

点地梅属 10 种植物叶绿体基因组密码子偏好性分析

郭玉朋^{1*}, 史明伟¹, 马 虎¹, 马彦军², 姚步青³

1. 青海民族大学生态环境与资源学院, 青海省特色经济植物高值化利用重点实验室, 青海省蕨麻产业研究院, 青海 西宁 810007
2. 甘肃农业大学林学院, 甘肃 兰州 730070
3. 中国科学院 西北高原生物研究所, 青海省寒区恢复生态学重点实验室, 青海 西宁 810008

摘要: **目的** 研究点地梅属 *Androsace* L. 物种叶绿体基因组蛋白编码基因密码子的偏好性及其影响因素。**方法** 测序了点地梅属 1 个物种 (西藏点地梅) 的叶绿体基因组, 其他 9 物种的序列从 NCBI 下载。根据序列筛选条件, 共得到 41 条共有序列。用 CodonW 1.4.2 计算相关参数, Python 整理数据, R 做相关性分析并绘图。**结果** 点地梅属各物种密码子第 1、2、3 位置碱基组成 GC 含量表现为 $GC_1 > GC_2 > GC_3$, 其中第 3 位 GC 含量最小。各位点 GC 含量相关性为 GC_1 、 GC_2 极显著, 而两者与 GC_3 相关性都不显著, 说明 3 个位置 GC 含量变化机制不同。3 个位置 GC 含量都小于 50%, 表明密码子倾向于使用 A/U 碱基, 尤其是第 3 位。从 10 个物种中共鉴定到 13~17 个最优密码子。所有基因的有效密码子数 (effective number of codons, ENC) 均大于 35, 说明密码子使用整体偏性较弱。中性绘图、ENC-plot 及 PR2-plot 分析显示, 造成密码子使用偏性的主要因素是自然选择, 但突变压对大部分基因的密码子偏性也有重要影响。系统发育及聚类分析显示密码子偏性在物种间是保守的。**结论** 点地梅属叶绿体基因组蛋白编码基因倾向于使用 A/U 碱基编码的密码子, 自然选择是影响密码子偏好性的主要因素。研究结果能为点地梅属的分子进化、系统发育及叶绿体基因工程提供基础。

关键词: 点地梅属; 密码子偏性; 最优密码子; 突变压; 自然选择

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)04-1355-11

DOI: 10.7501/j.issn.0253-2670.2025.04.023

Analysis of codon bias in chloroplast genomes of ten plants in *Androsace*

GUO Yupeng¹, SHI Mingwei¹, MA Hu¹, MA Yanjun², YAO Buqing³

1. Qinghai Provincial Key Laboratory of High Value Utilization of Characteristic Economic Plants, Qinghai Tibet Plateau *Potentilla anserine* Industry Research Institute, College of Ecological Environment and Resources, Qinghai Minzu University, Xining 810007, China
2. College of Forestry, Gansu Agricultural University, Lanzhou 730070, China
3. Key Laboratory of Restoration Ecology of Cold Area in Qinghai Province, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China

Abstract: Objective To study the codon preference of chloroplast genome protein encoding genes and influencing factors in *Androsace*. **Methods** The chloroplast genome of *Androsace mariae* was sequenced, and the sequences of the other nine species were downloaded from NCBI. According to the sequence screening conditions, a total of 41 common sequences were obtained. CodonW 1.4.2 was used to calculate the relevant parameters, Python was used to process the data, R was used for correlation analysis and drawing figures. **Results** The GC content at the first, second and third positions of the codon was $GC_1 > GC_2 > GC_3$, and the GC content at the third position was the minimum. The correlation of GC content between GC_1 and GC_2 was extremely significant, but not significant for GC_3 between GC_1 or GC_2 , indicating that the change mechanism of GC content at the three sites was different. The GC content of all three sites was less than 50%, indicating that the codons tended to use A/U bases, especially the third position. A total of 13—17 optimal codons were identified from 10 species. Effective number of codons (ENC) of all genes were greater than 35,

收稿日期: 2024-09-09

基金项目: 青海民族大学研究生专业示范思政课程建设项目 (SK-2022-09)

*通信作者: 郭玉朋, 青海民族大学副教授, 主要从事药用植物资源研究。Tel: 17697266518 E-mail: mlplml@sina.com

indicating that the overall bias was weak. Neutral-plot, ENC-plot and PR2-plot analysis showed that the main factor causing codon bias was natural selection, but mutation pressure also had an important influence on the codon bias of most genes. Phylogenetic and cluster analysis showed that codon bias was conserved in species. **Conclusion** The protein coding genes of the chloroplast genome tend to use codons encoded by A/U bases, and natural selection is the main factor affecting codon preference. The results of this study can provide the basis for molecular evolution, phylogeny and chloroplast genetic engineering.

Key words: *Androsace* L.; codon bias; optimal codons; mutation pressure; natural selection

点地梅属 *Androsace* L. 是报春花科 (Primulaceae) 中的 1 个属, 约有 100 种, 广泛分布于北半球温带地区。在我国有 73 种 7 变种, 主要分布在青藏高原, 仅有部分种类分布于西北至内蒙古、东北以及秦岭以南各省区。该属是青藏高原地区垫状植物中分布面积最广、种类最多的一个类群^[1]。该属植物含有许多有用的次生代谢物, 如黄酮类、三萜皂苷类、酚苷、甾醇类等, 用药历史悠久, 可用于治扁桃体炎、咽喉炎、口腔炎、急性结膜炎、热性水肿、跌扑损伤以及咽喉肿痛等症^[2-4]。

叶绿体是植物细胞中的光合细胞器, 主要进行光合作用, 吸收光能将 CO₂ 转变成有机物。叶绿体也具有感知外界环境胁迫信号的能力, 参与像氨基酸、脂类、萜类、酚类等多种物质的代谢^[5]。在进化中叶绿体尽管将一部分基因转移到了核基因组, 但仍有自己的遗传体系, 能合成蛋白质, 是一种半自主型细胞器。随着测序技术的进步, 近年来大量植物的叶绿体基因组序列得以公布, 为研究叶绿体分子生物学提供了便利, 其中相当一部分植物有很高的药用价值。叶绿体基因组由于拷贝数多、序列保守等特性, 在群体遗传、进化、系统发育及物种鉴定中得到了广泛的应用^[6]。

基因经转录、翻译转变成蛋白质的过程中, 碱基以三联密码的形式翻译成氨基酸。一个氨基酸可能有几个密码子与其对应, 即密码子存在简并性。氨基酸对密码子的使用并不随机, 可能更偏好于使用某种密码子, 这称为密码子偏好性。不同物种密码子偏好性不同, 甚至同一物种, 不同基因密码子偏好性也可能不同^[7]。密码子偏好性是物种在长期进化中形成的。影响的主要因素是自然选择、突变和遗传漂变, 至于哪种因素起主要作用仍是研究的热点和争论的焦点^[5]。对叶绿体密码子偏性的研究在叶绿体基因工程中有重要应用。例如, 通过使用最优密码子对外源基因进行改造, 能实现外源基因在叶绿体的高效、稳定表达; 而且密码子偏性研究结果也可以为外源基因选择合适的受体物种; 密码子偏好性研究也对探索物种生命进化模式及基因

的选择性表达有重要意义^[8]。

目前有关点地梅属植物的研究主要集中在物种鉴定、系统发育、化学成分分析及药用价值等^[1,2,9-13]。关于叶绿体基因组的研究主要集中在测序、基因组特征描述及系统发育^[14-16], 而关于叶绿体密码子偏好性的研究还未见报道。已公布的叶绿体基因组序列共有 10 个, 本研究主要分析了这 10 种的叶绿体密码子偏好模式及影响因素。研究结果将为点地梅属植物的叶绿体基因工程、进化及系统发育提供参考。

1 材料与仪器

1.1 材料

样品采自青海省西宁市 (36°34'37"N, 101°48'27"E), 由甘肃农业大学林学院马彦军教授鉴定为西藏点地梅 *A. mariae* Kanitz, 送样南京集思慧远公司测序, 经组装、注释后, 上传至 NCBI (NC_051991.1)。其他点地梅属植物叶绿体基因组序列从 NCBI 下载, 包括景天点地梅 *A. bulleyana* G. Forr. (NC_034641.1)、秦巴点地梅 *A. laxa* C. M. Hu et Y. C. Yang (NC_039347.1); 峨眉点地梅 *A. paxiana* R. Knuth (NC_034640.1)、直立点地梅 *A. erecta* Maxim. (NC_057637.1)、北点地梅 *A. septentrionalis* Linn. (NC_080503.1)、东北点地梅 *A. filiformis* Retz. (NC_080502.1)、莲叶点地梅 *A. henryi* Oliv. (NC_080506.1)、腋花点地梅 *A. axillaris* (Franch.) Franch. (NC_080505.1)、贵州点地梅 *A. kouytchensis* Bonati (NC_080504.1)。

1.2 仪器

DK-S24 型电热恒温水浴锅 (上海精宏有限公司); iCEM-24R 型高速冷冻离心机 (杭州奥盛有限公司); One Drop OD-1000⁺型超微量分光光度计; JY300C 型电泳仪 (北京君意有限公司); Tanon-1600 型凝胶成像仪; Illumina NovaSeq 6000 型测序仪。

2 方法

2.1 叶绿体基因组图谱绘制

将西藏点地梅叶绿体基因组 gb 格式注释文件上传至 <https://chlorobox.mpimp-golm.mpg.de/OGD>

raw.html 网站, 用 OGDRAW 工具绘制叶绿体基因组圈图^[17]。

2.2 序列数据筛选

从叶绿体基因组挑选蛋白质编码基因, 编码区(coding sequence, CDS)长度不小于 300 bp^[18], 且以 ATG 为起始密码子, TAA、TAG、TGA 为终止密码子, 去除重复及内部有无法识别碱基的基因^[19]。以所有物种叶绿体基因组共有基因用于本研究各项分析^[5-6]。

2.3 密码子碱基组成分析

用 Codon W 1.4.2 计算 CDS 的相对同义密码子使用度(relative synonymous codon usage, RSCU)、有效密码子数(effective number of codons, ENC)等参数^[18]。用自己编写的 Python 程序计算密码子第 1、2、3 位及整个序列的 GC 含量(GC₁、GC₂、GC₃ 及 GC_{all})。用 R 的 psych 包做各变量间相关分析。

2.4 中性绘图分析

中性绘图用来衡量自然选择作用对密码子使用模式的影响^[19]。以 GC₃ 为横坐标, GC₁₂ (GC₁ 和 GC₂ 的平均值) 为纵坐标。如果 GC₁₂ 和 GC₃ 显著相关, 拟合直线回归系数接近 1, 说明 3 个位置上碱基组成无差异, 密码子使用偏好主要受突变影响; 反之, GC₁₂ 和 GC₃ 不相关, 拟合直线回归系数接近 0, 说明密码子 3 个位置碱基组成明显不同, 密码子偏好更多受自然选择影响^[19-21]。

2.5 ENC 绘图分析

ENC 绘图(ENC-plot) 分析各基因的密码子使用偏性及碱基组成对密码子偏性的影响。以 ENC 为纵坐标, GC₃ 为纵坐标, 绘制散点图, 并辅以标准曲线。标准曲线计算公式: $ENC = 2 + GC_3 + 29/[GC_3^2 + (1 - GC_3)^2]$ 。落在标准曲线上及下部附近的基因, 表示密码子偏性主要受突变影响, 而密码子偏性主要受选择影响的基因会落在远离标准曲线的下方^[18]。

2.6 最优密码子确定

以 ENC 值为标准, 对基因排序, 高、低各选 10%, 分别建立高、低表达基因库。计算高、低基因表达库 RSCU 差值($\Delta RSCU$), 以 $RSCU > 1$ 为高频密码子, $\Delta RSCU \geq 0.8$ 为高表达密码子, 将同时符合上述 2 个条件的密码子确定为最优密码子^[6]。

2.7 PR2 分析

PR2 (parity rule 2) 分析通过计算每个基因密码

子第 3 位上各种碱基的含量, 确定碱基组成对密码子使用偏性的影响。以 $G_3/(G_3 + C_3)$ 为横坐标, $A_3/(A_3 + T_3)$ 为纵坐标绘制散点图。为避免由第 3 位碱基对 A/T 或 T/A, G/C 或 C/G 的突变不均衡对密码子偏性的影响, 只选择丝氨酸(TCA、TCC、TCG、TCT)、亮氨酸(CTA、CTC、CTG、CTT)、脯氨酸、精氨酸(CGA、CGC、CGG、CGT)、苏氨酸、缬氨酸、丙氨酸和甘氨酸共 8 个氨基酸的密码子^[19,22]。

2.8 系统发育和聚类分析

系统发育分析以各物种共有基因序列绘制邻接树(neighbor-joining, NJ)。用 MAFFT 比对^[23], Gblocks 对比对结果修剪^[24], MEGA 11 画树, Bootstrap (BS) 检验值 1000 次^[25]。用 R 以密码子 RSCU 值做聚类分析, 以平方欧式距离计算基因间距离, 以平均距离聚类^[21,26]。本研究中热图用 R 的 pheatmap 包绘制, 其他图用 ggplot2 包绘制。

3 结果与分析

3.1 西藏点地梅叶绿体基因组特征

西藏点地梅叶绿体基因组全长 151 958 bp, 包括 1 个大单拷贝区(LSC, 83 292 bp), 1 个小单拷贝区(SSC, 16 744 bp), 2 个反向重复区(IR, 25 961 bp), 共 133 个基因, 87 个编码蛋白质, 8 个编码 rRNA, 37 个 tRNA (图 1)。

3.2 密码子组成特征

点地梅属已测序叶绿体基因组共 10 个, 长度变化从 149 456~153 547 bp (表 1)。经筛选, 每个叶绿体基因组得到符合条件的 41 个共有基因。密码子 3 个位点的平均 GC₁、GC₂、GC₃ 分别为 47.32%~47.57%, 39.26%~39.53%, 27.53%~28.14%。GC_{all} 平均含量为 38.04%~38.41% (表 1)。GC₁>GC₂>GC₃, 说明密码子各位点 GC 含量并不是均匀分布的, 而且这些值都小于 50%, 说明密码子更倾向于使用 A/U 碱基, 尤其是以 A/U 作为结尾碱基。不同基因间 GC 含量差异明显, 最低的是 *ycf1*, 28.84%, 最高的是 *rps11*, 44.70%, 相差 15.86% (图 2)。各物种间 3 个位点及 GC_{all} 变化很小。各物种 ENC 值为 46.80~48.04, 变化范围较小(表 1), 且均大于 35, 说明各物种叶绿体基因的密码子使用偏好性较弱。

各物种密码子组成参数相关性分析显示(表 2), 所有物种中 GC_{all} 与 GC₁、GC₂ 及 GC₃ 呈极显著相关, GC₁ 与 GC₂ 也呈极显著相关。GC₁ 与 GC₃ 在大

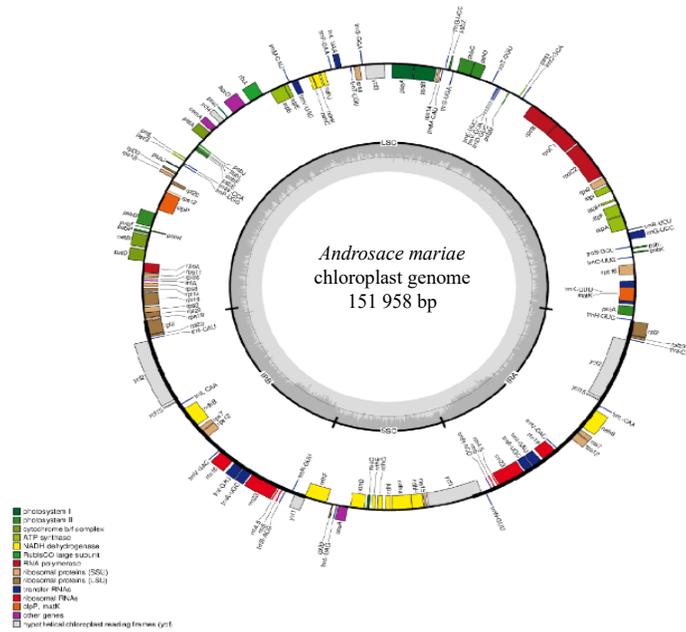


图 1 西藏点地梅叶绿体基因组环状图

Fig. 1 Map of *A. mariae* chloroplast genome

表 1 点地梅属植物叶绿体基因组密码子偏性基本参数

Table 1 Basic parameters of codon usage bias of chloroplast genome in *Androsace*

物种	GenBank ID	长度/bp	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC _{all} /%	ENC
腋花点地梅	NC_080505.1	151 013	47.48	39.43	27.59	38.17	47.07
景天点地梅	NC_034641.1	152 110	47.32	39.26	27.53	38.04	47.60
直立点地梅	NC_057637.1	153 547	47.53	39.35	27.65	38.18	47.65
东北点地梅	NC_080502.1	151 728	47.44	39.49	27.10	38.01	46.87
莲叶点地梅	NC_080506.1	149 456	47.47	39.40	27.60	38.16	47.25
贵州点地梅	NC_080504.1	150 792	47.39	39.35	27.30	38.01	46.80
秦巴点地梅	NC_039347.1	151 942	47.49	39.42	27.82	38.24	48.04
西藏点地梅	NC_051991.1	151 958	47.40	39.31	27.78	38.16	47.69
峨眉点地梅	NC_034640.1	150 528	47.46	39.43	27.38	38.09	47.09
北点地梅	NC_080503.1	150 709	47.57	39.53	28.14	38.41	47.53

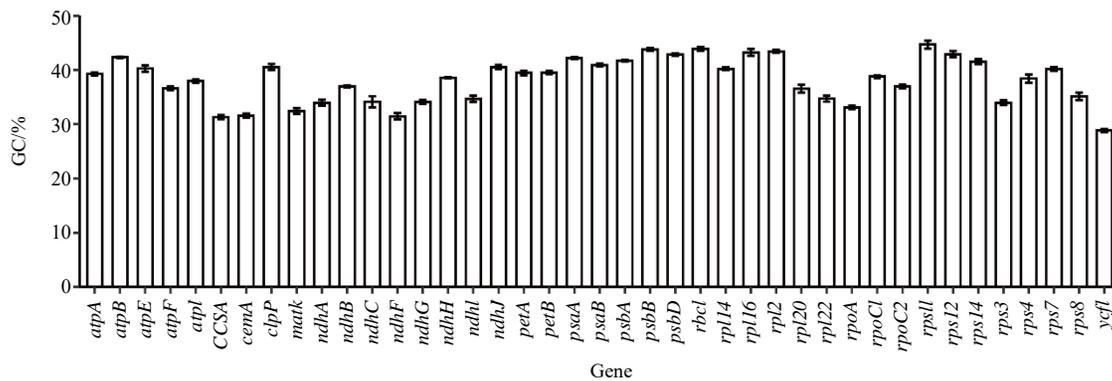


图 2 点地梅属叶绿体蛋白编码基因 GC 含量

Fig. 2 GC content of chloroplast protein coding genes in *Androsace*

表 2 密码子组成参数相关性分析

Table 2 Correlation analysis of codon composition parameters

物种	变量	相关系数				
		GC ₁	GC ₂	GC ₃	GC _{all}	ENC
腋花点地梅	GC ₂	0.542**				
	GC ₃	0.371*	0.275			
	GC _{all}	0.871**	0.825**	0.598**		
	ENC	0.175	-0.026	0.137	0.116	
	L _{aa}	-0.225	-0.273	-0.062	-0.261	-0.012
景天点地梅	GC ₂	0.539**				
	GC ₃	0.341*	0.231			
	GC _{all}	0.867**	0.815**	0.579**		
	ENC	0.197	-0.191	0.205	0.072	
	L _{aa}	-0.189	-0.247	-0.083	-0.238	-0.043
直立点地梅	GC ₂	0.521**				
	GC ₃	0.291	0.101			
	GC _{all}	0.864**	0.803**	0.501**		
	ENC	0.129	-0.217	0.235	0.027	
	L _{aa}	-0.206	-0.251	-0.096	-0.261	-0.085
东北点地梅	GC ₂	0.502**				
	GC ₃	0.365*	0.194			
	GC _{all}	0.870**	0.794**	0.573**		
	ENC	0.173	-0.109	0.333*	0.136	
	L _{aa}	-0.201	-0.253	-0.045	-0.238	0.017
莲叶点地梅	GC ₂	0.552**				
	GC ₃	0.367*	0.261			
	GC _{all}	0.874**	0.826**	0.589**		
	ENC	0.175	-0.087	0.143	0.088	
	L _{aa}	-0.242	-0.273	-0.074	-0.273	-0.043
贵州点地梅	GC ₂	0.528**				
	GC ₃	0.354*	0.285			
	GC _{all}	0.862**	0.824**	0.600**		
	ENC	0.195	-0.107	0.159	0.095	
	L _{aa}	-0.232	-0.272	-0.051	-0.262	0.043
秦巴点地梅	GC ₂	0.528**				
	GC ₃	0.273	0.209			
	GC _{all}	0.858**	0.822**	0.530**		
	ENC	0.109	-0.219	0.152	-0.006	
	L _{aa}	-0.203	-0.258	-0.115	-0.265	-0.094
西藏点地梅	GC ₂	0.544**				
	GC ₃	0.271	0.217			
	GC _{all}	0.857**	0.825**	0.544**		
	ENC	0.165	-0.246	0.220	0.033	
	L _{aa}	-0.202	-0.254	-0.118	-0.261	-0.072
峨眉点地梅	GC ₂	0.530**				
	GC ₃	0.354*	0.298			
	GC _{all}	0.862**	0.825**	0.605**		
	ENC	0.201	-0.051	0.044	0.092	
	L _{aa}	-0.226	-0.271	-0.057	-0.258	-0.015
北点地梅	GC ₂	0.556**				
	GC ₃	0.348*	0.282			
	GC _{all}	0.868**	0.838**	0.582**		
	ENC	0.227	-0.135	0.191	0.102	
	L _{aa}	-0.211	-0.277	-0.061	-0.256	0.026

*P<0.05 **P<0.01。

部分物种中呈显著相关, GC₂与GC₃在所有物种中相关性不显著。各位置GC含量相关性表明,第1位与第2位碱基组成相似,与第3位在部分物种中存在一定相似性,但第2位与第3位差异较大。ENC与GC₁、GC₂及GC_{all}相关性都未达到显著,只有在1个物种中(东北点地梅)与GC₃相关性达到显著,说明碱基组成对密码子偏性有影响,但不是主要因素。ENC与蛋白质长度(L_{aa})之间的相关性在所有物种中都未达到显著,说明基因长度对ENC没有影响。

根据RSCU值分析各密码子使用频率时,由于Trp及Met的密码子只有1个,不存在偏性,去除这2个密码子。密码子RSCU值大于1,表示使用频率较高,反之表示使用频率较低。10个物种中每个密码子RSCU值接近,说明各物种对密码子偏好类似(图3)。RSCU最大值是UUA(2.22),最小值是CUC(0.3),这2个极值都存在于东北点地梅。RSCU值大于1的30个,小于1的32个,其中大于1且以A/U结尾的29个,只有1个以G结尾(UUG),小于1且以A/U结尾的只有3个(CUA、UGA和AUA)。这表明点地梅属各物种叶绿体基因组基因偏好使用以A/U结

尾的同义密码子。

3.3 最优密码子分析

根据ENC排序,对建立的基因高、低表达库密码子使用进行分析,共筛选获得了25个最优密码子(表3)。其中,腋花点地梅最少(13个),直立点地梅最多(17个),其他物种最优密码子数介于两者之间。有6个是所有物种都有的共有最优密码子:UGU、GUA、GGU、GCU、AAA和UCU。25个最优密码子中,有24个以A/U结尾,只有1个以G结尾(编码亮氨酸的UUG密码子)。

3.4 中性绘图分析

中性绘图结果显示,点地梅属10个物种中叶绿体基因都分布于对角线上方(图4),且大部分基因离对角线较远。各物种的回归系数都较小,在0.3326~0.5771,最小的是直立点地梅,最大的是北点地梅。相关系数的平方(*r*²)为0.0509~0.1399,最小的是直立点地梅,最大的是峨眉点地梅。较小的回归系数表明点地梅叶绿体基因密码子GC₁₂相对于GC₃中性突变的程度较小,同时也表明GC₁₂与GC₃进化模式存在较大差异。

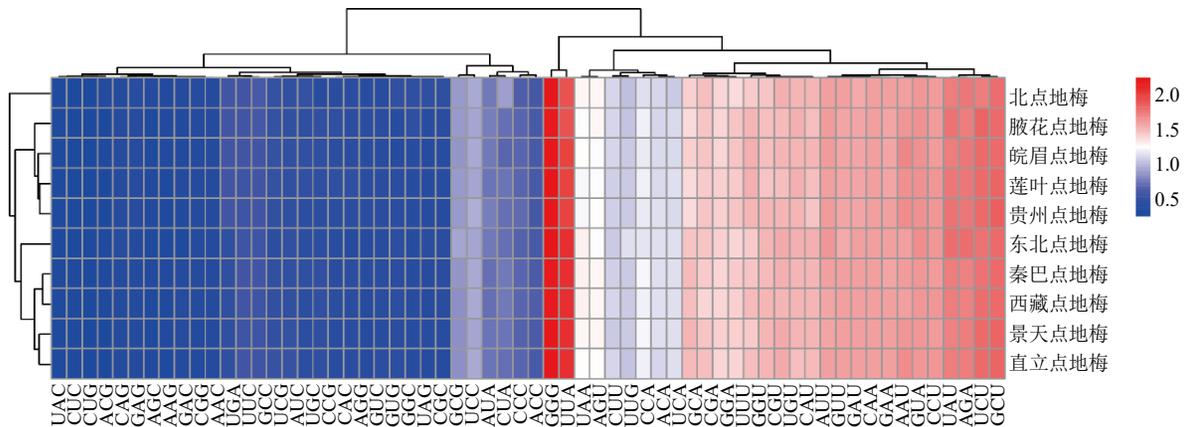


图3 各物种叶绿体基因密码子RSCU

Fig. 3 RSCU values of chloroplast gene codon in each species

表3 点地梅属各物种叶绿体基因组最优密码子

Table 3 Optimal codons of chloroplast genome in each species of *Androsace*

物种	最优密码子	数量
腋花点地梅	UCU、UGU、UUA、UAA、CUA、AUU、AAU、AAA、GUU、GCU、GGU、GUA、GAA	13
景天点地梅	UCU、UGU、UUA、UAA、CUU、CGU、AUU、AAU、AAA、GUU、GCU、GAU、GGU、GUA、GAA	15
直立点地梅	UCU、UAU、UGU、UUA、UAA、CCU、CGU、AUU、AAU、AAA、GUU、GCU、GAU、GGU、GUA、GAA、GGA	17
东北点地梅	UCU、UGU、UUA、UAA、UUG、CCU、CGU、CCA、AUU、ACU、AAA、GCU、GGU、GUA、GAA	15
莲叶点地梅	UCU、UGU、UAA、UUG、CCU、CUA、CGA、AUU、AAA、GUU、GCU、GGU、GUA、GAA	14
贵州点地梅	UCU、UGU、UUA、CUU、CCU、CAA、ACU、ACA、AAA、GUU、GCU、GGU、GUA、GAA	14
秦巴点地梅	UCU、UGU、UUA、CGU、CGA、AUU、AAU、AAA、GUU、GCU、GAU、GGU、GUA	13
西藏点地梅	UCU、UAU、UGU、UUA、UAA、CUU、CGU、CGA、AUU、AAU、AAA、GUU、GCU、GGU、GUA、GAA	16
峨眉点地梅	UCU、UGU、UUA、UUG、CCU、CGA、AUU、ACU、AAA、GUU、GCU、GGU、GUA、GAA	14
北点地梅	UCU、UGU、UUG、CCU、CGU、CGA、AUU、ACU、ACA、AAA、GUU、GCU、GGU、GUA、GAA	15

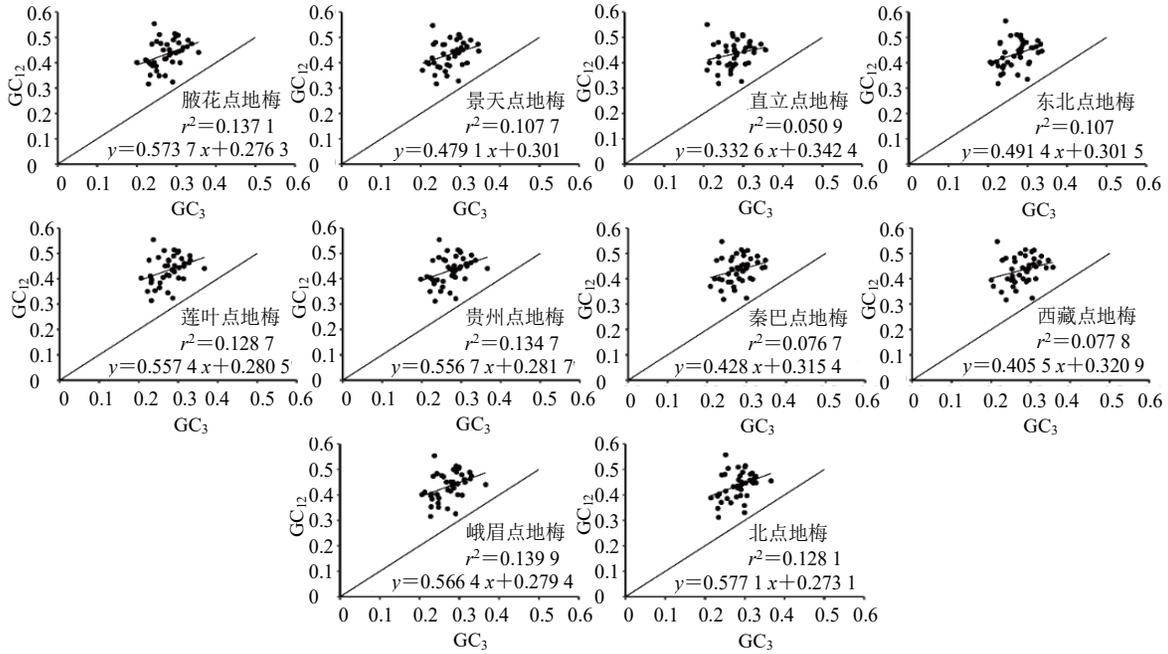


图4 中性绘图分析

Fig. 4 Analysis of neutrality plot

3.5 ENC-plot 绘图分析

ENC-plot 图显示各基因 GC3s 值主要分布在 0.2~0.4, ENC 值分布在 35~60 (图 5)。各基因间 ENC 值差异较大, 表明密码子偏性相差较大。从图中可以看出有相当一部分基因分布于标准曲线附近, 大体呈对称分布。分布于标准曲线上或下方

近的基因密码子偏性主要受突变压影响。分布于曲线下方, 且离曲线较远的基因受到了更强的选择作用影响。

ENC 比值频数分布定量显示了 ENC 观察值与期望值之间的差异 (图 6)。图中显示大部分基因位于 -0.05~0.05, 占 57%, 而分布在 -0.05~0.05

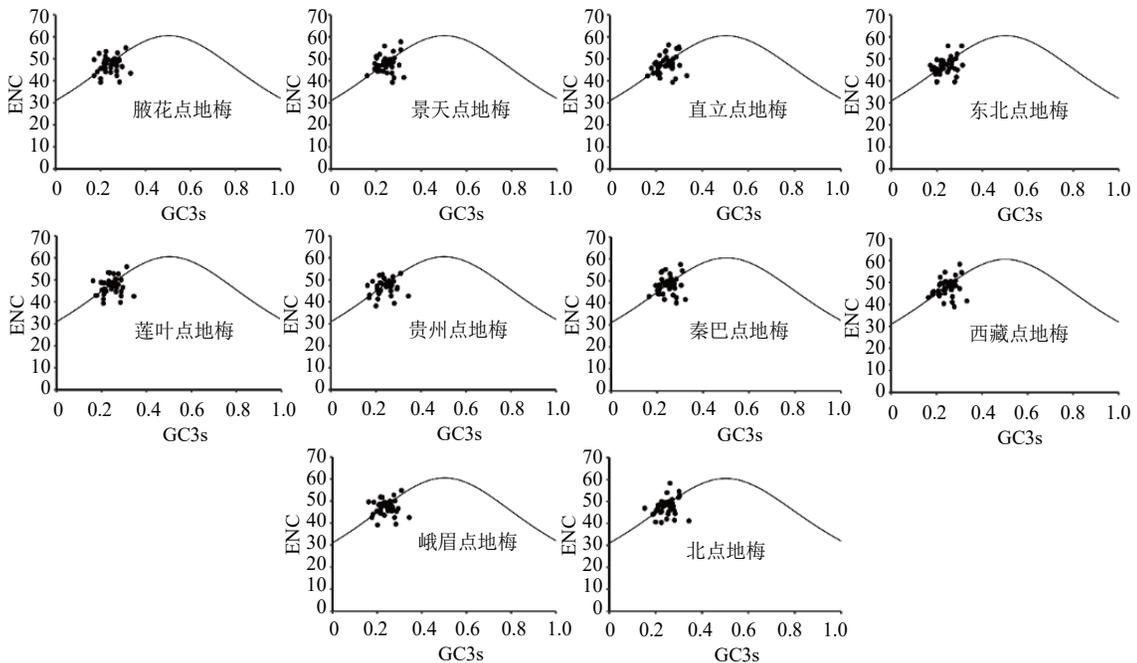


图5 ENC-plot 分析

Fig. 5 Analysis of ENC-plot

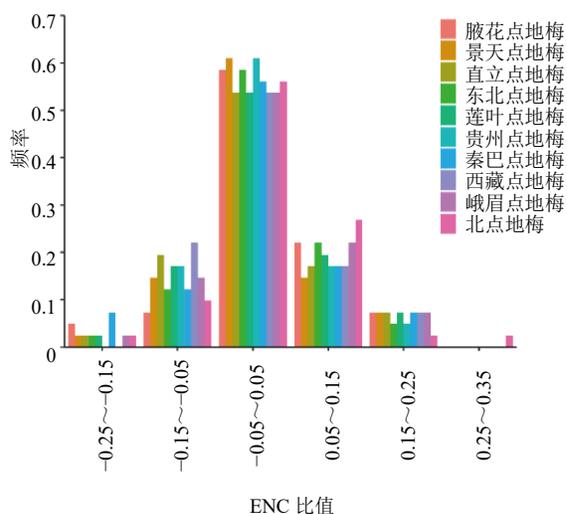


图 6 ENC 比值频数分布

Fig. 6 Frequency distribution of ENC ratio

区间之外的基因只有 43%。各物种分布频数相似。ENC 比值频数分布结果也表明大部分基因位于标准曲线附近，而与 ENC 期望值差异较大的，也就是离标准曲线较远的只有少部分基因。

ENC-plot 图及 ENC 比值频数分布结果表明点地梅叶绿体基因密码子偏性差异与 GC3s 差异有关，即大量基因的密码子偏性受到突变的影响。

3.6 PR2-plot 绘图分析

PR2-plot 图分析了 4 个密码子氨基酸第 3 位碱基使用的情况 (图 7)。有部分基因围绕中心点附近分布，但大部分分布在 4 个区域，且不均匀，位于下方的基因明显多于上方，其中左下方最多，这说明氨基酸使用的密码子中第 3 位 T 高于 A，C 高于 G，T+C 的频率高于 A+G，即第 3 位使用嘧啶的频率高于嘌呤。如果突变和选择作用对 DNA 两条

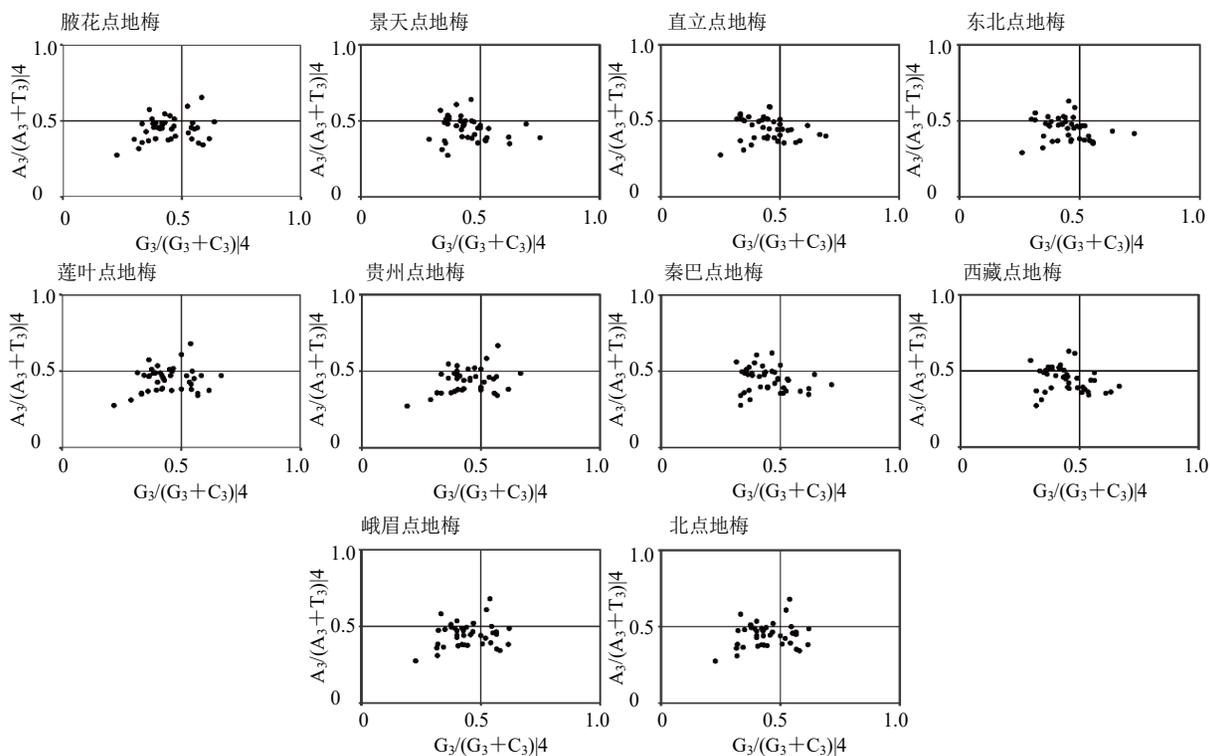


图 7 PR2-plot 分析

Fig. 7 Analysis of PR2-plot

链的作用不存在偏性，理论上在平衡时，应该 A=T，G=C，也就是基因应该分布在中心点附近，而离中心点越远，说明突变和选择作用对 DNA 两条链作用的偏性程度越大。

3.7 聚类分析

为分析系统发育与密码子使用模式之间的关系，以各物种 41 个基因的 CDS 序列构建了 NJ 及以密

码子 RSCU 的聚类分析 (图 8)。聚类分析表明，秦巴点地梅、景天点地梅、西藏点地梅和直立点地梅形成一组，密码子使用模式接近，其中秦巴点地梅和景天点地梅在组内又聚成一小组，说明这 2 个物种的密码子使用模式最相似；峨眉点地梅、腋花点地梅、莲叶点地梅及贵州点地梅形成另一组，相互之间密码子使用模式接近，其中最为相似的是峨眉

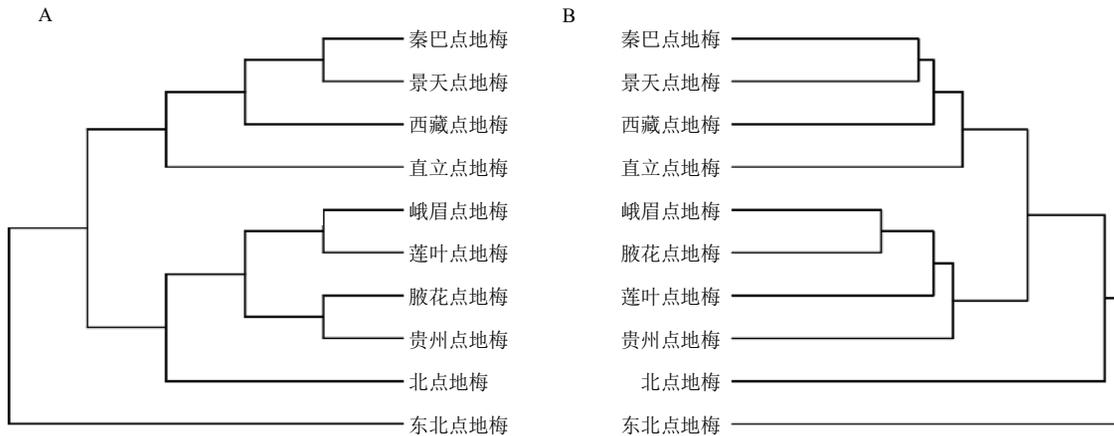


图 8 基于基因 CDS 的系统发育分析 (A) 和基于密码子 RSCU 的聚类分析 (B)

Fig.8 Phylogenetic analysis based on gene CDS (A) and cluster analysis based on codon RSCU (B)

点地梅和腋花点地梅；这 2 组聚成一大组和北点地梅并列，最外侧是东北点地梅。

系统发育显示了这些物种之间的亲缘关系，与聚类分析结果比较显示，亲缘关系越近，密码子使用模式也越近。其中，秦巴点地梅、景天点地梅、西藏点地梅和直立点地梅的系统发育分支关系与聚类分组完全相同；而其他物种的系统发育与聚类结果之间也基本相似。聚类与系统发育的比较表明，聚类结果虽然不能真正反映物种间的进化关系，但可以作为一种参考方法用于物种分类及系统发育研究中。

4 讨论

密码子各位点 GC 含量是受突变压与选择共同作用的结果。相对于第 1 和第 2 位，密码子第 3 位突变一般不会造成编码氨基酸的改变，是一个近中性位点^[20]。除非在高表达基因中，某些氨基酸第 3 位点突变，而与之相对应的 tRNA 丰度较低，影响到翻译速率或准确度时，才会受到选择作用^[22]，因此该位点就 GC 含量而言，主要反映的是突变压的大小，而第 1 和第 2 位的突变通常会改变氨基酸，要受到选择作用限制^[20]。本研究中的点地梅属植物叶绿体碱基组成显示，在蛋白编码基因的密码子组成中 $GC_1 > GC_2 > GC_3$ ，显示出较小的 GC_3 含量，只有 27.53%~28.14%，说明该属植物向 GC 方向的突变压较小，而且各基因间变动范围基本在 0.2~0.4，说明在叶绿体基因组内 GC 含量的异质性较小。这与在其他叶绿体基因组密码子碱基组成研究中的结论类似^[5,19,21,27-30]，而且这一特性与早期单细胞

生物研究中的结果类似，但与高等生物基因组 GC 含量有很大异质性有明显不同^[20]。

本研究的中性绘图结果显示各基因 GC_3 含量的变化与 GC_{12} 线性关系很弱，而回归系数所体现的是突变压对密码子偏性的影响程度。弱的线性关系及小的回归系数说明突变压不是影响密码子偏性的主要因素^[19,30]。密码子组成参数的相关性分析也显示在 GC 含量上 GC_1 和 GC_2 是呈极显著相关的，而与 GC_3 相关性较低，这也说明突变压不是影响密码子偏性的主要因素，在 3 个位点中，第 1、2 位的变化机制与第 3 位明显不同，选择作用是影响点地梅属植物叶绿体蛋白编码基因密码子偏好性的主要因素。ENC-plot 的分析结果显示基因间有较大的 ENC 变化，各基因密码子偏性差异较大。但 ENC 值都大于 35，说明尽管各基因间偏性差异随大，但却没有偏性很强的基因。处于标准曲线下方越远的基因受到选择作用的影响越强^[18]。结合中性绘图结果发现，尽管选择作用在影响点地梅属植物叶绿体基因密码子偏性中是主要因素，但大多数基因只受到了较弱的选择作用。ENC-plot 图也显示还有许多基因分布在标准曲线上或下方附近，比值频数分布图也显示有大量基因分布在 -0.05~0.05。因此，在密码子偏性中，突变压也起了很大作用，一些其他植物叶绿体密码子偏性的研究结果也显示了类似的情况^[19, 30-31]。PR2-plot 图显示大部分基因离中心点的距离并不远，说明突变和选择作用对 DNA 两条链的作用偏性程度较弱，而有个别基因偏离中心点较远，说明受到了较强的选择作用^[22]，主要是编

码光合作用中有重要功能的蛋白基因, 其功能受选择作用限制, 例如 *psbD* 基因编码蛋白是光系统中更新最快的一个蛋白, 与光合作用的光抑制有关, 受到光抑制的光系统, 需要及时更换该蛋白以修复光系统^[32-33]。

在进化上密码子使用模式是保守的, 一旦 1 个基因的密码子偏性模式建立, 通常与一种基础代谢途径相关, 在进化中不会轻易改变^[22]。因此, 亲缘关系较近的物种有相似的密码子偏性, 从本研究的聚类分析及系统发育分析可以清楚的看到这一点。亲缘关系近的, 密码子使用模式也聚为一类, 有许多相似的研究结果验证了这种情况^[21,30,34-35]。本研究的其他分析中, 无论是碱基组成、中性绘图、ENC-plot 及 PR2-plot 也都体现出点地梅属各物种间尽管个别基因密码子偏性模式有小的差异, 但大多数基因是相似的, 表现出了相似的密码子使用模式。

本研究揭示的点地梅属叶绿体密码子偏性模式及影响因素表明, 该属大部分基因密码子偏性较弱, 叶绿体基因组 GC 含量异质性较低, 共鉴定了 13~17 个最优密码子, 这些密码子大多以 A/U 结尾。影响密码子偏性的主要因素是选择作用, 但大多数基因仅受到了较弱的选择作用, 突变也是影响密码子偏性的重要因素, 只有少数光合系统及核糖体基因由于功能限制, 受到了较强的选择作用, 表现出较低的 ENC 值。本研究的结果将为点地梅属分子进化、系统发育及叶绿体基因工程提供有用参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 王玉金, 李小娟, 郝刚, 等. 点地梅属的分子系统学、生物地理学和垫状形态的趋同进化 [J]. 植物分类学报, 2004, 42(6): 481-499.
- [2] 何希瑞, 魏桂芳, 姚宏, 等. 点地梅属植物化学成分与药理活性研究进展 [J]. 中国实验方剂学杂志, 2012, 18(11): 296-299.
- [3] 黄景, 何开家, 苏华, 等. 点地梅提取物对大鼠股骨创伤愈合的影响 [J]. 重庆医学, 2017, 46(19): 2608-2611.
- [4] 殷志琦, 王磊, 李承花, 等. 点地梅中活性三萜皂苷的研究 [A] // 中国化学会第十届全国天然有机化学学术会议论文集 [C]. 广州: 中国化学会, 2014: 231.
- [5] 马孟莉, 张薇, 孟衡玲, 等. 豆蔻属药用植物叶绿体基因组密码子偏性分析 [J]. 中草药, 2021, 52(12): 3661-3670.
- [6] 冯展, 江媛, 郑燕, 等. 肉苁蓉属植物叶绿体基因组密码子偏好性分析 [J]. 中草药, 2023, 54(5): 1540-1550.
- [7] Parvathy S T, Udayasuriyan V, Bhadana V. Codon usage bias [J]. *Mol Biol Rep*, 2022, 49(1): 539-565.
- [8] 李显煌, 杨生超, 辛雅萱, 等. 灯盏花叶绿体基因组密码子偏好性分析 [J]. 云南农业大学学报: 自然科学, 2021, 36(3): 384-392.
- [9] 黄荣福. 雅江点地梅 (报春花科) 3 个居群的核型和倍性变化 [J]. 西北植物学报, 2001, 21(3): 526-531.
- [10] 王鑫晶, 王建娜, 萨仁高娃, 等. 蒙药点地梅 HPLC 指纹图谱研究 [J]. 中国中医药信息杂志, 2021, 28(8): 98-102.
- [11] 易冉, 刘瑶, 刘紫薇, 等. 河南点地梅属一新记录种: 细蔓点地梅 [J]. 河南农业大学学报, 2018, 52(4): 618-620.
- [12] 黄先丽, 王晓静, 贾献慧. 点地梅的挥发油成分分析 [J]. 食品与药品, 2009, 11(2): 32-34.
- [13] 向彬, 杨杰, 周春花, 等. 点地梅的抗炎镇痛作用研究 [J]. 中国医药科学, 2014, 4(3): 31-33.
- [14] Peng C, Yu C C, Xing Y W. The complete chloroplast genome of *Androsace erecta* (Primulaceae) and its phylogenetic implication [J]. *Mitochondrial DNA B Resour*, 2021, 6(7): 1987-1989.
- [15] Ren T, Yang Y C, Zhou T, et al. Comparative plastid genomes of *Primula species*: Sequence divergence and phylogenetic relationships [J]. *Int J Mol Sci*, 2018, 19(4): 1050.
- [16] Guo Y P, Ma R, Xie J, et al. The complete chloroplast genome of *Androsace mariae* [J]. *Mitochondrial DNA B Resour*, 2021, 6(2): 376-377.
- [17] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47(W1): W59-W64.
- [18] Wright F. The 'effective number of codons' used in a gene [J]. *Gene*, 1990, 87(1): 23-29.
- [19] 尚明照, 刘方, 华金平, 等. 陆地棉叶绿体基因组密码子使用偏性的分析 [J]. 中国农业科学, 2011, 44(2): 245-253.
- [20] Sueoka N. Directional mutation pressure and neutral molecular evolution [J]. *Proc Natl Acad Sci USA*, 1988, 85(8): 2653-2657.
- [21] 代江鹏, 蔡一鸣, 刘巧珍, 等. 甘草属 7 种植物叶绿体基因组密码子使用偏好性分析 [J]. 中草药, 2023, 54(9): 2907-2916.
- [22] Sueoka N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons [J]. *J Mol Evol*, 1995, 40(3): 318-325.
- [23] Katoh K, Misawa K, Kuma K I, et al. MAFFT: A novel method for rapid multiple sequence alignment based on

- fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30(14): 3059-3066.
- [24] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis [J]. *Mol Biol Evol*, 2000, 17(4): 540-552.
- [25] Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11 [J]. *Mol Biol Evol*, 2021, 38(7): 3022-3027.
- [26] 高守舆, 李钰莹, 杨志青, 等. 白羊草叶绿体基因组密码子使用偏好性分析 [J]. *草业学报*, 2023, 32(7): 85-95.
- [27] 夏晞, 彭劲谕, 王大玮, 等. 3种榕属叶绿体基因组密码子偏好性分析 [J]. *西北林学院学报*, 2022, 37(5): 88-94.
- [28] 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.
- [29] 丁祥青, 陈丝雨, 陈佳婷, 等. 11种金花茶叶绿体基因组密码子偏好性分析 [J]. *福建农林大学学报: 自然科学版*, 2023, 52(4): 473-479.
- [30] Liu H B, Lu Y Z, Lan B L, *et al.* Codon usage by chloroplast gene is bias in *Hemiptelea davidii* [J]. *J Genet*, 2020, 99: 8.
- [31] Zhang P P, Xu W B, Lu X, *et al.* Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species [J]. *Physiol Mol Biol Plants*, 2021, 27(12): 2727-2737.
- [32] Takahashi S, Bauwe H, Badger M. Impairment of the photorespiratory pathway accelerates photoinhibition of photosystem II by suppression of repair but not acceleration of damage processes in *Arabidopsis* [J]. *Plant Physiol*, 2007, 144(1): 487-494.
- [33] Nishiyama Y, Allakhverdiev S I, Murata N. Protein synthesis is the primary target of reactive oxygen species in the photoinhibition of photosystem II [J]. *Physiol Plant*, 2011, 142(1): 35-46.
- [34] 季凯凯, 宋希强, 陈春国, 等. 木兰科叶绿体基因组的密码子使用特征分析 [J]. *中国农业科技导报*, 2020, 22(11): 52-62.
- [35] 黄祥, 楚光明, 郑新开, 等. 睡莲属叶绿体基因组密码子偏好性及系统发育分析 [J]. *中国农业科技导报*, 2022, 24(4): 75-84.

[责任编辑 时圣明]