

• 农业生物技术 •

蔓赤车叶绿体基因组结构及系统进化分析

颜丽, 杨雪莲*, 吴永飞, 王霞, 胡小京

贵州大学农学院园艺系, 贵州 贵阳 550025

颜丽, 杨雪莲, 吴永飞, 王霞, 胡小京. 蔓赤车叶绿体基因组结构及系统进化分析[J]. 生物工程学报, 2023, 39(7): 2914-2925.

YAN Li, YANG Xuelian, WU Yongfei, WANG Xia, HU Xiaojing. Chloroplast genomic characterization and phylogenetic analysis of *Pellionia scabra*[J]. Chinese Journal of Biotechnology, 2023, 39(7): 2914-2925.

摘要: 蔓赤车(*Pellionia scabra*)属于荨麻科赤车属,是一种具有高营养价值的优质野菜。本研究以蔓赤车为材料,基于高通量技术方法完成叶绿体基因组测序、组装注释、结构解析及构建系统发育树,以此深入研究蔓赤车叶绿体基因组特征。结果表明,蔓赤车叶绿体基因组大小为 153 220 bp, GC 含量为 36.4%,属于典型的四分体结构。共注释到 130 个基因,其中 85 个蛋白编码基因, 37 个转运 RNA 基因, 8 个核糖体 RNA 基因;其中,有 15 个基因包含 1 个内含子, 2 个基因包含 2 个内含子, *rps12* 存在反式剪接情况。蔓赤车叶绿体基因组可分为光合作用(43 个)、自我复制(64 个)、其他编码蛋白(7 个)以及未知功能(4 个)4 大类基因。蔓赤车叶绿体基因组共检测出 51 073 个密码子,其中编码亮氨酸(Leu)的密码子占比最大,密码子偏向使用 A 和 U 两种碱基。检测到 72 个简单重复序列位点,其中有 58 个单核苷酸、12 个二核苷酸、1 个三核苷酸和 1 个四核苷酸 4 种不同类型的简单重复序列(simple sequence repeats, SSRs)。蔓赤车 IRb/SSC 边界存在 *ycf1* 基因扩张现象。系统进化树显示,蔓赤车(*Pellionia scabra* OL800583)与庐山楼梯草(*Elatostema stewardii* MZ292972)、盘托楼梯草(*Elatostema dissectum* MK227819)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum* MN189961)亲缘关系最密切。基于蔓赤车叶绿体基因组的研究,旨在为蔓赤车物种鉴定、遗传进化及基因组学研究奠定理论基础。

关键词: 蔓赤车; 叶绿体基因组; 简单重复序列; 密码子偏好性; IR 区边界分析; 系统发育

资助项目: 国家自然科学基金(31860225); 贵州省人才项目(黔科合基础[2019]1408 号; 黔科合平台人才[2018]5781 号)

This work was supported by the National Natural Science Foundation of China (31860225) and Guizhou Provincial Talent Project (Qian Kehe Foundation [2019] No. 1408, Qian Kehe Platform Personnel [2018] No. 5781).

*Corresponding author. E-mail: yxl4684221@126.com

Received: 2022-10-14; Accepted: 2023-01-10; Published online: 2023-01-13

Chloroplast genomic characterization and phylogenetic analysis of *Pellionia scabra*

YAN Li, YANG Xuelian*, WU Yongfei, WANG Xia, HU Xiaojing

Department of Horticulture, College of Agriculture, Guizhou University, Guiyang 550025, Guizhou, China

Abstract: *Pellionia scabra* belongs to the genus *Pellionia* in the family of *Urticaceae*, and is a high-quality wild vegetables with high nutritional value. In this study, high-throughput techniques were used to sequence, assemble and annotate the chloroplast genome. We also analyzed its structure, and construct the phylogenetic trees from the *P. scabra* to further study the chloroplast genome characteristics. The results showed that the chloroplast genome size was 153 220 bp, and the GC content was 36.4%, which belonged to the typical tetrad structure in *P. scabra*. The chloroplast genome encodes 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes in *P. scabra*. Among them, 15 genes contained 1 intron, 2 genes contained 2 introns, and *rps12* had trans-splicing, respectively. In *P. scabra*, chloroplast genomes could be divided into four categories, including 43 photosynthesis, 64 self-replication, other 7 coding proteins, and 4 unknown functions. A total of 51 073 codons were detected in the chloroplast genome, among which the codon encoding leucine (Leu) accounted for the largest proportion, and the codon preferred to use A and U bases. There were 72 simple sequence repeats (SSRs) in the chloroplast genome of *P. scabra*, containing 58 single nucleotides, 12 dinucleotides, 1 trinucleotide, and 1 tetranucleotide. The *ycf1* gene expansion was present at the IRb/SSC boundary. The phylogenetic trees showed that *P. scabra* (OL800583) was most closely related to *Elatostema stewardii* (MZ292972), *Elatostema dissectum* (MK227819) and *Elatostema laevissimum* var. *laevissimum* (MN189961). Taken together, our results provide worthwhile information for understanding the identification, genetic evolution, and genomics research of *P. scabra* species.

Keywords: *Pellionia scabra*; chloroplast genome; simple sequence repeat; codon preference; boundary analysis of IR region; phylogeny

蔓赤车(*Pellionia scabra*), 又名岩苋菜^[1]、糙叶赤车使者^[2], 是荨麻科赤车属的多年生草本或亚灌木状植物, 常分布于我国华南、西南、华中、华东、台湾地区, 越南和日本亦有种植。蔓赤车具有较高的药用及食用价值, 据记载蔓赤车具有清热解毒、散瘀消肿、凉血止血之功效, 主治目赤肿痛、疔腮、蛇缠疮、牙痛、扭挫伤、妇女闭经、疮疖肿痛、烧烫伤、毒蛇咬伤、外伤出血等^[3]病症。此外, 蔓赤车营养价

值较高, 富含多种矿质元素和氨基酸, 是一种符合蔬菜营养标准可开发利用的优质野菜^[4]。迄今为止, 国内外蔓赤车的相关研究主要是针对其内含物及营养成分分析, 对蔓赤车叶绿体基因组相关分析的报道较少。

叶绿体是半自主性细胞器, 具有独立自主的遗传性, 可作为 DNA 条形码应用于植物系统发育^[5]、种群分析^[6]等研究, 是分子系统发育研究的重要资源。叶绿体基因组最常见的结构为

双链环状结构,且大多数是由2个反向重复序列(inverted repeat, IR)分别将大单拷贝区域(large single copy, LSC)和小单拷贝(small single copy, SSC)区域隔开的保守环状四分体结构。相较线粒体与核基因组,叶绿体在结构、基因数量和基因组成上更为保守,进化速度也相对适中^[7]。此外,叶绿体基因属于直系同源基因,且叶绿体基因组在被子植物中为母系遗传,不易发生重组,相对于核基因组,序列较为保守,便于组装,具有更易解析等特点^[8]。现已有薄壳山核桃^[9]、香花枇杷^[10]、橄榄^[11]、杧果^[12]、单叶铁线莲^[13]、连翘^[14]等多种植物叶绿体基因组特征分析的报道,为相关植物的物种鉴定、遗传改良等提供了重要的理论依据。本研究对蔓赤车的叶绿体进行测序和组装,分析其特征、分类和系统发育关系,旨在为该物种的叶绿体基因工程、分子遗传学研究、资源保护及开发利用奠定理论基础。

1 材料与方法

1.1 试验材料

蔓赤车植物样本来自于贵州省植物园(N 26°37'20", E 106°43'29"),取其新鲜、无病虫害的健康叶片放入取样袋,带回实验室于-80 °C 冷藏保存。

1.2 试验方法

1.2.1 蔓赤车 DNA 提取及测序

将 150 mg 新鲜叶片用液氮快速冷冻,研磨成粉后采用十六烷基三甲基溴化铵(cetyltrimethylammonium bromide, CTAB)法^[15]提取全基因组 DNA,经电泳检测和浓度测定合格后送至惠通生物有限公司检测,经检测合格后用超声波破碎仪处理,在序列的末端进行修复、3'端加上 A 尾、加测序接头,纯化、PCR 扩增后构建好完整的蔓赤车测序文库,质检合

格后,用 Illumina 高通量测序 Nova Seq 6000 进行测序,获取原始数据。

1.2.2 蔓赤车叶绿体全基因组的组装、注释

使用 NCBI 上已公布的蔓赤车近缘叶绿体数据与蛋白编码基因序列作为参考,采用 SPAdes 软件对测序数据进行基因组拼接^[16],分别用 BLASTn 与 Exonerate 进行比对分析,选择并确定候选序列组装结果。使用 PRICE 和 MITObim^[17]对目标序列进行延伸合并拼接。使用 Bowtie2^[18]将原始测序 reads 进行回帖,用软件 SPAdes 进行重拼接,直至完成环状基因组。叶绿体基因组序列在 PGA (plastid genome annotator)软件^[19]中注释并完成,检查修正信息后用软件 Chloroplot^[16]绘制图谱。蔓赤车叶绿体基因组已上传至 NCBI (<https://www.ncbi.nlm.nih.gov/>),审核后获得其登录号 OL800583。

1.2.3 蔓赤车叶绿体简单重复序列和密码子偏好性分析

利用 MISA (Micro Satellite Identification Tool)提供的 Perl 程序^[20]分析简单重复序列(simple sequence repeat, SSR),在 WINDOWS 命令窗口运行程序。参数设置中的单核苷酸最低重复次数阈值为 10,二核苷酸为 5,三、四、五、六核苷酸均为 4。Codon W1.4.2 软件^[21]计算密码子偏好性(relative synonymous codon usage, RSCU)。

1.2.4 蔓赤车叶绿体基因组边界分析

通过在线软件 IRscope (<http://irscope.shinyapps.io/iapp/>)分析蔓赤车和其 4 个近缘种庐山楼梯草(*Elatostema stewardii*)、盘托楼梯草(*Elatostema dissectum*)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum*)和深裂号角树(*Cecropia pachystachya*)的 IR 区的收缩和扩张情况^[22],并绘制对比图。

1.2.5 蔓赤车叶绿体基因组系统发育分析

本文多重比较和系统发育分析的蔓赤车近缘 28 个物种的叶绿体基因组下自 NCBI 数据库, 然后构建最大似然(maximum likelihood, ML)树。利用 MAFFT 7.037^[23]和 FFT-NS-2 将每个叶绿体基因组中的共有蛋白编码基因与蔓赤车的基因进行比对。运行 Model-finder var 1.6 获得最佳模型, 选择 TVM++F+I+G4 模型^[24]。使用 RAxML var 8.2.9 基于 ML 法, 参数设为 1 000 构建系统发育树^[25]。

2 结果与分析

2.1 蔓赤车叶绿体基因组结构基本特征

蔓赤车叶绿体基因组结构分成 4 个部分, 两个反向重复序列(IRA 与 IRB)、大单拷贝区(LSC)与小单拷贝区(SSC)(图 1)。蔓赤车的叶绿体基因组长度为 153 220 bp, 据统计总 GC 含量为 36.4%, 各区域长度及 GC 含量分别为: LSC 长 84 480 bp (39.1%)、SSC 长 17 568 bp (35.2%), IRA 和 IRB 长 25 586 bp (42.1%)。

蔓赤车叶绿体基因组平均测序深度为 746 X, 共编码 130 个基因。其中, 蛋白编码基因 85 个、核糖体 RNA (ribosomal RNA, rRNA) 基因数为 8 个、转运 RNA (transfer RNA, tRNA) 基因数为 37 个。蔓赤车叶绿体基因内含子分析发现(表 1), 大部分基因不具有内含子, 15 个基因含有 1 个内含子, 2 个基因(*clpP*、*ycf3*)含有两个内含子, 此外 1 个基因(*rps12*)存在反式剪接情况。根据注释结果将不同的蛋白编码基因功能大致分 4 类, 包括参与光合作用的基因(43 个), 参与自我复制的基因(64 个), 其他基因以及一些功能未知的基因(表 1)。

2.2 蔓赤车叶绿体基因组 SSR 位点分析

对蔓赤车叶绿体全基因组 SSRs 进行鉴定,

分析结果识别到 72 个 SSR 位点(表 2), 其中, 58 个单核苷酸, 重复占比最多(占比 80.56%), 包括 21 个 A、2 个 C、35 个 T; 双核苷酸总共 12 个(占比 16.67%), 包括 7 个 AT、4 个 TA 和 1 个 TC; 三核苷酸只有 1 个 AAT (占比 1.39%); 四核苷酸也只有 1 个 TTTA (占比 1.39%)。

2.3 蔓赤车叶绿体基因组密码子偏好分析

植物叶绿体基因组密码子偏好性的成因是不同密码子对应的 tRNA 在细胞里的丰度不一样, tRNA 丰度越高, 其对应的密码子的使用频率也会越高。测序结果表明(表 3), 蔓赤车叶绿体基因组检测出 51 073 个密码子, 其中编码亮氨酸(Leu)的密码子数量最多, 为 5 165 个, 占总密码子数的 10.11%; 色氨酸(Trp)的密码子数量最少, 为 672 个, 占总密码子数的 1.32%。蔓赤车叶绿体基因组密码子具有明显的偏好性, RSCU 等于 1 的氨基酸有 3 个, 分别为甲硫氨酸(Met)、色氨酸(Trp)及甘氨酸(Gly) GGU, RSCU 大于 1 的密码子种类有 31 个, 其中只有 3 个密码子碱基以 C 结尾, 2 个以 G 结尾, 剩余以 A/U 结尾。

2.4 蔓赤车叶绿体基因组边界分析

叶绿体基因组序列长度变化通常是受 IR 区域的某个基因收缩或扩张影响的。对蔓赤车与其 4 个近缘种庐山楼梯草(*Elatostema stewardii*)、盘托楼梯草(*Elatostema dissectum*)、光叶楼梯草(*Elatostema laevisissimum*)和深裂号角树(*Cecropia pachystachya*)叶绿体基因组进行边界比较分析, 结果表明荨麻科 5 种植物基因组组成和结构大致相似(图 2)。其中, 各物种都有 2 个基因位于 IR 区边界处, 分别是 LSC/IRb 的 *rps19* 和 IRa/LSC 的 *trnH*。荨麻科 5 种植物在基因组长度上 *ycf1* 和 *ndhF* 基因缺失以及 *rps19* 基因收缩等特征上存在差异, 如蔓赤车与光叶

楼梯草的叶绿体基因组长度相差 2 996 bp, 深裂号角树 IRa 和 IRb 区存在 *ycf1* 和 *ndhF* 基因缺失, 光叶楼梯草 IRb 区具有 *rsp19* 基因收缩,

仅有蔓赤车和深裂号角树 IRb 区存在 *ycf1* 基因的收缩和扩张, 仅深裂号角树 IRa 存在 *ndhF* 基因的收缩和扩张。

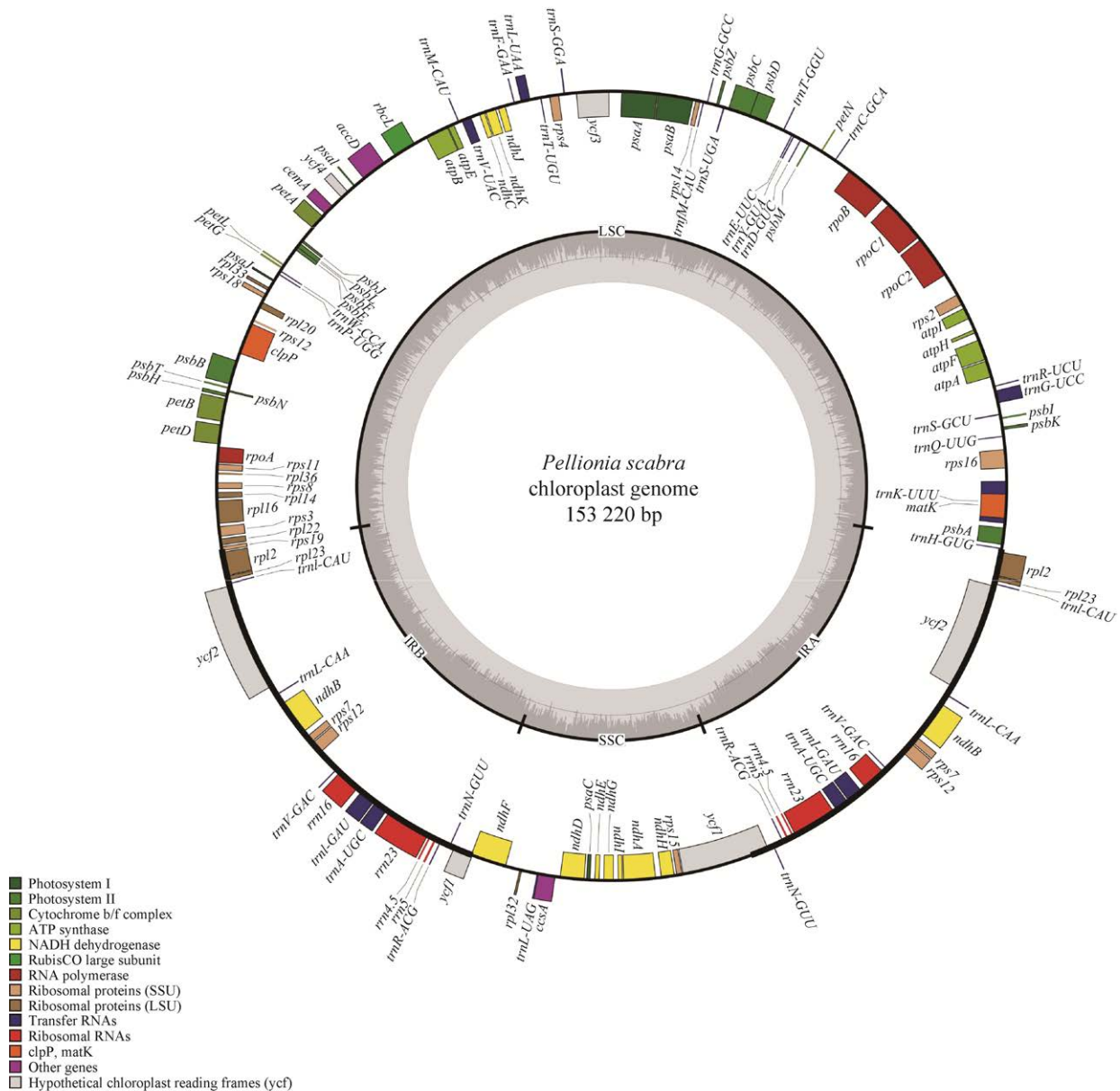


图 1 蔓赤车叶绿体基因组图谱 图的中心显示物种名称及长度; 基因组的 GC 含量和 AT 含量分别在 inner 圈以深灰色和浅灰色绘制; 不同功能的基因以不同的颜色编码

Figure 1 Map of the chloroplast genome of *Pellionia scabra*. The species name and length are depicted in the center of the plot; The GC content and AT content of the genome are plotted on dark and light gray in the inner circle, respectively; Genes with different functions are coded in different colors.

表 1 蔓赤车叶绿体基因组注释信息

Table 1 Gene annotation of the chloroplast genome of *Pellionia scabra*

| Gene category | Gene group | Gene name |
|------------------------------|------------------------------|--|
| Genes for photosynthesis | I Photosystem I | <i>psaA, psaB, psaC, psaI, psaJ</i> |
| | Photosystem II | <i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i> |
| | Cytochrome b/f complex | <i>petA, petB, petD, petG, petL, petN</i> |
| | ATP synthase | <i>atpA, atpB, atpE, atpF*, atpH, atpI</i> |
| | NADH dehydrogenase | <i>ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i> |
| Self replication | Ribosomal RNAs | <i>rrn4.5¹, rrn5¹, rrn16¹, rrn23¹</i> |
| | Ribosomal proteins (SSU) | <i>rps2, rps3, rps4, rps7, rps8, rps11, rps12[#], rps14, rps15, rps16, rps18, rps19</i> |
| | Ribosomal proteins (LSU) | <i>rpl2*, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i> |
| | RNA polymerase | <i>rpoA, rpoB, rpoC1*, rpoC2</i> |
| | Transfer RNAs | tRNA-Lys*, tRNA-Gln, tRNA-Ser, tRNA-Gly*, tRNA-Arg, tRNA-Cys, tRNA-Asp, tRNA-Tyr, tRNA-Glu, tRNA-Thr, tRNA-Ser, tRNA-Gly, tRNA-Met, tRNA-Ser, tRNA-Thr, tRNA-Leu, tRNA-Phe, tRNA-Val, tRNA-Met, tRNA-Trp, tRNA-Pro, tRNA-Ile, tRNA-Leu*, tRNA-Val*, tRNA-His, tRNA-Ile* ¹ , tRNA-Ala* ¹ , tRNA-Arg ¹ , tRNA-Asn ¹ , tRNA-Leu, tRNA-Asn, tRNA-Arg, tRNA-Ala, tRNA-Ile, tRNA-His |
| Other genes | Other genes | <i>accD, ccsA, cema, clpP**, matK, rbcL, infA</i> |
| Proteins of unknown function | Conserved open reading frame | <i>ycf1¹, ycf2, ycf3**, ycf4</i> |

*: Genes containing one intron; **: Genes containing two introns; #: *Trans*-spliced genes; ¹: Genes in IR region with two copies present.

表 2 蔓赤车叶绿体简单重复序列(SSR)信息

Table 2 The number of SSRs identified in the chloroplast genome of *Pellionia scabra*

| Repeat unit | Type | Number | Total (%) |
|-------------|------|--------|-----------|
| 1 | A | 21 | 80.56 |
| | C | 2 | |
| | T | 35 | |
| 2 | AT | 7 | 16.67 |
| | TA | 4 | |
| | TC | 1 | |
| 3 | AAT | 1 | 1.39 |
| 4 | TTTA | 1 | 1.39 |

2.5 蔓赤车叶绿体基因组系统进化树分析

叶绿体基因组的聚类分析, 对植物发育进化研究具有重要意义。本研究以青柠为外类群,

基于荨麻科 28 个物种叶绿体基因组构建发育树表明, 28 种荨麻科植物聚在 1 个分支上, 说明荨麻科之间关系亲近。蔓赤车(*Pellionia scabra* OL800583)与庐山楼梯草(*Elatostema stewardii* MZ292972)、盘托楼梯草(*Elatostema dissectum* MK227819)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum* MN189961)聚到一支上, 自展值为 100, 说明蔓赤车(*Pellionia scabra* OL800583)与庐山楼梯草(*Elatostema stewardii* MZ292972)、盘托楼梯草(*Elatostema dissectum* MK227819)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum* MN189961)亲缘关系最近(图 3)。

表 3 蔓赤车叶绿体基因组各氨基酸及相对密码子偏好情况

Table 3 Statistics of amino acid and relative codon bias in *Pellionia scabra* chloroplast genome

| Amino acid | Codon | Number | RSCU | Ratio (%) | Amino acid | Codon | Number | RSCU | Ratio (%) |
|------------|-------|--------|------|-----------|------------|-------|--------|------|-----------|
| Phe | UUU | 2 401 | 1.26 | 7.49 | Ser | UCU | 1 135 | 1.47 | 6.61 |
| | UUC | 1 425 | 0.74 | | | UCC | 840 | 1.09 | |
| Leu | UUA | 1 220 | 1.42 | 10.11 | UCA | 855 | 1.11 | 4.34 | |
| | UUG | 1 056 | 1.23 | | UCG | 546 | 0.71 | | |
| | CUU | 1 083 | 1.26 | | Pro | CCU | 607 | | 1.10 |
| | CUC | 595 | 0.69 | | CCC | 585 | 1.06 | | |
| | CUA | 771 | 0.90 | | CCA | 665 | 1.20 | | |
| | CUG | 440 | 0.51 | | CCG | 360 | 0.65 | | |
| Ile | AUU | 1 781 | 1.23 | 8.50 | Thr | ACU | 670 | 1.20 | 4.38 |
| | AUC | 1 004 | 0.69 | | ACC | 567 | 1.01 | | |
| Met | AUA | 1 558 | 1.08 | 1.67 | ACA | 640 | 1.14 | 2.83 | |
| | AUG | 852 | 1.00 | | ACG | 360 | 0.64 | | |
| Val | GUU | 737 | 1.28 | 4.51 | Ala | GCU | 437 | 1.21 | 2.24 |
| | GUC | 434 | 0.75 | | GCC | 348 | 0.96 | | |
| | GUA | 737 | 1.28 | | GCA | 412 | 1.14 | | |
| | GUG | 394 | 0.68 | | GCG | 250 | 0.69 | | |
| Tyr | UAU | 1 552 | 1.38 | 4.41 | Cys | UGU | 678 | 1.19 | 1.83 |
| | UAC | 698 | 0.62 | | UGC | 466 | 0.81 | | |
| TER | UAA | 1 278 | 1.27 | 4.07 | TER | UGA | 936 | 0.93 | 1.32 |
| | UAG | 801 | 0.80 | | Trp | UGG | 672 | 1.00 | |
| His | CAU | 916 | 1.39 | 2.57 | Arg | CGU | 355 | 0.69 | 2.91 |
| | CAC | 399 | 0.61 | | CGC | 243 | 0.47 | | |
| Gln | CAA | 1 036 | 1.41 | 2.88 | CGA | 520 | 1.00 | 2.46 | |
| | CAG | 436 | 0.59 | | CGG | 370 | 0.71 | | |
| Asn | AAU | 1 826 | 1.40 | 5.12 | Ser | AGU | 728 | 0.94 | 3.17 |
| | AAC | 788 | 0.60 | | AGC | 528 | 0.68 | | |
| Lys | AAA | 2 165 | 1.37 | 6.18 | Arg | AGA | 1 022 | 1.97 | 4.13 |
| | AAG | 990 | 0.63 | | AGG | 598 | 1.15 | | |
| Asp | GAU | 944 | 1.43 | 2.59 | Gly | GGU | 526 | 1.00 | 4.13 |
| | GAC | 377 | 0.57 | | GGC | 358 | 0.68 | | |
| Glu | GAA | 1 298 | 1.38 | 3.68 | GGA | 724 | 1.37 | 0.95 | |
| | GAG | 581 | 0.62 | | GGG | 499 | 0.95 | | |

3 讨论

叶绿体作为植物光合作用的细胞器,还参与了植物氨基酸、蛋白质等的生物合成,拥有独立完整的基因组。被子植物质体 DNA 一般是母系遗传,进化过程中不会出现基因重组的现

象,因而通过对其序列结构组成、特征及变异分析,可以很好地揭示物种系统发育过程^[26]。

植物叶绿体基因组序列长度介于 $(1.20 \times 10^5 - 1.80 \times 10^5)$ bp,编码基因数为 100–130 个^[27],结构为双链环状 DNA 分子结构,由 4 部分构成,包含 LSC、SSC 和 2 个 IR,其中 2 个 IR 区序列

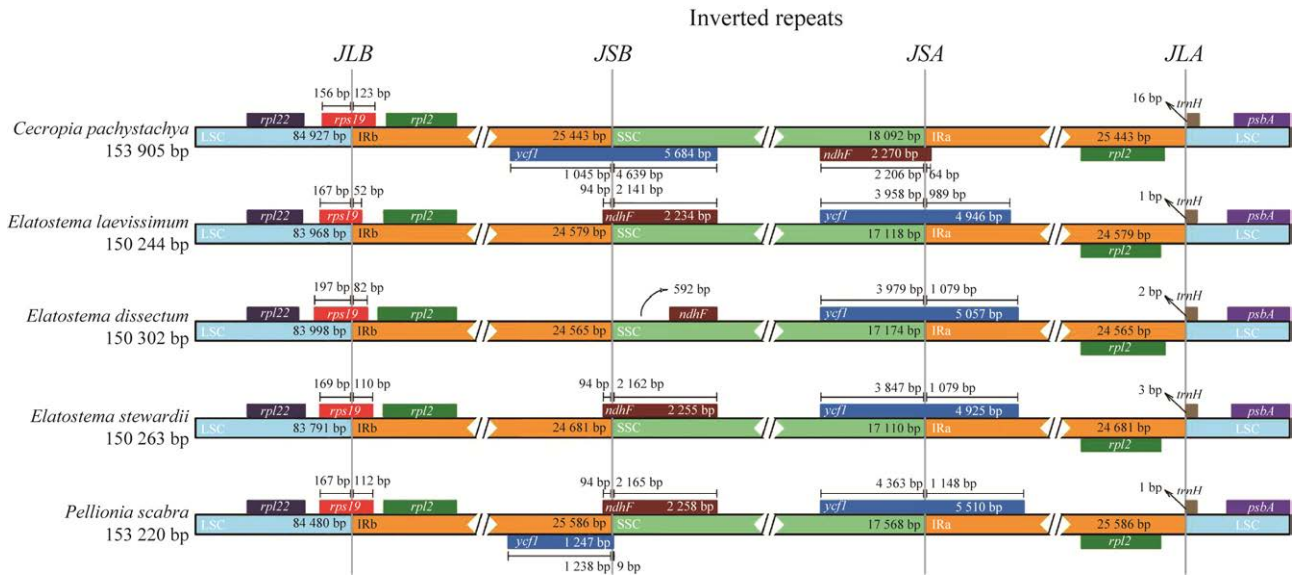


图 2 荨麻科 5 种植物的叶绿体基因组反向重复序列的连接位点 不同颜色表示叶绿体基因组不同结构区域及基因类型；JLB (IRb/LSC)、JSB (IRb/SSC)、JSA (SSC/IRa)和 JLA (IRa/LSC)表示基因组上每个相应的两个区域之间的连接位点

Figure 2 Junction sites of the inverted repeats of chloroplast genomes from five *Urticaceae* plants. Different colors indicate different structural regions and genes types of chloroplast genomes; JLB (IRb/LSC), JSB (IRb/SSC), JSA (SSC/IRa) and JLA (IRa/LSC) denote the junction sites between each corresponding two regions on the genome.

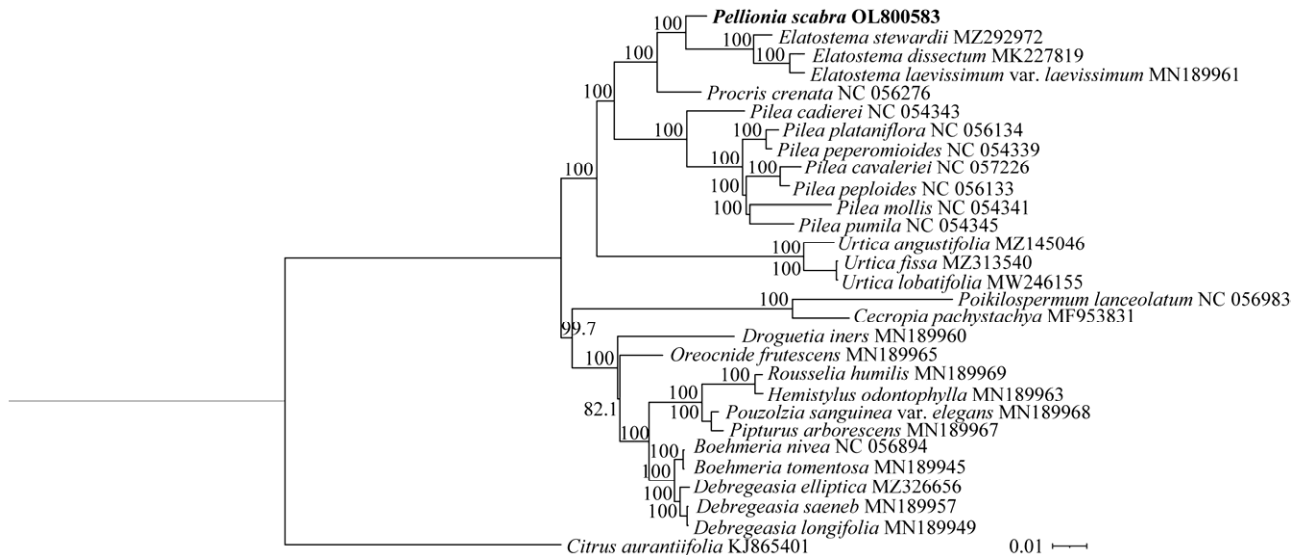


图 3 蔓赤车和 28 个物种完整叶绿体基因组序列的系统发育关系进化树 图中尺标代表进化距离，支上的数字为 1 000 次自展值支持率(%)，蔓赤车位置加粗标出

Figure 3 Cladogram illustrating the phylogenetic relationships of *Pellionia scabra* and 28 species based on complete chloroplast genome sequences. The scale represents the evolutionary distance, the 1 000 bootstrap value (%) was shown above the branch, and the position of *Pellionia scabra* is in bold.

相同,方向相反^[28]。本研究的蔓赤车叶绿体基因组长度为 153 220 bp,基因总数为 130 个,是典型的四分体结构,可见蔓赤车叶绿体基因组大小和结构与上述被子植物研究结果基本相符,说明蔓赤车叶绿体基因组较为恒定,保守性较高。蔓赤车的叶绿体基因组功能注释的 118 个基因主要分为光合作用(43 个)、自我复制(64 个)、其他编码蛋白(7 个)以及未知功能(4 个) 4 大类基因,这与大部分物种的划分一致^[29]。

密码子偏好性可在一定程度上反映基因乃至物种的起源和进化方式,并对基因功能及其编码蛋白表达有一定影响^[30]。本研究对蔓赤车叶绿体基因组共检测出 51 073 个密码子,其中编码亮氨酸(Leu)的密码子占比最大,密码子偏向使用 A 和 U 两种碱基,该结果与滇重楼^[31]、沙枣^[32]、花椒^[33]等叶绿体基因密码子偏好性相似。这可能是由于突变和自然选择等诸多因素导致,陆奇丰等^[34]在梧桐研究中证实自然选择对密码子偏好性有影响,其他影响因素有待进一步研究。

简单重复序列(SSR)是整个基因组中 1–6 bp 的重复序列,也是高等真核生物基因组的重要组成部分,在植物物种鉴定、群体遗传多态性和遗传分析研究中具有重要作用^[29]。本研究表明,蔓赤车叶绿体基因组中检测到 58 个单核苷酸、12 个双核苷酸、1 个三核苷酸和 1 个四核苷酸 SSRs。其中,双核苷酸 SSRs 的重复单元以 AT/TA/TC 为主,这与前人报道的其他被子植物叶绿体基因组的 SSRs 序列构成一致^[35],进而验证了叶绿体基因组的简单重复序列主要由短 poly A 和 poly T 构成,不是由 C 或 G 串联重复^[36]。叶绿体基因组的 IR 区域的扩张与收缩,是自然进化中的一般现象^[37]。本研究发现,荨麻科多数物种叶绿体基因组位于 IRb/SSC 边

界重叠区的 *ycf1* 和 *ndhF* 这 2 个基因,还具有不同程度的扩张与收缩现象;盘托楼梯草和深裂号角叶绿体基因组的 IRb/SSC 边界无基因重叠,蔓赤车、深裂号角树、光叶楼梯草在 IRb/IRa 区有基因收缩和扩张的现象。说明荨麻科植物 LSC/IRb、IRb/SSC、SSC/IRa 和 IRa/LSC 的边界变化主要是受边界扩张与收缩变化机制的影响^[38]。

叶绿体基因组能有效进行植物物种鉴定和系统亲缘关系分析,尤其在近缘物种间具有更理想的效果^[39]。本研究基于叶绿体基因组序列库,对青柠这一外类群及荨麻科 28 个物种系统发育树的构建,发现荨麻科植物是一个单系类群;蔓赤车(*Pellionia scabra* OL800583)与庐山楼梯草(*Elatostema stewardii* MZ292972)、盘托楼梯草(*Elatostema dissectum* MK227819)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum* MN189961)属于姊妹群,具有最近的亲缘关系,并且自展值为 100%,这为以后蔓赤车在荨麻科系统位置和亲缘关系确定、遗传多样性分析和群体遗传结构研究提供了理论依据,为该物种鉴定及进一步探索个体水平的遗传差异分析提供理论基础。

4 结论

本研究通过高通量测序法对蔓赤车的完整叶绿体基因组进行组装和注释,蔓赤车叶绿体基因组全长为 153 220 bp (85 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因),属于典型的四分体结构;其绝大多数的基因与光合作用及基因转录翻译相关;共 72 个 SSR 位点,以单碱基(80.56%)重复为主,偏向使用 A 和 U 两种碱基;蔓赤车叶绿体基因组的 IRb/SSC 边界存在 *ycf1* 基因扩张现象;蔓赤车(*Pellionia scabra* OL800583)与庐山楼梯草(*Elatostema stewardii*

MZ292972)、盘托楼梯草(*Elatostema dissectum* MK227819)、光叶楼梯草(*Elatostema laevissimum* var. *laevissimum* MN189961)亲缘关系最近。该研究为荨麻科物种分子标记开发、种间鉴定、该物种的演化过程、个体水平的遗传差异分析和遗传改良提供理论依据。

REFERENCES

- [1] 刘克明. 湖南植物志[M]. 长沙: 湖南科学技术出版社, 2001, (2): 289-290.
LIU KM. Flora of Hunan[M]. Changsha: Hunan Science and Technology Press, 2001, (2): 289-290 (in Chinese).
- [2] YANG YP, SHIH BL, LIU H. Flora of Taiwan[M]. 2nd ed. Taipei: The Editorial Committee of the Flora of Taiwan, 1996, 2: 231.
- [3] 吴贻谷. 中华本草[M]. 上海: 上海科学技术出版社, 1999.
WU YG. Chinese Materia Medica[M]. Shanghai: Shanghai Science & Technology Publishers, 1999 (in Chinese).
- [4] 王蕾, 廖文波, 邱文智. 野菜资源蔓赤车(*Pellionia scabra*)营养成分研究[J]. 中山大学学报(自然科学版), 2013, 52(6): 119-123.
WANG L, LIAO WB, QIU WZ. Nutritious components of the wild vegetable *Pellionia scabra* from south China[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2013, 52(6): 119-123 (in Chinese).
- [5] CHEN SL, YAO H, HAN JP, LIU C, SONG JY, SHI LC, ZHU YJ, MA XY, GAO T, PANG XH, LUO K, LI Y, LI XW, JIA XC, LIN YL, LEON C. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species[J]. PLoS One, 2010, 5(1): e8613.
- [6] ZHA X, WANG XY, LI JR, GAO F, ZHOU YJ. Complete chloroplast genome of *Sophora alopecuroides* (Papilionoideae): molecular structures, comparative genome analysis and phylogenetic analysis[J]. Journal of Genetics, 2020, 99(1): 13.
- [7] DONG WP, XU C, CHENG T, LIN K, ZHOU SL. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales[J]. Genome Biology and Evolution, 2013, 5(5): 989-997.
- [8] 李倩, 郭其强, 高超, 李慧娥. 贵州威宁红花油茶
- 叶绿体基因组特征分析[J]. 园艺学报, 2020, 47(4): 779-787.
- LI Q, GUO QQ, GAO C, LI HE. Characterization of complete chloroplast genome of *Camellia weiningensis* in Weining, Guizhou Province[J]. Acta Horticulturae Sinica, 2020, 47(4): 779-787 (in Chinese).
- [9] 蒋思思, 袁军, 周文君, 钮根花, 周俊琴. 薄壳山核桃(*Carya illinoensis*)叶绿体基因组及其特征分析[J]. 园艺学报, 2022, 49(8): 1772-1784.
JIANG SS, YUAN J, ZHOU WJ, NIU GH, ZHOU JQ. Complete chloroplast genome sequence and characteristics analysis of *Carya illinoensis*[J]. Acta Horticulturae Sinica, 2022, 49(8): 1772-1784 (in Chinese).
- [10] 屈亚亚, 辛静, 冯发玉, 董章宏, 瞿绍宏, 王海洋, 李斌, 辛培尧. 香花枇杷质体基因组序列密码子偏好性分析[J]. 西北林学院学报, 2021, 36(4): 138-144, 158.
QU YY, XIN J, FENG FY, DONG ZH, QU SH, WANG HY, LI B, XIN PY. Codon usage bias in chloroplast genome of *Eriobotrya fragrans* champ. ex Benth[J]. Journal of Northwest Forestry University, 2021, 36(4): 138-144, 158 (in Chinese).
- [11] 赖瑞联, 陈瑾, 冯新, 覃振师, 陈义挺, 沈朝贵, 田奇琳, 吴如健. 橄榄叶绿体基因组密码子偏好性特征[J]. 福建农林大学学报(自然科学版), 2022, 51(4): 502-509.
LAI RL, CHEN J, FENG X, QIN ZS, CHEN YT, SHEN CG, TIAN QL, WU RJ. Codon usage preference of chloroplast genome of *Canarium album*[J]. Journal of Fujian Agriculture and Forestry University (Natural Science Edition), 2022, 51(4): 502-509 (in Chinese).
- [12] 辛雅萱, 黎若竹, 李鑫, 陈丽琼, 唐军荣, 屈亚亚, 杨琳懿, 辛培尧, 李云芳. 杠果叶绿体基因组密码子使用偏好性分析[J]. 中南林业科技大学学报, 2021, 41(9): 148-156, 165.
XIN YX, LI RZ, LI X, CHEN LQ, TANG JR, QU YY, YANG LY, XIN PY, LI YF. Analysis on codon usage bias of chloroplast genome in *Mangifera indica*[J]. Journal of Central South University of Forestry & Technology, 2021, 41(9): 148-156, 165 (in Chinese).
- [13] 韦海忠, 潘丽芹, 田盛野, 汤紫依, 何海叶, 张慧娟, 蒋明. 单叶铁线莲叶绿体基因组的序列特征与系统发育分析[J]. 中草药, 2022, 53(12): 3766-3773.
WEI HZ, PAN LQ, TIAN SY, TANG ZY, HE HY, ZHANG HJ, JIANG M. Chloroplast genome sequence characterization and phylogenetic analysis of *Clematis henryi*[J]. Chinese Traditional and Herbal Drugs, 2022,

- 53(12): 3766-3773 (in Chinese).
- [14] 宋芸, 贾孟君, 曹亚萍, 李政, 贺嘉欣, 王勇飞, 张鑫瑞, 乔永刚. 连翘叶绿体基因组特征分析[J]. 园艺学报, 2022, 49(1): 187-199.
SONG Y, JIA MJ, CAO YP, LI Z, HE JX, WANG YF, ZHANG XR, QIAO YG. Analysis on chloroplast genomic characteristics of *Forsythia suspensa*[J]. Acta Horticulturae Sinica, 2022, 49(1): 187-199 (in Chinese).
- [15] TAI TH, TANKSLEY SD. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue[J]. Plant Molecular Biology Reporter, 1990, 8(4): 297-303.
- [16] ZHENG SY, POCZAI P, HYVÖNEN J, TANG J, AMIRYOUSEFI A. Chloroplot: an online program for the versatile plotting of organelle genomes[J]. Frontiers in Genetics, 2020, 11: 576124.
- [17] DUAN YZ, SHEN YH, KANG FR, WANG JW. Characterization of the complete chloroplast genomes of the endangered shrub species *Prunus mongolica* and *Prunus pedunculata* (Rosales: Rosaceae)[J]. Conservation Genetics Resources, 2019, 11(3): 249-252.
- [18] YAO Z, YOU FM, N'DIAYE A, KNOX RE, MCCARTNEY C, HIEBERT CW, POZNIAK C, XU W. Evaluation of variant calling tools for large plant genome re-sequencing[J]. BMC Bioinformatics, 2020, 21(1): 360.
- [19] QU XJ, MOORE MJ, LI DZ, YI TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes[J]. Plant Methods, 2019, 15: 50.
- [20] THIEL T, MICHALEK W, VARSHNEY R, GRANER A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)[J]. Theoretical and Applied Genetics, 2003, 106(3): 411-422.
- [21] SHIELDS DC, SHARP PM. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases[J]. Nucleic Acids Research, 1987, 15(19): 8023-8040.
- [22] AMIRYOUSEFI A, HYVÖNEN J, POCZAI P. IRscope: an online program to visualize the junction sites of chloroplast genomes[J]. Bioinformatics, 2018, 34(17): 3030-3031.
- [23] KATOH K, ROZEWICKI J, YAMADA KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization[J]. Briefings in Bioinformatics, 2019, 20(4): 1160-1166.
- [24] KALYAANAMOORTHY S, MINH BQ, WONG TKF, von HAESLER A, JERMIIN LS. ModelFinder: fast model selection for accurate phylogenetic estimates[J]. Nature Methods, 2017, 14(6): 587-589.
- [25] GUINDON S, DUFAYARD JF, LEFORT V, ANISIMOVA M, HORDIJK W, GASCUEL O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0[J]. Systematic Biology, 2010, 59(3): 307-321.
- [26] 张庆滢, 陈璇, 郭孟璧, 郭蓉, 许艳萍, 杨明, 郭鸿彦. 野生大麻叶绿体基因组分子多态标记的筛选与开发[J]. 分子植物育种, 2017, 15(3): 979-985.
ZHANG QY, CHEN X, GUO MB, GUO R, XU YP, YANG M, GUO HY. Screening and development of chloroplast polymorphic molecular markers on wild hemp (*Cannabis sativa* L.)[J]. Molecular Plant Breeding, 2017, 15(3): 979-985 (in Chinese).
- [27] ZHANG TW, FANG YJ, WANG XM, DENG X, ZHANG XW, HU SN, YU J. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes[J]. PLoS One, 2012, 7(1): e30531.
- [28] SHINOZAKI K, OHME M, TANAKA M, WAKASUGI T, HAYASHIDA N, MATSUBAYASHI T, ZAITA N, CHUNWONGSE J, OBOKATA J, YAMAGUCHI-SHINOZAKI K, OHTO C, TORAZAWA K, MENG BY, SUGITA M, DENO H, KAMOGASHIRA T, YAMADA K, KUSUDA J, TAKAIWA F, KATO A, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression[J]. The EMBO Journal, 1986, 5(9): 2043-2049.
- [29] 苏丹丹, 刘玉萍, 刘涛, 郑长远, 张雨, 王亚男, 秦娜, 苏旭. 苦马豆叶绿体基因组结构及其特征分析[J]. 植物研究, 2022, 42(3): 446-454.
SU DD, LIU YP, LIU T, ZHENG CY, ZHANG Y, WANG YN, QIN N, SU X. Structure of chloroplast genome and its characteristics of *Sphaerophysa salsula*[J]. Bulletin of Botanical Research, 2022, 42(3): 446-454 (in Chinese).
- [30] TILLICH M, LEHWARK P, PELLIZZER T, ULBRICHT-JONES ES, FISCHER A, BOCK R, GREINER S. GeSeq-versatile and accurate annotation of organelle genomes[J]. Nucleic Acids Research, 2017, 45(W1): W6-W11.
- [31] 王宇飞, 江媛, 杨成金, 王婧, 徐志超, 段宝忠. 滇重楼叶绿体基因组特征及密码子偏好性分析[J]. 分

- 子植物育种, 2021, 19(22): 7448-7458.
- WANG YF, JIANG Y, YANG CJ, WANG J, XU ZC, DUAN BZ. Analysis of chloroplast genome characteristic and codon usage bias in *Paris polyphylla* var. *yunnanensis*[J]. Molecular Plant Breeding, 2021, 19(22): 7448-7458 (in Chinese).
- [32] 王婧, 王天翼, 王罗云, 张建国, 曾艳飞. 沙枣叶绿体全基因组序列及其使用密码子偏性分析[J]. 西北植物学报, 2019, 39(9): 1559-1572.
- WANG J, WANG TY, WANG LY, ZHANG JG, ZENG YF. Assembling and analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias[J]. Acta Botanica Boreali-Occidentalia Sinica, 2019, 39(9): 1559-1572 (in Chinese).
- [33] 王超, 赵懿琛, 赵德刚. 朝仓花椒转录组密码子偏好性分析[J]. 分子植物育种, 2020, 18(21): 6989-6996.
- WANG C, ZHAO YC, ZHAO DG. Preference analysis of transcriptome codon in *Zanthoxylum piperitum* DC. var. *inerme makino*[J]. Molecular Plant Breeding, 2020, 18(21): 6989-6996 (in Chinese).
- [34] 陆奇丰, 骆文华, 黄至欢. 两种梧桐叶绿体基因组密码子使用偏性分析[J]. 广西植物, 2020, 40(2): 173-183.
- LU QF, LUO WH, HUANG ZH. Codon usage bias of chloroplast genome from two species of *Firmiana marsili*[J]. Guihaia, 2020, 40(2): 173-183 (in Chinese).
- [35] KUANG DY, WU H, WANG YL, GAO LM, ZHANG SZ, LU L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics[J]. Genome, 2011, 54(8): 663-673.
- [36] TONTI-FILIPPINI J, NEVILL PG, DIXON K, SMALL I. What can we do with 1 000 plastid genomes?[J]. The Plant Journal, 2017, 90(4): 808-818.
- [37] HANSEN DR, DASTIDAR SG, CAI ZQ, PENAFLORE C, KUEHL JV, BOORE JL, JANSEN RK. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae)[J]. Molecular Phylogenetics and Evolution, 2007, 45(2): 547-563.
- [38] KANE NC, CRONK Q. Botany without borders: barcoding in focus[J]. Molecular Ecology, 2008, 17(24): 5175-5176.
- [39] ZHOU JG, CHEN XL, CUI YX, SUN W, LI YH, WANG Y, SONG JY, YAO H. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species[J]. International Journal of Molecular Sciences, 2017, 18(9): 1839.

(本文责编 郝丽芳)