昆明 650224

2 西南林业大学 西南山地森林资源保育与利用教育部重点实验室, 云南 昆明 650224

李连星, 彭劲谕, 王大玮, 段安安. 长爪柃[木衣]叶绿体基因组特征系统发育及密码子偏好性分析. 生物工程学报, 2022, 38(1): 328-342.

LI LX, PENG JY, WANG DW, DUAN AA. Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*. Chin J Biotech, 2022, 38(1): 328-342.

摘 要: 长爪柃[木衣] (*Docynia longiunguis* Q. Luo & J. L. Liu) 是我国特有的柃[木衣]属植物, 具有较高的食药价值。对其叶绿体基因组进行分析, 有助于阐明柃[木衣]属内的系统发育关系, 为长爪柃[木衣]资源的开发利用及进一步研究奠定基础。结合其近缘种云南柃[木衣]叶绿体基因组数据, 在进行全序列比对后, 对其系统发育、密码子偏好性等进行分析。长爪柃[木衣]叶绿体基因组序列总长为 158 914 bp (GenBank 登录号为 MW367027), 总 GC 含量为 36.7%, 其中大的单拷贝区 (large single-copy, LSC) 长度为 87 020 bp, 小的单拷贝区 (small single-copy, SSC) 长度为 19 156 bp, 反向重复区 (inverted repeats, IRs) 长度为 26 369 bp。共注释了 102 个功能性基因, 包括 72 个蛋白编码基因、26 个编码 tRNA 基因和 4 个编码 rRNA 基因。构建系统发育树的最佳模型为 TVM+F+R2。系统发育分析结果表明, 长爪柃[木衣]与柃[木衣] (*Docynia indica* (Wall.) Dcne.) 聚为一支, 柃[木衣]属物种与苹果属 (*Malus*) 聚为一支。对长爪柃[木衣]及近缘种叶绿体基因组序列进行比对分析, *trnY*(GUA)-*psbD*、*ndhC-trnV*(UAC)、*accD-psaI*、*psbZ-trnFM*(CAU) 和 *ndhF-trnL* 等区域的变异较大, 核酸多样性分析则表明有 11 处 Pi 值>0.01 的高变区域, 且都位于 LSC 区及 SSC 区。除长爪柃[木衣]外, 其他序列中均有 *trnH* 基因位于 IRs/LSC 区交界处且都没有越过边界。密码子偏好分析显示, 长爪柃[木衣]叶绿体基因中异亮氨酸的密码子编码数量最多, 达到了 1 205 个。长爪柃[木衣]与山荆子 (*Malus baccata* (L.) Borkh.)、三叶海棠 (*Malus sieboldii* (Regel) Rehd.)、湖北海棠 (*Malus hupehensis* (Pamp.) Rehd.) 及木瓜 (*Chaenomeles sinensis* (Thouin) Koehne) 的亲缘关系最近; 其叶

Received: April 16, 2021; Accepted: July 20, 2021; Published online: September 1, 2021

Supported by: National Natural Science Foundation of China (32060350)

Corresponding author: WANG Dawei. Tel: +86-871-63863345; E-mail: wangdawei@swfu.edu.cn

基金项目: 国家自然科学基金 (32060350)

绿体基因密码子更偏好于使用 A/T 结尾; 长爪栒子[木衣]叶绿体基因组与其他蔷薇科植物叶绿体基因组在 4 个边界区域基因分布显示出较大差异, 与同属的云南栒子[木衣]及栒子[木衣]叶绿体基因组差异相对较小。长爪栒子[木衣]叶绿体基因组的组装注释、系统发育分析及序列比对分析, 为该物种的资源鉴定、开发和利用提供了理论依据。

关键词: 长爪栒子[木衣]; 栒子[木衣]属; 叶绿体基因组; 系统发育; IR 边界分析; 密码子偏好性分析

Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*

LI Lianxing¹, PENG Jinyu¹, WANG Dawei^{1,2}, DUAN An'an¹

1 Key Laboratory for Forest Genetic and Tree Improvement & Propagation in Universities of Yunnan Province, Southwest Forestry University, Kunming 650224, Yunnan, China

2 Key Laboratory for Forest Resource Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming 650224, Yunnan, China

Abstract: *Docynia longiunguis* is a plant uniquely present in China and is of high edible and medicinal value. The analysis of its chloroplast genome will help clarify the phylogenetic relationship among *Docynia* and facilitate the development and utilization of *D. longiunguis* resources. Based on the alignment of chloroplast genome sequences of related species, the phylogeny and codon preference were analyzed. The total length of *D. longiunguis* chloroplast genome sequence was 158 914 bp (GenBank accession number is MW367027), with an average GC content of 36.7%. The length of the large single-copy (LSC), the small single-copy (SSC), and inverted repeats (IRs) are 87 020 bp, 19 156 bp, and 26 369 bp, respectively. A total of 102 functional genes were annotated, including 72 protein-coding genes, 26 tRNA genes, and 4 rRNA genes. The best model for constructing phylogenetic tree was TVM+F+R2. *D. longiunguis* and *Docynia indica* were clustered into a single group, while *Docynia* and *Malus* were clustered into a single group. Comparison of the chloroplast genome sequences of *D. longiunguis* and its five related species revealed that *trnY* (*GUA*)-*psbD*, *ndhC-trnV* (*UAC*), *accD-psaI*, *psbZ-trnfM* (*CAU*), *ndhF-trnL* gene regions varied greatly. The nucleic acid diversity analysis showed that there were 11 high variation areas with nucleotide variability > 0.01, all were located in the LSC and SSC regions. Except for *D. longiunguis*, the *trnH* genes in other sequences were located at the IRs/LSC junction and did not cross the boundary. Codon preference analysis showed that *D. longiunguis* chloroplast genome has the largest number of isoleucine (Ile) codons, up to 1 205. *D. longiunguis* has the closest genetic relationship with *Malus baccata*, *Malus sieboldii*, *Malus hupehensis* and *Chaenomeles sinensis*. Its chloroplast genome codon prefers to end with A/T. The chloroplast genome of *D. longiunguis* and other Rosaceae chloroplast genomes showed great differences in gene distribution in four boundary regions, while relatively small differences from the chloroplast genomes of *Docynia*

delavayi and *D. indica* of the same genus were observed. The genome annotation, phylogenetic analysis and sequence alignment of chloroplast genome of *D. longiunguis* may facilitate the identification, development and utilization of this species.

Keywords: *Docynia longiunguis*; *Docynia*; chloroplast genome; phylogenetic analysis; IR boundary analysis; codon preference

叶绿体基因组中包含着诸多与植物光合作用相关的关键基因^[1]。凭借其单亲遗传的特性,叶绿体基因组在植物系统发育研究方面发挥着重要作用,是系统发育研究的理想工具^[2-3]。基于 Illumina 平台的高通量测序技术的出现,大大地减少了测序的成本及时间,基因组数据越来越多地被用于研究系统发育关系,使得拥有完整的核和细胞器基因组序列的物种数量迅速增加^[4]。使用高通量测序技术可以更高效地满足对动植物线粒体或叶绿体的测序等研究领域的研究需求^[5-9]。随着测序技术的发展,大量植物的叶绿体基因组得到公布,通过对以高通量测序为基础得到的叶绿体基因组序列为研究材料进行拼装注释^[10-15],为不同植物物种鉴定^[16]、谱系地理学^[17]、进化动力学^[18]等方面的研究提供了数据支撑^[19]。

长爪柎[木衣] (*Docynia longiunguis* Q. Luo & J. L. Liu) 为蔷薇科 (Rosaceae) 柎[木衣]属 (*Docynia*) 植物,于 2009 年 3 月被首次发现,经观察发现与同属的柎[木衣] [*Docynia indica* (Wall.) Dcne.]和云南柎[木衣] [*Docynia delavayi* (Franch.) Schneid.]在花的大小和花瓣基部等特征方面有显著区别^[20]。柎[木衣]属植物是非常具有利用价值的野生资源,用途十分广泛^[21]。除了其果实可以食用外,其树叶的内含物是很好的抗氧化剂,提取物有降血糖血脂的作用;其根、树皮、果还可用于入药,可治疗烫伤、骨折等^[22-24]。此外,柎[木衣]植物的果实与苹果

(*Malus pumila*) 等水果相比,其常量及微量元素含量更为丰富^[25]。目前对柎[木衣]属其他两个种的研究主要集中在次生代谢物功能分析^[24]、有效成分提取^[26]以及生物学特性和物候特征^[27]等方面。

本研究利用高通量测序技术对长爪柎[木衣]叶绿体基因组进行测序,并以已发表的云南柎[木衣]等近缘种的叶绿体基因组序列为参考,拼接得到长爪柎[木衣]叶绿体基因组全序列,以此进行系统发育、序列对比、密码子偏好性分析;并揭示长爪柎[木衣]与其他蔷薇科植物的进化关系及其在系统发育中的地位,对了解柎[木衣]属主要谱系之间的系统发育关系有一定的意义,为长爪柎[木衣]资源鉴定、开发和利用提供理论依据。

1 材料与方法

1.1 实验材料

所用长爪柎[木衣]样本来源于四川省凉山州西昌市 (102°15' E, 27°51' N),海拔为 1 712 m。采摘新鲜幼嫩的叶片,液氮中速冻后存于 -80 °C 冰箱保存备用。

1.2 方法

1.2.1 基因组 DNA 提取及测序

以长爪柎[木衣]的幼嫩叶片为材料,参照张鲁杰的改良 CTAB 法^[28],提取长爪柎[木衣]基因组 DNA。使用分光光度计检测所提样品的 DNA 浓度,并用 1%的琼脂糖凝胶电泳检测

提取的 DNA 纯度。送至安诺优达基因科技 (北京) 有限公司使用 Illumina 测序平台进行高通量测序。

1.2.2 序列拼接及注释

长爪柃[木衣]叶绿体基因组的序列拼接通过 MEGA 5.1 软件进行, 以 NCBI 上已公布的云南柃[木衣] (*D. delavayi*) 叶绿体基因组序列为参考序列 (KX499860) 进行组装。并将组装好的序列用 Genious 8.1.3 软件进行注释, 安装该软件中的 MAFFT 插件, 将该序列与其他近缘种序列进行比对注释^[29]。对所有的蛋白编码基因组的相关内容进行检查, 并将其修正至准确位置, 进而得到完整的注释结果^[30]。利用在线软件 OGDRAW^[31]绘制长爪柃[木衣]叶绿体基因组物理图谱。注释完成后, 提交至 NCBI 数据库, 登录号为 MW367027。

1.2.3 叶绿体基因组基因特征分析

叶绿体基因组完成拼接和注释后, 在 Geneious 软件中完成其特征分析^[32], 包括基因组全长、各个区域长度及 GC 含量、基因数目和类型、基因拷贝情况等。

1.2.4 系统发育分析

从 NCBI 数据库下载已报道的蔷薇科 12 个属 24 个种的叶绿体基因组并以鼠李科 (Rhamnaceae) 枣属的枣 (*Ziziphus jujuba*) 作为外类群, 与测序得到的长爪柃[木衣]叶绿体基因组进行聚类分析, 并研究其系统发育关系。使用 MAFFT 对序列进行比对, 利用 IQ-TREE v1.6.1 软件并使用最大似然法 (maximum likelihood, ML) 进行系统发育树的构建, 选择最佳模型 TVM+F+R2, 进化树用 1 000 次重复 bootstrap 检验, 构建系统发育树^[33]。并对长爪柃[木衣]的系统进化关系进行分析。具体所选物种见表 1。

表 1 用于系统发育分析的 26 种植物叶绿体基因组
Table 1 Chloroplast genomes of 26 plants for phylogenetic analysis

Number	Species	GenBank Accession No.	Families
1	<i>Docynia longiunguis</i>	MW367027	Rosaceae
2	<i>Docynia delavayi</i>	KX499860	Rosaceae
3	<i>Docynia indica</i>	MN088849	Rosaceae
4	<i>Pyrus communis</i>	MN577870	Rosaceae
5	<i>Malus baccata</i>	MK896774	Rosaceae
6	<i>Malus sieboldii</i>	MT593044	Rosaceae
7	<i>Malus hupehensis</i>	MK020147	Rosaceae
8	<i>Eriobotrya japonica</i>	KT633951	Rosaceae
9	<i>Eriobotrya malipoensis</i>	MN577881	Rosaceae
10	<i>Eriobotrya obovata</i>	MN577882	Rosaceae
11	<i>Eriobotrya salwinensis</i>	MN577883	Rosaceae
12	<i>Chaenomeles sinensis</i>	MN577871	Rosaceae
13	<i>Cotoneaster buxifolius</i>	MN577892	Rosaceae
14	<i>Cotoneaster microphyllus</i>	MN577873	Rosaceae
15	<i>Cotoneaster taylorii</i>	MN577872	Rosaceae
16	<i>Cotoneaster rubens</i>	MN577895	Rosaceae
17	<i>Cotoneaster salicifolius</i>	MN577863	Rosaceae
18	<i>Cotoneaster silvestrii</i>	MN577894	Rosaceae
19	<i>Crataegus kansuensis</i>	MF784433	Rosaceae
20	<i>Cydonia oblonga</i>	KX499857	Rosaceae
21	<i>Dichotomanthes tristaniicarpa</i>	MN577869	Rosaceae
22	<i>Photinia villosa</i>	MN061989	Rosaceae
23	<i>Pourthiaea arguta</i>	MN061991	Rosaceae
24	<i>Rhaphiolepis indica</i>	MN577864	Rosaceae
25	<i>Rhaphiolepis salicifolia</i>	MN577876	Rosaceae
26	<i>Ziziphus jujuba</i>	KU351660	Rhamnaceae

1.2.5 叶绿体基因组序列的比对分析

选取枇杷 (*Eriobotrya japonica*, KT633951)、柃[木衣] (*D. indica*, MN088849)、西洋梨 (*Pyrus communis*, MN577870)、湖北海棠 (*Malus hupehensis*, MK020147) 及云南柃[木衣] (*D. delavayi*, KX499860) 5 个物种, 使用 mVISTA 软件与长爪柃[木衣]叶绿体基因组序列进行可视化分析, 并进一步分析其序列之间的异同。

将拼接好的6个样本序列以.fasta的格式上传至在线网站 mVISTA, 在 Shuffle-LAGAN 算法模式下进行比对。输出可视性文件并进行比较分析^[32]。利用 DNAsp v.5.0 软件计算核苷酸多态性 (nucleotide polymorphism, Pi), 将参数设置为 Window length 大小为 600 bp, 每个位点之间的距离为 200 bp, 构建多态性位点折线图^[34]。

1.2.6 叶绿体基因组边界区的收缩扩张

将藏南栒子 (*Cotoneaster taylorii*, MN577872)、中华落叶石楠 (*Pourthiaea arguta*, MN061991)、移[木衣] (*D. indica*, MN088849)、云南移[木衣] (*D. delavayi*, KX499860)、西洋梨 (*Pyrus communis*, MN577870)、湖北海棠 (*Malus hupehensis*, MK020147)、榲桲 (*Cydonia oblonga*, KX499857)、甘肃山楂 (*Crataegus kansuensis*, MF784433) 及木瓜 (*Chaenomeles sinensis*, MN577871) 9个物种与长爪移[木衣]叶绿体基因组进行 IR/SC 边界分析。将长爪移[木衣]及其近缘种叶绿体基因组序列保存为.gb 格式, 上传至在线分析网站 IRscope 与长爪移[木衣]比对, 进行叶绿体基因组 IR 区边界图的绘制, 将叶绿体基因组边界区域上的基因可视化, 并通过比较4个边界区域的基因分布对 IR 区进行收缩和扩张的分析。

1.2.7 密码子偏好性分析

使用 CodonW 软件对长爪移[木衣]叶绿体基因组 CDS 的密码子数量、密码子使用频率和同义密码子相对使用度进行分析。

2 结果与分析

2.1 叶绿体基因组的基本特征

长爪移[木衣]叶绿体基因组结构特征与大多数被子植物叶绿体基因组类似, 呈典型四段式结构, 总长度为 158 914 bp, 包括反向重复区 IRa 和 IRb 各 26 369 bp, 大的单拷贝 (LSC)

区和小的单拷贝 (SSC) 区长度分别为 87 020 bp 和 19 156 bp (图 1)。总 GC 含量为 36.7%, AT 含量为 63.3%。各区段具体的碱基组成见表 2。

通过分析注释, 长爪移[木衣]叶绿体基因组共有 102 个功能性基因 (剔除重复后), 包括 72 个蛋白编码基因、26 个 tRNA 基因和 4 个 rRNA 基因。其中 15 个基因有两个拷贝, 包括 7 个蛋白编码基因 (*rps7*、*rps12*、*rpl2*、*rpl23*、*psaI*、*ndhB*、*ycf2*), 4 个 tRNA 基因 (*trnL-CAA*、*trnN-GUU*、*trnR-ACG*、*trnV-GAC*) 和 4 个 rRNA 基因 (*rrn16*、*rrn23*、*rrn4.5*、*rrn5*)。具体见表 3。

2.2 长爪移[木衣]叶绿体基因组系统发育分析

本研究为了确定长爪移[木衣]在蔷薇科的系统位置, 从 NCBI 上下载了 23 个蔷薇科各物种的叶绿体基因组作为参考进行全局比对, 并以鼠李科枣属的枣 (*Ziziphus jujuba*) 为外类群, 构建了 ML 系统发育树 (图 2)。

结果表明, 聚类的支持率较高, 只有少部分分支的检验分值未达到 100%。所有蔷薇科物种聚为一支, 鼠李科枣属的枣成一支。所有蔷薇科物种分为 4 大进化支, 第 1 支由牛筋条属的牛筋条、榲桲属的榲桲和石楠属的毛叶石楠及中华落叶石楠聚在一起, 表明蔷薇科中牛筋条属、榲桲属、石楠属的亲缘关系较近; 第 2 支由栒子属的藏南栒子、华中栒子、柳叶栒子、小叶栒子、红花栒子、黄杨叶栒子以及梨属的西洋梨聚在一起, 说明栒子属与梨属亲缘关系较近; 第 3 支由移[木衣]属的长爪移[木衣]、移[木衣]及云南移[木衣]和苹果属的三叶海棠、山荆子、湖北海棠以及木瓜海棠属的木瓜组成; 第 4 分支由枇杷属的枇杷、麻栗坡枇杷、倒卵叶枇杷、怒江枇杷和石斑木属的石斑木以及柳叶石斑木组成。聚类分析结果表明, 苹果属与移[木衣]属亲缘关系最近, 其次是枇杷属和木瓜海棠属。

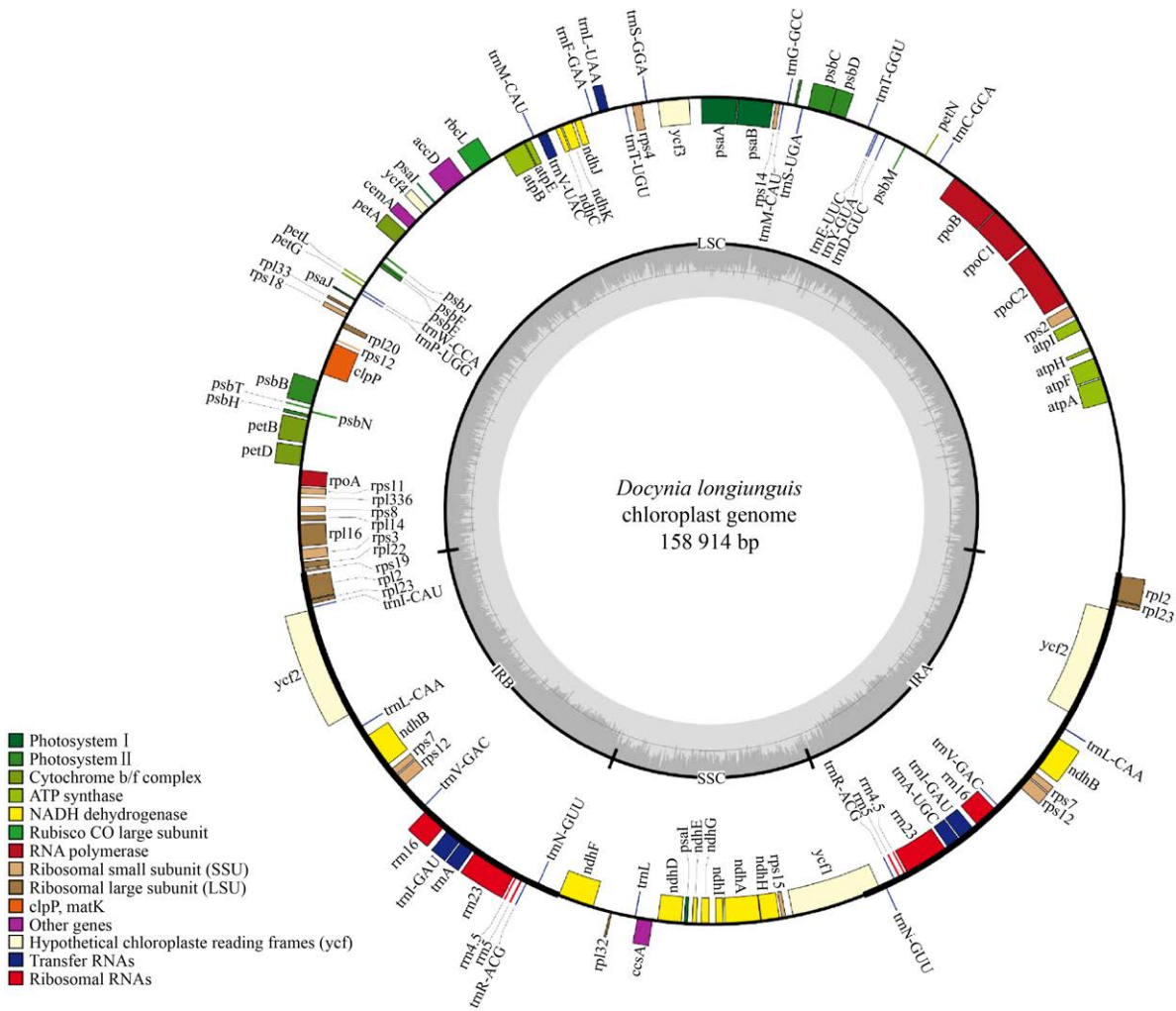


图 1 长爪柃[木衣]叶绿体基因组图谱 外圈基因逆时针方向转录,内圈基因顺时针方向转录,不同功能的基因以不同的颜色来表示,图中内圈深色部分表示 GC 含量,浅色部分表示 AT 含量

Figure 1 Gene map of *Docynia longiunguis* chloroplast genome. Genes drawn outside the circle are transcribed anti-clockwise, while genes shown inside the circle are transcribed clockwise. Genes with different functions are represented by different colors. The dark grey of the inner circle indicates GC content, and the light grey indicates AT content.

表 2 长爪柃[木衣]叶绿体基因组碱基组成

Table 2 Base composition of chloroplast genome in *Docynia longiunguis*

Region	T (U) (%)	A (%)	C (%)	G (%)	Length (bp)	GC (%)
LSC	33.4	32.0	17.8	16.85	87 020	34.6
SSC	34.8	34.8	15.9	14.50	19 156	30.4
IRA	28.7	28.6	22.1	20.6	26 369	42.7
IRB	28.7	28.6	20.6	22.1	26 369	42.7
Total	32.0	31.2	18.8	18.0	158 914	36.7

2.3 长爪柃[木衣]叶绿体基因组序列比对分析

为研究长爪柃[木衣]与其他蔷薇科植物在叶绿体基因组序列上的异同,对长爪柃[木衣]及其近缘种进行全序列比对分析(图 3)。结果显示,6 个蔷薇科植物叶绿体基因组之间具有高度的相似性,但同时也存在着一些差异。其中,变异较大的区域为 *trnY (GUA)-psbD*、

表 3 长爪柊[木衣]叶绿体基因组基因列表

Table 3 List of chloroplast genes of *Docynia longiunguis*

Gene function	Gene category	Gene name	
Self replication	Ribosomal RNAs	<i>rrn16^c, rrn23^c, rrn4.5^c, rrn5^c</i>	
	tRNA	<i>trnA^a, trnA-UGC^a, trnC-GCA, trnD-GUC</i>	
	RNA	<i>trnE-UUC, trnF-GAA, trnM-CAU, trnG-GCC</i>	
	Transfer RNAs		<i>trnI-CAU, tRNI-GAU, trnI-GAU, trnL, trnL-CAA^c</i>
			<i>trnL-UAA, trnM-CAU, trnN-GUU^c, trnP-UGG</i>
			<i>trnR-ACG^c, trnS-GGA, trnS-UGA, trnT-GGU</i>
			<i>trnT-UGU, trnV-GAC^c, trnV-UAC, trnW-CCA</i>
			<i>trnY-GUA</i>
			<i>rps2, rps3, rps4, rps7^c, rps8, rps11, rps12^{bc}</i>
		<i>rps14, rps15, rps18, rps19</i>	
		<i>rpl2^{ac}, rpl14, rpl16^a, rpl20, rpl22, rpl23^c, rpl32, rpl33, rpl36</i>	
		<i>rpoA, rpoB, rpoC1^a, rpoC2</i>	
	Genes for photosynthesis	Photosystem I	<i>psaA, psaB, psaI^c, psaJ</i>
Photosystem II		<i>psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbM, psbN, psbT, psbZ</i>	
Cytochrome b/f complex		<i>petA, petB^a, petD^a, petG, petL, petN</i>	
ATP synthase		<i>atpA, atpB, atpE, atpF^a, atpH, atpI</i>	
ATP-dependent protease subunit p gene		<i>clpP</i>	
Rubisco CO large subunit		<i>rbcL</i>	
Dehydrogenase		<i>ndhA^a, ndhB^{ac}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
Other genes	Envelop membrane protein	<i>cemA</i>	
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>	
	c-type cytochrome synthesis ccsA gene	<i>ccsA</i>	
	Protease gene	<i>clpP^b</i>	
Genes of unknown function	Conserved open reading frames	<i>ycf1, ycf2^c, ycf3^b, ycf4</i>	

Note: a means the gene contains 1 intron; b means the gene contains 2 introns; c means the gene contains 2 copies.

ndhC-trnV (UAC)、*accD-psaI*、*psbZ-trnM (CAU)* 和 *ndhF-trnL*。

六种蔷薇科植物叶绿体基因组序列经过 MAFFT 比对后通过 DNAsp 软件分析, 结果显示 IR 区的核酸变异度显著低于 LSC 区和 SSC 区 (图 4)。其中, 位于 LSC 区的 *petA-psbJ* 基因间区的 Pi 值为最高达到了 0.024 33。图中有 3 处 $0.015 < \text{Pi} < 0.02$ 的热点, 分别是 *petD-rpoA*、*rpl33-rpl14* 基因间区以及 *ycf1* 基因区; 有 7 处 $0.01 < \text{Pi} < 0.015$ 的热点, 分别是 *psbZ-rps14*、*accD-psaI*、*trnR (ACG)-trnA (UGC)*、*rps12-rps7*

基因间区及 *rpoB*、*ycf3*、*rpl22* 三处基因区, 表现出了较高水平的核苷酸多样性。以上结果说明在叶绿体基因组中 LSC 区及 SSC 区为变异发生的高频区域。

2.4 长爪柊[木衣]叶绿体基因组 IR 区边界的收缩和扩张分析

基因组 IR 区边界的收缩和扩张分析结果表明, 参与对比的 10 种植物 IR 区长度变化不大 (26 249–26 399 bp)。图 5 中长爪柊[木衣]、柊[木衣]和椴椴的 *rps19* 基因处于 LSC, 其余近缘种的 IRb 区扩张到了 *rps19* 基因中, 且扩

张长度均为 120 bp。在这 10 种植物叶绿体基因组中, 都有 *ycf1* 基因的存在, 其中一些物种的 *ycf1* 基因延伸至 SSC 区域几个碱基, 其余均位于 IRb 区中。所有物种在 SSC/Ira 边界上的基因分布十分相似, IR 区均延伸至 *ycf1* 基因内部, 除椴椴 *Cydonia oblonga* 的 *ycf1* 基

因长度较短为 2 297 bp 外, 其他种的 *ycf1* 基因长度均大于 5 600 bp。所有 10 种植物的 IRb 区均延伸进位于 SSC 区域中的 *ndhF* 基因中 12–14 bp。除长爪柃[木衣]外, 其他序列中均有 *trnH* 基因位于 IRa/LSC 区交界处且都没有越过边界。

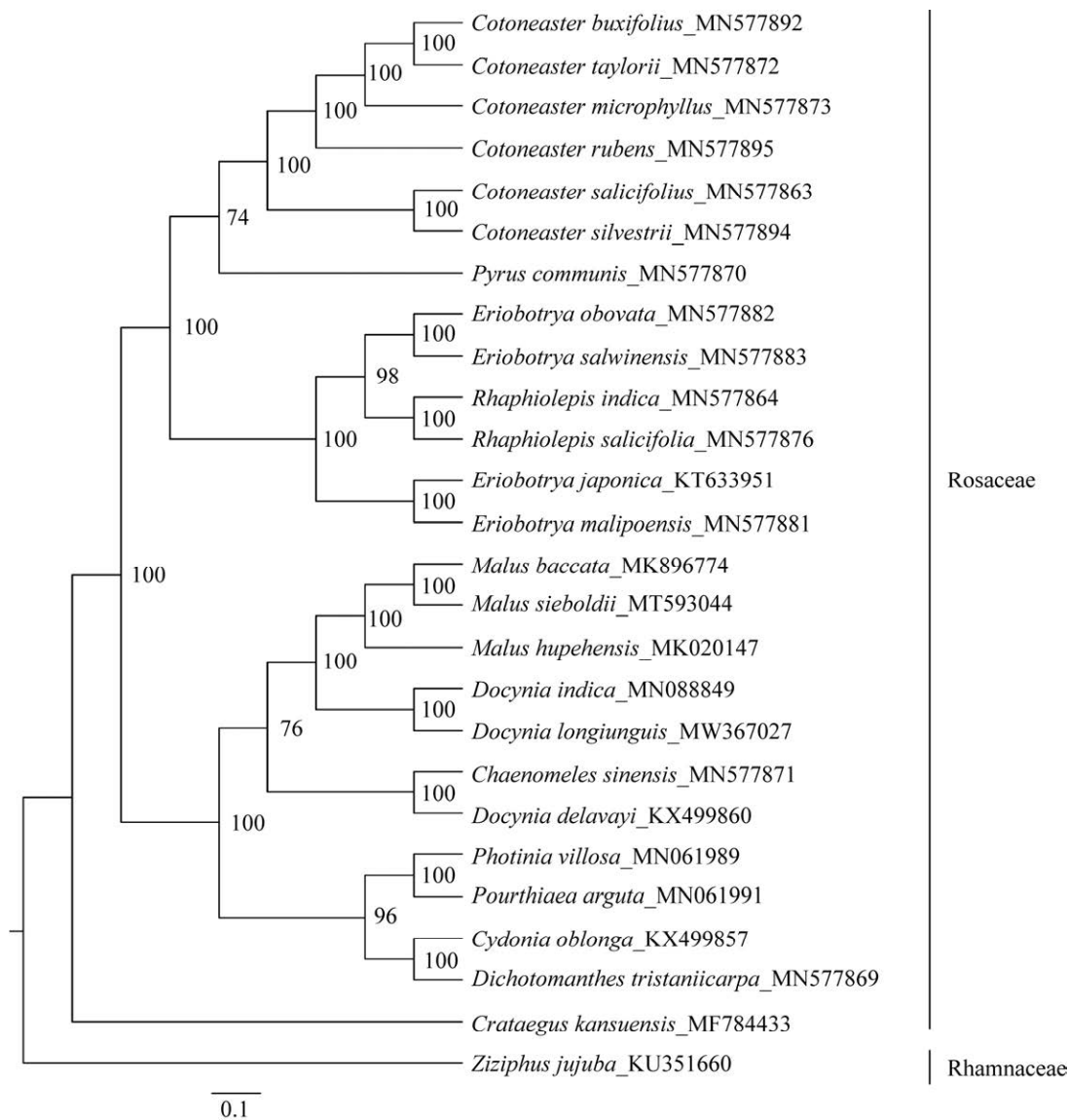


图 2 基于叶绿体基因组序列用 ML 法构建的 24 个物种的系统进化树 支上的数字为 1 000 次自展支持率 (%)

Figure 2 Phylogenetic tree of 24 species constructed by ML using chloroplast genome sequences. The 1 000 bootstrap value (%) was shown above the branch.

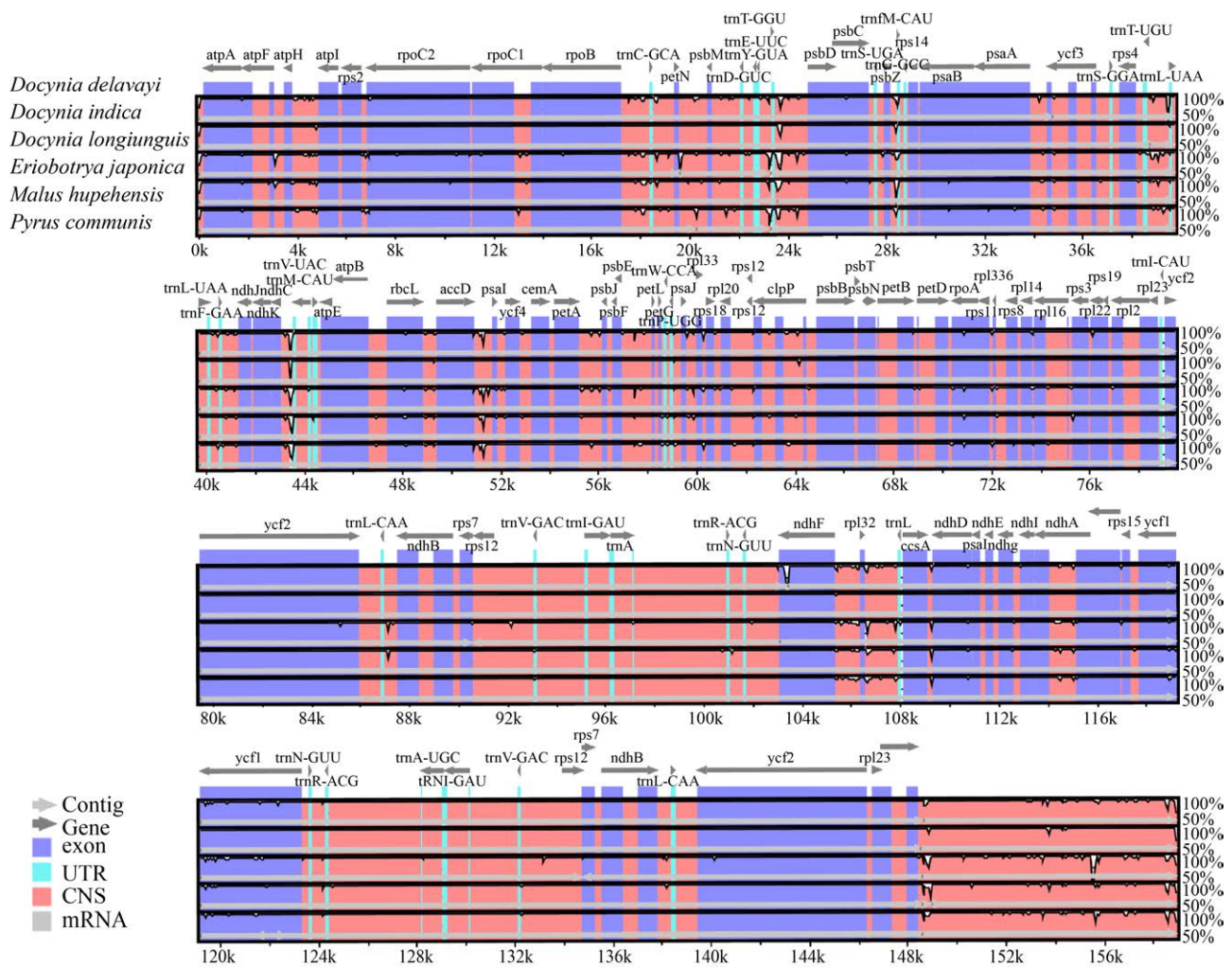


图3 六种蔷薇科植物叶绿体基因组序列比对 灰色箭头表示基因的方向,纵坐标表示从50%到100%的一致度

Figure 3 Sequence alignment of chloroplast genome of 6 Rosaceae plants. The gray arrows above the alignment indicate the genes' orientations. The vertical scale represents the percentage of identity, ranging from 50% to 100%.

2.5 密码子偏好性分析

长爪柊[木衣]叶绿体基因密码子偏好性分析发现,其叶绿体基因组中RSCU值>1的密码子为34个(图6),且其中有27个密码子是以A/T结尾,说明长爪柊[木衣]叶绿体基因组的密码子偏向于以A/T结尾,不偏好于以C/G结尾。分析结果表明,异亮氨酸(Ile)的密码子编码数量最多,达到了1205个,占比约为8.77%;而

半胱氨酸(Cys)的密码子数量最少为294个,占比为1.93%。编码亮氨酸(Leu)、苯丙氨酸(Phe)、苏氨酸(Thr)、赖氨酸(Lys)和丝氨酸(Ser)的密码子所占比例分别为8.76%、6.77%、5.82%、5.81%和5.46%。在所有的密码子中,使用次数最多的为赖氨酸的AAA,次数为545,相反使用次数最少的密码子是精氨酸(Arg)的CGC,为60次;偏好于以UAA作为终止密码子。

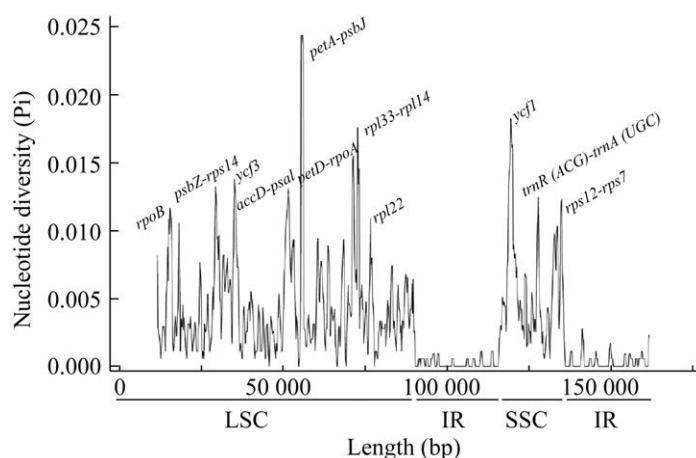


图 4 滑动窗口分析蔷薇科植物叶绿体基因组核酸多样性 x 轴: 窗口中点的位置; y 轴: 每个窗口的核苷酸多样性

Figure 4 Sliding window analysis for the nucleotide diversity. x -axis: position of the midpoint of a window. y -axis: nucleotide diversity of each window.

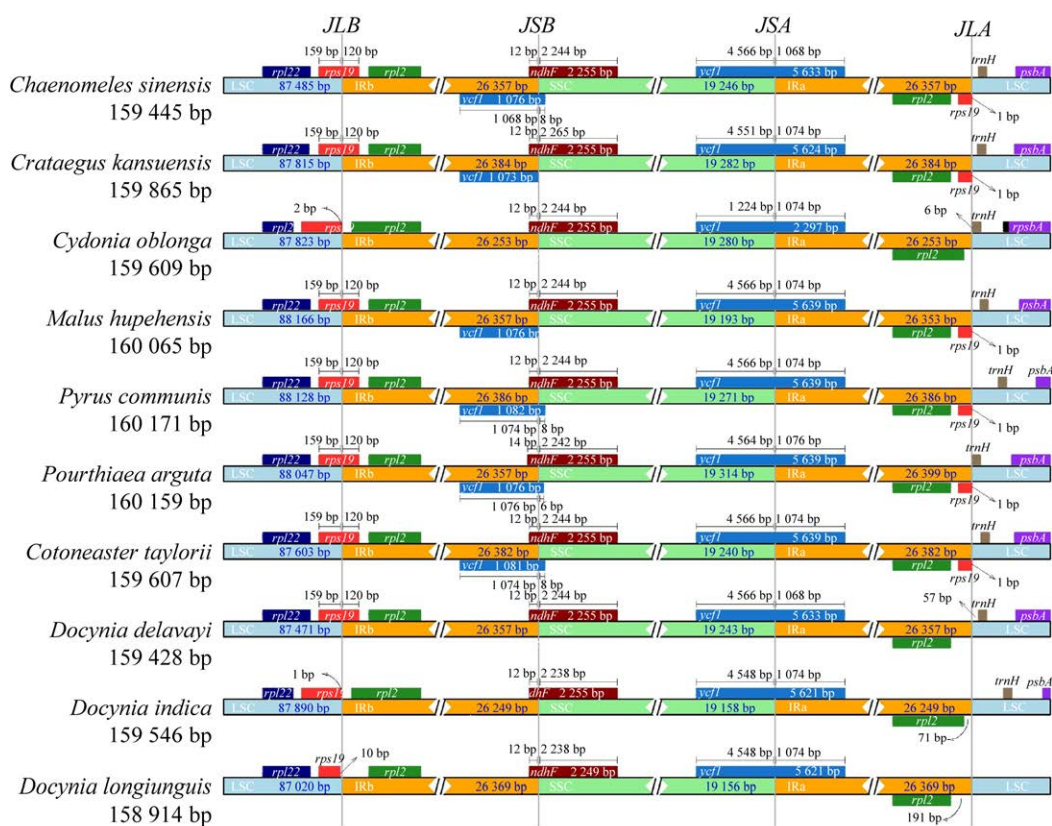


图 5 十种蔷薇科植物叶绿体基因组 IR/SC 边界收缩和扩张情况比较 不同基因使用不同颜色来表示, 基因和边界之间的间隙由碱基长度 (bp) 表示

Figure 5 Comparison of IR/SC boundary contraction and expansion in chloroplast genome of 10 Rosaceae plants. Different genes are denoted by colored boxes. The gaps between the genes and the boundaries are indicated by the base lengths (bp).

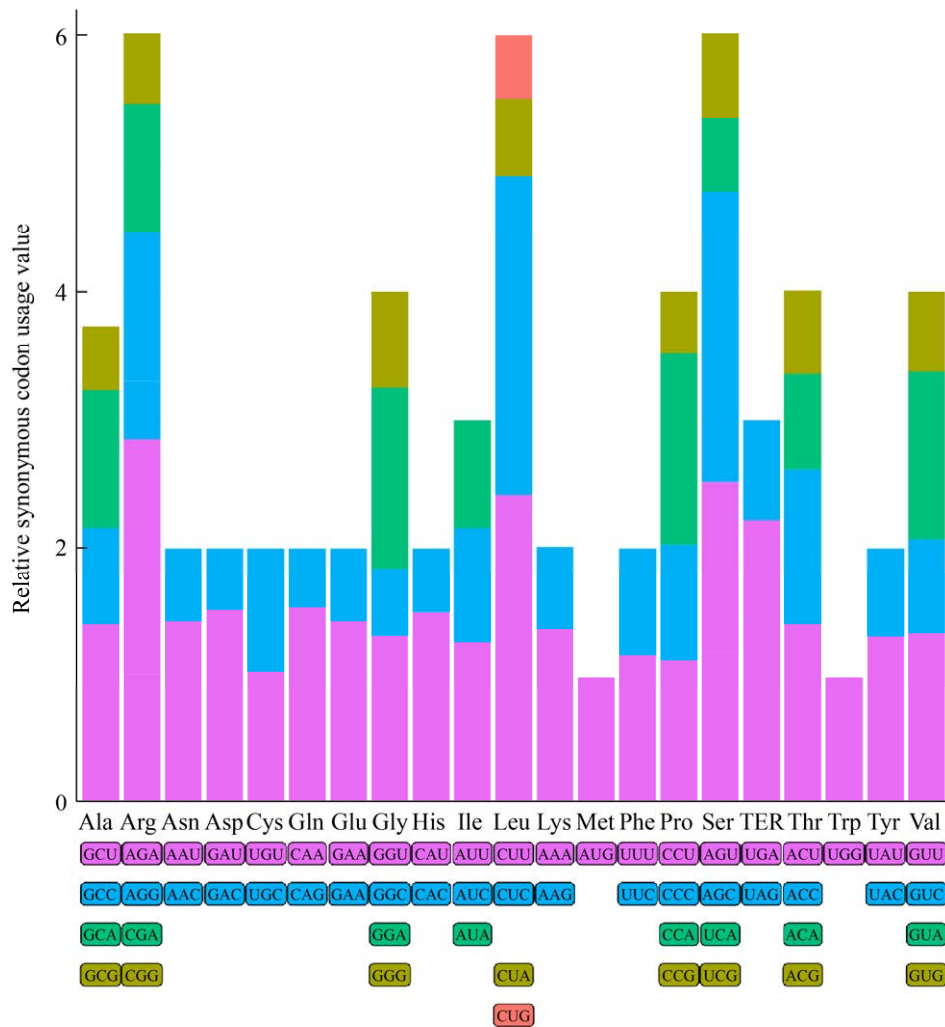


图 6 长爪柃[木衣]叶绿体基因密码子偏好性分析 x 轴代表密码子家族

Figure 6 Codon preference analysis of *D. longiunguis* chloroplast genome. The x-axis represents codon families.

3 讨论

本研究利用高通量测序技术对长爪柃[木衣]叶绿体基因组进行了测序,在拼接组装后得到了完整的叶绿体基因组。在此基础上,进行了系统发育、序列比对、边界扩张收缩、密码子偏好性等分析。长爪柃[木衣]叶绿体全基因组长度为 158 914 bp,其中 LSC 长度为 87 020 bp,SSC 长度为 19 156 bp,IRs 的长度为 26 369 bp,与被子植物叶绿体基因组基本特征相吻合^[35]。

长爪柃[木衣]叶绿体全基因组的 GC 含量为 36.75%,AT 含量为 63.25%,AT 含量明显高于 GC 含量,与柃[木衣]属其他两种植物有细微的差距(柃[木衣] GC 含量为 36.64%,云南柃[木衣] GC 含量为 36.66%)^[36-37]。因为在基因组中 GC 含量越高、AT 含量越低,DNA 密度就越大,序列则较为保守且不易变性^[38]。因此根据这一特性可推测长爪柃[木衣]叶绿体基因组序列具有一定的易变性,可能会导致其基因序列比其他物种的叶绿体基因组序列更加容易产生变异。

使用最大似然法对近缘种的叶绿体基因组与测序得到的长爪柊[木衣]叶绿体基因组进行聚类分析。由构成的 ML 系统发育树可得,除槲栎、牛筋条这一分支外各分支支持率均较高。聚类结果表明,柊[木衣]属与木瓜属及苹果属植物聚为一支,这与前人对蔷薇科植物所作的聚类结果一致^[39-40],说明以该方法得到的聚类分析结果是可靠的。该研究结果为进一步确定蔷薇科各属植物的起源及亲缘关系提供了一定的依据。

IR 区和 SC 区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^[41],不同物种的叶绿体基因组的 IR/SC 边界位置变化是不同的,而这些变化在同一科的物种上也会有所不同^[41]。对 10 种蔷薇科叶绿体基因组 IR/SC 边界区域的分析结果表明,这些叶绿体基因组的 IR 区域都存在扩张或收缩的现象。长爪柊[木衣]的 *rps19* 基因没有跨过 LSC/IR 边界处(距离边界处 10 bp),这表明了长爪柊[木衣]的 IR 区相对于其他近缘种有较为明显的收缩^[42],这种收缩可能是造成长爪柊[木衣]叶绿体基因组的长度短于其他近缘种序列的长度的主要原因。这一结果与 Cheng 等^[43]对草莓 (*Fragaria × ananassa*) 及其近缘种叶绿体基因组进行边界分析得到的结果相似;但是这种情况在对白芥属 (*Sinapis*) 植物^[44]及姜科 (*Zingiberaceae*) 植物^[45]的研究里并未出现。由此推测造成这一差异的原因可能是植物进化过程中 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^[46]。

本研究将长爪柊[木衣]及其 5 个近缘种叶绿体基因组序列进行比对分析,发现参试的叶绿体基因组序列之间存在着高度的相似性,存在着几个变异较大的区域。这些处于热点区域的差异位点对 DNA 条形码的开发有重要的参

考价值,从而可以应用于种质资源鉴定及其他研究^[47]。在核酸多样性分析中可知,所发现的变异位点 Pi 值范围在 0.005 6–0.024 33 之间,同时筛选出 11 个变异位点,其中 8 个位于 LSC 区,3 个位于 SSC 区,而两个 IR 区的核酸变异度则较低。特别是 *ycf1* 基因处的核酸变异度较高,其原因可能是因为该基因跨越了 IR 区与 SSC 区的交界处,也可能是存在着重复序列^[48]。这与对日本落叶松 (*Larix kaempferi*) 及龙血树 (*Dracaena draco*) 等植物的分析中发现的 IR 区核酸变异度较高的情况有差异^[49-50],说明叶绿体基因组在不同区域和不同物种间有着不同的进化速率^[51],这些高变区域可以作为对柊[木衣]属植物进行进化分析研究的依据^[52]。

密码子偏好性广泛存在于自然界植物基因组中,是在长期的进化发育中演变而来,会对其造成影响的因素有很多,包括突变、选择、基因长度、基因功能等^[53]。通过密码子偏好性分析结果发现,RSCU 值>1 的密码子为 34 个,而第 3 位是 A/T 的密码子达到了 27 个,占总量的 79.4%,说明长爪柊[木衣]叶绿体基因密码子偏好于以 A/T 结尾。这一特点与沙枣 (*Elaeagnus angustifolia*)、阔叶猕猴桃 (*Actinidia latifolia*)、羽叶丁香 (*Syringa pinnatifolia*) 等植物叶绿体基因组的特征都基本一致,说明不同物种之间密码子偏好性有着一定的相似性,而这一特性的产生可能与变异、自然选择及其他方面有关系^[54-56]。

4 结论

本研究首次使用高通量测序手段对长爪柊[木衣]样本进行测序,并进行序列的组装拼接,得到完整的长爪柊[木衣]叶绿体基因组序列。研究发现长爪柊[木衣]叶绿体基因组长度为 158 914 bp,呈典型的环状四分体结构;去除重复后该叶绿体基因组共有 104 个功能性基因。

长爪柃[木衣]叶绿体基因组更偏向于使用含有A/T的密码子。对长爪柃[木衣]叶绿体基因组进行组装注释、系统发育分析及序列比对等分析,对了解柃[木衣]属各个种之间的系统发育关系有重要意义,有利于对柃[木衣]属植物资源的开发利用,为该物种的下一步研究提供了理论依据。

REFERENCES

- [1] Bansal KC, Saha D. Chloroplast genomics and genetic engineering for crop improvement. *Agric Res*, 2012, 1(1): 53-66.
- [2] Wei W, Youliang Z, Li C, et al. PCR-RFLP analysis of cpDNA and mtDNA in the genus *Houttuynia* in some areas of China. *Hereditas*, 2005, 142(2005): 24-32.
- [3] Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*, 2016, 17(1): 134.
- [4] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*, 2008, 26(10): 1135-1145.
- [5] Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*, 2020, 21(10): 597-614.
- [6] Paaajanen P, Kettleborough G, López-Girona E, et al. A critical comparison of technologies for a plant genome sequencing project. *Gigascience*, 2019, 8(3): giy163.
- [7] Yildiz G, Ozkilinc H. First characterization of the complete mitochondrial genome of fungal plant-pathogen *Monilinia laxa* which represents the mobile intron rich structure. *Sci Rep*, 2020, 10: 13644.
- [8] Priest SJ, Yadav V, Heitman J. Advances in understanding the evolution of fungal genome architecture. *F1000 Res*, 2020, 9: 776.
- [9] Cui H, Ding Z, Zhu Q, et al. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Sci Rep*, 2021, 11(1): 1595.
- [10] Henriquez CL, Abdullah, Ahmed I, et al. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics*, 2020, 112(3): 2349-2360.
- [11] Park I, Song JH, Yang SY, et al. Comparative analysis of Actaea chloroplast genomes and molecular marker development for the identification of authentic cimicifugae rhizoma. *Plants*, 2020, 9(2): 157.
- [12] Zheng G, Wei L, Ma L, et al. Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships. *Plant Mol Biol*, 2020, 102(6): 659-676.
- [13] 周晓君, 张凯, 彭正锋, 等. 矮牡丹与芍药属其他5个种叶绿体基因组特征的比较. *林业科学*, 2020, 56(4): 82-88.
Zhou XJ, Zhang K, Peng ZF, et al. Comparative analysis of chloroplast genome characteristics between *Paeonia jishanensis* and other five species of *Paeonia*. *Sci Silvae Sin*, 2020, 56(4): 82-88 (in Chinese).
- [14] 苏宁, 何兆峰, 欧平和, 等. 小麦属植物叶绿体基因组结构的比较分析. *麦类作物学报*, 2020, 40(1): 55-64.
Su N, He ZF, Ou PH, et al. Comparative analysis of the chloroplast genome in *Triticum* species. *J Triticeae Crop*, 2020, 40(1): 55-64 (in Chinese).
- [15] 路东晔, 张磊, 张国盛, 等. 杨柳科植物叶绿体基因组结构及变异分析. *西北农林科技大学学报(自然科学版)*, 2020, 48(2): 87-94, 101.
Lu DY, Zhang L, Zhang GS, et al. Chloroplast genome structure and variation of Salicaceae plants. *J Northwest A F Univ (Nat Sci Ed)*, 2020, 48(2): 87-94, 101 (in Chinese).
- [16] 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. *世界中医药*, 2020, 15(5): 702-708, 716.
Jiang WJ, Guo MY, Pang XH. Application of chloroplast genome in identification and phylogenetic analysis of medicinal plants. *World Chin Med*, 2020, 15(5): 702-708, 716 (in Chinese).
- [17] 解兵斌. 罗汉果的谱系地理学研究[D]. 桂林: 广西师范大学, 2019.
Xie BB. Phylogeography study of *Siraitia grosvenorii*[D]. Guilin: Guangxi Normal University, 2019 (in Chinese).
- [18] Androsiuk P, Jastrzębski JP, Paukzto Ł, et al. Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. *Sci Rep*, 2020, 10(1): 11522.
- [19] 段义忠, 张凯. 沙冬青属植物叶绿体基因组对比和系统发育分析. *西北植物学报*, 2020, 40(8): 1323-1332.
Duan YZ, Zhang K. Comparative analysis and phylogenetic evolution of the complete chloroplast genome of *Ammopiptanthus*. *Acta Bot Boreali-Occidentalia Sin*, 2020, 40(8): 1323-1332 (in Chinese).

- [20] 罗强, 刘建林, 蔡光泽. 中国柃(木衣)属(*Docynia* Dcne.)一新种——长爪柃(木衣). 植物研究, 2011, 31(4): 389-391.
Luo Q, Liu JL, Cai GZ. *Docynia longiunguis* Q.Luo et J.L.Liu, a new species of *Docynia* Dcne. from China. Bull Bot Res, 2011, 31(4): 389-391 (in Chinese).
- [21] 彭珍华. 云南柃衣有效成分提取纯化工艺的研究[D]. 昆明: 昆明理工大学, 2010.
Peng ZH. Study on extraction and purification process of effective components from *Docynia delavayi*[D]. Kunming: Kunming University of Science and Technology, 2010 (in Chinese).
- [22] 慈晓彤, 石辰, 王大玮, 等. 云南柃[木衣]叶片总RNA提取方法的比较与改进. 西北林学院学报, 2020, 35(3): 95-99.
Ci XT, Shi C, Wang DW, et al. Comparison and improvement of total RNA extraction methods from *Docynia delavayi* leaves. J Northwest For Univ, 2020, 35(3): 95-99 (in Chinese).
- [23] Zhang XY, Yi K, Chen J, et al. Purified phlorizin from *Docynia indica* (wall.) Decne by HSCCC, compared with whole extract, phlorizin and non-phlorizin fragment ameliorate obesity, insulin resistance, and improves intestinal barrier function in high-fat-diet-fed mice. Molecules, 2018, 23(10): E2701.
- [24] 刘海霞, 刘刚, 张晓喻, 等. 柃[木衣]属植物多酚的含量测定与比较. 食品科学, 2014, 35(24): 295-300.
Liu HX, Liu G, Zhang XY, et al. Determination of polyphenol contents in *Docynia* Dcne. Food Sci, 2014, 35(24): 295-300 (in Chinese).
- [25] 韩本勇, 任英. 云南柃依的研究概况. 民营科技, 2014(6):267, 280.
Han BY, Ren Y. An overview study of the *Docynia delavayi*. Private Technology, 2014(6): 267, 280 (in Chinese).
- [26] 刘刚, 吴京, 朱明君, 等. 云南柃[木衣]叶提取物片段抗氧化活性及成分分析. 西南师范大学学报(自然科学版), 2014, 39(9): 73-81.
Liu G, Wu J, Zhu MJ, et al. On evaluation of antioxidant activities and screening of potential antioxidant compounds from *Docynia delavayi* (Franch.) Schneid. Leaves. J Southwest China Norm Univ (Nat Sci Ed), 2014, 39(9): 73-81 (in Chinese).
- [27] 李福寿, 李美珍, 李剑. 云南多依的生物学特性和物候特征研究. 中国园艺文摘, 2010, 26(4): 94-95.
Li FS, Li MZ, Li J. Study of biological characteristics of *Docynia delavayi* (Fr.) Schneid. Chin Hortic Abstr, 2010, 26(4): 94-95 (in Chinese).
- [28] 张鲁杰, 夏秀英, 徐娜, 等. 高效提取越橘成熟组织基因组 DNA 的方法. 华北农学报, 2008, 23(S2): 205-208.
Zhang LJ, Xia XY, Xu N, et al. An efficient method for genomic DNA extraction from mature tissues of blueberry (*Vaccinium* spp.). Acta Agric Boreali-Sin, 2008, 23(S2): 205-208 (in Chinese).
- [29] 陈韵. 菊属、亚菊属叶绿体 DNA 的基因组结构与进化分析[D]. 南京: 南京农业大学, 2017.
Chen Y. Comparative analysis of structural organization and evolution of *Chrysanthemum* and *Ajania* chloroplast genomes[D]. Nanjing: Nanjing Agricultural University, 2017 (in Chinese).
- [30] 郑祎, 张卉, 王钦美, 等. 大花君子兰叶绿体基因组及其特征. 园艺学报, 2020, 47(12): 2439-2450.
Zheng Y, Zhang H, Wang QM, et al. Complete chloroplast genome sequence of *Clivia miniata* and its characteristics. Acta Horti Sin, 2020, 47(12): 2439-2450 (in Chinese).
- [31] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Res, 2017, 45(w1): W6-W11.
- [32] 穆雨农. 基于叶绿体全基因组对中国姬蕨属的系统分类学研究[D]. 哈尔滨: 哈尔滨师范大学, 2020.
Mu YN. Phylogenic and taxonomic research of *Hypolepis* in China based on the complete chloroplast genome[D]. Harbin: Harbin Normal University, 2020 (in Chinese).
- [33] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol, 2013, 30(12): 2725-2729.
- [34] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009, 25(11): 1451-1452.
- [35] Jansen RK, Raubeson LA, Boore JL, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol, 2005, 395: 348-384.
- [36] Wang DW, Shi C, Tang HY, et al. The complete chloroplast genome sequence of *Docynia indica* (Wall.) Decne. Mitochondrial DNA Part B, 2019, 4(2): 3046-3048.
- [37] Zhang SD, Jin JJ, Chen SY, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol, 2017, 214(3): 1355-1367.
- [38] 毕彧. 百合属的比较叶绿体基因组学研究[D]. 长春: 吉林农业大学, 2017.
Bi Y. Comparative chloroplast genomics of the genus *Lilium*[D]. Changchun: Jilin Agricultural University,

- 2017 (in Chinese).
- [39] Jian HY, Zhang SD, Zhang T, et al. Characterization of the complete chloroplast genome of a critically endangered decaploid rose species, *Rosa praelucens* (Rosaceae). *Conserv Genet Resour*, 2018, 10(4): 851-854.
- [40] Zhao YH, Lu DX, Han RB, et al. The complete chloroplast genome sequence of the shrubby cinquefoil *Dasiphora fruticosa* (Rosales: Rosaceae). *Conserv Genet Resour*, 2018, 10(4): 675-678.
- [41] Wang WB, Yu H, Wang JH, et al. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int J Mol Sci*, 2017, 18(11): 2288.
- [42] Kyalo CM, Li ZZ, Mkala EM, et al. The first glimpse of *Streptocarpus ionanthus* (Gesneriaceae) phylogenomics: analysis of five subspecies' chloroplast genomes. *Plants*, 2020, 9(9): 456.
- [43] Cheng H, Li J, Zhang H, et al. The complete chloroplast genome sequence of strawberry (*Fragaria × ananassa* Duch.) and comparison with related species of Rosaceae. *PeerJ*, 2017, 5: e3919.
- [44] Du X, Zeng T, Feng Q, et al. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene*, 2020, 731: 144340.
- [45] An W, Li J, Yang Z, et al. Characteristics analysis of the complete *Wurfbainia villosa* chloroplast genome. *Physiol Mol Biol Plants*, 2020, 26(4): 747-758.
- [46] Dong WL, Wang RN, Zhang NY, et al. Molecular evolution of chloroplast genomes of orchid species: insights into phylogenetic relationship and adaptive evolution. *Int J Mol Sci*, 2018, 19(3): E716.
- [47] Dong W, Liu J, Yu J, et al. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One*, 2012, 7(4): e35071.
- [48] Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res*, 2004, 11(4): 247-261.
- [49] Celiński K, Kijak H, Wiland-Szymańska J. Complete chloroplast genome sequence and phylogenetic inference of the canary Islands dragon tree (*Dracaena draco* L.). *Forests*, 2020, 11(3): 309.
- [50] Chen SF, Ishizuka W, Hara T, et al. Complete chloroplast genome of Japanese larch (*Larix kaempferi*): insights into intraspecific variation with an isolated northern limit population. *Forests*, 2020, 11(8): 884.
- [51] 唐萍, 彭程. 叶绿体基因组进化的速率和方式. *生物学通报*, 2010, 45(6): 8-10.
Tang P, Peng C. The rate and mode of chloroplast genome evolution. *Bull Biol*, 2010, 45(6): 8-10 (in Chinese).
- [52] Park I, Yang S, Kim WJ, et al. Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species identification. *Molecules*, 2019, 24(6): E1038.
- [53] Xu C, Cai X, Chen Q, et al. Factors affecting synonymous Codon usage bias in chloroplast genome of *Oncidium Gower Ramsey*. *Evol Bioinform*, 2011, 7: 271-278.
- [54] 王婧, 王天翼, 王罗云, 等. 沙枣叶绿体全基因组序列及其使用密码子偏性分析. *西北植物学报*, 2019, 39(9): 1559-1572.
Wang J, Wang TY, Wang LY, et al. Assembling and analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias. *Acta Bot Boreali-Occidentalia Sin*, 2019, 39(9): 1559-1572 (in Chinese).
- [55] 王宇, 周俊良, 唐冬梅, 等. 阔叶猕猴桃叶绿体基因组特征及密码子偏好性分析. *种子*, 2020, 39(5): 13-19.
Wang Y, Zhou JL, Tang DM, et al. Analysis of chloroplast genome characteristics and codon preference in broad-leaf kiwifruit. *Seed*, 2020, 39(5): 13-19 (in Chinese).
- [56] 张靖雯. 濒危植物羽叶丁香叶绿体全基因组及系统发育研究[D]. 杨凌: 西北农林科技大学, 2019.
Zhang JW. The complete chloroplast genome and phylogenetic analysis of the endangered species *Syringa pinnatifolia* (Oleaceae)[D]. Yangling, China: Northwest A&F University, 2019 (in Chinese).

(本文责编 郝丽芳)