

祁连圆柏叶绿体基因组序列特征分析

赵明¹, 张宏斌¹, 李伟², 赵祐¹, 刘建海¹, 赵国生¹, 赵兴鹏¹, 吕东^{1,①}

(1. 甘肃省祁连山水源涵养林研究院, 甘肃 张掖 734000; 2. 北京林业大学生物科学与技术学院, 北京 100083)

摘要: 基于 Oxford Nanopore PromethION 测序仪对祁连圆柏 (*Juniperus przewalskii* Komarov) 叶绿体基因组的测序结果, 对原属于圆柏属 (*Sabina* Mill.)、现属于刺柏属 (*Juniperus* Linn.) 的祁连圆柏和高山柏 (*Juniperus squamata* Buchanan-Hamilton ex D. Don) 等 5 个树种的叶绿体基因组进行了序列对比和反向重复区 (IR) 边界分析, 并对柏科 (Cupressaceae) 15 个树种进行了系统发育分析。结果表明: 祁连圆柏叶绿体基因组为环状四分体结构, 但四分体结构不明显, 基因组全长 127 317 bp, 共编码 118 个基因; 密码子共 23 708 个, 且偏好使用 A 或 U; 简单重复序列共 289 个, 以 AA 或 TT 为主; 长重复序列共 38 个, 其中正向重复序列最为丰富。参与比较分析的 5 个树种的叶绿体基因组具有高度相似性, 存在显著变异的区域主要集中在非编码序列; 5 个树种的 IR 区表现出明显缩减现象, 长度在 257~265 bp 之间。系统发育分析结果显示: 原属于圆柏属的物种与刺柏属物种亲缘关系较近, 且 2 个属内的部分物种在系统发育树中的位置出现交错现象, 其中祁连圆柏与高山柏聚为一支, 亲缘关系最近。综合研究结果表明: 祁连圆柏叶绿体基因组偏小, 表现为不典型的四分体结构, IR 区发生明显缩减; 在叶绿体基因组编码的氨基酸序列中, Leu 数量最多, AAA (编码 Lys) 是使用最多的密码子; 原属于圆柏属的物种叶绿体基因组具有高度保守性, 高变异区域只存在于单拷贝区域, 鉴定出的 *accD*、*ycf1* 和 *ycf2* 3 个存在显著变异的基因可作为高变异区, 为界定密切相关的分类群提供丰富的系统发育信息。

关键词: 祁连圆柏; 叶绿体基因组; 刺柏属; 序列比对; 系统发育树

中图分类号: Q943.2; Q949.66⁺7; S718.49 文献标志码: A 文章编号: 1674-7895(2023)03-0001-11

DOI: 10.3969/j.issn.1674-7895.2023.03.01

Analysis on chloroplast genome sequence characteristics of *Juniperus przewalskii* ZHAO Ming¹, ZHANG Hongbin¹, LI Wei², ZHAO Hu¹, LIU Jianhai¹, ZHAO Guosheng¹, ZHAO Xingpeng¹, LYU Dong^{1,①} (1. Gansu Province Academy of Qilian Water Resource Conservation Forests Research Institute, Zhangye 734000, China; 2. College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China), *J. Plant Resour. & Environ.*, 2023, 32(3): 1-11

Abstract: Based on the sequencing result of chloroplast genome of *Juniperus przewalskii* Komarov by using Oxford Nanopore PromethION sequencer, sequence alignment and inverted repeat (IR) region boundary analysis were conducted for chloroplast genomes of five tree species including *J. przewalskii* and *Juniperus squamata* Buchanan-Hamilton ex D. Don, which originally belong to *Sabina* Mill. and now belong to *Juniperus* Linn., and phylogenetic analysis was performed for 15 tree species in Cupressaceae. The results show that the chloroplast genome of *J. przewalskii* is a loop tetrad structure, but the tetrad structure is not evident, and the total length of genome is 127 317 bp, encoding 118 genes in total; there are 23 708 codons in total, and A or U is preferred; there are 289 simple repeat sequences in total, and are mainly AA or TT; there are 38 long repeat sequences in total, in which forward repeat sequence is the most abundant. There are high similarities among chloroplast genomes of five tree species participated in comparative analysis, and the regions with significant variation are mainly concentrated in noncoding

收稿日期: 2022-12-02

基金项目: 甘肃省科技计划项目创新基地和人才计划(22JR10KA026); 国家自然科学基金项目(31860210; 31860221)

作者简介: 赵明(1977—), 男, 甘肃张掖人, 本科, 高级工程师, 主要从事林木遗传育种及良种繁育方面的研究。

①通信作者 E-mail: jiuwubaba@163.com

引用格式: 赵明, 张宏斌, 李伟, 等. 祁连圆柏叶绿体基因组序列特征分析[J]. 植物资源与环境学报, 2023, 32(3): 1-11.

sequence; IR regions of five tree species show an obvious reduction phenomenon, and the lengths are within 257–265 bp. The phylogenetic analysis result shows that species originally belong to *Sabina* has a relatively close genetic relationship with species in *Juniperus*, and the positions of some species in two genera in the phylogenetic tree show an interlaced phenomenon, in which *J. przewalskii* and *J. squamata* are clustered into one branch, and their genetic relationship is the closest. The comprehensive study result shows that the chloroplast genome of *J. przewalskii* is relatively small, which shows a nontypical tetrad structure, and the IR regions are obviously reduced; in the amino acid sequences encoded by chloroplast genome, Leu is the most abundant amino acid, and AAA (encoding Lys) is the most-used codon; the chloroplast genomes of species originally belong to *Sabina* are highly conserved, high variation regions only exist in single copy region, and three identified genes with significant variation namely *accD*, *ycf1*, and *ycf2* can be regarded as high variation regions, which can provide rich phylogenetic information for defining closely-associated taxonomic groups.

Key words: *Juniperus przewalskii* Komarov; chloroplast genome; *Juniperus* Linn.; sequence alignment; phylogenetic tree

祁连圆柏 (*Juniperus przewalskii* Komarov) 为柏科 (Cupressaceae) 刺柏属 (*Juniperus* Linn.) 乔木, 原属于圆柏属 (*Sabina* Mill.), 是中国特有树种, 主要分布于中国青海、甘肃河西走廊及南部、四川北部地区, 常生于海拔 2 600~4 000 m 的阳坡地带^[1]。祁连圆柏具有适应性强、耐寒、耐瘠薄和根系发达等优点, 是祁连山干旱、半干旱地区造林的主要常绿针叶树种之一^[2]。目前, 有关祁连圆柏的研究主要集中在树木年轮气候学^[3]、生理生态学^[4]、苗木生产和造林技术^[5]以及生化成分^[6]等方面, 关于物种群落特征的研究也有涉及, 例如海拔^[7]和坡向^[8]等地形因子对群落物种多样性的影响。但目前未见其叶绿体基因组信息的有关报道, 开展祁连圆柏及相关物种的叶绿体基因组研究有利于柏科植物进化分类学和种群遗传学的研究。

叶绿体作为绿色植物典型的细胞器, 遗传信息丰富, 且具有单亲遗传的特性, 有助于区分分类群和推断进化关系^[9]。被子植物以及大部分裸子植物, 包括苏铁科 (Cycadaceae)、银杏科 (Ginkgoaceae) 和买麻藤科 (Gnetaceae) 植物等, 其叶绿体基因组是一个典型的环状四分体结构, 包括 1 个大单拷贝区 (LSC)、1 个小单拷贝区 (SSC) 和 2 个互补的反向重复区 (IR)^[10]。但松柏类植物在进化过程中, IR 区发生了显著的缩减和丢失现象, IR 区的缩减和丢失也成为松柏类植物的叶绿体特征之一^[11]。在之前的研究中, 对于刺柏属和圆柏属的分开与合并一直存在争议, 有学者将其分为圆柏属和刺柏属^[12], 也有学者将刺柏属分为刺柏组 (Sect. *Juniperus*)、圆柏组 (Sect. *Sabina*) 和叙利亚刺柏组 (Sect. *Caryocedrus*)^[13-14]。

目前, 随着多个圆柏属物种更名并划分进刺柏属, 有关这 2 个属的划分又出现了新的转折点。随着测序技术的发展, 柏科植物中已有多个物种的叶绿体基因组被报道^[15-16]。完整的叶绿体基因组序列为叶绿体基因组特征及系统发育关系的研究提供了参考数据, 进而有助于了解物种的分类、进化和遗传多样性。

鉴于此, 作者利用 Oxford Nanopore 技术完成了祁连圆柏叶绿体基因组的测序, 描述了祁连圆柏叶绿体基因组特征, 并与 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中已报道的 4 个原属于圆柏属、现属于刺柏属的物种叶绿体基因组进行了比较。此外, 基于柏科 15 个树种的叶绿体基因组进行了系统发育分析, 以期为后续刺柏属植物的分子标记开发以及柏科植物分类和遗传多样性等方面的研究奠定基础。

1 材料和方法

1.1 材料

供试材料祁连圆柏分布于中国祁连山脉 (东经 94°10'~103°04'、北纬 35°50'~39°19'), 树龄约 8 a, 于 2022 年 3 月随机采集 5 个单株的新梢嫩叶, 每株采集 12 枚叶片, 混合后置于干冰中保存。样品储存于北京林业大学林木分子设计育种高精尖创新中心 (编号为 YL-382072)。

1.2 方法

1.2.1 总 DNA 提取与测序 利用改良 CTAB 法^[17]提取祁连圆柏 DNA 备用。DNA 样品检测合格后, 使用超声波破碎仪 (功率 100 W, 美国 Covaris 公司) 随机打断, 经末端修复、加 A 尾、加测序接头、纯化、PCR

扩增等步骤完成 DNA 文库制备,质检合格的 DNA 文库经 Oxford Nanopore PromethION 测序仪(英国 Oxford Nanopore Technologies 公司)测序,测序工作由武汉贝纳科技服务有限公司完成。

1.2.2 基因组组装与注释 利用 Flye v2.8.3 在线工具(<https://github.com/fenderglass/Flye>)^[18]对得到的有效数据进行基因组拼接,以欧洲刺柏(*Juniperus communis* Linn.) (GenBank 登录号为 NC_035068.1) 的叶绿体基因组作为近缘参考基因组进行比对,基于比对情况确定候选序列组装结果,以获得完整的叶绿体基因组。将祁连圆柏完整的叶绿体基因组数据上传至 NCBI 数据库,GenBank 登录号为 OM780071。利用 CPGAVAS2 v2 注释软件(<http://47.96.249.172:16019/analyzer/annotate/>)^[19]进行基因注释及基因组图谱的绘制。根据基因组中 G 和 C 所占的比例计算 GC 含量。

1.2.3 密码子偏好性分析 利用 CodonW v1.4.4 在线工具(<http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html/>)对祁连圆柏叶绿体基因组密码子偏好性进行分析,估算相对同义密码子使用度(RSCU),其中对长度大于 300 bp,且以 ATG、TTG、CTG、ATT、ATC、CTG 和 ATA 为起始,TGA、TAG 和 TAA 为终止的基因序列进行密码子偏好性分析;RSCU 大于 1.00 表示偏好使用该密码子,RSCU 等于 1.00 表示使用密码子无偏好性,RSCU 小于 1.00 表示不偏好使用该密码子^[20]。

1.2.4 简单重复序列和长重复序列分析 利用微卫星标记鉴定工具(MISA)(<https://webblast.ipk-gatersleben.de/misa/>)^[21]搜索祁连圆柏叶绿体基因组中的简单重复序列,用于识别单核苷酸、双核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸的最小重复数分别为 8、4、4、3、3 和 3;2 个简单重复序列之间的间距不超过 100 bp 的序列为复合重复序列(compound repeat sequence)。利用 Vmatch 在线工具(<http://www.vmatch.de/>)查找叶绿体基因组中的长重复序列片段,设置最小重复单元长度为 30 bp。

1.2.5 柏科 5 个树种叶绿体基因组的比较 在 NCBI 数据库下载铺地柏(*Juniperus procumbens* (Endlicher) Siebold ex Miquel)、落基山圆柏(*Juniperus scopulorum* Sarg.)、圆柏(*Juniperus chinensis* Linn.) 和高山柏(*Juniperus squamata* Buchanan-Hamilton ex D. Don) 4 个原属于圆柏属、现属于刺柏属的树种的叶绿体基因

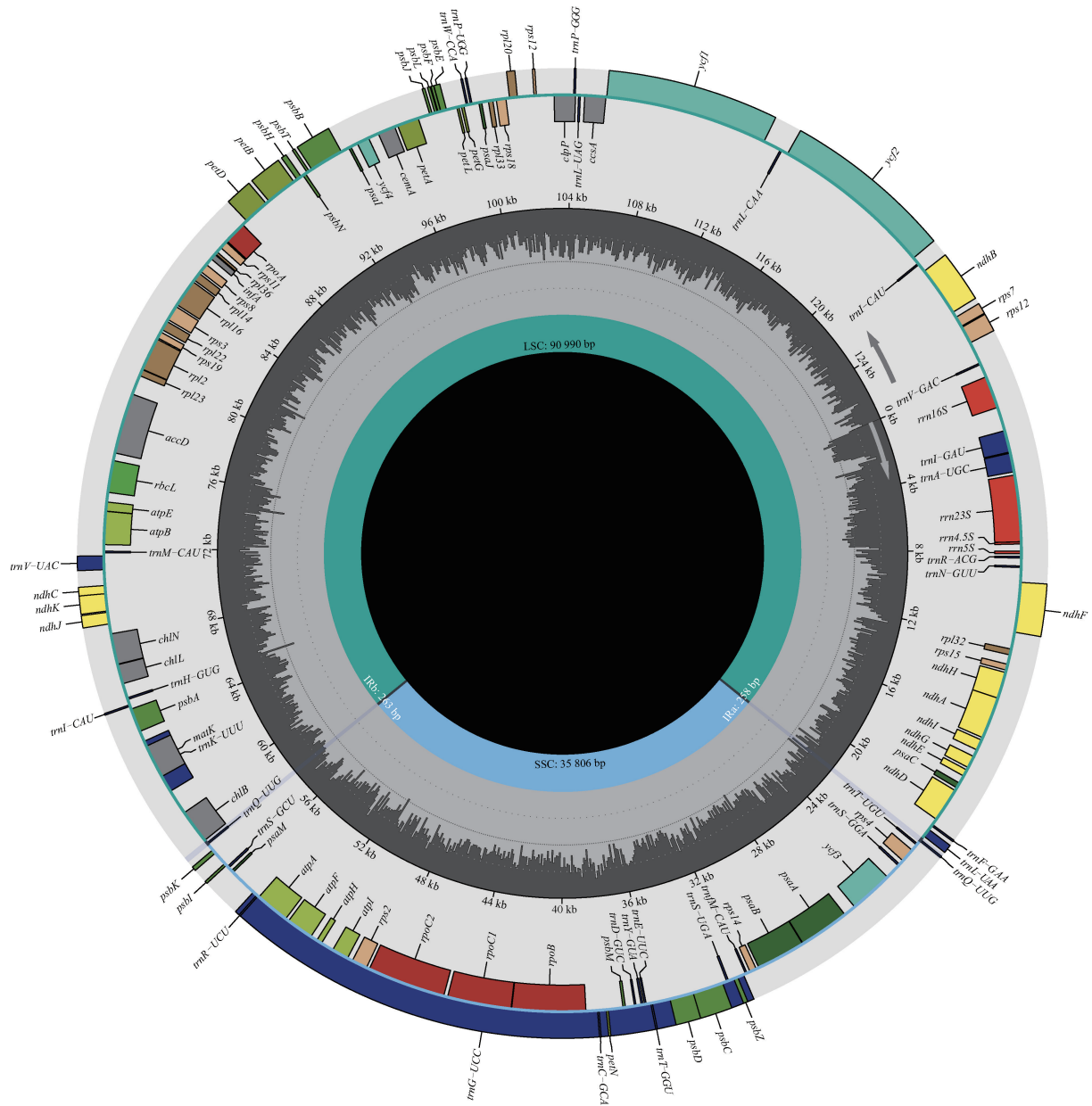
组序列,利用 mVISTA v7 在线工具(<https://genome.lbl.gov/vista/mvista/submit.shtml>)^[22]进行多序列比对,以祁连圆柏叶绿体基因组为参考,在 Shuffle-LAGAN 模式下比较上述柏科 5 个树种叶绿体基因组的差异。在 JSHYCloud 平台(<http://cloud.geneioneer.com:9929/>)上利用 Organelle Analysis 程序进行 IR 区边界的差异分析。

1.2.6 柏科植物的系统发育分析 在 NCBI 数据库中选取松科(Pinaceae)树种赤松(*Pinus densiflora* Sieb. et Zucc.)作为外类群,分析刺柏属的大果圆柏(*Juniperus tibetica* Komarov)和北美圆柏(*J. virginiana* Linn.),原属于圆柏属的祁连圆柏、铺地柏、落基山圆柏、圆柏和高山柏,翠柏属(*Calocedrus* Kurz)的岩生翠柏(*C. rupestris* Aver., T. H. Nguyễn et P. K. Lôc)和台湾翠柏[*C. formosana* (Florin) Florin],扁柏属(*Chamaecyparis* Spach)的红桧(*C. formosensis* Matsum.)和日本花柏[*C. pisifera* (Sieb. et Zucc.) Endl.],柏木属(*Cupressus* Linn.)的岷江柏木(*C. chengiana* S. Y. Hu)和千香柏(*C. duclouxiana* Hickel),以及崖柏属(*Thuja* Linn.)的崖柏(*T. sutchuenensis* Franch.)和北美香柏(*T. occidentalis* Linn.)的系统发育关系。利用 TBtools v0.665 在线工具(<https://github.com/CJ-Chen/TBtools/>)^[23]在上述柏科 15 个树种及赤松的完整序列中提取编码序列,利用 MAFFT v7.037 软件对上述处理后的序列进行多序列比对^[24]。利用 IQ-TREE v1.6.12 软件中的 ModelFinder 模块选出最适合的替代模型为 TVM+F+R2,然后基于此模型推断具有 1 000 个自举重复序列的最大似然(ML)系统发育树^[25],最后利用 iTOL v5 在线工具(<https://itol.embl.de/>)^[26]对得到的系统发育树进行编辑。

2 结果和分析

2.1 祁连圆柏叶绿体基因组特征

测序结果显示:共得到 31 885 698 条 reads,将低质量 reads(平均质量小于 7 的序列)过滤后得到 31 885 188 条 reads,拼接后得到的祁连圆柏叶绿体基因组为环状四分体结构(图 1),包括 1 个大单拷贝区(LSC)、1 个小单拷贝区(SSC)以及 2 个反向重复区(IR),但四分体结构不明显,主要表现为 IR 区发生明显缩减现象。叶绿体基因组全长 127 317 bp,总



■: 光系统 I Photosystem I ; ■: 光系统 II Photosystem II ; ■: 细胞色素 b/f 复合体 Cytochrome b/f complex ; ■: ATP 合成酶 ATP synthase ; ■: NADH 脱氢酶 NADH dehydrogenase ; ■: 核酮糖-1,5-二磷酸羧化酶大亚基 Large subunit of rubisco ; ■: RNA 聚合酶 RNA polymerase ; ■: 核糖体小亚基 Small subunit of ribosome ; ■: 核糖体大亚基 Large subunit of ribosome ; ■: rRNA ; ■: tRNA ; ■: 其他基因 Other genes ; ■: 保守开放阅读框 Conserved open reading frame. LSC: 大单拷贝区 Large single copy region; SSC: 小单拷贝区 Small single copy region; IR: 反向重复区 Inverted repeat region. 内圈深灰色表示 GC 含量 The dark gray part in the inner circle represents GC content.

图 1 祁连圆柏叶绿体基因组图谱
Fig. 1 Chloroplast genome map of *Juniperus przewalskii* Komarov

GC 含量为 35.1%;其中,LSC 区和 SSC 区长度分别为 90 990 和 35 806 bp,而 2 个 IR 区长度仅为 258 和 263 bp。LSC 区、SSC 区和 IR 区的 GC 含量分别为 34.2%、36.8%和 34.2%。

基因注释结果(表 1)显示:祁连圆柏叶绿体基因组共编码 118 个基因,包括 82 个蛋白质编码基因、32

个 tRNA 基因和 4 个 rRNA 基因。共有 14 个基因含有内含子,包括 8 个蛋白质编码基因和 6 个 tRNA 基因,分别为 *petB*、*petD*、*atpF*、*rpoC1*、*rps12*、*rpl2*、*rpl16* 和 *ycf3* 以及 *trnA-UGC*、*trnG-UCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA* 和 *trnV-UAC* 基因,其中,*rps12* 和 *ycf3* 基因含有 2 个内含子,其余 12 个基因含有 1 个内含子。

表1 祁连圆柏叶绿体基因组注释信息

Table 1 Annotation information of chloroplast genome of *Juniperus przewalskii* Komarov

基因 ¹⁾ Gene ¹⁾	数量 Number	功能 Function
光合作用基因 Photosynthesis genes		
<i>psaA, psaB, psaC, psaI, psaJ, psaM</i>	6	光系统 I Photosystem I
<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15	光系统 II Photosystem II
<i>petA, petB * , petD * , petG, petL, petN</i>	6	细胞色素 b/f 复合体 Cytochrome b/f complex
<i>atpA, atpB, atpE, atpF * , atpH, atpI</i>	6	ATP 合成酶 ATP synthase
<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	11	NADH 脱氢酶 NADH dehydrogenase
<i>rbcL</i>	1	核酮糖-1,5-二磷酸羧化酶大亚基 Large subunit of rubisco
自我复制基因 Self-replication genes		
<i>rpoA, rpoB, rpoC1 * , rpoC2</i>	4	RNA 聚合酶 RNA polymerase
<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12 ** , rps14, rps15, rps18, rps19</i>	11	核糖体小亚基 Small subunit of ribosome
<i>rpl2 * , rpl14, rpl16 * , rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>	9	核糖体大亚基 Large subunit of ribosome
<i>rrn4.5S, rrn5S, rrn16S, rrn23S</i>	4	rRNA
<i>trnA-UGC * , trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-UCC * , trnH-GUG, trnI-CAU (×2), trnI-GAU * , trnK-UUU * , trnL-CAA, trnL-UAA * , trnL-UAG, trnM-CAU, trnN-GUU, trnP-GGG, trnP-UGG, trnQ-UUG (×2), trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-UGU, trnT-GGU, trnV-GAC, trnV-UAC * , trnW-CCA, trnY-GUA</i>	32	tRNA
其他基因 Other genes		
<i>accD</i>	1	乙酰辅酶 A 羧化酶亚基 Subunit of acetyl-CoA carboxylase
<i>ccsA</i>	1	细胞色素 C Cytochrom C
<i>cemA</i>	1	囊膜蛋白 Envelop membrane protein
<i>clpP</i>	1	蛋白酶 Protease
<i>infA</i>	1	翻译起始因子 Translational initiation factor
<i>matK</i>	1	成熟酶 Maturase
<i>chlB, chlL, chlN</i>	3	叶绿素合成 Chlorophyll synthesis
未知功能基因 Unkown function genes		
<i>ycf1, ycf2, ycf3 ** , ycf4</i>	4	

¹⁾ * 含有 1 个内含子 Containing one intron; ** : 含有 2 个内含子 Containing two introns; (×2) : 2 份拷贝 Two copies.

2.2 祁连圆柏叶绿体基因组密码子偏好性分析

祁连圆柏叶绿体基因组共对应 23 708 个密码子,编码 20 种氨基酸。结果(表 2)显示:编码 Leu 的密码子数量最多,为 2 574 个;而编码 Pro、Tyr、Cys、His、Gln、Met 和 Trp 的密码子数量均少于 1 000 个。从相对同义密码子使用度(RSCU)看,蛋白质编码基因中有 30 种密码子的 RSCU 值大于 1,除 UUG 外均以 A 或 U 结尾,说明祁连圆柏叶绿体基因组蛋白质编码基因具有密码子偏好性,且偏好使用 A 或 U。

2.3 祁连圆柏叶绿体基因组简单重复序列和长重复序列分析

祁连圆柏叶绿体基因组简单重复序列统计结果见表 3,长重复序列统计结果见表 4。结果(表 3)显示:共检测到 21 类简单重复序列,共 289 个,长度在

8~22 bp 之间,其中,双核苷酸重复序列的数量最多,为 143 个,以 AA 或 TT 为主,说明祁连圆柏叶绿体基因组具有明显的碱基偏好性。其后按数量由高至低依次为单核苷酸重复序列(112)、四核苷酸重复序列(20)、三核苷酸重复序列(10)、五核苷酸重复序列(2)和六核苷酸重复序列(2),其中单核苷酸重复序列也以 A 或 T 为主。此外,还检测到 138 个复合重复序列。

结果(表 4)显示:祁连圆柏叶绿体基因组中检测到 3 类长重复序列,共 38 个,长度在 30~224 bp 之间,其中,正向重复(F)序列的数量最多,为 28 个,串联重复(T)和回文重复(P)序列的数量分别为 6 和 4 个。长重复序列的长度主要集中在 30~40 bp,这些序列占总序列数的 61%。

表2 祁连圆柏叶绿体基因组相对同义密码子使用度
Table 2 Relative synonymous codon usage of chloroplast genome of *Juniperus przewalskii* Komarov

氨基酸 Amino acid	密码子 Codon	数量 Number	相对同义密码子使用度 Relative synonymous codon usage	氨基酸 Amino acid	密码子 Codon	数量 Number	相对同义密码子使用度 Relative synonymous codon usage	氨基酸 Amino acid	密码子 Codon	数量 Number	相对同义密码子使用度 Relative synonymous codon usage
Phe	UUU	969	1.43	His	CAU	409	1.59	Val	AAG	373	0.49
	UUC	383	0.57		CAC	104	0.41		GUU	429	1.41
Ser	UCU	523	1.96	Arg	CGU	326	1.45	GUC	145	0.48	
	UCC	209	0.78		CGC	107	0.48	GUA	436	1.43	
	UCA	327	1.22		CGA	273	1.22	GUG	210	0.69	
	UCG	122	0.46		CGG	73	0.33	Ala	GCU	648	1.92
	AGU	322	1.21		AGA	447	1.99		GCC	161	0.48
Leu	AGC	100	0.37	Gln	AGG	120	0.53	GCA	422	1.25	
	UUA	863	2.01		CAA	675	1.57	GCG	120	0.36	
	UUG	522	1.22	Ile	CAG	185	0.43	Asp	GAU	842	1.61
	CUU	514	1.20		AUU	1 010	1.48		GAC	207	0.39
	CUC	169	0.39	AUC	338	0.50	Glu	GAA	1 161	1.60	
	CUA	362	0.84	AUA	694	1.02		GAG	292	0.40	
	Pro	CUG	144	0.34	Met	AUG	540	1.00	Gly	GGU	521
CCU		435	1.79	Thr		ACU	484	1.77		GGC	140
CCC		164	0.67	ACC	173	0.63	GGA	623	1.70		
CCA		250	1.03	ACA	318	1.16	GGG	181	0.49		
CCG		123	0.51	ACG	118	0.43	Tyr	UAU	705	1.58	
Trp	UGG	422	1.00	Asn	AAU	853		1.56	UAC	188	0.42
	Cys	UGU	196	1.44	AAC	243	0.44				
		UGC	76	0.56	Lys	AAA	1 163	1.51			

表3 祁连圆柏叶绿体基因组简单重复序列统计结果
Table 3 Statistical result of simple repeat sequences in chloroplast genome of *Juniperus przewalskii* Komarov

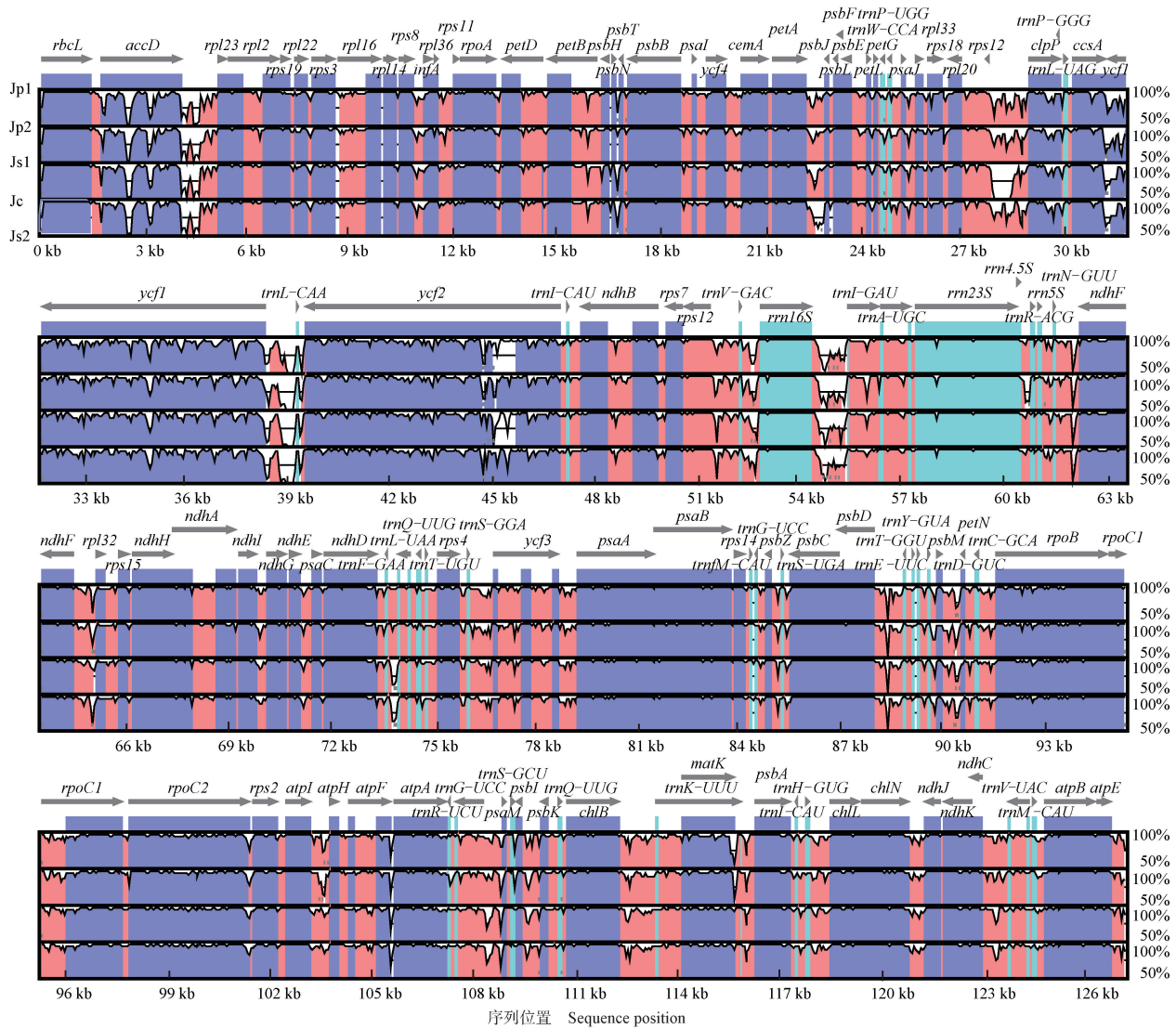
类型 Type	重复频率 Repeat frequency	数量 Number	类型 Type	重复频率 Repeat frequency	数量 Number	
A/T	8	48	AG/CT	4	11	
	9	24		5	1	
	10	17	CC/GG	4	2	
	11	14	AAA/TTT	4	6	
	12	1		5	1	
	13	2	AAG/CTT	6	1	
	14	3		7	1	
	15	1	AAT/ATT	4	1	
	C/G	8	1	AAAA/TTTT	4	7
		9	1	AAAC/GTTT	4	1
	AA/TT	4	72	AAAG/CTTT	4	2
		5	31	AAGT/ACTT	4	1
		6	3	ACAC/GTGT	4	1
		7	4	ACCT/AGGT	4	1
	AC/GT	4	3	ATAT/ATAT	4	6
6		1	ATCC/ATGG	4	1	
AT/AT	4	6	AAAAA/TTTTT	4	1	
	5	3	ACTAT/AGTAT	4	1	
	6	3	AAGAAG/CTTCTT	4	2	
	7	3				

表4 祁连圆柏叶绿体基因组长重复序列统计结果
Table 4 Statistical results of long repeat sequences in chloroplast genome of *Juniperus przewalskii* Komarov

类型 Type	不同长度序列的数量 Number of different length sequences			
	30-40 bp	41-50 bp	51-60 bp	≥61 bp
串联重复 Tandem repeat	5	0	0	1
回文重复 Palindromic repeat	2	0	0	2
正向重复 Forward repeat	16	7	1	4

2.4 柏科5个树种叶绿体基因组的比较

2.4.1 基因组序列比对 柏科5个树种叶绿体基因组的序列比对结果(图2)显示:祁连圆柏叶绿体基因组在基因组长度、基因内容和排列与铺地柏、落基山圆柏、圆柏和高山柏的叶绿体基因组几乎相同。5个树种叶绿体基因组在序列上整体比较保守,大部分编码基因以及所有 rRNA 变异不明显。存在显著变异的区域主要集中在基因非编码序列,编码序列(外显子)中存在显著变异的区域位于 *accD*、*ycf1* 和 *ycf2* 3个基因的位置。总体来看,由于反向重复区(IR)的缩减,高变异区域只存在于单拷贝区域,且大单拷贝区(LSC)明显高于小单拷贝区(SSC)。



Jp1: 祁连圆柏 *Juniperus przewalskii* Komarov; Jp2: 铺地柏 *Juniperus procumbens* (Endlicher) Siebold ex Miquel; Js1: 落基山圆柏 *Juniperus scopulorum* Sarg.; Jc: 圆柏 *Juniperus chinensis* Linn.; Js2: 高山柏 *Juniperus squamata* Buchanan-Hamilton ex D. Don. ■: 基因 Gene; ■: 外显子 Exon; ■: 非翻译区 Untranslated region; ■: 非编码序列 Noncoding sequence; ■: mRNA. 右侧百分比表示每个位点碱基在不同参考序列间的最高 (100%) 和最低 (50%) 保守性 The percentages on the right represent the highest (100%) and lowest (50%) conservation of each site base among different reference sequences.

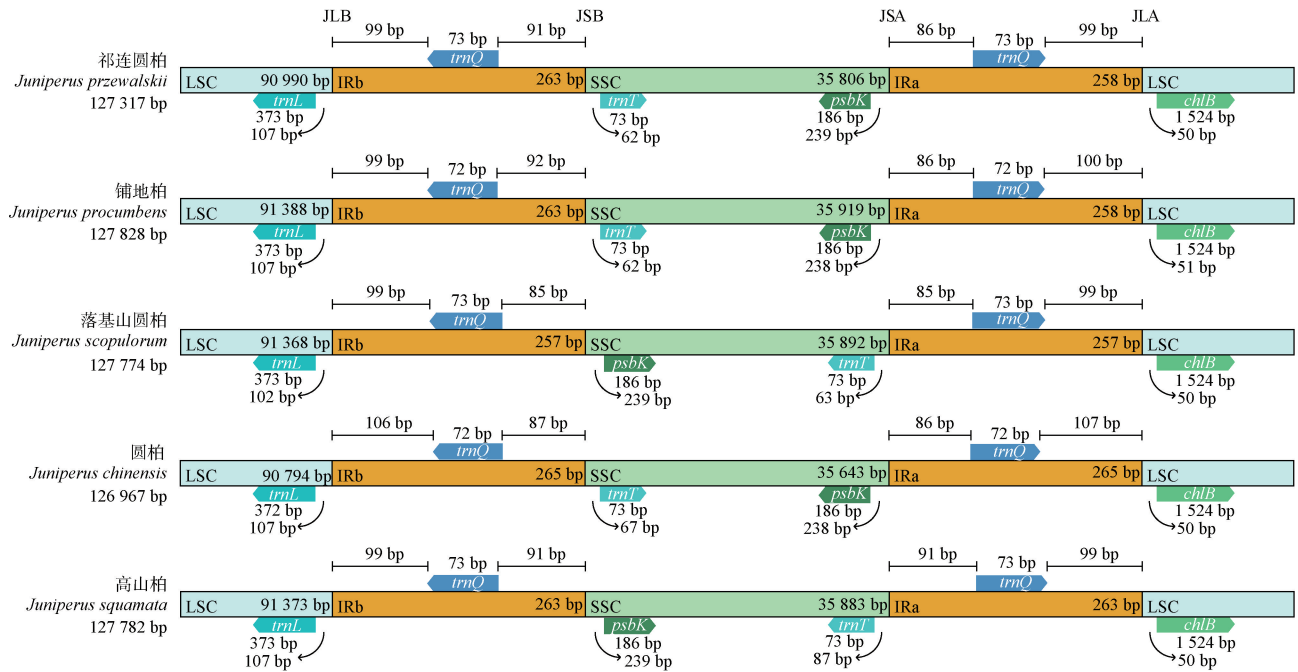
图2 柏科5个树种叶绿体基因组的序列比对结果
 Fig. 2 Sequence alignment results of chloroplast genomes of five tree species of Cupressaceae

2.4.2 叶绿体基因组 IR 区边界的差异分析 结果 (图3)显示:柏科5个树种叶绿体基因组 LSC 区长度在 90 794~91 388 bp 之间,SSC 区长度在 35 643~35 919 bp 之间,IR 区不完整,发生明显缩减,长度在 257~265 bp 之间。5个树种叶绿体基因组 IRa 区和 IRb 区内均仅保留 *trnQ* 基因,其与 LSC 区和 SSC 区的边界包括 *trnL*、*trnT*、*psbK* 和 *chlB* 基因。5个树种间叶绿体基因组 JSB 边界 (IRb 区与 SSC 区的边界) 和 JSA 边界 (SSC 区与 IRa 区的边界) 差异较大,其中,

祁连圆柏与铺地柏和圆柏的叶绿体基因组 IR 区边界较为相似。

2.5 柏科植物系统发育分析

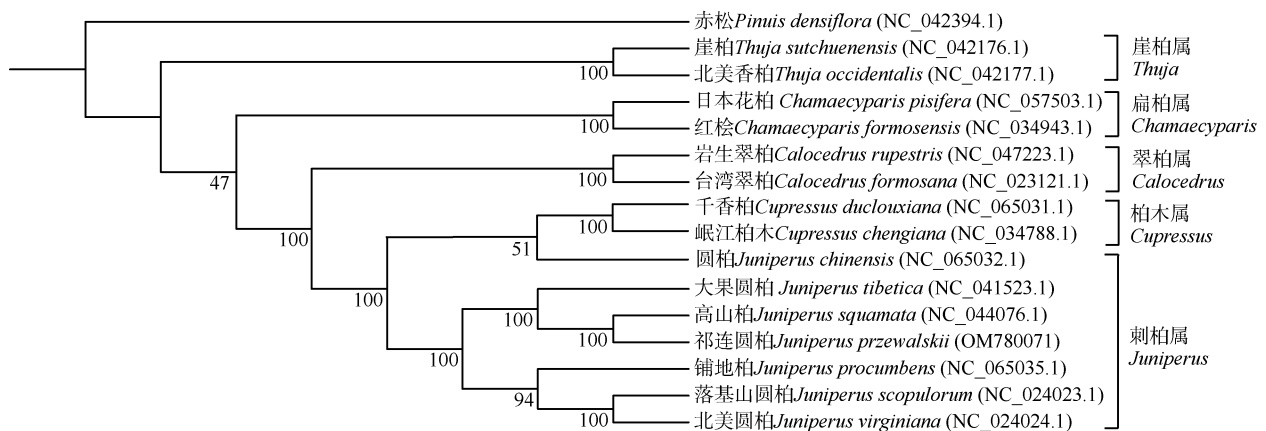
以赤松为外类群,基于 15 个树种的叶绿体基因组编码序列,构建系统发育树。结果 (图4)显示:总体上看,各属物种聚在同一支。除圆柏外,原属于圆柏属的高山柏、祁连圆柏、铺地柏和落基山圆柏与刺柏属的大果圆柏和北美圆柏聚为一支,其中与祁连圆柏亲缘关系最近的是高山柏。



JLB: LSC 区和 IRb 区的边界 Boundary of LSC and IRb regions; JSB: IRb 区和 SSC 区的边界 Boundary of IRb and SSC regions; JSA: SSC 区和 IRa 区的边界 Boundary of SSC and IRa regions; JLA: IRa 区和 LSC 区的边界 Boundary of IRa and LSC regions.

图 3 柏科 5 个树种叶绿体基因组大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 边界的比较

Fig. 3 Comparison on the boundaries of large single copy region (LSC), small single copy region (SSC) and inverted repeat region (IR) in chloroplast genomes among five tree species of Cupressaceae



进化树分支上的数据为自展支持率,括号内编号为 GenBank 登录号 The data on the branches of phylogenetic tree are bootstrap values, and the Nos. in brackets are the GenBank accession numbers.

图 4 基于叶绿体基因组编码序列的柏科植物的系统发育树

Fig. 4 Phylogenetic tree of Cupressaceae species based on coding sequences of chloroplast genomes

3 讨论和结论

3.1 祁连圆柏叶绿体基因组特征

利用下一代测序 (NGS) 技术获得完整叶绿体基

因组序列的方法简单、成本低^[27]。本研究利用 Oxford Nanopore 技术获得了祁连圆柏的完整叶绿体基因组序列。祁连圆柏叶绿体基因组全长 127 317 bp, 共编码 118 个基因, 包括 82 个蛋白质编码基因、32 个 tRNA 基因和 4 个 rRNA 基因。叶绿体基因组

表现为四分体结构,包括1个大单拷贝区(LSC)、1个小单拷贝区(SSC)和2个反向重复区(IR),这也是陆生植物叶绿体基因组结构的显著特征^[28-30]。但祁连圆柏叶绿体基因组的IR区发生缩减现象,长度仅为258和263 bp,使其基因组四分体结构不明显。而种子植物的叶绿体基因组长度为120~160 kb^[31],IR区的缩减可能是导致祁连圆柏叶绿体基因组偏小的主要原因。

3.2 祁连圆柏叶绿体基因组密码子偏好性

叶绿体基因组的密码子使用模式和核苷酸组成可为叶绿体基因组的基因工程提供理论基础^[26]。在祁连圆柏叶绿体基因组编码的氨基酸序列中,Leu数量最多(对应的密码子有2 574个),数量最少的氨基酸是Cys(对应的密码子仅有272个)。AAA(编码Lys)是使用最多(1 163)的密码子。除了Met和Trp外,其余氨基酸均有2种及以上的同义密码子,其中Leu、Arg和Ser的密码子数量最多,均有6种。此外,祁连圆柏叶绿体基因组的蛋白质编码基因中有30种密码子的相对同义密码子使用度(RSCU)大于1,且大多数密码子偏好以A或U结尾,不偏好以C或G结尾,这也是植物叶绿体基因组中的一种常见现象^[32-33]。

3.3 祁连圆柏叶绿体基因组重复序列特征

简单重复序列被广泛用作研究遗传多样性和群体结构的分子标记^[34-35]。在祁连圆柏中,共发现21类简单重复序列,主要以AA或TT双核苷酸重复序列为主。这表明双核苷酸重复序列在祁连圆柏叶绿体基因组中是最丰富的,且有碱基偏好性。本文首次报道祁连圆柏叶绿体基因组的简单重复序列,在基因组资源缺乏的情况下,叶绿体基因组简单重复序列可以作为鉴定遗传变异的标记,并且可为祁连圆柏及其近缘物种的进化研究提供新的思路。此外,祁连圆柏叶绿体基因组共38个长重复序列,长度均大于30 bp,且多数长重复序列长度为30~40 bp,其中正向重复(F)序列最丰富。长度大于等于30 bp的长重复序列在导致基因组变异的插入、缺失错配和重排方面发挥重要作用^[36-37]。

3.4 柏科不同树种叶绿体基因组的高变异区及反向重复区

mVISTA分析结果表明:原属于圆柏属的祁连圆柏、铺地柏、落基山圆柏、圆柏和高山柏5个树种的叶绿体基因组具有高度保守性,在基因组长度、基因内

容和排列上均没有太大差异。叶绿体基因组中编码序列比非编码序列更加保守,显著变异多集中在非编码序列^[38-39]。本研究从这5个树种叶绿体基因组中鉴定出3个存在显著变异的基因,分别是*accD*、*ycf1*和*ycf2*基因。这些高变异区可提供丰富的系统发育信息,用于界定密切相关的分类群^[40-41]。

松柏类叶绿体基因组的反向重复区(IR)在进化过程中发生了不同的丢失事件^[42]。虽然松柏类的叶绿体基因组不包含典型的IR区,但是该类植物仍然演化出了与叶绿体基因组重排相关的特异性IR区。例如:柏科刺柏属的4个树种^[43]均包含*tmq-IR*,*tmq-IR*能够介导36 kb区段的倒位。本研究结果也表明:与大多数陆生植物一样,原属于圆柏属的5个树种的叶绿体基因组为环状四分体结构,但5个树种的IR区并不完整,大小仅为257~265 bp,且叶绿体基因组的IRa区和IRb区内均仅保留*tmq*基因,反映出这些树种在进化过程中发生明显缩减,失去了大量的反向重复拷贝。

3.5 柏科不同树种系统发育分析

叶绿体系统发育基因组学已经成功地用于解决类群之间的系统发育关系^[44-46],并且使用完整的叶绿体基因组对系统发育分析具有更好的优势。刺柏属和圆柏属的分组一直存在争议^[12-14]。本研究通过使用完整的叶绿体基因组进行系统发育分析,也验证了这2个属之间具有密不可分的关系。例如:落基山圆柏与北美圆柏聚为一支,大果圆柏与高山柏和祁连圆柏聚为一支。此外,在本研究中,圆柏与柏木属物种聚为一支。推测产生这种结果的原因有2个方面:一方面,柏木属与刺柏属物种之间存在很大的相似性^[47];另一方面,“长分支吸引”效应对系统发育关系有影响^[48]。在柏木属和刺柏属植物叶绿体基因组的比较分析中,也提出将圆柏划入柏木属的观点^[49]。有关柏科内不同属物种的进一步划分,还需要随着测序技术的发展收集足够的的数据,以获得更加可靠的系统发育关系。

3.6 结论

祁连圆柏叶绿体基因组全长127 317 bp,总GC含量为35.1%;共包含23 708个密码子,偏好使用A或U;289个简单重复序列中以AA或TT双核苷酸重复序列最为丰富,38个长重复序列中以正向重复(F)序列最为丰富。原属于圆柏属的5个树种叶绿体基因组的长度、基因内容和排列具有高度相似性,且反

向重复区 (IR) 在进化过程中发生明显缩减现象。原属于圆柏属的物种与刺柏属物种亲缘关系较近,且2个属内的部分物种在系统发育树中的位置出现交错现象,其中,祁连圆柏与高山柏亲缘关系最近。

参考文献:

- [1] 中国科学院中国植物志编辑委员会. 中国植物志: 第七卷[M]. 北京: 科学出版社, 1978: 323.
- [2] 陆淑莲. 祁连圆柏的研究现状及展望[J]. 农家参谋, 2019(4): 119.
- [3] 金敏艳, 李进军, 车宗玺, 等. 祁连山中部祁连圆柏年内径向生长对气候因子的响应[J]. 生态学报, 2020, 40(21): 7699-7708.
- [4] 崔英, 王占林, 张得芳, 等. 土壤含水量和光照对祁连圆柏光合生理指标的影响[J]. 西部林业科学, 2019, 48(5): 95-100.
- [5] 张斌运. 祁连圆柏育苗及抗旱造林技术研究[J]. 种子科技, 2021, 39(22): 111-112.
- [6] 施征, 白登忠, 雷静品, 等. 祁连圆柏光合色素与非结构性碳水化合物含量对海拔变化的响应[J]. 西北植物学报, 2012, 32(11): 2286-2292.
- [7] 张静, 陈志林, 侯晓巍, 等. 三江源区祁连圆柏群落物种多样性沿海拔梯度的变化格局[J]. 西北植物学报, 2020, 40(10): 1759-1767.
- [8] 田晓萍, 马力, 占玉芳. 祁连山北坡中段祁连圆柏群落学特征的研究[J]. 西北林学院学报, 2015, 30(2): 77-83.
- [9] MOORE M J, SOLTIS P S, BELL C D, et al. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(10): 4623-4628.
- [10] MOWER J P, VICKREY T L. Structural diversity among plastid genomes of land plants[J]. Advances in Botanical Research, 2018, 85: 263-292.
- [11] WU C S, WANG Y N, HSU C Y, et al. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny[J]. Genome Biology and Evolution, 2011, 3: 1284-1295.
- [12] 田培林. 基于针叶结构的青海省6种刺柏属植物亲缘关系研究[D]. 杨凌: 西北农林科技大学, 2022: 5-6.
- [13] ADAMS R P. Systematics of multi-seeded eastern hemisphere *Juniperus* based on leaf essential oils and RAPD DNA fingerprinting[J]. Biochemical Systematics and Ecology, 1999, 27: 709-725.
- [14] 毛康珊. 广义柏科的生物地理学研究[D]. 兰州: 兰州大学, 2010: 72.
- [15] YERMAGAMBETOVA M, ABUGALIEVA S, TURUSPEKOV Y, et al. Illumina sequencing data of the complete chloroplast genome of rare species *Juniperus seravschanica* (Cupressaceae) from Kazakhstan[J]. Data in Brief, 2023, 46: 108866.
- [16] ALMEREKOVA S, YERMAGAMBETOVA M, ABUGALIEVA S, et al. The complete chloroplast genome sequencing data of *Juniperus sabina* L. (Cupressaceae Bartl.) from Kazakhstan[J]. Data in Brief, 2022, 45: 108644.
- [17] 宋艳波, 吴国良, 牛洪斌. 改良 CTAB 法在核桃叶片基因组 DNA 提取中的应用研究[J]. 山西农业大学学报(自然科学版), 2011, 31(2): 109-112.
- [18] KOLMOGOROV M, YUAN J, LIN Y, et al. Assembly of long, error-prone reads using repeat graphs[J]. Nature Biotechnology, 2019, 37(5): 540-546.
- [19] SHI L C, Chen H M, JIANG M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer[J]. Nucleic Acids Research, 2019, 47(W1): W65-W73.
- [20] 童一涵, 郑倩, 杜新明, 等. 多齿红山茶叶绿体基因组序列特征分析[J]. 植物资源与环境学报, 2022, 31(5): 27-36.
- [21] BEIER S, THIEL T, et al. MISA-web: a web server for microsatellite prediction[J]. Bioinformatics, 2017, 33(16): 2583-2585.
- [22] FRAZER K A, PACTHER L, POLIAKOV A, et al. VISTA: computational tools for comparative genomics[J]. Nucleic Acids Research, 2004, 32: W273-W279.
- [23] CHEN C, CHEN H, ZHANG Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. Molecular Plant, 2020, 13(8): 1194-1202.
- [24] KATO H K, STANDLEY D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. Molecular Biology and Evolution, 2013, 30(4): 772-780.
- [25] NGUYEN L-T, SCHMIDT H A, VON HAESELER A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. Molecular Biology and Evolution, 2015, 32(1): 268-274.
- [26] LETUNIC I, BORK P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation[J]. Nucleic Acids Research, 2021, 49: W293-W296.
- [27] HEATHER J M, CHAIN B. The sequence of sequencers: the history of sequencing DNA[J]. Genomics, 2016, 107(1): 1-8.
- [28] YIK M H-Y, KONG B L-H, SIU T Y, et al. Differentiation of *Hedyotis diffusa* and common adulterants based on chloroplast genome sequencing and DNA barcoding markers[J]. Plants, 2021, 10: 161.
- [29] ZHU B, FENG Q, YU J, et al. Chloroplast genome features of an important medicinal and edible plant: *Houttuynia cordata* (Saururaceae)[J]. PLoS ONE, 2020, 15(9): e0239823.
- [30] ALZAHIRANI D A, YARADUA S S, ALBOKHARI E J, et al. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae[J]. BMC Genomics, 2020, 21: 393.
- [31] 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 Molecular Biology, 2011, 76: 273-297.

- [32] ZHANG F J, WANG T, SHU X C, et al. Complete chloroplast genomes and comparative analyses of *L. chinensis*, *L. anhuiensis*, and *L. aurea* (Amaryllidaceae) [J]. International Journal of Molecular Sciences, 2020, 21: 5729.
- [33] WEN F, WU X Z, LI T J, et al. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China[J]. BMC Genomics, 2021, 22: 161.
- [34] SINGH R B, MAHENDERAKAR M D, JUGRAN A K, et al. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers[J]. Gene, 2020, 753: 144800.
- [35] RAMZAN M, SARWAR S, KAUSER N, et al. Assessment of inter simple sequence repeat (ISSR) and simple sequence repeat (SSR) markers to reveal genetic diversity among *Tamarix* ecotypes [J]. Journal of King Saud University, 2020, 32: 3437-3446.
- [36] WENG M L, BLAZIER J C, GOVINDU M, et al. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates[J]. Molecular Boilogy and Evolution, 2014, 31: 645-659.
- [37] ASAF S, KHAN A L, KHAN M A, et al. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species [J]. PLoS ONE, 2018, 13 (3): e0192966.
- [38] DONG F, LIN Z C, LIN J, et al. Chloroplast genome of rambutan and comparative analyses in Sapindaceae [J]. Plants, 2021, 10: 283.
- [39] WANG J H, MOORE M J, WANG H, et al. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies [J]. Gene, 2021, 765: 145103.
- [40] LI X W, YANG Y, HENRY R J, et al. Plant DNA barcoding: from gene to genome [J]. Biological Reviews, 2015, 90: 157-166.
- [41] ZENG C X, HOLLINGSWORTH P M, YANG J, et al. Genome skimming herbarium specimens for DNA barcoding and phylogenomics[J]. Plant Methods, 2018, 14: 43.
- [42] WU C S, CHAW S M. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers [J]. Plant Biotechnology Journal, 2014, 12: 344-353.
- [43] GUO W, GREWE F, COBO-CLARK A, et al. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution[J]. Genome Biology and Evolution, 2014, 6(3): 580-590.
- [44] RUHSAM M, RAI H S, MATHEWS S, et al. Does complete plastidgenome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? [J]. Molecular Ecology Resources, 2015, 15(5): 1067-1078.
- [45] NJUGUNA W, LISTON A, CRONN R, et al. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing [J]. Molecular Phylogenetics and Evolution, 2013, 66: 17-29.
- [46] MA P F, ZHANG Y X, ZENG C X, et al. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae) [J]. Systematic Biology, 2014, 63(6): 933-950.
- [47] ZHU A, FAN W, ADAMS R P, et al. Phylogenomic evidence for ancient recombination between plastid genomes of the *Cupressus-Juniperus-Xanthocyparis* complex (Cupressaceae) [J]. BMC Evolutionary Biology, 2018, 18: 137.
- [48] IVANOVA Z, SABLOK G, DASKALOVA E, et al. Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response [J]. Frontiers in Plant Science, 2017, 8: 204.
- [49] CHEN C, XIA X, PENG J Y, et al. Comparative analyses of six complete chloroplast genomes from the genus *Cupressus* and *Juniperus* (Cupressaceae) [J]. Gene, 2022, 837: 146696.

(责任编辑:郭严冬)

《植物资源与环境学报》启事

为了扩大科技期刊的信息交流、充分实现信息资源共享,《植物资源与环境学报》已先后加入“中国学术期刊(光盘版)”、“万方数据——数字化期刊群”和“中文科技期刊数据库”等网络文献资源数据库,凡在本刊发表的论文将编入数据库供上网交流、查阅及检索,作者的著作权使用费与本刊稿酬一次性给付,不再另付。如作者不同意将文章收编入数据库,请在来稿时声明,本刊将做适当处理。

《植物资源与环境学报》的投稿网址为 <http://zwzy.cnbg.net>; 联系电话: 025-84347014; E-mail: zwzybjb@163.com; QQ: 2219161478。

《植物资源与环境学报》编辑部
2023-05