

小秦艽叶绿体基因组学及谱系地理学研究

姜 帆, 剡心怡, 蒲 洋, 周 蕾, 周 涛*

西安交通大学药学院, 陕西 西安 710061

摘 要: **目的** 以小秦艽 *Gentiana dahurica* 为研究对象, 对其叶绿体基因组进行组装和序列分析, 基于叶绿体基因序列对我国不同地区野生小秦艽进行谱系地理学研究, 分析其遗传多样性及群体遗传结构, 探讨小秦艽居群遗传特征及地理分布格局间的成因及关联, 推测其冰期避难所及群体动态历史。**方法** 收集来自青海、甘肃、陕西和内蒙古等 6 个省份共 27 个野生小秦艽居群样本进行 DNA 重测序, 经组装及注释后获取所有个体叶绿体基因组, 并以此开展比较叶绿体基因组分析、系统发育分析、种群动态历史及谱系地理学分析。**结果** 共获得 81 个小秦艽叶绿体基因组, 均为典型环状四分体结构, 全长 148 792~149 000 bp, 注释得到 134 个基因 (85 个蛋白编码基因), 不同居群小秦艽叶绿体基因组 IR/SC 边界保守。叶绿体基因组非编码区变异明显高于编码区。小秦艽居群遗传多样性 ($h_s=0.632$) 和总遗传多样性 ($h_T=0.923$) 较高。系统发育分析结果表明小秦艽所有居群分为 2 个遗传分支, 但遗传距离与地理位置间无明显相关性。历史上小秦艽的群体规模总体较为恒定, 未出现明显的种群扩张现象。基于小秦艽非编码序列共识别到 27 种单倍型, 居群不存在明显的谱系地理结构。扩散廊道分析结果显示, 青藏高原东北边缘-祁连山-六盘山-吕梁山-太行山为小秦艽历史迁移的重要通道。**结论** 小秦艽叶绿体基因组结构高度保守, 基因数量稳定, 发现的 2 个突变热点可作为小秦艽特定区域种质资源分子鉴定的候选标记。野生小秦艽可能起源于青藏高原地区, 并以祁连山-六盘山-吕梁山-太行山等山脉作为潜在迁移廊道并维持了其不同居群的遗传连接度。将为药用植物小秦艽种质保护和资源开发利用提供理论基础。

关键词: 小秦艽; 叶绿体基因组; 遗传多样性; 谱系地理学; 迁移廊道

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2026)19-7819-11

DOI: 10.7501/j.issn.0253-2670.2026.19.025

Chloroplast genomics and phylogeography of *Gentiana dahurica*

JIANG Fan, YAN Xinyi, PU Yang, ZHOU Lei, ZHOU Tao

School of Pharmacy, Xi'an Jiaotong University, Xi'an 710061, China

Abstract: Objectives *Gentiana dahurica* was selected as the focal species to assemble and sequence analyze its chloroplast genome, investigate the phylogeographic structure of wild populations across China based on chloroplast gene sequences, assess genetic diversity and population genetic structure, explore the causes and correlations between the genetic characteristics and geographical distribution patterns of the *G. dahurica* population, and infer its glacial refugia and the dynamic history of the population. **Methods** A total of 27 wild populations of *G. dahurica* were collected from six provinces, including Qinghai, Gansu, Shaanxi, Inner Mongolia, ect. All individuals were subjected to DNA resequencing. Following genome assembly and annotation, the chloroplast genomes of all individuals were obtained for comparative genomic analyses, phylogenetic analysis, population dynamic history, and phylogeographic analyses. **Results** In total, 81 complete chloroplast genomes were generated, all exhibiting the typical quadripartite circular structure with lengths ranging from 148 792 to 149 000 bp. A total of 134 genes were annotated, including 85 protein-coding genes. The IR/SC boundaries of the chloroplast genomes of *G. dahurica* were conserved across populations. Variation in non-coding regions was markedly higher than in coding regions. *G. dahurica* exhibited high within-population genetic diversity ($h_s = 0.632$) and total genetic diversity ($h_T = 0.923$). Phylogenetic analyses assigned all populations into two genetic clades, but no significant correlation was detected between genetic distance and geographic location. Population size remained relatively stable over historical periods, with no evidence of strong population expansion. Based on the consensus of non-coding sequences of *G. dahurica*, 27 haplotypes were identified. Based on the consensus of non-coding sequences of *G. dahurica*, 27 haplotypes were identified. No clear phylogeographic structure was observed among populations. Dispersal corridor analyses suggested that the northeastern Qinghai-Tibet plateau-Qilian

收稿日期: 2026-03-21

基金项目: 国家自然科学基金面上项目 (32270386); 陕西省重点研发计划 (2025SF-YBXM-501)

作者简介: 姜 帆, 女, 硕士研究生, 研究方向为分子生药学。E-mail: jiangfan6881999@163.com

*通信作者: 周 涛, 男, 副教授, 从事药用植物分子遗传与资源及中药材种质资源评价研究。E-mail: zhou tao196@mail.xjtu.edu.cn

mountains-Liupan mountains-Lvliang mountains-Taihang mountains constituted an important historical migration pathway for *G. dahurica*. **Conclusions** The chloroplast genome of *G. dahurica* is highly conserved in structure and stable in gene content. Two mutation hotspots identified here may serve as candidate markers for molecular identification of germplasm resources. Wild *G. dahurica* likely originated on the Qinghai-Tibet plateau, and mountain chains such as the Qilian, Liupan, Lvliang, and Taihang ranges may have acted as historical migration corridors that facilitated genetic connectivity among populations. This study provides a theoretical foundation for germplasm conservation and development and utilization of this medicinal plant.

Key words: *Gentiana dahurica* Fischer; chloroplast genome; genetic diversity; phylogeography; migration corridor

叶绿体是植物进行光合作用和能量代谢的关键细胞器,在被子植物中遵循母系遗传规律,通常为典型的双链环状四分体结构,包括一个大单拷贝区(large single copy, LSC)、一个小单拷贝区(small single copy, SSC)和一对反向重复区(IRa/b)^[1-2]。相较于核基因组,叶绿体基因组具有序列与结构高度保守、进化速率适中等特点,随着高通量测序技术的发展及成本的降低,植物叶绿体基因组的获取已变得更加高效和普遍。目前,基于叶绿体基因组的分析在药用植物研究中得到了广泛运用,例如在吴茱萸、芍药、细辛等类群中,叶绿体基因组数据已成功用于区分不同品种乃至不同地区的中药材^[3-6]。此外,基于叶绿体基因组开发的 DNA 条形码因其高成功率、操作简便和重复性好等特点,在中药材的种属鉴定中应用日益广泛,相关案例如北柴胡、败酱草、风藤和山茱萸等^[7]。同时,叶绿体基因组也为谱系地理学研究提供了重要的遗传标记,并将其用于揭示类群的进化历史、遗传分化及其与气候变化、地质事件之间的关联,进而阐释物种现存分布格局的形成机制,该方法已在狭叶金粟兰、具鳞水柏枝等若干物种的谱系地理学研究中得到有效应用^[8-9]。因此,叶绿体基因组以其结构稳定、信息丰富和易于分析的特点,在植物系统学、资源鉴定与进化生物学等领域发挥着不可替代的作用。

小秦艽 *Gentiana dahurica* Fischer, 又名达乌里秦艽,隶属于龙胆科(Gentianaceae)龙胆属 *Gentiana* L.秦艽组,多分布于我国西北及青藏高原地区海拔 2 600~4 200 m 的高寒草甸或山地灌丛中,是重要的高山药用植物资源之一。小秦艽中含有多种环烯醚萜类、黄酮类及苷类化合物,具有清热解毒、祛风除湿、舒筋活络等药理作用,是传统中药秦艽的原植物之一,在中药材市场及民族药资源中具有重要地位^[10-11]。已有多项研究证实,秦艽属药材的化学表型与地理来源密切相关。牧其尔等对来自陕西宝鸡、内蒙古呼伦贝尔和东乌旗的 40 个批次秦艽花的 HPLC 指纹图谱研究显示,龙胆苦苷含量地域差异显著,聚类分析可将不同省份居群明显

区分^[12]。李霞等^[13]进一步收集了 7 个不同产区和 4 个不同基原的秦艽药材,发现不同产地或基原的样品中马钱苷酸、獐牙菜苦苷和龙胆苦苷的含量存在一定差异。这些研究表明不同地理区域或居群的秦艽药材在品质上存在差异,然而对于这种差异形成的遗传成因目前仍然尚未可知,极大程度地限制了其药用资源的合理利用。此外,由于长期过度采挖已导致小秦艽野生居群规模急剧下降,亟需对其野生资源进行遗传多样性评估并开展针对性的遗传保护。

目前仅有少数研究关注小秦艽种质资源遗传评估及鉴定。Zhou 等^[14]虽完成小秦艽全叶绿体基因组测序,并与秦艽组其他物种进行系统发育比对,但未对不同地区的小秦艽叶绿体基因组进行比较分析,以及可能用于鉴定小秦艽种质资源的潜在的 DNA 条形码。刘倩等^[15]基于内转录间隔区(ITS)探究了内蒙古不同地区小秦艽的亲缘关系,发现地理距离与其亲缘关系并无相关性。王聪聪^[16]利用叶绿体 *trnM-trnV* 片段分析了不同地区小秦艽单倍型差异,但该研究的居群取样量有限,不足以揭示小秦艽整体遗传多样性。综上,现有研究受限于低分辨率标记与片段化采样,既未能揭示小秦艽野生居群的遗传多样性特征,也未能追溯其冰期避难所及扩散路线。

本研究旨在通过对多个地区野生小秦艽叶绿体基因组的测序、组装与比较分析,系统解析其基因组结构特征、SSR 和长重复序列分布规律,评估种内遗传多样性及群体结构,并结合单倍型网络与地理信息推测历史扩散路径。研究结果不仅将为小秦艽的分子鉴定、系统进化研究提供分子基础,也为野生药用植物的保护、优良品种筛选及资源合理利用提供科学依据。

1 材料

本研究采集了青海、甘肃、陕西、山西、宁夏和内蒙古等多个地区 27 个居群,共计 81 份野生样品,经西安交通大学周涛副教授鉴定为小秦艽 *G. dahurica* Fischer。将采集的叶片装入取样袋,用硅胶干燥处理,带回实验室,用于后续总 DNA 提取。实验材料详细信息见表 1。

表 1 样本采集地点

Table 1 Sample collection locations

编号	产地	编号	产地
SXQS	山西省晋城市沁水县	GSYJ	甘肃省临夏自治州永靖县
SXFS	山西省忻州市繁峙县	GSXH	甘肃省甘南自治州夏河县
SNZD	陕西省延安市志丹县	GSTZ	甘肃省武威市天祝县
SNLC	陕西省宝鸡市麟游县	GSSD	甘肃省张掖市山丹县
QHTR	青海省同仁市瓜什则乡	GSSB	甘肃省张掖市肃南县长沟寺
QHTJ	青海省海西自治州天峻县	GSSA	甘肃省张掖市肃南县老虎沟村
QHDT	青海省西宁市大通县	GSML	甘肃省张掖市民乐县
NXPY	宁夏固原市彭阳县	GSLK	甘肃省临夏自治州康乐县
NXJY	宁夏固原市泾源县	GSHZ	甘肃省甘南自治州合作市
NMXW	内蒙古锡林郭勒盟西乌珠穆沁旗	GSHL	甘肃省庆阳市华池县
NMJN	内蒙古乌兰察布市丰镇市	GSHB	甘肃省平凉市华亭市策底镇
NMGY	内蒙古包头市固阳县	GSHA	甘肃省平凉市华亭市马峡镇
GSZN	甘肃省甘南自治州卓尼县	GSDX	甘肃省临夏自治州东乡族自治县
GSYZ	甘肃省兰州市榆中县		

2 方法

2.1 DNA 提取与测序

干燥叶片经组织研磨机研磨后，使用改良后的十六烷基三甲基溴化铵（CTAB）法从样本中提取 DNA。使用琼脂糖凝胶电泳和 NanoDrop 2000 检测 DNA 浓度和纯度。使用 BGI Optimal DNA Library Prep Kit 制备测序文库并基于 DNBSEQ-T7 测序平台进行高通量测序，测序读长为 150 bp。

2.2 叶绿体基因组组装与注释

测序得到的原始数据使用 fastp V0.23.4^[17]软件进行质控和过滤以剔除低质量、过短序列和含有过多不明碱基的序列和接头进而得到高质量测序数据。然后将质控后数据利用 GetOrganelle V1.7.0^[18]软件进行叶绿体基因组重头组装，并使用 CPGAVAS2（<http://47.96.249.172:16019/analyzer/home>）^[19]在线平台对组装完成的叶绿体基因组进行注释。将注释完成后的叶绿体基因组导入 Geneious Prime^[20]软件中进行人工校正。使用 Chloroplast（<https://irscope.shinyapps.io/Chloroplast/>）在线工具绘制小秦艽叶绿体基因组图谱。

2.3 重复序列鉴定

利用 MISA-web^[21]（<https://webblast.ipk-gatersleben.de/misa/>）在线平台鉴定小秦艽叶绿体基因组中的简单重复序列，参数设置为单、双、三、四、五和六核苷酸重复次数分别不小于 10、5、4、3、3 和 3。使用 TRF V4.09 软件鉴定串联重复序列，规定匹配权重、错配和插入缺失惩罚值分别为 2、7、7，Minscore 和 MaxPeriod 分别为 80 和 500。基

于 REputer^[22]在线软件（<https://bibiserv.cebitec.uni-bielefeld.de/reputer>）识别潜在的散在重复序列（包括正向重复、回文重复、互补重复和反向重复），软件参数设置为“Hamming Distance=3，最小序列一致性≥90%，碱基重复数量≥30 bp”。将所得结果整理后利用 ChiPlot 在线绘图网站进行结果可视化。

2.4 叶绿体基因组比较分析

利用在线平台 mVISTA（<https://genome.lbl.gov/vista/mvista/about.shtml>）将 81 个小秦艽个体的叶绿体基因组序列比对及可视化。使用 IRscope 软件分析秦艽不同个体 IR/SC 边界区域的基因特征和 IR 区结构扩张/收缩现象。使用 DnaSP V6^[23]统计了小秦艽叶绿体基因组全序列和蛋白编码基因的核苷酸多样性，并使用 R V4.5.1 包进行可视化。

2.5 系统发育分析

以 NCBI 下载龙胆草（NC_053842）和条纹华龙胆（MN199149）叶绿体基因组序列作为外类群，进行 81 个小秦艽全叶绿体基因组系统发育分析。使用 IQ-TREE V2.4.0^[24]基于最大似然法构建系统发生树，调用 MFP 选择最适合的进化模型，设置 1 000 次 Bootstrap。最后使用 FigTree V1.4.5 对获得的系统进化树进行美化。

2.6 贝叶斯天际线图分析

利用 BEAST V2.7.7 软件的 BDSKY 包进行贝叶斯天际线图（Bayesian skyline plot, BSP）分析，计算小秦艽有效种群大小的动态变化。使用 81 个小秦艽叶绿体基因组非编码区数据，使用 modeltest V3.7 筛选最优模型，通过层次似然比检验

(hierarchical likelihood ratio tests, hLRTs) 和赤池信息准则 (akaike information criterion, AIC) 2 种方法评估了 56 个候选模型。AIC 准则选择了 TIMeF 模型 (AIC=6.446 7), 然而, 模型不确定性分析显示 TIMeF 的 Akaike 权重仅为 0.188 0。层次似然比检验结果显示, 小秦艽碱基频率存在显著不均一性 ($P < 0.001$), 且位点间进化速率异质性显著 ($P < 0.001$), 而不变位点比例的支持度较低 ($P = 0.092$)。尽管 hLRT 推荐模型为 F81+G, 但考虑到 HKY 模型在仅增加一个参数的情况下即可同时刻画碱基频率不等和转换/颠换速率差异, 具有更高的生物学合理性与稳健性, 因此选择 HKY+G4 模型作为碱基替换模型。采用松弛对数正态分子钟 (relaxed clock log normal) 和溯祖贝叶斯天际线先验 (coalescent Bayesian skyline prior) 计算节点年龄。所采用的碱基替换率为被子植物叶绿体 DNA 的平均碱基替换率, 即 2.0×10^{-9} 替换/年/位点。独立运行 2 次, 每次包括 3×10^7 代 MCMC, 每 10 000 代采样 1 次。通过 LogCombiner V2.7.7 组合 2 次运行所得的 tree 和 log 文件。利用 Tracer V1.7.27 评估各参数是否收敛, 收敛标准为各参数的有效样本数均 > 200 且后验分布呈单峰分布。使用 Tracer 中贝叶斯天际线重建绘图。

2.7 失配分布及中性检验

利用 Dnasp V6.12 进行失配分布检验 (mismatch distribution), 当物种群体大小保持相对稳定时, 失配分布曲线呈现多峰, 如果历史上经历扩张或者持续增长, 则一般会呈现单峰。计算 Tajima's D 值进行中性检验, 推测小秦艽是否符合中性进化模式。利用在线绘图网站 Datawrapper (<https://app.datawrapper.de/plxgoniawr>) 绘制失配分布图。

2.8 谱系地理学分析

使用 DnaSP V6.12 软件识别小秦艽叶绿体基因组非编码序列单倍型, 并计算每个居群的单倍型多样性 (haplotype diversity, H_d) 和核苷酸多样性 (nucleotide diversity, π)。利用 Permut V2.0 软件计算居群内的平均基因多样性 (mean gene diversity within populations, h_s)、总基因多样性 (total gene diversity, h_T) 和遗传分化系数 (G_{ST} 和 N_{ST})。此外, 使用 Arlequin V3.5.2^[25] 软件对来自不同居群的小秦艽进行分子方差分析 (analyses of molecular variance, AMOVA), 分析遗传分化在种群间和种群

内所占比例, 并使用 1 000 次模拟检验显著性。利用 PopART V1.7 软件绘制叶绿体 DNA 单倍型网络图 (median-joining network)。利用 ArcGIS V10.8 绘制叶绿体 DNA 单倍型的地理分布图。使用 GenAlex V6.503^[26] 软件对小秦艽居群的地理距离和遗传距离进行 Mantel 相关性分析, 使用 R 进行可视化。

2.9 扩散路径分析

基于小秦艽 27 个居群的分布点以及非编码区单倍型分布情况, 采用最小成本路径分析方法 (least cost path, LCP) 将物种实际分布点遗传数据与 maxent 模拟的潜在分布数据相结合, 对小秦艽的潜在迁移路径进行计算与可视化。该方法基于以下假设: 物种出现概率越高的区域, 其对应的迁移成本越低, 从而模拟更符合实际的扩散路径。利用 SDMtoolbox 工具包中 “Landscape Connectivity” 模块下 “Calculate Least-Cost Corridors and Paths” 进行计算^[27]。将所有样点间的迁移路径结果叠加分析, 得到小秦艽在末次盛冰期 (last glacial maximum, LGM)、末次间冰期 (last interglacial, LIG) 和当代的潜在扩散通道图。预测模型使用的环境变量包含 19 个气候变量和 3 个地形因子, 从 WorldClim (<https://www.worldclim.org/data/worldclim21.html>) 全球气候数据库下载, 分辨率均为 30 s。

3 结果与分析

3.1 叶绿体基因组基本特征

组装完成的 81 个小秦艽个体的叶绿体基因组均呈现典型的环状四分体结构, 包括 LSC 区、SSC 区和 2 个 IR 区 (图 1)。小秦艽叶绿体基因组序列全长为 148 792~14 900 bp, 其中 LSC 的长度为 81 117~81 280 bp, SSC 的长度为 17 075~17 104 bp, IR 区的长度为 25 284~25 349 bp。

小秦艽叶绿体基因组序列总 GC 含量为 35.7%~35.8%, IR 区的 GC 含量最高 (42.7%), 其次是 LSC 区 (33.4%) 和 SSC 区 (28.5%)。每个叶绿体基因组均注释得到 134 个基因, 包括 85 个蛋白质编码基因、37 个 tRNA 基因、8 个 rRNA 基因以及 4 个假基因。其中 18 个基因在 IR 区重复, 包括 7 个蛋白质编码基因 (*rpl2*、*rpl23*、*rps7*、*rps12*、*ndhB*、*yef2*、*yef15*)、7 个 tRNA 基因 (*trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG*、*trnV-GAC*) 和 4 个 rRNA 基因 (*rrn4.5*、*rrn5*、*rrn16*、*rrn23*)。

IR/SC 边界分析结果 (图 2) 显示, 81 个体小秦艽叶绿体基因组的 IR/SC 边界区基因呈高度一

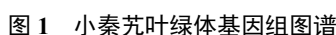


Fig. 1 Chloroplast genome maps for *G. dahurica*

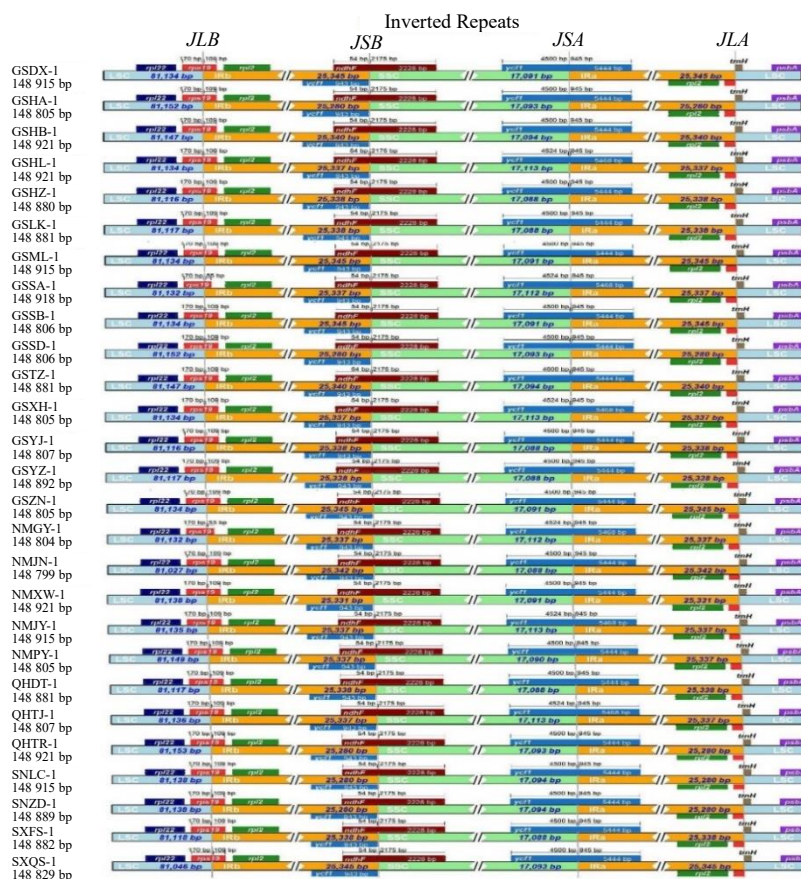


图2 小秦艽叶绿体基因组 IR/SC 边界分析 (部分)

Fig. 2 Analysis of IR/SC boundary regions in chloroplast genome of *G. dahurica* (partial)

利用 MISA 软件在 81 个小秦艽叶绿体基因组中共检测到了 30~38 个 SSR 位点, 包括 17~23 个单核苷酸重复 (以 A/T 碱基重复为主)、3 个二核苷酸重复、1~3 个三核苷酸重复、6 个四核苷酸重复

以及 3~5 个六核苷酸重复类型, 其中只有 GSHA、GSLK、QHTR 和 SXFS 居群中检测到 1 个五核苷酸重复。大多数 SSRs 分布在基因间隔区或内含子区, 只有少数 SSRs 位于蛋白质编码区。所有叶绿体基因组中的 SSR 分布情况大致相似, 一半以上位于 LSC 区, 其余分布于 SSC 或 IR 区。利用 REPuter 对小秦艽叶绿体基因组进行重复序列分析, 共鉴定到 23~25 个重复序列, 包括 11~12 个正向重复以

及 12~13 个回文重复, 其中仅有甘肃合作居群的一个个体 (GSHZ-4) 检测到一个反向重复序列以及 1 个互补重复序列 (图 3)。TRF 软件在小秦艽检测叶绿体基因组中检测到 25~31 个串联重复序列。

3.2 叶绿体基因组比较分析

以 GSDX-1 为参考序列, 使用 mVISTA 软件对 81 个小秦艽叶绿体基因组进行序列比较 (图 4)。结果显示, 小秦艽种内的序列相似性极高, 其中非编

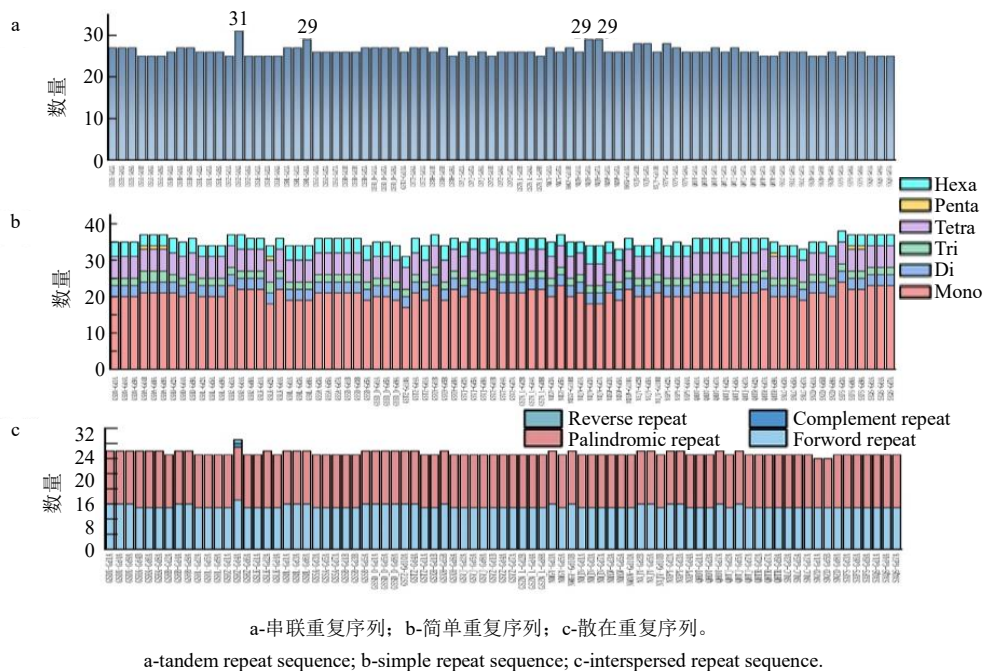


图 3 小秦艽叶绿体基因组的重复序列分析

Fig. 3 Repetitive sequence analysis of chloroplast genomes of *G. dahurica*

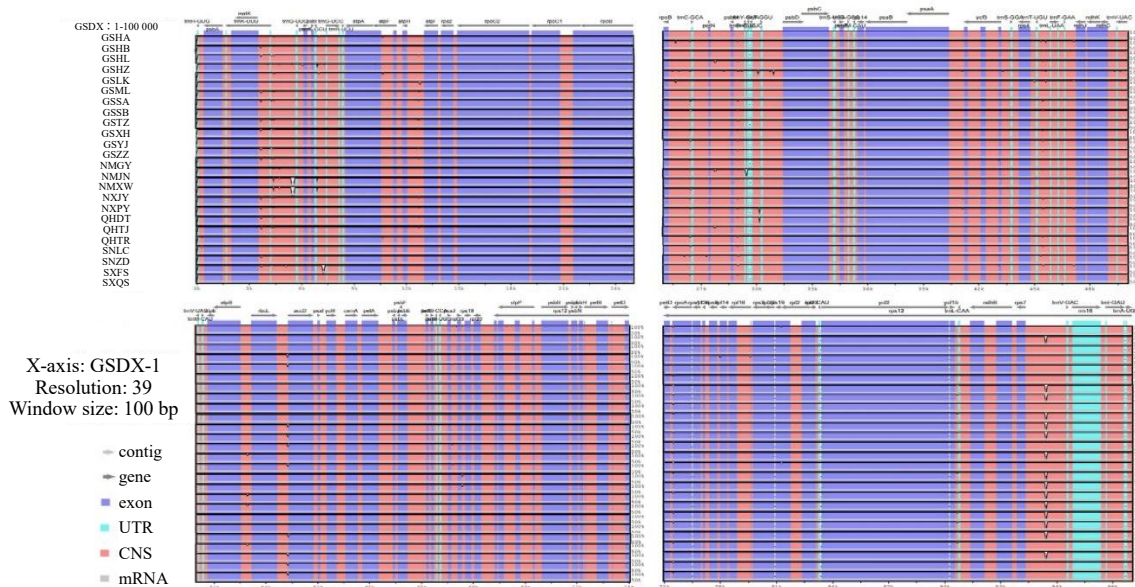


图 4 小秦艽叶绿体基因组 mVISTA 比对图 (部分)

Fig. 4 mVISTA-based alignment of chloroplast genomes of *G. dahurica* (partial)

码序列的变异明显高于编码区序列。使用 DnaSP 软件计算了 81 个小秦艽个体间的叶绿体基因组全序列核苷酸多样性 (P_i), 共检测到 1 189 个核苷酸多态性位点, 范围在 0.000 25~0.019 56 (图 5)。发现 *rpoB*~*trnC-GCA*、*rpl20*~*rps12*、*ndhD*~*psaC* 和 *ycf15*~*trnL-CAA* 处存在多态性较高的基因位点, 这 4 个片段均位于非编码区。其中在 *rpoB*~*trnC-GCA* 基因间隔区一个显著的 indel 变

异, 表现为 18 bp 完整序列与 6 bp 缺失型, 完整型在 GSHA 居群中固定存在, 而其余所有居群中均为缺失型, 仅在 GSLK 居群中发现一个完整型个体 (表 2)。在 *ycf15*~*trnL-CAA* 基因间隔区的序列变异包含 3 种分型, 完整型序列仅在 NMJN 居群的 2 个个体中被检出, 而其余 79 个个体在该位点均表现为缺失型序列。这些位点可作为鉴别小秦艽不同地理居群的潜在分子标记。

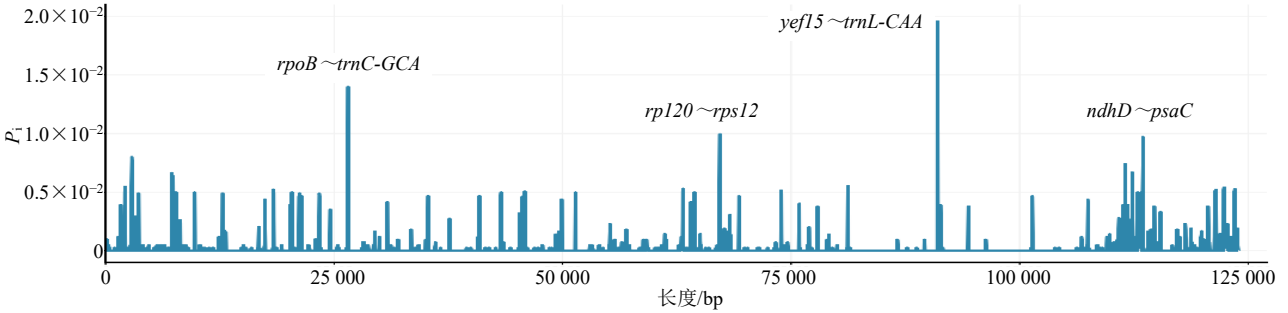


图 5 小秦艽叶绿体基因组的核酸多态性分析
Fig. 5 Nucleic polymorphism analysis of cloroplast genomes of *G. dahurica*

表 2 *rpoB*~*trnC-GCA* 与 *ycf15*~*trnL-CAA* 间隔区的 indel 多态性及其居群分布

Table 2 Indel polymorphisms and their population distribution in *rpoB*—*trnC-GCA* and *ycf15*—*trnL-CAA* intergenic regions

片段	分型	序列	所在居群 (个数)
<i>rpoB</i> ~ <i>trnC-GCA</i>	完整型	GATGAGAGATTTGATGAG	GSJA (3)、GSLK (1)
	缺失型	GAT-----GAG	其他 (77)
<i>ycf15</i> ~ <i>trnL-CAA</i>	完整型	TGAT...CAAAAT	NMJN (2)
	缺失型	TGAT...C-----或 ATCA...C-----	其他 (79)

3.3 系统发育分析和种群动态历史

以龙胆草 *G. cruciata* L. 和粗茎秦艽 *G. crassicaulis* Duthie ex Burk. 为外类群, 构建 81 个小秦艽叶绿体基因组的系统发育树 (图 6)。结果显示, 除外类群外, 小秦艽所有个体被划分为 3 个遗传分支。将小秦艽居群按照地理位置划分为东部、中部和西部居群, 其中东部居群包括内蒙古、山西省居群, 中部居群包含宁夏、陕西以及甘肃东部的部分居群, 西部居群为青海和甘肃西部居群。发现各居群在系统发育树的每个分支中几乎均有分布, 仅在 A 分支中不存在 East 居群, 这表明小秦艽并没有按照地理分布进行聚类。值得注意的是, 系统发育树基部类群主要分布于青藏高原及中部地区, 这表明青藏高原可能作为小秦艽的重要起源与早期分化中心之一, 随后其居群向在不断伴随气候震荡下向北进一步扩散形成了当今的分布格局。

通过贝叶斯天际线图 (图 7) 可以看出, 小秦

艽居群仅在 0.5 百万年前经历有效种群大小扩张, 其余阶段 (>1 Ma) 种群规模未发生剧烈的变化, 处于动态平衡状态。中性检验结果显示, Tajima's *D* 的值为不显著的负值 (-1.462 39, $P>0.10$)。失配分布分析结果 (图 8) 显示小秦艽实际观测曲线为多峰, 同时期望值是一条平滑下降的曲线, 与观测曲线不符, 表明小秦艽未经历明显的快速扩张事件, 分析结果与 BSP 结果一致。以上结果可以看出, 小秦艽的有效群体大小 0.5 Ma 之前保持相对稳定, 仅在更早期出现轻微波动, 且置信区间较宽, 表明历史上小秦艽的群体规模总体较为恒定。

3.4 谱系地理学分析

本研究中发现小秦艽叶绿体基因组变异主要分布在非编码区, 因此选择 81 个叶绿体基因组的非编码区进行比对进行单倍型分析, 共检测到 42 593 个单核苷酸变异位点, 在不考虑缺失情况下共识别到 27 种叶绿体单倍型。其中, 单倍型 Hap_1、Hap_5、

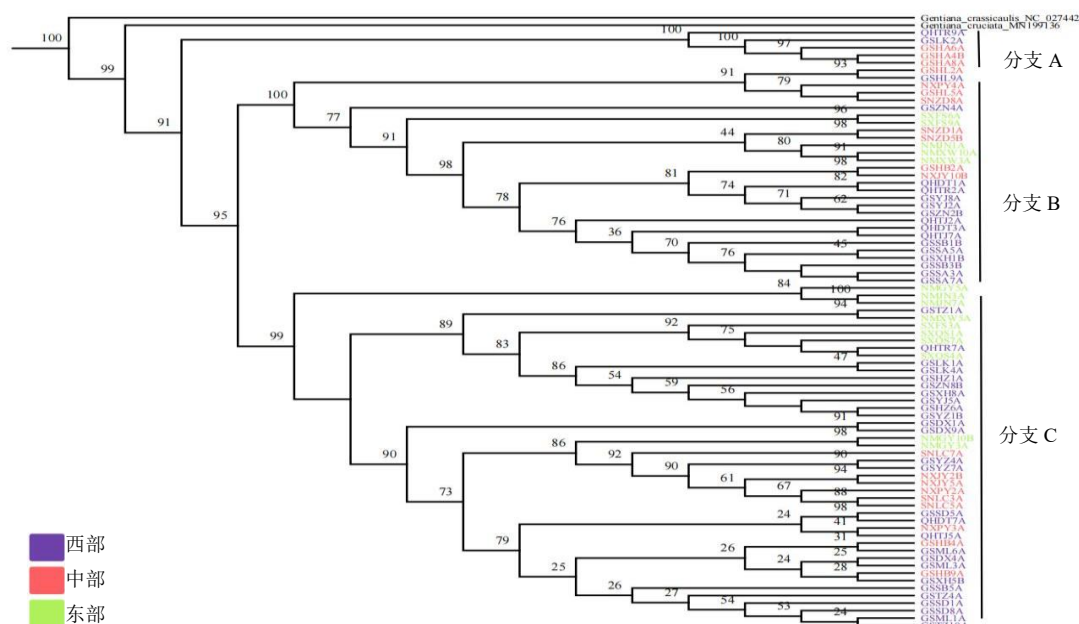
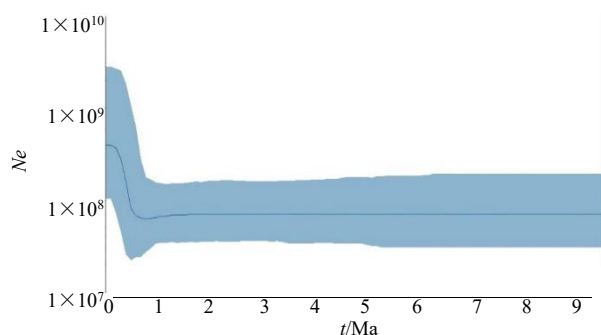


图 6 基于叶绿体全基因组构建的系统发育树

Fig. 6 Phylogenetic tree based on complete chloroplast genome



中间粗蓝线表示有效种群大小 (N_e) 的中值, 上、下细蓝线表示有效种群大小的 95% 最高后验密度 (HPD) 置信区间。

The bold blue line indicates the median values of effective population size (N_e). The upper and lower blue lines represent the confidence intervals of the 95% highest posterior density (HPD) interval of N_e .

图 7 小秦艽居群的贝叶斯天际线图

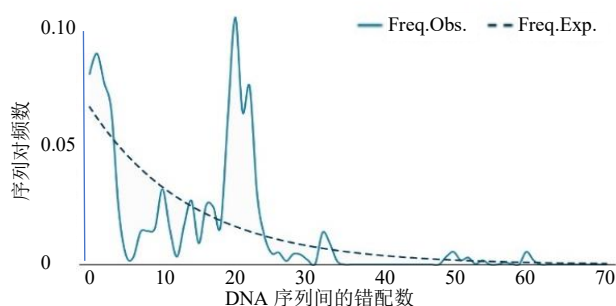
Fig. 7 Bayesian skyline plot for *G. dahurica* populations

图 8 小秦艽失配分布

Fig. 8 Mismatch distribution plot of *G. dahurica*

Hap_10 和 Hap_22 是多数居群的共享单倍型, 结合其较高频率及广泛分布的特征, 推测 Hap_1 可能为祖先单倍型。该单倍型位于网络的主干位置, 呈现出祖先型单倍型的典型特征。超半数单倍型为私有单倍型 (55.6%), 有 20 个居群中包含多种单倍型, 达到了总居群的 74%, 单倍型多样性 (H_d) 平均值为 0.917, 具有较高的平均居群内遗传多样性 ($h_s=0.632$) 和较高的总遗传多样性 ($h_T=0.923$)。

基因流系数 N_{ST} 考虑了单倍型之间的遗传距离, 居群遗传分化系数 G_{ST} 则仅考虑单倍型距离。由于小秦艽 N_{ST} 略高于 G_{ST} ($N_{ST}=0.330 > G_{ST}=0.300$), 且分子方差分析 (AMOVA) 显示仅 50.22% 的遗传变异分配至居群间 (表 3), 表明小秦艽所有居群不存在显著的谱系地理结构, 遗传距离与地理距离的 Mantel 检验 ($P=0.32 > 0.05$) 也验证了这一结果 (图 9)。

3.5 扩散廊道分析

通过结合物种分布模型与最小成本路径分析, 可以推测小秦艽在不同历史时期的可能迁移路线。结果表明, 在不同时期甘肃与宁夏居群的祁连山附近遗传连通性较高, 小秦艽显示出沿山脉和高原的扩散模式, 以六盘山为交流中心的青藏高原东北部边缘-祁连山-六盘山-吕梁山-太行山是小秦艽迁移过程中的重要走廊。

表 3 基于小秦艽 cpDNA 单倍体序列变异的分子变异分析

变异来源	df	SS	VP/%
种间	26	474.494	50.22
组内	54	380.667	49.78
总数	80	855.160	

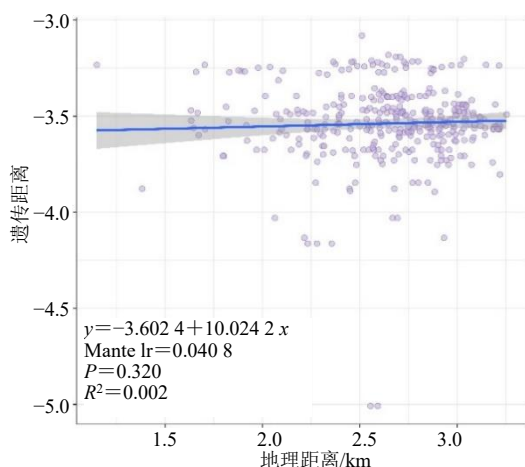


图 9 小秦艽遗传距离与地理距离的 Mantel 检验

Fig. 9 Mantel test of genetic distance with geographic distance of *G. dahurica*

4 讨论

本研究从叶绿体基因组层面评估了小秦艽叶绿体基因组的结构特征、种内遗传多样性格局及潜在演化历史，为该药用植物的资源评价与利用提供了重要分子依据。小秦艽叶绿体基因组大小差异较小，全长介于 148 792~14 900 bp，IR/SC 边界未检测到显著扩张或变异，表明其在种内演化过程中保持较高稳定性。

全叶绿体基因组序列比较识别出了 4 个高变异片段 ycf15~trnL-CAA、rpoB~trnC-GCA 和 ndhD~psaC，均位于非编码区，可以作为潜在分子标记用于 GSHA、NMJN 等小秦艽居群的种质鉴定。长重复序列通常在叶绿体基因组结构变异与进化中具有关键作用^[28-29]。小秦艽以正向重复和回文重复为主。值得注意的是，反向重复和互补重复仅在甘肃合作居群的个体（GSHZ-4）被检测到，这一独特的重复序列类型可能反映了该个体叶绿体基因组的特异性结构变异。

群体历史分析未检测到近期明显的快速扩张信号，表明小秦艽种群规模在较长时期内相对稳定，仅在约 0.5 Ma 的中更新世中期出现了轻微上升

趋势，该时间点恰处于中更新世气候转型（Mid-Pleistocene Transition, MPT, 1.2~0.7 Ma）之后，是全球气候波动由 41 kyr 周期向 100 kyr 长周期冰期—间冰期振荡过渡并趋于稳定的重要阶段^[30-31]。小秦艽在我国主要分布于北部的高海拔地区，多项研究显示这些地区在 MPT 之后呈现持续变冷、干旱加剧和冬季风增强等特点^[32-35]。同时，青藏高原的隆升，如“昆黄运动”（1.1~0.6 Ma）等也显著影响着地表环境以及生物多样性的演化^[36]，这些环境变化可能促进了小秦艽种群扩散与数量增长。

物种冰期避难所中的居群通常具备较高的遗传多样性和特有单倍型，相较之下，由后期迁移或扩散形成的居群更易受到遗传漂变或奠基者效应的影响，从而导致单倍型多样性降低^[37]。因此，拥有较高单倍型和核苷酸多样性并包含独特单倍型的居群，往往被视为物种的潜在起源中心或第四纪冰期的避难所^[38]。小秦艽谱系地理学分析识别到 27 种单倍型，但种内未形成明显的谱系地理结构。其中 hap_1、hap_22 和 hap_10 位于中心位置，且超半数的单倍型以它们为核心呈星状辐射分布，结合其较高的出现频率，推测这三者为小秦艽的祖先单倍型。祖先单倍型主要分布于青藏高原东北缘，该区域在第四纪冰期循环中为许多山地植物（如三叶木通、花椒属等）提供了关键的避难所功能^[39-40]。结合单倍型网络分析、系统发育关系以及青藏高原东北缘已知的地质历史和避难所背景，本研究支持该区域不仅作为小秦艽在冰期的重要避难所，更可能是该物种的一个关键起源与分化中心。

种群遗传结构和分化通常是多种因素综合作用的结果，其中地理距离往往通过隔离效应限制基因流动，从而导致群体间的遗传分化^[9,40-42]。然而，小秦艽的地理距离与遗传距离之间并无显著相关性，类似的结果也曾在五味子和香椿中被报道^[43]。这一格局可能源于多重过程的共同作用。首先，其相对恒定的种群规模削弱了遗传漂变的作用^[44]，同时中更新世气候转型后冰期—间冰期循环加剧，可能促进跨区域的扩散与基因交流。其次，秦艽组内近缘物种之间存在杂交潜力^[45-46]，叶绿体捕获现象可能进一步重塑叶绿体遗传结构。最后，其依赖昆虫授粉并具备一定程度的远距离种子扩散潜力，可能维持跨区域的低水平基因交流。这些因素均可能导致不同地区小秦艽之间存在较高的遗传相似性，从而弱化了距离隔离模式。

综上,本研究不仅丰富了我国传统药材小秦艽的叶绿体基因组资源,揭示了其种内遗传结构与演化历史,也为其药用资源的分子鉴定、种质保护及后续开发利用提供了重要参考。未来结合核基因组、生态因子以及代谢组数据开展综合研究,将有助于更全面理解该物种的适应进化机制及药用性状形成规律,从而更有效地促进其资源利用与分子育种研究。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 朱玲, 同栩瑶, 王强, 等. 太白贝母叶绿体全基因组序列特征及其系统发育分析 [J]. 中草药, 2025, 56(19): 7157-7167.
- [2] Freudenthal J A, Pfaff S, Terhoeven N, *et al.* A systematic comparison of chloroplast genome assembly tools [J]. *Genome Biol*, 2020, 21(1): 254.
- [3] 张万祥. 基于叶绿体基因组的吴茱萸基原鉴定与药材质量研究 [D]. 长沙: 湖南中医药大学, 2023.
- [4] An K L, Yin G Y, Wang X, *et al.* Intraspecific identification and genetic diversity analysis of *Paeonia lactiflora* based on chloroplast genes *rpoB* and *psbK-psbI* [J]. *Mol Biol Rep*, 2025, 52(1): 165.
- [5] 赵俊禧, 姜霞, 石盼盼, 等. 基于叶绿体基因组序列筛选酸枣仁特异性引物的研究 [J]. 中药材, 2025, 48(5): 1119-1122.
- [6] Li L, Sun W, Zhang M F, *et al.* Identification of *Asari Radix et Rhizoma* derived from three close *Asarum* species using the complete chloroplast genome [J]. *Phytomedicine*, 2025, 147: 157160.
- [7] 王馨, 尹光耀, 张志飞, 等. 北柴胡特异 DNA 条形码筛选及种质资源鉴定 [J]. 中草药, 2023, 54(17): 5703-5715.
- [8] 刘浩宇. 具鳞水柏枝叶叶绿体基因组及谱系地理学研究 [D]. 西宁: 青海师范大学, 2024.
- [9] 姚淑婷. 岸栖植物狭叶金粟兰的谱系地理学研究 [D]. 桂林: 广西师范大学, 2024.
- [10] 宝钢, 图诺木拉, 朝格巴达拉夫, 等. 龙胆属秦艽组植物化学成分及药理作用研究进展 [J]. 中成药, 2024, 46(8): 5691-5712.
- [11] Li Y, Zhang J, Fan J Y, *et al.* Tibetan medicine Bang Jian: A comprehensive review on botanical characterization, traditional use, phytochemistry, and pharmacology [J]. *Front Pharmacol*, 2023, 14: 1295789.
- [12] 牧其尔, 白艳秋, 乌吉斯古楞, 等. HPLC 法同时测定蒙药秦艽花中龙胆苦苷、异荭草苷、牡荆素的含量及其指纹图谱的建立 [J]. 中医药导报, 2025, 31(12): 83-89.
- [13] 李霞, 何志鹏, 王仕宝, 等. 基于 HPLC 的秦艽药材特征指纹图谱构建及主成分分析 [J]. 甘肃农业大学学报, 2023, 58(3): 115-124.
- [14] 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): 1962.
- [15] 刘倩, 黄艺晓, 臧二欢, 等. 内蒙古不同产地小秦艽 ITS 序列比较研究 [J]. 中医药导报, 2023, 29(6): 55-59.
- [16] 王聪聪. 秦艽组四种植物谱系地理与物种分化研究 [D]. 西安: 西北大学, 2013.
- [17] Chen S F, Zhou Y Q, Chen Y R, *et al.* Fastp: An ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): i884-i890.
- [18] 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.
- [19] 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.
- [20] Kearse M, Moir R, Wilson A, *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28(12): 1647-1649.
- [21] Beier S, Thiel T, Münch T, *et al.* MISA-web: A web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33(16): 2583-2585.
- [22] Kurtz S, Choudhuri J V, Ohlebusch E, *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Res*, 2001, 29(22): 4633-4642.
- [23] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J C, *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. *Mol Biol Evol*, 2017, 34(12): 3299-3302.
- [24] 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.
- [25] Excoffier L, Lischer H E L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows [J]. *Mol Ecol Resour*, 2010, 10(3): 564-567.
- [26] Peakall R, Smouse P E. GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research: An update [J]. *Bioinformatics*, 2012, 28(19): 2537-2539.
- [27] Brown J L, Bennett J R, French C M. SDMtoolbox 2.0: The next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution

- model analyses [J]. *Peer J*, 2017, 5: e4095.
- [28] Xu S J, Teng K, Zhang H, *et al.* Chloroplast genomes of four *Carex* species: Long repetitive sequences trigger dramatic changes in chloroplast genome structure [J]. *Front Plant Sci*, 2023, 14: 1100876.
- [29] Yao X, Tan Y H, Liu Y Y, *et al.* Chloroplast genome structure in *Ilex* (Aquifoliaceae) [J]. *Sci Rep*, 2016, 6: 28559.
- [30] Mukhin D, Gavrilov A, Loskutov E, *et al.* Bayesian data analysis for revealing causes of the middle pleistocene transition [J]. *Sci Rep*, 2019, 9: 7328.
- [31] 王婷, 孙有斌, 刘星星. 中更新世气候转型: 特征、机制和展望 [J]. 科学通报, 2017, 62(33): 3861-3872.
- [32] 曾琳. 中国东北地区 1.2Ma 以来的黄土地层和古气候记录 [D]. 南京: 南京大学, 2016.
- [33] 胡云壮, 杨吉龙, 李影, 等. 3Ma 以来华北地区植被演化特征及其驱动因素 [J]. 地质学报, 2020, 94(10): 3120-3129.
- [34] 盛美, 李宗耀, 仝亚博, 等. 青藏高原东北缘西宁黄土的磁学记录及其环境意义 [J]. 地球学报, 2025, 46(6): 1082-1094.
- [35] 赵世锦. 早更新世以来中国西北植被对气候突变转型的响应 [D]. 北京: 中国地质大学, 2024.
- [36] 丁林, 熊中玉, 田小龙, 等. 从造山带到青藏高原的隆升: 地质演化与气候变化的机理连通性 [J]. 科学通报, 2026, 71(2): 494-517.
- [37] 王璐. 漆树谱系地理学和景观基因组学研究 [D]. 南京: 南京林业大学, 2023.
- [38] 潘凤. 刺梨的遗传多样性及谱系地理学研究 [D]. 贵阳: 贵州大学, 2022.
- [39] 董晴. 基于叶绿体基因和 ITS 序列的三叶木通谱系地理学研究 [D]. 雅安: 四川农业大学, 2023.
- [40] 黄勤琴. 基于叶绿体全基因组序列的花椒属系统发育关系及其演化历史分析 [D]. 重庆: 重庆三峡学院, 2023.
- [41] 张倩, 张德全. 白花刺楸断野生居群的叶绿体全基因组特征解析 [J]. 广西植物, 2022, 42(10): 1750-1761.
- [42] Liu Y J, Wu M L, Xu X, *et al.* Genetic diversity and phylogeography of the endemic species *Chimonobambusa utilis* growing in southwest China: Chloroplast DNA sequence and microsatellite marker analyses [J]. *Front Plant Sci*, 2022, 13: 943225.
- [43] Ye J W, Zhang Z K, Wang H F, *et al.* Phylogeography of *Schisandra chinensis* (Magnoliaceae) reveal multiple refugia with ample gene flow in Northeast China [J]. *Front Plant Sci*, 2019, 10: 199.
- [44] 金永玲. 五趾跳鼠 (*Orientallactaga sibirica*) 群体遗传结构及表型分化研究 [D]. 呼和浩特: 内蒙古农业大学, 2024.
- [45] 郭伟娜, 魏朔南. 秦艽的生物学研究 [J]. 中国野生植物资源, 2008, 27(4): 1-5.
- [46] 李小娟. 麻花艽和管花秦艽(龙胆科)之间自然杂交的 RAPD 分析 [J]. 吉林农业, 2010(9): 53-55.

[责任编辑 时圣明]