

卵叶远志叶绿体基因组序列特征与系统发育分析

罗 瑶¹, 胡本祥^{1,2}, 张 晗¹, 史嘉周¹, 姬海月^{1,2}, 净易尧¹, 陈晓颖¹, 王帮庆³, 颜永刚¹, 赵 璠⁴, 李艳茸⁴, 彭 亮^{1*}

1. 陕西中医药大学 药学院/陕西省秦岭中草药应用开发工程技术研究中心/“秦药”研发重点实验室, 陕西 咸阳 712046
2. 陕西国际商贸学院, 陕西 咸阳 712046
3. 汉中市科技资源统筹中心, 陕西 汉中 723000
4. 榆林市第五医院, 陕西 榆林 719000

摘要: 目的 解析卵叶远志 *Polygala sibirica* 叶绿体因组的序列特征和系统发育关系。方法 利用 CTAB 法提取卵叶远志叶片基因组 DNA, 通过 Illumina HiSeq 平台测序, 再用 GetCellel 组装叶绿体基因组, 通过 MEGA11 构建邻接法 (neighbor-joining, NJ 法) 系统发育树。结果 卵叶远志的叶绿体基因组全长 165 192 bp, GC 含量为 36.70%, 共编码 124 个基因, 其中包含 8 个 rRNA 基因、42 个 tRNA 基因和 74 个蛋白编码基因; 相对同义密码子使用度显示, 67.29% 的密码子的使用度 > 1, 密码子偏好 A、T 结尾; 共检测到 287 个简单重复序列 (simple sequence repeats, SSR), 其中以单核苷酸重复次数最多, 占 80.61%, 单核苷酸主要由 A 和 T 组成, 这表明在碱基形成过程中 A 和 T 被频繁使用; 边界分析显示, 9 种远志属植物叶绿体基因组边界较为稳定, 卵叶远志、香港远志 *P. hongkongensis*、远志 *P. tenuifolia*、瓜子金 *P. japonica* 和合叶草 *P. subopposita* 基因结构存在相似性; 变异分析显示, 9 种远志属植物叶绿体基因组变异幅度较小; 系统发育分析结果显示, 卵叶远志、香港远志、瓜子金、远志、合叶草和西南远志 *P. crotalarioides* 因较高的支持率可聚成一支。结论 构建了高质量的卵叶远志全叶绿体基因组, 为卵叶远志的遗传结构和遗传多样性研究奠定了基础。

关键词: 远志; 卵叶远志; 叶绿体基因组; 序列特征; 系统发育

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)18-6065-09

DOI: 10.7501/j.issn.0253-2670.2023.18.024

Chloroplast genome sequence characteristics and phylogenetic analysis of *Polygala sibirica*

LUO Yao¹, HU Ben-xiang^{1,2}, ZHANG Han¹, SHI Jia-zhou¹, JI Hai-yue^{1,2}, JING Yi-yao¹, CHEN Xiao-ying¹, WANG Bang-qing³, YAN Yong-gang¹, ZHAO Fan⁴, LI Yan-rong⁴, PENG Liang¹

1. College of Pharmacy, Shaanxi University of Chinese Medicine/Shaanxi Engineering Research Center for Application and Development of Chinese Medicine in Qinling Mountains/Key Laboratory of “Qin Medicine” Research and Development, Xianyang 712046, China
2. Shaanxi Institute of International Trade & Commerce, Xianyang 712046, China
3. Hanzhong Science and Technology Coordination Center, Hanzhong 723000, China
4. Yulin No.5 Hospital, Yulin 719000, China

Abstract: Objective To identify the sequence characteristics and phylogenetic relationships of *Polygala sibirica* chloroplast genome.

Methods The genomic DNA of leaves was extracted by CTAB method and sequenced by the Illumina HiSeq platform, and the chloroplast genome was assembled by GetCellel, and the neighbor-joining (NJ) phylogenetic tree was constructed by MEGA11.

Results The total length of the chloroplast genome of *P. sibirica* was 165 192 bp, with a GC value of 36.70%. The chloroplast genome

收稿日期: 2023-03-03

基金项目: 国家自然科学基金资助项目 (82003899); 陕西省科技厅社发攻关一般项目 (2021SF-364); 公益性行业 (中医药) 科研专项经费项目 (201507002-1-08); 陕西中医药大学校级课题项目 (2020GP28); 陕西中医药大学“秦药”品质评价及资源开发学科创新团队项目 (2019-QN01)

作者简介: 罗 瑶 (1997—), 女, 在读硕士研究生, 从事中药资源质量评价与分子生药学研究。E-mail: 2207256840@qq.com

*通信作者: 彭 亮 (1985—), 男, 副教授, 从事中药资源评价与开发利用、分子生药学研究。E-mail: ppengliang@126.com

of *P. sibirica* has 124 coding genes, with the numbers of CDS, tRNA, and rRNA were 74, 42, and 8, respectively; relative synonymous codon usage showed that 67.29% of codons had a usage level large than 1, and codons preferred A and T endings. A total of 287 simple sequence repeats (SSRs) were detected, and the number of single nucleotide repetitions was the highest, accounting for 80.61%. Mononucleotides are mainly composed of A and T, indicating that A and T are frequently used during base formation. Boundary analysis showed that the chloroplast genome boundaries of nine species of *Polygala* were relatively stable, and the gene structures of *P. sibirica*, *P. hongkongensis*, *P. tenuifolia*, *P. japonica* and *P. subopposita* were similar. Variation analysis showed that the chloroplast genome variation of nine species of *Polygala* was small. The results of phylogenetic analysis showed that *P. sibirica*, *P. hongkongensis*, *P. japonica*, *P. tenuifolia*, *P. subopposita* and *P. crotalarioides* could be clustered into one branch due to their high support rates.

Conclusion The high-quality whole chloroplast genome of *P. sibirica* was constructed, which laid a foundation for the study of genetic structure and genetic diversity of *P. sibirica*.

Key words: *Polygala tenuifolia* Willd.; *Polygala sibirica* L.; chloroplast genome; sequence characterization; phylogenetic analysis

远志为远志科远志属多年生草本植物远志 *Polygala tenuifolia* Willd.或卵叶远志 *P. sibirica* L.的干燥根^[1]。远志是我国 42 种重点保护的三级野生品种和 85 种传统出口大宗药材之一,也是目前临床益智药处方中使用度名列前 3 位的单味中草药,被视为养命之要药^[2]。远志药用历史悠久,始载于《神农本草经》,目前已从远志药材中分离和鉴定出 140 多种化合物,包括皂苷类、吡啶类、寡糖酯类、生物碱类等^[3]。其中,远志皂苷类、吡啶类以及寡糖酯类作为远志药材中最主要的药效成分,具有广泛的药理活性,如抗衰老、神经保护、抗抑郁、催眠镇静、抗炎、抗病毒、抗肿瘤、抗氧化、抗心律失常等^[4]。

叶绿体 (chloroplast, cp) 是古代细菌内共生体的后代,系植物细胞中的重要细胞器,负责光合作用和新陈代谢等,为植物生存和生长提供必要的能量^[5]。高等植物的叶绿体基因组中存在典型的环状双链结构,其大小在 120~180 kb,由 4 部分构成,包含 1 个小单拷贝区 (small single copy, SSC)、1 个大单拷贝区 (large single copy, LSC) 和 2 个反向重复区 (inverted repeats, IRs)^[6-7]。叶绿体基因组 DNA (cpDNA) 独立于核基因组,表现出半自主的遗传特征,可作为物种鉴定、系统发育、起源进化等研究的依据^[8-9]。研究发现,基于叶绿体基因组的分子条形码具有极佳的物种鉴定潜力^[10-11],且完整的叶绿体基因组序列可以为植物物种和种群水平上的准确鉴定提供可靠的条形码^[12-13]。同时,通过比较叶绿体基因组序列为发现序列变异和识别突变热点区域提供了机会,还可以检测出基因缺失和重复事件。因此,从叶绿体基因组序列获得的突变热点区域和简单重复序列 (simple sequence repeats, SSR) 可以作为物种鉴定和群体遗传学的

有效分子标记^[14]。随着二代测序技术 (next generation sequencing, NGS) 的普及和发展,叶绿体基因组数据库日益丰富。在解析植物类群系统发育关系研究中逐渐得到广泛应用,迄今已有 1000 多种植物完成了叶绿体基因组的测序,在植物系统发育地位、物种鉴定、进化分析等解析方面发挥重要作用^[15]。

目前,远志研究多集中于远志及其药材,而有关卵叶远志 *P. sibirica* L.的研究,仅有少数是关于资源调查、化学成分和药理作用等方面^[16-17],尚缺乏其叶绿体基因组及遗传背景信息的系统研究与分析。据此,本研究采用高通量测序技术以获得卵叶远志叶绿体全基因组序列信息,并利用生物信息学相关软件,分析其序列特征、基因组组成及系统发育关系,以期为该药用植物的遗传结构和遗传多样性研究奠定研究和理论基础,也为将来远志属植物的亲缘关系解析及新药源寻找等提供支撑。

1 材料

卵叶远志采自陕西省太白县大贯子 (N 107.2883205°, E 33.99384660°, 海拔 1302 m),经陕西中医药大学胡本祥教授鉴定为远志科卵叶远志 *P. sibirica* L.,取卵叶远志的新鲜叶片,液氮速冻后存于-80 °C 冰箱,用于 DNA 提取。植物凭证样本保存于陕西中医药大学标本馆 (PS20211001)。

2 方法

2.1 基因组 DNA 提取和测序

取卵叶远志的新鲜幼嫩叶片,通过改良十六烷基三甲基溴化铵法 (cetyltrimethylammonium bromide, CTAB)^[18]提取基因组 DNA,经电泳检测和浓度测定后,置-20 °C 冰箱留存备用。超声处理基因组 DNA,通过末端修复、加 A 尾、加测序接头、纯化、PCR 扩增等步骤完成文库构建,再通过

Illumina 高通量测序平台 Illumina HiSeq 测序, 以此来获得序列原始数据。

2.2 叶绿体基因组组装与注释

Illumina 数据使用 Trimmomatic v0.36^[19]以去除接头和低质量数据。参数设置如下: 前导: 20; 尾随: 20; 滑动窗口: 4:15; 最小镜头: 36; 平均合格: 20。处理完后的高质量 clean data, 使用 GetCellel 对叶绿体基因组进行组装, 拼接软件对优化序列进行多个 Kmer 参数的拼接, 长度设置为 107、117 和 127 bp。使用在线网站 CHLOROBX (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) 对组装好的叶绿体基因组进行注释^[20]。使用 OGDRAW 绘制叶绿体全基因组图谱^[21]。最终注释的叶绿体基因组上传于 GenBank 中, 注册号为 OP870084。

2.3 叶绿体基因组序列比较分析

为消除氨基酸组成对密码子使用的影响, 采用 MEGA11^[22]分析同义密码子使用量、相对同义密码子使用值 (relative synonymous codon usage, RSCU)、碱基组成和密码子含量的变化特征。使用 SSRHunter 软件 (<http://www.biosoft.net>) 鉴定叶绿体基因组中的 SSR^[23-24]。参数设置为单核苷酸至六核苷酸 8、5、4、4、4、4。SC/IR 边界使用 IRSCOPE^[25]进行作图分析。同样, 通过 mVISTA^[26] (<https://genome.lbl.gov/vista/mvista/submit.shtml>) 做全基因组对比, 分析时勾选全局

对比 (Shuffle-LAGAN), 对远志属 9 种叶绿体基因组序列做同源性对比分析。

2.4 系统发育分析

从 NCBI (<https://www.ncbi.nlm.nih.gov/>) 上下载远志科远志属植物远志 (NC_050829)、香港远志 *P. hongkongensis* (NC_066219.1)、瓜子金 *P. japonica* (NC_052912.1)、密花远志 *P. karensium* (NC_056971.1)、西南远志 *P. crotalarioides* (NC_060367.1)、合叶草 *P. subopposita* (OP_612815.1)、黄花倒水莲 *P. fallax* (NC_052911.1)、荷包山桂花 *P. arrillata* (NC_056971.1), 选取楝科香椿属植物香椿 *Toona sinensis* (NC_059806.1) 和楝科非洲楝属植物非洲楝 *Khaya senegalensis* (NC_037362.1) 作为外类群, 共 11 个物种完整的叶绿体基因组序列通过在线软件 MAFFT version7 (<https://mafft.cbrc.jp/alignment/server/>) 进行多序列对比^[27], 利用 MEGA11 构建邻接法 (neighbor-joining, NJ) 系统发育树^[28], 自展次数为 1000^[28]。

3 结果与分析

3.1 叶绿体基因组的结构

卵叶远志叶绿体基因组具有典型的四分体结构 (图 1), 由 LSC 区域 (165、192~248、865), SSC 区域 (120、409~128、457), IRA (128、458~165、191) 和 IRB (83、674~120、408) 组成。叶绿体基因组全长 165 192 bp, GC 含量约为 36.7% (表 1)。

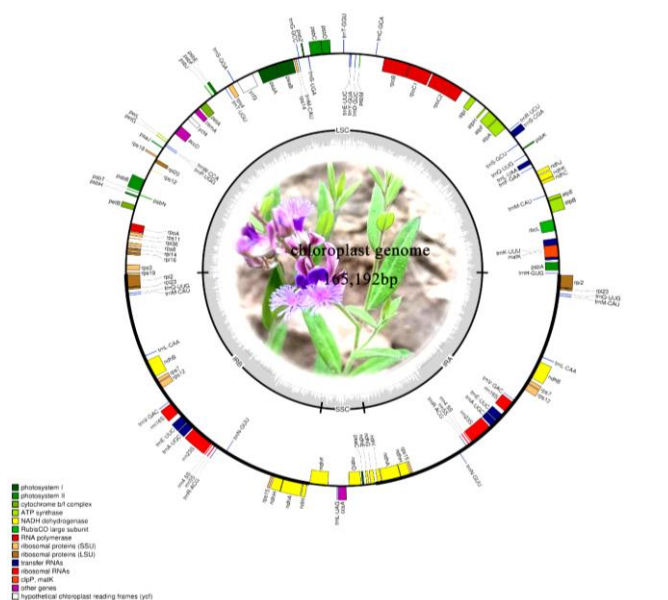


图 1 卵叶远志叶绿体基因组图谱

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

表 1 卵叶远志叶绿体基因组碱基组成

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

区域	A/%	T/%	C/%	G/%	GC/%	碱基长度/bp
IRA	29.2	31.1	20.9	18.8	39.8	36 734
IRB	31.1	29.2	18.8	20.9	39.8	36 735
SSC	38.4	32.1	15.3	14.3	29.6	8049
LSC	32.0	32.2	18.0	16.8	34.8	83 674
总量	31.5	31.8	18.7	18.0	36.7	165 192

3.2 基因组成

基因注释结果显示, 卵叶远志叶绿体基因组共注释得到 124 个基因, 包括 8 个核糖体 RNA (rRNA) 基因、42 个转运 RNA (tRNA) 基因和 74 个蛋白

编码基因 (coding sequence, CDS) (表 2)。注释的基因中存在 1 个内含子的为 *trnA-UGC*、*trnA-UGC*、*trnI*、*trnK-UUU*、*trnL-UAA*、*trnS-CGA* 等, 存在 2 个内含子的基因为 *ycf3* 和 *rps12*。

表 2 卵叶远志叶绿体基因组上的基因

Table 2 Genes located on chloroplast genome of *P. sibirica*

基因类别	基因名称	数量
tRNA	<i>trnA-UGC</i> *(2)、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnA-UGC</i> *(2)、 <i>trnC</i> 、 <i>GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnE-UUC</i> 、 <i>trnF-GAA</i> 、 <i>trnG-GCC</i> 、 <i>trnH-GUG</i> 、 <i>trnI</i> *(2)、 <i>trnK-UUU</i> *、 <i>trnL-CAA</i> (2)、 <i>trnL-UAA</i> *、 <i>trnL-UAG</i> 、 <i>trnM-CAU</i> (4)、 <i>trnN-GUU</i> (2)、 <i>trnP-UGG</i> 、 <i>trnQ-UUG</i> (3)、 <i>trnR-ACG</i> (2)、 <i>trnR-UCU</i> 、 <i>trnS-CGA</i> *、 <i>trnS-GCU</i> 、 <i>trnS-GGA</i> 、 <i>trnS-UGA</i> 、 <i>trnT-GGU</i> 、 <i>trnT-UGU</i> 、 <i>trnV-GAC</i> (2)、 <i>trnW-CCA</i> 、 <i>trnY-GUA</i>	42
rRNA	<i>rrn4.5</i> (2)、 <i>rrn5S</i> (2)、 <i>rrn16S</i> (2)、 <i>rrn23S</i> (2)	8
核糖体蛋白小亚基	<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> (2)、 <i>rps18</i> 、 <i>rps19</i>	13
核糖体蛋白大亚基	<i>rpl2</i> *(2)、 <i>rpl14</i> 、 <i>rpl16</i> *、 <i>rpl20</i> 、 <i>rpl23</i> (2)、 <i>rpl36</i>	8
RNA 聚合酶	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> *、 <i>rpoC2</i>	4
NADH 脱氢酶亚基	<i>ndhA</i> *(2)、 <i>ndhB</i> *(2)、 <i>ndhC</i> 、 <i>ndhD</i> 、 <i>ndhE</i> 、 <i>ndhF</i> 、 <i>ndhG</i> 、 <i>ndhH</i> (2)、 <i>ndhI</i> (2)、 <i>ndhJ</i> 、 <i>ndhK</i>	15
光系统 I 亚基	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaJ</i>	4
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>psbJ</i> 、 <i>psbK</i> 、 <i>psbM</i> 、 <i>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>	13
色素细胞 b/f 复合物亚基	<i>petA</i> 、 <i>petB</i> *、 <i>petG</i> 、 <i>petL</i>	4
ATP 合成酶亚基	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> *、 <i>atpH</i> 、 <i>atpI</i>	6
Rubisco 大亚基	<i>rbcL</i>	1
成熟酶	<i>matK</i>	1
被膜蛋白	<i>cemA</i>	1
乙酰 CoA 羧化酶	<i>accD</i> *	1
色素细胞 C 合成酶	<i>ccsA</i>	1
未知功能蛋白	<i>ycf3</i> **、 <i>ycf4</i>	2
合计		124

*-内含子为 1 **-内含子为 2

*-one intron **-two introns

3.3 密码子偏好分析

卵叶远志叶绿体密码子研究发现, 64 个蛋白编码基因共编码 35 541 个密码子。亮氨酸 (Leu) 使用频率最高, 异亮氨酸 (Ile) 次之, 半胱氨酸 (Cys) 使用度最低, 分别为 3775 (10.62%)、3075 (8.65%)、423 (1.19%)。密码子使用偏好显示, 卵叶远志叶绿体基因组中

RSCU > 1 的密码子占总量的 67.29%, 表现出以 A、T 结尾的特征 (表 3)。

3.4 SSR 分析

卵叶远志叶绿体基因组中鉴定出单、二、三、四、五、六核苷酸 SSRs 位点共 287 个, 其数量分别为 240、28、4、8、3、4。其中, 单核苷酸重复次数最多, 占比 80.61% (表 4)。

表3 卵叶远志密码子信息
Table 3 Codon usage in *P. sibirica*

密码子	氨基酸	数量	RSCU	密码子	氨基酸	数量	RSCU
GCA	Ala	593	1.12	CCC	PRO	324	0.88
GCC	Ala	306	0.58	CCG	PRO	241	0.66
GCG	Ala	249	0.47	CCU	PRO	509	1.39
GCU	Ala	961	1.82	CAA	Gln	911	1.48
UGC	Cys	122	0.58	CAG	Gln	322	0.52
UGU	Cys	301	1.42	AGA	Arg	588	1.65
GAC	Asp	235	0.38	AGG	Arg	264	0.74
GAU	Asp	987	1.62	CGA	Arg	479	1.34
GAA	Glu	1291	1.48	CGC	Arg	160	0.45
GAG	Glu	450	0.52	CGG	Arg	139	0.39
UUC	Phe	691	0.66	CGU	Arg	514	1.44
UUU	Phe	1405	1.34	AGC	Ser	208	0.49
GGA	Gly	995	1.51	AGU	Ser	503	1.19
GGC	Gly	292	0.44	UCA	Ser	458	1.08
GGG	Gly	472	0.72	UCC	Ser	403	0.95
GGU	Gly	871	1.32	UCG	Ser	239	0.56
CAC	His	238	0.56	UCU	Ser	730	1.72
CAU	His	613	1.44	ACA	Thr	502	1.13
AUA	Ile	953	0.93	ACC	Thr	341	0.77
AUC	Ile	578	0.56	ACG	Thr	209	0.47
AUU	Ile	1544	1.51	ACU	Thr	723	1.63
AAA	Lys	1291	1.50	GUA	Val	707	1.40
AAG	Lys	426	0.50	GUC	Val	245	0.49
CUA	Leu	429	0.68	GUG	Val	281	0.56
CUC	Leu	268	0.43	GUU	Val	782	1.55
CUG	Leu	210	0.33	UGG	Trp	655	1.00
CUU	Leu	772	1.23	UAC	Tyr	283	0.42
UUA	Leu	1304	2.07	UAU	Tyr	1064	1.58
UUG	Leu	792	1.26	UAA	Stop	169	1.19
AUG	Met	801	1.00	UAG	Stop	124	0.87
AAC	Asn	335	0.45	UGA	Stop	134	0.94
AAU	Asn	1161	1.55	CCA	Pro	394	1.07

表4 卵叶远志叶绿体基因组的SSRs

Table 4 Distribution of SSRs within *P. sibirica* chloroplast genome

核苷酸类型	重复序列	数量	占比/%
单核苷酸	A/T	230	80.13
	C/G	10	3.48
二核苷酸	AG/CT	5	1.74
	AT/AT	23	8.01
三核苷酸	AAG/CTT	1	0.34
	AAT/ATT	3	1.04
四核苷酸	AAAG/CTTT	1	0.34
	AAAT/ATTT	5	1.74
	ATCC/ATGG	2	0.69
五核苷酸	AAAAT/ATTTT	2	0.69
	AATCT/AGATT	1	0.34
六核苷酸	AAAGAT/ATCTTT	1	0.34
	AAGATC/ATCTTG	1	0.34
	AAGATG/ATCTTC	2	0.69

3.5 边界分析

选取NCBI上已发布的8种远志属植物叶绿体基因组SC/IR进行边界分析,显示远志属植物叶绿体基因组存在4个边界。如图2所示,LSC/IRb区

域中,除密花远志 *P. karensium* L.及西南远志 *P. crotalarioides* L.外,其余7种均位于rps19基因编码区。其中,向LSC区域扩张1bp的香港远志 *P. hongkongensis* L.、远志、瓜子金和合叶草;扩张2bp的为西伯利亚远志;扩张7bp的为黄花倒水莲和荷包山桂花。IRb/SSC区有4种在ndhF基因编码区,5种在ndhF基因非编码区,其中香港远志和荷包山桂花分别向IRb扩张10bp和3bp。SSC/IRa区,所有物种都在ndhG基因和ndhI基因之间的非编码区,且均向SSC区扩张,该区域无剧烈变化。在IRa/LSC,除西南远志处于ycf2基因编码区外,其他8种均在rpl2基因和trnH基因之间的非编码区,均向IRa区有不同程度的扩张。边界分析表明,远志属植物叶绿体基因组在其IR边界具有一定差异,总体而言基因边界变异幅度较小,叶绿体基因组较为保守。显示出西伯利亚远志、香港远志、远志、瓜子金和合叶草在SC/IR边界的结构上存在相似性。

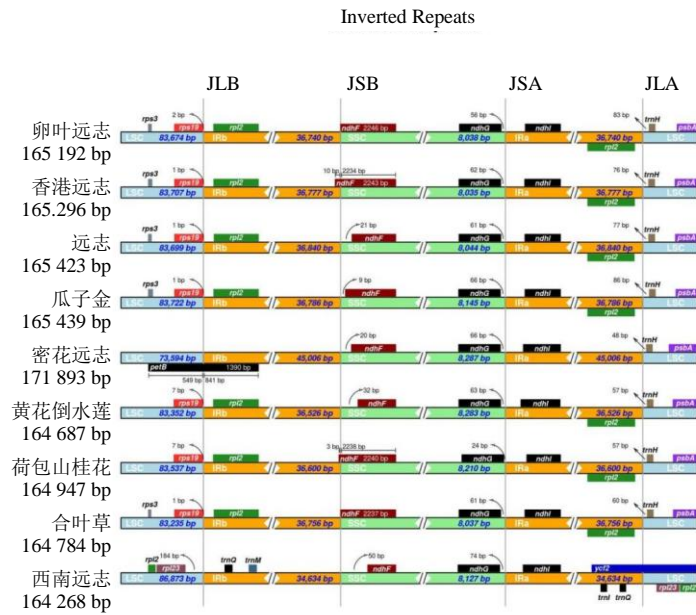


图2 9种远志属植物叶绿体基因组的IR/SC边界变化情况

Fig. 2 Changes of IR/SC boundary of chloroplast genomes of nine *Polygala* species

3.6 远志属植物叶绿体基因组变异分析

以卵叶远志 (GenBank 注册号为 NC056970) 为注释, 对9种远志属植物进行叶绿体基因组全序列对比分析 (图3)。结果表明, 9种远志属植物叶绿体基因组的4个基因区间组成较为一致, 差异性较小。从

4大区段来看, LSC区差异性最大, 变异程度最高, 而IRA区则差异性最小, 最为保守。从非基因编码区和基因编码区来看, 非基因编码区变异程度较高, 基因编码区较为保守, 但在 *ycf1*、*ycf2*、*ycf3*、*ndhF* 和 *ndhD* 等基因编码区变异程度较大, 存在显著差异。

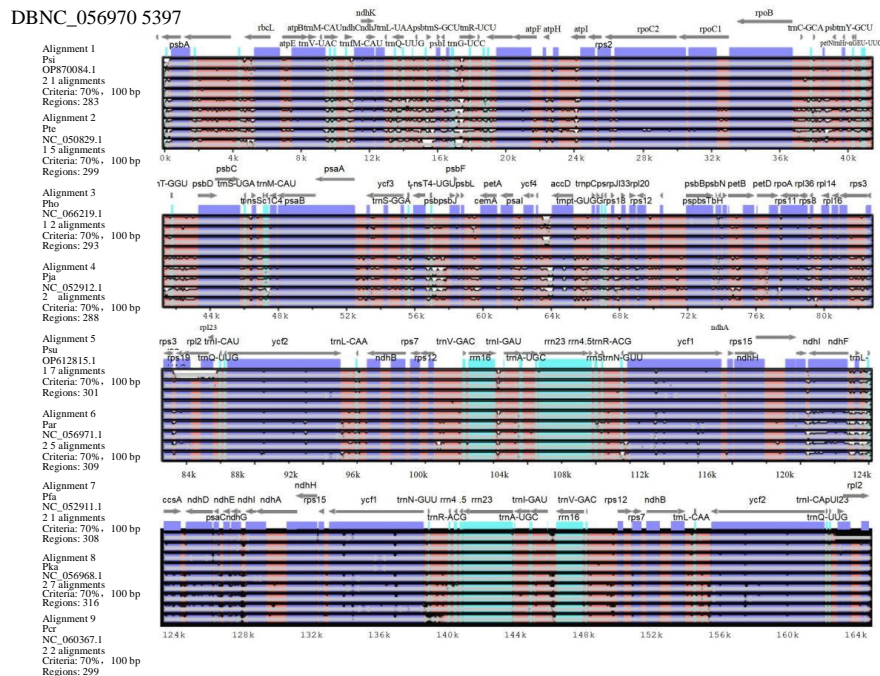


图3 远志属植物叶绿体基因组全局比对分析

Fig. 3 Global alignment analysis of *Polygala* chloroplast genomes

3.7 系统发育分析

利用邻接法构建系统发育树(图4),结果显示11个物种可分为远志科和外类群的楝科。远志科的9种远志属植物以100%支持率构成一个分

支,与楝科区分开。同时,远志属中香港远志、卵叶远志、瓜子金、远志、合叶草和西南远志可聚成一支,而密花远志、荷包山桂花和黄花倒水莲则聚为姐妹支。

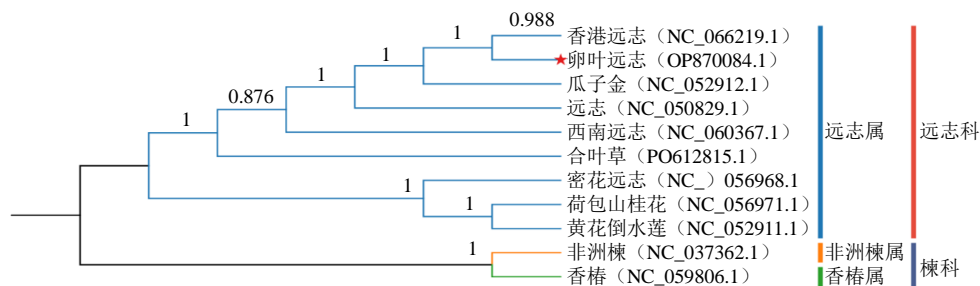


图4 基于叶绿体全基因组序列构建NJ系统进化树

Fig. 4 Phylogenetic analysis based on chloroplast genome sequences by neighbor-joining (NJ) tree

4 讨论

卵叶远志作为远志中药材的2种基原植物之一,具有重要的药用价值和经济价值。卵叶远志和远志分布和产地也基本相同,远志药材的质量标准中并未区分卵叶远志与远志的种间区别,也未客观确定各自的质量指标,有关药用远志的研究也主要集中在远志方面。近年来,对卵叶远志的研究也在逐渐加深^[29],但目前尚缺乏关于其遗传信息的研究。基于此,本研究完成了卵叶远志叶绿体基因组的测序、组装和注释工作,并分析了其结构、GC含量、基因组成、密码子使用度、SSR等,着重对9种远志属植物进行了叶绿体基因组边界分析、变异分析及系统发育分析等。研究发现,卵叶远志叶绿体基因组为经典的环状四分结构,全长为165 192 bp,其GC含量为36.75%,共编码124个基因,其中含有一个内含子的基因有14个,含有2个内含子的基因有2个。同时,和大多数被子植物叶绿体基因组一样^[30],卵叶远志中也存在未知功能基因,如ycf3和ycf4基因,仍有待进行突破性研究。

密码子是生物体中连接氨基酸、蛋白质和遗传物质的重要核心元素,在生物个体遗传信息的传递中起重要作用^[31]。大多数氨基酸不仅由一种密码子编码,也可以同时由2~3种密码子编码,生物体对于不同的密码子的使用具有一定偏好性^[32]。密码子偏好的研究使用为蛋白质表达及其相应功能的研究提供了可靠的信息。密码子偏好也可作为微调基因表达的手段^[33],双子叶植物密码子偏好以A/T结尾,单子叶植物偏好则与双子叶植物不同,偏向G/C^[34]。对卵叶远志密码子分析可知,卵叶远志密

码子RSCU>1的有67.29%,偏好为以A、T结尾。

目前,SSR是主要存在于基因外部和基因非编码区的共显性标记,也称微卫星,可用于群体遗传学研究的分子标记^[35]。在卵叶远志叶绿体基因组中,检测出的SSRs位点共有287个,其中单核苷酸重复次数最多,占80.61%。单核苷酸主要由A和T组成,这也表明卵叶远志叶绿体基因组在碱基形成过程中A和T被频繁使用。

远志属植物叶绿体基因组的SC/IR边界,虽存在部分扩张或收缩情况,但大体来说,整个基因组仍然较为保守,卵叶远志、远志、瓜子金、香港远志和合叶草基因结构存在相似性。在对其进行全序列比对时,LSC区和非基因编码区变异程度较高,IRa区和大部分基因组编码区保守程度较高。本研究通过全序列对比,结合边界分析与系统发育结果,得出卵叶远志与香港远志聚为一支,且与远志、瓜子金具有较高的亲缘关系。与王星璐等^[36]构建的进化树存在一些差异,分析其系统发育关系可知,瓜子金和卵叶远志聚为一支,二者又与香港远志聚类到一起,瓜子金、卵叶远志、香港远志和远志为姊妹类群,表明四者亲缘关系最近关系密切,推测差异原因可能与样本的采集环境及变异有关。

本研究通过文库构建、高通量测序、组装和序列分析,系统解析了卵叶远志叶绿体基因组,分析了其叶绿体基因组序列特征和系统发育关系,有助于深入开展后续分子标记、DNA条形码技术等研究,为进一步探究该药用植物卵叶远志的遗传结构和遗传多样性奠定了基础。同时,也增添了卵叶远志在分子生药学方面的部分研究,为加快对卵叶远志的

研究以及在基因层面鉴别该种属提供了相关依据。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国药典 [S]. 一部. 2020:163.
- [2] 薛英, 李晓伟, 李震宇, 等. 采用 UPLC/Q-TOF MS 与 NMR 代谢组学技术研究生长年限对远志药材质量的影响 [J]. 药学学报, 2015, 50(3): 340-347.
- [3] 彭亮, 颜永刚, 陈莹, 等. 茉莉酸甲酯诱导下远志幼苗转录组分析及三萜类生物合成途径关键酶基因挖掘 [J]. 中草药, 2020, 51(9): 2517-2529.
- [4] 姚辛敏, 周晓洁, 周妍妍, 等. 远志化学成分及药理作用研究进展 [J]. 中医药学报, 2022, 50(2): 103-107.
- [5] Brunkard J O, Runkel A M, Zambryski P C. Chloroplasts extend stromules independently and in response to internal redox signals [J]. *Proc Natl Acad Sci U S A*, 2015, 112(32): 10044-10049.
- [6] 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.
- [7] Zhang X, Gu C R, Zhang T X, et al. Chloroplast (Cp) Transcriptome of *P. davidiana* Dode × *P. bolleana* Lauch provides insight into the Cp drought response and *Populus* Cp phylogeny [J]. *BMC Evol Biol*, 2020, 20(1): 51.
- [8] Lilly J W, Havey M J, Jackson S A, et al. Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants [J]. *Plant Cell*, 2001, 13(2): 245-254.
- [9] 刘畅, 曾宪法, 杨小英, 等. 多叶勾儿茶叶叶绿体基因组比较分析及系统发育研究 [J]. 中草药, 2023, 54(8): 2558-2565.
- [10] Amirouf A, Hyvönen J, Poczar P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae [J]. *PLoS One*, 2018, 13(4): e0196069.
- [11] Dong W P, Xu C, Li C H, et al. ycf1, the most promising plastid DNA barcode of land plants [J]. *Sci Rep*, 2015, 5: 8348.
- [12] Neubig K M, Whitten W M, Carlswald B S, et al. Phylogenetic utility of ycf1 in orchids: A plastid gene more variable than matK [J]. *Plant Syst Evol*, 2009, 277(1): 75-84.
- [13] Hernández-León S, Gernandt D S, Pérez de la Rosa J A, et al. Phylogenetic relationships and species delimitation in pinus section trifoliae inferred from plastid DNA [J]. *PLoS One*, 2013, 8(7): e70501.
- [14] Dong W P, Liu J, Yu J, et al. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding [J]. *PLoS One*, 2012, 7(4): e35071.
- [15] 韦海忠, 潘丽芹, 田盛野, 等. 单叶铁线莲叶绿体基因组的序列特征与系统发育分析 [J]. 中草药, 2022, 53(12): 3766-3773.
- [16] 罗露, 胡本祥, 曹福麟, 等. 远志两种不同基原植物远志与卵叶远志的红外光谱学研究 [J]. 时珍国医国药, 2019, 30(8): 1895-1899.
- [17] 王永艳, 蔡小艳, 刘彦刚, 等. 卵叶远志的研究进展 [J]. 中国医药导报, 2016, 13(8): 33-35.
- [18] McPherson H, van der Merwe M, Delaney S K, et al. Capturing chloroplast variation for molecular ecology studies: A simple next generation sequencing approach applied to a rainforest tree [J]. *BMC Ecol*, 2013, 13: 8.
- [19] Bolger A M, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [20] 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.
- [21] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47(W1): W59-W64.
- [22] Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11 [J]. *Mol Biol Evol*, 2021, 38(7): 3022-3027.
- [23] Chen P, Gao G, Yu C M, et al. Data set for transcriptome analysis of *Apocynum venetum* L [J]. *Data Brief*, 2018, 20: 1739-1744.
- [24] 李强, 万建民. SSRHunter, 一个本地化的 SSR 位点搜索软件的开发 [J]. 遗传, 2005, 27(5): 808-810.
- [25] Amirouf A, Hyvönen J, Poczar P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [26] Frazer K A, Pachter L, Poliakov A, et al. VISTA: Computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, 32: W273-W279.
- [27] Katoh K, Misawa K, Kuma K, et al. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30(14): 3059-3066.
- [28] Minh B Q, Nguyen M A T, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap [J]. *Mol Biol Evol*, 2013, 30(5): 1188-1195.

- [29] 吕鼎豪, 李香串, 牛俊峰. 卵叶远志根茎结构特征及内含物鉴别研究 [J]. 中国野生植物资源, 2020, 39(5): 23-28.
- [30] 林新茂, 黄家权. 仙人掌科叶绿体与核基因组密码子偏好性分析 [J]. 分子植物育种, 2023, 26(5): 1-25.
- [31] 冯展, 江媛, 郑燕, 等. 肉苁蓉属植物叶绿体基因组密码子偏好性分析 [J]. 中草药, 2023, 54(5): 1540-1550.
- [32] 高梦琦, 邹建珍, 霍小位, 等. 基于转录组数据分析药用大黄的密码子使用偏好性 [J]. 中草药, 2021, 52(20): 6344-6349.
- [33] Quax T E, Claassens N J, Söll D, *et al.* Codon bias as a means to fine-tune gene expression [J]. *Mol Cell*, 2015, 59(2): 149-161.
- [34] Wang L J, Roossinck M J. Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants [J]. *Plant Mol Biol*, 2006, 61(4/5): 699-710.
- [35] Fernandez-Pozo N, Menda N, Edwards J D, *et al.* The Sol genomics network (SGN): From genotype to phenotype to breeding [J]. *Nucleic Acids Res*, 2015, 43: D1036-D1041.
- [36] 王星璐, 李慧娟, 赵伟, 等. 远志叶绿体基因组序列特征与系统发育分析 [J]. 中草药, 2023, 54(11): 3655-3665.

[责任编辑 时圣明]