

• 药材与资源 •

蒙古黄芪叶绿体全基因组特征解析及亲缘性分析

李俊霖¹, 郭淑红², 张强², 张丽君³, 田洪岭^{2*}, 张琼^{1*}

1. 山西医科大学药学院, 山西 太原 030001

2. 山西农业大学经济作物研究所, 山西 太原 030031

3. 山西农业大学 农业基因资源研究中心, 山西 太原 030031

摘要: 目的 以蒙古黄芪 *Astragalus membranaceus* var. *mongholicus* 为材料, 解析其叶绿体基因组的序列特征, 并进行系统发育分析。方法 利用 Illumina NovaSeq 测序平台对蒙古黄芪叶绿体全基因组进行测序, 完成其组装、注释、特征分析, 构建系统发育树。结果 蒙古黄芪叶绿体基因组全长为 123 349 bp, GC 含量为 34.09%, 由于缺失反向重复 IR 区, Circos 图呈现出非典型四分体结构, 属于 IRLC 群体。获得注释基因 109 个, 其中蛋白编码基因 76 个, rRNA 基因 4 个, tRNA 基因 29 个。蒙古黄芪叶绿体基因组的密码子偏好性较弱, 密码子偏向使用 A 碱基和 U 碱基。MISA 检测出 263 个 SSR 位点, 以 A/T 重复为主; 叶绿体全基因组比较分析可知 17 种豆科植物叶绿体基因组的基因区间组成差异较小; 但 *trnF-GAA~trnT-UGU*、*trnM-CUA~psbC*、*trnE~trnD*、*psbM~rpoB*、*ycf1* 和 *ycf2* 等编码区存在差异性; 系统发育树显示, 蒙古黄芪与胶黄芪 *A. gummifer* 聚为一类, 与阿纳卡黄芪 *A. nakaianus*、膜荚黄芪 *A. membranaceus* 具有一定亲缘性。结论 可为开展黄芪属遗传多样性研究、DNA 条形码构建和分子鉴定标记开发提供参考。

关键词: 蒙古黄芪; 叶绿体基因组; 序列比对; IRLC; 系统发育

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)07-2366-09

DOI: 10.7501/j.issn.0253-2670.2024.07.022

Characteristics and phylogenetic analysis of *Astragalus membranaceus* var. *mongholicus* chloroplast whole genome

LI Junlin¹, GUO Shuhong², ZHANG Qiang², ZHANG Lijun³, TIAN Hongling², ZHANG Qiong¹

1. School of Pharmacy, Shanxi Medical University, Taiyuan 030001, China

2. The Industrial Crop Institute, Shanxi Agricultural University, Taiyuan 030031, China

3. Center for Agricultural Genetic Resources Research, Shanxi Agricultural University, Taiyuan 030031, China

Abstract: Objective Using *Astragalus membranaceus* var. *mongholicus* as material to analyze sequence characteristics of chloroplast genome and performe phylogenetic analysis. **Methods** Illumina NovaSeq sequencing platform was used to sequence the whole chloroplast genome of *A. membranaceus* var. *mongholicus*, and completing its assembly, annotation and characterization, and constructing a phylogenetic tree. **Results** The full chloroplast genome of *A. membranaceus* var. *mongholicu* was 123 349 bp in length, and its GC content was 34.09%. Due to the deletion of the inverted repeat IR region, the Circos map showed an atypical tetrad structure, belonging to the IRLC population. A total of 109 genes were annotated, including 76 protein-coding genes, four rRNA genes, and 29 tRNA genes. The codon preference of the chloroplast genome of *A. membranaceus* var. *mongholicu* was weak, and it preferred to use A and U bases. A total of 263 SSRs were detected by MISA, most of which consisted of A and T. Comparative analysis of the whole chloroplast genomes showed that the gene region composition of the chloroplast genomes of 17 legume species was relatively little different. However, variations were observed in coding regions such as *trnF-GAA~trnT-UGU*, *trnM-CUA~psbC*, *trnE~trnD*, *psbM~rpoB*, *ycf1* and *ycf2*. The phylogenetic tree showed that the *A. membranaceus* var. *mongholicu* was grouped with *A. gummifer* and had certain genetic relationships with *A.*

收稿日期: 2023-11-06

基金项目: 国家重点研发专项 (2019YFC1710800); 国家中药材产业技术体系浑源综合试验站 (CARS-21-03); 山西农业大学校育种工程 (YZGC056); 山西省现代农业产业技术体系建设专项资金资助 (2021-11)

作者简介: 李俊霖 (1997—), 女, 硕士研究生, 研究方向为天然药物化学。E-mail: 15123581751@126.com

通信作者: 田洪岭, 男, 研究员, 硕士生导师, 从事道地药材品种选育、栽培技术及资源综合利用研究。E-mail: thl2003@163.com

张琼, 女, 教授, 硕士生导师, 从事天然药物化学研究。E-mail: zhangqiong_83@hotmail.com

nakaianus and *A. membranaceus*. **Conclusion** The findings of the research provided a reference for the subsequent study on genetic diversity of *Astragalus* L., the construction of DNA barcodes, and the development of molecular identification markers.

Key words: *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao; chloroplast genome; sequence alignment; IRLC; phylogeny

黄芪为豆科植物蒙古黄芪 *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao 或膜荚黄芪 *A. membranaceus* (Fisch.) Bge. 干燥的根, 是人们熟知的一种药食同源型植物, 在中国的山西、内蒙古、东北等地广泛栽培^[1]。黄芪具有补气升阳、固表止汗、利水消肿的功效, 常用于气虚乏力、食少便溏、中气下陷等疾病^[2]。目前, 黄芪的化学成分主要包括多糖、黄酮类和三萜类等, 其药理作用广泛, 包括增强免疫功能、抗肿瘤、保护心脑血管系统、保护内脏、调节机体代谢、保护神经系统等^[3]。

在植物进化中, 叶绿体内的遗传信息具有母系遗传的特性^[4]。叶绿体基因组的遗传信息相对于核基因组更加独立和保守, 表现出半自主的遗传特征, 其编码基因的数量、结构、组成和排列顺序在大部分情况下保持一致, 极少发生重组变异^[5]。因此, 叶绿体基因组在植物的物种鉴定、系统发育和物种起源的研究中得到了广泛应用^[6-8]。高等植物叶绿体基因组通常由 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 repeats, IRs) 组成, 形成一种典型环式双链 DNA 结构, 大小通常为 120~180 kb^[9-10]。高通量测序技术的发展为研究药

用植物叶绿体基因组提供了更加快捷、准确的方法。通过对蒙古黄芪近缘物种进行叶绿体基因组的比较分析, 可以有效地确定不同种群之间的亲缘关系, 研究物种起源、分化和进化过程^[11-12]。

目前, 关于黄芪叶绿体基因组的研究集中在叶绿体基因组密码子偏好性分析^[13], 叶绿体基因组的基因丢失和倒位^[14], 鲜有黄芪叶绿体全基因组结构特征及系统发育分析的报道。本研究完成了蒙古黄芪叶绿体基因组测序、组装和注释的工作, 分析其结构、GC 含量、基因组成、密码子使用度、简单重复序列、叶绿体基因组序列变异, 并构建系统发育树, 为进一步研究黄芪的遗传结构、遗传多样性及亲缘性奠定理论基础。

1 材料

测序对象采自山西农业大学经济作物研究所 (37°24'05"E, 111°78'65"N) 选择生长状况良好、干净、无病虫害的蒙古黄芪嫩叶, 放置-80 °C 冰箱备用, 凭证标本存放于山西农业大学经济作物研究所。蒙古黄芪及其近缘物种的叶绿体全基因组序列来源于 NCBI 数据库, 测序获得蒙古黄芪叶绿体基因组 (*Astragalus. mongholicus* chloroplast genome, AMCP) (表 1)。

表 1 植物样品信息

Table 1 Information of plant samples

物种	来源	GenBank 登录号	碱基长度/bp
草木樨状黄芪 <i>A. melilotoides</i> Pall.	NCBI	NC_072247	123 729
蒙古黄芪 <i>A. membranaceus</i> (Fisch.) Bge. var. <i>mongholicus</i> (Bge.) Hsiao	NCBI	NC_029828	123 582
细叶黄芪 <i>A. melilotoides</i> Pall. var. <i>tenuis</i> Ledeb.	NCBI	OP723862	123 012
乳白黄芪 <i>A. galactites</i> Pall.	NCBI	NC_058825	126 117
斜茎黄芪 <i>A. laxmannii</i> Pall.	NCBI	NC_052923	122 844
阿纳卡黄芪 <i>A. nakaianus</i> L.	NCBI	NC_028171	123 633
膜荚黄芪 <i>A. membranaceus</i> (Fisch.) Bge.	NCBI	KX255662	123 623
加拿大黄芪 <i>A. canadensis</i> L. var. <i>brevidens</i> (Gandog.) Barneby	NCBI	NC_060799	123 013
糙叶黄芪 <i>A. scaberrimus</i> Wang et Tang ex S. B. Ho	NCBI	NC_058245	123 492
胶黄芪 <i>A. gummifer</i> L.	NCBI	NC_047251	125 425
背扁黄芪 <i>A. complanatus</i> R. Br I ex Bunge	NCBI	NC_065023	124 131
弯花黄芪 <i>A. flexus</i> Fisch.	NCBI	ON550404	125 082
酒泉黄芪 <i>A. jiuquanensis</i> S. B. Ho	NCBI	ON550402	118 399
岩生黄芪 <i>A. lithophilus</i> Kar. et Kir.	NCBI	ON550399	123 196
舌尖黄芪 <i>A. oxyglottis</i> Steven	NCBI	ON550396	123 067
苜蓿 <i>Medicago sativa</i> L.	NCBI	KU321683	128 574
AMCP	测序	OR712437	123 349

2 方法

2.1 DNA 提取和测序

采用植物 DNA 提取试剂盒 (Tiangen Biotech 有限公司, 中国) 提取蒙古黄芪嫩叶片 DNA, 用 1% 琼脂糖凝胶电泳和微量分光光度计 (Nanodrop 2000, 美国) 检测提取的 DNA 质量和浓度。将符合要求的样品进行测序, 测序工作由上海派森诺生物科技股份有限公司完成。

2.2 高质量序列获取

研究采用全基因组鸟枪法 (whole genome shotgun, WGS) 策略, 借助二代测序技术 (next-generation sequencing, NGS), 使用 Illumina NovaSeq 测序平台, 构建名为 Lib_Name 的文库。本研究使用 fastp 进行数据质量控制, 滤过生成高质量序列。采用 AdapterRemoval (version 2)^[15] 去除 3' 端的接头污染, 采用滑动窗口法进行质量过滤, 计算窗口内碱基的平均 Q 值, 若 Q 值 < 20 , 删除窗口内的碱基; 若 Q 值 ≥ 20 , 则停止滑动。若双末端中任意 1 条 reads 的长度 ≤ 50 bp 和双末端中 N 碱基的个数 ≥ 5 , 则去除该双末端序列, 以确保数据集中包含的序列都具有足够的长度和足够的质量。

2.3 叶绿体基因组组装与注释

采用 GetOrganelle v1.7.7.0 软件, 进行叶绿体 DNA 序列的拼接。将拼接得到的完整的叶绿体基因组序列上传至 Geseq 网站 (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) 进行功能注释。在 Organellar Genome DRAW (OGDRAW) 绘制叶绿体全基因组图谱, 把最终注释的叶绿体基因组上传 GenBank 中获得登录号 (OR712437)。

2.4 密码子和简单重复序列分析

运用 CodonW 分析叶绿体基因组的密码子, 获得有效密码子 (effective number of codon, N_c)、GC 含量、相对同义密码子使用值 (relative synonymous codon usage, RSCU) 和第 3 位上的 GC 碱基的含量 (GC3s)。使用 MISA 检测叶绿体基因组的简单重复序列 (simple sequence repeats, SSRs)。参数设置为单核苷酸序列, 重复 ≥ 8 个单位; 二、三核苷酸重复序列, 重复 ≥ 4 个单位; 四、五及以上核苷酸重复序列, 重复 ≥ 3 个单位, 且 2 个 SSRs 之间的距离 ≥ 100 bp^[16]。

2.5 叶绿体全基因组比较分析

通过 mVISTA (<https://genome.lbl.gov/vista/>

index.shtml) 做全基因组对比, 用 shuffle-LAGAN 模式检测变异情况, 以蒙古黄芪 (OR712437) 为参照, 与已公布的 16 种豆科植物的叶绿体全基因组序列进行全基因组对比及差异性分析。

2.6 系统发育分析

为了解蒙古黄芪在黄芪属的系统发育位置, 从 NCBI 数据库中下载叶绿体全基因组, 共计 17 个物种, 分别为草木樨状黄芪 (NC_072247)、蒙古黄芪 (NC_029828)、细叶黄芪 (OP723862)、乳白黄芪 (NC_058825)、斜茎黄芪 (NC_052923)、阿纳卡黄芪 (NC_028171)、膜荚黄芪 (KX255662)、加拿大黄芪 s (NC_060799)、糙叶黄芪 (NC_058245)、胶黄芪 (NC_047251)、背扁黄芪 (NC_065023)、弯花黄芪 (ON550404)、酒泉黄芪 (ON550402)、岩生黄芪 (ON550399)、尖舌黄芪 (ON550396)、此外苜蓿 (KU321683) 被用作外群 (表 1)。利用 MEGA 11 软件运用邻接法 (neighbor-joining method, NJ) 构建系统发育树, 设置 Bootstrap 为 1 000^[17]。

3 结果与分析

3.1 叶绿体基因组的结构

对蒙古黄芪的叶绿体基因组分析发现该结构呈现非典型四分体结构, 由 3 个区域构成: 1 个 LSC、SSC 和 1 个 IRs, 故蒙古黄芪属于反转区缺失植物类群 (inverse region losing cloud, IRLC) (图 1)。测序获得叶绿体基因组全长为 123 349 bp, GC 含量为 34.09%, 获得注释基因 109 个, 包括 4 个核糖体 RNA (rRNA) 基因、29 个转运 RNA (tRNA) 基因和 76 个蛋白编码基因 (coding sequence, CDS) (表 2)。蒙古黄芪叶绿体基因组上共编码 29 个 tRNA, 其中含 1 个内含子的有 4 个 (*trnA-UGC*、*trnE-UUC*、*trnK-UUU*、*trnL-UAA*)。

3.2 密码子偏好性分析

从注释的叶绿体密码子分析发现, 64 个蛋白共编码了 41 116 个密码子。由于遗传密码子具有简并性, 色氨酸 (Trp) 和甲硫氨酸 (Met) 只有 1 个密码子, 其余均有 2 个及 2 个以上的密码子, RSCU > 1 的密码子共 29 个, 占总量的 62.42%, 其中除了精氨酸 (Arg) 密码子 AGA 的 RSCU 值 > 2 , 其余的密码子的 RSCU 值均在 1~2, 表明蒙古黄芪叶绿体基因组中不存在偏好性极强的密码子 (表 3)。其中使用频率最高的是亮氨酸 (Leu), 使用频率为 9.71%, 其次是异亮氨酸 (Ile), 使用频率为 9.45%, 最低的是色氨酸 (Trp), 使用频率为 1.48%。

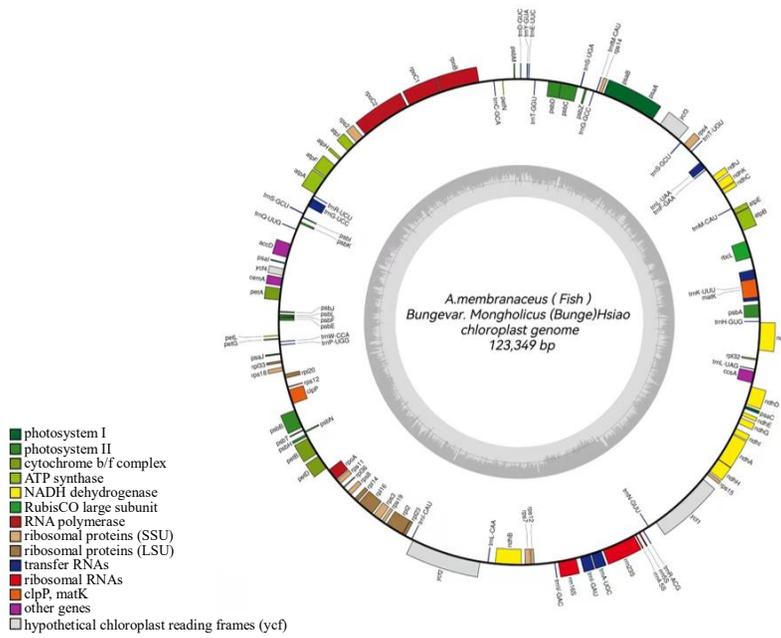


图 1 蒙古黄芪 AMCP 叶绿体基因组的基因圈图

Fig. 1 Gene map of chloroplast genome in AMCP

表 2 蒙古黄芪 AMCP 叶绿体基因组上的基因

Table 2 Genes located on chloroplast genome in AMCP

基因分类	基因功能	基因名称
自我复制基因	核糖体 RNA 转运 RNA	<i>rrn4.5S</i> , <i>rrn5S</i> , <i>rrn16S</i> , <i>rrn23S</i> <i>trnA-UGC*</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC*</i> , <i>trnF-GAA</i> , <i>trnM-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU</i> , <i>trnI-GAU*</i> , <i>trnK-UUU*</i> , <i>trnL-CAA</i> , <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GCU</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
光合作用基因	核糖体蛋白大亚基	<i>rpl2*</i> , <i>rpl14</i> , <i>rpl16*</i> , <i>rpl20</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	RNA 聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoCI*</i> , <i>rpoC2</i>
	核糖体蛋白小亚基	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12*</i> , <i>rps14</i> , <i>rps15</i> , <i>rps18</i> , <i>rps19</i>
	ATP 合成酶	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
其他基因	光系统 II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	NADH 脱氢酶	<i>ndhA*</i> , <i>ndhB*</i> , <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>
	细胞色素 b/f 复合体	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	光系统 I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>ycf3**</i>
未知	Rubisco 亚基	<i>rbcL</i>
	蛋白酶	<i>clpP*</i>
	成熟酶	<i>matk</i>
	c 型细胞色素合成基因	<i>ccsA</i>
未知	包膜蛋白	<i>cemA</i>
	乙酰辅酶 a 羧化酶的亚单位	<i>accD</i>
	保守的开发阅读框	<i>ycf1</i> , <i>ycf2</i> , <i>ycf4</i>

*内含子为 1, **内含子为 2。

*intron 1, **intron 2.

除了色氨酸 (Trp) 和甲硫氨酸 (Met) 外, 大多数氨基酸密码子具有偏好性 (图 2)。把 16 种黄芪属植物和 1 种苜蓿属植物叶绿体基因组进行密码子偏好性比较分析, 可知 Nc 值范围在 53.03~54.13, 说

明叶绿体基因组的密码子偏好性较弱。叶绿体基因组的 GC3s 含量为 20.62%~23.44%, GC 含量范围均低于 50%, 说明豆科植物叶绿体基因的密码子偏向使用 A 和 U 这 2 种碱基 (表 4)。

表 4 17 种豆科植物叶绿体基因组密码子使用的总体特征

Table 4 Overall characteristics of codon usage of 17 legume species chloroplast genomes

物种	T3s	C3s	A3s	G3s	Nc	GC3s	GC
草木樨状黄芪	0.427 7	0.212 8	0.427 8	0.233 1	54.00	0.326 0	0.350 0
蒙古黄芪	0.425 6	0.206 3	0.440 7	0.219 1	53.04	0.314 0	0.350 0
细叶黄芪	0.422 5	0.208 8	0.433 7	0.231 0	53.87	0.322 0	0.350 0
乳白黄芪	0.428 1	0.211 6	0.432 6	0.229 5	53.82	0.323 0	0.349 0
斜茎黄芪	0.428 0	0.210 2	0.434 4	0.220 5	53.03	0.318 0	0.351 0
阿纳卡黄芪	0.425 2	0.210 1	0.437 3	0.223 9	53.55	0.319 0	0.351 0
酒泉黄芪	0.425 7	0.218 1	0.430 0	0.226 8	53.73	0.327 0	0.351 0
岩生黄芪	0.425 1	0.217 9	0.429 1	0.226 5	53.69	0.327 0	0.351 0
舌尖黄芪	0.426 1	0.208 9	0.433 2	0.228 5	53.85	0.321 0	0.350 0
弯花黄芪	0.420 0	0.217 4	0.426 2	0.231 7	54.07	0.329 0	0.351 0
膜荚黄芪	0.420 7	0.209 1	0.431 0	0.234 8	53.69	0.326 0	0.351 0
加拿大黄芪	0.428 1	0.210 9	0.435 4	0.220 9	53.33	0.318 0	0.350 0
糙叶黄芪	0.427 4	0.221 9	0.434 2	0.206 2	53.09	0.318 0	0.350 0
胶黄芪	0.423 9	0.219 6	0.426 6	0.234 4	54.09	0.332 0	0.351 0
背扁黄芪	0.425 5	0.214 2	0.428 9	0.232 3	54.13	0.327 0	0.354 0
苜蓿	0.420 4	0.220 3	0.431 3	0.224 1	53.93	0.328 0	0.354 0
AMCP	0.423 8	0.219 3	0.429 0	0.226 4	54.02	0.328 0	0.351 0

3.3 简单重复序列 SSR 分析

共检测到 AMCP SSR 位点 263 个, 包括 149 个单核苷酸重复序列、90 个二核苷酸重复序列、12 个三核苷酸重复序列、11 个四核苷酸重复序列、1 个五核苷酸重复序列。测序的蒙古黄芪以单核苷酸 A/T 重复为主, 占比总 SSR 位点数的 56.27% (表 5)。同时对 16 种黄芪属植物和 1 种苜蓿属植物叶绿体基因组进行简单重复序列分析, 检测到 SSR 位点 241~268, 其中 A/T 占比范围为 55.11%~71.72%, 说明可知豆科植物多数以 A/T 重复为主, 少有其他类型的重复 (表 6)。

表 5 蒙古黄芪叶绿体基因组的 SSRs

Table 5 SSRs in chloroplast genome of AMCP

重复序列	总数	占比/%
A/T	148	56.27
C/G	1	0.38
AC/GT	1	0.38
AG/CT	11	4.18
AT/AT	78	29.66
AA/TT	12	4.56
AAAC/GTTT	1	0.38
AAAG/CTTT	1	0.38
AAAT/ATTT	3	1.14
AAGT/ACTT	1	0.38
AATT/AATT	1	0.38
ACCT/AGGT	1	0.38
AGAT/ATCT	3	1.14
AATAT/ATATT	1	0.38

表 6 17 种豆科植物叶绿体基因组的 SSR 比较分析

Table 6 Comparative analysis of SSR of 17 legume species chloroplast genomes

物种	总数	核苷酸						主要重复	A/T 占比/%
		1	2	3	4	5	6		
草木樨状黄芪	254	140	84	14	14	1	1	A/T	55.12
蒙古黄芪	262	149	89	12	11	1	0	A/T	56.49
细叶黄芪	245	142	77	10	14	1	1	A/T	57.96
乳白黄芪	267	158	87	10	11	1	0	A/T	58.43
斜茎黄芪	244	144	78	7	13	2	0	A/T	58.61
阿纳卡黄芪	267	152	86	15	12	2	0	A/T	56.55
酒泉黄芪	241	161	61	2	13	3	1	A/T	65.98
岩生黄芪	254	152	82	7	11	2	0	A/T	59.45
舌尖黄芪	247	139	82	16	9	1	0	A/T	55.06
弯花黄芪	268	149	94	9	10	3	3	A/T	55.22
膜荚黄芪	266	152	86	14	12	2	0	A/T	56.77
加拿大黄芪	258	148	81	16	13	0	0	A/T	56.59
糙叶黄芪	254	152	79	12	10	1	0	A/T	58.66
胶黄芪	265	149	90	13	12	1	0	A/T	55.85
背扁黄芪	243	163	73	3	3	0	1	A/T	66.26
苜蓿	244	178	52	8	4	2	0	A/T	71.72
AMCP	263	149	90	12	11	1	0	A/T	56.27

3.4 叶绿体全基因组比较分析

以蒙古黄芪 (OR712437) 的叶绿体基因组序列作为参考, 对其余 16 种豆科植物进行叶绿体基因组进行比较分析 (图 3)。结果表明, 17 条叶绿体基因组的 4 个基因区大致相同, 差异性较小。从非编码区和编码区看, 非编码区域序列变异高

于编码区域, 但在 *ycf1*、*ycf2* 等基因编码区变异程度较大。*trnF-GAA*~*trnT-UGU*、*trnfM-CUA*~*psbC*、*trnE*~*trnD*、*psbM*~*rpoB*、*trnR-UCU*~*trnQ-UUG*、*petL*~*rpl133*、*rps7*~*trnV-GAC*、*rrn5s*~*trnGUU*、*trn-UAG*~*ndhF* 等区间的存在较大的差异。

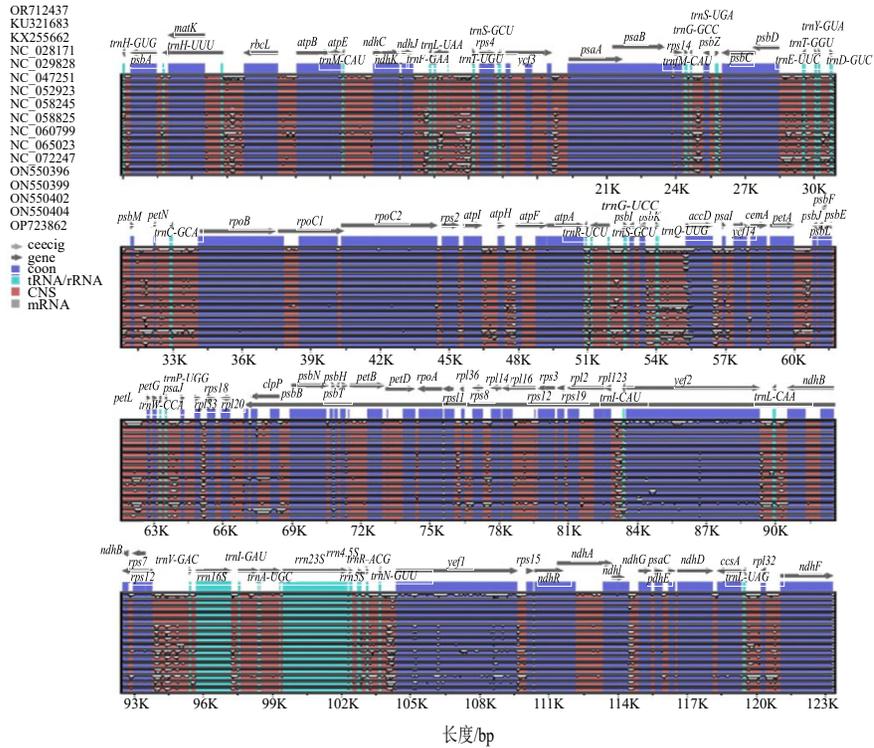


图 3 17 种豆科植物叶绿体基因组对比分析

Fig. 3 Alignment analysis of 17 legume species chloroplast genomes

3.5 系统发育分析

通过构建黄芪属的系统进化树, 了解蒙古黄芪 AMCP 在黄芪属中的近缘关系, 将其和 16 种豆科植物的叶绿体基因组进行亲缘性分析。结果显示豆科 2 个属的 17 种植物具有明显的进化关系, 可以聚为 3 大类, 第一支由细叶黄芪 (OP723862)、加拿大黄芪 (N_C060799)、草木樨状黄芪 (N_C072247)、岩生黄芪 (ON550399) 共 14 个物种组成, 第二支由背扁黄芪 (N_C065023) 和糙叶黄芪 (N_C058245) 组成, 苜蓿 (KU321683) 单独聚为一类。由系统进化树可知, 蒙古黄芪 (OR712437)、胶黄芪 (N_C047251)、阿纳卡黄芪 (N_C028171)、膜荚黄芪 (KX255662) 亲缘关系最近, 其中蒙古黄芪 (OR712437) 与胶黄芪 (N_C047251) 的自展支持率为 100%, 说明两者的遗传关系最近 (图 4)。

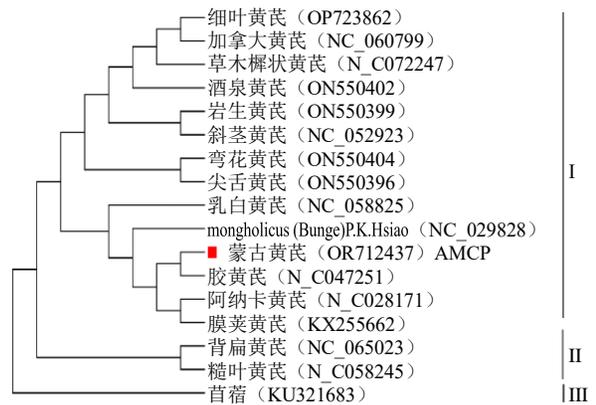


图 4 基于 17 种物种叶绿体全基因的 NJ 系统发育树

Fig. 4 NJ phylogenetic tree based on 17 species of complete chloroplast genomes

4 讨论

大多数被子植物的叶绿体基因组呈现典型的四分体结构, 而部分豆科植物叶绿体基因组由于在进

化过程中发生多次重排,从而导致缺失1个IR区被归为IRLC类群^[18]。本研究测序获得蒙古黄芪叶绿体基因组缺少1个IR区,呈现出非典型的四分体结构,故蒙古黄芪属于豆科蝶形花亚科的IRLC类群,与雷万钧^[16]研究的结论一致。同时,蒙古黄芪叶绿体基因组上存在未知功能的基因,如 $ycf1$ 、 $ycf2$ 和 $ycf4$,还有待研究。本研究测序获得的蒙古黄芪有29个tRNA,与2016年测序获得蒙古黄芪的tRNA在数量和种类上存在差异。这反映出蒙古黄芪基因组的多样性,也为应对不同的生物学需求,从而导致tRNA的数量和种类发生变化,以满足特定的代谢和蛋白质合成需求,同时反应出叶绿体基因组的检测手段不断发展完善。

密码子偏好性分析在植物叶绿体基因组的蛋白质编码基因过程中起着重要的作用,与突变、自然选择和随机遗传漂变等分子进化现象密切相关^[19]。大多数氨基酸可以同时被多种密码子编码,表现出不同生物对密码子的使用具有一定的偏好性^[20]。蒙古黄芪叶绿体基因组中64个蛋白编码基因共编码出41116个密码子,除色氨酸(Trp)和甲硫氨酸(Met)的RSCU=1,有29个氨基酸的RSCU>1。17种豆科植物的Nc值范围为53.04~54.13,说明叶绿体基因组的密码子使用偏好性较弱。叶绿体基因组的GC含量和GC3s含量均小于50%,可知豆科植物叶绿体基因组的密码子偏向使用A和U这两种碱基,这与Nie等^[21]关于双子叶植物叶绿体基因组中的密码子偏好性分析的结论一致,表明密码子使用在双子叶植物遗传进化中具有保守性和普遍适用性。其中,GC含量和GC3s含量相似可以推断出这17个物种在进化过程中存在一定的关联。

SSR也被称为微卫星,高度变异且数量丰富,具有保守性、多态性高、分布广泛等特点,广泛应用于物种鉴定,遗传多样性、分子辅助育种等方面^[22-24]。本研究通过对蒙古黄芪(OR712437)和其余的16种豆科叶绿体全基因组序列进行简单重复序列分析,发现以A/T碱基重复为主,这与张洁等^[25]和向如双等^[26]关于豆科植物重复序列研究结论相吻合。通过了解重复序列的分布特点可以进一步研究豆科植物基因组的整体结构和基因的稳定性,从而用于研究种间和种内的亲缘关系和统发育关系。

叶绿体基因组为系统发育提供了很多重要的分子片段^[27],多种基因、内含子和基因间隔区已经被使用在不同分类阶元的系统发育重建过程中,如

$atpB$ 、 $atpB-rbcL$ 、 $matK$ 、 $ndhF$ 、 $rbcL$ 等^[28]。植物DNA条形码常用于植物鉴定、植物分类的研究中,如 $atpF$ 、 $matK$ 、 $psb-I$ 、 $rbcL$ 、 $rpoB$ 、 $rbcL$ 、 $trnH-psbA$ 等片段是研究植物DNA条形码的重要选择^[29-30]。常晶茹等^[31]用 $ITS2$ 、 $matK$ 、 $rbcL$ 和 $psbA$ 共4种片段对吉林产的10种黄芪属植物进行DNA条形码与聚类分析,其结论为单独用某一片段只能鉴别出部分植物,不能把10种黄芪属植物全部鉴别出来。本研究基于mVISTA的叶绿体基因组序列对比分析得到 $trnF-GAA\sim trnT-UGU$ 、 $trnM-CUA\sim psbC$ 、 $trnE\sim trnD$ 、 $psbM\sim rpoB$ 、 $ycf1$ 和 $ycf2$ 等编码区的差异性大,也可以用于黄芪属植物DNA条形码的开发,为黄芪属药材品种鉴别、种质资源评价以及种苗筛选提供有效工具,达到解决种源混杂的目的。

本研究采用邻接法构建系统发育树,所获得系统发育树中各类群间的拓扑结构与经典分类学的观点一致:外类群苜蓿单独聚为一类,剩下的黄芪属聚为一大类,其中胶黄芪(N_C047251)、阿纳卡黄芪(N_C028171)在《中国植物志》没有收录,对于两者属于何种亚属目前缺少研究。根据本研究,蒙古黄芪(OR712437)、胶黄芪、阿纳卡黄芪、膜荚黄芪(KX255662)聚为一类,可以推测出胶黄芪、纳凯黄芪与蒙古黄芪(OR712437)、膜荚黄芪的遗传关系相近,与Tian等^[32]聚类结果相似,为后续研究胶黄芪、阿纳卡黄芪的分类提供参考。

本研究通过测序、组装和文库构建,对蒙古黄芪叶绿体全基因组特征进行描述,分析了黄芪属植物叶绿体基因组的序列特征并构建系统发育树,为进一步探究黄芪属植物的遗传结构、遗传多样性和亲缘性奠定了基础。同时,丰富了黄芪属药用植物的叶绿体基因组数据,为黄芪属的物种鉴别、群体结构分类和进化发育等研究提供科学参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 侯敏娜,侯少平,杨枫玉,等.不同产地蒙古黄芪补气主成分分析比较[J].化学与生物工程,2021,38(12):64-67.
- [2] 中国药典[S].一部.2020:315.
- [3] 马艳春,胡建辉,吴文轩,等.黄芪化学成分及药理作用研究进展[J].中医药学报,2022,50(4):92-95.
- [4] 王玲,董文攀,周世良,等.被子植物叶绿体基因组的结构变异研究进展[J].西北植物学报,2015,32(6):1282-1288.
- [5] 丁怡宁,毕光耀,胡赛文,等.红花不同花色变异类型

- 叶绿体基因组特征与系统进化分析 [J]. 中草药, 2023, 54(1): 262-271.
- [6] Li D, Liu Y T, Qin X M. Rapid quantitative analysis of 12 chemical constituents in wild-simulated and cultivated *Astragalus Radix* based on UHPLC-MS [J]. *Chin Herb Med*, 2022, 14(3): 464-469.
- [7] 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.
- [8] 刘畅, 曾宪法, 杨小英, 等. 多叶勾儿茶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.
- [9] Twyford A D, Ness R W. Strategies for complete plastid genome sequencing [J]. *Mol Ecol Resour*, 2017, 17(5): 858-868.
- [10] 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.
- [11] Ren F M, Wang L Q, Zhuo W, et al. Complete chloroplast genome of the rare medicinal vegetable *Allium hookeri* [J]. *Mitochondrial DNA B Resour*, 2022, 7(1): 6-7.
- [12] Yi S Y, Lu H B, Wang W, et al. The chloroplast genome of wild *Saposhnikovia divaricata*: Genomic features, comparative analysis, and phylogenetic relationships [J]. *Genes*, 2022, 13(5): 931.
- [13] 王涛, 田茜, 赵宇亮, 等. 黄芪叶绿体基因组密码子偏好性分析 [J]. 分子植物育种, 2023, 23(3): 15-42.
- [14] Lei W J, Ni D P, Wang Y J, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus* [J]. *Sci Rep*, 2016, 6: 21669.
- [15] Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging [J]. *BMC Res Notes*, 2016, 9: 88.
- [16] 雷万钧. 蒙古黄芪叶绿体基因组研究 [D]. 太原: 山西农业大学, 2016.
- [17] Minh B Q, Nguyen M A T, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap [J]. *Mol Biol Evol*, 2013, 30(5): 1188-1195.
- [18] 田春育, 李志勇, 刘倩, 等. 苜蓿属不同物种叶绿体基因组结构比较及亲缘关系分析 [J]. 中国草地学报, 2021, 43(10): 1-8.
- [19] Dong S J, Ying Z Q, Yu S S, et al. Complete chloroplast genome of *Stephania tetrandra* (Menispermaceae) from Zhejiang Province: Insights into molecular structures, comparative genome analysis, mutational hotspots and phylogenetic relationships [J]. *BMC Genomics*, 2021, 22(1): 880.
- [20] 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.
- [21] Nie X J, Deng P C, Feng K W, et al. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family [J]. *Plant Mol Biol Report*, 2014, 32(4): 828-840.
- [22] Yuan Q J, Zhang Z Y, Hu J, et al. Impacts of recent cultivation on genetic diversity pattern of a medicinal plant, *Scutellaria baicalensis* (Lamiaceae) [J]. *BMC Genet*, 2010, 11: 29.
- [23] Powell W, Morgante M, McDevitt R, et al. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines [J]. *Proc Natl Acad Sci U S A*, 1995, 92(17): 7759-7763.
- [24] Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers [J]. *Nucleic Acids Res*, 1989, 17(16): 6463-6471.
- [25] 张洁, 陆嘉惠, 王倩倩, 等. 新疆产黄甘草叶绿体基因组特征及甘草属药用植物叶绿体基因组比较与系统发育分析 [J]. 药学报, 2022, 57(5): 1516-1525.
- [26] 向如双, 孙伟, 段宝忠, 等. 刺果甘草全基因组 Survey 及叶绿体基因组特征分析 [J]. 世界科学技术—中医药现代化, 2022, 24(5): 1781-1790.
- [27] 王银环. 豆科的叶绿体系统发育基因组学 [D]. 昆明: 云南大学, 2017.
- [28] Peterson P M, Romaschenko K, Johnson G. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees [J]. *Mol Phylogenet Evol*, 2010, 55(2): 580-598.
- [29] Kress W J, Erickson D L. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region [J]. *PLoS One*, 2007, 2(6): e508.
- [30] Newmaster S G, Subramanyam R. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae) [J]. *Mol Ecol Resour*, 2009, 9(Suppl s1): 172-180.
- [31] 常晶茹, 姚萱航, 张雪薇, 等. 黄芪属植物 DNA 条形码与聚类分析的研究 [J]. 中草药, 2022, 53(22): 7201-7206.
- [32] Tian C, Li X, Wu Z, et al. Characterization and comparative analysis of complete chloroplast genomes of three species from the genus *Astragalus* (Leguminosae) [J]. *Front Genet*, 2021, 12: 705482.