远志叶绿体基因组序列特征与系统发育分析

王星璐[1]，李慧娟[1]，赵 伟[1]，李彦青[2]，秦雪梅[1]，张福生[1*]

1. 山西大学 中医药现代研究中心，山西 太原 030006
2. 山西中医药大学基础医学院，山西 晋中 030619

摘 要：目的 解析远志 *Polygala tenuifolia* 叶绿体基因组信息序列特征和确定其在远志属的系统位置。**方法** 获得远志叶绿体基因组序列，借助 GeSeq、Chloroplot、MISA、REPute、Tandem repeats finder、CodonW、Geneious、IRscope、MAFFT 7 和 IQtree2.0.5 等生物信息学工具进行序列分析、密码子偏好分析、远志属基因组比较分析和系统发育研究。**结果** 远志的叶绿体基因组全长 165 423 bp，为典型的环状四段式结构；包括 1 个大单拷贝区（large single copy，LSC；83 699 bp），1 个小单拷贝区（small single copy，SSC；8044 bp）和 1 对反向重复区（inverted repeats，IRs；36 840 bp），该基因组共注释到 135 个基因，包括 8 个 rRNA 基因、38 个 tRNA 基因和 89 个蛋白编码基因。该基因组中共检测到 161 个 SSR 位点，223 条散在重复序列，90 条串联重复序列；亮氨酸（Leu）是远志叶绿体基因组中使用次数最高的氨基酸（10.21%），同义密码子相对使用频次（RSCU）＞1 的密码子有 35 种且均以 A/U 结尾；序列长度、基因组成以及 GC 含量等方面相对保守。IR 边界分析除了西南远志 *Polygala crotalarioides* 和密花远志 *Polygala karensium* 外，其余远志属边界高度保守；基于远志科 9 个物种的叶绿体基因组数据同时选择远志属的近缘类鸦胆子 *Brucea javanica* 和九里香 *Murraya exotica* 作为外类群，采用最大似然法和邻接法分别构建系统发育树，支持率均为 100，显示远志 *P. tenuifolia* 与卵叶远志 *Polygala sibirica*、瓜子金 *Polygala japonica* 和香港远志 *Polygala hongkongensis* 聚类到一起，能够反映远志属的亲缘关系。**结论** 远志与卵叶远志、瓜子金和香港远志的亲缘关系最近。首次采用生物信息学分析方法对远志叶绿体基因组进行了全面、深度解析，结果将为远志属药用植物的遗传多样性分析以及新品系的遗传育种研究等提供理论依据。

关键词：远志；叶绿体基因组；序列特征；密码子偏好性；系统发育分析

中图分类号：R286.12 **文献标志码：**A **文章编号：**0253 - 2670(2023)11 - 3655 - 11

DOI: 10.7501/j.issn.0253-2670.2023.11.027

Characterization and phylogenetic analysis of complete chloroplast genome of *Polygala tenuifolia*

WANG Xing-lu[1], LI Hui-juan[1], ZHAO Wei[1], LI Yan-qing[2], QIN Xue-mei[1], ZHANG Fu-sheng[1]

1. Modern Research Center for Traditional Chinese Medicine, Shanxi University, Taiyuan 030006, China
2. College of Basic Medical Sciences, Shanxi University of Chinese Medicine, Jinzhong 030619, China

Abstract: Objective To analyze the chloroplast genome information of *Polygala tenuifolia* and determine its phylogenetic position in *Polygala* genus. **Methods** The chloroplast genome sequence of *P. tenuifolia* was obtained. Bioinformatics tools of GeSeq, Chloroplot, MISA, REPute, Tandem repeats finder, CodonW, Geneious, IRscope, MAFFT 7 and IQtree2.0.5 were used for sequence analysis, codon preference analysis, genome comparative analysis and phylogenetic study of *Polygala*. **Results** The chloroplast genome of *P. tenuifolia* was 165 423 bp in length, which was a typical circular four-segment structure, including a large single copy region (LSC; 83 699 bp), a small single copy region (SSC; 8044 bp) and a pair of reverse repeat region (IRs; 36 840 bp). The total of 135 genes were annotated, including eight rRNA genes, 38 tRNA genes and 89 protein coding genes. The total of 161 SSR location, 223 scattered repeats and 90 tandem repeats were detected in the genome. Leucine (Leu) was the most frequently used amino acid

收稿日期：2022-10-03

基金项目：山西省科技厅应用基础研究项目（201901D111039）；山西省科技成果转化引导专项（202104021301064）；吕梁市引进高层次科技人才重点研发项目（2021RC-2-13）；山西省基础研究计划（20210302123312）

作者简介：王星璐（1995—），女，河北石家庄人，硕士研究生，研究方向药用植物次生代谢物调控研究。E-mail: 2642421651@qq.com

***通信作者：**张福生（1978—），男，博士，副教授，硕士生导师，从事中草药次生代谢物的生物合成途径研究。Tel: (0351)7019178 E-mail: ample1007@163.com

(10.21%) in the chloroplast genome of *P. tenuifolia*, and there were 35 codons with relative synonymous codon usage (RSCU)＞1, all of which ended in A/U, the chloroplast genome of was relatively conservative in terms of sequence length, gene composition and GC content. Except *P. crotalarioides* and *P. karensium*, the boundaries of other genera were highly conservative by IR boundary analysis. Based on the chloroplast genome data of nine species in the family Polydiaceae, we selected the related member of the genus *Polygala*: *Brucea javanica* and *Murraya exotica* as the outgroup. The phylogenetic trees were constructed by maximum likelihood and neighbor-joining method, with support rates of 100. The results showed that *P. tenuifolia* was clustered with *P. sibirica*, *P. japonica* and *P. hongkongensis*, which could reflect the genetic relationship of *Polygala*. **Conclusion** *P. tenuifolia* had the closest relationship with *P. sibirica*, *P. japonica*, *P. hongkongensis*. In this study, the chloroplast genome of *P. tenuifolia* was analyzed comprehensively and deeply for the first time by using bioinformatics analysis method, which will provide theoretical basis for genetic diversity analysis of medicinal plants and genetic breeding research of new strains of *Polygala*.

Key words: *Polygala tenuifolia* Willd.; chloroplast genome; sequence characterization; codon preference; phylogenetic analysis

在植物中核基因组的复杂性使得低拷贝基因的筛选比较困难，目前仅局限在基因组相对较小的模式物种或具有重要经济生态价值的物种中，整体应用范围较为有限。植物线粒体基因组具有在植物类群中变异很大、存在基因组之间水平转移的外源基因和进化速率较慢等特性，使其在系统发育研究中的应用受到了限制[1]。叶绿体基因组作为第 2 大基因组，属于相对独立于核基因组之外的基因组[2]，其基因组完整且相对独立。大多数被子植物的叶绿体含有大量遗传信息，同时具有相对分子质量小、结构简单，进化速率中等、突变率较低、遗传稳定、成本低及开发难度低、广泛分布着微卫星序列和线粒体 DNA 等特点[3-4]，在一定程度上弥补了线粒体和核基因组的部分缺点与不足。

叶绿体起源于古细菌入侵植物细胞[5]，是一种拥有自身遗传物质的多功能细胞器[6]，为质体中的一种类型。叶绿体基因组的结构大部分为共价闭合的双链状环形分子，也有少部分是线性分子或者多聚体。对于高等植物而言，叶绿体存在于细胞质的基质中，其形状呈现为绿色椭圆球型或者是圆球形[7]。典型的环式双链叶绿体基因组结构，包括 4 个部分，即大单拷贝区（large single copy，LSC），小单拷贝区（small single copy，SSC）和 2 个反向重复区（inverted repeats，IRs）；其中，2 个反向重复区把基因组分隔为大单拷贝区和小单拷贝区[8]。LSC 区长 81～90 kb，SSC 区介于 18～20 kb；2 个 IR 区序列基本一致，大小介于 20～30 kb，是叶绿体基因组进化过程中延展或缩小的区域[9]。大多数陆生植物的叶绿体基因组中有 110～130 个基因[10]。近年来，由于叶绿体基因组大小、结构和基因种类一般较为保守，叶绿体全基因组的物种鉴定及系统进化研究成为植物系统分类学的一个新趋势，为研究药用植物系统进化及进行物种鉴定提供可靠工具[11]。目前，叶绿体基因组技术已经广泛应用于药用植物的研究中[12-14]。

远志植物为双子叶植物芸香目远志科一种一年生或多年生草本、灌木或小乔木，约 500 种，广布于全世界，我国有 42 种 8 变种，广布于全国各地，而以西南和华南地区最盛[15]。远志属远志 *Polygala tenuifolia* Willd.最早记载于《华氏中藏经》[16]，药用价值广泛，历史悠久，具有较大的药用研究和开发利用价值。近年来由于过度采挖远志药材，导致野生药材资源严重不足，远志已被收入《国家重点保护野生药材物种名录》，保护级别为III级。目前关于远志的研究主要集中于其化学成分与药理作用等方面。研究发现，远志具有改善学习记忆、抗氧化、抗抑郁、抗衰老、镇静催眠、益智和抗肿瘤以及祛痰镇咳、影响药物代谢以及抗炎等[17-24]作用。分子生物方面研究主要集中在远志随机扩增多态性 DNA 分析[25]、遗传多样性[26]、远志谱系地理学[27-28]、序列鉴定[29-30]等方面；对叶绿体基因组的研究报道较少，虽有对西南远志 *P. crotalarioides* Buch.-Ham. ex DC.[31]和远志 *P. tenuifolia* Willd.[32]叶绿体基因组进行测序、拼装和注释的报道，但是报道的远志[32]仅限于描述了序列全长和系统发育位置等一些基本信息，未对其展开详细的分析描述。

目前，远志属物种的叶绿体基因组研究十分缺乏，其相关的亲缘关系值得深入研究，NCBI 数据库中有远志 *P. tenuifolia* Willd.（NC_050829.1）[32]完整叶绿体基因组数据，但是仅记载了叶绿体基因组总长（165 423 bp）等基础信息，也未进行深入挖掘研究，相关信息依旧不清楚。本研究利用生物信息学相关软件，分析其叶绿体基因组的构成，并对

其进行基因组装和注释，分析序列特征，密码子偏好性以及系统发育，阐明远志叶绿体基因组结构特征以及其物种之间的亲缘关系。远志叶绿体基因组的研究是对远志药材种质资源遗传多样性的进一步了解，有利于远志药材优良品种的选育和种质资源的评价及其有效保护与合理利用。

1 材料与方法

1.1 样品收集

从美国国立生物技术信息中心（National Center of Biotechnology Information，NCBI）数据库（https://www.ncbi.nlm.nih.gov/）检索远志科远志属物种的叶绿体全基因组序列信息。检索到远志科远志属 8 个物种以及齿果草属 1 个叶绿体全基因组序列信息，远志科亲缘关系较近的芸香目苦木科鸦胆子 *Brucea javanica* (L.) Merr. 和芸香科九里香 *Murraya exotica* L. Mant.完整叶绿体全基因组序列。下载检索到的物种的叶绿体基因组名称、基因组登录号（表 1）。

表 1 远志科、苦木科和芸香科叶绿体 GenBank 登录号

Table 1 GenBank accession number of chloroplast in Polygalaceae, Simaroubaceae and Rutaceae

种	GenBank 号	种	GenBank 号
远志 *P. tenuifolia*	NC_050829.1	瓜子金 *P. japonica*	NC_052912.1
卵叶远志 *P. sibirica*	NC_056970.1	黄花倒水莲 *P. fallax*	NC_052911.1
黄花远志 *P. arillata*	MN243714.1	西南远志 *P. crotalarioides*	NC_060367.1
香港远志 *P. hongkongensis*	MZ707521.1	密花远志 *P. karensium*	NC_056968.1
齿果草 *Salomonia cantoniensis*	NC_056969.1	九里香 *Murraya exotica*	MW722359.1
鸦胆子 *Brucea javanica*	NC_063730.1		

1.2 叶绿体基因组注释、图谱绘制以及基本特征分析

远志属远志的叶绿体全基因组序列（登录号为 NC_050829.1）通过 GeSeq：（https://chlorobox.mpimp-golm.mpg.de/geseq.htm）[33]和 Plastid Genome Annotator（PGA）[34]软件进行基因注释，将结果对比矫正，去除错误及冗余注释。通过 Chloroplot（https://irscope.shinyapps.io/Chloroplot/）[35]在线绘制工具绘制。通过 GeSeq 初步注释和 Chloroplot 画图相结合对叶绿体基因组的总长度及各个区域（LSC、SSC、IR）的长度、基因组成（蛋白编码基因、tRNA 基因、rRNA 基因）、碱基组成、GC（AT）含量进行统计和比较分析，解析远志叶绿体基因组序列的基本特征。

1.3 叶绿体全基因组重复序列检测

利用 Perl 语言通过 MISA 软件（http://pgrc.ipk-gatersleben.de/misa/misa.html）[36]完成简单重复序列（simple sequence repeats，SSRs）检测，参数设置为单核苷酸重复单元不少于 10 个，二核苷酸重复单元不少于 5 个，三核苷酸和四核苷酸重复单元不少于 4 个，五核苷酸和六核苷酸重复单元不少于 3 个，且 2 个 SSRs 之间的最小距离为 100 bp[37]。

叶绿体全基因组中的散在重复序列（dispersed repeats）利用 REPute 软件（https://bibiserv.cebitec.uni-bielefeld.de/reputer）[38]进行检测，正向重复（forward repeats，F）、反向重复（reverse repeats，R）、互补重复（complement repeats，C）、回文重复（palindromic repeats，P）。参数设置最小重复长度（minimal repeat size）设置为 30，hamming 距离（hamming distance）为 3，最大计算重复次数（maximum computed P repeats）5000[39]。串联重复序列（tandem repeats）利用 Tandem repeats finder 软件（https://tandem.bu.edu/trf/trf.html）进行检测。参数设置选择默认值[40]。

1.4 密码子使用分析

采用 CodonW（http://codonw.sourceforge.net）软件[41]分析密码子使用情况。对远志的叶绿体基因组同义密码子相对使用频次（relative synonymous codon usage，RSCU）进行分析和统计。当 RSCU＞1 时，表明该密码子使用频率较高；RSCU＝1 时，说明该密码子无偏好性；RSCU＜1 时，表明该密码子使用频率较低[42]。

1.5 基因组比较分析

采用 Geneious[43]软件统计远志属 8 个物种的叶绿体基因组序列的 4 个边界（SSC、LSC、IRa 和 IRb 区域）长度和基因数目类型、GC 含量等信息，并用 EXCEL 计算各自的 GC 值。使用 IRscope（https://irscope.shinyapps.io/irapp/）[44]可视化工具，比较远志属叶绿体基因组 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。

1.6 基于叶绿体基因组序列的系统进化分析

为了确定远志的系统发育位置，下载 NCBI 数据库中收录的远志科上述所有 9 种植物的叶绿体全基因组序列。与本研究远志叶绿体全基因组序列共同构建序列矩阵，同时选择远志属的近缘类鸦胆子和九里香作为外类群。利用 MAFFT 7 软件（https:// mafft.cbrc.jp/alignment/software/）[45]完成序列比对。利用 IQtree2.0.5

（http://www.iqtree.org/）[46]以最大似然法（maximum likelihood，ML）进行系统发育关系重建分析，核苷酸替换模型选择 GTR＋G，其他参数选择默认设置，进行 1000 次自展重复分析，计算系统发育树各节点的靴带支持率（bootstrap values，BS）。并利用 MEGA X 软件，使用邻接法（neighbor-joining，NJ）构建系统进化树，并进行系统发育树的比较。

2　结果与分析

2.1　远志叶绿体基因组的基本特征

远志的叶绿体基因组与典型的陆地植物叶绿体基因组基本特征一样，均为典型的四分体结构，远志属远志叶绿体基因组呈双链环状四分体结构（图 1），大小为 165 423 bp，包括大单拷贝区、小单拷贝区以及 2 个反向重复区，各区长度分别为 83 699、8044、36 840 bp。远志叶绿体基因组的总 GC 量为 36.7%，IR 区的 GC 量为 39.7%，高于 LSC 区（34.7%）及 SSC 区（29.4%）。整体而言，IR 区的保守性高于 LSC 区和 SSC 区，这可能是 IR 区的 GC 量高于 LSC 区和 SSC 区而造成的。计算各分区的碱基组成（表 2）。

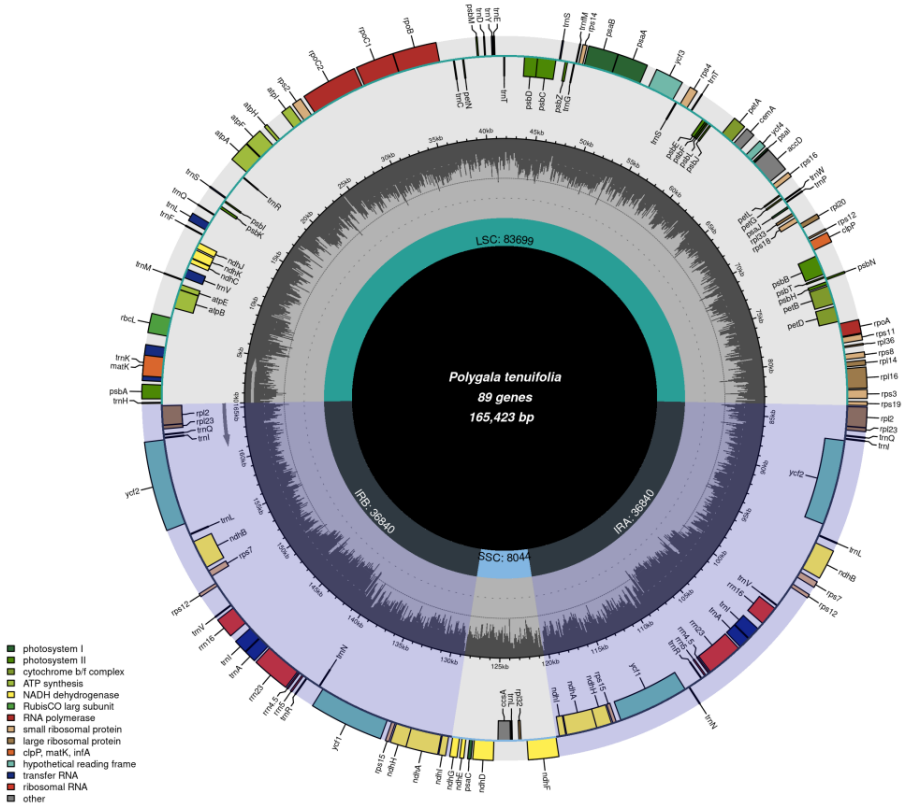


图 1　远志叶绿体基因组图谱

Fig. 1　Chloroplast genome map of *P. tenuifolia*

表 2　远志叶绿体基因组碱基组成

Table 2　Base composition of chloroplast genome of *P. tenuifolia*

区域	A/%	T/%	C/%	G/%	长度/bp	GC/%
LSC	32.0	33.3	18.0	16.7	83 699	34.7
SSC	38.6	32.3	15.0	14.4	8044	29.4
IRa	29.2	31.1	20.9	18.9	36 840	39.7
IRb	31.1	29.2	18.8	20.9	36 840	39.7
总数	31.5	31.8	18.7	18.0	165 423	36.7

2.2　远志叶绿体基因组的基因注释及归类分析

远志叶绿体基因组共注释得到 135 个基因，包括 8 个 rRNA 基因、38 个 tRNA 基因和 89 个蛋白编码基因。根据其功能可分为以下 4 个大类（表 3）。第 1 类：48 个与光合作用相关的基因，包括 5 个光合系统I基因、15 个光合系统 II 基因、6 个细胞色素 b/f 复合体基因、6 个 ATP 合酶基因、1 个二磷酸核糖体羧化酶、15 个 NADH 氧化还原酶基因。第 2 类：除 rRNA 基因和 tRNA 基因外，有 30 个与自我复制相关的基因，包括 10 个核糖体蛋白大亚基基因、16 个核

表 3 远志叶绿体基因组基因功能注释与分类

Table 3 Gene functional annotation and classification of *P. tenuifolia* chloroplast genome

基因作用分类	基因分组	基因名称
光合作用相关基因	光合系统I基因	*psaA*、*psaB*、*psaC*、*psaI*、*psaJ*
	光合系统II基因	*psbA*、*psbB*、*psbC*、*psbD*、*psbE*、*psbF*、*psbL*、*psbM*、*psbH*、*psbI*、*psbJ*、*psbK*、*psbN*、*psbT*、*psbZ*
	细胞色素 b/f 复合物基因	*petA*、*petB*、*petD*、*petG*、*petL*、*petN*
	ATP 合酶基因	*atpA*、*atpB*、*atpE*、*atpF*、*atpH*、*atpI*
	二磷酸核糖体羧化酶	*rbcL*
	NADH 氧化还原酶	*ndhA*（2）、*NdhB*（2）、*ndhC*、*ndhD*、*ndhE*、*ndhF*、*ndhG*、*ndhH*（2）、*NdhI*（2）、*ndhJ*、*ndhK*
自我复制基因	RNA 聚合酶	*rpoA*、*rpoB*、*rpoC1*、*rpoC2*
	核糖体蛋白大亚基基因	*rpl2*（2）、*rpl14*、*rpl16*、*rpl20*、*rpl23*（2）、*rpl32*、*rpl33*、*rpl36*
	核糖体蛋白小亚基基因	*rps2*、*rps3*、*rps4*、*rps7*（2）、*rps8*、*rps11*、*rps12*（3）、*rps14*、*rps15*（2）、*rps16*、*rps18*、*rps19*
	转运 RNA	*trnA*（2）、*trnC*、*trnD*、*trnE*、*trnF*、*trnH*、*trnG*、*trnI*（4）、*trnK*、*trnL*（4）、*trnM*（2）、*trnN*（2）、*trnP*、*trnQ*（3）、*trnR*、*trnS*（3）、*trnT*（2）、*trnR*（2）、*trnV*（3）、*trnW*、*trnY*
	核糖体 RNA	*rrn4.5*（2）、*rrn5*（2）、 *rrn16* （2）、*rrn23*（2）
其他基因	成熟酶基因	*MatK*
	包裹膜蛋白基因	*cemA*
	蛋白酶	*clpP*
	乙酰辅酶 A 基因亚基	*accD*
	蛋白酶基因	*clpP*
	成熟酶基因	*ccsA*
未知基因	开放阅读框	*ycf1*（2）、*ycf2*（2）、*ycf3*、*ycf4*

糖体蛋白小亚基基因和 4 个 RNA 聚合酶亚基基因。第 3 类：6 个其他编码蛋白质的基因。第 4 类：6 个功能未知的基因。在这些基因中，有 12 个蛋白编码基因（*rpl2*、*rpl23*、*ycf2*、*rps12*、*rps7*、*rps15*、*ycf1*、*ndhA*、*ndhB*、*ndhH*、*ndhI*、*ndhN*）、7 个 tRNA 编码基因（*trnQ*、*trnI*、*trnL*、*trnV*、*TrnA*、*trnN*、*trnR*）和 4 个 rRNA 编码基因（*rrn4.5*、*rrn5*、*rrn16*、*rrn23* ）位于 IR 区。

2.3 重复序列统计分析

叶绿体基因组 SSR 分析在远志叶绿体基因组中搜索到 161 个 SSR 位点。其中，29 个为复合型 SSRs；单核苷酸重复 Motif 位点最多为 113 个，二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复基序分别有 30、3、9、4、2 个（表 4）。SSR 的类型以 A/T 为主，无含有 G/C 组成的位点，占总重复类型的 70.18%；其次为 AT/AT，共有 25 个（15.52%），AG/CT（3.11%），其余总数占总重复类型的 11.18%。A/T 和 AT/AT 占简单重复序列位点重复单元总数量的 85.71%，因此远志叶绿体基因组富含 AT。从分布区段上看，34 个位点位于 IR 区段，115 个位于 LSC 区段，12 个位于 SSC 区段，说明远志叶绿体基因组 SSRs 分布的不均匀性。

使用 REPuter 软件在远志叶绿体基因组中共检测到 223 个散在重复序列，包括正向重复序列（F 型）、反向重复（R 型）、回文重复序列（P 型）和互补重复（C 型）4 种类型。其中，正向重复序列（F 型）89 个，反向重复（R 型）13 个，回文重复序列（P 型）109 个和互补重复（C 型）12 个。长度为 30～39 bp 的重复序列中包括 61 个正向重复，12 个反向重复，72 个回文重复和 12 个互补重复。其分布数量相对较多（图 2）。发现 90 条长度范围在 2～54 bp 的串联重复序列。串联重复序列，4 个区域均有分布，其中，LSC 区 35 条，重复次数 1.9～26.8 次，TRB 区 24 条，重复次数由 1.9～12.2 次，SSC 区 1 条，重复次数由 2 次，TRA 区 30 条，重复次数由 1.9～8.1 次。

2.4 叶绿体基因组密码子偏好性分析

用 CodonW 软件分析远志叶绿体基因组蛋白编

表 4　远志叶绿体基因组 SSRs 位点类型及数量

Table 4　Type and number of SSRs of *P. tenuifolia* chloroplast genome

重复类型	重复序列	重复序列个数	总数
单核苷酸	T	63	113
	A	50	
二核苷酸	AT	13	30
	TA	12	
	AG	2	
	CT	2	
	TC	1	
三核苷酸	TCT	1	3
	AAT	1	
	TTA	1	
四核苷酸	ATTT	2	9
	AGAA	1	
	AAAT	1	
	AAGG	1	
	TAAA	1	
	TTTA	1	
	ATCC	1	
	GGAT	1	
五核苷酸	TTATT	1	4
	TTAGA	1	
	TTTTA	1	
	ATAAA	1	
六核苷酸	TGAAGA	1	2
	TCTTCA	1	

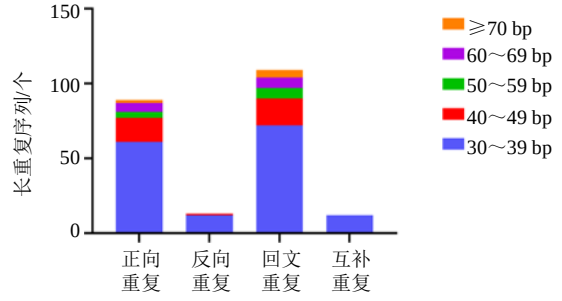


图 2　远志叶绿体基因组散在重复序列类型与数量

Fig. 2　Type and number of long repeats in *P. tenuifolia* chloroplast genome

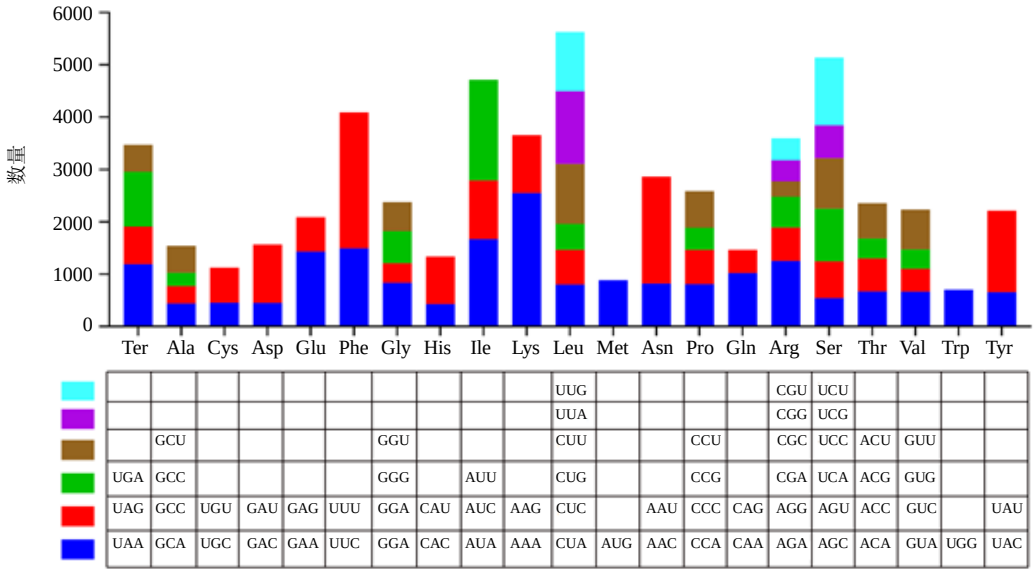
码序列密码子的使用偏好性（图 3），共预测了 55 141 个密码子，其中编码亮氨酸（Leu）的密码子数量最多（5629 个，10.21%），编码色氨酸（Trp）的密码子数量最少（705 个，1.28%）。在远志叶绿体基因组中有 20 种氨基酸，除甲硫氨酸（Met）和色氨酸（Trp）使用 1 个密码子 AUG 和 UGG 外，其余氨基酸均有 2～6 个同义密码子。亮氨酸（Leu）、精氨酸（Arg）、丝氨酸（Ser）使用 6 个同义密码子，使用频率最高的密码子分别为 UUA、UCU、AGA；丙氨酸（Ala）、甘氨酸（Gly）、脯氨酸（Pro）、苏氨酸（Thr）、缬氨酸（Val）分别使用 4 个同义密码子，使用频率最高的密码子分别为 GGA、CCA、GUU、CCU、ACU；异亮氨酸（Ile）使用 3 个同义密码子，使用频率最高的密码子是 AUU；半胱氨酸（Cys）、天冬氨酸（Asp)、谷氨酸（Glu）、苯丙氨酸（Phe）、组氨酸（His）、赖氨酸（Lys）、天冬酰胺（Asn）、谷氨酰胺（Gln）、酪氨酸（Tyr）均使用 2 个同义密码子，但使用频率并不相同，这 9 个氨基酸使用频率较高的密码子分别为 UUU、AAA、AAU、UAU、UUC、GAA、GAU、AAG、CAA。结果表明，远志叶绿体基因偏好使用含有 A、U 碱基的密码子，密码子第 3 位也偏好以 A 和 U 结尾。

相对同义密码子使用度（RSCU）分析表明，在所有被编码的密码子中，RSCU＞1 的密码子共有 35 个（AGA、UCU、UUA 等），其中 27 个密码子以 A/U 碱基结尾，以 A 结尾的占 51.85%，以 U 结尾的占 48.15%，故这些密码子的结尾具有 A/U 偏好性。RSCU＜1 的密码子共有 29 个（CGC、AAC、GAC 等），以 G/C 结尾的共有 23 个，且以 C 结尾占 52.13%和以 G 结尾的占 47.83%（表 5）。

2.5　基因组特征及 IR 边界比较分析

从 NCBI 下载了远志、瓜子金、卵叶远志、黄花倒水莲、黄花远志、西南远志、香港远志和密花远志 8 种远志属植物的叶绿体基因组，它们的全长分别为 165 423、165 439、165 397、164 687、164 947、164 268、165 296、171 893 bp。黄花倒水莲叶绿体基因组的 GC 值最大，为 36.9%，其余 GC 含量均在 36.7%～36.9%（表 6）。瓜子金的 LSC 最长，为 83 722 bp、密花远志的 SSC 最长为 8409 bp。

在远志属 8 种植物的叶绿体基因组边界区域中，远志叶绿体基因组共存在 4 个边界，远志与其他 7 种植物的 LSC/IRb、IRb/SSC、SSC/IRa 和 IRa/LSC 边界及基因分布（图 4）。除密花远志和西南远志外，JLB（LSC/IRb）边界两侧分布均有 *rpsl2* 和 *rps19* 基因，密花远志边界位于 *petB* 基因中，是其独有的基因，该基因的保守长度为 1390 bp，西南远志 JLB 边界有 *rpl23* 基因存在，距离边界 184 bp。其中远志中的 *rps19* 基因距离 JLB（LSC/IRb）边界长度为 1 bp，其位于 LSC 区域内，*rps12* 位于 IRb 区域内。



x 轴代表密码子家族，密码子使用频率绘制在 *y* 轴上

the x-axis represents codon families, frequency of codon usage is plotted on the y-axis.

图 3　远志叶绿体基因组的密码子使用频率

Fig. 3　Codon usage frequency in chloroplast genome of *P. tenuifolia*

表 5　远志绿体基因组蛋白编码序列 RSCU 分析

Table 5　RSCU of protein coding region in chloroplast genome of *P. tenuifolia*

氨基酸	密码子	数目	RSCU	氨基酸	密码子	数目	RSCU
Ala	GCA*	440	1.14	Pro	CCA*	809	1.25
	GCC	333	0.86		CCC*	656	1.01
	GCG	252	0.65		CCG	422	0.65
	GCU*	516	1.34		CCU*	699	1.08
Cys	UGC	454	0.81	Gln	CAA*	1022	1.40
	UGU*	671	1.19		CAG	443	0.60
Asp	GAC	452	0.58	Arg	AGA*	1251	2.08
	GAU*	1114	1.42		AGG*	640	1.07
Glu	GAA*	1430	1.37		CGA	589	0.98
	GAG	661	0.63		CGC	293	0.49
Phe	UUC	1495	0.73		CGG	402	0.67
	UUU*	2599	1.27		CGU	426	0.71
Gly	GGA*	837	1.41	Ser	AGC	543	0.63
	GGC	371	0.62		AGU	701	0.82
	GGG*	614	1.03		UCA*	1006	1.17
	GGU	556	0.94		UCC*	964	1.13
His	CAC	431	0.65		UCG	629	0.73
	CAU*	904	1.35		UCU*	1296	1.51
Ile	AUA*	1669	1.06	Thr	ACA*	671	1.14
	AUC	1121	0.71		ACC*	627	1.06
	AUU*	1923	1.22		ACG	381	0.65
Lys	AAA*	2550	1.39		ACU*	678	1.15
	AAG	1109	0.61	Val	GUA*	666	1.19
Leu	CUA	802	0.85		GUC	435	0.78
	CUC	661	0.7		GUG	371	0.66
	CUG	500	0.53		GUU*	763	1.37
	CUU*	1143	1.22	Trp	UGG*	705	1.00
	UUA*	1388	1.48	Tyr	UAC	656	0.59
	UUG*	1135	1.21		UAU*	1559	1.41
Asn	AAC	822	0.57	终止密码子 TER	UAA*	1191	1.21
	AAU*	2043	1.43		UAG	719	0.73
Met	AUG*	887	1.00		UGA*	1045	1.06

*RSCU＞1 高频密码子

*RSCU > 1 high frequency code

表 6　8 种远志属植物叶绿体基因组的特征

Table 6　Chloroplast genome characteristics of eight *Polygala* plants

植物名称	叶绿体基因组长度/bp	GC 值/%	LSC 长度/bp	SSC 长度/bp	IR 长度/bp
远志	165 423	36.7	83 699	8044	73 680
瓜子金	165 439	36.7	83 722	8145	73 572
卵叶远志	165 397	36.7	83 689	8118	73 590
黄花倒水莲	164 687	36.9	83 352	8283	73 052
黄花远志	164 947	36.8	83 537	8210	73 200
西南远志	164 268	36.7	82 831	8246	73 191
香港远志	165 296	36.7	83 707	8035	36 777
密花远志	171 893	36.8	73 594	8409	89 890

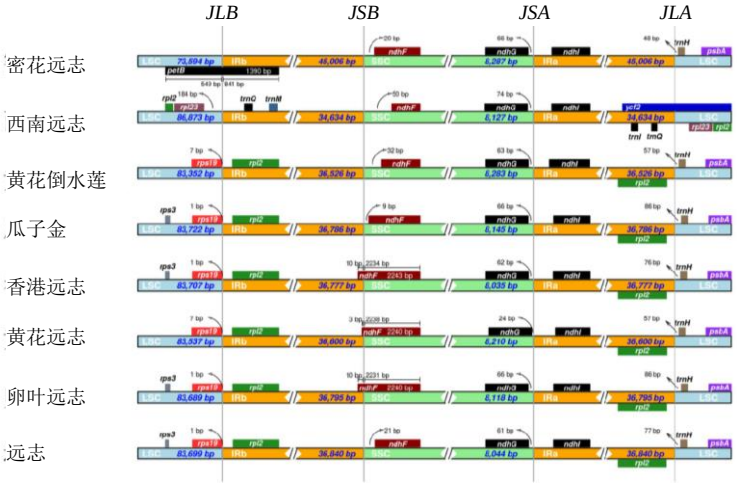


图 4　远志属 8 个物种叶绿体基因组的 IR 与 SC 边界比较

Fig. 4　Boubdary comparison of IR and SC region of chloroplast genomes in nine species of *Polygala*

ndh F 基因位于 JSB（IRb/SSC）的边界，8 种植物在该位置与基因存在间隙或重叠存在。其中该基因在远志中距 SSC 区仅 21 bp，其中香港远志和卵叶远志 *ndh F* 扩张最大。香港远志、黄花远志和卵叶远志的 JSB（IRb/SSC）边界均位于 *ndh F*，其余远志属物种均在 SSC 区域，该基因相对保守长度均为 2240 bp。另一个 *ndhG* 基因位于 JSA（SSC/IR）交界处，均位于 SSC 区域中。除了黄花远志距离边界 24 bp，其余距离边界 60～74 bp。

ndh I 基因位于 JSB（IRb/SSC）的边界，均位于 IRa 区域中。西南远志 JLA（IRa/LSC）边界分布有 *ycf2* 基因，密花远志分布有 *trnH* 基因。而其他 6 种植物在该处均为 *trnH* 和 *rpl2* 基因。其中，*trnH* 基因均分布在 LSC 区域，*rpl2* 基因均分布在 IRa 区域。IRA-LSC 边界均与 *trnH* 基因存在着 48～86 bp 的间隙。总体来说，远志属的西南远志和密花远志叶绿体基因组和边界基因与其余 6 个叶绿体基因组相互之间具有明显的差异。

2.6　基于叶绿体基因组的远志系统发育分析

利用最大似然法构建系统发育分析结果显示，11 种植物被分为 2 大类，即远志科和外类群（苦木科和芸香科），见图 5。包括远志在内的远志属 8 个物种和一个齿果草属以 100%的支持率构成一个单系分支与芸香科和苦木科区分开来。同时，这 9 个物种可以进一步划分为 2 个次级单系分支，其中，同来自远志属的黄花远志、黄花倒水莲、密花远志和齿果草属齿果草 4 个物种构成一个单系的支持率为 100%，齿果草属和部分远志属聚类到一块值得进一步研究，侧面说明叶绿体基因组序列也并不能完全解决远志属内部网状进化等问题，仍需和其他基因组学联合分析。结合 IR 边界分析可知远志、瓜子金、卵叶远志和香港远志的亲缘关系最近。本研究的远志叶绿体基因组序列可为远志科后续开展遗传多样性研究提供重要信息。

邻接法的结果与最大似然法构树除了齿果草和密花远志位置互换外，其余结果一致，见图 6。远志科仍分为 2 大类同来自远志属的远志和卵叶远志、瓜子金和香港远志 4 个物种分为一类，黄花倒水莲、黄花远志、密花远志和齿果草属齿果草聚到一类。仍然可以说明远志和卵叶远志、瓜子金和香港远志关系密切。

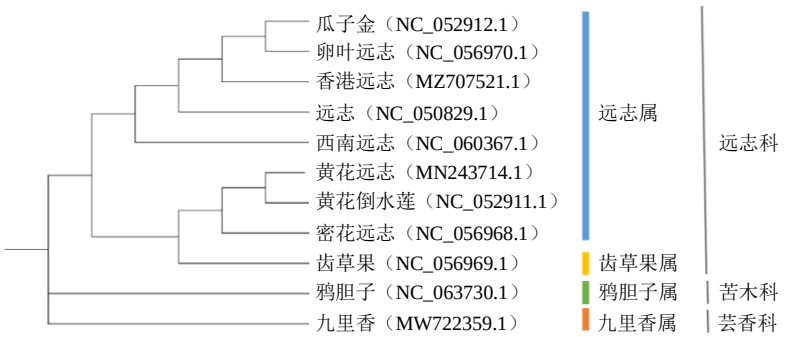


图 5　基于叶绿体全基因组序列构建 ML 系统进化树

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

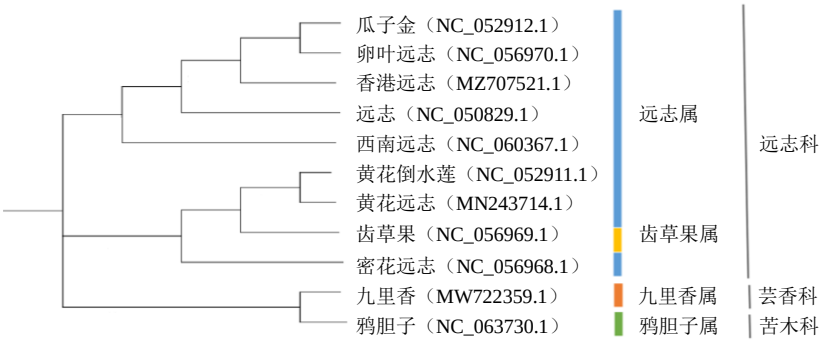


图 6　基于叶绿体全基因组序列构建 NJ 系统进化树

Fig. 6　Phylogenetic analysis based on chloroplast genome sequences by neighbor-joining (NJ) tree

3　讨论

《中国药典》2020 年版中将远志科植物远志 *P. tenuifolia* L.的干燥根作为远志的主流基原药材[47]进行商品流通，其药用价值显著。近年来，面临远志的野生资源状况不容乐观，植物资源日益贫乏的危险。本实验基于叶绿体基因组技术进行序列分析、密码子偏好分析、远志属基因组比较分析以及远志属植物系统发育关系研究为远志在野生资源保护、分子育种及远志属药用植物的物种鉴定等领域提供了宝贵的基因资源。

本实验结果发现，远志叶绿体基因组长为 165 423 bp，GC 含量为 36.7%，共编码 135 个基因，这与先前报道的远志属其他植物西南远志[31]的叶绿体基因组结构特征类似，表明远志的叶绿体基因组结构相对保守。与韩国学者 Lee 等[32]发表的文章相比较，本研究将系统发育树聚焦到远志科植物上，且增加对其重复序列，密码子偏好性，远志属基因组比较，IR 边界内容的分析，补充和完善前者文章分析量少的缺点，从而对其叶绿体基因组有了更加全面的了解。

SSR 在植物叶绿体基因组中十分常见，其类型、数目及分布都因植物不同而异，被广泛应用于植物群体遗传多样性和系统发育研究以及分子标记研究等内容[48-49]。本研究通过分析远志叶绿体基因组，共检测到 161 个 SSR 位点，其主要位于 LSC 区域；并且单核苷酸 A/T 碱基在 SSR 位点中出现频率最高，调研文献推测原因是在大部分植物中 A/T 含量均高于 G/C 含量，即 A/T 类型的 SSR 在植物中最多[37,39,42]。本研究通过对远志叶绿体基因组中 SSR 的数量、组成进行分析，为后续进一步研究分子标记、群体遗传分析以及作物育种提供参考。

CUB 是基因组中重要的进化特征，RSCU 是作为密码子偏好性的指标之一[50]。研究表明，自然界存在的 20 种氨基酸中，除 Met 和 Trp 由唯一密码子编码外，其他氨基酸均对应 2～6 个同义密码子，由于同义密码子在机体内的使用频率存在差异，导致植物密码子的出现频率不同，在不同物种翻译的过程中。存在突变和自然选择等多种因素，表现出一定的偏好性[51]。本研究中密码子偏好性分析表

明，亮氨酸（Leu）是远志叶绿体基因组中占比最高的氨基酸，并且 RSCU＞1 的 27 种密码子均以 A/U 结尾，这与其他高等植物相似。

通过对远志属基因组特征分析，可知黄花倒水莲叶绿体基因组的 GC 值最大，为 36.9%，其余 GC 含量均在 36.7%～36.9%。瓜子金的 LSC 最长为 83 722 bp、密花远志的 SSC 最长为 8409 bp，通过 IR 边界分析远志、卵叶远志、瓜子金、香港远志、黄花远志、黄花倒水莲亲缘关系较近，其物种结构更加相似。相对保守的 IR 区域的收缩与扩张现象代表着植物的进化，不同植物的叶绿体基因组大小与其密切相关。分析 IR-LSC/SSC 区域的边界信息，对叶绿体基因组结构的差异、物种进化等有进一步了解[52]。通过对 8 个物种的 IR 边界研究发现，在密花远志叶绿体基因组的 IR 长度最大，其叶绿体基因组长度 8 个物种中也最大，远志属植物叶绿体基因组除了西南远志和密花远志外，基本边界变化呈现规律性。

基于远志科 9 个物种的叶绿体基因组数据构建最大似然树和邻接法两种构建系统发育树，支持率均为 100，其进化树的拓扑结构略有不同，与 Ma 等[31]构建的进化树存在一些差异，分析其系统发育关系可知，远志、卵叶远志、瓜子金和香港远志为姊妹类群，表明四者亲缘关系最近，另外远志科与外类群其他属能够很好的区分。原因首先基于叶绿体不同数据集构建的进化树，本研究相比其数据量更为丰富些。其次叶绿体基因片段可能会丢失某些重要的信息，难以解决物种多、分类较难的大科的系统进化问题。并且基于叶绿体的不同数据集构建的进化树相比于其他方法构建的进化树支持率更高，可靠性更强，为远志的系统分类地位和种间进化关系研究奠定一定的理论基础。

本研究对远志叶绿体基因组进行研究，并分析了其叶绿体基因组结构特征，挖掘其叶绿体基因组的重复序列位点、分析其密码子偏好性和远志属叶绿体基因组 IR 与 SC 边界分析并利用叶绿体基因组数据构建系统发育树，有助于后续分子标记、DNA 条形码技术等研究的深入进展，提供了远志叶绿体基因组信息支持。揭示远志属物种之间的系统关系，为药用远志的资源筛选、鉴定、保存及遗传多样性分析等后续研究提供了分子依据，为产业化应用奠定基础，进而为保护远志物种资源提供有力保障。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 张韵洁, 李德铢. 叶绿体系统发育基因组学的研究进展 [J]. 植物分类与资源学报, 2011, 33(4): 365-375.

[2] 乔永刚, 贺嘉欣, 王勇飞, 等. 药用植物苦参的叶绿体基因组及其特征分析 [J]. 药学学报, 2019, 54(11): 2106-2112.

[3] 杨琬卿, 刘嘉灏, 解秋风, 等. 乌头叶绿体基因组密码子偏好性分析 [J]. 分子植物育种, 2008, 12: 36.

[4] 毕毓芳, 温星, 潘雁红, 等. 叶绿体 DNA 条形码在林木中的应用及研究进展 [J]. 分子植物育种, 2020, 18(16): 5444-5452.

[5] Dyall S D, Brown M T, Johnson P J. Ancient invasions: From endosymbionts to organelles [J]. *Science*, 2004, 304(5668): 253-257.

[6] Chiba Y. Cytochemical studies on chloroplasts I [J]. *CYTOLOGIA*, 1951, 16(3): 259-264.

[7] 韩增杰. 西洋参叶绿体基因组的生物信息学研究 [D]. 昆明: 昆明理工大学, 2016.

[8] Cheng H, Li J F, Zhang H, *et al*. The complete chloroplast genome sequence of strawberry (*Fragaria* × *ananassa* Duch.) and comparison with related species of Rosaceae [J]. *PeerJ*, 2017, 5: e3919.

[9] 倪梁红, 赵志礼, 米玛. 药用植物叶绿体基因组研究进展 [J]. 中药材, 2015, 38(9): 1990-1994.

[10] Wicke S, Schneeweiss G M, DePamphilis C W, *et al*. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function [J]. *Plant Mol Biol*, 2011, 76(3/4/5): 273-297.

[11] 田星, 刘莹莹, 张颖敏, 等. 藜芦属药用植物的叶绿体基因组比较分析和系统发育研究 [J]. 中草药, 2022, 53(4): 1127-1137.

[12] Takano A, Okada H. Phylogenetic relationships among subgenera, species, and varieties of Japanese *Salvia* L. (Lamiaceae) [J]. *J Plant Res*, 2011, 124(2): 245-252.

[13] Kuang D Y, Wu H, Wang Y L, *et al*. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): Implication for DNA barcoding and population genetics [J]. *Genome*, 2011, 54(8): 663-673.

[14] 武立伟, 崔英贤, 聂丽萍, 等. 细茎石斛叶绿体全基因组序列特征及系统发育分析 [J]. 药学学报, 2020, 55(5): 1056-1066.

[15] 中国科学院中国植物志编辑委员会. 马其云等编著. 中国植物志-拉丁名索引: 1959～1992 [M]. 北京: 科学出版社, 1997: 263.

[16] 华佗. 华氏中藏经 [M]. 北京: 商务印书馆出版, 1960: 47.

[17] 桑旭星, 杨依, 方芳. 远志寡糖酯类化合物药理活性研究进展 [J]. 中国药学杂志, 2017, 52(18): 1576-1579.

[18] 张陶珍, 荣巍巍, 李清, 等. 远志的研究进展 [J]. 中草药, 2016, 47(13): 2381-2389.

[19] 刘婉婉, 许璐, 董宪喆, 等. 开心散类方对慢性应激大鼠行为学及中枢单胺类神经递质的影响 [J]. 中国中药杂志, 2015, 40(11): 2180-2185.

[20] Zhou H, Xue W, Chu S F, *et al*. Polygalasaponin XXXII,

a triterpenoid saponin from *Polygalae Radix*, attenuates scopolamine-induced cognitive impairments in mice [J]. *Acta Pharmacol Sin*, 2016, 37(8): 1045-1053.

[21] 柴智, 张娟娟, 孙胜杰, 等. 远志总皂苷抗衰老与免疫调节作用研究 [J]. 中华中医药杂志, 2018, 33(2): 704-707.

[22] Hua L, Yao X. A water-soluble polysaccharide from the roots of *Polygala tenuifolia* suppresses ovarian tumor growth and angiogenesis *in vivo* [J]. *Int J Biol Macromol*, 2018, 107: 713-718.

[23] 彭汶铎. 远志皂苷 H 对离体平滑肌与心脏的作用 [J]. 中国药学杂志, 1999, 34(4): 241-243.

[24] 彭汶铎, 许实波. 四种远志皂苷的镇咳和祛痰作用 [J]. 中国药学杂志, 1998, (8): 45.

[25] 王光志, 万德光. 远志种质资源遗传多样性随机扩增多态性 DNA 分析 [J]. 时珍国医国药, 2009, 20(7): 1834-1835.

[26] 刘超. 传统中药远志和西伯利亚远志的分类和遗传多样性研究 [D]. 昆明: 云南大学, 2015.

[27] 韩雪婷, 房敏峰, 李忠虎, 等. 基于叶绿体DNA trnL内含子序列变异的远志谱系地理学研究 [J]. 中草药, 2014, 45(22): 3311-3316.

[28] 韩雪婷. 药用植物远志的谱系地理学研究 [D]. 西安: 西北大学, 2014.

[29] 樊杰, 薛同同, 王佳丽, 等. 远志及其混淆品的 ITS2 序列鉴别 [J]. 时珍国医国药, 2020, 31(10): 2406-2408.

[30] 马孝熙, 任伟超, 孙伟, 等. 远志药材及其混伪品的 DNA 条形码鉴定 [J]. 世界科学技术—中医药现代化, 2014, 16(8): 1719-1724.

[31] Ma J Y, Wang J M, Li C Y, *et al*. The complete chloroplast genome characteristics of *Polygala crotalarioides* Buch. -Ham. ex DC. (Polygalaceae) from Yunnan, China [J]. *Mitochondrial DNA B*, 2021, 6(10): 2838-2840.

[32] Lee D H, Cho W B, Park B J, *et al*. The complete chloroplast genome of *Polygala tenuifolia*, a critically endangered species in Korea [J]. *Mitochondrial DNA B*, 2020, 5(2): 1919-1920.

[33] Tillich M, Lehwark P, Pellizzer T, *et al*. GeSeq - versatile and accurate annotation of organelle genomes [J]. *Nucleic Acids Res*, 2017, 45(W1): W6-W11.

[34] Qu X J, Moore M J, Li D Z, *et al*. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes [J]. *Plant Methods*, 2019, 15: 50.

[35] Zheng S Y, Poczai P, Hyvönen J, *et al*. Chloroplot: An online program for the versatile plotting of organelle genomes [J]. *Front Genet*, 2020, 11: 576124.

[36] Beier S, Thiel T, Münch T, *et al*. MISA-web: A web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33(16): 2583-2585.

[37] 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.

[38] Lohse M, Drechsel O, Kahlau S, *et al*. Organellar GenomeDRAW: A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucleic Acids Res*, 2013, 41: W575-W581.

[39] 热伊汉古丽·图尔迪, 慕丽红, 田新民. 扁果草叶绿体基因组特征分析 [J]. 生物工程学报, 2022, 38(8): 2999-3013.

[40] Benson G. Tandem repeats finder: A program to analyze DNA sequences [J]. *Nucleic Acids Res*, 1999, 27(2): 573-580.

[41] 李冉郡, 武立伟, 辛天怡, 等. 大黄药材基原物种叶绿体基因组分析与特异 DNA 条形码开发 [J]. 药学学报, 2022, 57(5): 1495-1505.

[42] 兰朝辉, 田徐芳, 师玉华, 等. 五月艾 *Artemisia indica* 叶绿体基因组结构及系统发育分析 [J]. 中国中药杂志, 2022, 47(22): 6058-6065.

[43] 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.

[44] Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.

[45] Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program [J]. *Bioinformatics*, 2010, 26(15): 1899-1900.

[46] 李述成, 郭生虎, 贝盏临. 小檗属植物叶绿体基因组序列结构及系统发育分析 [J]. 中草药, 2022, 53(3): 818-826.

[47] 中国药典 [S]. 一部. 2020: 163.

[48] Zhang Y J, Du L W, Liu A, *et al*. The complete chloroplast genome sequences of five *Epimedium* species: Lights into phylogenetic and taxonomic analyses [J]. *Front Plant Sci*, 2016, 7: 306.

[49] 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.

[50] Bellgard M, Schibeci D, Trifonov E, *et al*. Early detection of G + C differences in bacterial species inferred from the comparative analysis of the two completely sequenced *Helicobacter pylori* strains [J]. *J Mol Evol*, 2001, 53(4): 465-468.

[51] 冯瑞云, 梅超, 王慧杰, 等. 籽粒苋叶绿体基因组密码子偏好性分析 [J]. 中国草地学报, 2019, 41(4): 8-15.

[52] 赵容, 尹舒悦, 姜诚湜, 等. 马兜铃科药用植物叶绿体基因组比较分析 [J]. 中国中药杂志, 2022, 47(11): 2932-2937.

[责任编辑 时圣明]