

桔梗叶绿体比较基因组学分析及系统发育研究

张 瑜, 杜晨晖*, 詹海仙, 尚彩玲, 李瑞锋, 原淑佳

山西中医药大学中药与食品工程学院, 山西 晋中 030619

摘要:目的 以桔梗 *Platycodon grandiflorus* 为材料, 分析其叶绿体基因组特征, 探究不同地区桔梗叶绿体基因组的差异及桔梗科其他物种的系统发育关系。方法 利用 Illumina NovaSeq 测序平台对桔梗叶绿体全基因组进行测序, 完成其组装、注释和特征分析, 采用生物信息学方法对不同地区桔梗进行比较基因组分析和系统发育分析。结果 桔梗叶绿体基因组全长 172 770 bp, 呈现典型的环状四分体结构, 总 GC 含量为 38.10%, 注释到 139 个基因, 其中蛋白质编码基因 95 个, 核糖体 RNA 8 个和转运 RNA 36 个。经序列分析鉴定出 139 个 SSR 位点, 大部分重复由 A 和 T 组成。该叶绿体基因组密码子偏好性 A/U 大于 G/C。边界分析表明, 不同地区桔梗的 JLA 边界区域存在差异。对比不同地区桔梗叶绿体基因组序列发现 21 个变异区间, 包括 *ycf1*、*psbC*、*rps18* 和 *rpoB* 等编码区以及 *rpl32-trnL*、*trnS-psbZ* 和 *trnN-ycf1* 等非编码区。基于最大似然法 (maximum likelihood method, ML) 对桔梗及其他 17 种桔梗科植物进行系统发育分析, 发现桔梗科物种形成一个单系群, 各属物种聚为一束, 支持率达 100%。结论 桔梗科物种聚为一支与传统相符合, 不同地区桔梗叶绿体基因组序列存在显著差异, 为后期开展分子鉴定及群体遗传学研究提供提供科学依据。

关键词: 桔梗; 叶绿体基因组; 序列比对; 系统发育; SSR

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)15-4981-11

DOI: 10.7501/j.issn.0253-2670.2023.15.023

Comparative and phylogeny analysis of *Platycodon grandiflorus* complete chloroplast genomes

ZHANG Yu, DU Chen-hui, ZHAN Hai-xian, SHANG Cai-ling, LI Rui-feng, YUAN Shu-jia

College of Traditional Chinese Medicine and Food Engineering, Shanxi University of Chinese Medicine College, Jinzhong 030619, China

Abstract: Objective To explore the differences of chloroplast genomes of Jiegeng [*Platycodon grandiflorus* (Jacq.) A. DC.] from different regions and phylogenetic relationship between it and its relatives, the complete chloroplast genome of *P. grandiflorus* was sequenced and assembled. **Methods** We sequenced the chloroplast genome of *P. grandiflorus* using Illumina NovaSeq sequencing platform, and assembled, annotated and characterized them, compared the chloroplast genomes sequences using bioinformatics methods, and analyzed the phylogenetic relationships. **Results** The full chloroplast genome of *P. grandiflorus* was 172 770 bp in length and its GC content was 38.10%, with a typical circular tetrad structure. A total of 139 genes were annotated, including 95 protein-coding genes, 8 rRNA genes, and 36 tRNA genes. A total of 139 SSRs were detected, most of which consisted of A and T. The codon preference of A/U was greater than that of G/C. The comparison of boundary region showed that there were differences in inverted repeat region (IR) of two different regions of *P. grandiflorus*. Analyses of sequences from the two different regions of *P. grandiflorus* showed that 21 diverse regions are found in the protein coding regions such as *psbC*, *rps18*, *rpoB*, *ycf1*, and in the intergenic regions such as *rpl32-trnL*, *trnS-psbZ*, *trnN-ycf1*. Based on the maximum likelihood method (ML) for phylogenetic analysis of *P. grandiflorus* and 17 other Campanulaceae plants, it was found that Campanulaceae species formed a single monophyletic group. The clustering of plants of the same genus was in line, and the support rate reached 100%. **Conclusion** The clustering of Campanulaceae species was in line with tradition. There were significant differences in chloroplast genome sequence of *P. grandiflorus* among different regions. These results will provide an important basis for the further development of molecular

收稿日期: 2023-03-10

基金项目: 基于基因组学的“孟桔梗”道地性研究 (SZY-YQZX-2019006); 山西省应用基础研究计划项目 (201901D211540); 山西道地药材品质形成机制创新团队 (2022TD2009); 山西中医药大学中药资源学 (2023XKJS-23)

作者简介: 张 瑜 (1982—), 女, 讲师, 研究方向为中药资源与开发。E-mail: 15364887652@163.com

*通信作者: 杜晨晖, 教授, 硕士生导师, 研究方向为中药鉴定学。E-mail: dch@sxtcm.edu.cn

identification and population genetics.

Key words: *Platycodon grandiflorus* (Jacq.) A. DC.; complete chloroplast genome; sequence alignment; phylogenetic analysis; SSR

药用植物桔梗 *Platycodon grandiflorus* (Jacq.) A. DC. 隶属于桔梗科桔梗属, 为多年生草本, 别名包袱花、铃当花、道拉基^[1]。在《中国药典》2020年版中桔梗以根入药, 具有宣肺利咽、祛痰排脓等功效, 多用于咳嗽痰多、胸闷不畅、咽痛音哑、肺痛吐脓^[2]。桔梗的主要化学成分为含桔梗皂甙, 此外含有丰富的多糖、各种氨基酸及矿物质等, 国家卫生部将桔梗列为药食两用的中药材^[3]。桔梗科有60~70个属, 中国产16个属, 而桔梗属仅有1个种桔梗^[4]。桔梗产自东北、华北、华东、华中各省, 韩国、朝鲜、日本等地区也有分布^[5]。山西省阳泉市盂县山地面积广, 海拔高, 土壤中富含磷钾, 为桔梗的生长提供了得天独厚的自然环境。桔梗作为山西盂县地道中药材, 因其品质好被百姓誉为“小人参”。目前, 药用植物桔梗的叶绿体基因组虽然已经报道^[6], 但是对于不同地区桔梗的叶绿体基因组组装、基因组特征和系统进化分析等未见报道。

叶绿体基因组一般由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 repeat, IR) 组成闭环双链 DNA 结构, 基因组大小多为150~200 kb^[7-9]。叶绿体是植物细胞内半自主性的细胞器, 拥有相对独立的基因组, 依赖于母系遗传, 具有结构稳定、序列保守、间隔区变异位点丰富, 分子进化速率慢等特点, 广泛应用于亲缘关系、系统进化关系及遗传多样性研究^[10-12]。

随着测序技术的不断发展以及测序成本的降低, 已报道多种重要的药用植物叶绿体基因组序列, 如人参 *Panax ginseng* C. A. Meyer^[13]、银杏 *Ginkgo biloba* L.^[14]、厚朴 *Houpoa officinalis* (Rehder & E. H. Wilson) N. H. Xia & C. Y. Wu^[15]、红豆杉 *Taxus wallichiana* var. *chinensis* (Pilger) Florin^[16]、肉苁蓉 *Cistanche deserticola* Ma^[17]、三七 *Panax notoginseng* (Burkill) F. H. Chen ex C. Y. Wu & K. M. Feng^[18]、石斛 *Dendrobium nobile* Lindl.^[19]、丹参 *Salvia miltiorrhiza* Bunge^[20]等。本研究基于高通量测序技术, 对桔梗开展叶绿体全基因组测序、组装及注释, 获得其叶绿体基因组的全长序列信息; 利用生物信息学手段比较分析不同地区桔梗的结构特征、简单重复序列 (simple sequence repeats, SSRs) 位点、

边界分析、序列变异程度等, 同时构建系统进化树, 为深入研究桔梗的鉴定、分子标记开发及系统发育关系提供数据支撑。

1 材料

桔梗新鲜叶片材料采自山西省阳泉市盂县桔梗种植基地 (38°13'55"E, 113°6'16"N), 凭证标本号为 XZ-2020-1。叶片装入取样袋带回实验室, 液氮速冻后放于-80 °C冰箱保存备用。样品由山西中医药大学杜晨晖教授鉴定, 凭证标本存放于山西中医药大学标本馆。桔梗及其近缘物种叶绿体基因组序列来源于 NCBI 数据库, 实验材料详细信息见表 1。

表 1 植物样品来源

Table 1 Plant samples used in this study

物种	来源	GenBank 登录号
桔梗 <i>Platycodon grandiflorus</i>	GenBank	NC_035624
桔梗 (标本号 XZ-2020-1)	山西盂县	MZ202358
沙参 <i>Adenophora stricta</i>	GenBank	NC_036223
<i>Adenophora erecta</i>	GenBank	NC_036222
展枝沙参 <i>Adenophora divaricata</i>	GenBank	NC_036221
薄叶芥苳 <i>Adenophora remotiflora</i>	GenBank	NC_026999
<i>Campanula takesimana</i>	GenBank	NC_026203
西南风铃草 <i>Campanula pallida</i>	GenBank	NC_063742
<i>Campanula zangezura</i>	GenBank	NC_057269
丽江黄钟花 <i>Cyananthus flavus</i>	GenBank	MT074354
<i>Cyananthus obatus</i>	GenBank	NC_063739
小党参 <i>Codonopsis minima</i>	GenBank	NC_036311
羊乳 <i>Codonopsis lanceolata</i>	GenBank	MH018574
毛细钟花 <i>Leptocodon hirsutus</i>	GenBank	NC_049093
蓝花参 <i>Wahlenbergia marginata</i>	GenBank	NC_063740
<i>Lobelia kauaensis</i>	GenBank	MF061196
<i>Lobelia gibberoa</i>	GenBank	NC_035374
<i>Lobelia acrochila</i>	GenBank	NC_035366
铜锤玉带草 <i>Pratia nummularia</i>	GenBank	NC_035381
铜锤玉带草	GenBank	NC_035367
拟南芥 <i>Arabidopsis thaliana</i>	GenBank	NC_000932
人参 <i>Panax ginseng</i>	GenBank	NC_006290

2 方法

2.1 提取 DNA 及测序

采用北京天根生化植物 DNA 提取试剂盒 (Tiangen Biotech 有限公司, 中国) 提取叶片 DNA, 1%琼脂糖凝胶电泳和微量分光光度计 (Nanodrop 2000, 美国) 检测 DNA 质量及浓度。以 1 μg DNA 起始量建库, 随机打断检测合格的基因组总 DNA,

构建长度约 350 bp 的插入片段文库。构建好的文库经过质检后,利用 Illumina NovaSeq 平台进行双末端测序,序列读长为 150 bp,测序工作由北京上锋基因生物科技有限公司完成。

2.2 序列拼接、注释及绘制物理图谱

利用 Trimmomatic 软件^[21]对测序得到的原始序列 (raw reads) 进行质控 (参数设置选择默认值), 滤过去除接头 (adapter) 和低质量序列 (reads), 得到高质量待分析序列 (clean reads)。参考序列为桔梗 (NC_035624), 运用 NOVOPlasty 软件^[22]对桔梗叶绿体基因组组装, 并利用 GapCloser 软件对组装结果进行内洞修补。然后使用 CPGAVAS 软件^[23]进行叶绿体基因组注释, 并运用 Apollo 软件对注释结果进行人工校准, 最终结果提交至 GenBank 获得登录号 (MZ202358)。运用 Organellar Genome DRAW 在线软件^[24]绘制桔梗叶绿体基因组物理图谱。

2.3 简单重复序列分析

利用 MISA 软件检测桔梗叶绿体基因组的 SSRs^[25], 参数设置为: 单核苷酸重复单元 ≥ 10 , 二核苷酸重复单元 ≥ 5 , 三核苷酸和四核苷酸重复单元 ≥ 4 , 四核苷酸、五核苷酸和六核苷酸重复单元 ≥ 3 , 且 2 个 SSRs 之间的最小距离为 100 bp。

2.4 密码子偏好性分析

使用 CodonW 软件^[26] (<http://codonw.sourceforge.net/>) 获得桔梗叶绿体基因组的有效密码子数 (effective number of codon, ENC)、密码子的第 3 个碱基出现 G/C 的概率 (GC3s)、相对同义密码子使用度 (relative synonymous codon usage, RSCU)。以 ENC 值为参考, 对桔梗叶绿体基因组编码序列从大到小排序, 筛选最高和最低两端各 10% 的基因, 分别建立高、低表达基因库, 以 $\Delta RSCU$ (高表达基因 RSCU - 低表达基因 RSCU) ≥ 0.08 为条件, 筛选出高表达密码子; 以 RSCU > 1 为条件, 筛选出高频密码子。同时满足高频和高表达条件的密码子确定为最优密码子^[27]。

2.5 叶绿体基因组 IR 边界的收缩和扩张分析

IR 区域相对保守, IR 边界的膨胀和收缩被认为是被子植物叶绿体全基因组大小变化的主要机制^[28]。本研究使用 IRscope 在线软件 (<https://ir-scope.shinyapps.io/irapp/>) 获得并比较分析桔梗及桔梗科其他 4 种植物叶绿体基因组的 IRA/IRB、LSC 和 SSC 和边界基因的序列长度, 探讨桔梗科植物叶绿体基

因组 IR 边界的收缩和扩张特征^[29]。

2.6 叶绿体基因组序列变异分析

本研究基于 LAGAN 模型, 利用 mVISTA 软件对中国地区桔梗 (MZ202358) 与韩国地区桔梗 (NC_035624) 叶绿体全基因组进行序列比对和变异分析^[30]。

2.7 系统发育分析

从 GenBank 获取紫茎泽兰 (NC_036222)、展枝沙参 (NC_036221)、薄叶芥苳 (NC_026999) 等桔梗科 17 种 (19 个) 物种和 2 个外类群物种的完整叶绿体基因组信息, 分析获得 66 个共有蛋白质编码基因: *atpA*、*atpB*、*atpE*、*atpF*、*atpH*、*atpI*、*ccsA*、*cemA*、*matK*、*ndhA*、*ndhC*、*ndhD*、*ndhF*、*ndhG*、*ndhH*、*ndhI*、*ndhJ*、*petA*、*petB*、*petD*、*petG*、*petL*、*petN*、*psaB*、*psaC*、*psaI*、*psaJ*、*psbA*、*psbB*、*psbD*、*psbE*、*psbF*、*psbH*、*psbI*、*psbK*、*psbL*、*psbM*、*psbN*、*psbT*、*psbZ*、*rbcL*、*rpl14*、*rpl16*、*rpl2*、*rpl20*、*rpl32*、*rpl33*、*rpl36*、*rpoA*、*rpoB*、*rpoC1*、*rpoC2*、*rps11*、*rps12*、*rps14*、*rps15*、*rps16*、*rps18*、*rps19*、*rps2*、*rps3*、*rps4*、*rps7*、*rps8*、*ycf1*、*ycf2*。利用在线软件 MAFFT7 (<https://mafft.cbrc.jp/alignment/server/>) 将上述物种的共有编码基因进行比对, 比对结果使用 IQ-TREE multicore 2.0.5 (<http://www.iqtree.org/>) 软件^[31], 最大似然法 (maximum likelihood method, ML) 构建系统发育树, 构树参数设置: -mMFP- B1000-*alrt* 1000, 优构树模型: GTR+F+R3, Bootstrap value 为 1000。

3 结果与分析

3.1 叶绿体基因组的基本特征

与大多数被子植物相同, 桔梗的叶绿体基因组呈现典型的四分体结构, 由 2 个大小相同、方向相反的反向重复区 (IRa、IRb)、1 个 LSC 区和 1 个 SSC 区组成 (图 1)。桔梗的叶绿体基因总长度为 172 770 bp, 其中 LSC 区长 79 115 bp, SSC 区长 7841 bp, IRa 区长 42 907 bp, IRb 区长 42 907 bp。桔梗叶绿体基因组整体 GC 含量为 38.10%, 其中蛋白编码区的 GC 含量是 38.11%, 4 个区段中 GC 含量最高的是 IR 区 (39.53%), 其次是 LSC 区 (37.25%) 和 SSC 区 (31.03%)。

桔梗叶绿体基因组共注释到 139 个基因, 包括 95 个蛋白编码基因、36 个 tRNA 基因和 8 个 rRNA 基因 (表 2)。蛋白编码基因的长度为 90 594 bp, 占整个基因组长度的 52.44%; tRNA 基因的长度

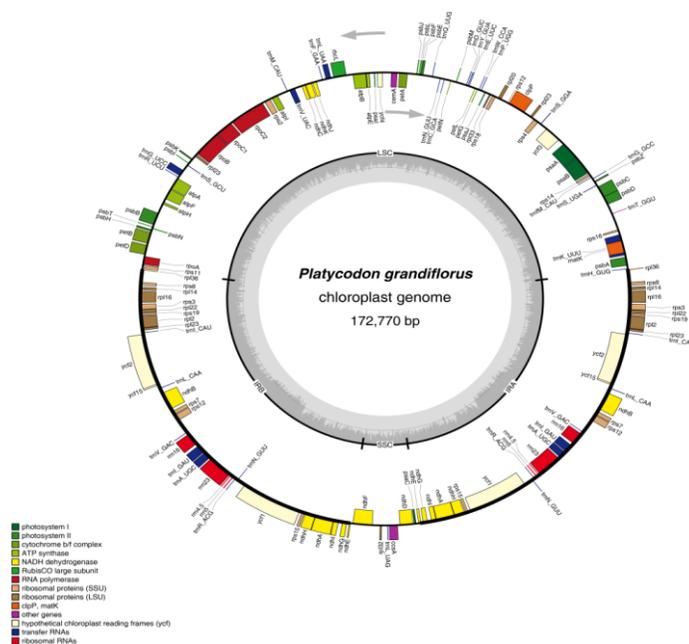


图 1 桔梗叶绿体基因组物理图谱

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

表 2 桔梗叶绿体基因组注释信息

Table 2 Annotation information of chloroplast genomes of *P. grandiflorus*

种类	基因群类型	基因
自我复制	核糖体 RNA	<i>rrn23S</i> (×2), <i>rrn16S</i> (×2), <i>rrn5S</i> (×2), <i>rrn4.5S</i> (×2)
	转运 RNA	<i>trnH-GUG</i> , <i>trnK-UUU</i> , <i>trnQ-UUG</i> , <i>trnS-GCU</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnY-GUA</i> , <i>trnE-UUC</i> , <i>trnR-UCU</i> , <i>trnS-UGA</i> , <i>trnG-GCC</i> , <i>trnM-CAU</i> , <i>trnS-GGA</i> , <i>trnL-UAA</i> , <i>trnF-GAA</i> , <i>trnW-CCA</i> , <i>trnP-UGG</i> , <i>trnL-CAA</i> (×2), <i>trnV-GAC</i> (×2), <i>trnA-UGC</i> (×2), <i>trnR-ACG</i> (×2), <i>trnN-GUU</i> (×3), <i>trnL-UAG</i> , <i>trnG-UCC</i> , <i>trnI-CAU</i> (×4), <i>trnV-UAC</i> , <i>trnJm-CAU</i>
	核糖体蛋白大亚基	<i>rpl14</i> (×2), <i>rpl16</i> (×2), <i>rpl2</i> (×2), <i>rpl20</i> , <i>rpl22</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i> (×2)
	RNA 聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoCl</i> , <i>rpoC2</i>
	核糖体蛋白小亚基	<i>rps11</i> , <i>rps12</i> (×2), <i>rps14</i> , <i>rps15</i> (×2), <i>rps16</i> , <i>rps18</i> , <i>rps19</i> (×2), <i>rps2</i> , <i>rps3</i> (×2), <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> (×2)
光合系统	ATP 合成酶	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
	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> (×2), <i>ndhB</i> (×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> (×2), <i>ndhH</i> (×2), <i>ndhI</i> (×2), <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>pet</i> , <i>petN</i>
	光系统 I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> <i>psaJ</i>
	Rubisco 亚基	<i>rbcL</i>
其他基因	细胞色素 c 合成酶	<i>ccsA</i>
	包膜蛋白	<i>cemA</i>
	成熟酶	<i>matK</i>
未知	保守的开放阅读框	<i>ycf1</i> (×2), <i>ycf15</i> (×2), <i>ycf2</i> (×2), <i>ycf3</i> , <i>ycf4</i>

为 2713 bp, 占总长度的 1.57%; rRNA 基因的长度为 9176 bp, 占总长度的 5.31%。非编码区主要包括内含子、假基因 (*trnT-GGU*, *rpl23*, *clpP*, *ndhE* 等) 和基因间隔区, 占整个基因组长度的 40.68%。在桔梗叶绿体基因组注释到 4 个具有内含子的基因, 包括 *trnT_GGU*, *rpl23*, *clpP*, *ndhE*。

3.2 密码子的偏好性

桔梗叶绿体基因组的密码子使用偏性中, 叶绿

体基因组的 ENC 值为 51.24, 明显高于 35, 表明密码子偏好性较弱。GC3s 含量为 28.61%, 表明密码子第 3 位偏好于 A 和 U 结尾。通过最优密码子分析显示 (表 3), 有 30 个密码子 RSCU>1, 25 个密码子 $\Delta RSCU \geq 0.08$, 筛选发现同时满足这 2 个条件的最优密码子有 9 个, 分别为 UUG、AUU、UCA、AGC、CCU、GCU、CAU、CGU 和 GAU, 其中以 A/U 结尾的有 7 个, 以 C/G 结尾的有 2 个。

表3 桔梗叶绿体基因组最优密码子

Table 3 Optimal codons in chloroplast genome of *P. grandiflorus*

氨基酸	密码子	高表达基因		低表达基因		ΔRSCU	RSCU
		数目	RSCU	数目	RSCU		
Phe	<u>UUU</u>	23	1.14	15	1.50	-0.36	1.32
	UUC	9	0.56	5	0.50	0.06	0.68
Leu	<u>UUA</u>	23	1.68	22	2.06	-0.38	1.73
	UUG*	19	1.39	14	1.31	0.08	1.20
	<u>CUU</u>	21	1.54	16	1.50	0.04	1.39
	CUC*	5	0.37	2	0.19	0.18	0.42
	CUA	10	0.73	9	0.84	-0.11	0.88
	CUG*	4	0.29	1	0.09	0.20	0.37
Ile	AUU*	44	1.47	21	1.26	0.21	1.43
	AUC	17	0.57	12	0.72	-0.15	0.64
	AUA	29	0.97	17	1.02	-0.05	0.93
Val	<u>GUU</u>	23	1.46	17	1.55	-0.09	1.52
	GUC*	8	0.51	3	0.27	0.24	0.53
	<u>GUA</u>	23	1.46	20	1.82	-0.36	1.38
	GUG*	9	0.57	4	0.36	0.21	0.57
Ser	<u>UCU</u>	11	1.27	9	1.42	-0.15	1.83
	UCC	10	1.15	7	1.11	0.04	0.88
	UCA*	8	0.92	4	0.63	0.29	1.10
	UCG	4	0.46	5	0.79	-0.33	0.55
	AGU	13	1.50	12	1.89	-0.39	1.76
	AGC***	6	0.69	1	0.16	0.53	1.01
	CCU**	20	1.54	12	1.23	0.31	1.53
Pro	CCC*	13	1.00	9	0.92	0.08	0.71
	<u>CCA</u>	13	1.00	11	1.13	-0.13	1.11
	CCG	6	0.46	7	0.72	-0.26	0.65
	<u>ACU</u>	12	1.12	14	1.56	-0.44	1.60
	ACC***	14	1.30	7	0.78	0.52	0.70
Ala	<u>ACA</u>	12	1.12	13	1.44	-0.32	1.22
	ACG*	5	0.47	2	0.22	0.25	0.48
	GCU*	32	1.47	13	1.27	0.20	1.62
	GCC*	16	0.74	5	0.49	0.25	0.71
	<u>GCA</u>	19	0.87	16	1.56	-0.69	1.15
Tyr	<u>GCG*</u>	20	0.92	7	0.68	0.24	0.52
	<u>UAU</u>	30	1.62	17	1.89	-0.27	1.57
	UAC*	7	0.38	1	0.11	0.27	0.43
His	CAU*	15	1.67	5	1.43	0.24	1.46
	CAC	3	0.33	2	0.57	-0.24	0.54
Gln	<u>CAA</u>	21	1.17	13	1.44	-0.27	1.49
	CAG*	15	0.83	5	0.56	0.27	0.51
Asn	<u>AAU</u>	43	1.56	20	1.54	0.02	1.50
	AAC	12	0.44	6	0.46	-0.02	0.50
Lys	<u>AAA</u>	42	1.50	38	1.55	-0.05	1.51
	AAG	14	0.50	11	0.45	0.05	0.49
Asp	GAU*	20	1.29	9	1.20	0.09	1.50
	GAC	11	0.71	6	0.80	-0.09	0.50
Glu	<u>GAA</u>	41	1.44	24	1.78	-0.34	1.48
	GAG**	16	0.56	3	0.22	0.34	0.52
Cys	<u>UGU</u>	6	1.20	6	2.00	-0.80	1.44
	UGC*	4	0.80	0	0.00	0.80	0.56
Arg	CGU*	15	1.30	14	1.15	0.15	1.10
	CGC*	7	0.61	4	0.33	0.28	0.42
	<u>CGA</u>	17	1.48	25	2.05	-0.57	1.20
	CGG	4	0.35	4	0.33	0.02	0.50
	<u>AGA</u>	21	1.83	22	1.81	0.02	1.20
	AGG*	5	0.43	4	0.33	0.10	0.44
Gly	<u>GGU</u>	32	1.39	22	1.69	-0.30	1.26
	GGC	6	0.26	3	0.23	0.03	0.56
	<u>GGA</u>	33	1.43	20	1.54	-0.11	1.49
	GGG**	21	0.91	7	0.54	0.37	0.69

下划线表示 RSCU > 1; *表示该密码子 ΔRSCU ≥ 0.08, **表示 ΔRSCU ≥ 0.3, ***表示 ΔRSCU ≥ 0.5; 加粗标注的密码子为最优密码子

The underline indicates RSCU > 1; *represents the codon ΔRSCU ≥ 0.08, **represents ΔRSCU ≥ 0.3, and ***represents ΔRSCU ≥ 0.5; The codons highlighted in bold are the optimal codons

3.3 SSR 位点分析

共检测到桔梗叶绿体基因组 139 个 SSR 位点，包括 115 个单核苷酸重复序列、6 个二核苷酸重复序列、5 个三核苷酸重复序列、11 个四核苷酸重复序列和 2 个六核苷酸重复序列。其中，60 个 SSR 位于基因间隔区 (IGS)，65 个 SSR 位于外显子 (exon)，14 个 SSR 位于内含子 (intron)。叶绿体不同区域的 SSR 分布情况见图 2，大部分 SSR 位于 LSC 和 IRs 区域内，只有少数位于 SSC 区域内。SSR 中重复占比最大的是单核苷酸，约 82.73%，而二核苷酸、三核苷酸、四核苷酸和六核苷酸则占比减小，分别占 4.32%、3.60%、7.91% 和 1.44%。桔梗叶绿体基因组中的 SSR 主要是有 A 和 T 组成的，占有所有重复序列的 83.45%，其中 A/T 碱基构成的单核苷酸重复 108 条，AT/AT 组成的二核苷酸重复序列 5 条，ATT/TAA 组成的三核苷酸

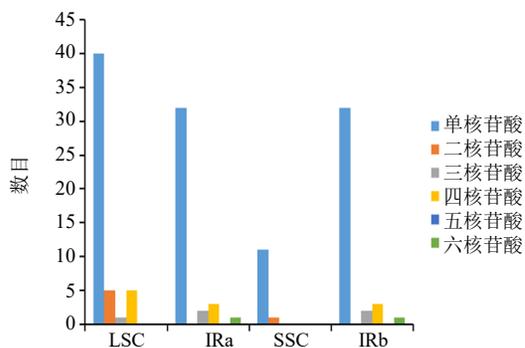


图 2 桔梗叶绿体基因组中 SSR 位点类型及数量

Fig. 2 Type and number of SSR loci of *P. grandiflorus* chloroplast genome

重复序列 2 条，AAAT 组成的四核苷酸重复序列 1 条。

桔梗科 5 种植物叶绿体基因组序列中，总共检测到 SSR 位点 103~143 个 SSR 位点 (表 4)，其中单核苷酸重复为 90~117 个、二核苷酸重复为 6~8 个、三核苷酸重复 5~14 个、四核苷酸重复 1~11 个、五核苷酸重复 0~4 个、六核苷酸重复 0~2 个。这些 SSR 主要分布于叶绿体基因组的 LSC 区 (36.70%~65.12%)，编码基因序列中分布的 SSR 数量仅占总数的 32.56%~54.37%。

3.4 桔梗及桔梗科部分物种叶绿体全基因组特征比较

桔梗科部分植物叶绿体全基因组特征的比较分析见表 5，桔梗及其近缘物种共 5 种 (6 个) 植物叶绿体基因组的序列长度范围为 159 759~172 770 bp，其中沙参的叶绿体基因组长度最短，而桔梗的叶绿体基因组最长，不同地区的桔梗叶绿体基因组长度相差 952 bp。6 个物种叶绿体基因组总 GC 含量为 38.10%~39.03%，IR 区的 GC 含量 (39.53%~51.00%) 均显著高于 LSC 区 (36.86%~38.18%) 和 SSC 区 (31.03%~35.42%)。其中桔梗的 GC 含量最低，铜锤玉带草的 GC 含量最高。不同地区桔梗全基因组的总 GC 含量相差不大，分别是 38.10% 和 38.12%。

利用 Geneious 11.0.3 软件获取桔梗及 4 种近缘桔梗科植物的 IR 区、LSC 区和 SSC 区序列。结果表明，IR 区的序列长度范围为 20 200~85 814 bp，其中沙参的 IR 区最短，桔梗的 IR 区最长，不同地区桔梗的 IR 区的长度相差 948 bp。LSC 区序列长

表 4 桔梗科物种叶绿体基因组中 SSR 信息统计

Table 4 SSR information statistics of chloroplast genomes of Campanulaceae species

类型和区域	名称	桔梗 (MZ202358)	桔梗 (NC_035624)	铜锤玉带草 (NC_035381)	沙参 (NC_036223)	小党参 (NC_036311)	毛细钟花 (NC_049093)
类型	单核苷酸重复	115	111	90	98	96	117
	二核苷酸重复	6	6	6	6	8	7
	三核苷酸重复	5	5	6	14	9	8
	四核苷酸重复	11	11	1	8	6	7
	五核苷酸重复	0	0	0	2	1	4
	六核苷酸重复	2	2	0	1	0	0
区域	LSC	51	51	49	84	62	72
	SSC	12	12	16	32	8	17
	IR	76	72	38	12	50	54
	CDS	65	65	56	42	49	57
	合计	139	135	103	128	120	143

表 5 桔梗科植物叶绿体全基因组基本特征

Table 5 Basic characteristics of chloroplast genomes of Campanulaceae species

基本结构特征	桔梗 (MZ202358)	桔梗 (NC_035624)	铜锤玉带草 (NC_035381)	沙参 (NC_036223)	小党参 (NC_036311)	毛细钟花 (NC_049093)
基因组长度/bp	172 770	171 818	165 428	159 759	169 321	169 524
IR 区长度/bp	85 814	84 866	74 682	20 200	76 286	76 286
LSC 区长度/bp	79 115	79 061	82 764	112 321	85 506	85 326
SSC 区长度/bp	7841	7840	7982	27 238	8065	7912
基因组 GC/%	38.10	38.12	39.03	38.54	38.28	38.24
LSC 区 GC/%	37.25	37.25	38.18	37.05	36.86	36.88
IR 区 GC/%	39.53	39.58	40.74	51.00	40.50	40.40
SSC 区 GC/%	31.03	31.03	31.98	35.42	32.44	32.03

度范围为 79 061~112 321 bp，其中桔梗的 LSC 区长度最短 (79 061 bp)，沙参 LSC 区长度最长 (112 321 bp)，不同地区的桔梗 LSC 区的长度相差 54 bp。SSC 区序列长度范围介于 7840~27 238 bp，其中沙参的 SSC 区最长 (27 238 bp)，桔梗的 SSC 区最短 (7840 bp)，不同地区桔梗的 SSC 区的长度仅相差 1 bp。

3.5 桔梗及桔梗科部分物种的叶绿体基因组边界分析

桔梗及桔梗科 5 个属共 5 种植物叶绿体基因组的 IR-LSC 和 IR-SSC 边界比较显示 (图 3)，桔梗与其余 4 种桔梗科植物叶绿体全基因组相比，边界处基因不完全相同，基因的长度也有差异。LSC/IRb 边界 (JLB) 边界扩张范围显示，桔梗与其余 4 种桔梗科植物相比，JLB 边界侧翼基因为 *rps11* 和 *rpl36* 基因，而铜锤玉带草、毛细钟花和小党参 JLB 边界处具有相同的基因，左翼为 *rps19* 基因，右翼为 *rpl2* 基因，但扩张程度稍有差异，沙参的 JLB 边界位于

trnL 基因左翼。SSC/IRb (JSB) 边界扩张范围显示，桔梗、毛细钟花与铜锤玉带草 JSB 边界位于 *ndhE* 基因，但扩张程度不同。小党参 JSB 边界位于 *ndhE* 基因的右翼，距离 1 bp，而沙参 JSB 边界位于 *ndhF* 基因的左翼。SSC/IRa (JSA) 边界扩张范围显示，桔梗和小党参 JSA 边界位于 *ndhE* 基因内，但扩张程度稍有差异。毛细钟花和铜锤玉带草 JSB 边界位于 *ndhF* 基因右翼，距离分别为 273、139 bp，而沙参 JSB 边界位于 *clpP* 基因右翼。LSC/IRa (JLA) 边界扩张范围显示，桔梗与其余 4 种桔梗科植物 JLA 边界位于 *trnH* 基因左翼，但扩张程度有差异。不同地区的桔梗其边界具有高度的保守性，各边界侧翼基因完全相同，扩张程度略有差异。

3.6 叶绿体基因组序列变异分析

本研究测序的桔梗植物叶绿体基因组与韩国地区桔梗全基因组序列相似性比较结果见图 4。结果表明，2 种不同地区桔梗叶绿体基因组序列的整

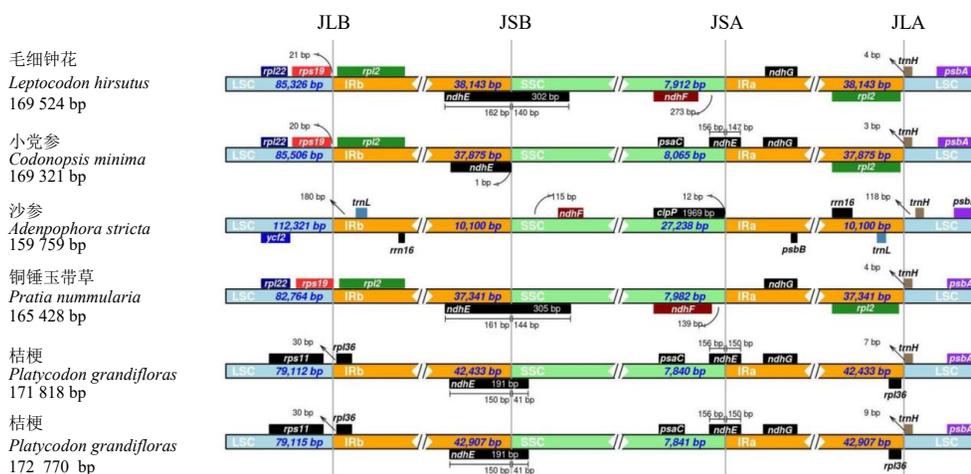


图 3 桔梗科物种叶绿体基因组边界序列差异

Fig. 3 Comparison of sequence divergence of boundary region among Campanulaceae species

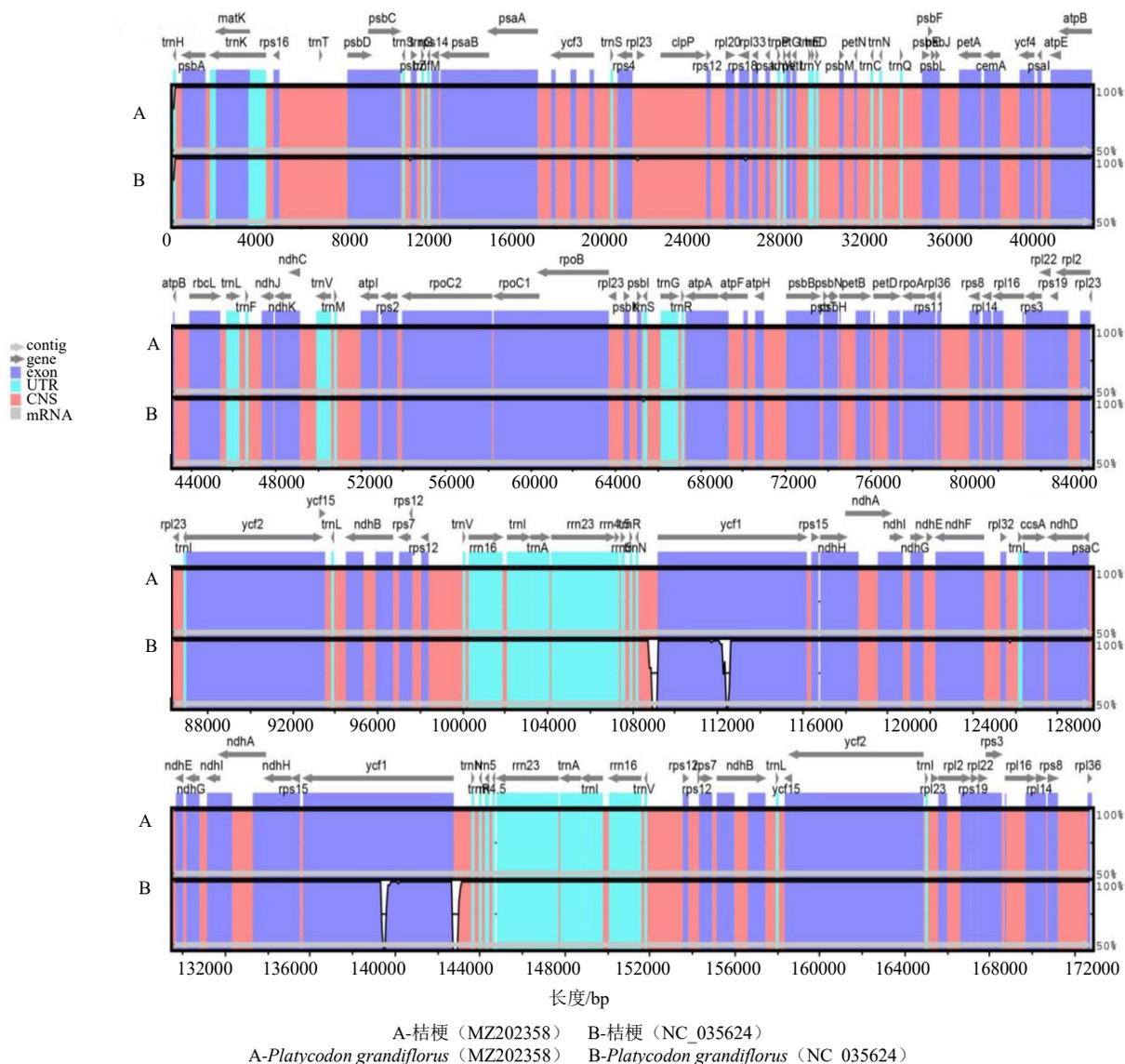


图 4 不同地区桔梗叶绿体全基因组比对分析

Fig. 4 Alignment analysis of chloroplast genomes of *P. grandiflorus* from different regions

体相似性较高，但仍存在明显的差异。变异程度最高的区域集中在 IR 区，编码区和非编码区都有不同程度的变异。*psbC*、*rps18*、*rpoB*、*ycf1* 等基因的编码区存在 7 个单碱基突变 (SNP)，*ycf1* 基因的编码区有 2 个插入突变。非编码区存在 7 个单碱基突变和 5 个插入突变，包括 *rpl32-trnL*、*trnS-psbZ*、*trnN-ycf1* 等，如新测序桔梗在 *trnN-ycf1* 基因间隔区 108 704 bp 处插入 255 个碱基。这些变异位点是不同地区桔梗鉴定的潜在分子标记，具有极高的研究价值。

3.7 系统进化分析

本研究以桔梗、2 个外类群和已公布的 17 种 (18 个) 桔梗科植物的 66 个共有蛋白编码基因进行比

对，将比对结果采用 IQ-TREE 软件 ML 法构建系统进化树 (图 5)。结果显示，19 个桔梗科物种构成一单系分支，显著区别于外类群拟南芥和人参，具有 100% 支持率。在系统树上，桔梗科物种可以分为 3 个主要分支。A 分支是半边莲亚科构成的单系分支，支持率为 100%。B 分支和 C 分支由桔梗亚科构成，支持率为 100%。分支 C 包括 2 个族，桔梗族和蓝钟花族。不同地区的桔梗聚为一小支，进一步与党参属、蓝钟花属、细钟花属物种聚为一支。分支 B 是风铃草族构成的单系分支，支持率为 100%。基于叶绿体基因组的系统进化树与已知的物种进化关系一致，为桔梗科物种系统进化关系的建立提供依据。

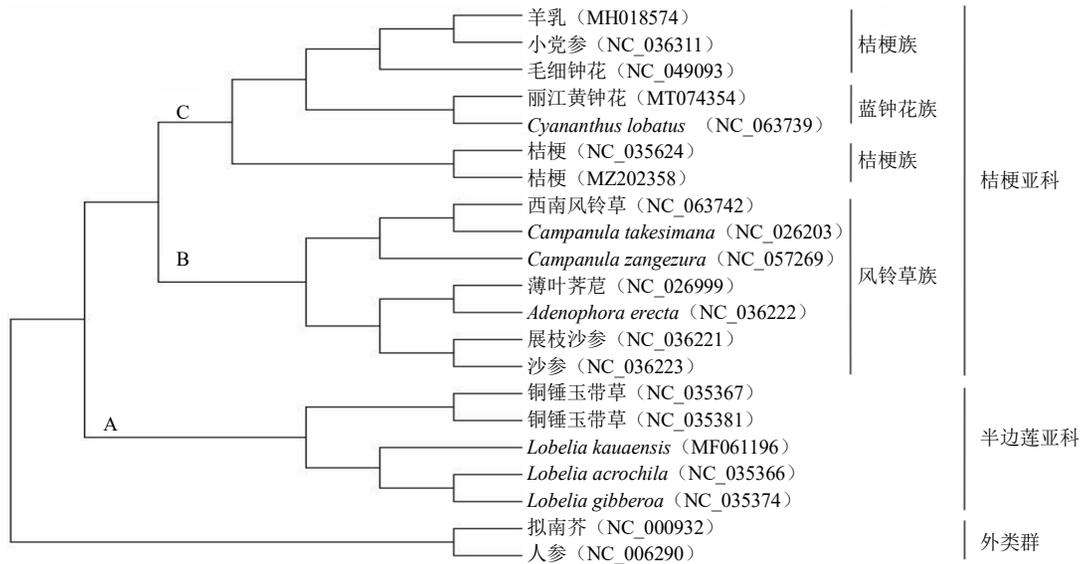


图 5 基于 66 个蛋白编码基因构建的 ML 树

Fig. 5 ML tree constructed based on 66 protein-coding genes

4 讨论

桔梗始载于《神农本草经》，是中医常用的一味止咳平喘的良药。本研究完成了桔梗叶绿体基因组的测序、组装和注释。结果表明，桔梗叶绿体基因组具有被子植物叶绿体基因组典型的四分体结构，包括 1 个 LSC 区、1 对 IR 区和 1 个 SSC 区，序列长 172 770 bp，其长度高于其他 4 个近缘种。不同地区桔梗叶绿体基因序列长度在 171 800~172 800 bp，IR 区的差异最大（952 bp），原因是本研究测序的桔梗位于 IR 区的 *ycf1* 基因存在明显的序列插入现象，导致其比已报道的韩国地区桔梗序列长。桔梗叶绿体基因组 GC 含量低于其他 4 个近缘种。不同地区桔梗叶绿体基因组 GC 含量基本一致（38.10%~38.12%），IR 区的 GC 含量均高于其他区域，可能与 IR 区含有高 GC 含量的 rRNA 基因有关。不同地区桔梗叶绿体基因组分别注释到 139 和 140 个基因，包括 95 个蛋白编码基因、36~37 个 tRNA 基因和 8 个 rRNA 基因，体现了叶绿体基因组的稳定性。

植物叶绿体基因组中 IR 区域的收缩和扩张是一种相对普遍的现象，使相同植物群体之间的整个叶绿体基因组之间出现差异^[32]。本研究中桔梗与其他 4 个近缘种 IR 边界分析显示，其 LSC/IRb 边界（JLB）、SSC/IRb 边界（JSB）、SSC/IRa 边界（JSA）和 LSC/IRa 边界（JLA）的侧翼基因类型和扩张程度有明显差异，对研究物种进化等具有重要意义。

不同地区桔梗叶绿体基因组边界基因和扩张程度完全相同，仅 JLA 边界与侧翼基因 *trnH* 的距离相差 2 bp。本研究中桔梗叶绿体基因组 SSR 位点有 139 个，以单核苷酸重复（占总数的 83.40%）为主要类型，且随着拷贝数目增加，SSRs 数量明显减少，这种现象在其它植物中也有报道^[33]。桔梗叶绿体基因组中的 SSR 主要分布于 LSC 和 IR 区，多数由碱基 A 和 T 组成，与石斛属部分物种的 SSR 碱基组成相似^[34]。不同地区桔梗的核苷酸重复类型基本相同，但 SSR 的数量有差异，原因是 IR 区的 SSR 数量存在差异。桔梗科不同属的物种叶绿体基因组中的 SSR 的数量和分布模式差异显著，这些 SSRs 为桔梗科物种鉴定和系统发育分析奠定基础。

已有研究表明，叶绿体基因组常应用于物种鉴定和系统发育关系研究^[10-12]。比对 2 个不同地区的桔梗叶绿体基因组序列编码区和非编码区都有不同程度的变异，*psbC*、*rps18*、*rpoB*、*ycf1* 等基因的编码区和 *rpl32-trnL*、*trnS-psbZ*、*trnN-ycf1* 等基因间隔区存在 14 个单碱基突变和 7 个插入缺失突变，基于 21 个差异区可开发特异 DNA 条形码，鉴定不同地区的桔梗，为优化种质资源及良种繁育奠定基础。

为进一步界定药用植物桔梗在桔梗科的系统位置，探讨桔梗属与其近缘属的系统发育关系。基于桔梗科 19 个物种 66 个共有编码基因构建系统发育树结果显示，所有的桔梗科物种形成一个单系群，支持率为 100%。桔梗族物种与蓝钟花族物种聚为

一组,该结果与 Li 等^[34]的研究结果类似。不同地区的桔梗聚为一小支,进一步与党参属、蓝钟花属、细钟花属物种聚为一支,说明桔梗属、党参属、蓝钟花属、细钟花属亲缘关系较近。有学者认为铜锤玉带属与半边莲属有密切的亲缘关系,应合并为半边莲属^[10]。铜锤玉带草原隶属于铜锤玉带属,但目前《中国植物志》已修订铜锤玉带草为半边莲属^[35]。图 5 显示铜锤玉带草与其他半边莲属物种互为姐妹群,支持“半边莲属与铜锤玉带属应合并成为半边莲属”的说法^[10]。本研究中桔梗科系统进化树与已知的物种进化关系一致。

本研究对桔梗叶绿体基因组重新测序,综合分析桔梗的叶绿体基因组序列、结构和特征,筛选出不同地区桔梗叶绿体基因组的差异序列,探讨桔梗科物种系统进化关系,结果不仅丰富了桔梗的叶绿体基因组信息,为叶绿体基因组序列中筛选特异性 DNA 条形码基因序列及系统发育关系研究提供了参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 赵琪,王玲,吴杨璐,等. 中药桔梗的本草考证 [J]. 中药材, 2022, 45(5): 1266-1272.
- [2] 中国药典 [S]. 一部. 2020: 284.
- [3] 蒋桃,祖矩雄,李兰岚. 药用型和食用型桔梗定向选育的研究进展 [J]. 中国中医药现代远程教育, 2018, 16(20): 158-160.
- [4] 刘自刚,张雁,王新军,等. 桔梗育种研究进展 [J]. 中草药, 2006, 37(6): 962-964.
- [5] 张燕,李阳,李倩,等. 桔梗种质资源研究新进展 [J]. 中国野生植物资源, 2017, 36(3): 53-56.
- [6] Hong C P, Park J, Lee Y, et al. accD nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae [J]. *BMC Genom*, 2017, 18(1): 1-13.
- [7] Twyford A D, Ness R W. Strategies for complete plastid genome sequencing [J]. *Mol Ecol Resour*, 2017, 17(5): 858-868.
- [8] Yu X Y, Zuo L H, Lu D D, et al. Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis [J]. *Gene*, 2019, 689: 141-151.
- [9] 李雪佩,孟静,张琳娜,等. 黄草乌与展毛黄草乌叶绿体全基因组结构的比较分析 [J]. 中药材, 2018, 41(8): 1812-1820.
- [10] Li C J, Lu J M, Li Y, et al. Comparative morphology of the leaf epidermis in *Lobelia* (Lobelioideae) from China [J]. *Microsc Res Tech*, 2017, 80(7): 763-778.
- [11] 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.
- [12] Yu B, Sun Y B, Liu X F, et al. Complete chloroplast genome sequence and phylogenetic analysis of *Camellia fraterna* [J]. *Mitochondrial DNA B*, 2020, 5(4): 3840-3842.
- [13] Kim K J, Lee H L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants [J]. *DNA Res*, 2004, 11(4): 247-261.
- [14] Jiao P, Qi Z, Jiang Z Z, et al. Characterization of the complete chloroplast genome of *Ginkgo biloba* L. (Ginkgoaceae), an endangered species endemic to China [J]. *Mitochondrial DNA B*, 2019, 4(2): 4128-4129.
- [15] 李西文,胡志刚,林小涵,等. 基于 454FLX 高通量技术的厚朴叶绿体全基因组测序及应用研究 [J]. 药学报, 2012, 47(1): 124-130.
- [16] Zhang Y Z, Ma J, Yang B X, et al. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): Loss of an inverted repeat region and comparative analysis with related species [J]. *Gene*, 2014, 540(2): 201-209.
- [17] 宋菊,龙月红,林丽梅,等. 五加科植物叶绿体基因组结构与进化分析 [J]. 中草药, 2017, 48(24): 5070-5075.
- [18] Biswal D, Konhar R, Debnath M, et al. Chloroplast genome sequence annotation of *Dendrobium nobile* (Asparagales: Orchidaceae), an endangered medicinal orchid from northeast India [J]. *PLoS Curr*, 2017: 78-89.
- [19] Hu J L, Zhao M A, Hou Z J, et al. The complete chloroplast genome sequence of *Salvia miltiorrhiza*, a medicinal plant for preventing and treating vascular dementia [J]. *Mitochondrial DNA B*, 2020, 5(3): 2460-2462.
- [20] 李冉郡,武立伟,辛天怡,等. 大黄药材基原物种叶绿体基因组分析与特异 DNA 条形码开发 [J]. 药学报, 2022, 57(5): 1495-1505.
- [21] Guo Y P, Yao B Q, Yuan M R, et al. The complete chloroplast genome and phylogenetic analysis of *Astragalus scaberrimus* Bunge 1833 [J]. *Mitochondrial DNA B*, 2021, 6(12): 3364-3366.
- [22] Wu M L, Li Q, Xu J, et al. Complete chloroplast genome of the medicinal plant *Amomum compactum*: Gene organization, comparative analysis, and phylogenetic relationships within Zingiberales [J]. *Chin Med*, 2018, 13(1): 1-12.
- [23] Skuza L, Gastineau R, Sielska A. The complete

- chloroplast genome of *Secale sylvestre* (Poaceae: Triticeae) [J]. *J Appl Genet*, 2022, 63(1): 115-117.
- [24] Zhu B, Gao Z M, Qian F, *et al.* The complete chloroplast genome of a purple Ethiopian rape (*Brassica Carinata*: Brassicaceae) from Guizhou Province, China and its phylogenetic analysis [J]. *Mitochondrial DNA B*, 2021, 6(7): 1821-1823.
- [25] Wang W B, Yu H A, Wang J H, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J]. *Int J Mol Sci*, 2017, 18(11): 2288.
- [26] 杨林, 谭萍, 姜月影, 等. 3种木莲属濒危药用植物叶绿体基因组密码子偏好性分析 [J]. 中国中医药信息杂志, 2022, 29(8): 93-99.
- [27] Wang Z J, Cai Q W, Wang Y, *et al.* Comparative analysis of *Codon* bias in the chloroplast genomes of Theaceae species [J]. *Front Genet*, 2022, 13: 824610.
- [28] Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.
- [29] Wu L W, Cui Y X, Wang Q, *et al.* Identification and phylogenetic analysis of five *Crataegus* species (Rosaceae) based on complete chloroplast genomes [J]. *Planta*, 2021, 254(1): 14.
- [30] Ma J Y, Wang J M, Li C Y, *et al.* The complete chloroplast genome characteristics of *Polygala crotalarioides* Buch.-Ham. ex DC. (Polygalaceae) from Yunnan, China [J]. *Mitochondrial DNA B*, 2021, 6(10): 2838-2840.
- [31] Mehmood F, Shahzadi I, Waseem S, *et al.* Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational hotspots [J]. *Genomics*, 2020, 112(1): 581-591.
- [32] 杨小英, 刘畅, 曾宪法, 等. 八角枫及其亚种叶绿体基因组序列结构及系统发育分析 [J]. 药学报, 2022, 57(10): 3229-3239.
- [33] 李卓蔚, 邱迁, 郎佳琪, 等. 尖刀唇石斛和翅梗石斛叶绿体全基因组分析 [J]. 中草药, 2022, 53(16): 5159-5169.
- [34] Li C J, Wang R N, Li D Z. Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications [J]. *PLoS One*, 2020, 15(5): e0233167.
- [35] 中国科学院中国植物志编辑委员会. 中国植物志(第七十三卷, 第二分册) [M]. 北京: 科学出版社, 1983: 56.
[责任编辑 时圣明]