

竹根七属 4 种药用植物叶绿体基因组特征比较分析

林 焰¹, 蔡丽滨², 林 蓉¹, 吴建国¹, 杨成梓^{1*}, 吴岩斌^{1*}

1. 福建中医药大学药学院, 福建 福州 350122

2. 漳浦县中医院, 福建 漳州 363200

摘要: 目的 解析散斑竹根七 *Disporopsis aspersa*、竹根七 *D. fuscopicta*、长叶竹根七 *D. longifolia*、深裂竹根七 *D. pernyi* 4 种竹根七属药用植物的叶绿体基因组特征与系统发育关系, 筛选属内潜在物种鉴别位点, 为该属植物的物种鉴定、种质资源保护、药材质量控制及开发利用提供分子基础与理论支撑。方法 以 4 种竹根七属药用植物的幼嫩叶片为材料进行高通量测序, 完成叶绿体基因组的组装与注释, 并运用生物信息学方法开展重复序列、IR 区边界、密码子使用偏好、核苷酸多态性、选择压力及系统发育分析。结果 4 种竹根七属药用植物的叶绿体基因组均呈典型的四分体环状结构, 基因组全长为 155 975~156 109 bp, 总 GC 含量均为 37.68%; 均注释得到 131 个基因, 包括 85 个蛋白编码基因、38 个 tRNA 基因和 8 个 rRNA 基因。4 个叶绿体基因组中的简单重复序列 (simple sequence repeats, SSR) 均以单核苷酸重复类型为主, 长重复序列 (long-sequence repeats, LSR) 均以回文重复为主。4 种植物的 IR 区边界整体较为保守, 密码子使用模式基本一致, 偏好使用以 A 或 T 结尾的密码子。核苷酸多态性分析表明, *rrn23S* 和 *trnT-UGU* 在基因区域中具有较高的核苷酸多态性值 (nucleotide diversity, P_i) 值, *rrn4.5S-rrn5S*、*ccsA-ndhD*、*rps11-rpl36* 和 *rps15-ycf1* 为变异程度较高的基因间区。选择压力分析显示, 大多数蛋白编码基因的 K_a/K_s 值小于 1, 主要受到净化选择约束; *ycf1*、*ycf4*、*psaI* 和 *cemA* 在部分物种组合中的 K_a/K_s 值大于 1, 可能存在与适应性进化相关的信号。系统发育分析显示, 4 种竹根七属植物聚集于同一分支, 且节点支持率较高, 并与黄精属构成姊妹群。结论 4 种竹根七属植物叶绿体基因组结构与基因组成总体保守, 但局部区域存在一定序列变异。筛选获得的候选高变区域可为竹根七属植物的物种鉴定和分子标记开发提供参考, 系统发育结果为该属植物的分类及亲缘关系研究提供了叶绿体基因组证据。

关键词: 竹根七属; 散斑竹根七; 竹根七; 长叶竹根七; 深裂竹根七; 叶绿体基因组; 种质资源; 种间差异; 选择压力; 系统发育分析

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2026)18-7419-12

DOI: 10.7501/j.issn.0253-2670.2026.18.028

Comparative analysis of chloroplast genome characteristics in four medicinal *disporopsis* species

LIN Yan¹, CAI Libin², LIN Rong¹, WU Jianguo¹, YANG Chengzi¹, WU Yanbin¹

1. School of Pharmacy, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China

2. Zhangpu Traditional Chinese Medicine Hospital, Zhangzhou 363200, China

Abstract: Objective To characterize the chloroplast genomes and phylogenetic relationships of four medicinal species of *Disporopsis*, namely *Disporopsis aspersa*, *D. fuscopicta*, *D. longifolia*, and *D. pernyi*, and to identify potential regions for species discrimination, thereby providing a molecular and theoretical basis for species identification, germplasm resource conservation, quality control of medicinal materials, and resource development and utilization of this genus. **Methods** Young leaves of the four medicinal *Disporopsis* species were subjected to high-throughput sequencing. The chloroplast genomes were assembled and annotated, followed by bioinformatic analyses of repeat sequences, inverted repeat (IR) boundaries, codon usage bias, nucleotide diversity, selection pressure, and phylogenetic relationships. **Results** The chloroplast genomes of the four medicinal *Disporopsis* species exhibited a

收稿日期: 2026-03-02

基金项目: 2023 年福建省本科高校教育教学研究项目 (FBJY20230030)

作者简介: 林 焰, 硕士研究生, 研究方向为中药资源与品质评价研究。E-mail: 1601442173@qq.com

*通信作者: 杨成梓, 教授, 研究方向为中药资源与品质评价。E-mail: tiebaojin@163.com

吴岩斌, 副研究员, 研究方向为中药药效物质及品质评价研究。E-mail: wxsq1@163.com

typical quadripartite circular structure, with genome sizes ranging from 155 975 to 156 109 bp and an overall GC content of 37.68%. A total of 131 genes were annotated in each genome, including 85 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. Simple sequence repeats (SSRs) were dominated by mononucleotide repeats, whereas long sequence repeats (LSRs) were dominated by palindromic repeats. The IR boundaries were generally conserved among the four species. Their codon usage patterns were highly similar, with a preference for codons ending in A or T. Nucleotide diversity analysis showed that *rrn23S* and *trnT-UGU* had relatively high Pi values among the genic regions, whereas *rrn4.5S-rrn5S*, *ccsA-ndhD*, *rps11-rpl36*, and *rps15-ycf1* were highly variable intergenic regions. Selection pressure analysis showed that most protein-coding genes had Ka/Ks values below 1 and were predominantly constrained by purifying selection. However, *ycf1*, *ycf4*, *psaI*, and *cemA* had Ka/Ks values above 1 in some species comparisons, suggesting possible signals associated with adaptive evolution. Phylogenetic analysis showed that the four *Disporopsis* species clustered within the same well-supported branch and formed a sister group with *Polygonatum*. **Conclusion** The chloroplast genomes of the four *Disporopsis* species are generally conserved in genome structure and gene composition, although local sequence variation is present. The candidate highly variable regions identified in this study may provide useful references for species identification and molecular marker development in *Disporopsis*. The phylogenetic results also provide chloroplast genomic evidence for investigating the classification and relationships of this genus.

Key words: *Disporopsis*; *Disporopsis aspersa* (Hua) Engl. ex K. Krause; *Disporopsis fuscipicta* Hance; *Disporopsis longifolia* Craib; *Disporopsis pernyi* (Hua) Diels; chloroplast genome; germplasm resources; interspecific variation; selection pressure; phylogenetic analysis

竹根七属 *Disporopsis* L. 隶属于天门冬科 (Asparagaceae), 包括散斑竹根七 *D. aspersa* (Hua) Engl. ex K. Krause、竹根七 *D. fuscipicta* Hance、长叶竹根七 *D. longifolia* Craib、深裂竹根七 *D. pernyi* (Hua) Diels 4 种, 我国均产, 主要分布于长江流域及南方诸省^[1-3]。该属植物多为林下阴生草本, 具有肉质根状茎, 具有清热解毒、散瘀止痛、祛风除湿等功效, 主治跌打损伤等病症, 在民间药用中具有一定应用基础, 具备良好的药用开发潜力^[4-6]。然而, 在资源采集、药材流通和民间使用过程中, 4 种竹根七属植物之间存在不同程度的名称混用、基原混杂甚至替代使用现象。这不仅增加了种质资源调查和药材鉴定的难度, 也可能进一步影响药材来源的准确性及其质量稳定性, 对后续的开发利用、质量评价与临床安全应用带来潜在风险。此外, 由于属内物种在营养器官和花部形态上较为接近, 不同种间常表现出较高的形态相似性, 而部分性状又易受环境、发育阶段及采收状态影响, 进一步增加了物种鉴别和药材基原界定的难度。

在中药资源研究由“经验识别”向“精准鉴定”转变的背景下, 建立可靠的分子鉴别体系已成为保障药材真实性和一致性的重要途径。叶绿体基因组

因具有结构相对保守、基因组成稳定、单倍体遗传和适中变异速率等特点^[7], 近年来已被广泛应用于植物物种鉴定、比较基因组学和系统发育研究。相较于传统 DNA 条形码片段, 完整叶绿体基因组能够提供更多变异位点和更高分辨率, 不仅有助于挖掘高变区和候选分子标记, 还能够有效解析近缘种之间的亲缘关系及基因组结构差异。因此, 基于叶绿体基因组开展竹根七属药用植物的比较分析, 对于澄清种间关系、筛选特异性鉴别位点以及促进药材准确识别具有重要意义^[8]。

基于此, 本研究选取 4 种竹根七属药用植物: 散斑竹根七、竹根七、长叶竹根七、深裂竹根七, 对其叶绿体基因组进行测序、组装、注释与比较分析, 系统评估其基因组结构特征、边界变异、重复序列分布、序列多样性及系统发育关系, 并进一步筛选可用于物种鉴别的高变区。研究结果可为竹根七属药用植物的准确鉴定、种质资源保护、质量控制及合理开发利用提供分子基础和理论依据。

1 材料

4 份测序植物样品收集来源于中国不同省市 (表 1), 每种各采集 1 个野生居群单株个体, 经福建中医药大学药学院杨成梓教授分别鉴定为散斑

表 1 样品收集信息

Table 1 Collection information of four *Disporopsis* samples

序号	种名	产地	经纬度	GenBank 登录号
1	散斑竹根七	江西省宜春市铜鼓县	28°30'23.00"N, 114°23'08.74"E	PZ785683
2	竹根七	福建省漳州市华安县	24°58'09.71"N, 117°29'36.24"E	PZ785684
3	长叶竹根七	云南省楚雄市大地基乡	24°59'20.99"N, 101°32'05.14"E	PZ785685
4	深裂竹根七	四川省绵阳市平武县	24°59'20.99"N, 101°32'05.14"E	PZ785686

竹根七 *D. aspersa* Engl. ex K. Krause、竹根七 *D. fuscopicta* Hance、长叶竹根七 *D. longifolia* Craib、深裂竹根七 *D. pernyi* (Hua) Diels。在福建中医药大学时珍园 (119°11'01"E, 26°04'33"N) 栽培稳定后, 采集约 5 g 生长健壮、无病虫害的新鲜幼嫩叶片, 经无菌水多次冲洗后, 用液氮迅速冷冻。

2 方法

2.1 DNA 提取与测序

本研究中实验样品的 DNA 提取与完整性检验、文库构建及聚类测序等委托康圣环球有限公司完成。DNA 提取采用改良的 CTAB 法 (十六烷基三甲基溴化铵) 提取散斑竹根七、竹根七、长叶竹根七、深裂竹根七植物的基因组总 DNA。采用 Qubit 3.0 精确定量 DNA 浓度, 琼脂糖凝胶电泳评估完整性及纯度, AATI 微流控电泳分析文库片段分布及插入片段大小, qRT-PCR 检测文库有效浓度。质检合格后, 按需定量混合以保证测序均衡性, 加载至测序芯片, 在 DNBSEQ-T7 型平台完成测序。

2.2 叶绿体基因组的组装与注释

使用 GetOrganelle v1.7.5^[9] 软件对 4 个竹根七属药用植物叶绿体基因组进行组装。提交至 CPGAVAS2^[10] 在线平台 (<http://47.96.249.172:16019/analyzer/home>) 对进行基因结构注释, 随后检查叶绿体基因组的注释信息并进行人工修正。最后借助 OGDRAW 在线平台^[11] (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 绘制叶绿体基因组图谱。

2.3 4 个叶绿体基因组的比较分析

利用 Misa 软件对竹根七属 4 种药用植物的质体基因组进行简单重复序列 (simple sequence repeats, SSR) 挖掘。参数设定如下: 重复次数中单核苷酸基元不低于 10 次, 二核苷酸不低于 5 次, 三核苷酸不低于 4 次, 而四至六核苷酸基元均以 3 次为最低阈值。在此基础上, 进一步对单核苷酸 (A/T)、二核苷酸 (如 AT/TA) 及三核苷酸等各类重复单元的分布情况进行分类汇总。采用 REPuter^[12] 对叶绿体基因组中的长重复序列 (long sequence repeats, LSRs) 进行分析, 识别正向重复 (forward repeats, F)、回文重复 (palindromic repeats, P)、反向重复 (reverse repeats, R) 和互补重复 (complementary repeats, C)。最小重复长度设置为 30 bp, Hamming distance 设置为 3, 统计不同物种

中各类长重复序列的数量及组成。

使用软件 CPStools^[13] 提取竹根七属质体基因组中的蛋白编码序列 (coding DNA sequence, CDS), 并对其密码子使用特征进行系统评估。统计各氨基酸的出现频次、使用频率以及相对同义密码子使用度 (relative synonymous codon usage, RSCU)。

利用在线工具 IRscope^[14], 对竹根七属 4 种药用植物叶绿体基因组的 IR/SC 衔接点进行了详细比对。鉴定大单拷贝区 (large single-copy region, LSC)、小单拷贝区 (small single-copy region, SSC) 与 2 个反向重复区 (inverted repeat a/inverted repeat b, IRa/IRb) 之间的边界特征, 评估各物种在演化过程中的边界扩张或收缩情况。

通过 CPStools^[13] 软件计算基因和基因间区的核苷酸多态性值 (P_i)。使用 Phylosuite v1.2.2^[15] 将提取物种共有蛋白编码基因序列, 利用 Calculator v3.0^[16] 计算各物种非同义替换位点替换次数 (K_a) 与同义替换位点替换次数 (K_s) 的比值 (K_a/K_s)。

2.4 系统发育树的构建

从 NCBI 公共数据库下载原广义百合科 (Liliaceae) 相关类群的叶绿体全基因组序列, 并选取薯蓣 *Dioscorea polystachya* 作为外类群^[17]。将下载的叶绿体基因组序列与本研究获得的 4 种竹根七属植物叶绿体基因组共同纳入分析, 利用 CPStools 提取各物种共有的蛋白编码基因 (protein-coding genes, PCGs)。采用 MAFFT 软件对串联序列进行多序列比对^[18], 随后使用 IQ-TREE 软件根据贝叶斯信息准则自动筛选最适核苷酸替换模型, 并在最佳模型下采用最大似然法 (maximum likelihood, ML) 构建系统发育树^[19]。通过 1 000 次自举重复检验评估各分支的支持率, 贝叶斯推断采用 MrBayes v3.2.7 软件^[20] 进行, 根据最适核苷酸替换模型设置分析参数, 运行 2 次独立的马尔可夫链蒙特卡罗分析, 每次包含 4 条链, 共运行 2 000 000 代, 每 1 000 代取样 1 次。舍弃前 25% 的样本作为老化样本, 其余样本用于构建 50% 多数一致树, 并计算各分支的贝叶斯后验概率。最终利用 ChiPlot 在线平台 (<https://www.chiplot.online/>)^[21] 对系统发育树进行可视化和编辑。

3 结果与分析

3.1 叶绿体基因组基本特征

竹根七属 4 种药用植物的叶绿体基因组均为双链环状 DNA, 呈典型的四分体结构 (图 1), 基因组全

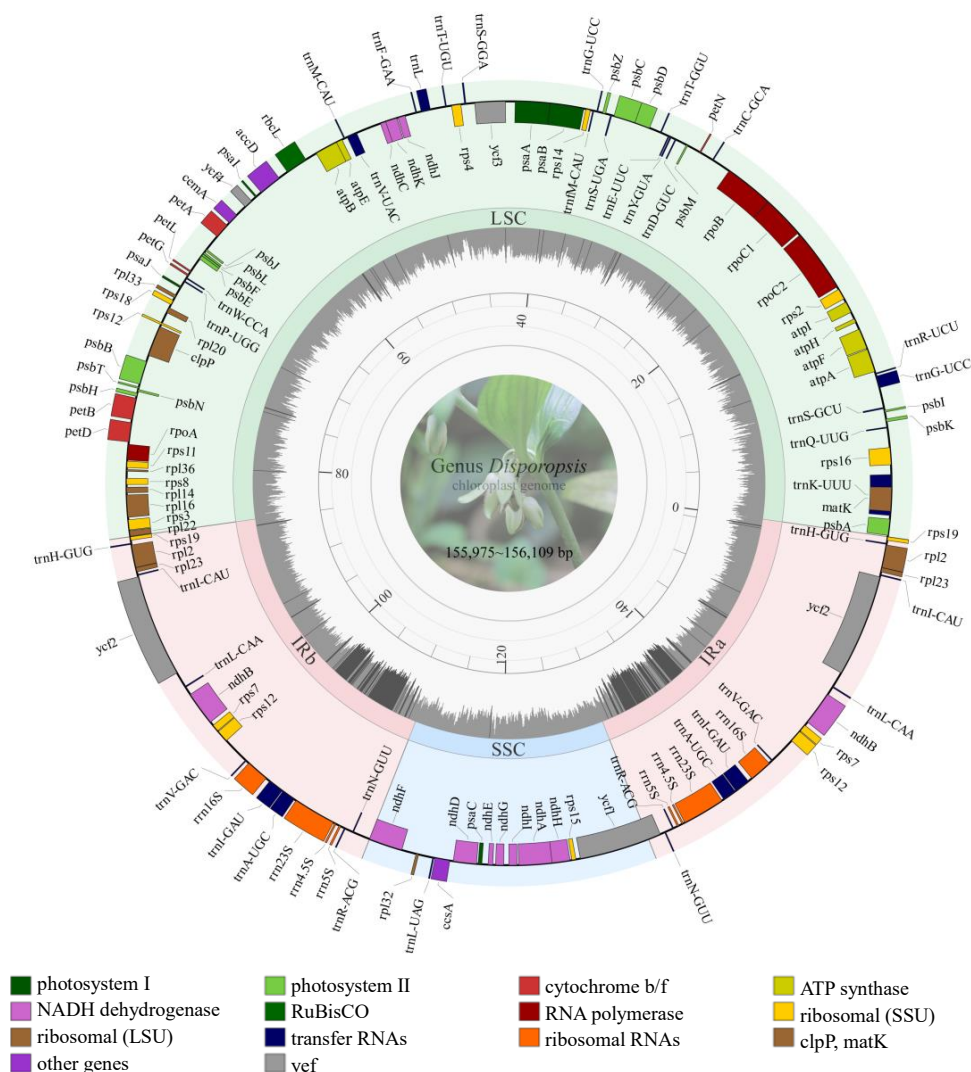


图 1 竹根七属 4 种药用植物叶绿体基因组图谱

Fig. 1 Chloroplast genome map of four medicinal *Disporopsis* species

长为 155 975~156 109 bp, 包含 1 个大单拷贝区 (LSC, 84 963~85 055 bp)、1 个小单拷贝区 (SSC, 18 513~18 528 bp) 以及 1 对反向重复区 (IRa/IRb, 各 26 242~26 265 bp) (表 2)。4 个叶绿体基因组的总 GC 含量均为 37.68%, 其中竹根七叶绿体基因组的 LSC、SSC、IRa 和 IRb 区 GC 含量分别为 35.68%、31.61%、43.07% 和 43.07%; 长叶竹根七分别为 35.69%、31.58%、

43.05% 和 43.05%; 深裂竹根七分别为 35.69%、31.63%、43.05% 和 43.05%; 散斑竹根七分别为 35.68%、31.63%、43.05% 和 43.05%。总体来看, 4 种植物各区域的 GC 含量差异较小, 均表现为 IR 区最高、LSC 区次之、SSC 区最低。结合各区域长度差异较小的特征, 表明竹根七属 4 种植物的叶绿体基因组在整体结构和碱基组成方面具有较高的保守性。

表 2 竹根七属 4 种药用植物叶绿体基因组结构

Table 2 Chloroplast genome structure of four medicinal *Disporopsis* species

种名	总长度/bp	LSC区长度/bp	IRa/IRb长度/bp	SSC区长度/bp	GC含量/%
散斑竹根七	156 067	85 055	26 244	18 524	37.68
竹根七	155 975	84 963	26 242	18 528	37.68
长叶竹根七	156 041	85 010	26 259	18 513	37.68
深裂竹根七	156 109	85 054	26 265	18 525	37.68

竹根七属 4 种药用植物叶绿体基因组的基因组成基本一致,均注释得到 131 个基因,其中包括 85 个蛋白编码基因、38 个 tRNA 基因和 8 个 rRNA 基因(表 3)。根据功能可将这些基因分为光合作用相关基因、自我复制相关基因和其他功能基因 3 类。其中,光合作用相关基因共 45 个,主要包括光系统 I、光系统 II、NADH 脱氢酶、细胞色素 b/f 复合体、Rubisco 大亚基和 ATP 合酶相关基因;自我复制相关基因共 76 个,包括核糖体大亚基蛋白基因、核糖体小亚基蛋白基因、RNA 聚合酶亚基基因、rRNA 基因和 tRNA 基因;其他功能基因共 10 个,包括

matK、*clpP*、*cemA*、*accD*、*ccsA* 以及 *ycf* 类基因。

基因结构分析表明,共有 18 个基因含有内含子。其中,*ndhB*、*ndhA*、*petB*、*petD*、*atpF*、*rpl16*、*rpl2*、*rps16*、*rpoC1*、*trnK-UUU*、*trnG-UCC*、*trnL*、*trnV-UAC*、*trnI-GAU* 和 *trnA-UGC* 15 个基因各含有 1 个内含子;*rps12*、*clpP* 和 *ycf3* 各含有 2 个内含子。此外,*ndhB*、*rpl2*、*rpl23*、*rps12*、*rps19*、*rps7*、4 个 rRNA 基因、部分 tRNA 基因及 *ycf2* 在 IR 区中以双拷贝形式存在。总体来看,4 种竹根七属植物叶绿体基因组的基因数量、功能组成及内含子分布高度一致,表明其基因组成和结构较为保守。

表 3 竹根七属 4 种药用植物叶绿体基因功能分类
Table 3 Functional classification of four medicinal *Disporopsis* species

基因类别	基因组	基因名称	数量
光合作用 基因	光系统I的亚基	<i>psaB</i> 、 <i>psaA</i> 、 <i>psaI</i> 、 <i>psaJ</i> 、 <i>psaC</i>	5
	光系统2的亚基	<i>psbA</i> 、 <i>psbK</i> 、 <i>psbI</i> 、 <i>psbM</i> 、 <i>psbD</i> 、 <i>psbC</i> 、 <i>psbZ</i> 、 <i>psbJ</i> 、 <i>psbL</i> 、 <i>psbF</i> 、 <i>psbE</i> 、 <i>psbB</i> 、 <i>psbT</i> 、 <i>psbN</i> 、 <i>psbH</i>	15
自我复制	NADH脱氢酶亚基	<i>ndhJ</i> 、 <i>ndhK</i> 、 <i>ndhC</i> 、 <i>ndhB(2)</i> [#] 、 <i>ndhF</i> 、 <i>ndhD</i> 、 <i>ndhE</i> 、 <i>ndhG</i> 、 <i>ndhI</i> 、 <i>ndhA</i> [#] 、 <i>ndhH</i>	13
	细胞色素b/f复合体亚基	<i>petN</i> 、 <i>petA</i> 、 <i>petL</i> 、 <i>petG</i> 、 <i>petB</i> [#] 、 <i>petD</i> [#]	6
	Rubisco大亚基	<i>rbcL</i>	1
	ATP合成酶亚基	<i>atpA</i> 、 <i>atpF</i> [#] 、 <i>atpH</i> 、 <i>atpI</i> 、 <i>atpE</i> 、 <i>atpB</i>	6
	大核糖体亚基蛋白	<i>rpl33</i> 、 <i>rpl20</i> 、 <i>rpl36</i> 、 <i>rpl14</i> 、 <i>rpl16</i> [#] 、 <i>rpl22</i> 、 <i>rpl2(2)</i> [#] 、 <i>rpl23(2)</i> 、 <i>rpl32</i>	11
	小核糖体亚基蛋白	<i>rps12(2)</i> ^{##} 、 <i>rps16</i> [#] 、 <i>rps2</i> 、 <i>rps14</i> 、 <i>rps4</i> 、 <i>rps18</i> 、 <i>rps11</i> 、 <i>rps8</i> 、 <i>rps3</i> 、 <i>rps19(2)</i> 、 <i>rps7(2)</i> 、 <i>rps15</i>	15
	RNA聚合酶亚基	<i>rpoC2</i> 、 <i>rpoC1</i> [#] 、 <i>rpoB</i> 、 <i>rpoA</i>	4
其他基因	核糖体RNA	<i>rrn16S(2)</i> 、 <i>rrn23S(2)</i> 、 <i>rrn4.5S(2)</i> 、 <i>rrn5S(2)</i>	8
	转运RNA	<i>trnK-UUU</i> [#] 、 <i>trnQ-UUG</i> 、 <i>trnS-GCU</i> 、 <i>trnG-UCC(2)</i> [#] 、 <i>trnR-UCU</i> 、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnY-GUA</i> 、 <i>trnE-UUC</i> 、 <i>trnT-GGU</i> 、 <i>trnS-UGA</i> 、 <i>trnM-CAU</i> 、 <i>trnS-GGA</i> 、 <i>trnT-UGU</i> 、 <i>trnL</i> [#] 、 <i>trnF-GAA</i> 、 <i>trnV-UAC</i> [#] 、 <i>trnM-CAU</i> 、 <i>trnW-CCA</i> 、 <i>trnP-UGG</i> 、 <i>trnH-GUG(2)</i> 、 <i>trnI-CAU(2)</i> 、 <i>trnL-CAA(2)</i> 、 <i>trnV-GAC(2)</i> 、 <i>trnI-GAU(2)</i> [#] 、 <i>trnA-UGC(2)</i> [#] 、 <i>trnR-ACG(2)</i> 、 <i>trnN-GUU(2)</i> 、 <i>trnL-UAG</i>	38
	成熟酶	<i>matK</i>	1
其他基因	蛋白酶	<i>clpP</i> ^{##}	1
	包膜蛋白	<i>cemA</i>	1
	乙酰辅酶A羧化酶	<i>accD</i>	1
	细胞色素C合成酶	<i>ccsA</i>	1
	保守开放阅读框	<i>ycf3</i> ^{##} 、 <i>ycf4</i> 、 <i>ycf2(2)</i> 、 <i>ycf1</i>	5

[#]含有 1 个内含子; ^{##}含有 2 个内含子。

[#]contains one intron; ^{##} contains two intron.

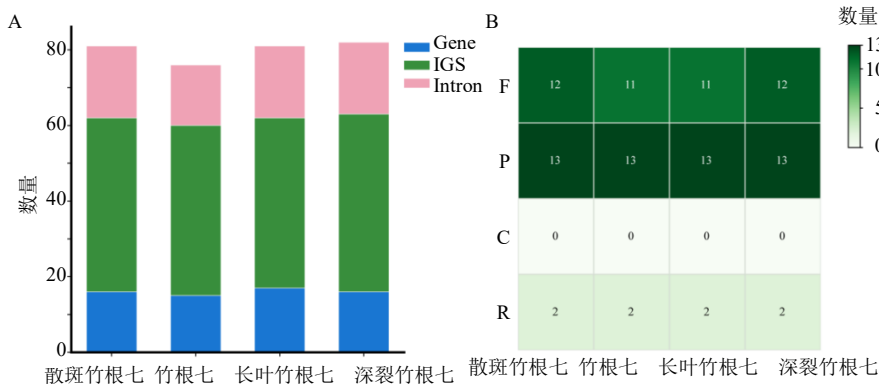
3.2 重复序列统计分析

已有研究报道^[22],被子植物叶绿体基因组中大量的单核苷酸重复序列有助于遗传变异。在竹根七属 4 种药用植物基因组中共检测到存在 320 个 SSR

位点,其中竹根七 76 个,长叶竹根七 81 个、深裂竹根七 82 个和散斑竹根七 81 个(表 4)。其中单核苷酸重复序列数量最多,在竹根七、长叶竹根七、深裂竹根七、散斑竹根七中占比分别为 55.26%、

表 4 竹根七属 4 种药用植物叶绿体基因组 SSR 分析
Table 4 SSR analysis of chloroplast genomes of four medicinal *Disporopsis* species

植物	单核苷酸	二核苷酸	三核苷酸	四核苷酸	五核苷酸	六核苷酸	总计
竹根七	42	16	6	9	2	1	76
长叶竹根七	50	14	4	10	2	1	81
深裂竹根七	52	15	4	9	2	0	82
散斑竹根七	51	15	4	9	2	0	81



A-SSR 在基因区 (Gene)、基因间区 (IGS) 和内含子区 (Intron) 中的数量分布; B-不同类型 LSR 的数量分布: F-正向重复; P-回文重复; C-互补重复; R-反向重复。

A-distribution of SSRs in genic regions (Gene), intergenic spacer regions (IGS), and intronic regions (Intron); B-number of different types of LSRs. F-forward repeats; P-palindromic repeats; C-complementary repeats; R-reverse repeats.

图 2 竹根七属 4 种药用植物叶绿体基因组中 SSR 和 LSR 的分布与数量

Fig. 2 Distribution and number of SSRs and LSRs in chloroplast genomes of four medicinal *Disporopsis* species

各物种分别含有 26~27 个。其中,竹根七和长叶竹根七均含有 26 个,深裂竹根七和散斑竹根七均含有 27 个。按照重复类型划分,共检测到 52 个回文重复 (palindromic repeats, P)、46 个正向重复 (forward repeats, F) 和 8 个反向重复 (reverse repeats, R),未检测到互补重复 (complementary repeats, C)。4 种植物均以回文重复数量最多,各含 13 个;正向重复分别为 11~12 个;反向重复均为 2 个 (图 2)。总体而言,4 种竹根七属植物的长重复序列数量和类型组成较为一致,进一步反映出其叶绿体基因组结构具有较高的保守性。

3.3 密码子偏好性分析

基于相对同义密码子使用度 (relative synonymous codon usage, RSCU) 分析竹根七属 4 种药用植物叶绿体基因组的密码子使用偏好性 (图 3)。结果显示,4 种植物在密码子偏好性上展现出高度的一致性,在 RSCU 值大于 1 的 31 个高频密码子中,以 A/T 结尾的类型高达 29 种 (其中 A 结尾 13 个, T 结尾 16 个),而 G 结尾仅 1 种 (GGA)。具体而言,TTA (Leu, RSCU 1.93~1.98)、AGA (Arg,

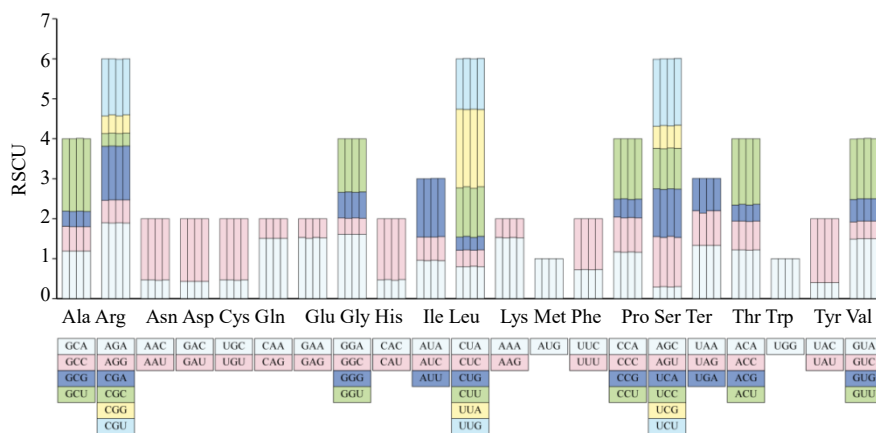
61.73%、63.41% 和 62.96%, 其次是二核苷酸,六核苷酸重复最少。从分布区域来看,经统计竹根七、长叶竹根七和散斑竹根七 SSR 主要位于基因组基因间隔区,其次是内含子区域,基因区域内分布的 SSR 数量最低 (图 2)。

此外,在 4 种植物叶绿体基因组中共鉴定到 106 个长重复序列 (long-sequence repeats, LSR),

RSCU 1.89~1.90)、GCT (Ala, RSCU 1.81~1.82)、TCT (Ser, RSCU 1.67~1.68)、ACT (Thr, RSCU 1.64~1.66)、GGA (Gly, RSCU 1.61) 以及 TAT (Tyr, RSCU 1.60) 等,以 A/T 结尾的密码子为各物种共有的最优密码子。反之,在低偏好性密码子 (RSCU < 1) 中,末位碱基则以 C 和 G 为主,如 CTG、CGC、AGC、GCG 等。这充分证明了该属植物叶绿体基因组强烈偏好使用以 A 或 T (U) 结尾的密码子,呈现出明显的 AT 偏倚特征^[23]。

3.4 IR 边界区分析

竹根七属 4 种药用植物叶绿体基因组的 LSC/IRb (JLB)、IRb/SSC (JSB)、SSC/IRa (JSA) 和 IRa/LSC (JLA) 边界及其邻近基因分布 (图 4)。4 种植物叶绿体基因组总长度为 155 975~156 109 bp,各区域的边界位置和邻近基因组成总体一致,表明其 IR 区边界结构较为保守。在 JLB 边界处, rpl22 位于 LSC 区,其中竹根七和长叶竹根七的 rpl22 距边界均为 33 bp,深裂竹根七和散斑竹根七均为 19 bp; rps19 位于 IRb 区,在竹根七和长叶竹根七中距边界均为 9 bp,在深裂竹根七和散斑竹根



自左至右分别代表散斑竹根七、竹根七、长叶竹根七和深裂竹根七。
From left to right: *D. aspera*, *D. fuscopicta*, *D. longifolia*, *D. pernyi*.

图 3 竹根七属 4 种药用植物叶绿体基因组相对同义密码子使用度

Fig. 3 Relative synonymous codon usage (RSCU) of chloroplast genomes in four medicinal *Disporopsis* species

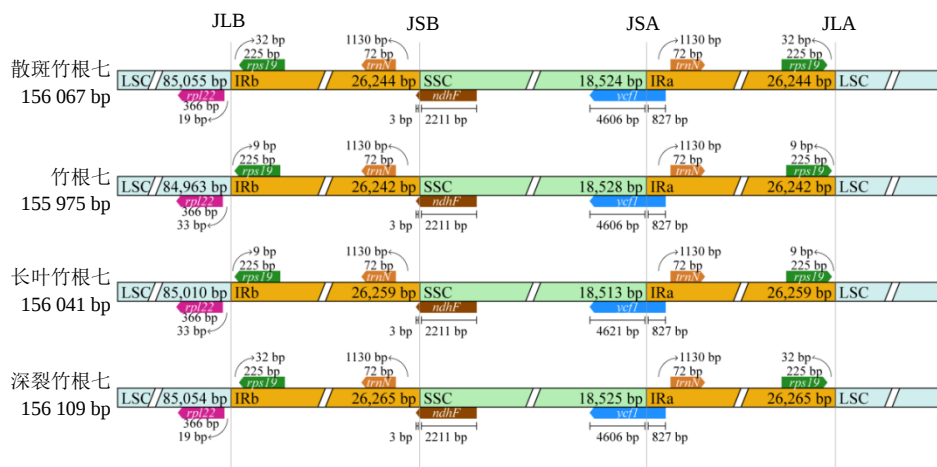


图 4 竹根七属 4 种药用植物叶绿体全基因组 IR 边界的比较示意图

Fig. 4 Comparison of IR boundaries in complete chloroplast genomes of four medicinal *Disporopsis* species

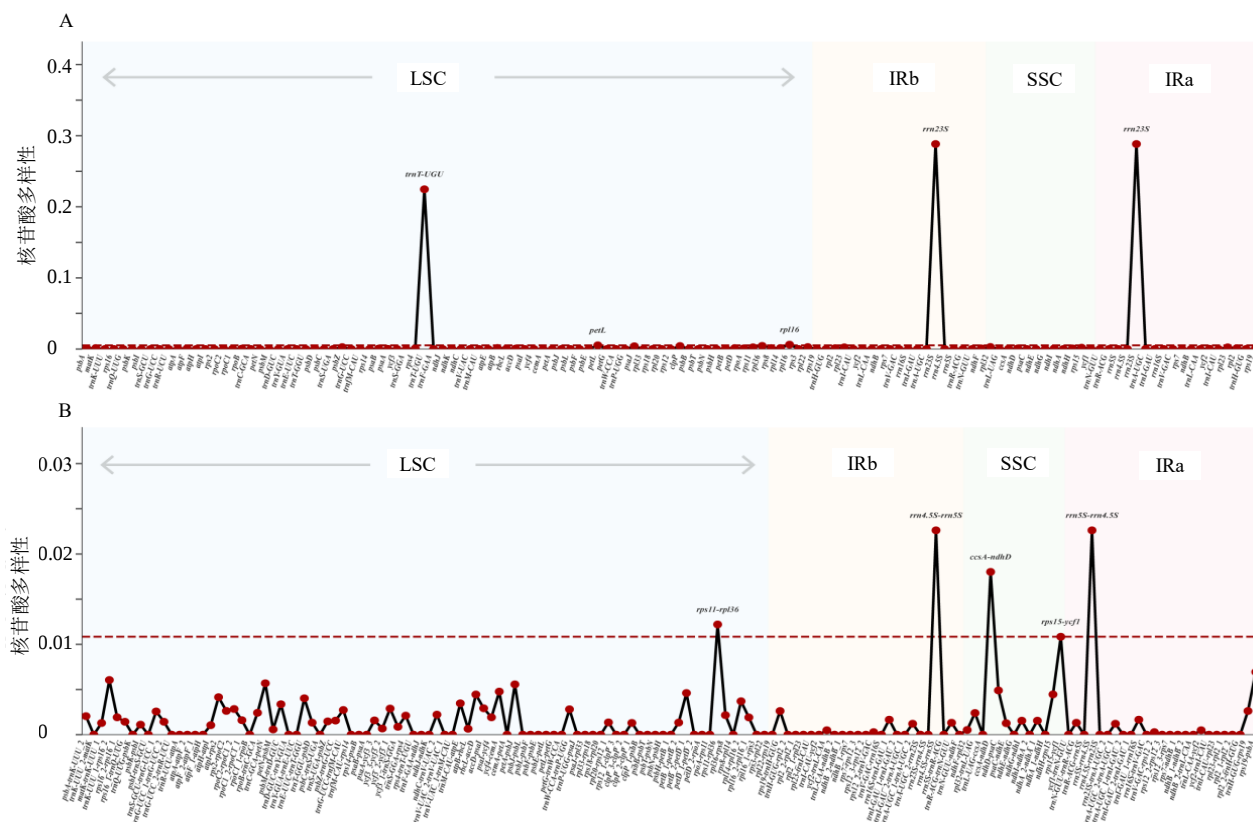
七中均为 32 bp。在 JSB 边界处, 4 种植物的 *ndhF* 均跨越 IRb/SSC 边界, 其中 3 bp 位于 IRb 区, 2211 bp 位于 SSC 区。JSA 边界均位于 *ycf1* 基因内部, 4 种植物的 *ycf1* 均有 827 bp 位于 IRa 区; 其位于 SSC 区的部分在竹根七、深裂竹根七和散斑竹根七中均为 4 606 bp, 在长叶竹根七中为 4 621 bp。在 JLA 边界处, *rps19* 均位于 IRa 区, 其中竹根七和长叶竹根七的 *rps19* 距边界均为 9 bp, 深裂竹根七和散斑竹根七均为 32 bp。总体而言, 4 种竹根七属植物仅在少数边界基因与连接点之间的距离上存在轻微差异, 未观察到明显的 IR 区扩张或收缩现象。

3.5 P_i 分析

为进一步分析 4 种竹根七属药用植物叶绿体基

因组的序列变异特征, 对共有基因和基因间区分别进行 P_i 分析。共获得 110 个共有基因和 126 个共有基因间区(图 5)。共有基因的 P_i 值为 0~0.288 32, 平均值为 0.005 24。其中, *rrn23S* 和 *trnT-UGU* 的 P_i 值分别为 0.288 32 和 0.224 64, 明显高于其他基因; 其余基因的 P_i 值均低于 0.01, 其中 *rpl16* 和 *petL* 相对较高, 分别为 0.006 13 和 0.005 21。

共有基因间区的 P_i 值为 0~0.022 62, 平均值为 0.001 54。以基因间区 P_i 值的平均值加 2 倍标准差作为筛选参考, 所得临界值约为 0.007 80, 因此将 $P_i > 0.008$ 的区域作为候选高变区域^[24-25], 共筛选出 *rrn4.5S-rrn5S*、*ccsA-ndhD*、*rps11-rpl36* 和 *rps15-ycf1* 4 个基因间区, 其 P_i 值分别为 0.022 62、



A-基因区的核苷酸多样性；B-基因间区的核苷酸多样性。

A-nucleotide diversity of genic regions; B-nucleotide diversity of intergenic regions.

图 5 竹根七属 4 种药用植物基因组 P_i 折线图

Fig. 5 Line plots of nucleotide diversity (P_i) in chloroplast genomes of four medicinal *Disporopsis* species

0.018 02、0.012 20 和 0.010 84。上述区域表现出较高的种间序列变异，可作为竹根七属植物分子鉴别和候选标记开发的潜在区域。

3.6 选择压力分析

为分析 4 种竹根七属药用植物叶绿体蛋白编码基因的进化特征，对共有蛋白编码基因进行两两比较，并计算 K_a/K_s 值（图 6）。结果显示，多数基因在不同物种组合中的 K_a/K_s 值为 0 或小于 1，表明竹根七属叶绿体蛋白编码基因总体受到净化选择约束，功能相对保守。其中，*rbcL*、*rpoC1*、*rpoB*、*rpl16*、*ndhD* 和 *ycf2* 等基因在部分物种组合中的 K_a/K_s 值介于 0~1，亦表现出净化选择特征。

少数基因在特定物种组合中的 K_a/K_s 值大于 1。*ycf1* 在深裂竹根七与长叶竹根七、长叶竹根七与竹根七，以及长叶竹根七与散斑竹根七组合中的 K_a/K_s 值分别为 1.562、2.344 和 1.562；*ycf4* 在长叶竹根七与竹根七组合中的 K_a/K_s 值为 1.543；*psaI* 在深裂竹根七与竹根七组合中的 K_a/K_s 值为 1.028；*cemA* 在散

斑竹根七与竹根七组合中的 K_a/K_s 值为 1.028。这些结果表明，上述基因在个别物种组合中可能具有较高的非同义替换积累，呈现潜在的选择信号。需要注意的是，4 个物种亲缘关系较近，部分基因的同义替换数量可能较少，个别较高的 K_a/K_s 值可能受到低 K_s 值的影响。总体而言，净化选择是竹根七属叶绿体蛋白编码基因进化的主要模式。其中，*ycf1*、*ycf4*、*psaI* 和 *cemA* 在部分物种组合中的 K_a/K_s 值大于 1，提示这些基因可能存在与适应性进化相关的信号。

3.7 系统发育分析

传统分类体系中，黄精属等相关类群曾被归入广义百合科，现代分类系统则将其归入天门冬科^[26]。因此，本研究选取 NCBI 数据库中已收录完整叶绿体基因组的原广义百合科代表类群进行系统发育分析。以薯蓣为外类群，基于叶绿体全基因组序列分别采用 ML 和贝叶斯推断法（bayesian inference, BI）构建系统发育树（图 7）。结果显示，ML 树与 BI 树的拓扑结构基本一致，主要分支均获得较高的支

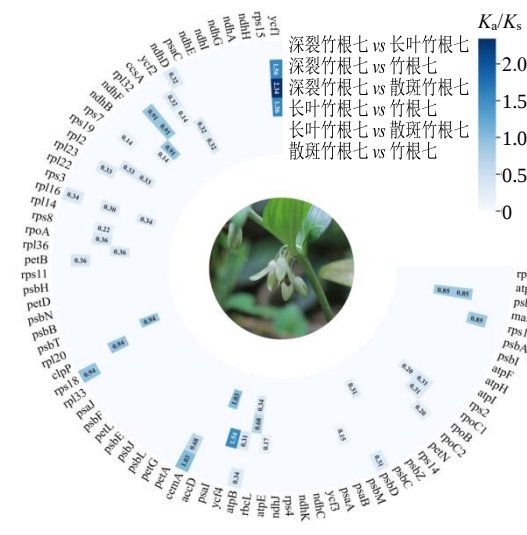
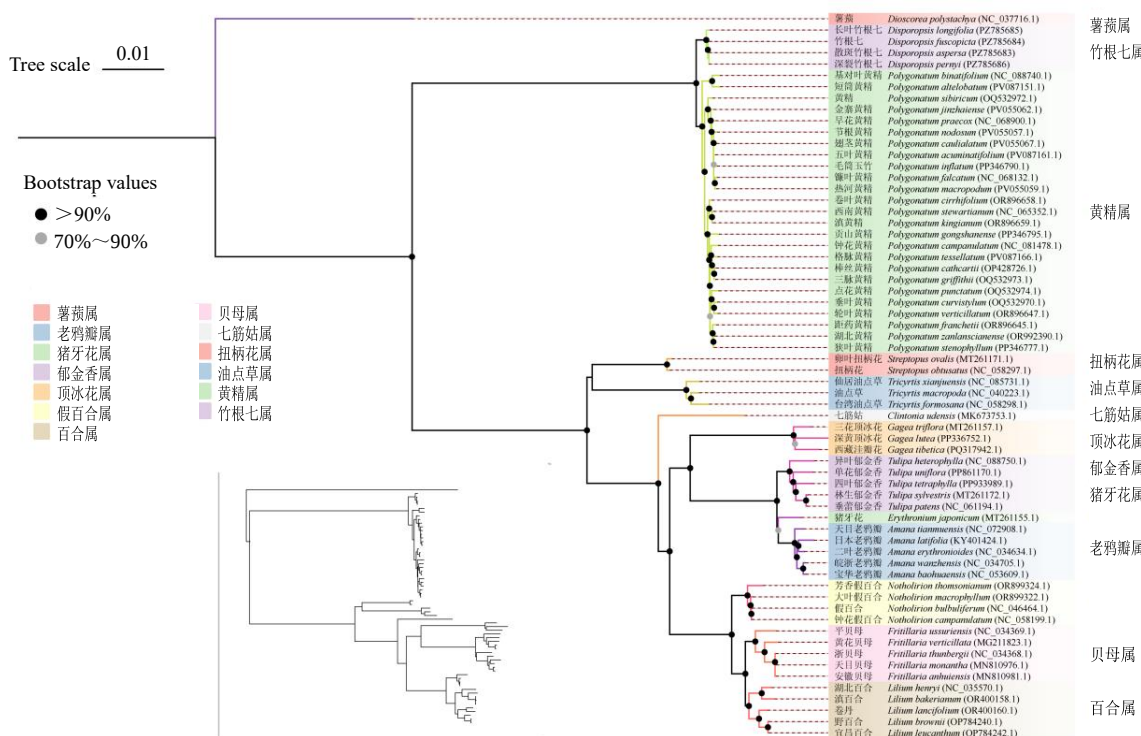


图 6 竹根七属 4 种药用植物内部分基因的 K_a/K_s 值

Fig. 6 K_a/K_s values of partial genes in four medicinal *Disporopsis* species



左下角为贝叶斯推断 (BI) 系统发育树, 其拓扑结构与 ML 树基本一致。

Bayesian inference (BI) tree is shown in the lower-left inset, and its topology is largely consistent with that of the ML tree.

图7 基于叶绿体全基因组序列构建的广义百合科 ML 系统发育树

Fig. 7 ML phylogenetic tree of *Liliaceae sensu lato* inferred from complete chloroplast genome sequences

持率。竹根七、长叶竹根七、深裂竹根七和散斑竹根七首先聚集于同一分支, 且该分支具有较高的自举支持率 (Bootstrap=98), 表明这 4 个物种具有较近的亲缘关系和较稳定的属级聚类模式。竹根七属分支与黄精属 *Polygonatum* 聚为姊妹群, 并与其他属明

显分离, 说明竹根七属与黄精属在系统发育上关系较近。

从枝长来看,竹根七属4个物种之间的分支长度总体较短,种间差异相对较小,表明其叶绿体基因组较为保守,序列分化程度有限,这也与同属近

缘物种之间通常具有较低遗传分歧的特征一致。整体拓扑结构清晰,主要节点均获得了较高支持率,说明基于叶绿体基因组数据构建的系统发育关系具有较高可靠性。此外,系统发育树还显示,竹根七属与黄精属共同构成一个相对独立的分支,与油点草属 *Tricyrtis*、七筋姑属 *Clintonia*、顶冰花属 *Gagea*、猪牙花属 *Erythronium*、老鸦瓣属 *Amana*、郁金香属 *Tulipa*、假百合属 *Notholirion*、百合属 *Lilium* 和贝母属 *Fritillaria* 等类群区分明显,进一步支持了竹根七属在该系统中的独立分类地位。

4 讨论

本研究对竹根七属 4 种药用植物的叶绿体基因组进行了比较分析。4 个叶绿体基因组均呈典型的四分体环状结构,基因组全长为 155 975~156 109 bp, LSC、SSC 和 IR 区长度差异较小,总 GC 含量均为 37.68%。不同区域的 GC 含量表现出一致规律,其中 IR 区最高, LSC 区次之, SSC 区最低。4 个物种的基因数量、功能分类和内含子组成也基本一致,说明竹根七属叶绿体基因组在整体结构和基因组成方面具有较高的保守性。叶绿体基因组承担光合作用、能量转换以及部分遗传信息表达功能,其中多数光合作用相关基因和核糖体蛋白基因受到较强的功能约束,这可能是近缘物种间叶绿体基因组结构和基因组成相对稳定的重要原因^[27]。

密码子使用偏好性分析显示,4 种竹根七属植物的相对同义密码子使用度分布较为一致。在 31 个 RSCU 值大于 1 的密码子中,多数以 A 或 T 结尾,表明其叶绿体蛋白编码基因具有较明显的 A/T 末端偏好。这一特征与 4 个叶绿体基因组较低的整体 GC 含量相吻合,也与多数高等植物叶绿体基因组的密码子使用特征相似^[28]。重复序列分析表明,4 种植物叶绿体基因组中的 SSR 均以单核苷酸重复类型为主,长重复序列则以回文重复和正向重复为主。不同物种之间的重复序列数量和类型组成总体接近,但其在编码区、内含子和基因间隔区中的分布存在一定差异。重复序列可能是叶绿体基因组局部序列变异的重要来源,相关区域可为后续种间差异分析和分子标记筛选提供基础数据。IR 边界比较结果显示,4 种竹根七属植物的 JLB、JSB、JSA 和 JLA 边界位置总体一致,仅在部分边界基因与连接点之间的距离上存在小幅差异。具体来说,4 个物种的 *ndhF* 基因在 JSB 边界处均向 IRb 区延伸 3 bp, JSA 边界均位于 *ycf1* 基因内部,表明属内物种的 IR 区

扩张和收缩程度较小。IR 区边界的相对稳定与 4 个叶绿体基因组长度接近的结果一致,进一步说明竹根七属叶绿体基因组整体结构较为保守。

尽管 4 个叶绿体基因组的整体结构较为一致,核苷酸多样性分析仍检测到若干变异相对较高的区域。在共有基因中, *rrn23S* 和 *trnT-UGU* 的 P_i 值明显高于其他基因;在基因间区中, *rrn4.5S-rrn5S*、*ccsA-ndhD*、*rps11-rpl36* 和 *rps15-ycf1* 的 P_i 值相对较高。这表明竹根七属叶绿体基因组具有“整体保守、局部变异”的特征。相比核心编码区域,部分基因间隔区受到的功能约束相对较弱,因而可能积累更多碱基替换和插入缺失变异。上述区域可作为竹根七属物种鉴别和分子标记开发的候选序列。 K_a/K_s 分析显示,大多数叶绿体蛋白编码基因在不同物种组合中的 K_a/K_s 值为 0 或小于 1,说明净化选择是竹根七属叶绿体蛋白编码基因进化的主要模式,相关基因的编码功能总体较为保守。少数基因在特定物种组合中的 K_a/K_s 值大于 1,其中 *ycf1*、*ycf4*、*psaI* 和 *cemA* 在部分比较组合中表现出较高的非同义替换比例,提示这些基因可能存在与适应性进化相关的信号。

与黄精属等天门冬科近缘类群已报道的叶绿体基因组变异热点相比,本研究筛选出的候选高变区域并未完全重合,仅部分区域在基因组位置上相邻或具有一定对应关系^[29-30]。这说明叶绿体基因组的变异热点可能受到研究类群、取样范围及筛选标准等因素影响,不同属之间不宜直接套用相同的鉴别位点。因此,本研究获得的高变区域现阶段更适合作为竹根七属内物种鉴别和分子标记开发的候选序列,而不能直接认定为竹根七属特异性标记。竹根七属不同物种在形态上具有一定相似性,药材经过干燥、切制等加工后,部分外部鉴别特征可能进一步减弱,从而增加基原鉴别难度^[31]。叶绿体基因组具有结构相对保守、拷贝数较高和易于扩增等特点,本研究筛选的候选高变区域可为传统形态鉴定和通用 DNA 条形码提供补充。然而,本研究尚未纳入不同地理居群、黄精属等近缘类群以及常见混伪品,也未开展引物设计和 PCR 验证。后续应扩大物种和居群取样范围,比较候选区域的种内稳定性与种间差异,并结合特异性引物设计、PCR 扩增和测序验证,评价其在药材基原鉴定和混伪品识别中的实际应用价值。

系统发育分析表明,竹根七、长叶竹根七、深

裂竹根七和散斑竹根七聚集在同一分支,且节点支持率较高,说明 4 个物种之间具有较近的亲缘关系。竹根七属与黄精属聚为姊妹群,并与所纳入的其他原广义百合科类群明显分离,表明竹根七属与黄精属的亲缘关系相对较近。该结果与现代分类系统中将竹根七属和黄精属归入天门冬科的处理基本一致,为二者亲缘关系的研究提供了叶绿体基因组证据。竹根七属 4 个物种之间的分支长度总体较短,表明其叶绿体基因组序列分化程度相对较低,这与 4 个物种在基因组结构、基因组成及 GC 含量等方面较为一致的结果相吻合。尽管 4 个物种的叶绿体基因组整体较为保守,基于叶绿体全基因组序列构建的系统发育树仍能对其进行区分,说明叶绿体基因组数据可为竹根七属近缘物种的亲缘关系分析和分类鉴定提供参考。后续仍需增加不同地理居群和个体,并结合近缘种及常见混伪品,对候选高变区域的种内稳定性和种间鉴别能力进行进一步验证。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Flora of China Editorial Committee. *Flora of China* [M]. Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 2000: 232-234.
- [2] 中国科学院中国植物志编辑委员会. 中国植物志 [M]. 北京: 中国植物志科学出版社, 1978: 80-86.
- [3] 何顺志, 徐文芬. 贵州中草药资源研究 [M]. 贵阳: 贵州科技出版社, 2007: 735.
- [4] 国家中医药管理局《中华本草》编委会. 中华本草-5 [M]. 上海: 上海科学技术出版社, 1999: 5034.
- [5] 黄泰康. 现代本草纲目 [M]. 北京: 中国医药科技出版社, 2001: 56.
- [6] 汪毅. 精编中草药图谱-4: 便携本 [M]. 第 2 版. 贵阳: 贵州科技出版社, 2012: 85.
- [7] Zhu Z Y, Liu J, Tang S Y, *et al.* Sequencing, assembly, and comparative evolutionary analysis of the chloroplast genome of kenaf (*Hibiscus cannabinus* L.) [J]. *Genes*, 2025, 16(12): 1519.
- [8] Zhang Z, Zhang Y, Song M, *et al.* Species identification of dracaena using the complete chloroplast genome as a super-barcode[J]. *Front Pharmacol*, 2019, 10: 1441.
- [9] Jin J J, Yu W B, Yang J B, *et al.* GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes [J]. *Genome Biol*, 2020, 21(1): 241.
- [10] Shi L C, Chen H M, Jiang M, *et al.* CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. *Nucleic Acids Res*, 2019, 47(W1): W65-W73.
- [11] 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.
- [12] Kurtz S. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Res*, 2001, 29(22): 4633-4642.
- [13] Fuglsang A. Codon optimizer: A freeware tool for codon optimization [J]. *Protein Expr Purif*, 2003, 31(2): 247-249.
- [14] Amiryousefi A, Hyvönen J, Pocai P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [15] 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.
- [16] Zhang Z. KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences [J]. *Genom Proteom Bioinform*, 2022, 20(3): 536-540.
- [17] Wang J, Qian J, Jiang Y, *et al.* Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Polygonatum* and tribe polygonateae [J]. *Front Plant Sci*, 2022, 13: 882189.
- [18] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.
- [19] Minh B Q, Schmidt H A, Chernomor O, *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era [J]. *Mol Biol Evol*, 2020, 37(5): 1530-1534.
- [20] Ronquist F, Teslenko M, van der Mark P, *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space [J]. *Syst Biol*, 2012, 61(3): 539-542.
- [21] Xie J, Chen Y, Cai G, *et al.* TreePlot: a web server for displaying, editing, and annotating phylogenetic trees [J]. *Nucl Acids Res*, 2023, 51(W1): W238-W242.
- [22] Weising K, Gardner R C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms [J]. *Genome*, 1999, 42(1): 9-19.
- [23] Jia X B, Wei J Q, Chen Y W, *et al.* Codon usage patterns and genomic variation analysis of chloroplast genomes provides new insights into the evolution of Aroideae [J]. *Sci Rep*, 2025, 15: 4333.
- [24] Li B Y, Huang K, Chen X L, *et al.* Comparative and phylogenetic analysis of chloroplast genomes from four

- species in *Quercus section Cyclobalanopsis* [J]. *BMC Genom Data*, 2024, 25(1): 57.
- [25] 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.
- [26] The angiosperm phylogeny group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV [J]. *Botan J Linnean Society*, 181(1): 1-20.
- [27] Maier U G, Zauner S, Woehle C, *et al.* Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes [J]. *Genome Biol Evol*, 2013, 5(12): 2318-2329.
- [28] Yao J C, Zheng Z H, Xu T, *et al.* Chloroplast genome sequencing and comparative analysis of six medicinal plants of *Polygonatum* [J]. *Ecol Evol*, 2025, 15(1): e70831.
- [29] Yan M X, Dong S J, Gong Q Y, *et al.* Comparative chloroplast genome analysis of four *Polygonatum* species insights into DNA barcoding, evolution, and phylogeny [J]. *Sci Rep*, 2023, 13: 16495.
- [30] Yao J C, Zheng Z H, Xu T, *et al.* Chloroplast genome sequencing and comparative analysis of six medicinal plants of *Polygonatum* [J]. *Ecol Evol*, 2025, 15(1): e70831.
- [31] Hu Y F, Wang S Q, Xu Z Z, *et al.* Chloroplast genome comparison and taxonomic reassessment of *Polygonatum sensu Lato* (Asparagaceae): Implications for molecular marker development in traditional medicinal plants [J]. *BMC Genom*, 2025, 26(1): 796.

[责任编辑 时圣明]