

## 桔梗特异DNA条形码筛选、种质资源鉴定及遗传多样性分析

王馨<sup>1</sup>, 石玥<sup>3</sup>, 满金辉<sup>1</sup>, 黄钰莹<sup>1</sup>, 张晓芹<sup>1</sup>, 安克露<sup>1</sup>, 何高洁<sup>1</sup>, 刘子齐<sup>4</sup>,  
关范圆<sup>1</sup>, 郑语嫣<sup>1</sup>, 王晓晖<sup>2\*</sup>, 魏胜利<sup>1\*</sup>

(1. 北京中医药大学中药学院, 北京 102488; 2. 北京中医药大学, 北京中医药研究院, 中药现代研究中心, 北京 102488; 3. 北京中医药大学生命科学院, 北京 102488; 4. 黑龙江诺初中药材种植有限公司, 黑龙江 哈尔滨 150600)

**摘要:** 中药材桔梗为桔梗科植物桔梗的干燥根, 具有多种药理作用, 是常用的大宗中药材。本研究利用 Illumina HiSeq X Ten 平台对 6 份来自不同产地的桔梗进行叶绿体基因组测序, 筛选特异性 DNA 条形码, 基于特异性 DNA 条形码对不同产区的桔梗样品的种质资源和遗传多样性进行分析。6 份桔梗叶绿体基因组全长为 172 260~172 275 bp, 均呈现典型的环状四分体结构, 编码 141 个基因。根据比较基因组学分析和扩增效率分析发现 *trnG-UCC* 和 *ndhG\_ndhF* 可作为潜在的桔梗种内种质资源鉴定的特异性 DNA 条形码。对来自 9 省 15 个产地 305 份桔梗样品的 *trnG-UCC* 和 *ndhG\_ndhF* 进行 PCR 扩增和序列分析, 结果表明 *trnG-UCC* 和 *ndhG\_ndhF* 分别有 5 和 11 个变异位点, 分别鉴定到 5 和 7 个单倍型; 两段序列联合分析鉴定 13 个单倍型 (Hap1~Hap13), 其中占比最多的是 Hap4, 其次是 Hap1。3 个产地拥有的特异单倍型, 可作为该产地的 DNA 分子标签与其他产地的桔梗种质资源进行区分。单倍型多样性、核苷酸多样性和遗传距离分别为 0.94、 $4.79 \times 10^{-3}$  和 0.000 0~0.020 3, 表明桔梗在物种水平上有较高的遗传多样性, 种内各单倍型之间亲缘关系比较接近。本研究为桔梗产地鉴定、种质资源保护利用和分子育种等工作奠定基础。

**关键词:** 桔梗; 叶绿体基因组; DNA 条形码; 种质资源; 遗传多样性

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2024)01-0243-10

## Specific DNA barcodes screening, germplasm resource identification, and genetic diversity analysis of *Platycodon grandiflorum*

WANG Xin<sup>1</sup>, SHI Yue<sup>3</sup>, MAN Jin-hui<sup>1</sup>, HUANG Yu-ying<sup>1</sup>, ZHANG Xiao-qin<sup>1</sup>, AN Ke-lu<sup>1</sup>,  
HE Gao-jie<sup>1</sup>, LIU Zi-qi<sup>4</sup>, GUAN Fan-yuan<sup>1</sup>, ZHENG Yu-yan<sup>1</sup>, WANG Xiao-hui<sup>2\*</sup>, WEI Sheng-li<sup>1\*</sup>

(1. School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 102488, China; 2. Modern Research Center for Traditional Chinese Medicine, Beijing Institute of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 102488, China; 3. School of Life Sciences, Beijing University of Chinese Medicine, Beijing 102488, China; 4. Heilongjiang Nuochu Medical Materials Planting Co., Ltd., Harbin 150600, China)

**Abstract:** *Platycodonis Radix* is the dry root of *Platycodon grandiflorum* of Campanulaceae, which has a variety of pharmacological effects and is a commonly used bulk Chinese medicine. In this study, the chloroplast genome sequences of six *P. grandiflorum* from different producing areas has been sequenced with Illumina HiSeq X Ten platform. The specific DNA barcodes were screened, and the germplasm resources and genetic diversity were analyzed according to the specific barcodes. The total length of the chloroplast genome of 6 *P. grandiflorum* samples was 172 260–172 275 bp, and all chloroplast genomes showed a typical circular tetrad structure and encoded

收稿日期: 2023-04-23; 修回日期: 2023-07-25.

基金项目: 国家重点研发计划项目(2022YFX3501505); 桔梗等 4 种精准药材批次分子防伪技术研究项目(2020110031009385).

\*通讯作者 E-mail: wangxhui2014@163.com; wsl7491@126.com

DOI: 10.16438/j.0513-4870.2023-0497

141 genes. The comparative genomics analysis and results of amplification efficiency demonstrated that *trnG-UCC* and *ndhG\_ndhF* were the potential specific DNA barcodes for identification the germplasm resources of *P. grandiflorum*. A total of 305 *P. grandiflorum* samples were collected from 15 production areas in 9 provinces, for which the fragments of *trnG-UCC* and *ndhG\_ndhF* were amplified and the sequences were analyzed. The results showed that *trnG-UCC* and *ndhG\_ndhF* have 5 and 11 mutation sites, respectively, and 5 and 7 haplotypes were identified, respectively. The combined analysis of the two sequences formed 13 haplotypes (named Hap1-Hap13), and Hap4 is the main genotype, followed by Hap1. The unique haplotypes possessed by the three producing areas can be used as DNA molecular tags in this area to distinguish from the germplasm resources of *P. grandiflorum* from other areas. The haplotype diversity, nucleotide diversity and genetic distance were 0.94,  $4.79 \times 10^{-3}$  and 0.000 0–0.020 3, respectively, suggesting that the genetic diversity was abundant and intraspecific kinship was relatively close. This study laid a foundation for the identification of *P. grandiflorum*, the protection and utilization of germplasm resources, and molecular breeding.

**Key words:** *Platycodon grandiflorum*; chloroplast genome; DNA barcode; germplasm resource; genetic diversity

中药材桔梗 (*Platycodonis Radix*) 具有宣肺利咽, 祛痰排脓的功效, 是最常用的大宗药材之一。《中华人民共和国药典》(2020年版, 一部) 规定中药材桔梗来源于桔梗科药用植物桔梗 *Platycodon grandiflorum* (Jacq.) A. DC. 的干燥根<sup>[1]</sup>。现代药理学作用研究表明其具有多种药理活性, 如抗肿瘤、抗菌消炎、免疫调节和降血压降血脂等<sup>[2-5]</sup>作用。近年来随着桔梗需求量增大, 需要不断扩大桔梗的生产规模, 因此对桔梗的种质资源鉴定和遗传多样性分析有利于后续桔梗优良品种选育和利用。

叶绿体是被子植物光合作用的器官, 同时也是半自主细胞器, 拥有自己的基因组, 依赖母系遗传, 称为叶绿体基因组。叶绿体基因组的结果是双链环状构型, 包括一个小的单拷贝区 (SSC) 和一个大单拷贝区 (LSC); 两区域被一对反向重复区域 (IRA 和 IRB) 区分开, 为典型的四分体结构。叶绿体基因组相对分子量较小, 基因组成和结构相对保守, 包含的大量遗传信息被广泛应用于植物分子进化及系统发育的研究<sup>[6,7]</sup>。随着高通量测序技术和生物信息学的快速发展, 多种植物的叶绿体基因组已经解析。目前桔梗科的桔梗、半边莲 *Lobelia chinensis*、沙参 *Adenophora stricta* 等<sup>[8-10]</sup>物种的叶绿体基因组已有研究报道, 但是种内叶绿体比较基因组学相关研究鲜有报道。

2003年, Hebert等<sup>[11]</sup>首次提出DNA条形码是标准化的、较短的DNA序列, 可作为条形码实现对物种进行快速、准确地鉴定, 是传统性状鉴定和理化鉴定方法的有效补充。如 *matK* 能够鉴别大黄不同基原及混伪品<sup>[12]</sup>。随着叶绿体基因组测序技术和比较叶绿体基因组学的发展, 通过比较叶绿体基因组序列差异获得可应用于鉴别不同物种、不同基原甚至同一物种的不同产地的特异DNA条形码也变得切实可行, 如利用叶绿体基因

组及比较基因组学方法获得特异性条形码 *rps16-trnQ*、*psaA\_ycf3*、*psbE\_petL*、*ndhF\_rpl32* 以及 *trnT\_trnL* 用于鉴定唐古特大黄、药用大黄和掌叶大黄<sup>[13]</sup>。基于叶绿体基因组测序和比较基因组学发现黄芩 *petA\_psbJ*、*ycf4\_cemA* 和 *trnH\_psbA* 三个基因片段突变位点最多, 可用作特异DNA条形码, 利用3段叶绿体基因组DNA的联合分析对黄芩的野生居群展开研究, 共发现29个变异位点, 形成50个单倍型<sup>[14]</sup>。*rbcL* 序列可鉴别不同产地的山慈菇<sup>[15]</sup>。*psbA\_trnH* 序列可鉴别何首乌的道地种源和其他种源<sup>[16]</sup>。

*ITS*、*psbA\_trnH*、*matK*、*trnG-S*、*rbcL* 等条形码为桔梗科植物的分类鉴定与系统学研究提供了重要依据<sup>[17]</sup>。Zhao等<sup>[18]</sup>验证了 *ITS* 序列可对桔梗药材及其易混品进行鉴别。Wu等<sup>[19]</sup>通过 *ITS2* 序列对桔梗进行研究发现其种内遗传性较为丰富。Nie等<sup>[20]</sup>研究发现, *rbcL* 序列可用于桔梗科内不同属间的鉴别。目前基于桔梗核基因的种质资源鉴别已较为深入, 不仅能对桔梗及易混品进行鉴别, 还发现了桔梗种内具有较高的遗传多样性。而基于桔梗叶绿体基因的研究较少, 仅发现了可用于桔梗科的鉴别序列 *rbcL*, 基于桔梗叶绿体比较基因组学的种内特异DNA条形码研究还未见报道, 因此本研究选取不同产地的6个桔梗样品进行叶绿体测序, 进行比较基因组学研究, 筛选特异性DNA条形码, 并利用筛选的特异DNA条形码对9省15产地305份桔梗样品进行种质资源鉴定和遗传多样性分析, 弥补了目前研究的空缺, 为后续桔梗种质筛选、质量控制及相关研究奠定基础。

## 材料与方法

**植物材料** 用于叶绿体全基因组测序的6份桔梗样品分别采自内蒙古自治区巴彦淖尔市、黑龙江省哈

尔滨市、河北省保定市安国市、陕西省商洛市商州区、安徽省亳州市谯城区和河南省洛阳市;用于种质资源鉴定的15个产地305份样品来源如表1所示。所用样品经北京中医药大学魏胜利教授鉴定为桔梗植株 *Platycodon grandiflorum* (Jacq.) A. DC.。

**基因组DNA提取与测序** 利用FastPure® Plant DNA Isolation Mini Kit (诺唯赞)提取植物总DNA, 1.0%的琼脂糖凝胶电泳检测DNA质量, Nano Drop one (Thermo Fisher公司)检测DNA浓度。检测合格的基因组总DNA构建插入片段长度约350 bp的文库, 利用Illumina HiSeq X Ten平台进行序列读长为150 bp的双端测序, 并对原始序列进行4步处理, 分别是去除质量值连续≤20的碱基数达到40%的reads、去除含N的碱基数目总和达到10%的reads、去除adapter污染和去除duplication污染, 得到高质量待分析序列(clean reads)。

**叶绿体基因组组装、拼接和注释** 使用NOVO-Plasty将待分析序列组装成完整的叶绿体基因组, 采用PGA软件进行组装结果进行注释(默认参数)。采用BWA将待分析序列比对桔梗参考基因组(NC\_035624)的叶绿体基因组序列, 利用在线软件tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>)确定所有tRNA基因的边界, 并通过CLC Sequence Viewer 8人工检查确保组装无误。利用Organellar Genome DRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>)在线绘制叶绿体全基因组图谱。

**叶绿体基因组比较分析及单倍型多样性分析** 利用mVISTA在线软件对测序获得6份叶绿体基因组进行全局比对分析, 利用Dna SP 6软件检测6条叶绿体基因组的核苷酸多样性(Pi)。通过Dna SP 6软件对15个产区的单倍型进行多样性分析。

**高变区域扩增、遗传距离分析与进化树构建** 为

利用高变区基因对不同产区的种质资源进行分析, 以高变区基因为模板设计引物(*ndhG\_ndhF*-F: 5'-TCAC TTTTTCGAAGGGGGAATC-3'; *ndhG\_ndhF*-R: 5'-TCT AGAGACTAAAATTTTCAC-3'; *trnG-UCC*-F: 5'-TAGT GGTAAGTGTGATT-3'; *trnG-UCC*-R: 5'-GGCTAG GGGTTATAGTCGA-3'), 对收集到的15个产区305份样品进行PCR扩增分析。PCR混合体系共50 μL, 包括ddH<sub>2</sub>O 37 μL、10×Buffer 5 μL、dNTP 4 μL、引物-F 1 μL、引物-R 1 μL、DNA模板0.5 μL、0.1% BSA 1 μL和TaKaRa Taq 0.5 μL。扩增程序为93 °C 3 min, 50 °C 2 min, 30个循环(93 °C 30 s, 45.9 °C 45 s, 70 °C 45 s), 70 °C 5 min, 4 °C保存。通过1.0%的琼脂糖凝胶电泳和FastPure® Gel DNA Extraction Mini Kit (诺唯赞)对PCR产物进行纯化, 纯化后的产物送北京六合华大公司进行双向测序。

利用DNAMAN和Chromas软件对测序结果进行核对和单倍型汇总。将汇总后的单倍型序列通过mafft软件进行比对。比对后的序列通过MEGA X软件计算单倍型间的遗传距离, 构建邻接法系统进化树, 设置bootstrap重复值为1 000。

## 结果与分析

### 1 叶绿体全基因组测序、序列组装与拼接注释

将桔梗1~6 (*P. grandiflorum* 1~6)的测序结果过滤去除低质量序列和接头序列等杂质, 分别得到29 841 912条(4.48 Gb)、27 638 910条(4.14 Gb)、29 934 124条(4.49 Gb)、28 810 240条(4.32 Gb)、28 895 078条(4.33 Gb)和33 150 576条(4.97 Gb) clean reads。将6份桔梗叶绿体全基因组组装拼接后均得到完整的环状四分体结构, 序列总长度分别为172 272、172 260、172 267、172 268、172 271和172 275 bp。环状四分体结构由1个LSC、1个SSC和2个反向重复区

**Table 1** Source of 305 samples of *P. grandiflorum* from 15 regions

Source	Sample (number)	Longitude	Latitude	Altitude/m
Houma, Linfen, Shanxi	SXLFHM (85)	E110°52'10.40"	N38°13'88.26"	713
Louyang, Henan	HNLY (14)	E112°43'44.68"	N34°63'04.10"	585
Baoding, Hebei	HBBD (4)	E113°45'32.00"	N38°14'29.00"	25
Yiyuan, Zibo, Shandong	SDZB (26)	E118°17'10.50"	N36°18'53.60"	154
Qiaocheng, Bozhou, Anhui	AHBZ (16)	E115°60'59.64"	N33°76'49.58"	36
Dalian, Liaoning	LNDL (12)	E121°44'00.00"	N39°01'00.00"	29
Shangzhou, Shangluo, Shannxi	SXSL (6)	E109°92'44.18"	N33°87'86.34"	1 037
Chaoyang, Liaoning	LNCY (31)	E120°16'46.44"	N41°40'91.13"	307
Jingyu, Baishan, Jilin	JLBS (26)	E127°16'00.00"	N42°48'00.00"	132
Kalaqinqi, Chifeng, Neimenggu	NMGCF (18)	E118°65'34.58"	N42°02'92.66"	652
Jian, Jilin	JLJA (18)	E125°45'00.00"	N40°52'00.00"	136
Chifeng, Neimenggu	NMGCF2 (20)	E118°65'34.58"	N42°02'92.66"	652
Hongshan, Chifeng, Neimenggu	NMGHS (18)	E118°57'33.37"	N42°15'56.92"	742
Chishang, Boshan, Zibo, Shandong	SDCS (6)	E117°51'42.19"	N36°29'41.03"	230
Zhangjiakou, Hebei	HBZJK (5)	E118°17'10.50"	N36°18'53.60"	154



**Table 2** Gene composition of six chloroplast genomes of *P. grandiflorum*. \*: Genes with one intron; \*\*: Genes with two introns. (2) means genes with two copies

Category	Gene group	Gene name	
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaj</i>	
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	
	Subunits of NADH ehydrogenase	<i>ndhA*(2), ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG(2), ndhH(2), ndhI(2), ndhJ, ndhK</i>	
	Subunits of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	
	Large subunit of rubisco	<i>rbcL</i>	
	Self-replication	Proteins of large ribosomal subunit	<i>rpl14(2), rpl16*(2), rpl2*(2), rpl20, rpl22(2), rpl23(2), rpl32, rpl33, rpl36(2)</i>
Proteins of small ribosomal subunit		<i>rps11, rps12**(2), rps14, rps15(2), rps16*, rps18, rps19(2), rps2, rps3(2), rps4, rps7(2), rps8(2)</i>	
Subunits of RNA polymerase		<i>rpoA, rpoB, rpoC1*, rpoC2</i>	
Ribosomal RNAs		<i>rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)</i>	
Transfer RNAs		<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnY-M-CAU</i>	
Other genes		Maturase	<i>matK</i>
		Protease	<i>clpP**(2)</i>
		Envelope membrane protein	<i>cemA</i>
		c-type cytochrome synthesis gene	<i>ccsA</i>
Genes of unknown function		Conserved hypothetical	<i>ycf1(2), ycf2(2), ycf3**, ycf4</i>
	chloroplast ORF		

*ycf3*、*ndhG*、*ndhF*等23个片段具有不同程度的种内变异,与mVISTA结果一致。综合上述结果分析表明,*ycf1*、*rps4*、*ndhJ*、*ropB*、*trnG-UCC*、*rpl2*、*rps18*、*rpl23*、*rps16*、*trnT-GGU*、*psaA*、*ycf3*、*rpl23*、*clpP*、*psbZ*、*trnG-GCC*、*rps8*、*rpl36*、*ndhG*、*ndhF*等23个片段可以作为桔梗种内种质资源鉴定的潜在的特异DNA条形码。对23个基因设计引物,进行PCR扩增,发现*trnG-UCC*和*ndhG*、*ndhF*的扩增效率为100%,因此本研究选择*trnG-UCC*和*ndhG*、*ndhF*为桔梗种内鉴定的特异DNA条形码,进行后续研究。

### 3 不同产地桔梗样品的 *trnG-UCC* 和 *ndhG*、*ndhF* 的扩增及分析

以15个产地305份桔梗样品总DNA为模板对筛选的2个叶绿体基因组高变区进行PCR扩增,并对测序结果进行分析。利用设计的*trnG-UCC*引物进行PCR扩增,全部样品扩增到500~750 bp之间,对PCR产物纯化后测序,与PG1 (PG: *Platycodon grandiflorum*)的*trnG-UCC*序列对比可判断其为*trnG-UCC*序列;利用设计的*ndhG*、*ndhF*引物进行PCR扩增,全部样品扩增到500~750 bp之间,对PCR产物纯化后测序,与PG1的*ndhG*、*ndhF*序列对比可判断其为*ndhG*、*ndhF*序列。

基因*trnG-UCC*共检测到5个变异位点,共形成5个单倍型(表3)。其中位于430和597 bp的2个位点突变类型为插入,位于141、344和374 bp的3个位点为单点

突变。基因*ndhG*、*ndhF*共检测到11个变异位点,共形成7个单倍型(表4)。其中位于589和591 bp的2个位点突变类型为插入和缺失突变;位于588、590和592~596 bp的7个位点突变类型为缺失突变;其他2个位点为单点突变,分别位于401和427 bp。对两个基因联合分析共形成了13个单倍型,占比最多分布最广的单倍型为Hap4,占全部样品的45.90%,其次是Hap1,占比为30.39%。如表5所示,山西省临汾市侯马市拥有特异单倍型Hap3、Hap12和Hap13;吉林省白山市靖宇县拥有特异单倍型Hap6;吉林省集安市拥有特异性单倍型Hap8,特异单倍型可作为该产地特有的种质资源加以扩繁,亦可作为鉴定桔梗产地来源的DNA分子标签。

**Table 3** Nucleotide position of *trnG-UCC*. \*: Nucleic acid is the same as GHap1; -: Nucleic acid deletion

No.	Nucleotide position							<i>trnG-UCC</i> haplotype
	141	344	374	430	-	597	-	
1	T	G	G	A	-	T	-	GHap1
2	C	*	*	*	*	*	*	GHap2
3	C	A	*	*	*	*	*	GHap3
4	C	*	A	*	*	*	*	GHap4
5	C	*	*	A	A	T	T	GHap5

### 4 遗传多样性分析

通过DNAsp 6对联合分析的单倍型进行多样性分析(表6)。15个产地的居群变异位点数(S)和单倍

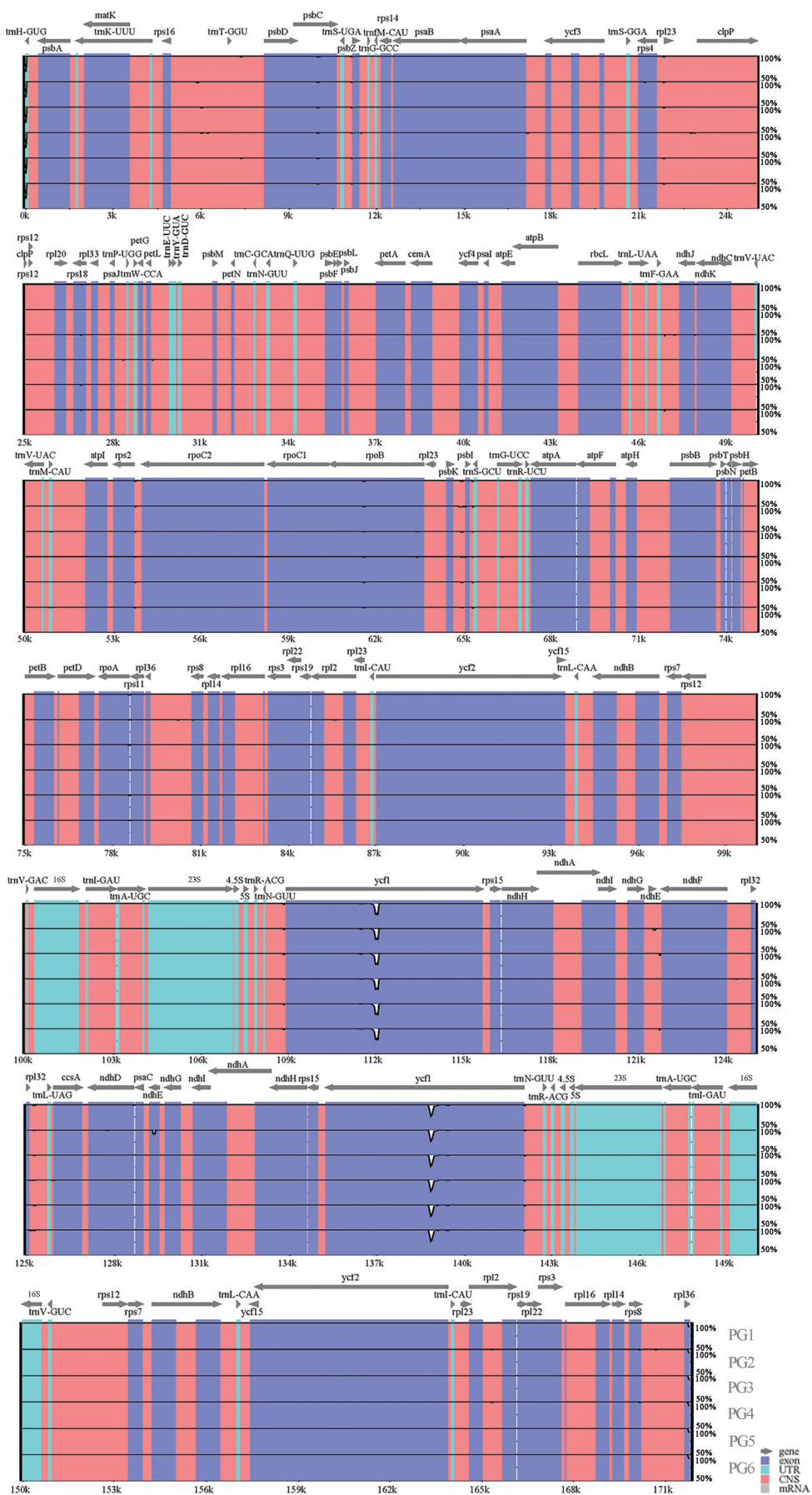


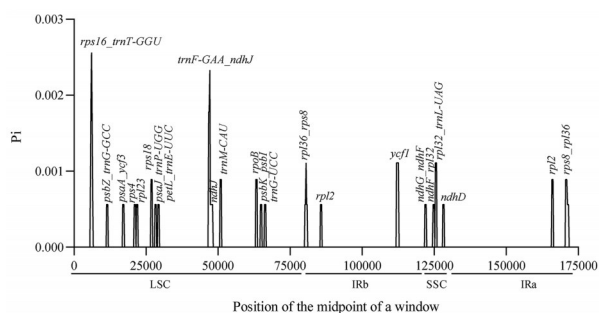
Figure 2 Global comparison of six chloroplast genomes of *P. grandiflorum* (PG)

**Table 4** Nucleotide position of *ndhG\_ndhF*. \*: Nucleic acid is the same as FHap1; -: Nucleic acid deletion

No.	Nucleotide position																<i>ndhG_ndhF</i> haplotype	
	401	427	588	589	-	-	-	-	590	591	-	-	592	593	594	595		596
1	C	T	A	T	-	-	-	-	A	T	-	-	T	A	T	A	T	FHap1
2	*	*	-	-	*	*	*	*	-	-	*	*	*	*	*	*	*	FHap2
3	*	*	*	*	*	*	*	*	*	T	A	T	*	*	*	*	*	FHap3
4	*	G	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	FHap4
5	*	*	*	*	*	*	*	*	*	-	*	*	-	-	-	-	-	FHap5
6	*	*	*	T	A	T	A	T	*	*	*	*	*	*	*	*	*	FHap6
7	T	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	FHap7

**Table 5** Haplotype distribution table for different source. \*: Unique haplotype

Source	Haplotype distribution
SXLFHM	Hap1, Hap2, Hap3*, Hap4, Hap9, Hap10, Hap11, Hap12*, Hap13*
HNLV	Hap1, Hap4, Hap5, Hap7, Hap10
HBBD	Hap1
SDZB	Hap1, Hap4, Hap10
AHBZ	Hap1, Hap4, Hap7
LNDL	Hap1, Hap4
SXSL	Hap1, Hap2
LNCY	Hap1, Hap4, Hap5, Hap7, Hap11
JLBS	Hap1, Hap4, Hap6*, Hap7, Hap9
NMGCF	Hap1, Hap4, Hap9
JLJA	Hap1, Hap4, Hap8*, Hap10
NMGCF2	Hap1, Hap4, Hap7
NMGHS	Hap1, Hap4, Hap7, Hap9, Hap11
SDCS	Hap4, Hap10
HBZJK	Hap1, Hap2, Hap4

**Figure 3** Nucleotide diversity of six chloroplast genome of *P. grandiflorum*. Window length: 600 bp; Step size: 200 bp. Pi: Nucleotide diversity

型数 (h) 分别为 19 和 9 个, 单倍型多样性 ( $H_d$ ) 和核苷酸多样性 (Pi) 分别为 0.94 和  $4.79 \times 10^{-3}$ ,  $H_d$  的变化范围在 0.00~1.00 之间, Pi 的变化范围在  $0 \sim 7.40 \times 10^{-3}$  之间, 表明其遗传多样性较为丰富。为检验桔梗居群的扩张情况, 对联合分析的单倍型进行中性检验, 结果显示, 15 个居群的 Tajima's D、Fu&Li's D 分别为 -0.93 和 -0.49, 且差异均未达显著水平 ( $P > 0.10$ ), 表明桔梗的进化遵循中性进化模式。

## 5 遗传距离分析

使用 Mega X 对 2 个基因联合分析的单倍型进行

**Table 6** Genetic diversity of *P. grandiflorum* from 15 regions. S: Number of segregating sites; h: Number of haplotypes;  $H_d$ : Haplotype diversity; Fu&Li's D: Fu&Li's D neutral test; Tajima's D: Tajima's D neutral test

Population	S	h	$H_d$	Pi ( $\times 10^{-3}$ )	Fu&Li's D	Tajima's D
SXLFHM	23	7	0.94	5.96	-1.12	-1.17
HNLV	10	4	0.90	3.21	-0.60	-0.60
HBBD	0	1	0.00	0.00	-	-
SDZB	9	3	1.00	4.38	-	-
AHBZ	1	2	0.67	0.47	-	-
LNDL	1	2	1.00	0.71	-	-
SXSL	8	2	1.00	5.79	-	-
LNCY	7	4	0.90	2.17	-0.75	-0.75
JLBS	8	3	0.80	2.45	-0.81	-0.81
NMGCF	1	2	0.67	0.47	-	-
JLJA	19	4	1.00	7.40	-0.85	-0.85
NMGCF2	1	2	0.67	0.47	-	-
NMGHS	2	3	0.80	0.72	0.24	0.24
SDCS	9	2	1.00	6.57	-	-
HBZJK	9	3	1.00	4.34	-	-
Total	19	9	0.94	4.79	-0.49	-0.93

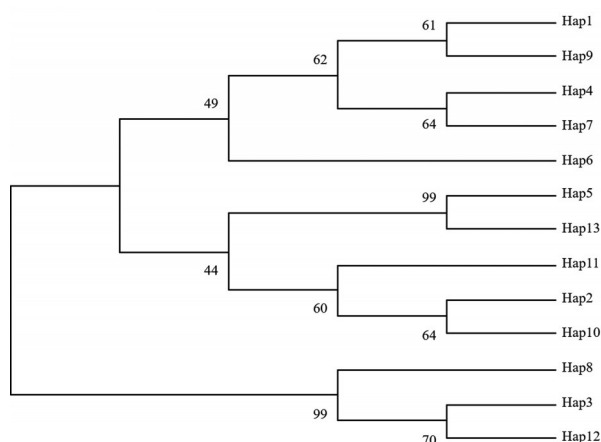
遗传距离分析 (表 7), 结果显示, 13 个单倍型的遗传距离为 0.000 0~0.020 3, 平均遗传距离为 0.007 5, 表明桔梗样品间亲缘关系差别不大。最大遗传距离 0.020 3 存在于 Hap5 和 Hap8 之间, 亲缘关系最远; 最小遗传距离为 0.000 0 存在于多个单倍型之间, 分别在 Hap4 和 Hap6 之间, Hap4 和 Hap7 之间, Hap6 和 Hap7 之间, Hap10 和 Hap11 之间, 亲缘关系最近。占比较多的 Hap1 和 Hap4 之间遗传距离为 0.000 7, 亲缘关系较近。吉林省白山市的特异单倍型 Hap6 与山西省临汾市侯马县的特异单倍型 Hap3 之间的遗传距离为 0.011 9, 亲缘关系相对较远; 吉林省集安市的特异单倍型 Hap8 与山西省临汾市侯马县的特异单倍型 Hap13 之间的遗传距离为 0.019 5, 亲缘关系相对较远; 吉林省白山市的特异单倍型 Hap6 与吉林省集安市的特异单倍型 Hap8 之间的遗传距离为 0.012 3, 亲缘关系相对较远。

## 6 系统进化分析

将 2 个基因联合分析后的单倍型序列对比后构建 NJ Tree, 从而进一步分析各单倍型的亲缘关系。如图 4 所示, Hap3、Hap8 和 Hap12 聚在同一支上, 其他单倍

**Table 7** Genetic distance of haplotypes by joint analysis

	Hap1	Hap2	Hap3	Hap4	Hap5	Hap6	Hap7	Hap8	Hap9	Hap10	Hap11	Hap12	Average
Hap1													
Hap2	0.002 9												
Hap3	0.012 6	0.010 2											
Hap4	0.000 7	0.003 6	0.013 4										
Hap5	0.008 0	0.006 5	0.015 8	0.007 3									
Hap6	0.000 7	0.002 2	0.011 9	0.000 0	0.003 7								
Hap7	0.000 7	0.003 6	0.013 4	0.000 0	0.007 3	0.000 0							0.007 5
Hap8	0.011 4	0.014 8	0.007 9	0.012 2	0.020 3	0.012 3	0.012 2						
Hap9	0.000 0	0.002 9	0.012 6	0.000 7	0.008 0	0.000 7	0.000 7	0.011 4					
Hap10	0.002 9	0.000 7	0.009 4	0.003 7	0.005 9	0.002 2	0.003 7	0.014 9	0.002 9				
Hap11	0.002 2	0.000 7	0.009 4	0.002 9	0.005 8	0.001 5	0.002 9	0.014 0	0.002 2	0.000 0			
Hap12	0.014 8	0.012 5	0.005 5	0.015 6	0.018 0	0.014 2	0.015 6	0.010 2	0.014 8	0.011 7	0.011 7		
Hap13	0.007 3	0.005 8	0.015 0	0.008 0	0.000 7	0.004 4	0.008 0	0.019 5	0.007 3	0.005 1	0.005 1	0.017 2	

**Figure 4** Haplotypes NJ Tree of combined analysis of two genes

型聚为一支,同一支的单倍型之间亲缘关系较近。占比较多的 Hap1 和 Hap4 分布在同一单系分支,亲缘关系较近。与遗传距离分析结果一致,遗传距离最大的 Hap5 和 Hap8 分布于两个不同的单系分支,遗传距离最近的 Hap4 和 Hap7、Hap4 和 Hap6、Hap6 和 Hap7、Hap10 和 Hap11 位于同一分支。

吉林省白山市的特异单倍型 Hap6 与山西省临汾市侯马县的特异单倍型 Hap3 位于进化树的不同分支,亲缘关系相对较远;吉林省集安市的特异单倍型 Hap8 与山西省临汾市侯马县的特异单倍型 Hap13 位于进化树的不同分支,亲缘关系相对较远;吉林省白山市的特异单倍型 Hap6 与吉林省集安市的特异单倍型 Hap8 位于进化树的不同分支,亲缘关系相对较远,与遗传距离的分析结果一致。

## 讨论

桔梗叶绿体基因均具有典型的双链环状四分体结构,由 4 部分构成,分别为 LSC、SSC 和 2 个方向相反的 IR<sup>[8]</sup>。比对 NCBI 中的 4 条桔梗叶绿体全基因组序列

(NC035624、MZ202358、OK317683 和 KX352464),发现桔梗基因组长度为 171 818~173 345 bp,编码基因数量为 138~139 个,蛋白质编码基因数量为 93~94 个,tRNA 编码基因数量为 37 个,rRNA 编码基因数量为 8 个。本研究所测 6 份桔梗叶绿体全基因组结构和大小基本符合上述研究结果。

目前已有研究<sup>[21,22]</sup>利用叶绿体测序和比较基因组进行筛选高变区作为特异 DNA 条形码用于物种的鉴定,而验证特异性 DNA 条形码的实验比较少。如有研究者<sup>[21]</sup>通过比较叶绿体基因组学推荐将 *psbI\_trnS*、*clpP-ex1\_psbB*、*rpl33\_rps18*、*psbC\_trnS*、*psaJ\_rpl33* 等叶绿体基因组高变区作为鉴定石豆兰属药用植物潜在的特异性 DNA 条形码;通过叶绿体基因组学分析筛选 *yef1*、*rpl32\_trnL-UAG*、*trnK-UUU\_rps16*、*psbK\_trnQ-UUG*、*petN\_psbM* 作为龙胆属植物鉴定的高变片段<sup>[22]</sup>。目前也有对叶绿体基因组高变区作为鉴定物种特异 DNA 条形码进行的验证研究,如基于叶绿体基因组筛选的秦艽的两个高变区基因 *trnT\_trnL* 和 *yef1*,可作为特异性 DNA 片段用于长梗秦艽、全萼秦艽的鉴别<sup>[23]</sup>;高变区 *ccsA* 和 *trnC-GCA\_petN* 可作为特异 DNA 条形码鉴别不同产地北苍术<sup>[24]</sup>,高变区 *atpI* 和 *atpB-rbcL* 可作为特异 DNA 条形码鉴别用于刺五加种内鉴别<sup>[25]</sup>。但是也有相关报道证明利用叶绿体基因组测序和比较基因组学筛选的特异 DNA 条形码后扩增效果不理想,如豆蔻属的特异性 DNA 片段 *atpH\_atpI* 扩增效果不理想,不适合豆蔻属的鉴定<sup>[26]</sup>。综合文献研究表明,在筛选特异 DNA 条形码后,需要经过实际扩增效率验证。虽然桔梗的叶绿体基因组已经有报道,但是不同地区桔梗样品叶绿体基因组研究及种内比较基因组学目前还没有研究。本研究对来自不同地区的 6 份桔梗样品进行基因组测序,进行比较基因组学分析,结果表明 *yef1*、*rps4*、*ndhJ*、*ropB*、*trnG-UCC*、*rpl2*、*rps18*、*rpl23*、*rps16\_trnT-GGU*、*psaA\_yef3*、*rpl23\_clpP*、*psbZ\_trnG-*

*GCC*、*rps8\_rpl36*、*ndhG\_ndhF*等23个DNA片段可作为桔梗潜在的特异性DNA片段,根据片段的大小及扩增效率选择*trnG-UCC*和*ndhG\_ndhF*作为特异性DNA片段用于鉴定桔梗种质资源。

目前DNA条形码技术也在桔梗科药用植物鉴定方面取得显著进展,如*ITS*、*psbA\_trnH*、*matK*、*trnG-S*、*rbcL*等DNA片段可以进行桔梗科植物分类鉴定<sup>[17-20]</sup>,但是目前基于叶绿体基因对桔梗种内资源鉴别的研究没有报道。本研究基于叶绿体基因组测序筛选得到2条DNA条形码*trnG-UCC*和*ndhG\_ndhF*对来自9省15个产地305份桔梗样品进行种质资源鉴定,结果表明两基因联合分析共有13个单倍型,15个居群中主要分布的单倍型为Hap1和Hap4。在15个居群中发现3个居群具有特异单倍型,通过遗传距离和系统发育分析发现特异单倍型间亲缘关系相对较远,可作为该地特有的种质资源与其他产地的桔梗种质资源进行高效鉴别,为鉴别不同产地桔梗的研究奠定了基础。桔梗15个居群的遗传多样性丰富,但各单倍型之间遗传距离较小,桔梗种内亲缘关系较近,且主流单倍型Hap1和Hap4之间亲缘关系相对较近。因此本研究结合种内叶绿体基因组比较分析、扩增效率、鉴别能力,发现*trnG-UCC*和*ndhG\_ndhF*为桔梗不同产地种质资源鉴定的特异性DNA片段,发现部分地区独有的桔梗种质资源,分析发现桔梗种内遗传多样性较丰富,为桔梗品种产地鉴别和选育奠定基础。

**作者贡献:**王馨负责实验、数据分析和论文撰写;石玥和满金辉参与实验;黄钰莹和张晓芹参与论文数据分析;安克露、何高洁、刘子齐、关范圆、郑语嫣负责样品收集;王晓晖和魏胜利负责设计论文的实验思路,指导学生实验,论文撰写及修改。

**利益冲突:**无任何利益冲突。

## References

- [1] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [M]. Beijing: China Medical Science Press, 2020, 1: 289.
- [2] Li Y, Wu Y, Xia Q, et al. *Platycodon grandiflorus* enhances the effect of DDP against lung cancer by down regulating PI3K/Akt signaling pathway [J]. Biomed Pharmacother, 2019, 120: 109496.
- [3] Qu Y, Zhou L, Wang C. Effects of platycodin D on IL-1 $\beta$ -induced inflammatory response in human osteoarthritis chondrocytes [J]. Int Immunopharmacol, 2016, 40: 474-479.
- [4] Wang C, Cheng G, Yang S, et al. Protective effects of *Platycodon grandiflorus* polysaccharides against apoptosis induced by carbonyl cyanide 3-chlorophenylhydrazone in 3D4/21 cells [J]. Int J Biol Macromol, 2019, 141: 1220-1227.
- [5] Lin YC, Lin CH, Yao HT, et al. *Platycodon grandiflorum* (PG) reverses angiotensin II-induced apoptosis by repressing IGF-IIR expression [J]. J Ethnopharmacol, 2017, 205: 41-50.
- [6] Yang XY, Liu C, Zeng XF, et al. Sequence structure and phylogenetic analysis of the chloroplast genomes of *Alangium chinense* (Lour.) Harms and its different subspecies [J]. Acta Pharm Sin (药学报), 2022, 57: 3229-3239.
- [7] Li ZW, Qiu Q, Lang JQ, et al. Sequence analysis of complete chloroplast genome of *Dendrobium heterocarpum* and *Dendrobium trigonopus* [J]. Chin Tradit Herb Drugs (中草药), 2022, 53: 5159-5169.
- [8] Hong CP, Park J, Lee Y, et al. *accD* nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae [J]. BMC Genomics, 2017, 18: 607.
- [9] Knox EB, Li C. The East Asian origin of the giant lobelias [J]. Am J Bot, 2017, 104: 924-938.
- [10] Cheon KS, Kim KA, Yoo KO. The complete chloroplast genome sequences of three *Adenophora* species and comparative analysis with Campanuloid species (Campanulaceae) [J]. PLoS One, 2017, 12: e0183652.
- [11] Hebert PD, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes [J]. Proc Biol Sci, 2003, 270: 313-321.
- [12] Zhang XQ, Liu CS, Yan XL, et al. Sequence analysis and identification of a chloroplast *matK* gene in *Rhei Radix et Rhizoma* from different botanical origins [J]. Acta Pharm Sin (药学报), 2013, 48: 1722-1728.
- [13] Li RJ, Wu LW, Xin TY, et al. Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of *Rhei Radix et Rhizoma* [J]. Acta Pharm Sin (药学报), 2022, 57: 1495-1505.
- [14] Jiang D. Study on Genetic and Chemical Basis of the Trueness of *Scutellaria* (黄芩道地性的遗传和化学物质基础研究) [D]. Beijing: Beijing University of Chinese Medicine, 2018.
- [15] Niu XL, Wei NN, Shao M, et al. Analysis of the *rbcL* sequence of *Cremastrae Pseudobulbus Pleiones Pseudobulbus* from Miao Yao in Guizhou [J]. Mol Plant Breeding (分子植物育种), 2019, 17: 4717-4721.
- [16] Zhang HY, Yuan LL, Rao QH, et al. Sequence analysis of *psbA\_trnH* gene in chloroplasts of *Polygonum multiflorum* from different habitats [J]. Chin Tradit Herb Drugs (中草药), 2018, 49: 1146-1149.
- [17] Zhou Z, Hong D, Niu Y, et al. Phylogenetic and biogeographic analyses of the Sino-Himalayan endemic genus *Cyananthus* (Campanulaceae) and implications for the evolution of its sexual system [J]. Mol Phylogenet Evol, 2013, 68: 482-497.
- [18] Zhao SY, Liu R, Feng H, et al. DNA barcode molecular identification of *Platycodon grandiflorum* and its easily mixed products [J]. J Hebei Univ Nat Sci Ed (河北大学学报自然科学版), 2021, 41: 60-67.
- [19] Wu B, Li YB, Rao JB, et al. Study on genetic diversity of *Platy-*

- codon grandiflorum* based on *ITS2* barcode [J]. China J Chin Mater Med (中国中药杂志), 2015, 40: 1075-1078.
- [20] Nie CP, Li YY, Chen J, et al. Discussion about the phylogeny of Campanulaceae based on the partial sequence of *rbcL* gene [J]. Genomics Appl Biol (基因组学与应用生物学), 2017, 36: 3091-3095.
- [21] Yang JP, Zhu ZL, Fan YJ, et al. Comparative plastomic analysis of three Bulbophyllum medicinal plants and its significance in species identification [J]. Acta Pharm Sin (药学报), 2020, 55: 2736-2745.
- [22] Hu HS, Zhang DQ. DNA super-barcoding of several medicinal species in *Gentiana* from Yunnan province [J]. China J Chin Mater Med (中国中药杂志), 2021, 46: 5260-5269.
- [23] Dong BR, Zhao ZL, Ni LH, et al. Molecular markers based upon whole chloroplast genomes and identifying alpine *Gentiana waltonii* and *G. lhassica* (Gentianaceae) [J]. Acta Pharm Sin (药学报), 2021, 56: 2584-2591.
- [24] Yin GY, Yuan L, Wang X, et al. Screening of specific DNA barcode, identification of germplasm resources, and analysis of genetic diversity of *Atractylodes chinensis* [J]. Acta Pharm Sin (药学报), 2023, 58: 1693-1704.
- [25] Zhang ZF, Zhao ZZ, Wang X, et al. Specific DNA barcodes, germplasm resources, and genetic diversity of *Acanthopanax senticosus* [J]. China J Chin Mater Med (中国中药杂志), 2023, 48: 1229-1237.
- [26] Cui YX. Analysis of Chloroplast Genome Structure of Prochloroplast Plants of *Amomi Fructus*, *Lycii Fructus*, *Crataegi Fructus* and *Zingiberis Rhizoma Recens* (药食两用药材砂仁、枸杞、山楂和姜基原植物叶绿体基因组结构解析) [D]. Beijing: Peking Union Medical College, 2020.