

星毛唐松草叶绿体基因组特征及系统发育分析

侯 哲, 娄晓鸣, 李 昂, 黄长兵^①

(苏州农业职业技术学院, 江苏 苏州 215008)

摘要: 对星毛唐松草 (*Thalictrum cirrhosum* Lévl.) 叶绿体基因组的碱基组成和基因类型等特征进行分析, 并与同属近缘种黏唐松草 (*Thalictrum viscosum* C. Y. Wu)、丝叶唐松草 (*Thalictrum foeniculaceum* Bunge)、朝鲜唐松草 (*Thalictrum coreanum* H. Lévl.) 和芸香唐松草 [*Thalictrum thalictroides* (Linn.) Eames et Boivin] 的叶绿体基因组进行比较和系统发育分析。结果显示: 星毛唐松草叶绿体基因组中 AT 含量 (61.6%) 明显高于 GC 含量 (38.4%), 反向重复区 (IR) 的 GC 含量最高 (43.2%); 注释到的 133 个基因中 20 个基因含有 2 份拷贝, 17 个基因含有 2 个内含子, 仅 *clpPI* 基因含有 3 个内含子; 根据基因对应的产物, 133 个基因可分为 17 类, 其中, 蛋白质编码基因中核糖体蛋白小亚基和光系统 II 亚基的编码基因数量最多 (14); 叶绿体基因组中共检测到 58 个 SSR 位点, 其中, 由 T、A 和 C 组成的单核苷酸 SSR 位点最多 (51), 占总 SSR 位点数的 88%。星毛唐松草及其 4 个近缘种叶绿体基因组的大单拷贝区 (LSC)、小单拷贝区 (SSC) 和 IR 区分别包含 409、84 和 260 个多态性位点, 核苷酸多态性平均值分别为 0.009 4、0.014 0 和 0.002 1; IR 区序列变异较小, 表明 IR 区相对保守。星毛唐松草及其 4 个近缘种的叶绿体基因组结构、基因数量和 GC 含量均较为相似, 非编码序列变异程度均明显高于编码序列, 但在 IR 区的边界处有一定差异。聚类分析结果显示: 唐松草属 (*Thalictrum* Linn.) 植物聚为一支, 其中, 星毛唐松草与黏唐松草的亲缘关系最近。综合研究表明: 星毛唐松草叶绿体基因组的碱基组成偏向使用 A 或 T, *trnV-UAC* 基因、*ndhF* 和 *ndhD* 基因以及 *ycf1* 基因对应的 3 个区域为星毛唐松草及其 4 个近缘种的核苷酸多态性热点区域; 星毛唐松草及其 4 个近缘种的叶绿体基因组特征较为相似, *ycf1*、*ndhF*、*rpl2* 和 *rps19* 基因的收缩或扩张可能是唐松草属植物叶绿体基因组 IR 区存在差异的主要原因; 叶绿体基因组可为唐松草属植物的分类及鉴定提供可靠依据。

关键词: 星毛唐松草; 叶绿体基因组; 序列比对; 系统发育分析

中图分类号: Q943.2; Q949.746.5; S567.2 文献标志码: A 文章编号: 1674-7895(2023)03-0024-09
DOI: 10.3969/j.issn.1674-7895.2023.03.03

Analyses on characteristics of chloroplast genome and phylogeny of *Thalictrum cirrhosum* HOU Zhe, LOU Xiaoming, LI Ang, HUANG Changbing^① (Suzhou Polytechnic Institute of Agriculture, Suzhou 215008, China), *J. Plant Resour. & Environ.*, 2023, 32(3): 24-32

Abstract: Characteristics including base composition and gene type of chloroplast genome of *Thalictrum cirrhosum* Lévl. were analyzed, and comparison and phylogenetic analysis with those of related species in the same genus namely *Thalictrum viscosum* C. Y. Wu, *Thalictrum foeniculaceum* Bunge, *Thalictrum coreanum* H. Lévl., and *Thalictrum thalictroides* (Linn.) Eames et Boivin were carried out. The results show that AT content (61.6%) in chloroplast genome of *T. cirrhosum* is obviously higher than GC content (38.4%), and GC content in the inverted repeat region (IR) is the highest (43.2%); 20 out of 133 annotated genes contain two copies, and 17 genes contain two introns, while only *clpPI* gene contains three introns; based on the products of genes, 133 genes can be divided into 17 categories, in which, the number of genes that encode small subunit of ribosome protein and subunit of photosystem II are the most

收稿日期: 2022-10-10

基金项目: 四川省自然科学基金项目(2022NSFSC1141); 江苏省特色花卉工程中心基金项目(苏发改高技发[2020]1460号); 江苏省省级作物种质资源库(球宿根花卉)项目(JS-ZW-K05); 苏州农业职业技术学院博士提升计划(BS[2022]08)

作者简介: 侯 哲(1991—), 男, 山西运城人, 博士, 讲师, 主要从事园林植物抗逆性分子机制方面的研究。

^①通信作者 E-mail: cbhuang@szai.edu.cn

引用格式: 侯 哲, 娄晓鸣, 李 昂, 等. 星毛唐松草叶绿体基因组特征及系统发育分析[J]. 植物资源与环境学报, 2023, 32(3): 24-32.

(14) among protein-coding genes; there are 58 SSR loci in the chloroplast genome, in which, the single nucleotide SSR loci consisting of T, A, and C are the most (51), which accounts for 88% of total number of SSR loci. The large single copy region (LSC), small single copy region (SSC), and IR region of the chloroplast genomes of *T. cirrhosum* and its four related species contain 409, 84, and 260 polymorphic loci, respectively, and the averages of nucleotide polymorphism are 0.009 4, 0.014 0, and 0.002 1, respectively; the sequence variation of IR region is small, indicating that the IR region is relatively conservative. The structures, gene numbers, and GC contents of chloroplast genomes of *T. cirrhosum* and its four related species are all relatively similar, and the variation degree of non-coding sequences is obviously higher than that of coding sequences, but there are some differences in the boundaries of IR regions. The cluster analysis result shows that *Thalictrum* Linn. species can be clustered into one branch, in which, *T. cirrhosum* has the closest genetic relationship with *T. viscosum*. The comprehensive study result shows that the base composition of chloroplast genome of *T. cirrhosum* prefers A or T, the three regions corresponding to *trnV-UAC* gene, *ndhF* and *ndhD* genes, and *ycf1* gene were nucleotide polymorphism hotspots of *T. cirrhosum* and its four related species; the characteristics of chloroplast genomes of *T. cirrhosum* and its four related species are relatively similar, and contraction or expansion of *ycf1*, *ndhF*, *rpl2*, and *rps19* genes are the main reason for the differences in IR regions of chloroplast genomes of *Thalictrum* species; the chloroplast genome can provide reliable basis for the classification and identification of *Thalictrum* species.

Key words: *Thalictrum cirrhosum* Lévl.; chloroplast genome; sequence alignment; phylogenetic analysis

星毛唐松草 (*Thalictrum cirrhosum* Lévl.) 为毛茛科 (Ranunculaceae) 唐松草属 (*Thalictrum* Linn.) 草本植物, 一般分布于中国云南和四川一带^[1]。唐松草属植物在全世界约有 150 种, 中国有 76 种, 其中约 30 种富含生物碱、三萜、酚类和黄酮等化合物, 具有较高的药用价值, 根和茎常作为传统的民间药物, 用于治疗流感和结膜炎等疾病^[2-3]。目前, 关于唐松草属植物的研究主要集中在化学成分分析等方面^[4], 而唐松草属植物许多类群在分类系统学上尚存在着一定的争议, 唐松草属植物系统发育关系的研究有待进一步探讨。近年来, 基于叶绿体基因组构建的系统发育树在揭示不同物种系统发育关系方面取得了一定的进展^[5]。

叶绿体基因组是高等植物独立的遗传物质^[6], 长度一般在 120~160 kb, 组装成功的叶绿体基因组通常具有闭合的环状结构, 包含 1 个序列较长的大单拷贝区 (LSC)、1 个序列较短的小单拷贝区 (SSC) 和 2 个反向重复区 (IRa 和 IRb), 其中, IRa 区与 IRb 区的碱基序列一致但方向相反^[7]。叶绿体基因组有 110~130 个基因^[8], 其编码区的基因序列相对保守, 与其他质体基因组相比, 叶绿体基因组的进化速率较低, 结构较为稳定, 常被用于植物的群体进化、群体遗传学和系统发育关系的研究^[9-12]。迄今为止, NCBI (<https://www.ncbi.nlm.nih.gov/>) 数据库中已包含大量的植物叶绿体基因组序列, 其中唐松草属仅有几个

物种的叶绿体基因组公布。本项目组前期对星毛唐松草的叶绿体基因组进行了测序^[13], 但星毛唐松草在毛茛科唐松草属内的系统发育位置尚不明确, 形态学方面的研究也未见相关报道, 不利于唐松草属植物后续研究工作的开展。鉴于此, 本文在项目组前期测序结果的基础上对星毛唐松草叶绿体基因组序列特征进行了分析, 并与已发表的唐松草属其他植物的叶绿体基因组进行比对和系统发育分析, 以期对唐松草属植物后续的群体遗传学、种质资源保护和系统发育研究提供理论基础。

1 材料和方法

1.1 材料

供试材料星毛唐松草为云南省漾濞县大平地山 (东经 99°57'29.30"、北纬 25°40'11.71") 原生分布单株, 于 2022 年 3 月从各植株上随机采集新梢嫩叶 3~5 枚, 共采集 3 株, 混合后装入含有硅胶的自封袋带回实验室保存、备用。

1.2 方法

1.2.1 序列碱基组成和基因类型分析 基于前期获得的星毛唐松草叶绿体基因组数据 (GenBank 登录号 OL711693)^[13], 使用南京集思慧远生物科技有限公司开发的在线程序 (<http://cloud.genepioneer.com:9929/#/tool/alltool?type=4>) 中的序列 GC 含量统计

工具对星毛唐松草叶绿体基因组中大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 的 A、T、C 和 G 组成进行统计和分析,使用 GeSeq^[14] 软件和 tRNAscan-SE^[15] 软件对星毛唐松草叶绿体基因组进行注释。

1.2.2 简单重复序列 (SSR) 分析 使用 MISA 在线工具 (<https://webblast.ipk-gatersleben.de/misa/>) 分析星毛唐松草叶绿体基因组的简单重复序列,其中,单核苷酸的重复次数设置为 10,双核苷酸的重复次数设置为 5,三核苷酸的重复次数设置为 4,四核苷酸、五核苷酸和六核苷酸的重复次数均设置为 3,其余参数采用默认值。

1.2.3 核苷酸多态性 (Pi) 分析 为确定唐松草属植物叶绿体基因组的多态性热点区域 (hotspot),在 NCBI 数据库中选择星毛唐松草及其近缘种黏唐松草 (*Thalictrum viscosum* C. Y. Wu) (GenBank 登录号 NC058831)、丝叶唐松草 (*Thalictrum foeniculaceum* Bunge) (GenBank 登录号 NC053570)、朝鲜唐松草 (*Thalictrum coreanum* H. Lév.) (GenBank 登录号 NC026103) 和芸香唐松草 [*Thalictrum thalictroides* (Linn.) Eames et Boivin] (GenBank 登录号 NC039433) 的叶绿体基因组序列,使用 DnaSP v6.12.03 软件,采用滑动窗口法对星毛唐松草及其近缘种叶绿体基因组的核苷酸多态性进行分析,步长设置为 200 bp,窗口长度设置为 600 bp^[16]。峰值越大表示核苷酸多态性越高。

1.2.4 与近缘种叶绿体基因组的比较 为了解星毛唐松草叶绿体基因组与亲缘关系相近的同属植物叶绿体基因组间的差别,以黏唐松草叶绿体基因组为参考基因组,使用 mVISTA 在线工具 (<https://genome.lbl.gov/vista/mvista/submit.shtml>) 对星毛唐松草和上述 4 个近缘种叶绿体基因组进行比对分析 (Shuffle-LAGAN 模式),窗口长度设置为 100 bp,若各物种的序列图完全一致则代表各基因组间无分化,序列图中峰值越大表示变异程度越高;使用 IRscope 在线工具 (<https://irscope.shinyapps.io/irapp/>) 对星毛唐松草及其 4 个近缘种叶绿体基因组 IR 区边界的收缩与扩张情况进行分析。

1.2.5 唐松草属植物系统发育分析 从 NCBI 数据库中下载唐松草属目前已知的 12 条叶绿体基因组序列,以同科近缘属的尾囊草 [*Urophysa henryi* (Oliv.) VUbr.] (GenBank 登录号 NC039744) 和距瓣尾囊草

(*Urophysa rockii* Ulbr.) (GenBank 登录号 NC039742) 为外类群,使用 MAFFT 软件^[17] 对所有物种的叶绿体基因组序列进行比对,人工调整后,使用 MEGA7.0 软件^[18] (基于 Tamura-Nei 模型),采用最大似然法 (maximum likelihood, ML) 构建系统发育树,bootstrap 值设置为 1 000。

2 结果和分析

2.1 星毛唐松草叶绿体基因组碱基组成和基因类型分析

对星毛唐松草叶绿体基因组中大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 的碱基组成情况进行统计,结果见表 1。结果显示:星毛唐松草叶绿体基因组中 A、T、C 和 G 含量分别为 30.5%、31.1%、19.6% 和 18.8%,GC 含量为 38.4%,明显低于 AT 含量 (61.6%);其中,IR 区的 GC 含量最高 (43.2%),LSC 区次之 (36.6%),SSC 区的 GC 含量最低 (32.4%)。

表 1 星毛唐松草叶绿体基因组中碱基组成
Table 1 Composition of bases in chloroplast genome of *Thalictrum cirrhosum* Lévl.

区域 Region	碱基含量/% Base content				
	A	T	C	G	GC
大单拷贝区 Large single copy region	31.0	32.3	18.8	17.8	36.6
小单拷贝区 Small single copy region	34.1	33.6	17.1	15.3	32.4
反向重复区 Inverted repeat region	28.6	28.2	20.9	22.3	43.2
总体 Overall	30.5	31.1	19.6	18.8	38.4

基因组注释结果 (表 2) 显示:根据基因对应的产物,星毛唐松草叶绿体基因组 133 个基因可分为 17 类。蛋白质编码基因中核糖体蛋白小亚基和光系统 II 亚基的编码基因数量最多,均有 14 个;核酮糖-1,5-二磷酸羧化酶大亚基、成熟酶、蛋白酶、被膜蛋白、乙酰 CoA 亚基和细胞色素 C 合成酶的编码基因数量最少,均仅有 1 个。*trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnM-CAU*、*trnN-GUU*、*trnR-ACG*、*trnV-GAC*、*rrn4.5*、*rrn5*、*rrn16*、*rrn23*、*rps7*、*rps12*、*rpl2*、*rpl23*、*ndhB*、*ycf1*、*ycf2* 和 *ycf15* 这 20 个基因均有 2 份拷贝,*trnA-UGC*、*trnG-GCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnV-UAC*、*rps12*、*rps16*、*rpl2*、*rpl16*、*rpoC1*、*ndhA*、*ndhB*、*petB*、*petD*、*atpF* 和 *pafl* 这 17 个基因均含有 2 个

内含子, *clpP1* 基因含有 3 个内含子。

2.2 星毛唐松草叶绿体基因组 SSR 分析

星毛唐松草叶绿体基因组中共检测到 58 个 SSR 位点, 对不同重复频率的 SSR 位点数量进行统计, 结果(表 3) 显示: 单核苷酸 SSR 位点最多(51), 占总 SSR 位点数的 88%, 双核苷酸 SSR 位点有 7 个, 其他

核苷酸 SSR 位点未检测到。在单核苷酸 SSR 位点中, 由 T 组成的单核苷酸 SSR 位点最多(33), 由 A 组成的单核苷酸 SSR 位点次之(16), 由 C 组成的单核苷酸 SSR 位点仅 2 个, 未检测到由 G 组成的单核苷酸 SSR 位点, 表明星毛唐松草叶绿体基因组的碱基组成偏向使用 A 或 T。

表 2 星毛唐松草叶绿体基因组注释结果

Table 2 Results of chloroplast genome annotations of *Thalictrum cirrhosum* Lévl.

产物类别 Product category	基因 ¹⁾ Gene ¹⁾	数量 Number
tRNA	<i>trnA-UGC</i> (×2) **, <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-UCC</i> , <i>trnG-GCC</i> **, <i>trnH-GUG</i> , <i>trnI-CAU</i> (×2), <i>trnI-GAU</i> (×2) **, <i>trnK-UUU</i> **, <i>trnL-UAA</i> **, <i>trnL-UAG</i> , <i>trnL-CAA</i> (×2), <i>trnM-CAU</i> (×2), <i>trnN-GUU</i> (×2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-UCU</i> , <i>trnR-ACG</i> (×2), <i>trnS-GCU</i> , <i>trnS-UGA</i> , <i>trnS-GGA</i> , <i>trnT-UGU</i> , <i>trnT-GGU</i> , <i>trnV-UAC</i> **, <i>trnV-GAC</i> (×2), <i>trnW-CCA</i> , <i>trnY-GUA</i>	37
rRNA	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)	8
核糖体蛋白小亚基 Small subunit of ribosome protein	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (×2) **, <i>rps14</i> , <i>rps15</i> , <i>rps16</i> **, <i>rps18</i> , <i>rps19</i>	14
核糖体蛋白大亚基 Large subunit of ribosome protein	<i>rpl2</i> (×2) **, <i>rpl14</i> , <i>rpl16</i> **, <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl33</i> , <i>rpl36</i> , <i>rpl32</i>	11
RNA 聚合酶亚基 Subunit of RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> **, <i>rpoC2</i>	4
NADH 脱氢酶亚基 Subunit of NADH dehydrogenase	<i>ndhA</i> **, <i>ndhB</i> (×2) **, <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhF</i> , <i>ndhK</i>	12
光系统 I 亚基 Subunit of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
光系统 II 亚基 Subunit of photosystem 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>psbT</i> , <i>psbZ</i>	14
细胞色素亚基 Subunit of cytochrome	<i>petA</i> , <i>petB</i> **, <i>petD</i> **, <i>petG</i> , <i>petL</i> , <i>petN</i>	6
ATP 合成酶亚基 Subunit of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> **, <i>atpH</i> , <i>atpI</i>	6
核酮糖-1,5-二磷酸羧化酶大亚基 Large subunit of rubisco	<i>rbcL</i>	1
成熟酶 Maturase	<i>matK</i>	1
蛋白酶 Protease	<i>clpP1</i> ***	1
被膜蛋白 Envelope protein	<i>cemA</i>	1
乙酰 CoA 亚基 Subunit of acetyl-CoA	<i>accD</i>	1
细胞色素 C 合成酶 Synthase of cytochrome C	<i>ccsA</i>	1
未知功能蛋白 Unknown function protein	<i>ycf1</i> (×2), <i>ycf2</i> (×2), <i>ycf15</i> (×2), <i>infA</i> , <i>pafl</i> **, <i>paflII</i> , <i>pbfl</i>	10
合计 Total		133

¹⁾(×2): 2 份拷贝 Two copies; **: 含有 2 个内含子 Containing two introns; ***: 含有 3 个内含子 Containing three introns.

表 3 星毛唐松草叶绿体基因组 SSR 分析

Table 3 SSR analysis on chloroplast genome of *Thalictrum cirrhosum* Lévl.

重复单元 Repeat unit	不同重复频率的 SSR 位点数 Number of SSR loci with different repetition rates								总计 Total	
	6	7	10	11	12	13	14	15		16
A			9	4	3					16
C			1						1	2
T			17	6	3	2	4	1		33
AT	2	1								3
TA	2	2								4
总计 Total										58

2.3 星毛唐松草及其近缘种叶绿体基因组的核苷酸多态性分析

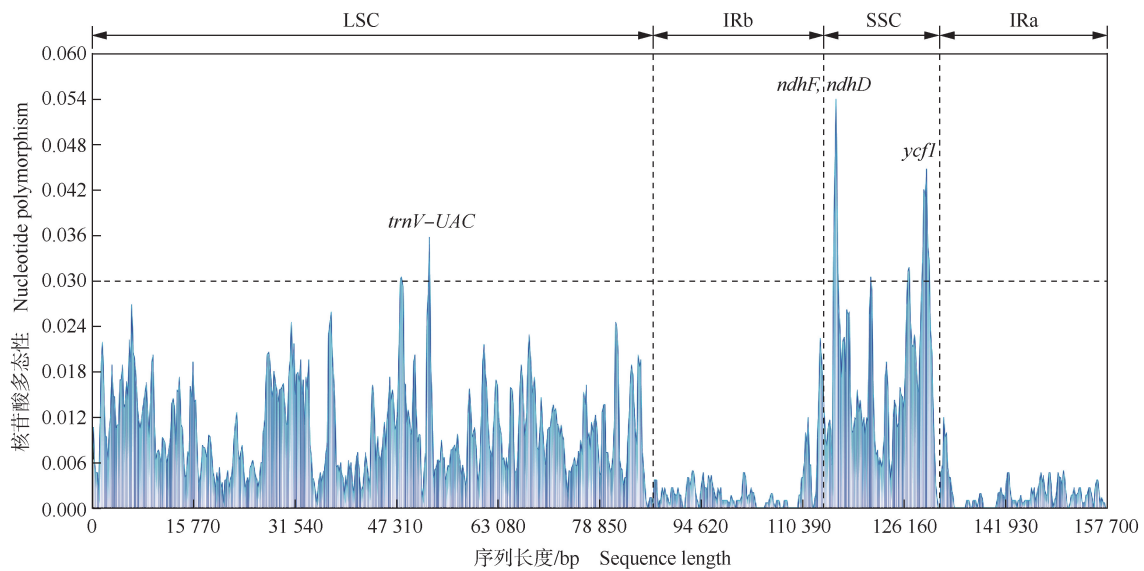
对星毛唐松草及其 4 个近缘种叶绿体基因组的核苷酸多态性进行分析, 结果见图 1。结果显示: 共检测到 753 个多态性位点, 其中大单拷贝区(LSC)包含的多态性位点最多(409), 核苷酸多态性平均值为 0.009 4; 小单拷贝区(SSC)包含的多态性位点最少(84), 核苷酸多态性平均值为 0.014 0; 反向重复区(IR)包含 260 个多态性位点, 核苷酸多态性平均值为 0.002 1。说明 IR 区序列变异较小, 表现相对保

守。图中峰值较高的 3 个区域为核苷酸多态性热点区域,分别对应 *trnV-UAC* 基因、*ndhF* 和 *ndhD* 基因以及 *ycf1* 基因,其中, *trnV-UAC* 基因(核苷酸多态性为 0.032 0)位于 LSC 区, *ndhF* 和 *ndhD* 基因(核苷酸多态性为 0.048 3)以及 *ycf1* 基因(核苷酸多态性为 0.040 0)位于 SSC 区。

2.4 星毛唐松草及其近缘种叶绿体基因组的比较

2.4.1 基因组特征和序列比对 星毛唐松草、黏唐松草、丝叶唐松草、朝鲜唐松草和芸香唐松草的叶绿体基因组特征见表 4。结果显示:星毛唐松草及其

4 个近缘种的叶绿体基因组结构相似,均为典型的四分体结构,其中,芸香唐松草的叶绿体基因组长度最短(154 924 bp),黏唐松草的叶绿体基因组长度最长(155 984 bp);朝鲜唐松草的大单拷贝区(LSC)长度最短(84 733 bp),黏唐松草的 LSC 区长度最长(85 339 bp);芸香唐松草的小单拷贝区(SSC)长度最短(17 497 bp),星毛唐松草的 SSC 区长度最长(17 657 bp)。星毛唐松草及其 4 个近缘种的叶绿体基因数量(132~134)和 GC 含量(38.0%~38.4%)也较为相近。



LSC: 大单拷贝区 Large single copy region; SSC: 小单拷贝区 Small single copy region; IRa, IRb: 反向重复区 Inverted repeat region.

图 1 星毛唐松草及其近缘种叶绿体基因组核苷酸多态性分析

Fig. 1 Analysis on nucleotide polymorphism of chloroplast genome of *Thalictum cirrhosum* Lévl. and its related species

表 4 星毛唐松草及其近缘种叶绿体基因组特征比较¹⁾

Table 4 Comparison on characteristics of chloroplast genomes among *Thalictum cirrhosum* Lévl. and its related species¹⁾

物种 Species	l_G /bp	l_{LSC} /bp	l_{SSC} /bp	l_{IR} /bp	n_1	n_2	n_3	n_4	$C_{GC}/\%$
星毛唐松草 <i>Thalictum cirrhosum</i>	155 969	85 324	17 657	26 494	133	88	37	8	38.4
黏唐松草 <i>Thalictum viscosum</i>	155 984	85 339	17 655	26 495	134	88	38	8	38.0
丝叶唐松草 <i>Thalictum foeniculaceum</i>	155 923	85 323	17 628	26 486	132	87	37	8	38.2
朝鲜唐松草 <i>Thalictum coreanum</i>	155 088	84 733	17 549	26 403	133	88	37	8	38.1
芸香唐松草 <i>Thalictum thalictroides</i>	154 924	84 899	17 497	26 273	134	89	37	8	38.1

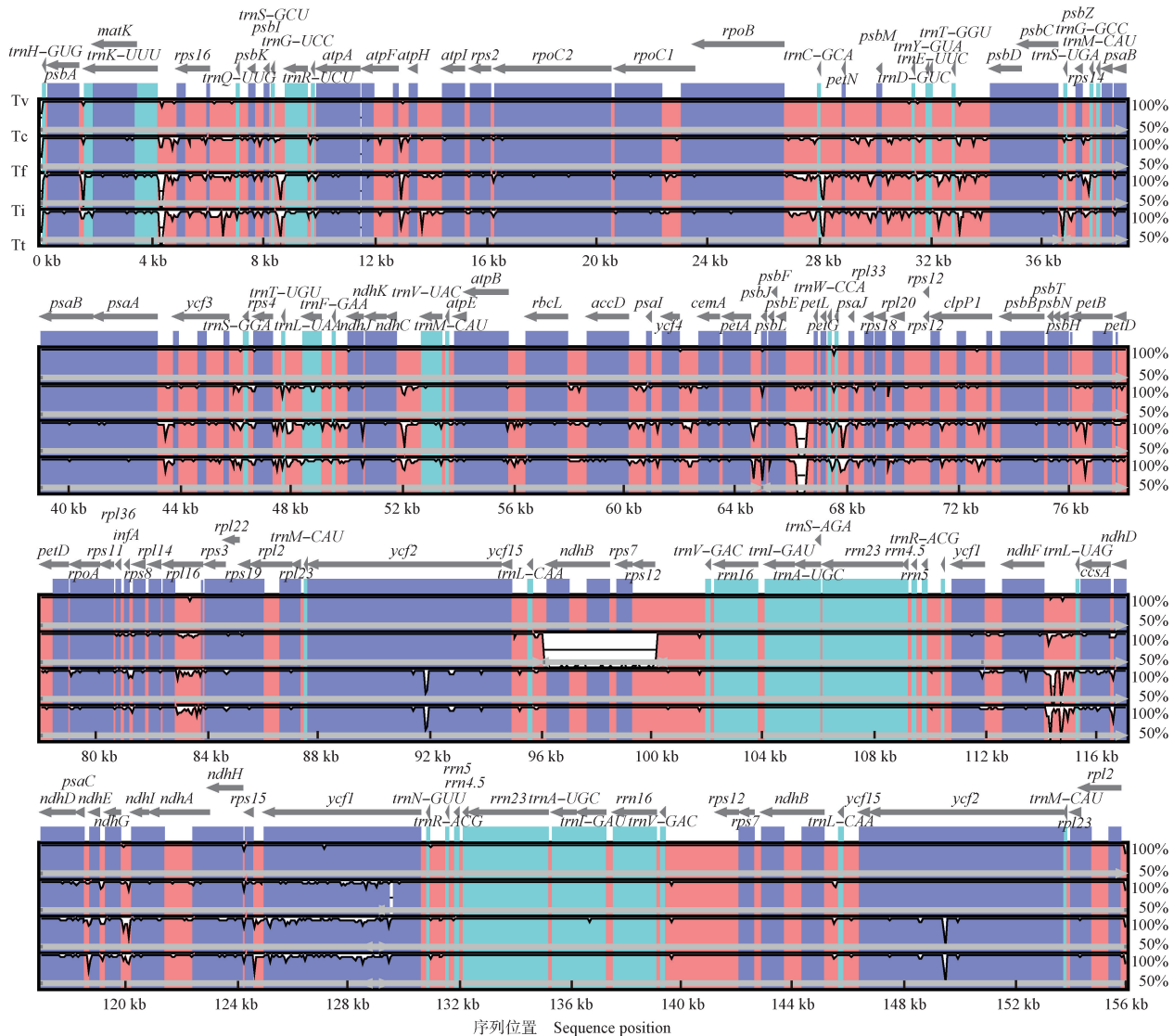
¹⁾ l_G : 基因组长度 Length of genome; l_{LSC} : 大单拷贝区长度 Length of large single copy region; l_{SSC} : 小单拷贝区长度 Length of small single copy region; l_{IR} : 反向重复区长度 Length of inverted repeat region; n_1 : 基因数量 Number of gene; n_2 : 蛋白质编码基因数 Number of protein-coding genes; n_3 : tRNA 对应的基因数 Number of genes corresponding to tRNA; n_4 : rRNA 对应的基因数 Number of genes corresponding to rRNA; C_{GC} : GC 含量 GC Content.

叶绿体基因组序列比对结果(图 2)显示:星毛唐松草及其 4 个近缘种叶绿体基因组的非编码序列变异程度均明显高于编码序列,而与反向重复区(IR)

相比,LSC 区和 SSC 区在物种间的差异较大。大多数序列的变异发生在基因间隔区,这些区域比其他区域表现出明显较高的分化程度。

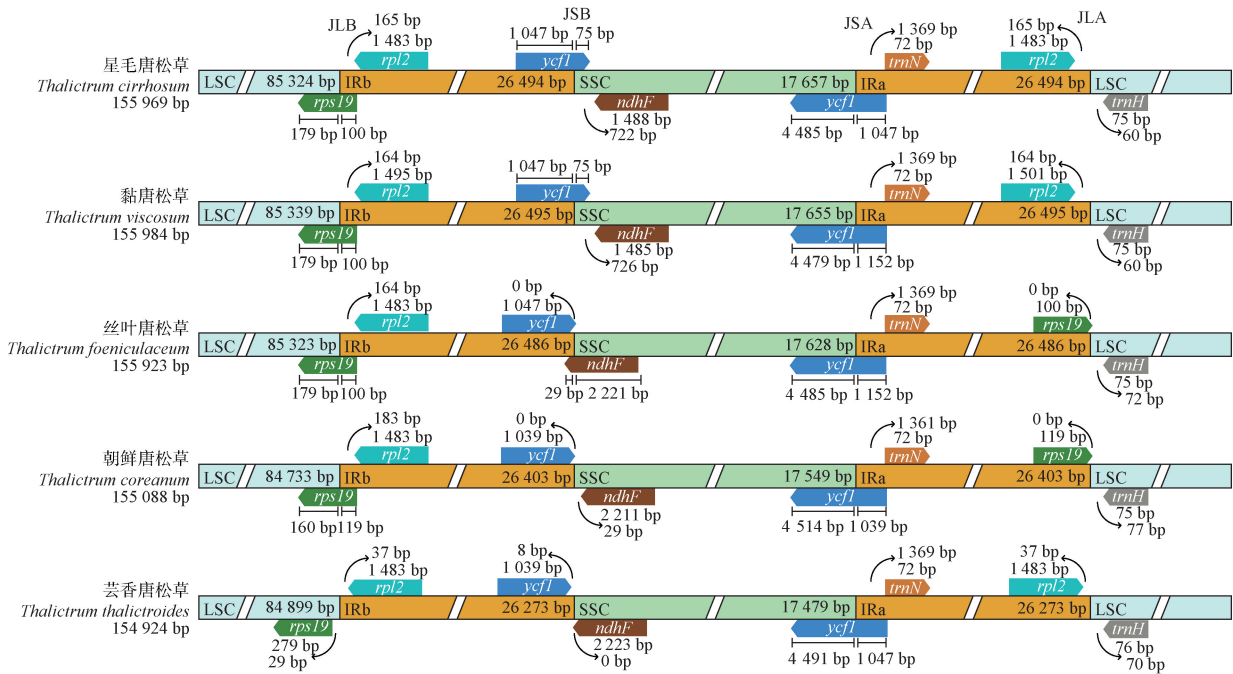
2.4.2 IR 区边界的收缩与扩张分析 星毛唐松草及其近缘种叶绿体基因组 LSC 区、SSC 区和 IR 区边界的比较结果见图 3。结果显示:星毛唐松草的叶绿体基因组中,IR 区边界与 4 个近缘种基本一致,星毛唐松草、黏唐松草、丝叶唐松草和朝鲜唐松草的 JLB 边界(LSC 区与 IRb 区的边界)均位于 *rps19* 基因内,而芸香唐松草 JLB 边界位于 *rps19* 基因和 *rpl2* 基因间;星毛唐松草与黏唐松草的 JSB 边界(IRb 区与 SSC 区的边界)位于 *ycf1* 基因内,丝叶唐松草的 JSB 边界位

于 *ndhF* 基因内,而朝鲜唐松草与芸香唐松草的 JSB 边界位于 *ycf1* 基因与 *ndhF* 基因间;5 种植物的 JSA 边界(SSC 区与 IRa 区的边界)均位于 *ycf1* 基因内;星毛唐松草、黏唐松草和芸香唐松草的 JLA 边界(IRa 区与 LSC 区的边界)均位于 *rpl2* 基因与 *trnH* 基因间,丝叶唐松草和朝鲜唐松草的 JLA 边界均位于 *rps19* 基因与 *trnH* 基因间。说明 5 种唐松草属植物叶绿体基因组的 JSA 边界差异最小,而 JSB 和 JLA 边界差异较大,可能在这些区域基因发生了收缩或扩张。



Tv: 黏唐松草 *Thalictrium viscosum* C. Y. Wu; Tc: 星毛唐松草 *Thalictrium cirrhosum* Lévl.; Tf: 丝叶唐松草 *Thalictrium foeniculaceum* Bunge; Ti: 朝鲜唐松草 *Thalictrium coreanum* H. Lévl.; Tt: 芸香唐松草 *Thalictrium thalictroides* (Linn.) Eames et Boivin. ◀: 重叠群 Contig; ▶: 基因 Gene; ■: 外显子 Exon; ■: 非翻译区 Untranslated region; ■: 非编码序列 Noncoding sequence; ■: mRNA. 右侧百分比表示每个位点碱基在不同参考序列间的最高(100%)和最低(50%)保守性 The percentages on the right represent the highest (100%) and lowest (50%) conservation of each site base among different reference sequences.

图 2 星毛唐松草及其近缘种叶绿体基因组的序列比对结果
Fig. 2 Sequence alignment results of chloroplast genomes among *Thalictrium cirrhosum* Lévl. and its related species



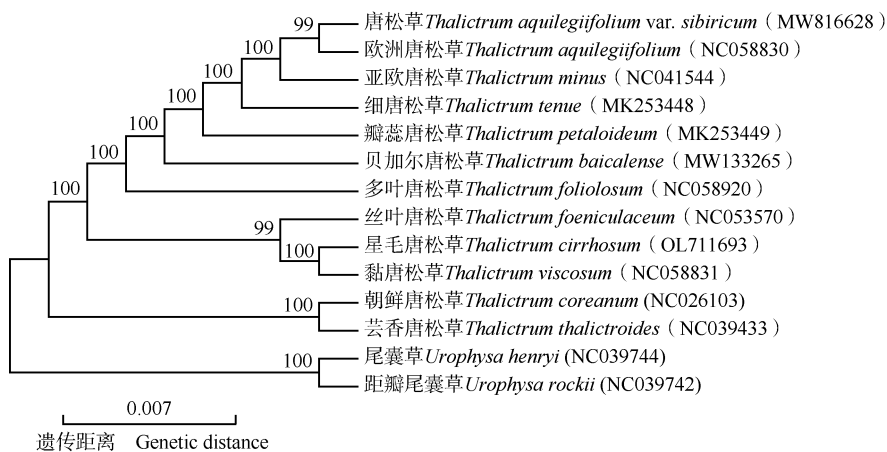
JLB: LSC 区和 IRb 区的边界 Boundary of LSC and IRb regions; JSB: IRb 区和 SSC 区的边界 Boundary of IRb and SSC regions; JSA: SSC 区和 IRa 区的边界 Boundary of SSC and IRa regions; JLA: IRa 区和 LSC 区的边界 Boundary of IRa and LSC regions.

图 3 星毛唐松草及其近缘种叶绿体基因组大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 边界的比较
 Fig. 3 Comparison on the boundaries of large single copy region (LSC), small single copy region (SSC), and inverted repeat region (IR) in chloroplast genomes among *Thalictrium cirrhosum* Lévl. and its related species

2.5 唐松草属植物系统发育分析

基于 NCBI 数据库中已知的唐松草属 12 种植物叶绿体基因组序列,以尾囊草和距瓣尾囊草为外类群构建系统发育树。结果(图 4)显示:14 种植物的遗传变异度为 0.007,共形成 11 个节点,其中 9 个节点

的自展支持率为 100%,2 个节点的自展支持率为 99%。从亲缘关系来看,星毛唐松草、黏唐松草和丝叶唐松草聚为一支,星毛唐松草与黏唐松草的亲缘关系最近;朝鲜唐松草和芸香唐松草聚为一支;欧洲唐松草 (*Thalictrium aquilegiifolium* Linn.)、唐松草



系统发育树分支上的数据为自展支持率,括号内编号为 GenBank 登录号 The data on the branches of phylogenetic tree are the bootstrap values, and the Nos. in brackets are GenBank accession numbers.

图 4 基于叶绿体基因组的唐松草属植物的系统发育树
 Fig. 4 Phylogenetic tree of *Thalictrium* Linn. species based on chloroplast genomes

(*Thalictrum aquilegifolium* var. *sibiricum* Linn.)、亚欧唐松草 (*Thalictrum minus* Linn.)、细唐松草 (*Thalictrum tenue* Franch.)、瓣蕊唐松草 (*Thalictrum petaloideum* Linn.)、贝加尔唐松草 (*Thalictrum baicalense* Turcz.) 和多叶唐松草 (*Thalictrum foliolosum* DC.) 聚为一支。

3 讨论和结论

星毛唐松草叶绿体基因组为典型的四分体结构,包含1个大单拷贝区(LSC)、1个小单拷贝区(SSC)和2个反向重复区(IRa和IRb),长度为155 969 bp,这符合大部分被子植物叶绿体基因组的特征^[19]。本文中,星毛唐松草叶绿体基因组IR区序列变异较小,而LSC区和SSC区序列变异较大,这与目前已经公布的唐松草属植物的叶绿体基因组类似^[20-21]。Šmarda等^[22]的研究结果显示:植物基因组的碱基组成(GC含量)会显著影响基因组的功能。本研究结果显示:唐松草属不同植物的叶绿体基因组,甚至同一个叶绿体基因组内不同区域的GC含量是不同的。星毛唐松草叶绿体基因组中,LSC区的GC含量为36.6%,SSC区的GC含量为32.4%,IR区的GC含量最高(43.2%)。这可能是由于IR区存在8个rRNA,而rRNA一般具有非常高的GC含量,且LSC区、SSC区和IR区间GC含量的这种差异是被子植物叶绿体基因组的典型特征^[23]。

尽管星毛唐松草叶绿体基因组与同属近缘种黏唐松草、丝叶唐松草、朝鲜唐松草和芸香唐松草的叶绿体基因组比较相似,但在IR区的边界处有一定差异,5种植物SSC区与IRa区的边界差异最小,而IRb区与SSC区以及IRa区与LSC区的边界差异较大,*ycf1*、*ndhF*、*rpl2*和*rps19*基因显示出较高的序列分化,在兰科(Orchidaceae)植物叶绿体基因组中也得到类似的结果^[24]。推测这4个基因可能在唐松草属植物的叶绿体基因组进化过程中发挥了至关重要的作用。mVISTA分析结果显示:非编码序列变异程度明显高于编码序列,这与大多数被子植物中发现的多样性模式相一致^[25],总的来说,星毛唐松草叶绿体基因组的序列变异集中在非编码序列,而编码序列的变异是有限的。许多研究也发现非编码序列富含多种序列调节因子,说明非编码序列可能在生物进化和关键的生物功能上扮演着十分重要的角色^[26]。

叶绿体基因组比较保守,在物种分类及鉴定过程中有十分重要的作用。例如:Liu等^[27]利用叶绿体基因组数据对樟科(Lauraceae)25属133种的191份材料进行分类鉴定和系统发育研究,探讨了传统叶绿体条形码、专属条形码以及叶绿体全基因组条形码对樟科物种的分辨情况,发现叶绿体全基因组条形码可以鉴定约60%的樟科物种,这与传统叶绿体条形码和专属条形码的鉴定率相比有大幅度的提高,且叶绿体全基因组在纠正物种错误鉴定以及发现隐存种与新物种方面具有十分明显的优势。基于唐松草属12种植物及2个外类群种类叶绿体基因组的系统发育树显示:发育树所有节点的自展支持率均大于等于99%,唐松草属所有植物聚为一个分支,而尾囊草属(*Urophysa* Ulbr.)植物聚为另一个分支,其中,星毛唐松草与黏唐松草的亲缘关系最近。此外,*trnV-UAC*基因、*ndhF*和*ndhD*基因以及*ycf1*基因对应的3个核苷酸多态性热点区域变异度较高,可以为唐松草属植物的群体遗传学提供可靠的分子标记。表明叶绿体基因组在解决唐松草属植物的分类及鉴定方面是可靠的。

综上所述,星毛唐松草叶绿体基因组共有133个基因,AT含量明显高于GC含量,IR区的GC含量最高;叶绿体基因组的碱基组成偏向使用A或T。星毛唐松草及其4个近缘种的叶绿体基因组特征较为相似,非编码序列变异程度均明显高于编码序列,*ycf1*、*ndhF*、*rpl2*和*rps19*基因的收缩和扩张可能是唐松草属植物的叶绿体基因组IR区存在差异的主要原因。星毛唐松草与唐松草属其他近缘植物可聚为一支,*trnV-UAC*基因、*ndhF*和*ndhD*基因以及*ycf1*基因对应的区域为核苷酸多态性热点区域,叶绿体基因组可为唐松草属植物的分类及鉴定提供可靠依据。

参考文献:

- [1] 宋建平,陈雅研,高从元,等.星毛唐松草生物碱的研究(I)[J].中草药,1992,23(3):160.
- [2] 王文采.中国唐松草属植物[M].北京:北京大学出版社,2018:1258.
- [3] LUO D, LV N, ZHU L J, et al. Isoquinoline alkaloids from whole plants of *Thalictrum cirrhosum* and their antirotavirus activity[J]. Chemistry of Natural Compounds, 2020, 56(3): 504-508.
- [4] 薛晶晶,李建勇,李冰洁,等.2种唐松草属植物异喹啉类生物碱的研究[J].中国中药杂志,2022,47(10):2676-2680.
- [5] 赵儒楠,褚晓洁,刘维,等.鹅耳枥属树种叶绿体基因组结构及变异分析[J].南京林业大学学报(自然科学版),2021,45

- (2): 25–34.
- [6] 杨家鑫. 掌叶木和伞花木的叶绿体全基因组研究及其系统位置分析[D]. 贵阳: 贵州大学, 2019: 4–7.
- [7] BRANLANT C, KROL A, MACHATT M A, et al. Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs [J]. *Nucleic Acids Research*, 1981, 9(17): 4303–4324.
- [8] PALMER J D. Comparative organization of chloroplast genomes[J]. *Annual Review of Genetics*, 1985, 19: 325–354.
- [9] ZHOU J G, CHEN X L, CUI Y X, et al. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species [J]. *International Journal Molecular Sciences*, 2017, 18: 1839.
- [10] 毛立彦, 黄秋伟, 龙凌云, 等. 7种睡莲属植物叶绿体基因组密码子偏好性分析[J]. *西北林学院学报*, 2022, 37(2): 98–107.
- [11] 朱 斌, 甘晨晨, 王洪程. 球花石斛(*Dendrobium thyrsiflorum*) 叶绿体基因组特征及亲缘关系解析[J]. *生物技术通报*, 2021, 37(5): 38–47.
- [12] 梁湘兰, 郭 松. 苦马豆叶绿体基因组密码子偏好性分析[J]. *西北林学院学报*, 2022, 37(2): 121–126.
- [13] 侯 哲, 娄晓鸣, 李 昂, 等. 星毛唐松草叶绿体基因组的测序及其密码子偏好性分析[J]. *西南林业大学学报(自然科学)*, 2023, 43(3): 1–11.
- [14] TILLICH M, LEHWARK P, PELLIZZER T, et al. GeSeq-versatile and accurate annotation of organelle genomes[J]. *Nucleic Acids Research*, 2017, 45(1): W6–W11.
- [15] CHAN P P, LOWE T M. tRNAscan-SE: searching for tRNA genes in genomic sequences [J]. *Methods in Molecular Biology*, 2019, 1962: 1–14.
- [16] ROZAS J, FERRER-MATA A, SÁNCHEZ-DELBARRIO J C, et al. DnaSP 6: DNA Sequence Polymorphism analysis of large data sets [J]. *Molecular Biology Evolution*, 2017, 34(12): 3299–3302.
- [17] NAKAMURA T, YAMADA K D, TOMII K, et al. Parallelization of MAFFT for large-scale multiple sequence alignments [J]. *Bioinformatics*, 2018, 34(14): 2490–2492.
- [18] KUMAR S, STECHER G, TAMURA K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets [J]. *Molecular Biology Evolution*, 2016, 33(7): 1870–1874.
- [19] ZHU A, GUO W, GUPTA S, et al. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates [J]. *New Phytologist*, 2016, 209: 1747–1756.
- [20] HE Y F, WANG R N, GAI X Y, et al. The complete chloroplast genome of *Thalictrum baicalense* Turcz. ex Ledeb [J]. *Mitochondrial DNA Part B*, 2021, 6(2): 437–438.
- [21] ZHAI W, DUAN X S, ZHANG R, et al. Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae [J]. *Molecular Phylogenetics and Evolution*, 2019, 135: 12–21.
- [22] ŠMARDÁ P, BUREŠ P, HOROVÁ L, et al. Ecological and evolutionary significance of genomic GC content diversity in monocots [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(39): E4096–E4102.
- [23] WU L W, NIE L P, XU Z C, et al. Comparative and phylogenetic analysis of the complete chloroplast genomes of three *Paeonia* section *Moutan* species (Paeoniaceae) [J]. *Frontiers in Genetics*, 2020, 11: 980.
- [24] HAN C Y, DING R, ZONG X Y, et al. Structural characterization of *Platanthera ussuriensis* chloroplast genome and comparative analyses with other species of Orchidaceae [J]. *BMC Genomics*, 2022, 23: 84.
- [25] ZHENG G, WEI L, MA L L, et al. Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships [J]. *Plant Molecular Biology*, 2020, 102: 659–676.
- [26] XIE J B, QIAN K C, SI J N, et al. Conserved noncoding sequences conserve biological networks and influence genome evolution [J]. *Heredity*, 2018, 120: 437–451.
- [27] LIU Z F, MA H, CI X Q, et al. Can plastid genome sequencing be used for species identification in Lauraceae? [J]. *Botanical Journal of the Linnean Society*, 2021, 197: 1–14.

(责任编辑: 郭严冬)