

龙胆科叶绿体基因组结构特征及其药用植物鉴定意义

董博然¹, 赵志礼^{1*}, 倪梁红¹, 吴靳荣¹, 丹珍卓嘎²

1. 上海中医药大学中药学院, 上海 201203

2. 西藏藏医药大学, 西藏 拉萨 850000

摘要: 我国龙胆科植物大多分布于青藏高原, 具有重要的药用价值。龙胆科植物基因组长度 137~154 kb; 常具有蛋白质编码基因 67~80 个、tRNA 基因 30 个、rRNA 基因 4 个; 系统发育树的拓扑结构与传统分类学观点基本一致; 叶绿体基因组序列具有属、组间的分辨率。基于目前文献报道的龙胆科 5 属 (龙胆属 *Gentiana*、獐牙菜属 *Swertia*、花锚属 *Halenia*、睡菜属 *Menyanthes* 及苔菜属 *Nymphoides*) 27 个分类群的叶绿体全基因组序列, 应用相关软件比较分析其结构特征, 以夹竹桃科 (Apocynaceae) 植物 *Pentalinon luteum* 叶绿体全基因组序列为外类群, 以贝叶斯法构建系统发育树, 深入分析系统发育关系, 评价其物种鉴定意义, 可为该科物种资源保护、遗传背景分析及中藏药生药学研究等提供参考资料。

关键词: 龙胆科; 龙胆属; 獐牙菜属; 花锚属; 睡菜属; 苔菜属; 药用植物; 叶绿体基因组; 系统学意义

中图分类号: R282.1 **文献标志码:** A **文章编号:** 0253-2670(2020)06-1641-09

DOI: 10.7501/j.issn.0253-2670.2020.06.033

Comparative analysis of complete chloroplast genome sequences within Gentianaceae and significance of identifying species

DONG Bo-ran¹, ZHAO Zhi-li¹, NI Liang-hong¹, WU Jin-rong¹, DANZHEN Zhuo-ga²

1. School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

2. Tibetan Traditional Medical College, Lhasa 850000, China

Abstract: The family Gentianaceae are found mostly on the Qinghai-Tibet Plateau in China, which have important medicinal properties. Based on 27 published complete chloroplast genome sequences from *Gentiana*, *Swertia*, *Halenia*, *Menyanthes*, and *Nymphoides* of Gentianaceae, the chloroplast genome structure was analyzed. The cp genome sizes of 27 taxa range from 137 to 154 kb, and they contain 101—114 unique genes, including 67—80 protein-coding genes, 30 tRNA genes and four rRNA genes. Also, a Bayesian phylogenetic tree was constructed based on the 27 cp genome sequences with *Pentalinon luteum* (Apocynaceae) as the outgroup. The tree was topologically consistent with the treatments of traditional taxonomy, and the cp genome sequences have genus- or section-level resolution. In addition, we reviewed the significance of species identification within the family. These cp genome sequences could provide basic data for the endangered species conservation, the genetic analysis and pharmacognostic researches of herbs from Gentianaceae.

Key words: Gentianaceae; *Gentiana* (Tourn.) L.; *Swertia* L.; *Halenia* Borkh.; *Menyanthes* (Tourn.) L.; *Nymphoides* Seguiet; medicinal plants; chloroplast genome; phylogenetic significance

龙胆科 (Gentianaceae) 隶属于双子叶植物纲捩花目 (Contortae)^[1], 约 80 属 700 种, 广泛分布于世界各地。其中第一大属为龙胆属 *Gentiana* (Tourn.) L., 约 400 种; 第二大属为獐牙菜属 *Swertia* L., 约 170 种, 两者种数占全科近 81%。我国龙胆科植物

有 22 属 400 多种, 大多分布于西南高山地区。其中龙胆属约 250 种, 隶属于 11 组 (section) 23 系 (series), 獐牙菜属约 80 种, 隶属于 7 组 7 系^[2]。

龙胆科植物具有重要的药用价值, 如常用中藏药材秦艽、龙胆、解吉 (ཉེ་མེ་ལྷོ་མེ) [3]、邦见 (པཎ་ཇེ་མེ) 及

收稿日期: 2019-12-04

基金项目: 藏医药区域协同创新中心科研项目 (2018XTCX005); 上海市卫生健康委员会中医传承和科技创新项目 (ZYCC2019012); 上海中医药大学预算内项目 (A1-GY01020902)

作者简介: 董博然 (1994—), 女, 硕士研究生, 研究方向为中药资源与品种鉴定。E-mail: dongborann@163.com

*通信作者 赵志礼, 教授, 博士生导师, 主要从事中药资源与品种鉴定工作。E-mail: zhilzhao@sohu.com

冈嘎穹 (ཁོང་ཁྱེད་) 来源于龙胆属; 当药、青叶胆、巴莴色保 (དཔལ་ལྷོ་མེ་འཕྲོ་) [4] 及桑蒂 (མཚོ་ཏིག་) 等来源于獐牙菜属; 还包括来源于扁蕾属 *Gentianopsis* Ma、花锚属 *Halenia* Borkh. 及肋柱花属 *Lomatogonium* A. Br. 等的藏药吉合蒂 (ཇི་ཁ་ཏིག་) [5] 等。

我国青藏高原是世界龙胆科植物分布中心之一, 深入探讨其各类群的生物学特性、物种形成机制及系统发育关系, 可为持续利用有限的高山药用植物资源, 进一步整理中藏药材品种, 构建快速、准确的物种 DNA 条形码等具重要的科学意义。当务之急是深入开展种质资源调查、物种遗传多样性分析等基础性工作。

叶绿体是质体的一种, 普遍存在于植物中, 是细胞内具有自主遗传信息的重要细胞器, 以母性遗传为主。自 1986 年烟草 *Nicotiana tabacum* L. [6] 和地钱 *Marchantia polymorpha* L. [7] 的叶绿体全基因组序列首次发表以来, 叶绿体的基因结构特征及其系统学意义也逐渐被人们熟知。

被子植物叶绿体基因组结构常为双链环状, 由 4 部分组成: 大单拷贝区 (LSC)、小单拷贝区 (SSC)、2 个反向重复区 (IRa、IRb 编码相同, 方向相反, 是叶绿体基因组进化过程中延展或缩小的区域, 也是叶绿体基因组大小差异的主要原因) [8]。叶绿体基因组的大小、结构和基因种类一般较为保守 [9-10], 基因组长度在 115~165 kb [11]。叶绿体基因组共编码 100~120 个基因, 其中蛋白质编码基因 70~88 个, tRNA 基因 30~32 个, rRNA 基因 4 个 [12-13]。同时, 由于叶绿体基因组结构、长度和基因种类的保守性 [14-15], 其编码区和非编码区进化速率与模式具有物种鉴别意义 [16], 现已广泛应用于遗传多样性分析, 物种 DNA 分子鉴定, 分子系统发育研究及药用植物遗传转化、分子育种等叶绿体基因工程方面 [17-18]。

本文基于现有文献中的龙胆科 5 属 (龙胆属、獐牙菜属、花锚属、睡菜属 *Menyanthes* (Tourn.) L. 及苳菜属 *Nymphaeoides* Seguiet) 27 个分类群的叶绿体全基因组序列, 比较分析其结构特征, 梳理系统发育关系, 评价其物种鉴定意义, 以期为该科物种资源保护、遗传背景分析及中藏药生药学研究等提供参考资料。

1 龙胆科植物叶绿体基因组基本结构

根据目前已报道的龙胆科植物叶绿体全基因组序列, 龙胆属包括秦艽组 (Sect. *Cruciata*)、多枝组

(Sect. *Monopodiae*) 2 组 18 个分类群, 獐牙菜属 3 种, 花锚属 1 种, 睡菜属 1 种, 苳菜属 4 种, 共 5 属 27 个分类群。以夹竹桃科 (Apocynaceae) 植物 *Pentalinon luteum* (L.) B. F. Hansen & Wunderlin 叶绿体全基因组序列为外类群, 进行叶绿体基因结构的比较分析 (表 1)。

龙胆科 27 个分类群的叶绿体基因组均由 LSC、SSC、IRa、IRb 4 个部分组成。对该科叶绿体基因组结构进行初步分析的同时, 对该科最大的 2 个属龙胆属、獐牙菜属植物做进一步 GC 含量、重复序列和分布位置的分析。

1.1 基因组长度

被子植物叶绿体基因组大小一般为 115~165 kb, 龙胆科植物叶绿体基因组长度在 137~154 kb, 与被子植物的长度特征相吻合 [12]。獐牙菜属 3 种植物的基因组长度在 150 kb 左右, 比龙胆属长 3~16 kb; 苳菜属约为 152 kb, 花锚属约为 153 kb, 睡菜属 (单种) 为 154 kb, 是龙胆科植物目前叶绿体基因组序列最长的物种。龙胆属秦艽组植物叶绿体基因组长度在 148 652~149 916 bp [9], 组内基因组长度差异小, 大叶秦艽的序列长度最长, 与同组物种存在 1 kb 左右的差异, 从叶绿体各区间比较可以看出, 这种差异主要出现在 LSC 区和 IR 区; 多枝组植物叶绿体基因组长度在 137 278~147 156 bp, 大部分植物比秦艽组植物少约 10 kb, 这主要与多枝组物种存在 *ndh* 基因部分序列的缺失有关。特殊序列长度出现在短柄龙胆, 短柄龙胆比其他多枝组物种长度大约多 10 kb, 其中 LSC 区多 1~2 kb, SSC 区差异更为明显, 约 5 kb, IR 区有 1~2 kb 的差异。獐牙菜属中轮叶獐牙菜和川西獐牙菜的序列长度存在约 1.8 kb 的差异, 毛萼獐牙菜的基因序列最短, 导致这个差异的原因出现在 LSC 区。

1.2 叶绿体基因数量

龙胆科植物叶绿体基因组常具有蛋白质编码基因 67~80 个, tRNA 基因 30 个, rRNA 基因 4 个。

不同类群的基因缺失或假基因情况: 龙胆属多枝组物种普遍存在 *ndh* 基因缺失, 包括 *ndhA*、*ndhC*、*ndhG*、*ndhH*、*ndhI*、*ndhJ*、*ndhK*; 苳菜属 4 个物种全部缺失 *rpl2* 基因, 獐牙菜属植物中仅毛萼獐牙菜缺失 *rpl33* 基因。由于叶绿体基因组边界结构原因, 龙胆科物种普遍存在 *ψrps19*、*ψycf1* 2 个假基因; 龙胆属秦艽组和獐牙菜属植物还存在 2 个假基因 *ψrps16*、*ψinfA*; 多枝组植物存在 *ψndhB*、*ψndhD*、

表 1 龙胆科 27 个分类群叶绿体基因组序列结构信息
Table 1 Chloroplast genome sequence structure of 27 taxa in Gentianaceae

属	组	分类群	登录号	LSC/bp	SSC/bp	IR/bp	总长度/bp
龙胆属	秦艽组	西藏秦艽 <i>G. tibetica</i> King ex Hook. f.	KU975374	81 163	17 070	25 266	148 765
		麻花艽 <i>G. straminea</i> Maxim.	KJ657732	81 240	17 085	25 333	148 991
		粗壮秦艽 <i>G. robusta</i> King ex Hook. f.	KT159969	81 164	17 081	25 333	148 911
		粗茎秦艽 <i>G. crassicaulis</i> Duthie ex Burk.	KJ676538	81 164	17 070	25 271	148 776
		黄管秦艽 <i>G. officinalis</i> H. Smith	MH261261	81 119	17 088	25 336	148 879
		管花秦艽 <i>G. siphonantha</i> Maxim. ex Kusnez.	MH261260	81 121	17 113	25 337	148 908
		小秦艽 <i>G. dahurica</i> Fisch.	MH261259	81 154	17 093	25 278	148 803
		长梗秦艽 <i>G. waltonii</i> Burk.	MK780032	81 068	17 029	25 304	148 705
		全萼秦艽 <i>G. lhassica</i> Burk.	MK790135	80 997	17 051	25 302	148 652
		大叶秦艽 <i>G. macrophylla</i> Pall.	KY856959	82 911	17 095	24 955	149 916
	多枝组	蓝玉簪龙胆 <i>G. veitchiorum</i> Hemsl.	MG192310	77 932	11 807	23 864	137 467
		短柄龙胆 <i>G. stipitata</i> Edgew.	MG192309	79 712	16 986	25 229	147 156
		华丽龙胆 <i>G. ornata</i> (Wallich ex G. Don) Grisebach	MG192308	77 816	11 353	24 108	137 385
		山景龙胆 <i>G. oreodoxa</i> H. Smith	MG192307	77 908	11 765	23 865	137 403
		倒锥花龙胆 <i>G. obconica</i> T. N. Ho	MG192306	77 754	11 794	23 865	137 278
		六叶龙胆 <i>G. hexaphylla</i> Maxim. ex Kusnez.	MG192305	77 922	11 771	23 865	137 423
		天蓝龙胆 <i>G. caelestis</i> (Marq.) H. Smith	MG192304	77 870	11 548	24 113	137 644
		线叶龙胆 <i>G. lawrencei</i> Burkill var. <i>farreri</i> (I. B. Balfour) T. N. Ho	KX096882	78 082	11 364	24 652	138 750
		毛萼獐牙菜 <i>S. hispidicalyx</i> Burk.	MH321887	80 727	17 911	25 425	149 488
		轮叶獐牙菜 <i>S. verticillifolia</i> T. N. Ho et S. W. Liu	MF795137	82 623	18 335	25 362	151 682
獐牙菜属		川西獐牙菜 <i>S. mussotii</i> Franch.	KU641021	83 567	18 342	25 761	153 431
花锚属		花锚 <i>H. corniculata</i> (L.) Cornaz	MK606372	83 252	18 372	25 787	153 198
睡菜属		睡菜 <i>M. trifoliata</i> L.	MH201540	86 609	18 068	24 818	154 313
苔菜属		<i>N. coronata</i> (Dunn) Chun ex Y.D. Zhou & G.W. Hu	MH201539	83 986	17 872	25 064	151 986
		<i>N. crenata</i> Kuntze	MH201538	84 242	17 811	25 086	152 225
		刺种苔菜 <i>N. hydrophylla</i> (Loureiro) Kuntze	MH201537	84 207	17 796	25 054	152 111
		<i>N. simulans</i> Aston	MH201534	84 114	17 821	25 048	152 031
金香藤属 (夹竹桃科)		<i>P. luteum</i>	KJ953909	84 614	17 907	25 766	154 053

$\psi ndhE$ 、 $\psi ndhF$ 假基因；花锚属植物花锚和獐牙菜属川西獐牙菜中 $\psi accD$ 是假基因，不参与编码蛋白质的过程。

1.3 编码蛋白基因长度差异

对龙胆科叶绿体编码蛋白基因进行分析，结果显示，共有 31 个基因的序列长度在不同物种中完全

一致。其他基因在亲缘关系较近的物种间长度基本一致,但存在属间差异。不同物种间长度变异最大的 *ycf1* 基因,该基因长度 5 400 bp 左右,各物种间最大变异长度为 165 bp。*ycf15* 基因变化范围为 129 (毛萼獐牙菜)~237 bp (粗茎秦艽),在不同物种中存在明显的长度差异。

从 9 个含内含子的编码基因来看,大部分基因编码区长度相对稳定,但在不同物种中内含子长度不尽相同,亲缘关系相近的物种间变化较小。*rps12*、*petB*、*petD*、*rpl16*、*ndhB* 5 个基因在部分物种中缺失内含子区间,并且编码区长度相对其他物种亦较短;线叶龙胆在 *rps12*、*petB*、*petD*、*rpl16* 4 个基因中内含子全部缺失;*ndhB* 基因在龙胆属多枝组多个物种中为假基因或缺失部分序列,而其他龙胆科植物该基因编码 511 个氨基酸,差异较为明显。

1.4 SC 与 IR 边界分析

龙胆科植物基因组 LSC、SSC、IRa、IRb 4 个部分的接口位置,基因类型基本一致。LSC-IRb 均为 *rps19*,且长度一致 (279 bp)。但该基因在 LSC 与 IRb 的分布比例存在组间或属间差异。IRb-SSC 边界区均存在 ψ *ycf1* 假基因,其中龙胆属秦艽组、花锚属、獐牙菜属与龙胆属多枝组短柄龙胆的 *ndhF* 基因也位于该边界区。 ψ *ycf1* 假基因的长度具有一定的属或组间区分。SSC-IRa 均为 *ycf1*,且大部分 *ycf1* 序列位于 SSC。IRa-LSC 边界均位于 ψ *rps19* 和 *tmH-GUG* 基因的交界位置,偶可见 2、6 bp 或 735 bp 的序列间隔。 ψ *rps19* 基因长度为 4~199 bp, *tmH-GUG* 基因长度为 74、75 bp,在物种间相对稳定 (图 1)。

2 龙胆属及獐牙菜属叶绿体基因组结构特征

2.1 GC 含量

用 EditSeq 7.1.0 软件对龙胆属、獐牙菜属植物叶绿体基因组序列的 GC 含量进行计算,结果显示,獐牙菜属 GC 含量最高,等于或大于 38.14%,龙胆属多枝组居中,在 37.93%~38.06%,而龙胆属秦艽组最低,在 37.67%~37.73%。

2.2 重复序列分析

利用在线分析软件 REputer^[10] (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) 对龙胆属、獐牙菜属 21 种植物叶绿体全基因组进行重复序列分析,参数设置:重复序列长度 $n \geq 30$,序列一致性 90%,Hamming distance 为 3,得到正向重复序列 (F)、反向重复序列 (R)、反向互补序列 (P)、正向互补序列 (C) 的数量和分布位置对比图 (图 2 和 3)。

由图 2 可知,龙胆属重复序列个数在 20~41,其中多枝组和獐牙菜属数量相近,秦艽组重复序列数最多,在 24~41,物种内 F、P 重复数约为 1:1,且除了西藏秦艽、粗茎秦艽及大叶秦艽外,其他秦艽组 7 种植物都只有 F、P 重复。西藏秦艽和粗茎秦艽存在 1 个 C 重复,说明 2 种植物的叶绿体基因结构存在一定的相似性。值得注意的是大叶秦艽的重复序列数为 41,明显高于其他秦艽组植物,且存在 3 个 R 重复。多枝组 8 个物种的重复序列数整体少于秦艽组植物,在 20~23, F、P 重复数也近似 1:1,只有蓝玉簪龙胆存在 1 个 R 重复。獐牙菜属 3 个物种 P 重复数都为 8; F 重复数彼此差异不大,重复序列的类型和数量相对稳定。

物种的 F、R、P、C 重复序列在叶绿体基因组中的分布主要包括 4 个位置:基因间区 (intergenic spacer, IGS)、编码序列 (coding sequence, CDS)、内含子 (intron) 及 CDS-IGS (位于编码序列、基因间区和内含子)。由图 3 可知,秦艽组 10 种与獐牙菜属 3 种的分布位置比例更为接近。獐牙菜属大部分的重复序列位于基因间区和内含子,三者各个位置重复序列分布的比例大致相同。龙胆属重复序列主要位于基因间区和编码区,多枝组植物中除了短柄龙胆之外,其他物种在基因间区分布比例都在 50%以上。除獐牙菜属 3 种植物外,重复序列分布在 CDS-IGS 区的比例大多在 10%以下,大叶秦艽最低;同时,多枝组该区分布比例都高于秦艽组植物。值得注意的是,西藏秦艽与粗茎秦艽、麻花艽与粗壮秦艽、小秦艽与黄管秦艽、华丽龙胆与倒锥花龙胆各部分分布比例完全一致,这可能与它们结构的相似性有关。除此之外,大叶秦艽、蓝玉簪龙胆、短柄龙胆的重复序列分布位置比例与同组其他物种有明显不同,这可能与它们的结构特征有关。

3 龙胆科系统进化树的构建与分析

利用 MrBayes 3.2.7^[19] 软件,以夹竹桃科植物 *P. luteum* 为外类群,构建基于龙胆科 5 属 27 个分类群叶绿体全基因组的贝叶斯系统发育树 (图 4),参数设置: $N_{st}=6$, Rates=gamma, Samplefreq=10。系统树中,外类群夹竹桃科及龙胆科的睡菜属及苔菜属首先被分出;其次,分为 2 大支,龙胆属聚为一支,另一支由獐牙菜属与花锚属组成。龙胆属下的分组情况与传统分类观点基本一致。值得注意的是,龙胆属多枝组的短柄龙胆与秦艽组植物聚为一支,在亲缘关系上似与秦艽组更近,有待进一步研究。

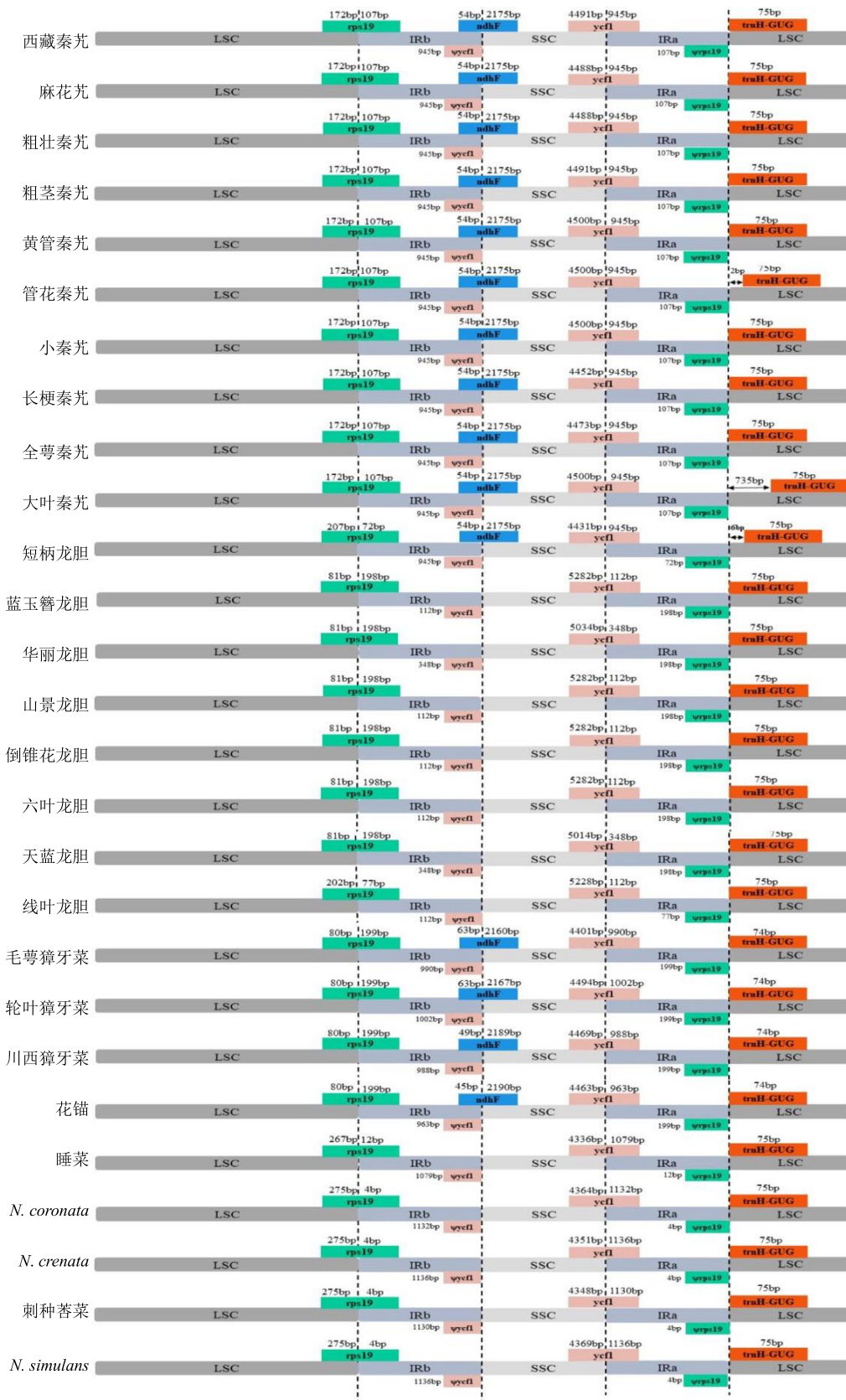


图 1 龙胆科 27 个分类群的叶绿体基因组 LSC、SSC、IRa 及 IRb 边界分析

Fig. 1 Comparison of borders of LSC, SSC, and IR regions in 27 Gentianaceae chloroplast genomes

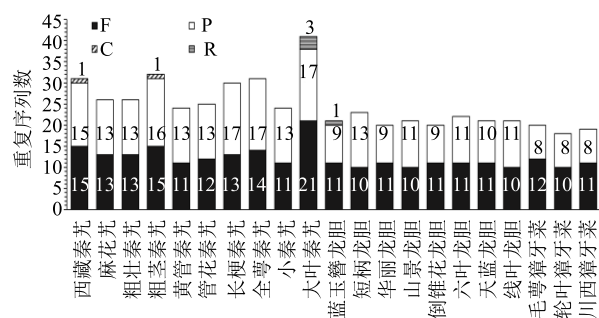


图2 龙胆属、獐牙菜属叶绿体基因组中F、R、P、C 4种重复序列数量对比

Fig. 2 Numbers of F, R, P, and C repeats of chloroplast genomes within *Gentiana* and *Swertia*

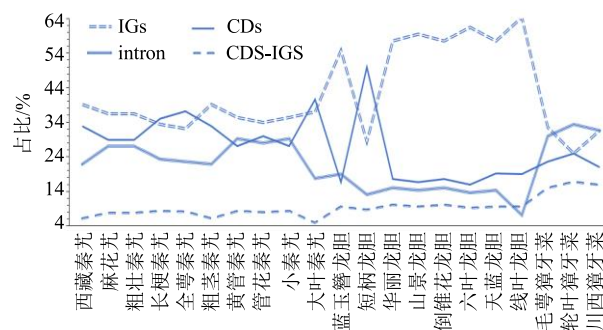
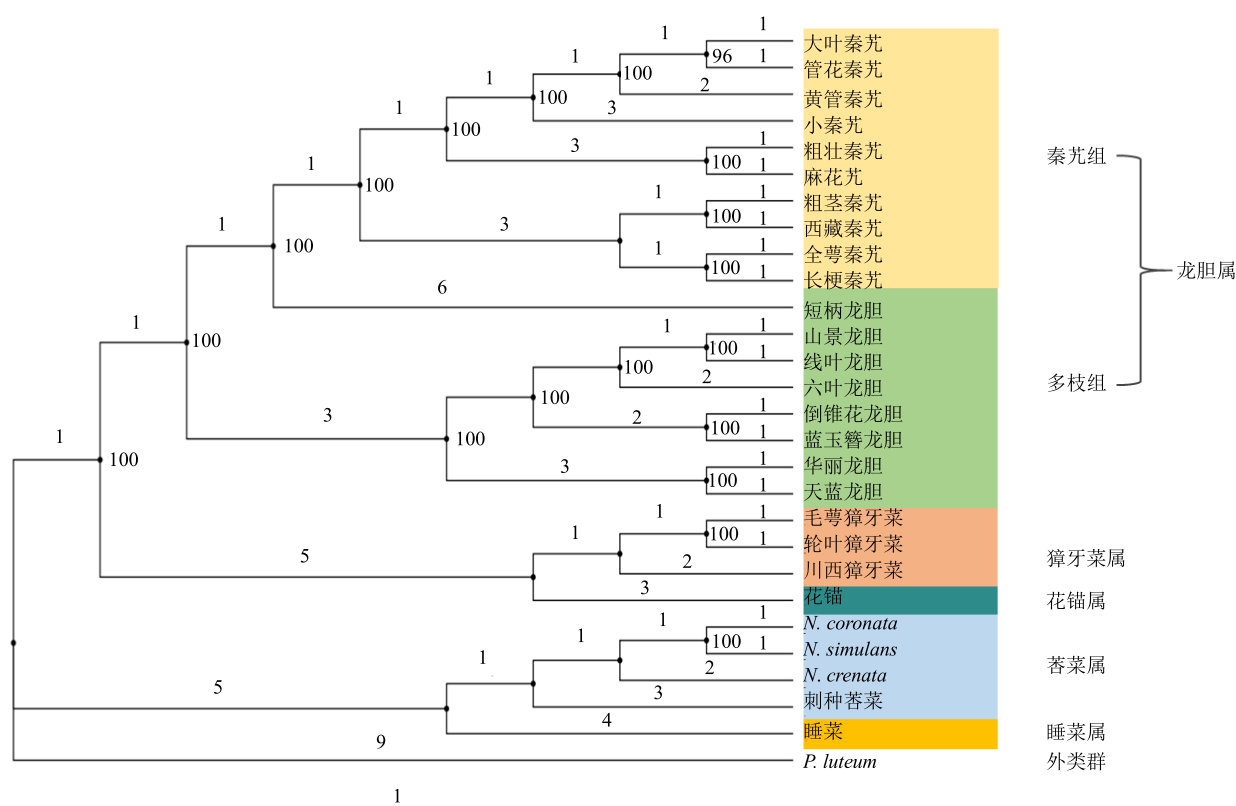


图3 龙胆属、獐牙菜属植物叶绿体基因组重复序列分布

Fig. 3 Distribution location of chloroplast repeat sequences within *Gentiana* and *Swertia*



节点后标记值为后验概率

The numbers after nodes indicate Bayes posterior probabilities

图 4 基于龙胆科叶绿体全基因组序列构建的贝叶斯树

Fig. 4 Bayesian phylogenetic tree based on complete genome sequences within Gentianaceae

4 结语与展望

4.1 基因组长度变化与基因缺失

与被子植物茄科 (Solanaceae)、蔷薇科 (Rosaceae) [20-23] 等相比, 龙胆科植物的叶绿体基因长度较短, 有 5~23 kb 的差异, 主要在 LSC 区, 这与大单拷贝区变异程度高是密不可分的 [24]。

就现有序列资料来看,龙胆属基因组长度最短,除大叶秦艽(149 916 bp)外,均在 149 kb 以下;

其中, 多枝组均低于 148 kb。基因组序列最长的为睡菜属, 长达 154 kb。基因组序列长度在龙胆科是否具有属或组一级分类学意义, 值得进一步探讨。

在龙胆属秦艽组和多枝组中, 基因组长度出现明显差异的原因在于多枝组植物 *ndh* 基因的缺失; 除短柄龙胆外, 多枝组植物缺失了大部分 *ndh* 基因的编码序列 (*ndhJ*、*ndhK*、*ndhC*、*ndhG*、*ndhI*、*ndhA* 和部分 *ndhE*、*ndhH* 及 *ndhB* 第 2 个外显子) [22,25],

缺失的 *ndh* 基因大部分位于 LSC、SSC 区。*ndh* 基因编码类囊体 NADH 复合物,与光合作用密切相关^[26-27],同时,被子植物中 *ndh* 基因缺失的情况普遍存在^[28-31],可以为研究物种自然选择和系统进化提供参考资料。在构建系统进化树时,基于 46 个编码区基因所构建的系统树显示短柄龙胆与秦艽组关系更为密切^[25],这与经典的植物分类学观点相违背,出现这种现象的原因也是值得关注的。

4.2 假基因现象

龙胆科植物假基因个数一般为 4 个,分别是 $\psi rps16$ 、 $\psi rps19$ 、 $\psi infA$ 及 $\psi ycf1$, $\psi infA$ 假基因的出现是由于物种进化过程中发生了转移或损失^[32-33]; $\psi rps19$ 和 $\psi ycf1$ 假基因的出现是由于位于叶绿体基因区域边界位置,产生了边界效应^[34]; $\psi rps16$ 假基因缺失第 2 个外显子,首次在麻花苳^[35]中发现,也是首次在菊分支 (APG IV) 非寄生物种中发现 $\psi rps16$ 假基因。此后在多个龙胆科植物叶绿体基因组研究中被发现,各物种间有着相同的结构,因此推测龙胆属植物 $\psi rps16$ 假基因现象可能是普遍存在的^[25,35-37]。川西獐牙菜除了以上 4 个假基因外,还存在 $\psi accD$ 假基因,这是 *accD* 在此物种中存在内部终止密码子的原因,这种现象也并非罕见,在桔梗科 (Campanulaceae) 和木犀科 (Oleaceae) 植物中都有报道^[38-39]。另外,龙胆属多枝组物种还存在 $\psi ndhB$ 、 $\psi ndhD$ 、 $\psi ndhE$ 、 $\psi ndhF$ 假基因现象。

4.3 龙胆科叶绿体基因组序列的分类学鉴定意义

基于龙胆科 5 属 27 个分类群叶绿体全基因组序列的结构分析,以及构建的贝叶斯系统发育树来看,叶绿体基因组显示了重要的分类学鉴定意义。

①科一级分辨率:龙胆科所有分类群序列均与外类群夹竹桃科植物 *P. luteum* 明显区分。②科内属一级分辨率:龙胆科 5 属间均能区分。睡菜属及苔菜属彼此亲缘关系更为接近,且与科内其他属关系较远,这在基因组结构特征上亦有证据支持。有学者将睡菜属与苔菜属从龙胆科中分出,与其他一些类群共同组成睡菜科 (Menyanthaceae)^[40],本文结果与该处理相吻合。③龙胆属秦艽组的鉴定:就目前获得的秦艽组 10 个物种的基因组资料来看,均可与属内多枝组相区别。但龙胆属多枝组短柄龙胆显示了与秦艽组各类群更为近缘的系统关系,有待于深入探究。④其他片段的鉴定意义:*ycf15* 基因位于叶绿体基因组 IR 区,长度一般为 250~550 bp,高度保守,就目前研究工作来看,其独特的序列结

构及其显现的丰富多样的遗传信息不仅具有一定的物种鉴定意义,而且在植物系统进化分析研究中有较大的潜力。同时,在构建龙胆科龙胆属药用植物 DNA 条形码工作中,*ycf15* 基因显示出一定的物种鉴定意义^[41]。

龙胆属中 *atpB-rbcL*、*trnL-trnF*、*matK*^[42-43]、*trnH-psbA*、*trnK-trnQ*、*ycf3-trnS* 和 *rpl32-trnL*^[25] 基因鉴别片段以及物种进化过程中阳性选择的基因片段 *accD*、*clpP*、*matK*^[44] 和 *ycf1*^[38] 为龙胆属物种的鉴别和进化关系提供了较多的数据支持;插入、缺失位点的标记也为其鉴定提供了依据。*accD*、*clpP*、*matK*、*ycf1* 4 个基因在物种进化中表现出积极选择的趋势,它们的进化形式、基因选择的原理、对环境的适应性都是值得探讨的问题。

相较于龙胆属序列资料,獐牙菜属目前测序数量较少,北温带獐牙菜 *S. perennis* L. 叶绿体基因 *trnL-trnF* 区和 *trnH-psbA* 间隔区存在的 36 个多态性位点,可以用于该植物不同地区基因进化的研究依据^[45],这也可作为獐牙菜属的系统进化研究提供思路。

龙胆科植物物种多样,且具有重要的药用价值。现有叶绿体基因组序列资料分析,已显示出独特的序列结构及其丰富多样的遗传信息。但就整个科的种类而言,仅测定 27 个基因组序列远远不够。下一步应补充该科其他分类群的基因数据,更为深入探讨该科植物系统发育关系,以期为该科物种资源保护及中藏药生药学研究等提供科学依据。

参考文献

- [1] 张美珍,邱莲卿. 中国植物志 (61 卷) [M]. 北京: 科学出版社, 1992.
- [2] 何延农. 中国植物志 (62 卷) [M]. 北京: 科学出版社, 1988.
- [3] Zhao Z L, Gaawe D, Wang Z T. Identification of medicinal plants used as Tibetan traditional medicine Jie-Ji [J]. *J Ethnopharmacol*, 2010, 132(1): 122-126.
- [4] 魏似婕,赵志礼,倪梁红,等. 藏药“巴葛色保”基原鉴定及其叶绿体全基因组序列分析 [J]. 药学报, 2018, 53(6): 1009-1015.
- [5] 杨永昌. 藏药志 [M]. 西宁: 青海人民出版社, 1991.
- [6] Shinozaki K, Ohme M, Tanaka M, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression [J]. *EMBO J*, 1986, 5(9): 2043-2049.
- [7] Ohyama K, Fukuzawa H, Kohchi T, et al. Chloroplast gene organization deduced from complete sequence of

- liverwort *Marchantia polymorpha* chloroplast DNA [J]. *Nature*, 1986, 322(6079): 572-574.
- [8] Palmer J, Stein D. Conservation of chloroplast genome structure among vascular plants [J]. *Curr Genet*, 1986, 10(11): 823-833.
- [9] 王玲, 董文攀, 周世良. 被子植物叶绿体基因组的结构变异研究进展 [J]. *西北植物学报*, 2012, 32(6): 1282-1288.
- [10] Kurtz S, Choudhuri J V, Ohlebusch E, *et al.* REPuter: the manifold applications of repeat analysis on a genomic scale [J]. *Nucl Acid Res*, 2001, 29(22): 4633-4642.
- [11] Hu Y H, Woeste K E, Zhao P. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny [J]. *Front Plant Sci*, 2017, doi: 10.3389/fpls.2016.01955.
- [12] Li X, Li Y, Zang M, *et al.* Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissia* [J]. *Int J Mol Sci*, 2018, 19(8): 1-17.
- [13] Daniell H, Lin C S, Yu M, *et al.* Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 1-29.
- [14] 朱婷婷, 张磊, 陈万生, 等. 1342 个植物叶绿体基因组分析 [J]. *基因组学与应用生物学*, 2017, 36(10): 4323-4333.
- [15] 王博, 高磊, 苏应娟, 等. 基于叶绿体基因组全序列分析真叶植物叶绿体基因的适应性进化 [J]. *中山大学学报: 自然科学版*, 2012, 51(3): 108-113.
- [16] Clegg M T, Gaut B S, Learn G H, *et al.* Rates and patterns of chloroplast DNA evolution [J]. *P Natl Acad Sci USA*, 1994, 91(15): 6795-6801.
- [17] 张韵洁, 李德铎. 叶绿体系统发育基因组学的研究进展 [J]. *植物分类与资源学报*, 2011, 33(4): 365-375.
- [18] 倪梁红, 赵志礼, 米玛. 药用植物叶绿体基因组研究进展 [J]. *中药材*, 2015, 38(9): 1990-1994.
- [19] Ronquist F, Huelsenbeck J P. MrBayes3: Bayesian phylogenetic inference under mixed models [J]. *Bioinformatics*, 2003, 19(12): 1572-1574.
- [20] Shim D, Raveendar S, Lee J R, *et al.* The complete chloroplast genome of *Capsicum frutescens* (Solanaceae) [J]. *Appl Plant Sci*, 2016, 4(5): 1-6.
- [21] Zhang X, Rong C X, Qin L, *et al.* Complete chloroplast genome sequence of *Malus hupehensis*: Genome structure, comparative analysis, and phylogenetic relationships [J]. *Molecules*, 2018, 23(11): 1-17.
- [22] Li W Q, Liu Y L, Yang Y, *et al.* Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros* [J]. *BMC Plant Biol*, 2018, 18(1): 1-11.
- [23] Samson N, Bausher M G, Lee S B, *et al.* The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: Organization and implications for biotechnology and phylogenetic relationships amongst angiosperms [J]. *Plant Biotechnol J*, 2010, 5(2): 339-353.
- [24] Straub S C K, Moore M J, Soltis P S, *et al.* Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae [J]. *Mol Phylogenet Evol*, 2014, 80(1): 169-185.
- [25] Sun S S, Fu P C, Gao Q B, *et al.* The complete plastome sequences of seven species in *Gentiana* Sect. *Kudoa* (Gentianaceae): Insights into plastid gene loss and molecular evolution [J]. *Front Plant Sci*, 2018, doi: 10.3389/fpls.2018.00493.
- [26] Stefanovic S, Olmstead R G. Down the slippery slope: Plastid genome evolution in Convolvulaceae [J]. *J Mol Evol*, 2005, 61(3): 292-305.
- [27] Martin M, Sabater B. Plastid *ndh* genes in plant evolution [J]. *Plant Physiol Biochem*, 2010, 48(8): 636-645.
- [28] Jansen R K, Cai Z, Raubeson L A, *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns [J]. *Proc Natl Acad Sci USA*, 2007, 104(49): 19369-19374.
- [29] Sun Y, Moore M J, Lin N, *et al.* Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family [J]. *BMC Genomics*, 2017, 18(1): 1-10.
- [30] Kim H T, Chase M W. Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae, Epidendroideae) [J]. *PLoS One*, 2017, 12(11): e0187318.
- [31] Kim H T, Kim J S, Moore M J B, *et al.* Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across Orchids and associated instability of the inverted repeat/small single-copy region boundaries [J]. *PLoS One*, 2015, 10(11): e0142215.
- [32] Zhou T, Chen C, Wei Y, *et al.* Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species [J]. *Front Plant Sci*, 2016, doi: 10.3389/fpls.2016.01512
- [33] Millen R S, Olmstead R G, Adams K L, *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus [J]. *Plant Cell*, 2001, 13(3): 645-658.
- [34] 李巧丽, 延娜, 宋琼, 等. 鲁桑叶绿体基因组序列及特征分析 [J]. *植物学报*, 2018, 53(1): 94-103.

- [35] Ni L H, Zhao Z L, Xu H X, *et al.* Chloroplast genome structures in *Gentiana* (Gentianaceae), based on three medicinal alpine plants used in Tibetan herbal medicine [J]. *Curr Genet*, 2016, 63(2): 241-251.
- [36] Zhou T, Wang J, Jia Y, *et al.* Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers [J]. *Int J Mol Sci*, 2018, 19(7): 1-15.
- [37] Fu P C, Zhang Y Z, Geng H M, *et al.* The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species [J]. *Peer J*, 2016, doi: 10.7717/peerj.2540.
- [38] Haberle R C, Fourcade H M, Boore J L, *et al.* Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes [J]. *J Mol Evol*, 2008, 66(4): 350-361.
- [39] Lee H L, Jansen R K, Chumley T W, *et al.* Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions [J]. *Mol Biol Evol*, 2007, 24(5): 1161-1180.
- [40] Hutchinson J. *The Families of Flowering Plants* [M]. Oxford: Oxford University Press, 1959.
- [41] 高娜娜, 赵志礼, 倪梁红. 植物叶绿体 *ycf15* 基因应用于药用植物鉴定的前景展望 [J]. 中草药, 2017, 48(15): 3210-3217.
- [42] Liu J, Yan H F, Ge X J. The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China [J]. *PLoS One*, 2016, 11(4): e0153008.
- [43] 倪梁红, 赵志礼, 熊波, 等. ITS 与叶绿体基因组多片段组合鉴定甘肃产秦艽基原植物的策略 [J]. 药学报, 2016, 51(5): 821-827.
- [44] Hilu K W, Liang G. The *matK* gene: Sequence variation and application in plant systematics [J]. *Am J Bot*, 1997, 84(6): 830-839.
- [45] Urbaniak J, Kwiatkowski P, Pawlikowski P. Phylogeography of *Swertia perennis* in Europe based on cp DNA markers [J]. *Peer J*, 2018, doi: 10.7717/peerj.5512.