

DOI: 10.11931/guihaia.gxzw202312023

王其刚, 曹世睿, 王慧纯, 等, 2025. 中甸刺玫的叶绿体基因组特征及种内变异 [J]. 广西植物, 45(1): 15–30.

WANG QG, CAO SR, WANG HC, et al., 2025. Chloroplast genome features and intraspecific variation of *Rosa praelucens* [J]. *Guihaia*, 45(1): 15–30.



中甸刺玫的叶绿体基因组特征及种内变异

王其刚¹, 曹世睿^{1,2}, 王慧纯¹, 马长乐², 晏慧君¹, 邱显钦¹, 景维坤¹, 蹇洪英^{1*}

(1. 云南省农业科学院花卉研究所, 昆明 650205; 2. 西南林业大学 园林园艺学院, 昆明 650224)

摘要: 中甸刺玫 (*Rosa praelucens*) 是云南香格里拉市的特有“极危”植物和国家二级重点保护植物, 也是著名的高山花卉和重要的十倍体月季种质资源, 种内存在丰富的表型多样性。为了澄清中甸刺玫种内表型变异的遗传背景, 该文利用二代测序技术对 40 个不同表型的中甸刺玫代表性个体的叶绿体基因组进行测序、组装注释和比较分析。结果表明: (1) 中甸刺玫的基因组序列长 157 173~157 261 bp, 植株间仅相差 88 bp, 共编码 132 个功能基因, 主要为与光合作用和自我复制相关的基因。全部基因共由 27 155 个密码子编码, 以 A-和 U-为末端的密码子较常见。(2) 中甸刺玫的叶绿体基因组共鉴定出 36 个重复序列和 73 个简单重复序列 (SSRs), 后者大部分为单核苷酸 SSRs, 主要位于大单拷贝 (LSC) 区的基因间隔区。(3) 中甸刺玫种内叶绿体全基因组的单倍型多样性为 0.928 ± 0.027 , 核酸多态性 (P_i) 为 0.000 12; 位于 LSC 的 *petN-trnD*、*psaA-ycf3* 等基因间隔区, 以及 *rps16* 和 *ycf1* 等基因的核酸多态性相对较高; 不同表型的代表性个体的叶绿体基因组间在结构上不存在大片段或基因的逆转或者丢失。该研究结果表明, 中甸刺玫种内在叶绿体基因组大小、序列和结构等方面均高度保守, 其种内丰富的表型多样性并非由叶绿体基因组变异而引起。

关键词: 中甸刺玫, 叶绿体基因组, 比较基因组, 简单重复序列, 核酸多态性, 密码子偏好

中图分类号: Q943 **文献标识码:** A **文章编号:** 1000-3142(2025)01-0015-16

Chloroplast genome features and intraspecific variation of *Rosa praelucens*

WANG Qigang¹, CAO Shirui^{1,2}, WANG Huichun¹, MA Changle²,
YAN Huijun¹, QIU Xianqin¹, JING Weikun¹, JIAN Hongying^{1*}

(1. Flower Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China; 2. College of Landscape and Horticulture, Southwest Forestry University, Kunming 650224, China)

Abstract: *Rosa praelucens* is a critically endangered alpine wild flower endemic to Shangri-La City of Yunnan Province. Rich in phenotypic diversity and with a high ploidy level of decaploid, *R. praelucens* is a very important rose germplasm resource. In order to clarify the genetic background of its phenotypic variation, the chloroplast genomes of 40

收稿日期: 2024-02-07 接受日期: 2024-03-05

基金项目: 国家自然科学基金 (31972443); 云南省高层次科技人才及创新团队选拔专项 (202305AS350002)。

第一作者: 王其刚 (1977—), 硕士, 研究员, 研究方向为月季遗传育种, (E-mail) 171068976@qq.com。

*通信作者: 蹇洪英, 博士, 研究员, 研究方向为月季种质资源研究与利用, (E-mail) ynwildflower@aliyun.com。

R. praelucens individual plants representing different phenotypes within the species were sequenced by using the Illumina HiSeq 2000 platform, and then assembled, annotated, compared and analyzed. The results were as follows: (1) Chloroplast genomes of *R. praelucens* were 157 173 – 157 261 bp in length, with a size difference of 88 bp among different individual plants. The genomes encoded 132 function genes, mainly related with photosynthesis and self-replication. A total of 27 155 codons, preferring using codon ending of A or U, were found in all the coding sequences. (2) A total of 36 repeats and 73 simple sequence repeats (SSRs) were detected in the chloroplast genome of *R. praelucens*. Most of the SSRs were mononucleotide type and located in the intergenic region of large single scale (LSC). (3) The haplotype diversity among the 40 chloroplast genomes was 0.928 ± 0.027 , and the nucleotide polymorphism (P_i) was 0.000 12. The intergenic region of *petN-trnD* and *psaA-ycf3*, gene *rps16* and *ycf1* were relatively more divergent. No reverse or loss of large DNA fragments and genes were found among the chloroplast genomes of different individuals. These results indicate that the chloroplast genomes are highly conserved in size, sequence and structure within *R. praelucens*. The rich intraspecific phenotypic diversity is not caused by the variation of chloroplast genomes among different individual plants.

Key words: *Rosa praelucens*, chloroplast genome, comparative genome, simple sequence repeats (SSRs), nucleotide polymorphism (P_i), codon preference

叶绿体在植物的生活史中具有重要作用 (Wicke et al., 2011)。大多数维管植物的叶绿体基因组长度约为 150 kb, 为保守的四分体结构, 包括一个大单拷贝区 (large single copy, LSC)、一个小单拷贝区 (small single copy, SSC) 和两个反向重复区 (inverted repeats, IRs) (Wicke et al., 2011; Shetty et al., 2016; Zhu et al., 2016)。高等植物的叶绿体基因组高度保守, 但某些类群的叶绿体基因组中存在大片段的倒置 (Sun et al., 2017)、大量重复序列 (Guisinger et al., 2011)、基因丢失或假基因化 (Ye et al., 2018) 以及 IR 区的扩张或收缩 (Li et al., 2017; Liu et al., 2017)。叶绿体基因组在大多数被子植物中为母系遗传 (Neale & Sederoff, 1989; Daniell et al., 2016)。与核基因组相比, 叶绿体基因组具有分子量低、结构简单、保守性强等特点, 还包含大量的重复序列, 包括简单重复序列 (simple sequence repeats, SSRs) (Cavalier-Smith, 2002)。因此, 被广泛应用于系统发育、DNA 条形码、基因工程和亲缘关系等研究 (Dong et al., 2018)。随着二代测序技术 (next-generation sequencing, NGS) 的发展, 越来越多的植物叶绿体全基因组序列被报道, NCBI 数据库迄今已公布了超过 8 500 个植物叶绿体基因组。

中甸刺玫 (*Rosa praelucens*) 是云南省香格里拉市的特有“极危”植物 (Ku & Robertson, 2003; 覃海宁等, 2017), 也是国家二级保护植物 ([http://www. forestry. gov. cn/main/3954/20210908/](http://www.forestry.gov.cn/main/3954/20210908/)

163949170374051.html)。中甸刺玫是著名的高山花卉 (Li & Zhou, 2005) 和重要的月季种质资源, 既耐寒 (邓菊庆等, 2013), 也高抗蚜虫 (范元兰等, 2021)。自 Jian 等 (2010) 发现其是蔷薇属野生种中唯一的最高倍性——十倍体 ($2n = 10x = 70$) 以来, 人们对中甸刺玫的生境和群落特征 (关文灵等, 2012)、繁育系统 (伍翔宇等, 2014)、种群现状 (周玉泉等, 2016)、系统位置 (王开锦等, 2018)、基于染色体荧光原位杂交的核型特征 (方桥等, 2020) 以及遗传多样性和遗传结构 (Jian et al., 2018a) 等进行了系统研究, 发现中甸刺玫种内存在丰富的表型多样性, 其中花色和花型变异尤其显著 (李树发等, 2013; Jian et al., 2018a)。

理解多倍性如何修饰表型性状是进化生物学的一个研究热点和主要目标 (Balao et al., 2011)。大量研究表明自然形成或人工诱导的多倍体植物都会产生遗传和表观遗传的改变, 从而改变基因的表达, 使其在遗传、生理和形态上产生分化, 形成新的表型 (Ramsey & Schemske, 2002)。对于十倍体的中甸刺玫而言, 其种内丰富的表型变异的机制尚不清楚。中甸刺玫的高倍性特征, 限制了多种分子技术手段在其遗传背景研究上的应用。Jian 等 (2017) 报道了中甸刺玫的叶绿体基因组大小、各分区的长度和 GC 含量、编码的基因数量等基本信息, 发现中甸刺玫的叶绿体基因组大小为 157 186 bp, 与同属的单瓣月季花 (*R. chinensis* var. *spontanea*) 等其他植物相比, 其叶绿体基因组

最大且在 LSC 区的 *psbM* 和 *trnD* 之间有一个长为 505 bp 的插入。在此基础上,为了探究中甸刺玫种内表型变异的遗传背景,本研究利用二代测序技术对 40 个不同表型的中甸刺玫代表性个体的叶绿体基因组进行了测序、组装和比较分析,探讨以下问题:(1)中甸刺玫叶绿体基因组的序列特征和密码子偏好性;(2)中甸刺玫种内不同表型个体在叶绿体基因组上是否存在较大的变异。以期为中甸刺玫的物种形成和保护提供更多的遗传信息,也为探讨其种内表型变异的分子机制提供叶绿体基因组方面的基础数据。

1 材料与方 法

1.1 研究材料

用于叶绿体基因组序列特征分析的中甸刺玫植株来源于香格里拉市小中甸镇和平村塘安培组(99°49'38.1" E, 27°32'16.68" N, 3 248 m),其原始序列已上传至 NCBI,序列号为 MG450565.1。其余 40 个不同表型的代表性植株的基本信息见表 1。于 2021 年 6 月底在野外采集当年生成健康叶片,立即用硅胶进行干燥, -18 °C 低温保存用于后续实验。

1.2 研究方法

1.2.1 基因组总 DNA 的提取、测序、组装及注释 使用改良的 CTAB 法进行中甸刺玫叶片的总 DNA 提取,达到建库测序要求的 DNA 送到北京诺禾致源科技股份有限公司用 Illumina Hiseq 2000 测序平台进行建库测序,每个样品得到约 3.5 Gb 的 150 bp 短片段原始序列(raw data),用 NGSQC Toolkit_v2.3.3 软件(Patel & Jain, 2012)按照默认参数对原始序列进行过滤筛选,得到高质量的有效序列(clean data)。使用 GetOrganelle (<https://github.com/Kinggerm/GetOrganelle>)进行 *de novo* 从头组装,得到叶绿体全基因组序列。将组装好的基因组序列使用 CpGAVAS(Liu et al., 2012)自动进行注释,用 Genious 9.1(Kearse et al., 2012)进行校对和调整每个注释基因的边界区域,采用 OGDRAW(Lohse et al., 2013)绘制叶绿体基因组物理图谱。

1.2.2 叶绿体基因组结构分析 利用 Geneious 软件对已上传到 NCBI、序列号为 MG450565.1 的中甸刺玫叶绿体全基因组进行叶绿体全基因组编码基因构成统计。用 MEGA6(Tamura et al., 2013)

进行密码子偏好分析,计算同义密码子相对使用值(relative synonymous codon usage values; RSCU)并统计 AT 含量。用重复序列分析软件 REPuter 软件(<https://bibiserv.cebitec.uni-bielefeld.de/reputer>)(Kurtz et al., 2001),搜索基因组中的正向重复(forward repeats)和反向重复(reverse repeats)序列。软件运行时,设置搜索的重复序列长度不小于 20 bp,序列一致性大于 85%。此外,利用 MISA(Beier et al., 2017)软件来鉴定简单重复序列,搜索的阈值设置为单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复次数分别不小于 10、5、4、3、3 和 3。

1.2.3 种内不同个体的叶绿体基因组序列比较

在 Geneious 软件中调用 Mauve 程序(Darling et al., 2004)对中甸刺玫 40 个代表性植株的叶绿体基因组进行比对,分析不同植株的叶绿体基因组间是否存在大片段的逆转或丢失。用 DNASP v5.10 软件(Librado & Rozas, 2009)计算种内叶绿体基因组的单倍型多样性和核苷酸多态性(P_i),筛选叶绿体基因组中的高变区。

2 结果与分析

2.1 中甸刺玫叶绿体基因组的结构特征

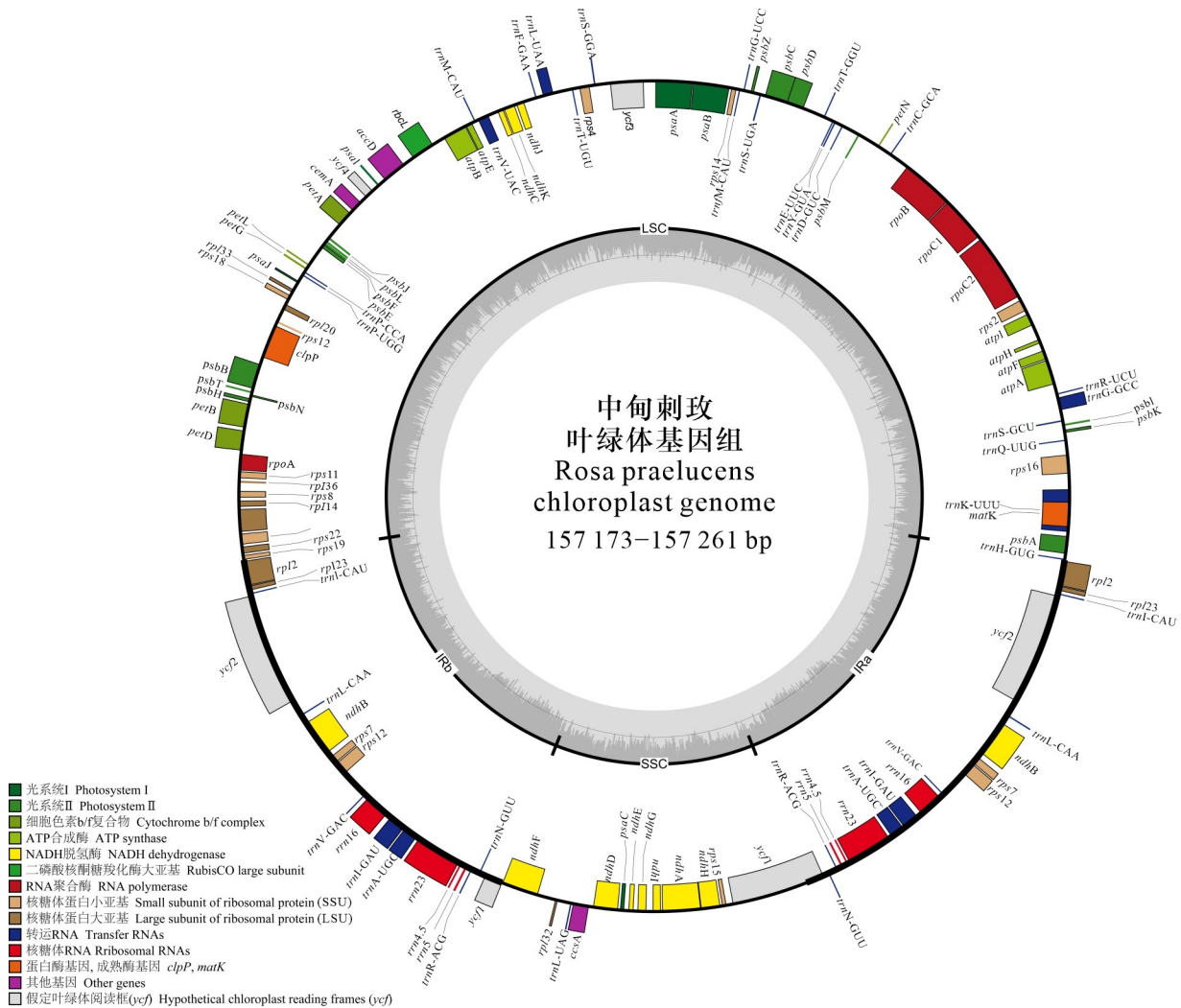
中甸刺玫种内不同表型的 40 个代表性植株的叶绿体基因组序列的长度、各分区长度、GC 含量、编码基因数目等基本信息见图 1 和表 2。中甸刺玫的基因组序列全长 157 173~157 261 bp,植株间相差 88 bp。基因组最大的是 7-1 号植株,为 157 261 bp,基因组最小的是 2-5 号植株,为 157 173 bp。LSC 区长为 86 300~86 353 bp,相差 53 bp,最长的是 7-1 号植株,最短的是 2-5 号植株;SSC 区长为 18 765~18 803 bp,相差 38 bp;反向重复 IR 区长度均为 26 054 bp,说明种内基因组大小的差异主要来源于 LSC 区和 SSC 区。基因组的 GC 含量在不同个体间无显著差异,全基因组的 GC 含量均为 37.2%,其中 IR 区的 GC 含量为 42.7%,LSC 区的 GC 含量为 35.2%,SSC 区的 GC 含量为 31.2%。

2.2 中甸刺玫叶绿体基因组的基因构成

由表 3 可知,中甸刺玫的叶绿体基因组共编码 132 个功能基因,包括 87 个蛋白质编码基因,37 个 tRNA 基因和 8 个 rRNA 基因。其中,与光合作

表 1 不同表型的中甸刺玫代表性植株的基本信息
Table 1 Basic information of different phenotypes of representative *Rosa praelucens* plants

| 编号 Code | 表型特征 Phenotypic characteristics | | | 位置 Location |
|------------|--|----------------------------|------------------------|---|
| | 花色 Flower color | 花直径 Flower size (cm) | 花瓣数 Petal number | |
| 0-1 | 白色 White | 12.5 | 5 | 格桑公司苗圃 Nursery of Gesang Flower Company |
| 0-2 | 白色 White | 10 | 5 | 格桑公司苗圃 Nursery of Gesang Flower Company |
| 0-3 | 粉白色 Pinkish white | 8~8.5 | 5~7 | 布呵谷老村 Buhegulao Village |
| 0-4 | 粉白色 Pinkish white | 6~7 | 5~6 | 布呵谷老村 Buhegulao Village |
| 1-1 | 玫红色 Rose red | 11 | 5~8 | 塘安培村 Tang'anpei Village |
| 1-2 | 粉红色 Pinkish red | 10~12 | 5~7 | 塘安培村 Tang'anpei Village |
| 1-3 | 粉红色 Pinkish red | 10~13 | 5~7 | 基公村 Jigong Village |
| 1-4 | 粉红色 Pinkish red | 10~12 | 5~7 | 基公村 Jigong Village |
| 1-5 | 粉红色 Pinkish red | 10 | 5~7 | 托木南村 Tuomunan Village |
| 2-1 | 玫红色 Rose red | 7~10.5 | 5 | 塘培村 Tangpei Village |
| 2-2 | 粉红色 Pinkish red | 6.5~9 | 5 | 塘培村 Tangpei Village |
| 2-3 | 粉红色 Pinkish red | 8~10 | 5 | 塘培村 Tangpei Village |
| 2-4 | 粉红色 Pinkish red | 8~10 | 5 | 塘培村 Tangpei Village |
| 2-5 | 粉红色 Pinkish red | 8~10 | 5 | 塘培村 Tangpei Village |
| 2-6 | 深粉红色 Dark pink | 7~10 | 5 | 塘培村 Tangpei Village |
| 2-7 | 桃红色 Crimson red | 9~11 | 5~11 | 塘培村 Tangpei Village |
| 3-1 | 深桃红芯白 Deep crimson red with white core | 8 | 5 | 热水塘村 Reshuitang Village |
| 3-2 | 深桃红芯白 Deep crimson red with white core | 7~10 | 5 | 碧古村 Bigu Village |
| 3-3 | 深桃红芯白 Deep crimson red with white core | 7~10 | 5 | 碧古村 Bigu Village |
| 4-1 | 浅粉红色 Light pink | 8.5~11 | 5 | 胡批村 Hupi Village |
| 4-2 | 粉色 Pink | 8~10 | 5 | 胡批村 Hupi Village |
| 4-3 | 粉色 Pink | 8~10 | 5 | 胡批村 Hupi Village |
| 4-4 | 粉色 Pink | 8~10 | 5~7 | 胡批村 Hupi Village |
| 5-1 | 粉红色 Reddish pink | 8~11 | 5 | 胡批村 Hupi Village |
| 5-2 | 粉红色 Reddish pink | 10~11 | 5 | 胡批村 Hupi Village |
| 6-1 | 深粉红色 Dark pink | 9~10 | 5 | 布呵谷老村 Buhegulao Village |
| 6-2 | 深粉红色 Dark pink | 7~10 | 5 | 布呵谷老村 Buhegulao Village |
| 6-3 | 深粉红色 Dark pink | 10~12 | 5 | 布呵谷老村 Buhegulao Village |
| 6-4 | 深粉红色 Dark pink | 9~11 | 5 | 布呵谷老村 Buhegulao Village |
| 6-5 | 深粉红色 Dark pink | 9~12 | 5 | 布呵谷老村 Buhegulao Village |
| 7-1 | 粉红色 Reddish pink | 9~11 | 5 | 乃司村 Naisi Village |
| 7-2 | 粉红色 Reddish pink | 9~11 | 5 | 乃司村 Naisi Village |
| 8-1 | 淡粉色 Light pink | 7~8 | 5 | 布呵谷老村 Buhegulao Village |
| 8-2 | 淡粉色 Light pink | 8~10 | 5~8 | 布呵谷老村 Buhegulao Village |
| 8-3 | 粉白色 Pinkish white | 11 | 5 | 布呵谷老村 Buhegulao Village |
| 8-4 | 粉色 Pink | 7~10 | 5~7 | 布呵谷老村 Buhegulao Village |
| 8-5 | 粉色 Pink | 7~10 | 5~7 | 布呵谷老村 Buhegulao Village |
| 8-6 | 粉白色 Pinkish white | 10 | 5 | 布呵谷老村 Buhegulao Village |
| 9-1 | 淡粉色 Light pink | 9~12 | 5 | 桃木谷村 Taomugu Village |
| 9-2 | 淡粉色 Light pink | 10~12 | 5 | 桃木谷村 Taomugu Village |



环内基因为正向转录, 环外基因为反向转录。

Genes inside the circle are forward transcribed, while those outside was reversely transcribed.

图 1 中甸刺玫叶绿体基因组图

Fig. 1 Chloroplast genome map of *Rosa praelucens*

用有关的基因有 45 个, 与自我复制相关的基因有 76 个, 功能未知的其他基因共 11 个。6 个蛋白编码基因(*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *ycf2*)、7 个 tRNAs (*trnA*-UGC、*trnI*-CAU、*trnI*-GAU、*trnL*-CAA、*trnN*-GUU、*trnR*-ACG、*trnV*-GAC) 和 4 个 rRNAs (*rrn16*、*rrn23*、*rrn4.5*、*rrn5*) 在 IR 区完全重复。在 132 个基因中, *petB*、*petD*、*ndhA*、*ndhB*、*rps16*、*rpl2*、*rpl16*、*rpoC1*、*trnA*-UGC、*trnI*-GAU、*trnK*-UUU、*trnL*-UAA、*trnV*-UAC 等 13 个基因有 1 个内含子, *ycf3* 和 *clpP* 这 2 个基因各有 2 个内含子, *rps12* 为反式剪接的基因, 其 5'端在 LSC 区, 而 3'端在 IR 区重复。

2.3 中甸刺玫叶绿体基因组密码子偏好性

中甸刺玫的叶绿体基因组密码子使用频率 (RSCU) 如表 4。全部基因由 27 155 个密码子编码, 其中亮氨酸 (leucine) 是使用频率最高的氨基酸, 共编码了其中的 2 765 个密码子, 占总数的 10.85%, 而组氨酸 (histidine) 是使用频率最低的氨基酸, 仅编码了其中的 530 个密码子, 占总数的 1.95%。以 A-和 U-为末端的密码子较常见, 除了 *trnL*-CAA、*trnS*-GGA、精氨酸 Arg-AGG 和甘氨酸 Gly-GGG 以外, 所有的首选同义密码子 (RSCU>1) 都是以 A 或 U 结尾。

表 2 中甸刺玫种内代表性植株的叶绿体基因组特征
Table 2 Characteristics of the chloroplast genomes of representative *Rosa praelucens* plants

| 编号 Code | 长度 Length (bp) | | | | GC 含量 GC content (%) | | | |
|------------|--------------------|--------------|--------------|-------------|-------------------------|--------------|--------------|-------------|
| | 总长 Total length | 大单拷贝区 LSC | 小单拷贝区 SSC | 反向重复区 IR | 全基因组 Total | 大单拷贝区 LSC | 小单拷贝区 SSC | 反向重复区 IR |
| 0-1 | 157 184 | 86 311 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 0-2 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 0-3 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 0-4 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 1-1 | 157 188 | 86 315 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 1-2 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 1-3 | 157 187 | 86 314 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 1-4 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 1-5 | 157 185 | 86 312 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-1 | 157 177 | 86 304 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-2 | 157 180 | 86 307 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-3 | 157 177 | 86 304 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-4 | 157 178 | 86 305 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-5 | 157 173 | 86 300 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-6 | 157 179 | 86 306 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 2-7 | 157 179 | 86 306 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 3-1 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 3-2 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 3-3 | 157 231 | 86 320 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 3-4 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 4-1 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 4-2 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 4-3 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 5-1 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 5-2 | 157 186 | 86 313 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 6-1 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 6-2 | 157 229 | 86 318 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 6-3 | 157 229 | 86 318 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 6-4 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 6-5 | 157 229 | 86 318 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 7-1 | 157 261 | 86 353 | 18 800 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 7-2 | 157 185 | 86 312 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-1 | 157 229 | 86 318 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-2 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-3 | 157 230 | 86 319 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-4 | 157 229 | 86 318 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-5 | 157 185 | 86 312 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 8-6 | 157 231 | 86 320 | 18 803 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 9-1 | 157 213 | 86 340 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |
| 9-2 | 157 188 | 86 315 | 18 765 | 26 054 | 37.2 | 35.2 | 31.2 | 42.7 |

表 3 中甸刺玫叶绿体基因组的基因构成
Table 3 Gene composition of *Rose praelucens* chloroplast genomes

| 功能 Function | 基因分组 Group of gene | 基因 Gene |
|--|--|--|
| 光合作用 Photosynthesis | ATP 合成酶 ATP synthase | <i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i> |
| | 细胞色素复合体 Cytochrome b/f complex | <i>petA</i> , <i>petB</i> *, <i>petD</i> *, <i>petG</i> , <i>petL</i> , <i>petN</i> |
| | NADH 合成酶 NADH dehydrogenase | <i>ndhA</i> *, <i>ndhB</i> ***, <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i> |
| | 光系统 I Photosystem I | <i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> |
| | 光系统 II Photosystem II | <i>psbB</i> , <i>psbT</i> , <i>psbH</i> , <i>psbN</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbF</i> , <i>psbE</i> , <i>psbZ</i> , <i>psbC</i> , <i>psbD</i> , <i>psbM</i> , <i>psbI</i> , <i>psbK</i> , <i>psbA</i> |
| | 二磷酸核酮糖氧化酶大亚基 RubisCO large subunit | <i>rbcL</i> |
| | 自我复制 Self-replication | 核糖体蛋白小亚基 Small subunit of ribosomal protein (SSU) |
| 核糖体蛋白大亚基 Large subunit of ribosomal protein (LSU) | | <i>rpl2</i> ***, <i>rpl14</i> , <i>rpl16</i> *, <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> **, <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i> |
| RNA 聚合酶 RNA polymerase | | <i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i> |
| 转运 RNA Transfer RNAs | | <i>trnA</i> -UGC***, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -UCC, <i>trnG</i> -GCC*, <i>trnH</i> -GUG, <i>trnI</i> -CAU**, <i>trnI</i> -GAU***, <i>trnK</i> -UUU*, <i>trnS</i> -GCU, <i>trnL</i> -CAA**, <i>trnL</i> -UAA*, <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU**, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -UCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnR</i> -ACG**, <i>trnT</i> -UGU, <i>trnT</i> -GGU, <i>trnV</i> -UAC*, <i>trnW</i> -CCA, <i>trnV</i> -GAC** |
| 其他 Others | 核糖体 RNA Ribosomal rRNAs | <i>rrn16</i> **, <i>rrn23</i> **, <i>rrn4.5</i> **, <i>rrn5</i> ** |
| | 假定叶绿体阅读框 Hypothetical chloroplast reading frames (<i>ycf</i>) | <i>ycf1</i> , <i>ycf2</i> **, <i>ycf3</i> ***, <i>ycf4</i> |
| | 乙酰-CoA-羧化酶亚基 Acetyl-CoA-carboxylase subunit | <i>accD</i> |
| | 蛋白酶基因 <i>clpP</i> | <i>clpP</i> ** |
| | 包膜蛋白 Envelope protein | <i>cemA</i> |
| | c 型细胞色素合成 Type c cytochrome synthesis | <i>ccsA</i> |
| | 成熟酶基因 <i>matK</i> | <i>matK</i> |
| 转录起始因子 Transcription initiation factor | <i>infA</i> | |

注: * 表示有 1 个内含子; ** 表示有 2 个内含子; # 表示反式剪接; * 表示在 IR 区有两个反向重复拷贝。

Note: * indicates genes with one intron; ** indicates genes with two introns; # indicates trans-spliced gene; * indicates genes with two copies in the IR region.

2.4 中甸刺玫叶绿体基因组的重复序列

重复序列分析结果表明, 中甸刺玫的叶绿体基因组中共有 33 个正向重复序列和 3 个反向重复

序列, 这些重复序列长度大多为 20~30 bp(表 5)。最长的重复序列分别位于 *rps12-trnV* (GAC) 基因间隔区和 *ndhA* 基因的含子区域。大多数重复序

表 4 中甸刺玫的密码子-反密码子识别模式及密码子偏好

Table 4 Condon-anticodon recognition patterns and codon reference of *Rosa praelucens* chloroplast genomes

| 氨基酸 Amino acid | 密码子 Codon | 数量 Count | 相对同义密码子 RSCU | 氨基酸 Amino acid | 密码子 Codon | 数量 Count | 相对同义密码子 RSCU |
|-------------------|--------------|-------------|-----------------|-------------------|--------------|-------------|-----------------|
| 苯丙氨酸 Phe | UUU(F) | 1 092 | 1.21 | 丝氨酸 Ser | UCU(S) | 504 | 1.26 |
| | UUC(F) | 706 | 0.79 | | UCC(S) | 422 | 1.06 |
| 亮氨酸 Leu | UUA(L) | 597 | 1.3 | | UCA(S) | 477 | 1.19 |
| | UUG(L) | 619 | 1.34 | UCG(S) | 342 | 0.86 | |
| | CUU(L) | 556 | 1.21 | 脯氨酸 Pro | CCU(P) | 262 | 1.1 |
| | CUC(L) | 319 | 0.69 | | CCC(P) | 207 | 0.87 |
| | CUA(L) | 393 | 0.85 | | CCA(P) | 272 | 1.14 |
| | CUG(L) | 281 | 0.61 | | CCG(P) | 211 | 0.89 |
| 异亮氨酸 Ile | AUU(I) | 925 | 1.35 | 苏氨酸 Thr | ACU(T) | 352 | 1.14 |
| | AUC(I) | 517 | 0.75 | | ACC(T) | 280 | 0.91 |
| | AUA(I) | 615 | 0.9 | | ACA(T) | 375 | 1.22 |
| 甲硫氨酸 Met | AUG(M) | 571 | 1 | | ACG(T) | 223 | 0.73 |
| 缬氨酸 Val | GUU(V) | 421 | 1.35 | 丙氨酸 Ala | GCU(A) | 239 | 1.24 |
| | GUC(V) | 208 | 0.67 | | GCC(A) | 176 | 0.91 |
| | GUA(V) | 372 | 1.2 | | GCA(A) | 223 | 1.16 |
| | GUG(V) | 244 | 0.78 | | GCG(A) | 134 | 0.69 |
| 酪氨酸 Tyr | UAU(Y) | 800 | 1.33 | 半胱氨酸 Cys | UGU(C) | 382 | 1.13 |
| | UAC(Y) | 405 | 0.67 | | UGC(C) | 292 | 0.87 |
| 终止子 Stop | UAA(*) | 488 | 1.07 | 终止子 Stop | UGA(*) | 445 | 0.97 |
| | UAG(*) | 439 | 0.96 | 色氨酸 Trp | UGG(W) | 542 | 1 |
| 组氨酸 His | CAU(H) | 352 | 1.33 | 精氨酸 Arg | CGU(R) | 218 | 0.72 |
| | CAC(H) | 178 | 0.67 | | CGC(R) | 141 | 0.46 |
| 谷氨酰胺 Gln | CAA(Q) | 547 | 1.32 | | CGA(R) | 304 | 1 |
| | CAG(Q) | 283 | 0.68 | CGG(R) | 261 | 0.86 | |
| 天冬酰胺 Asn | AAU(N) | 957 | 1.34 | 丝氨酸 Ser | AGU(S) | 362 | 0.91 |
| | AAC(N) | 470 | 0.66 | AGC(S) | 292 | 0.73 | |
| 离氨酸 Lys | AAA(K) | 1 015 | 1.26 | 精氨酸 Arg | AGA(R) | 526 | 1.73 |
| | AAG(K) | 590 | 0.74 | | AGG(R) | 370 | 1.22 |
| 天冬氨酸 Asp | GAU(D) | 628 | 1.46 | 甘氨酸 Gly | GGU(G) | 350 | 0.98 |
| | GAC(D) | 235 | 0.54 | | GGC(G) | 224 | 0.63 |
| 谷氨酸 Glu | GAA(E) | 691 | 1.3 | | GGA(G) | 458 | 1.28 |
| | GAG(E) | 375 | 0.7 | | GGG(G) | 400 | 1.12 |

列位于 LSC 区和 IR 区,还有 9 个重复序列在不同的区域开始,如第 1 号重复的 2 个序列分别开始于 IRb 区和 SSC 区。

2.5 中甸刺玫叶绿体基因组的简单重复序列

MISA 软件在中甸刺玫的叶绿体基因组中共找到 73 个简单重复序列(SSRs),其中单核苷酸 SSRs

表 5 中甸刺玫叶绿体基因组的重复序列
Table 5 Repeat sequences in *Rosa praelucens* chloroplast genomes

| 编号 Code | 重复第 1 个 开始处 Repeat start 1 | 重复类型 Repeat type | 重复长度 Size (bp) | 重复第 2 个 开始处 Repeat start 2 | 失配碱基 Mismatch (bp) | 所在具体位置 Location | 所在区域 Region |
|------------|----------------------------------|---------------------|----------------------|----------------------------------|--------------------------|------------------------------------|----------------|
| 1 | 100770 | F | 40 | 122818 | 0 | IGS; <i>ndhA</i> | IRb, SSC |
| 2 | 44997 | F | 39 | 100772 | 0 | <i>yef3</i> ; IGS | LSC, IRb |
| 3 | 44997 | F | 38 | 122820 | 0 | <i>yef3</i> ; <i>ndhA</i> | LSC, SSC |
| 4 | 59254 | F | 34 | 59285 | 0 | IGS | LSC |
| 5 | 123010 | R | 37 | 123023 | -3 | <i>ndhA</i> | SSC |
| 6 | 109739 | F | 32 | 109771 | -2 | IGS | IRb |
| 7 | 133696 | F | 32 | 133728 | -2 | IGS | IRa |
| 8 | 67436 | F | 25 | 67460 | 0 | IGS | LSC |
| 9 | 9822 | F | 27 | 37714 | -1 | <i>trnG</i> -GCC, <i>trnG</i> -UCC | LSC |
| 10 | 8380 | F | 29 | 36640 | -2 | <i>trnS</i> -GCU, <i>trnS</i> -UGA | LSC, IRb |
| 11 | 91108 | F | 29 | 91129 | -2 | <i>yef2</i> | IRb |
| 12 | 152341 | F | 29 | 152362 | -2 | <i>yef2</i> | IRa |
| 13 | 51039 | F | 22 | 51060 | 0 | IGS | LSC |
| 14 | 43196 | F | 25 | 147743 | -1 | IGS | LSC, IRa |
| 15 | 58610 | F | 25 | 58635 | -1 | <i>rbcL</i> ; IGS | LSC |
| 16 | 39962 | F | 30 | 42186 | -3 | <i>psaB</i> ; <i>psaA</i> | LSC |
| 17 | 10499 | F | 21 | 10520 | 0 | IGS | LSC |
| 18 | 8733 | F | 29 | 65172 | -3 | IGS | LSC |
| 19 | 83661 | F | 29 | 122808 | -3 | <i>rpl16</i> , <i>ndhA</i> | LSC, SSC |
| 20 | 10623 | F | 20 | 10647 | 0 | IGS | LSC |
| 21 | 13570 | F | 20 | 90104 | 0 | IGS, <i>yef2</i> | LSC, IRb |
| 22 | 28582 | F | 20 | 28602 | 0 | IGS | LSC |
| 23 | 33042 | R | 20 | 86327 | 0 | IGS | LSC, IRb |
| 24 | 67511 | F | 20 | 67530 | 0 | IGS | LSC |
| 25 | 120144 | F | 20 | 120164 | 0 | IGS | SSC |
| 26 | 133705 | F | 23 | 133737 | -1 | IGS | IRa |
| 27 | 8669 | F | 28 | 31214 | -3 | IGS | LSC |
| 28 | 12223 | F | 28 | 72420 | -3 | IGS, <i>clpP</i> | LSC |
| 29 | 83667 | F | 28 | 122814 | -3 | <i>rpl16</i> , <i>ndhA</i> | LSC, SSC |
| 30 | 7598 | R | 25 | 61186 | -2 | IGS | LSC |
| 31 | 83670 | F | 25 | 100769 | -2 | <i>rpl16</i> , IGS | IRb |
| 32 | 37582 | F | 22 | 37604 | -1 | IGS | LSC |
| 33 | 43764 | F | 22 | 73530 | -1 | IGS, <i>clpP</i> | LSC |
| 34 | 45662 | F | 22 | 142499 | -1 | <i>yef3</i> ; IGS | IRa |
| 35 | 109749 | F | 22 | 109781 | -1 | IGS | IRb |
| 36 | 5203 | F | 24 | 127726 | -2 | IGS; <i>yef1</i> | SSC |

注: F 表示正向重复; R 表示反向重复。

Note: F indicates forward repeat; R indicates reverse repeat.

(A/T/G/C)最多,共有 42 个;其次是二核苷酸类型(AG/AT/TA/TC),有 9 个,三核苷酸类型有 4 个,四核苷酸类型有 8 个,六核苷酸 SSRs 有 2 个,没有五核苷酸 SSRs。绝大部分为单纯类型 SSRs,复合类型的 SSRs 较少,二者分别为 65 个和 8 个,没有间接型 SSRs(表 6)。58 个 SSRs 位于 LSC 区,占全部 SSRs 的 79.5%,5 个 SSRs 位于 SSC 区,IRa 和 IRb 区各有 5 个 SSRs。只有 23 个 SSRs 位于基因中,其他均位于基因间隔区。单核苷酸 SSRs 中的 74%属于 A/T,这与 SSRs 主要由短腺嘌呤(adenine, A)或胸腺嘧啶(thymine, T)重复组成而很少有串联鸟嘌呤(guanine, G)和胞嘧啶(cytosine, C)组成的假说相一致。

2.6 中甸刺玫种内的叶绿体基因组序列差异

中甸刺玫种内不同个体的叶绿体基因组序列差异较小。所有代表性个体的全基因组序列比对分析共检测到 58 个变异位点共 22 个单倍型,单倍型多样性为 0.928 ± 0.027 ,核酸多态性为 0.000 12。种内叶绿体基因组在基因、基因间隔区的核酸多态性都较低,相对多态性较大的是位于 LSC 区的 *psbI-trnS* (GCU)、*trnS* (GCU)-*trnG* (UCC)、*trnG* (UCC)-*trnfM* (CAU)、*petN-trnD* (GUC)、*petA-psbJ*、*psaA-ycf3* 等基因间隔区,以及 *rps16* 和 *ycf1* 等基因(图 2)。Mauve 比对的结果表明,中甸刺玫种内不同代表性个体的叶绿体基因组在结构上并无显著差异,并且不存在大片段或基因的逆转或者丢失(图 3)。

3 讨论与结论

虽然植物的叶绿体基因组在基因组成和排列顺序上具有较高的保守性,但由于长期对不同环境的适应常导致同属植物的叶绿体基因组在大小上发生改变,从而产生结构重排以及 IR 区的收缩或扩张等(Daniell et al., 2016)。中甸刺玫不同表型 40 个代表性个体的叶绿体基因组大小为 157 173~157 261 bp,整个基因组共编码 132 个基因,主要与光合作用和自我复制相关。与同属的单瓣月季花(Jian et al., 2018b)、亮叶月季(*R. lucidissima*)(Zhao et al., 2019)、木香花(*R. banksiae*)(杨芳, 2019)、大花香水月季(*R. odorata* var. *gigantea*)(Yang et al., 2014)、金樱子(*R. laevigata*)(Yin et al., 2020)以及其他种(Chen et

al., 2019; Cui et al., 2022)的叶绿体基因组相比,中甸刺玫的基因组除在长度上明显多出约 500 bp 以外,在基因组的 GC 含量、基因构成和排列顺序上与其他种基本一致,表明蔷薇属植物的叶绿体基因组较保守,种间差异较小,主要在非编码区有序列长度的变化。

密码子是连接核酸和蛋白质的纽带,研究物种的密码子偏好并确定最优密码子,有助于设计基因表达载体来提高目的基因的表达量,在作物遗传育种和品种改良方面具有重要应用价值(Qi et al., 2015)。密码子偏好性分析显示中甸刺玫的叶绿体基因组编码密码子中,亮氨酸是使用频率最高的氨基酸,而组氨酸使用频率最低。此外,中甸刺玫所有的首选同义密码子都是以 A-或 U-结尾(RSCU > 1),由于 RSCU 大于 1 则表示实际频率高于其他同义密码子的使用频率(晁岳恩等, 2012),因此中甸刺玫的密码子偏好以 A-和 U-结尾。这与同属的单瓣月季花的密码子偏好性(Jian et al., 2018b)基本一致,为研究相关基因的分子进化和外源表达奠定了基础。

植物的叶绿体简单重复序列为单亲遗传且有较高的种内多态性,是物种进化和多态性的重要遗传标记(Cavalier-Smith, 2002),常被用作野生植物系统演化和居群遗传研究(Provan, 2000; Flannery et al., 2006)以及作物遗传图谱构建的分子标记(Powell et al., 1995; Xue et al., 2012)。由于 poly A 和 poly T 相比 poly C 和 poly G 可能具有更高的结构稳定性(Gragg et al., 2002),因此大多数植物的叶绿体基因组单核苷酸 SSRs 多为 poly A 和 poly T 结构。中甸刺玫叶绿体基因组中共有 73 个简单重复序列 SSRs,主要为单核苷酸 SSRs,大部分由短腺嘌呤(A)或胸腺嘧啶(T)重复组成,主要位于 LSC 区的基因间隔区,与同属的单瓣月季花(Jian et al., 2018b)基本一致,也与其他多种植物如山茶属(*Camellia*)(丁祥青等, 2022; 邓永彪等, 2024)、绢蒿属(*Seriphidium*)(Jin et al., 2023)等类似。

Jian 等(2018b)研究表明,蔷薇属内不同物种的叶绿体基因组间的核酸高变区主要位于 LSC 区的 *trnK-rps16*、*ps16-trnQ*、*trnS-trnG*、*atpF-atpH*、*rps2-rpoC2* 等多个基因隔区、*rps19* 和 *ycf1* 等基因的编码区,以及 *rpl2*、*rps16*、*ndhA* 等基因的内含子区域。中甸刺玫种内不同表型代表性个体的全基因组核

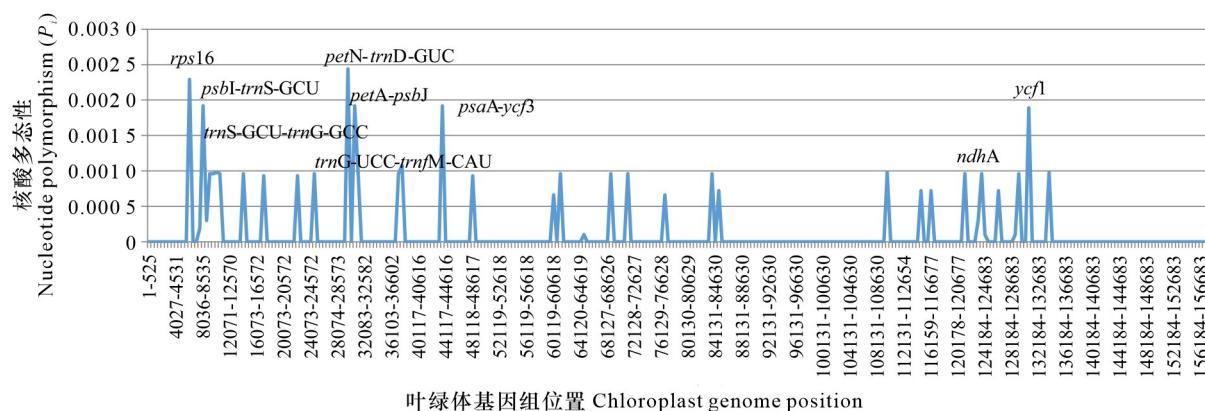
表 6 中甸刺玫叶绿体基因组中的简单序列重复

Table 6 Simple sequence repeats (SSRs) in *Rosa praelucens* chloroplast genomes

| 序号 Serial number | 类型 Type | 简单重复序列 SSR | 大小 Size (bp) | 开始位置 Start position | 结束位置 End position | 区域 Region | 基因 Gene |
|------------------------|------------|---|--------------------|---------------------------|-------------------------|--------------|-----------------|
| 1 | p1 | (A) 15 | 15 | 185 | 199 | LSC | |
| 2 | p1 | (A) 12 | 12 | 3764 | 3775 | LSC | <i>trnK-UUU</i> |
| 3 | p1 | (A) 15 | 15 | 4361 | 4375 | LSC | |
| 4 | p1 | (C) 11 | 11 | 6187 | 6197 | LSC | <i>rps16</i> |
| 5 | c | (T) 11 atacattctaaatttagaataattttcgaattattgtaaaaat caaattagttaactaaattctaaactaaataaacag(AATA) 3 | 104 | 6478 | 6581 | LSC | |
| 6 | p2 | (AG) 5 | 10 | 6791 | 6800 | LSC | |
| 7 | p1 | (A) 12 | 12 | 6971 | 6982 | LSC | |
| 8 | p2 | (AT) 6 | 12 | 7163 | 7174 | LSC | |
| 9 | p3 | (TAA) 4 | 12 | 8319 | 8330 | LSC | |
| 10 | p1 | (A) 10 | 10 | 8508 | 8517 | LSC | |
| 11 | p1 | (A) 20 | 20 | 8678 | 8697 | LSC | |
| 12 | c | (T) 12attttaatt(A) 11 | 34 | 9894 | 9927 | LSC | |
| 13 | p2 | (AT) 9 | 18 | 10115 | 10132 | LSC | |
| 14 | p3 | (TTA) 4 | 12 | 10368 | 10379 | LSC | |
| 15 | p1 | (T) 13 | 13 | 12228 | 12240 | LSC | |
| 16 | p1 | (C) 11 | 11 | 14297 | 14307 | LSC | |
| 17 | p1 | (T) 11 | 11 | 18416 | 18426 | LSC | <i>rpoC2</i> |
| 18 | p2 | (TA) 5 | 10 | 19785 | 19794 | LSC | <i>rpoC2</i> |
| 19 | p1 | (T) 10 | 10 | 26132 | 26141 | LSC | <i>rpoB</i> |
| 20 | p1 | (T) 10 | 10 | 28994 | 29003 | LSC | |
| 21 | p1 | (C) 14 | 14 | 29627 | 29640 | LSC | |
| 22 | c | (AAT) 4tc(T) 10 | 24 | 30565 | 30588 | LSC | |
| 23 | p1 | (A) 14 | 14 | 30722 | 30735 | LSC | |
| 24 | p1 | (T) 10 | 10 | 33042 | 33051 | LSC | |
| 25 | p2 | (TA) 5 | 10 | 36876 | 36885 | LSC | |
| 26 | c | (ATAA) 3tataa(AT) 5 | 27 | 37632 | 37658 | LSC | |
| 27 | p2 | (TA) 5 | 10 | 44657 | 44666 | LSC | <i>yef3</i> |
| 28 | p1 | (A) 10 | 10 | 45663 | 45672 | LSC | <i>yef3</i> |
| 29 | p1 | (A) 12 | 12 | 45979 | 45990 | LSC | |
| 30 | p1 | (T) 10 | 10 | 46522 | 46531 | LSC | |
| 31 | p1 | (A) 14 | 14 | 47885 | 47898 | LSC | |
| 32 | p1 | (A) 11 | 11 | 48476 | 48486 | LSC | |
| 33 | p4 | (TAAT) 3 | 12 | 48666 | 48677 | LSC | |
| 34 | p1 | (T) 14 | 14 | 48850 | 48863 | LSC | |
| 35 | c | (A) 13gaatcgaccgttcaagtattcaaaaattgcactaaaaa tgatagaaaaatcatagaaattgggacatg(TA) 5 | 91 | 48965 | 49055 | LSC | |
| 36 | p1 | (T) 10 | 10 | 49785 | 49794 | LSC | <i>trnL-UAA</i> |
| 37 | p4 | (TTTA) 3 | 12 | 51007 | 51018 | LSC | |
| 38 | c | (TA) 5t(TA) 5 | 21 | 53297 | 53317 | LSC | |
| 39 | c | (T) 10caagtgcggaacccaggaccagaagtagtaggattatttc ataataaaaatgtcgaaaatttttgcgaaaatgactgaaatcaa(AAAT) 3 | 112 | 56377 | 56488 | LSC | <i>atpB</i> |
| 40 | p4 | (AATT) 3 | 12 | 58966 | 58977 | LSC | |
| 41 | c | (T) 10ataggtatttagt(A) 10 | 33 | 61024 | 61056 | LSC | |
| 42 | p1 | (T) 12 | 12 | 61394 | 61405 | LSC | |
| 43 | p2 | (TC) 5 | 10 | 62863 | 62872 | LSC | <i>cemA</i> |
| 44 | p1 | (G) 10 | 10 | 64968 | 64977 | LSC | |

续表 6

| 序号 Serial number | 类型 Type | 简单重复序列 SSR | 大小 Size (bp) | 开始位置 Start position | 结束位置 End position | 区域 Region | 基因 Gene |
|------------------------|------------|---------------|--------------------|---------------------------|-------------------------|--------------|--------------|
| 45 | p1 | (T) 10 | 10 | 66847 | 66856 | LSC | |
| 46 | p1 | (T) 11 | 11 | 69049 | 69059 | LSC | |
| 47 | p1 | (A) 11 | 11 | 70255 | 70265 | LSC | |
| 48 | p1 | (T) 12 | 12 | 71742 | 71753 | LSC | |
| 49 | p1 | (T) 12 | 12 | 72425 | 72436 | LSC | <i>clpP</i> |
| 50 | p1 | (T) 11 | 11 | 73436 | 73446 | LSC | <i>clpP</i> |
| 51 | p2 | (AT) 6 | 12 | 74110 | 74121 | LSC | |
| 52 | p1 | (A) 12 | 12 | 79712 | 79723 | LSC | |
| 53 | p1 | (A) 14 | 14 | 79874 | 79887 | LSC | |
| 54 | p4 | (ATGT) 3 | 12 | 80012 | 80023 | LSC | <i>rpoA</i> |
| 55 | p1 | (T) 11 | 11 | 82081 | 82091 | LSC | |
| 56 | p1 | (A) 10 | 10 | 83136 | 83145 | LSC | |
| 57 | p1 | (A) 10 | 10 | 83991 | 84000 | LSC | <i>rpl16</i> |
| 58 | p1 | (T) 15 | 15 | 85420 | 85434 | LSC | |
| 59 | p3 | (TAT) 4 | 12 | 87117 | 87128 | IRb | <i>rpl2</i> |
| 60 | p6 | (TAGAAG) 4 | 24 | 94485 | 94508 | IRb | <i>ycf2</i> |
| 61 | p1 | (T) 10 | 10 | 102111 | 102120 | IRb | |
| 62 | p4 | (AGGT) 3 | 12 | 108333 | 108344 | IRb | <i>rrn23</i> |
| 63 | p4 | (TTTA) 3 | 12 | 110521 | 110532 | IRb | |
| 64 | p1 | (T) 10 | 10 | 121335 | 121344 | SSC | |
| 65 | p2 | (AT) 6 | 12 | 122168 | 122179 | SSC | |
| 66 | p1 | (A) 10 | 10 | 123061 | 123070 | SSC | <i>ndhA</i> |
| 67 | p1 | (T) 13 | 13 | 123308 | 123320 | SSC | <i>ndhA</i> |
| 68 | p1 | (T) 10 | 10 | 130333 | 130342 | SSC | <i>ycf1</i> |
| 69 | p4 | (AATA) 3 | 12 | 132966 | 132977 | IRa | |
| 70 | p4 | (CTAC) 3 | 12 | 135154 | 135165 | IRa | <i>rrn23</i> |
| 71 | p1 | (A) 10 | 10 | 141380 | 141389 | IRa | |
| 72 | p6 | (CTTCTA) 4 | 24 | 148992 | 149015 | IRa | <i>ycf2</i> |
| 73 | p3 | (ATA) 4 | 12 | 156372 | 156383 | IRa | <i>rpl2</i> |



滑窗大小 = 1 000 bp, 步长 = 1 000 bp.

Window size = 1 000 bp, and step length = 1 000 bp.

图 2 中甸刺玫种内叶绿体基因组的核酸多态性

Fig. 2 Nucleotide polymorphism (P_i) among *Rosa praelucens* chloroplast genomes



以 NCBI 序列号 MG450565.1 的中甸刺玫叶绿体基因组参照，用线连接的颜色相同的块表示局部共线块。

The chloroplast genome of *R. praelucens* with the NCBI accession No. MG450565.1 is used as the reference, the local collinear blocks are represented by blocks of the same color connected by line.

图 3 中甸刺玫种内代表性植株的叶绿体基因组 Mauve 比对

Fig. 3 Mauve alignment of the chloroplast genomes of representative *Rosa praelucens* plants

酸多态性较低,核酸多态性相对较高的是位于LSC区内的 *petN-trnD* (GUC)、*petA-psbJ*、*psaA-ycf3* 等少数几个基因间隔区,以及 *rps16* 和 *ycf1* 等少数几个基因。结合 Mauve 比对的结果,说明中甸刺玫种内的叶绿体基因组的基因序列和结构均高度保守,不存在大片段序列或基因的逆转或者丢失,其种内不同个体的表型变异并非由叶绿体基因组变异而引起。

综上所述,本研究在对中甸刺玫的叶绿体基因组进行简单报道的基础上,详细地分析了中甸刺玫叶绿体基因组的基因构成、密码子偏好以及简单重复序列等基本特征,并对种内具有不同表型的代表性个体的叶绿体基因组进行了比较基因组分析,结果表明,中甸刺玫种内的叶绿体基因组大小、序列和基因结构等方面均高度保守,不存在大片段序列或基因的逆转或者丢失,为中甸刺玫的保护和开发利用提供了叶绿体基因组方面基础数据。本研究也表明中甸刺玫种内丰富的表型变异并非由叶绿体基因组变异而引起,应结合中甸刺玫的高倍性特征,从染色体的数量和结构、基因表达以及表观遗传等角度进行深入系统的研究。

参考文献:

- BALAO F, HERRERA J, TALAVERA S, 2011. Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: A multivariate morphological approach [J]. *The New Phytologist*, 192(1): 256-265.
- BEIER S, THIEL T, MÜNCH T, et al., 2017. MISA-web: A web server for microsatellite prediction [J]. *Bioinformatics*, 33(16): 2583-2585.
- CAVALIER-SMITH ST, 2002. Chloroplast evolution: Secondary symbiogenesis and multiple losses [J]. *Current Biology*, 12(2): 62-64.
- CHAO YE, CHANG Y, WANG MF, et al., 2012. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species [J]. *Acta Agriculture Boreali-Sinica*, 27(4): 60-64. [晁岳恩, 常阳, 王美芳, 等, 2012. 7种作物叶绿体基因的密码子偏好性及聚类分析 [J]. *华北农学报*, 27(4): 60-64.]
- CHEN MR, ZHANG C, GAO XF. 2019. The complete chloroplast genome sequence of *Rosa pricei* (Rosaceae) [J]. *Mitochondrial DNA Part B-Resources*, 4(1): 1918-1919.
- CUI WH, DU XY, ZHONG MC, et al., 2022. Complex and reticulate origin of edible roses (*Rosa*, Rosaceae) in China [J]. *Horticulture Research*, 9(1): 678-691.
- DANIELL H, LIN CS, YU M, et al., 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biology*, 17(1): 1-29.
- DARLING ACE, MAU B, BLATTNER FR, et al., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements [J]. *Genome Research*, 14(7): 1394-1403.
- DENG JQ, JIAN HY, LI SB, et al., 2013. Cold tolerance of several wild *Rosa* resources endemic of Yunnan [J]. *Southwest China Journal of Agricultural Sciences*, 26(2): 273-277. [邓菊庆, 蹇洪英, 李淑斌, 等, 2013. 几种云南特有蔷薇资源的抗寒性研究 [J]. *西南农业学报*, 26(2): 273-277.]
- DENG YB, ZHANG J, LAN LL, et al., 2024. Analysis of chloroplast genome features of endangered and rare plant *Camellia minima* [J]. *Guihaia*, 44(1): 30-42. [邓永彪, 张进, 蓝伦礼, 等, 2024. 珍稀濒危植物越南小花金花茶的叶绿体基因组特征分析 [J]. *广西植物*, 44(1): 30-42.]
- DING XQ, LI WF, WU JL, et al., 2022. Chloroplast genome characteristics and genetic relationship of yellow *Camellia* [J]. *Journal Fujian Agriculturae Forestry University (Natural Science Edition)*, 52(3): 1-11. [丁祥青, 李文芳, 吴丽君, 等, 2022. 4种金花茶叶叶绿体基因组的比较分析 [J]. *福建农林大学学报(自然科学版)*, 52(3): 1-11.]
- DONG WP, XU C, WU P, et al., 2018. Resolving the systematic positions of enigmatic taxa: Manipulating the chloroplast genome data of Saxifragales [J]. *Molecular Phylogenetics and Evolution*, 126(9): 321-330.
- FAN YL, CHEN YC, JIAN HY, et al., 2021. Screening of *Rosa* germplasm resources with resistance to aphids [J]. *Journal of Yunnan University (Natural Science Edition)*, 43(3): 619-628. [范元兰, 陈宇春, 蹇洪英, 等, 2021. 蔷薇属抗蚜种质资源的筛选 [J]. *云南大学学报(自然科学版)*, 43(3): 619-628.]
- FANG Q, TIAN M, ZHANG T, et al., 2020. Karyotype analysis of *Rosa praelucens* and its closely related congeneric species based on FISH [J]. *Acta Horticulturae Sinica*, 47(3): 503-516. [方桥, 田敏, 张婷, 等, 2020. 中甸刺玫及其近缘种基于 FISH 的核型分析 [J]. *园艺学报*, 47(3): 503-516.]
- FLANNERY ML, MITCHELL FJ, COYNE S, et al., 2006. Plastid genome characterisation in *Brassica* and Brassicaceae

- using a new set of nine SSRs [J]. *Theoretical and Applied Genetics*, 113(7): 1221–1231.
- GRAGG H, HARFE BD, JINKS-ROBERTSON S, 2002. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frame shift intermediates by mismatch repair in *Saccharomyces cerevisiae* [J]. *Molecular and Cellular Biology*, 22(24): 8756–8762.
- GUAN WL, LI SF, SONG J, et al., 2012. Study on geographic distribution of *Rosa praelucens* endemic to Yunnan [J]. *Journal of West China Forestry Science*, 41(1): 88–93. [关文灵, 李世峰, 宋杰, 等, 2012. 云南特有濒危植物中甸刺玫的分布特征研究 [J]. *西部林业科学*, 41(1): 88–93.]
- GUISINGER MM, KUEHL JV, BOORE JL, et al., 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage [J]. *Molecular Biology and Evolution*, 28(1): 1543–1543.
- JIAN HY, LI SF, GUO JL, et al., 2018a. High genetic diversity and differentiation of an extremely narrowly distributed and critically endangered decaploid rose (*Rosa praelucens*): implications for its conservation [J]. *Conservation Genetics*, 19(4): 761–776.
- JIAN HY, ZHANG H, TANG KX, et al., 2010. Decaploidy in *Rosa praelucens* Byhouwer (Rosaceae) endemic to Zhongdian plateau, Yunnan, China [J]. *Caryologia*, 63(2): 162–167.
- JIAN HY, ZHANG SD, ZHANG T, et al., 2017. Characterization of the complete chloroplast genome of a critically endangered decaploid rose species, *Rosa praelucens* (Rosaceae) [J]. *Conservation Genetics Resources*, 10: 851–854.
- JIAN HY, ZHANG YH, YAN HJ, et al., 2018b. The complete chloroplast genome of a key ancestor of modern roses, *Rosa chinensis* var. *spontanea*, and a comparison with congeneric species [J]. *Molecules*, 23: 389.
- JIN GZ, LI WJ, SONG F, et al., 2023. Comparative analysis of complete *Artemisia* subgenus *Seriphidium* (Asteraceae: *Anthemideae*) chloroplast genomes: insights into structural divergence and phylogenetic relationships [J]. *BMC Plant Biology*, 23(1): 136–150.
- KEARSE M, MOIR R, WILSON A, et al., 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 28(12): 1647–1649.
- KU TC, ROBERTSON KR, 2003. *Rosa* (Rosaceae) [M] // WU ZY, RAVEN PH. *Flora of China*: Vol. 9. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press: 339–381.
- KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Research*, 29(22): 4633–4642.
- LI P, LU RS, XU WQ, et al., 2017. Comparative genomics and phylogenomics of East Asian tulips (*Amana*, Liliaceae) [J]. *Frontiers in Plant Science*, 8: 451.
- LI SF, LI CJ, JIAN HY, et al., 2013. Studies on phenotypic diversity of vulnerable *Rosa praelucens* endemic to Shangrila, Yunnan [J]. *Acta Horticulturae Sinica*, 40(5): 924–932. [李树发, 李纯佳, 蹇洪英, 等, 2013. 云南香格里拉特有易危植物中甸刺玫的表型多样性 [J]. *园艺学报*, 40(5): 924–932.]
- LI XX, ZHOU ZK, 2005. Endemic wild ornamental plants from North Western Yunnan [J]. *HortScience*, 40(6): 1612–1619.
- LIBRADO P, ROZAS J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 25(11): 1451–1452.
- LIU C, SHI L, ZHU Y, et al., 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J]. *BMC Genomics*, 13: 715.
- LIU LX, LI R, WORTH JRP, et al., 2017. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): Implications for understanding the evolution of *Fagales* [J]. *Frontiers in Plant Science*, 8: 968.
- LOHSE M, DRECHSEL O, KAHLAU S, et al., 2013. Organellar Genome-DRAW — A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucleic Acids Research*, 41(W1): 575–581.
- NEALE DB, SEDEROFF RR, 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine [J]. *Theoretical Applied Genetics*, 77(2): 212–216.
- PATEL RK, JAIN M, 2017. NGS QC toolkit: A toolkit for quality control of next generation sequencing data [J]. *PLoS ONE*, 7: e30619.
- POWELL W, MORGANTE M, MCDEVITT R, et al., 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 92(17): 7759–7763.
- PROVAN J, 2000. Novel chloroplast microsatellites reveal

- cytoplasmic variation in *Arabidopsis thaliana* [J]. *Molecular Ecology*, 9(12): 2183–2185.
- QI YY, XU WJ, XING T, et al., 2015. Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity [J]. *Evolutionary Bioinformatics Online*, 11: 65–77.
- QIN HN, YANG Y, DONG SY, et al., 2017. List of threatened species of higher plants in China [J]. *Biodiversity Science*, 25(7): 696–744. [覃海宁, 杨永, 董仕勇, 等, 2017. 中国高等植物受威胁物种名录 [J]. *生物多样性*, 25(7): 696–744.]
- RAMSEY J, SCHEMSKE DW, 2002. Neopolyploidy in flowering plants [J]. *Annual Review Ecology and Systematics*, 33(1): 589–639.
- SHETTY SM, SHAH MUM, MAKALE K, et al., 2016. Complete chloroplast genome sequence of *Musa balbisiana* corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains [J]. *Plant Genome*, 9(2): 1–14.
- SUN YX, MOORE MJ, LIN N, et al., 2017. Complete plastome sequencing of both living species of *Circaeasteraceae* (*Ranunculales*) reveals unusual rearrangements and the loss of the *ndh* gene family [J]. *BMC Genomics*, 18: 592.
- TAMURA K, STECHER G, PETERSON D, et al., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0 [J]. *Molecular Biology and Evolution*, 30(12): 2725–2733.
- WANG KJ, ZHANG T, WANG QG, et al., 2018. The phylogenetic position and hybrid origination of *Rosa praelucens* Byhouwer [J]. *Journal of Plant Genetic Resources*, 19(5): 1006–1015. [王开锦, 张婷, 王其刚, 等, 2018. 中甸刺玫的系统位置及杂交起源研究 [J]. *植物遗传资源学报*, 19(5): 1006–1015.]
- WICKE S, SCHNEEWEISS GM, DE PAMPHILIS CW, et al., 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function [J]. *Plant Molecular Biology*, 76(3–5): 273–297.
- WU XY, CHEN M, WANG QG, et al., 2014. Comparative study on the breeding systems of *Rosa praelucens* and *Rosa soulieana* [J]. *Acta Horticulturae Sinica*, 41(10): 2075–2084. [伍翔宇, 陈敏, 王其刚, 等, 2014. 中甸刺玫和川滇蔷薇的繁育系统比较研究 [J]. *园艺学报*, 41(10): 2075–2084.]
- XUE J, WANG S, ZHOU SL, 2012. Polymorphic chloroplast microsatellite loci in *Nelumbo* (*Nelumbonaceae*) [J]. *American Journal of Botany*, 99(6): 240–244.
- YANG F, 2019. Sequencing and structural analysis of chloroplast genome in *Rosa banksiae* [J]. *Genomics and Applied Biology*, 38(8): 3586–3594. [杨芳, 2019. 七里香蔷薇叶绿体基因组测序及结构分析 [J]. *基因组学与应用生物学*, 38(8): 3586–3594.]
- YANG JB, LI DZ, LI HT, 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs [J]. *Molecular Ecology Resources*, 14(5): 1024–1031.
- YE WQ, YAP ZY, LI P, et al., 2018. Plastome organization, genome-based phylogeny and evolution of plastid genes in *Podophylloideae* (*Berberidaceae*) [J]. *Molecular Phylogenetics and Evolution*, 127: 978–987.
- YIN XM, LIAO BS, GUO S, et al., 2020. The chloroplasts genomic analyses of *Rosa laevigata*, *R. rugosa* and *R. canina* [J]. *Chinese Medicine*, 15: 18.
- ZHAO L, ZHANG H, WANG QG, et al., 2019. The complete chloroplast genome of *Rosa lucidissima*, a critically endangered wild rose endemic to China [J]. *Mitochondrial DNA Part B-Resources*, 4(1): 1826–1827.
- ZHOU YQ, SU Q, ZHANG H, et al., 2016. Distribution and population quantitative dynamics of critically risked *Rosa praelucens* Byhouwer [J]. *Journal of Plant Genetic Resources*, 17(4): 649–654. [周玉泉, 苏群, 张颢, 等, 2016. 极危植物中甸刺玫的分布及种群数量动态 [J]. *植物遗传资源学报*, 17(4): 649–654.]
- ZHU A, GUO W, GUPTA S, et al., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates [J]. *The New Phytologist*, 209(4): 1747–1756.

(责任编辑 李 莉 王登惠)