

重瓣榆叶梅全叶绿体基因组遗传特征分析

段春燕, 王晓凌

(河南科技大学 农学院/牡丹学院, 河南 洛阳 471000)

摘要: 【目的】对重瓣榆叶梅 *Prunus triloba* ‘Multiplex’ 全叶绿体基因组序列进行测序, 探究其系统发育位置并分析其叶绿体基因组组成特点。【方法】以重瓣榆叶梅叶片为材料, 采用 2×CTAB 法提取叶绿体 DNA, 利用 Illumina NovaSeq 平台进行叶绿体基因组的测序, 组装、注释并分析其叶绿体基因组遗传特征。联合美国国家生物技术信息中心 (NCBI) 数据库数据, 基于全叶绿体基因组序列构建了重瓣榆叶梅系统进化关系。【结果】重瓣榆叶梅叶绿体基因组全长为 157 827 bp, NCBI 登录号 MT937181, 其结构为经典的四分体结构, 由 1 个大单拷贝区域 (LSC), 1 个小单拷贝区域 (SSC) 及反向重复区域 (IRa/IRb) 构成, 其序列长度分别为 86 032、19 023、26 386 bp。GC 和 AT 的总占比分别为 36.80% 和 63.20%。重瓣榆叶梅的完整叶绿体基因组序列注释到 132 个基因, 包括 tRNA 基因、编码蛋白基因、rRNA 基因, 分别为 37、87、8 个。重瓣榆叶梅叶绿体基因组共编码 26 678 个密码子和 236 个符合条件的 SSR 位点。SSR 位点中 A/T 碱基占优势, 碱基偏好性十分明显。【结论】系统进化树分析表明, 重瓣榆叶梅和榆叶梅 *P. triloba* 聚合成一支结构, 与同属植物长柄扁桃 *P. pedunculata* 亲缘关系较近。图 4 表 4 参 26

关键词: 重瓣榆叶梅; 全叶绿体基因组; 密码子偏好性; 重复序列; 系统发育

中图分类号: Q943.2 文献标志码: A 文章编号: 2095-0756(2024)03-0577-09

Genetic characteristics of whole chloroplast genome in *Prunus triloba* ‘Multiplex’

DUAN Chunyan, WANG Xiaoling

(College of Agriculture/Peony College, Henan University of Science and Technology, Luoyang 471000, Henan, China)

Abstract: [Objective] To investigate the phylogenetic location of *Prunus triloba* ‘Multiplex’ and analyze its chloroplast gene composition characteristics, the whole chloroplast genome sequence of *P. triloba* ‘Multiplex’ was sequenced. [Method] The leaves of *P. triloba* ‘Multiplex’ were used as materials to extract chloroplast DNA using 2×CTAB method. The whole chloroplast genome was sequenced by Illumina NovaSeq platform. We then assembled, annotated, and analyzed the genetic characteristics of the chloroplast genomic sequence. Combined with the NCBI database data, the phylogenetic relationship of *P. triloba* ‘Multiplex’ was constructed based on the whole chloroplast genome sequence. [Result] The total length of the chloroplast genome of *P. triloba* ‘Multiplex’ is 157 827 bp, with NCBI registration number MT937181. The structure is a classic tetrad structure, consisting of 1 LSC (large single copy region), 1 SSC (small single copy region), and IRa/IRb (reverse repeat region), with sequence lengths of 86 032, 19 023 and 26 386 bp, respectively. The total content of GC and AT is 36.80% and 63.20%, respectively. The complete chloroplast genome sequence of *P. triloba* ‘Multiplex’ was annotated to 132 genes, including 37 tRNA genes, 87 coding protein genes, and 8 rRNA genes, respectively. The chloroplast genome of *P. triloba* ‘Multiplex’ encodes a

收稿日期: 2023-09-30; 修回日期: 2023-12-13

基金项目: 河南省特色骨干学科建设项目 (17100001); 河南科技大学博士科研启动基金资助项目 (4024-13480114)

作者简介: 段春燕 (ORCID: 0000-0002-8939-6106), 副教授, 博士, 从事植物资源学研究。E-mail: 1445262292@

qq.com

total of 26 678 codons and 236 eligible SSR loci. The A/T base is dominant in the SSR site, and the base preference is very obvious. [Conclusion] The phylogenetic tree analysis showed that *P. triloba* ‘Multiplex’ and *P. triloba* aggregate into a branch structure, and are closely related to *P. pedunculata* in the same genus. The results of this study will provide new evidence and basic experimental data for species identification, phylogenetic evolution, and breeding and selection of *P. triloba* ‘Multiplex’. [Ch, 4 fig. 4 tab. 26 ref.]

Key words: *Prunus triloba* ‘Multiplex’; whole chloroplast genome; codon preference; repeated sequence; phylogeny

重瓣榆叶梅 *Prunus triloba* ‘Multiplex’ 属蔷薇科 Rosaceae 李属 *Prunus* 灌木或小乔木, 早春开花, 是中国特有园林植物^[1-2]。重瓣榆叶梅具有较强的抗盐碱能力, 广泛栽培于中国东北、华北和西北地区^[3]; 重瓣榆叶梅花和花粉富含蛋白质、生物酶、多糖、脂肪、氨基酸和维生素等物质, 可以用作保健品、药品、食品和营养性化妆品的原材料^[4-5]。种仁含油量较高, 可代郁李仁入药^[6]。重瓣榆叶梅于 1855 年被引入欧洲、美洲、大洋洲^[7], 广泛用于园林绿化。研究发现: 榆叶梅 *P. triloba* 与近缘种桃 *P. persica* 的核不对称性较接近, 榆叶梅为八倍体 ($2n=64$), 核型分类为 2B 或 2A, 核不对称系数高达 63.08%^[2]。但榆叶梅的物种发生还尚不清楚, 关于重瓣榆叶梅和榆叶梅系统发育关系仍是空白。

叶绿体基因组是植物叶绿体的独立遗传系统, 重组率低且保守, 一般为四分体结构, 由大单拷贝区 (LSC)、双向重复区 (IR) 和小单拷贝区 (SSC) 组成, 在植物物种鉴定、物种分化、DNA 条形码开发和系统分类研究中具备可信度^[8-11]。已有研究通过对 46 个类群的叶绿体 *atpB-rbcL* 构建进化树, 明确了小麦族狍草属 *Hystrix* 和赖草属 *Leymus* 的系统发育关系^[12]; 基于 *rbcL + matK + trnH-psbA* 组合序列和叶绿体全基因组研究白头翁属 *Pulsatilla* 物种的系统发育关系, 发现变异热点在物种识别和系统发育分析中建树结果相似^[13]; 基于叶绿体 *psbD* 基因构建的进化树表明: 花叶矢竹 *Pseudosasa japonica* f. *akebonosuji* 与毛竹 *Phyllostachys edulis* 的亲缘关系较近^[14]。目前, 已获得野生榆叶梅叶绿体基因组序列^[3], 但未见关于重瓣榆叶梅的叶绿体基因组序列报道。本研究通过 Illumina NovaSeq 高通量测序平台, 成功组装了重瓣榆叶梅叶绿体全基因组序列结构并分析其遗传多样性, 以期为重瓣榆叶梅系统发育、物种鉴定、资源开发研究提供理论依据。

1 材料与方法

1.1 材料

2020 和 2021 年 3—4 月, 对河南省洛阳市洛龙区 (34°38′09.94″N, 112°26′36.61″E) 重瓣榆叶梅种质资源进行调查, 采集重瓣榆叶梅植株的幼嫩新鲜叶片, 用纯净水和纸巾依次清洁叶片表面, 装入有变色硅胶颗粒的密封袋, 带回实验室保存以备。

1.2 DNA 提取

采用 2×CTAB 法提取重瓣榆叶梅总 DNA^[15]。叶片于液氮中研磨后用 CTAB 法提取植物总 DNA。以琼脂糖凝胶电泳 (琼脂质量分数为 1%, 电压为 120 V, 电泳时间为 25 min) 和微量核酸测定仪 (NanoDrop 2000) 检测样本 DNA 的完整性、质量和质量分数。样本 DNA 质量浓度为 56.16 mg·L⁻¹, 体积为 40 μL, 总量为 2.25 μg, $D(260/280)$ 为 1.89, $D(260/230)$ 为 1.02, 检测结果为 A 类。样品质量满足建库测序要求, 且总量满足 2 次或 2 次以上建库需要。

1.3 测序和序列组装、基因注释

已检测为 A 类的 DNA 样品用超声波机械打断, 经片段纯化、末端修复、3′加 A、连接测序接头后, 进行 PCR 扩增构建文库, 在 Illumina NovaSeq 平台进行测序得到原始序列 (raw reads), 对原始序列进行过滤: 去除带接头的序列, 过滤 N 含量超过 10% 的序列, 通过 BLAST 软件比对核酸序列数据库 (NT) 等质控得到过滤序列 (clean reads)。原始序列上传于数据库 GeneBank。

使用软件 SPAdes^[16]、aragorn v1.2.38、DOGMA^[17] 和 CPGview-RSG^[18] 进行测序数据组装和预测注释重瓣榆叶梅叶绿体的基因, 通过手工纠正, 成功获得其全叶绿体基因组序列并完成重瓣榆叶梅叶绿体基

因组的基因注释。用在线软件 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 绘制其叶绿体全基因组物理可视化图谱。

1.4 密码子

用 CodonW 1.4.2 对重瓣榆叶梅叶绿体基因组密码子数量和相对同义密码子使用度 (relative synonymous codon usage, RSCU) 进行分析。

1.5 重复序列

利用 MISA-web^[19] 检测重瓣榆叶梅的 cpSSR 位点, 参数设置为单核苷酸-8、双核苷酸-5、三核苷酸-3、四核苷酸-3、五核苷酸-3 和其他核苷酸-3。

1.6 系统发育分析

于美国国家生物技术信息中心 (NCBI) 数据库中下载 MK764428、HQ336405、MK798146、KF990036、KY101152、MK790138、MG602257、MK434918 和 MN661138 等多个物种的叶绿体基因组序列, 构建包括重瓣榆叶梅在内的蔷薇科李属植物的系统发育进化树, 观察并分析重瓣榆叶梅和其他物种间的亲缘关系和系统发育地位。系统发育进化树采用全叶绿体基因组数据进行分析: 用 MAFFT^[20](v7.427, auto 模式) 进行多序列比对, 将比对好的数据用 trimAl v1.4.rev15 修剪, 使用 RAxML v8.2.10, 选用 GTRGAMMA 模型, bootstrap=1 000, 构建最大似然进化树。

2 结果与分析

2.1 叶绿体基因组序列结构

重瓣榆叶梅全叶绿体基因组总长度为 157 827 bp, 数据已上传至 GeneBank, 登录号为 MT937181。基于 Illumina NovaSeq 6000 测序平台表明, 重瓣榆叶梅 (MT937181) 的读取总和 (ReadSum)、基本总和 (BaseSum) 分别为 15 307 141、4 592 142 300 bp, Q20 和 Q30 分别为 97.91% 和 93.90%。重瓣榆叶梅全叶绿体基因组为经典的四分体结构, 由 1 个大单拷贝区域 (LSC), 1 个小单拷贝区域 (SSC) 及反向重复区域 (IRa/IRb) 构成 (表 1~2 和图 1), 其序列长度分别为 86 032、19 023、26 386 bp。重瓣榆叶梅全叶绿体基因组中 GC 和 AT 的总占比分别为 36.80% 和 63.20%, GC 占比在 IR、LSC 和 SSC 区域存在较大的差异, IR 区域 GC 占比最高可达 42.58%; LSC 次之, GC 占比 34.64%; SSC 中 GC 占比最低仅 30.48%(表 1)。

表 1 重瓣榆叶梅叶绿体基因组结构组成

Table 1 Structure and composition of *P. triloba* 'Multiplex' chloroplast genome

区域	全叶绿体基因组		大单拷贝区(LSC)		小单拷贝区(SSC)		反向重复区(IRa)		反向重复区(IRb)	
	数量	占比/%	数量	占比/%	数量	占比/%	数量	占比/%	数量	占比/%
A	49 091	31.10	27 334	31.77	6 606	34.73	7 588	28.76	7 563	28.66
C	29 663	18.79	15 381	17.88	3 047	16.02	5 416	20.53	5 819	22.05
G	28 410	18.00	14 423	16.76	2 752	14.47	5 819	22.05	5 416	20.53
T	50 663	32.10	28 894	33.59	6 618	34.79	7 563	28.66	7 588	28.76
GC	58 073	36.80	29 804	34.64	5 799	30.48	11 235	42.58	11 235	42.58
总计	157 827	100.00	86 032	100.00	19 023	100.00	26 386	100.00	26 386	100.00

重瓣榆叶梅全叶绿体基因组中注释了 132 个基因 (图 1 和表 2), LSC 区、IR 区和 SSC 区的基因数量分别为 82 个 (62.12%)、19 个 (14.39%) 和 12 个 (9.09%), 蛋白编码基因有 87 个, 占比 65.91%, 核糖体 RNA(rRNA) 数量为 8 个, 占比 6.06%, 转运 RNA(tRNA) 数量为 37 个, 占比 28.03%。重瓣榆叶梅全叶绿体基因组序列的基因中, *ycf2* 长度为 6 834 bp, *ycf1* 为 5 589 bp, *rpoC2* 为 4 107 bp, 是排名在前 3 位的最大基因。rRNA 基因 *rrn4.5*、*rrn16*、*rrn5*、*rrn23* 在 IRs 区, 均有 2 个拷贝, 其基因长度分别为 103、1 491、121 和 2 809 bp。

重瓣榆叶梅全叶绿体基因组与其他植物的叶绿体基因类似, 大多数基因均不存在内含子且基因数量为 1。重瓣榆叶梅全叶绿体基因组有 19 个双拷贝类型基因, 占比为 14.39%, 包括: 5 个自我复制类型基因, 即 *ndhB*(NADH 脱氢酶亚基基因)、*rpl2*、*rpl23*、*rps12*、*rps7*; 4 个 rRNA 基因, 即 *rrn16*、*rrn23*、

rrn4.5、*rrn5*; 7个 tRNA 基因, 即 *trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG*、*trnV-GAC*; 3个未知功能蛋白基因, 即 *ycf1*、*ycf2* 和 *ycf15*。除 *atpF*、*ndhA*、*ndhB*、*petB*、*petD*、*rpl16*、*rpl2*、*rpoC1*、*rps12*、*rps16*、*trnA-UGC*、*trnG-GCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnV-UAC* 等 16 个基因各含 1 个内含子, *clpP*、*ycf3* 基因含 2 个内含子, 其余大部分基因不含内含子 (表 2)。

表 2 重瓣榆叶梅叶绿体全基因组所编码基因

Table 2 List of genes present in *P. triloba* 'Multiplex' chloroplast genome

基因分类	基因分组	基因名称
	光合系统 I 基因	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaI</i> 、 <i>psaJ</i>
	光合系统 II 基因	<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>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>
光合作用相关基因	NADH脱氢酶基因	<i>ndhA</i> *、 <i>ndhB</i> *(2)、 <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>
	细胞色素复合物质基因	<i>petA</i> 、 <i>petB</i> *、 <i>petD</i> *、 <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>
	ATP合酶基因	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> *、 <i>atpH</i> 、 <i>atpI</i>
	二磷酸核酮糖羧化酶大亚基基因	<i>rbcL</i>
自我复制相关基因	核糖体大亚基基因	<i>rpl14</i> 、 <i>rpl16</i> *、 <i>rpl2</i> *(2)、 <i>rpl20</i> 、 <i>rpl22</i> 、 <i>rpl23</i> (2)、 <i>rpl32</i> 、 <i>rpl33</i> 、 <i>rpl36</i>
	核糖体小亚基基因	<i>rps11</i> 、 <i>rps12</i> ***(2)、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps16</i> *、 <i>rps18</i> 、 <i>rps19</i> 、 <i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> (2)、 <i>rps8</i>
	RNA聚合酶亚基基因	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> *、 <i>rpoC2</i>
	核糖体RNA基因	<i>rrn16</i> (2)、 <i>rrn23</i> (2)、 <i>rrn4.5</i> (2)、 <i>rrn5</i> (2)
	转运RNA基因	<i>trnA-UGC</i> *(2)、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnE-UUC</i> 、 <i>trnF-GAA</i> 、 <i>trnG-GCC</i> *、 <i>trnG-UCC</i> 、 <i>trnH-GUG</i> 、 <i>trnI-CAU</i> (2)、 <i>trnI-GAU</i> *(2)、 <i>trnK-UUU</i> *、 <i>trnL-CAA</i> (2)、 <i>trnL-UAA</i> *、 <i>trnL-UAG</i> 、 <i>trnM-CAU</i> 、 <i>trnN-GUU</i> (2)、 <i>trnP-UGG</i> 、 <i>trnQ-UUG</i> 、 <i>trnR-ACG</i> (2)、 <i>trnR-UCU</i> 、 <i>trnS-GCU</i> 、 <i>trnS-GGA</i> 、 <i>trnS-UGA</i> 、 <i>trnT-GGU</i> 、 <i>trnT-UGU</i> 、 <i>trnV-GAC</i> (2)、 <i>trnV-UAC</i> *、 <i>trnW-CCA</i> 、 <i>trnY-GUA</i> 、 <i>trnY-M-CAU</i>
	成熟酶基因	<i>matK</i>
	依赖ATP的蛋白酶单元p基因	<i>clpP</i> **
其他基因	包裹膜蛋白基因	<i>cemA</i>
	乙酰辅酶A羧化酶亚基基因	<i>accD</i>
	c型细胞色素合成基因	<i>ccsA</i>
未知功能基因	保守开放阅读框	<i>ycf1</i> (2)、 <i>ycf15</i> (2)、 <i>ycf2</i> (2)、 <i>ycf3</i> **、 <i>ycf4</i>

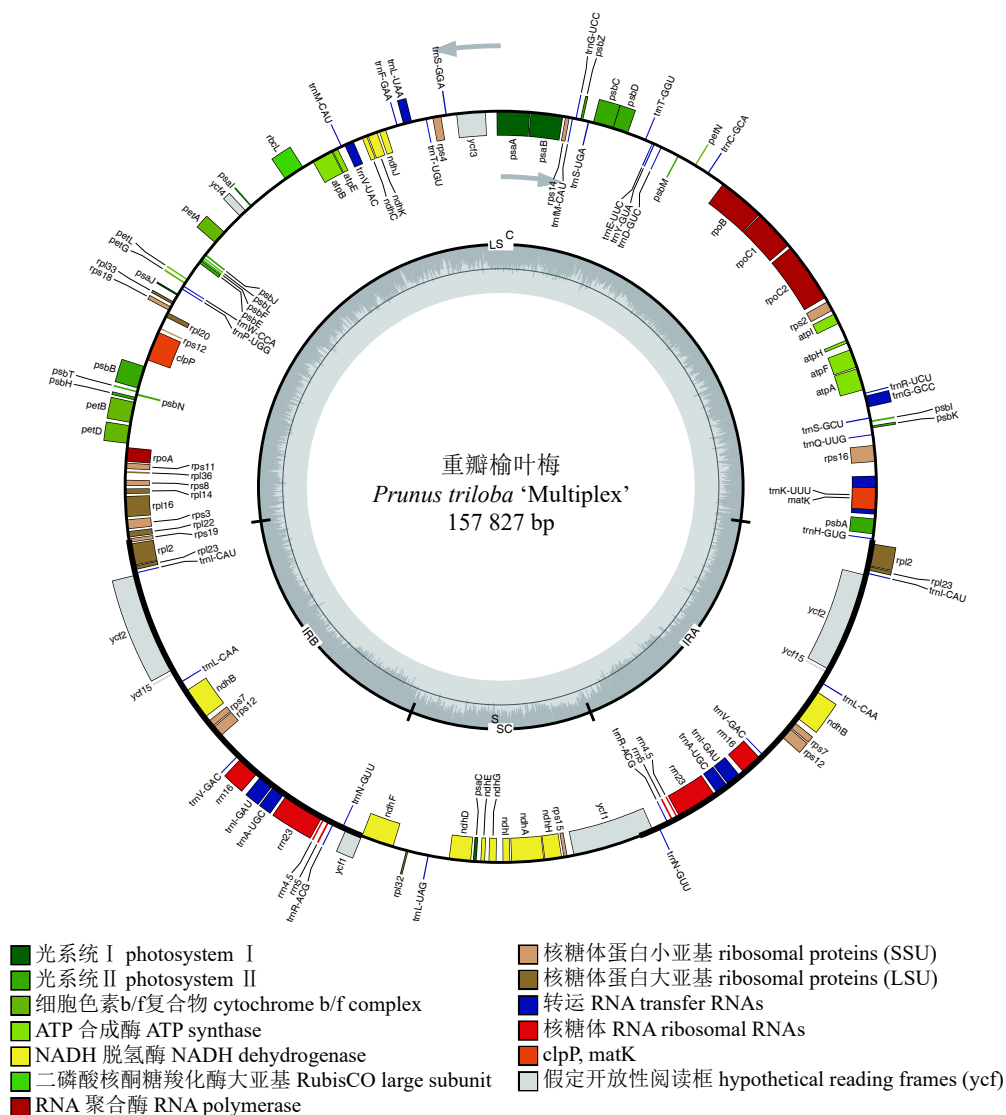
说明: * 代表基因有1个内含子; **代表基因有2个内含子; (2)代表基因的拷贝数为2。

2.2 密码子偏好性分析

MT937181 中发现共有 26 678 个编码密码子 (表 3)。其中, 亮氨酸 (Leu) 的编码密码子数量最多, 达 2 792 个; 蛋氨酸 (Met) 的编码密码子有 AUG、AUU、CUG、GUG 和 UUG 等, 其中 AUU 和 UUG 数量最少, 均只有 1 个。重瓣榆叶梅全叶绿体基因组的编码异亮氨酸 (Ile) 的 AUU 数量最多, 达 1 108 个。编码 Leu 的密码子最多, 占总量的 10.47%; 而最少的色氨酸 (Trp) 仅占 1.70%。此外, 31 种密码子的相对同义密码子使用度 (RSCU) > 1, 表明它们在重瓣榆叶梅全叶绿体基因组中是偏好密码子。在偏好密码子中有 29 种都是以 A 或者 U 结尾, 占 RSCU > 1 的密码子总量 93.55%, 只有编码 Leu 的 UUG 和编码 Met 的 AUG 这 2 种偏好密码子是以 G 结尾, 偏好密码子多选择密码子第 3 位是 A 或者 T, 体现出密码子 A/T 偏好性。RSCU < 1 的密码子有 36 个, 其中以 G/C 碱基结尾的有 33 个, 说明 RSCU < 1 的重瓣榆叶梅全叶绿体基因组的密码子更倾向以 G/C 碱基结尾; Trp 的 RSCU=1, 无密码子偏好性。

2.3 基因组重复序列情况

采用 MISA 软件对重瓣榆叶梅叶绿体基因组进行了 SSR 分析 (图 2 和表 4)。在重瓣榆叶梅叶绿体基因组中共查找到 236 个符合条件的 SSR 位点。单核苷酸重复单元有 146 个, 双核苷酸重复单元有 12 个, 三核苷酸重复单元有 62 个, 四核苷酸重复单元有 8 个, 五核苷酸重复单元有 1 个, 其他核苷酸重



圈外侧表示正向编码的基因，圈内侧表示反向编码的基因；彩色条表示不同的功能组；内部深灰色圈代表 GC 占比；浅灰色圈代表 AT 占比。

图 1 重瓣榆叶梅叶绿体基因组图谱

Figure 1 Gene map in *P. triloba* 'Multiplex' chloroplast genome

复单元有 7 个。叶绿体不同区域的 SSR 分布情况有差异，大部分 SSR 位于 LSC，一部分在 IRs 区域内，少数位于 SSC 区域。分布在 IR、LSC、SSC 区域的 SSR 位点分别为 38、158、40 个，占比分别为 16.10%、66.95%、16.95%。SSR 位点在不同功能的基因间分布也不均：114 个位于基因间隔区 (IGS)，82 个 SSR 位于外显子，40 个 SSR 位于内含子。从分布情况发现，SSR 多位于 LSC 和基因间隔区。从碱基的组成上看，SSR 中重复占比最大的是单核苷酸，约 61.86%。重瓣榆叶梅叶绿体基因组中的 SSR 主要是由 A 和 T 组成的，占有重复序列的 72.46%，其中 A/T 碱基构成的单核苷酸重复序列 137 条，AT/TA 组成的二核苷酸重复序列 11 条，A 与 T 组成的三核苷酸重复序列 18 条，A 与 T 组成的四核苷酸重复序列 5 条。

重瓣榆叶梅 SSR 位点在叶绿体基因组中多态性较高，具有 A/T 碱基偏好性。SSR 主要是 A 或 T、ATA、ATTA/TAAT/AT 复合重复、TTA/TAT 或 AT/TA，为 SSR 分子标记的开发打下基础。

2.4 基因组共线性与系统发育分析

选择重瓣榆叶梅与东京樱花 *P. yedoensis*、扁桃 *P. dulcis*、麦李 *P. glandulosa* 和中国李 *P. salicina* 的叶绿体基因组进行 Mauve 比对 (图 3)，结果表明：重瓣榆叶梅全叶绿体基因组与其余 4 个物种序列有良好的共线性关系，叶绿体基因组未检测到大片段的基因重排。5 个物种的基因组序列横向均分为 3 个部

表3 重瓣榆叶梅叶绿体基因组中的编码密码子

Table 3 Codon usage of *P. triloba* 'Multiplex'

符号	氨基酸	密码子	数量/个	RSCU	符号	氨基酸	密码子	数量/个	RSCU
*	Ter	UAA	52	1.793 1	M	Met	GUG	2	0.016 0
*	Ter	UAG	21	0.724 2	M	Met	UUG	1	0.008 0
*	Ter	UGA	14	0.482 7	M	Met	AUG	620	0.495 2
A	Ala	GCA	385	1.096 8	N	Asn	AAC	308	0.470 2
A	Ala	GCC	226	0.644 0	N	Asn	AAU	1 002	1.529 8
A	Ala	GCG	155	0.441 6	P	Pro	CCA	313	1.137 2
A	Ala	GCU	638	1.817 6	P	Pro	CCC	208	0.755 6
C	Cys	UGC	78	0.500 0	P	Pro	CCG	151	0.548 4
C	Cys	UGU	234	1.500 0	P	Pro	CCU	429	1.558 4
D	Asp	GAC	213	0.393 4	Q	Gln	CAA	725	1.547 4
D	Asp	GAU	870	1.606 6	Q	Gln	CAG	212	0.452 6
E	Glu	GAA	1 041	1.476 6	R	Arg	AGA	503	1.880 4
E	Glu	GAG	369	0.523 4	R	Arg	AGG	176	0.658 2
F	Phe	UUC	524	0.688 2	R	Arg	CGA	358	1.338 6
F	Phe	UUU	999	1.311 8	R	Arg	CGC	111	0.415 2
G	Gly	GGA	728	1.621 2	R	Arg	CGG	118	0.441 0
G	Gly	GGC	174	0.387 6	R	Arg	CGU	339	1.267 2
G	Gly	GGG	303	0.674 8	S	Ser	AGC	140	0.412 8
G	Gly	GGU	591	1.316 4	S	Ser	AGU	398	1.174 2
H	His	CAC	152	0.471 4	S	Ser	UCA	414	1.221 0
H	His	CAU	493	1.528 6	S	Ser	UCC	327	0.964 8
I	Ile	AUA	732	0.959 1	S	Ser	UCG	187	0.551 4
I	Ile	AUC	450	0.589 5	S	Ser	UCU	568	1.675 8
I	Ile	AUU	1 108	1.451 4	T	Thr	ACA	422	1.241 2
K	Lys	AAA	1 080	1.505 2	T	Thr	ACC	250	0.735 2
K	Lys	AAG	355	0.494 8	T	Thr	ACG	151	0.444 0
L	Leu	CUA	367	0.788 4	T	Thr	ACU	537	1.579 6
L	Leu	CUC	186	0.399 6	V	Val	GUA	557	1.537 6
L	Leu	CUG	188	0.403 8	V	Val	GUC	161	0.444 4
L	Leu	CUU	580	1.246 2	V	Val	GUG	206	0.568 8
L	Leu	UUA	909	1.953 6	V	Val	GUU	525	1.449 2
L	Leu	UUG	562	1.207 8	W	Trp	UGG	454	1.000 0
M	Met	AUU	1	0.008 0	Y	Tyr	UAC	205	0.400 0
M	Met	CUG	2	0.016 0	Y	Tyr	UAU	820	1.600 0

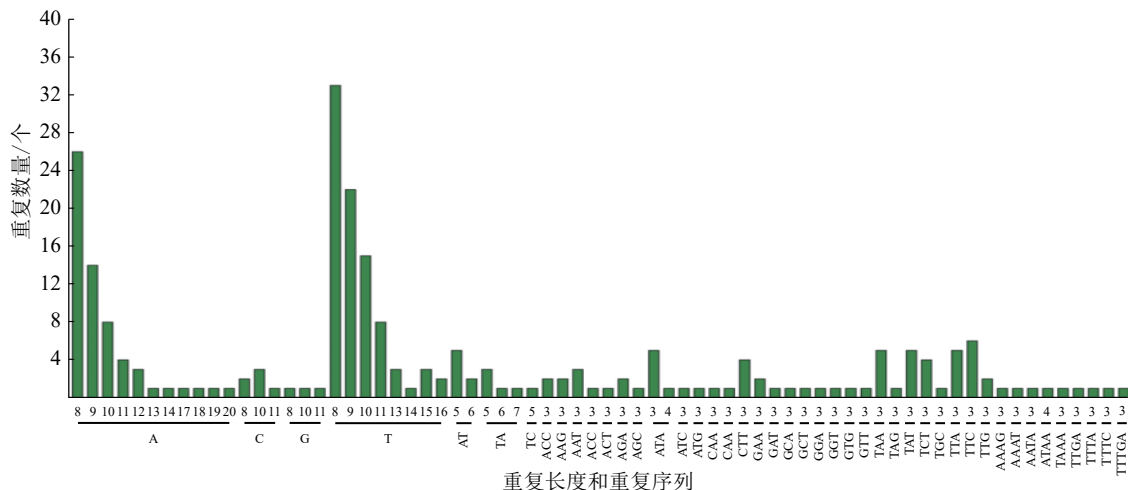
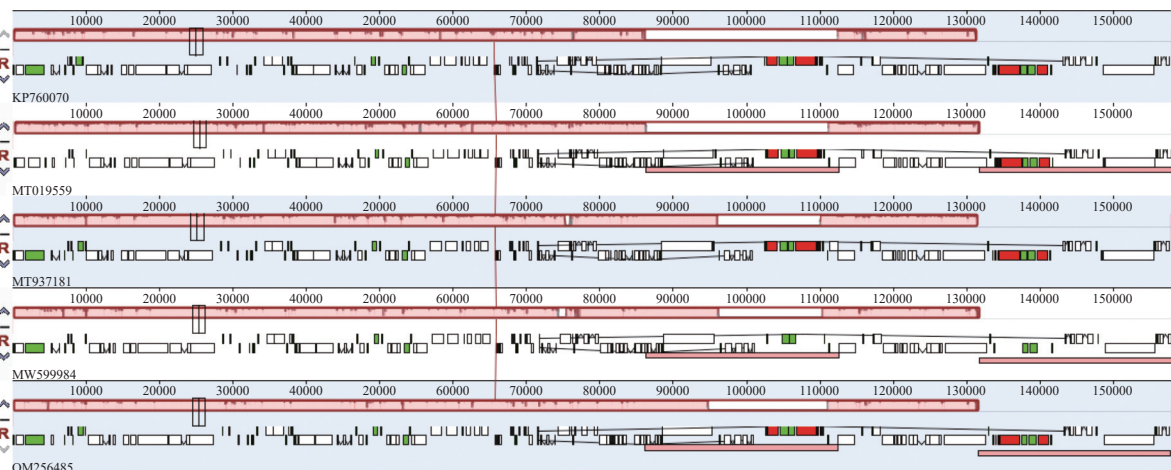


图2 重瓣榆叶梅叶绿体基因组中 SSR 位点类型及数量
Figure 2 Type and number of SSR loci of *P. triloba* 'Multiplex' chloroplast genome

分(各物种序列长度与图中序列长度不同), 从左到右为 LSC/IR/SSC, 纵向有一直竖线, 说明 5 个物种叶绿体基因组的基因组成和排列十分相似。LSC 和 SSC 中粉色局块斑纹较多, 白色多为 IR 区, 几乎为空白, 说明 IR 区域十分保守。76 915~81 087 bp 局块区间, 5 个物种的对比差异性最大。

重瓣榆叶梅全叶绿体基因组与已经公布的其他植物的叶绿体序列构建系统进化树(图 4)。结果发现: 进化树上各节点和分支聚类的支持率较高, 检验分值可达 100%, 说明聚类结果的可靠性较高。由聚类图可知重瓣榆叶梅与榆叶梅聚合在一个分支, 和长柄扁桃 *P. pedunculata* 亲缘关系最近。基于单拷贝核基因 *RPB2*、*Leafy* 内含子序列对榆叶梅及其近缘种系统发育学分析, 发现桃 *P. persica*、普通扁桃 *P. dulcis*、蒙古扁桃 *P. mongolica* 聚为一个类群; 榆叶梅、长柄扁桃、矮扁桃 *P. tenella* 聚为一个类群。矮扁桃的系统发育地位有争议, 可能与矮扁桃的取样和命名有极大关系, 需要进一步探讨。核基因和叶绿体基因组共同认为榆叶梅和长柄扁桃具有同源最近亲缘关系, 在进行榆叶梅品种选育和育苗育种时, 可充分考虑榆叶梅与长柄扁桃的系统发育关系。



粉色条带为局部共线性; 下方的方块为基因注释信息: 绿色为 tRNA 基因, 红色为 rRNA 基因, 白色为 CDS 基因。

图 3 重瓣榆叶梅的叶绿体全基因组共线性分析

Figure 3 Genomic collinearity of *P. triloba* 'Multiplex' chloroplast genome

3 讨论

被子植物的叶绿体基因组长度约 1.60×10^5 bp, 编码基因数约 130 个^[21-22]。本研究获得的重瓣榆叶梅全叶绿体基因组序列 MT937181 全长为 157 827 bp, 编码基因数为 132 个, 与一般被子植物相符合。重瓣榆叶梅为园林栽培植物, 分布广泛; 其他种为野生种, 分布和自然生境不同, 其中蒙古扁桃十分耐寒且仅分布于蒙古高原。将 4 个不同叶绿体基因组序列进行比较, 发现重瓣榆叶梅总序列、LSC 和 SSC 长度最小, 蒙古扁桃 SSC 和 IR 长度最大。推测可能在人工栽培和植物品种选育中, 重瓣榆叶梅由于外在表现性状和生长环境的优越, 其叶绿体基因组中基因种类增加, 但总序列长度减小; 而在野生环境分布狭窄或需要极端耐寒时, 蒙古扁桃为适应不良环境, 其 SSC 区和 IR 长度增大, 基因种类和数量发生较大变化, *ycf15* 基因缺失, 且增加了 1 个 *rps19* 基因。*ycf15* 属于未知功能基因, 在龙葵

表 4 重瓣榆叶梅叶绿体基因组中 SSR 位点所在区域情况

Table 4 Region of SSR loci of *P. triloba* 'Multiplex' chloroplast genome

区域	数量/个	占比/%	外显子/个	内含子/个	基因间区/个
IR	38	16.10	21	4	13
LSC	158	66.90	35	33	90
SSC	40	16.90	26	3	11
总计	236	100.00	82	40	114

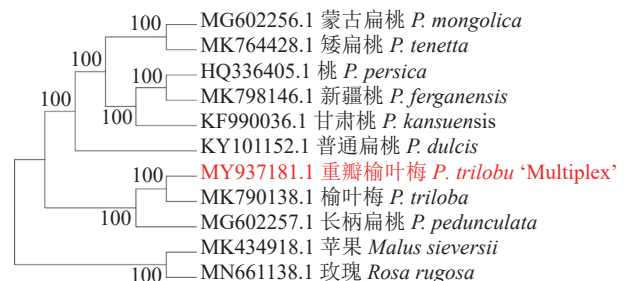


图 4 重瓣榆叶梅及其近缘种系统发育树

Figure 4 Phylogenetic tree of *P. triloba* 'Multiplex' and its related species

属 *Amborella* 和蔷薇类等原始被子植物中无功能, 在毛茛属 *Ranunculus* 植物已经完全丢失^[23]。 *rps4* 以及 *rps19* 为核糖体小亚基类基因, 核糖体在细胞中负责完成“中心法则”里“翻译”: 由 RNA 到蛋白质这一过程。核糖体小亚基通过与信使 RNA 结合, 转运 RNA 运送的氨基酸分子来合成多肽。 *ycf15*、*rps4* 以及 *rps19* 等 3 类基因的变化情况可能与植物对栽培选育、不同环境适应有直接关系。

自然界遗传密码子在 mRNA 翻译成蛋白质时具有重要决定作用。同义密码子指编码同一种氨基酸的不同密码子, 有时具有使用偏好性^[10, 24]。重瓣榆叶梅编码基因 RSCU > 1 密码子的第 3 个核苷酸具有明显的 A/U 偏好。简单重复序列 (SSR) 广泛分布于真核生物基因组中, 重瓣榆叶梅与榆叶梅 (MK790138)、长柄扁桃 (MG602257) 和蒙古扁桃 (MG602256) 叶绿体基因组的 SSR 位点均具有富集多聚 A 或多聚 T 情况。本研究系统进化分析表明: 重瓣榆叶梅与榆叶梅、长柄扁桃亲缘关系最近, 进化结果与传统分类学相符^[2-3, 24-26]。

4 结论

本研究成功获得了重瓣榆叶梅全叶绿体基因组序列, 其总长度为 157 827 bp, 具备经典的四分体结构, GeneBank 登录号为 MT937181, 编码基因数为 132 个。LSC、SSC 和 IRa/IRb 序列长度分别为 86 032、19 023 和 26 386 bp, GC 占比在 IR、LSC 和 SSC 区域存在较大的差异。重瓣榆叶梅叶绿体基因组与其他 3 种近缘种植物比较中, *ycf15*、*rps4* 以及 *rps19* 等 3 类基因有较大变化, 可能是栽培环境的选择影响。重瓣榆叶梅叶绿体基因组的 SSR 位点具有富集 A/T 情况。基于叶绿体基因组数据确定了重瓣榆叶梅和榆叶梅系统发育关系, 系统进化分析结果与传统植物分类学与形态学相符, 即重瓣榆叶梅与榆叶梅、长柄扁桃亲缘关系最近。

5 参考文献

- [1] 俞德浚, 陆玲娣, 谷粹芝, 等. 中国植物志: 第38卷[M]. 北京: 科学出版社, 1986.
YU Dejun, LU Lingdi, GU Cuizhi, et al. *Flora of China: Volume 38* [M]. Beijing: Science Press, 1986.
- [2] 董山平, 罗乐, 钟军珺, 等. 榆叶梅及其部分近缘种的染色体核型分析[J]. 西北植物学报, 2015, 35(3): 608 - 613.
DONG Shanping, LUO Le, ZHONG Junjun, et al. Analysis of karyotypes of *Prunus triloba* and four related species [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2015, 35(3): 608 - 613.
- [3] DUAN Chunyan, SHEN Yehua, ZHAO Guifang. Complete chloroplast genome characteristics of *Prunus triloba* Lindl [J]. *Mitochond DNA Part B Resources*, 2020, 5(1): 504 - 505.
- [4] 李姣姣, 陈媛梅. 重瓣榆叶梅花中多糖提取工艺及动态含量研究[J]. 中国农业大学学报, 2018, 23(1): 71 - 79.
LI Jiaojiao, CHEN Yuanmei. Optimal extraction technology and the dynamic content of polysaccharide in *Amygdalus triloba* f. *multiflex* flowers [J]. *Journal of China Agricultural University*, 2018, 23(1): 71 - 79.
- [5] 徐毅励, 陈媛梅. 重瓣榆叶梅花中蛋白质提取工艺及动态含量研究[J]. 西北林学院学报, 2015, 30(3): 191 - 196.
XU Yili, CHEN Yuanmei. Optimal extraction technology and dynamic content of protein in the flower of *Amygdalus triloba* f. *multiflex* [J]. *Journal of Northwest Forestry University*, 2015, 30(3): 191 - 196.
- [6] 周保昌, 贺斌, 李焯, 等. 经典名方中郁李仁的本草考证[J]. 中国实验方剂学杂志, 2023, 29(11): 1 - 11.
ZHOU Baochang, HE Bin, LI Ye, et al. Herbal textual research on pruni semen in famous classical formulas [J]. *Chinese Journal of Experimental Traditional Medical Formulae*, 2023, 29(11): 1 - 11.
- [7] 张强英, 张启翔, 程堂仁. 榆叶梅野生群体表型多样性的研究[J]. 中南林业科技大学学报, 2012, 32(5): 155 - 160.
ZHANG Qiangying, ZHANG Qixiang, CHENG Tangren. Study on phenotypic diversity of *Prunus triloba* wild populations [J]. *Journal of Central South University of Forestry & Technology*, 2012, 32(5): 155 - 160.
- [8] 洪森荣, 张牧彤, 徐子林, 等. ‘怀玉山’高山马铃薯叶绿体基因组特征及密码子使用偏好性分析[J]. 浙江农林大学学报, 2024, 41(1): 92 - 103.
HONG Senrong, ZHANG Mutong, XU Zilin, et al. Chloroplast genome characteristics and codon usage preference of *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’ [J]. *Journal of Zhejiang A&F University*, 2024, 41(1): 92 - 103.
- [9] 赵祺, 余佳兴, 秦宇雯, 等. 基于高通量测序的菜头肾叶绿体基因组的组装及序列分析[J]. 中草药, 2021, 52(6): 1744 - 1750.
ZHAO Qi, YU Jiaying, QIN Yuwen, et al. Assembly and sequence analysis of chloroplast genome of *Championella sarcorrhiza* based on high-throughput sequencing [J]. *Chinese Traditional and Herbal Drugs*, 2021, 52(6): 1744 - 1750.

- [10] 钱方, 高作敏, 胡利娟, 等. 海甘蓝(*Crambe abyssinica*)叶绿体基因组特征及其系统发育研究[J]. 生物技术通报, 2022, **38**(6): 174 – 186.
QIAN Fang, GAO Zuomin, HU Lijuan, *et al.* Characteristics of *Crambe abyssinica* chloroplast genome and its phylogenetic relationship in Brassicaceae [J]. *Biotechnology Bulletin*, 2022, **38**(6): 174 – 186.
- [11] 龚意辉, 周桂花, 李丽梅, 等. 黄桃叶绿体基因组的组装与序列分析[J]. 热带作物学报, 2022, **43**(8): 1545 – 1553.
GONG Yihui, ZHOU Guihua, LI Limei, *et al.* Assembly and sequence analysis of chloroplast genome of *Amygdalus persica* [J]. *Chinese Journal of Tropical Crops*, 2022, **43**(8): 1545 – 1553.
- [12] 刘静, 张海琴, 凡星, 等. 基于叶绿体 $atpB-rbcL$ 序列探讨猬草属和赖草属植物的系统发育和母系起源[J]. 草业学报, 2012, **21**(5): 77 – 85.
LIU Jing, ZHANG Haiqin, FAN Xing, *et al.* Phylogenetic relationships and maternal donor of *Hystrix* and *Leymus* species as revealed by chloroplast $atpB-rbcL$ sequence [J]. *Acta Prataculturae Sinica*, 2012, **21**(5): 77 – 85.
- [13] LI Qiujie, SU Na, ZHANG Ling, *et al.* Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae) [J/OL]. *Scientific reports*, 2020, **10**(1): 19781 [2023-08-20]. doi: 10.1038/s41598-020-76699-7.
- [14] 许冰清, 安苗苗, 姜可以, 等. 花叶矢竹叶叶绿体 $psbD$ 基因的克隆与功能分析[J]. 浙江农林大学学报, 2015, **32**(4): 557 – 565.
XU Bingqing, AN Miaomiao, JIANG Keyi, *et al.* Cloning and functional analysis of the chloroplast $psbD$ gene of *Phyllostachys japonicus* [J]. *Journal of Zhejiang A&F University*, 2015, **32**(4): 557 – 565.
- [15] 闫玖英, 马长青, 常博, 等. 改良CTAB法用于苹果果实基因组DNA的提取[J]. 分子植物育种, 2017, **15**(9): 3610 – 3615.
YAN Jiuying, MA Changqing, CHANG Bo, *et al.* A modified CTAB method for genomic DNA extraction from apple fruit [J]. *Molecular Plant Breeding*, 2017, **15**(9): 3610 – 3615.
- [16] BANKEVICH A, NURK S, ANTIPOV D, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing [J]. *Journal of Computational Biology*, 2012, **19**(5): 455 – 477.
- [17] WYMAN S K, JANSEN R K, BOORE J L. Automatic annotation of organellar genomes with DOGMA [J]. *Bioinformatics*, 2004, **20**(17): 3252 – 3255.
- [18] LIU Chang, SHI Linchun, ZHU Yingjie, *et al.* CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J/OL]. *BMC Genomics*, 2012, **13**: 715 [2023-08-25]. doi:10.1186/1471-2164-13-715.
- [19] BEIER S, THIEL T, MNCH T, *et al.* MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 2017, **33**(16): 2853 – 2858.
- [20] KATO H, STANDLEY D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability [J]. *Molecular Biology and Evolution*, 2013, **30**(4): 772 – 780.
- [21] 富贵, 刘晶, 李军乔. 密花香薷叶绿体基因组结构及系统进化分析[J]. 中草药, 2022, **53**(6): 1844 – 1853.
FU Gui, LIU Jing, LI Junqiao. Characterization of chloroplast genome structure and phyletic evolution of *Elsholtzia densa* [J]. *Chinese Traditional and Herbal Drugs*, 2022, **53**(6): 1844 – 1853.
- [22] 黄琼林. 高良姜叶绿体基因组测序与特征分析[J]. 热带作物学报, 2021, **42**(1): 1 – 6.
HUANG Qionglin. Complete sequencing and analysis of chloroplast genome from *Alpinia officinarum* Hance [J]. *Chinese Journal of Tropical Crops*, 2021, **42**(1): 1 – 6.
- [23] SHI Cao, LIU Yuan, HUANG Hui, *et al.* Contradiction between plastid gene transcription and function due to complex post transcriptional splicing: an exemplary study of $ycf15$ function and evolution in angiosperms [J/OL]. *PLoS One*, 2013, **8**(3): e59620 [2023-08-25]. doi: 10.1371/journal.pone.0059620.
- [24] 段春燕, 张凯, 段义忠. 长柄扁桃叶绿体基因组比较[J]. 生物工程学报, 2020, **36**(12): 2850 – 2859.
DUAN Chunyan, ZHANG Kai, DUAN Yizhong. Comparison of complete chloroplast genome sequences of *Amygdalus pedunculata* Pall [J]. *Chinese Journal of Biotechnology*, 2020, **36**(12): 2850 – 2859.
- [25] 赵一之. 关于榆叶梅的系统位置[J]. 内蒙古大学学报(自然科学版), 1996, **27**(1): 70 – 71.
ZHAO Yizhi. On systematic position of *Cerasus triloba*, [J]. *Journal of Inner Mongolia University Natural Science Edition*, 1996, **27**(1): 70 – 71.
- [26] WANG Li, GUO Zhihong, SHANG Qianhan, *et al.* The complete chloroplast genome of *Prunus triloba* var. *plena* and comparative analysis of *Prunus* species: genome structure, sequence divergence, and phylogenetic analysis [J]. *Brazilian Journal of Botany*, 2021, **44**: 85 – 95.