

甘草属 7 种植物叶绿体基因组密码子使用偏好性分析

代江鹏¹, 蔡一鸣¹, 刘巧珍¹, 高晓霞², 朱 爽^{1*}

1. 广东药科大学 生命科学与生物制药学院广东省生物技术候选药物研究重点实验室, 广东 广州 510006

2. 广东药科大学药学院, 广东 广州 510006

摘要: **目的** 探究甘草属 7 种植物叶绿体基因组密码子使用模式及影响因素。**方法** 从 NCBI 获取国内已有的 7 种甘草属植物叶绿体全基因组, 筛选得到 49 条共有蛋白编码序列, 使用 codonW 1.44、EMBOSS、SPSS 26.0 等分析密码子偏好性相关参数。

结果 7 种甘草属物种叶绿体基因组密码子不同位置的 GC 含量均小于 50%, 共得到了 29 个高频密码子, 其中有 28 个密码子第 3 位碱基为 A/U, 筛选得到的 17 个最优密码子第 3 位碱基均为 A/U。7 种甘草属植物叶绿体蛋白编码基因的 ENC 值均大于 35, 密码子偏好性较弱。PR2-plot 分析、ENC-plot 分析和中性绘图分析结果显示突变压力和自然选择均为这 7 种甘草属物种叶绿体基因组密码子偏好性形成的重要影响因素, 但自然选择对密码子偏好性形成的贡献更大。基于叶绿体编码序列构建的系统发育树与基于 RSCU 的聚类分析结果具有一定的相似性, 叶绿体基因组密码子偏好性与物种的系统发育关系存在一定的联系。**结论** 7 种甘草属植物叶绿体基因组密码子第 3 位倾向于使用 A/U 碱基, 自然选择是密码子偏好性形成的主要影响因素。

关键词: 叶绿体基因组; 甘草属; 密码子偏好性; 最优密码子; 粗毛甘草; 光果甘草; 胀果甘草; 刺果甘草; 圆果甘草; 甘草; 云南甘草

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)09-2907-10

DOI: 10.7501/j.issn.0253-2670.2023.09.023

Analysis of codon usage bias of chloroplast genome in seven *Glycyrrhiza* species

DAI Jiang-peng¹, CAI Yi-ming¹, LIU Qiao-zhen¹, GAO Xiao-xia², ZHU Shuang¹

1. Guangdong Province Key Laboratory for Biotechnology Drug Candidates, School of Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou 510006, China

2. School of Pharmacy, Guangdong Pharmaceutical University, Guangzhou 510006, China

Abstract: Objective To explore codon usage patterns and affecting factors of chloroplast genome in seven *Glycyrrhiza* species.

Methods A total of 49 common protein coding sequences were screened from the complete chloroplast genomes of seven *Glycyrrhiza* species in China downloaded from the NCBI database. The relative parameters of codon usage bias were analyzed by codonW 1.44, EMBOSS, SPSS 26.0. **Results** All of the GC content of different codon positions in seven *Glycyrrhiza* species were less than 50%. The third base of 28 codons were A/U among 29 high-frequency codons, and all the third base of 17 optimal codons were A/U. The ENC values of all the chloroplast genomic protein coding genes in seven *Glycyrrhiza* species were more than 35, showing the weak codon usage bias. The analysis of PR2-plot, ENC-plot and neutral plot analysis revealed that both of mutation pressure and natural selection were the important factors influencing the formation of codon usage preference, but natural selection contributed more in the process. The similarity between phylogenetic tree based on the chloroplast coding sequences and RSCU cluster analysis indicated that the codon usage preference was associated with the phylogenetic relationship.

Conclusion The third position of codon tends to use A/U base in seven *Glycyrrhiza* chloroplast genome, and natural selection is the main factor affecting the formation of codon usage bias.

Key words: chloroplast genome; *Glycyrrhiza* Linn.; codon usage bias; optimal codons; *G. aspera* Pall.; *G. glabra* Linn.; *G. inflata* Batal.; *G. pallidiflora* Maxim.; *G. squamulosa* Franch.; *G. uralensis* Fisch.; *G. yunnanensis* Cheng f. et L. K. Dai ex P. C. Li

收稿日期: 2022-10-06

基金项目: 广东省省级科技计划项目 (2017A020213014)

作者简介: 代江鹏, 男, 硕士研究生。Tel: 17852250626 E-mail: 1056525712@qq.com

*通信作者: 朱 爽, 硕士生导师, 副教授, 研究方向为中药分析与鉴定。Tel: (020) 39352021 E-mail: 15683727@qq.com

甘草属 *Glycyrrhiza* Linn. 为豆科 (Fabaceae) 多年生草本植物, 在全世界约有 20 种, 我国有 8 种, 主要分布于黄河流域以北的广大地区^[1]。甘草属植物具有极高药用价值和经济价值, 其药用部位为根和根茎, 是中国、日本、土耳其、埃及等国家的传统药用植物^[2]。甘草 *G. uralensis* Fisch.、胀果甘草 *G. inflata* Batal. 或光果甘草 *G. glabra* Linn. 为《中国药典》2020 年版规定药用甘草来源^[3]。甘草属植物含有甘草酸、甘草素等多种活性成分, 其提取物具有保肝、抗炎、抗病毒、抗肿瘤、神经保护等药理作用^[4]。其提取物除了药用外, 还可以用做食品添加剂, 如其主要成分甘草酸是一种高甜度、低热量、价廉易获得、安全可靠无毒的天然甜味剂, 被广泛应用于茶、糖果、甜点等食品饮料生产加工行业^[5]。近年来因甘草需求量的提高、环境污染以及不合理的开发利用使得野生甘草资源日益减少, 仅凭野生甘草难以满足市场需求, 除了需加强对野生甘草资源的保护外, 还应扩大甘草的人工栽培, 但由于所处生长环境的不同, 栽培甘草与野生甘草的化学成分含量存在较大差异^[6], 部分栽培甘草的化学成分含量甚至达不到《中国药典》的要求^[7], 最终就会导致药效的差异, 故加强对优质甘草的培育势在必行。

生物在长期的进化过程中受多种因素影响形成不均等使用同义密码子的现象称为密码子偏好性, 即有的密码子使用频率明显高于与其编码同一种氨基酸的同义密码子^[8]。密码子使用偏好性不仅反映物种的起源、进化和突变模式, 还和基因表达水平息息相关。依照密码子偏好性优化基因可提高外源基因在目标转基因受体中的表达水平^[9], 如寇莹莹等^[10]在植酸酶 (*phyA*) 基因在大豆中的表达研究中发现, 按照大豆密码子偏好性特点人工改造植酸酶 *phyA* 基因比原始 *phyA* 基因所表达的植酸酶活性更高。叶绿体是高等植物进行光合作用的主要场所, 为半自主细胞器, 有独立的基因组和蛋白质合成机制, 遗传方式为母系遗传^[11]。叶绿体基因组大小适中, 结构保守, 有大量遗传信息, 通过对不同物种的叶绿体基因组进行比较分析, 可以了解植物物种之间的系统发育关系, 探究其进化机制^[12]。此外, 与核基因转化相比, 叶绿体基因转化具有外源基因在叶绿体中的表达效率高、定点整合、无位置效应、遗传稳定等优点^[13]。近 10 年来, 随着测序技术的进步, 大量植物的叶绿体基因组序列被公布, 叶绿体密码子偏好性的研究逐渐引起科学界关注, 一些植物如豆蔻属^[14]、芒属^[15]、稻属^[16]、绞股

蓝属^[17]的叶绿体密码子偏好性已得到研究, 这些研究有助于构建成成熟稳定的叶绿体转基因系统和探讨叶绿体基因组的进化过程。但目前尚未有关于甘草属叶绿体基因组密码子偏好性的研究发表, 本研究分析了 7 种甘草属植物叶绿体基因组密码子使用情况, 探讨了影响密码子偏好性的因素, 为甘草属叶绿体基因组的进化研究和后续通过叶绿体转基因技术改良物种性状以培育优质栽培甘草新种奠定基础。

1 材料

从 NCBI 的 GenBank 数据库中下载粗毛甘草 *G. aspera* Pall.、光果甘草 *G. glabra* Linn.、胀果甘草 *G. inflata* Batal.、刺果甘草 *G. pallidiflora* Maxim.、圆果甘草 *G. squamulosa* Franch.、甘草 *G. uralensis* Fisch.、云南甘草 *G. yunnanensis* Cheng f. et L. K. Dai ex P. C. Li 的叶绿体基因组, 依据注释选取序列长度大于 300 bp 的蛋白编码序列 (coding sequence, CDS) 用于后续分析, 并确保所提取的序列以 ATG 为起始密码子, 以 TAA、TAG、TGA 为终止密码子, 且终止密码子未提前出现于序列中间。

2 方法

2.1 密码子偏好性相关参数的计算

使用 CodonW 1.4.4 和 EMBOSS 计算以下参数: 相对同义密码子使用度 (relative synonymous codons usage, RSCU)、有效密码子数 (effective number of codons, ENC)、密码子数 (N)、总 GC 含量 (GCall) 和密码子第 1 位、第 2 位、第 3 位的 GC 含量 (GC1、GC2、GC3), 并使用 SPSS 26.0 软件对各参数进行相关性分析。

2.2 最优密码子分析

将提取的蛋白编码基因的 ENC 值从高到低进行排序, 分别选取前 10% 及后 10% 的基因建立高低偏性库, 将高低偏性组 RSCU 值相减得到 $\Delta RSCU$ ($RSCU_{high} - RSCU_{low}$)。符合 $RSCU > 1$ 且 $\Delta RSCU > 0.08$ 的密码子为最优密码子^[14]。

2.3 PR2-plot 偏倚分析

研究表明, 密码子第 3 位碱基是影响其偏好性的关键因素, 本研究选用密码子第 3 位碱基以 A3/ (A3+T3) 为纵坐标, G3/(G3+C3) 为横坐标绘制散点图^[17], 进行 PR2-plot 图分析, 探究其偏好性除突变外是否受其他因素的影响。

2.4 ENC-plot 分析

以 GC3 为横坐标, ENC 为纵坐标绘制散点图, 并依据公式 $ENC = 2 + GC3 + 29/[GC3 + (1 - GC3)^2]$

绘制 ENC 期望值标准曲线,当密码子偏好性主要受突变压力影响时,基因的 ENC 值接近标准曲线或分布于标准曲线上,当密码子偏好性主要受自然选择影响时,基因的 ENC 值远离标准曲线^[19]。为更准确的衡量 ENC 理论值 (ENCexp) 和 ENC 实际值 (ENCobs) 的差异大小,计算每个基因 ENC 比值即 (ENCexp-ENCobs)/ENCexp, 并做频率分布图。

2.5 中性绘图分析

以 GC1、GC2 的平均值 GC12 为纵坐标, GC3 为横坐标绘制散点图并进行回归分析, 当回归系数接近 1 时,密码子偏好性更多受到基因突变的影响, 当回归系数接近 0 时,则更多受自然选择的影响^[18]。

2.6 系统发育分析

2.6.1 基于叶绿体编码序列的系统发育分析 从 NCBI 数据库中下载紫藤 *Wisteria sinensis* Sweet (NC_029406)、白花紫藤 *W. brachybotrys* (Lindl.) Rehd. et Wils. (MT120814)、多花紫藤 *W. floribunda* (Willd.) DC. (KM103376)、腺车轴草 *Trifolium glanduliferum* (NC_025744) 的叶绿体基因组, 并提取蛋白编码序列, 与 7 种甘草属物种的叶绿体蛋白编码基因一同用于后续系统发育分析。使用 Mafft v 7.49 对共有蛋白编码序列进行序列对比, 手动调整后

据集导入 MEGA v7.0 中采用邻接法 (neighbor-joining, NJ) 和最大似然法 (maximum likelihood, ML) 构建系统发育树, bootstrap 检验 1000 次。

2.6.2 基于 RSCU 的聚类分析 将 11 个物种的蛋白编码序列按 3.1 标准过滤后, 以每个物种作为一个分析对象, 以 RSCU 为变量 (剔除 AUG、UGG、UAA、UAG、UGA) 在 SPSS 26.0 中进行聚类分析^[20]。

3 结果与分析

3.1 叶绿体基因组密码子组成和偏好性相关参数分析

每个甘草属植物叶绿体基因组均包含 76 条蛋白编码序列, 按 3.1 标准甘草属每个物种各提取 49 条序列用于密码子偏好性分析, 如表 1 所示, 不同甘草属物种之间的密码子 GC1、GC2、GC3、GC4 含量接近, 且每个物种均有 GC1>GC2>GC3, 表明 GC 含量在密码子不同位置的不均匀分布, 不同基因 GC 含量也存在差异 (图 1), GC 含量最低的 ycf1 基因其平均含量只有 0.295, 而 GC 含量最高的 rps11 基因其平均含量为 0.437, GC1、GC2、GC3 均小于 0.5, 叶绿体基因组密码子倾向于使用 A/U 碱基。每个甘草属物种的平均 ENC 值均大于 45, 表明 7 种甘草属植物叶绿体基因组的密码子偏好性相对较弱。

表 1 甘草属植物叶绿体基因组密码子偏好性基本参数

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

物种	登录号	GC1	GC2	GC3	GC4	ENC	N
粗毛甘草	MT120754	0.465	0.387	0.271	0.374	46.78	20 036
光果甘草	MT120770	0.465	0.388	0.27	0.374	46.75	20 031
胀果甘草	MT120775	0.465	0.387	0.27	0.374	46.72	20 043
刺果甘草	MZ052084	0.465	0.387	0.271	0.374	47.04	20 047
圆果甘草	MT120787	0.466	0.387	0.272	0.375	47.12	20 024
甘草	MN709826	0.465	0.387	0.271	0.374	46.75	20 029
云南甘草	MT120792	0.466	0.387	0.271	0.375	46.93	20 033

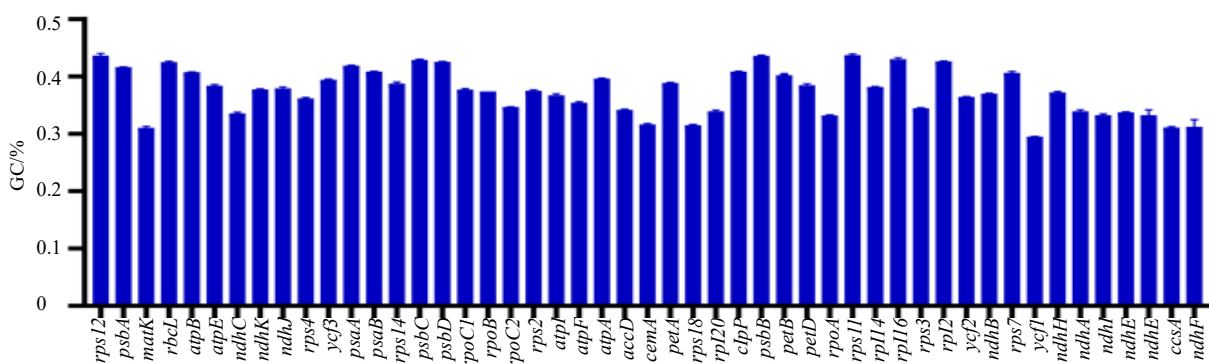


图 1 甘草属叶绿体蛋白编码基因 GC 含量

Fig. 1 GC content of chloroplast protein coding genes of *Glycyrrhiza* genome

对各参数进行相关性分析, 结果显示(表 2), GC1, GC2, GC12 与 GC3 相关性不显著, GC1 与 GC2 呈极显著正相关, 表明密码子第 1 位与第 2 位碱基组成相似, 而与密码子第 3 碱基组成有较大差异。GC1, GC2, GC12 与 ENC 相关性均未达显著水平,

而 GC3 与 ENC 呈显著相关, 表明 7 种甘草属植物密码子偏好性主要受密码子第 3 位碱基组成变化的影响, 第 1 位和第 2 位密码子碱基组成不是影响密码子偏好性的主要因素。此外, N 与 ENC 相关性不显著, 表明基因长度对密码子偏好性的影响较小。

表 2 甘草属物种密码子各参数相关性分析

Table 2 Correlation analysis of codon parameters in *Glycyrrhiza* species

物种名	项目	相关系数					
		GC1	GC2	GC12	GC3	GCall	ENC
粗毛甘草	GC2	0.576**					
	GC12	0.905**	0.870**				
	GC3	0.173	0.194	0.205			
	GCall	0.867**	0.841**	0.962**	0.465**		
	ENC	0.179	-0.028	0.093	0.339*	0.179	
	N	-0.177	-0.273	-0.249	0.188	-0.173	0.108
洋甘草	GC2	0.579**					
	GC12	0.905**	0.871**				
	GC3	0.191	0.194	0.216			
	GCall	0.870**	0.840**	0.962**	0.473**		
	ENC	0.174	-0.038	0.085	0.327*	0.167	
	N	-0.179	-0.272	-0.250	0.197	-0.171	0.112
胀果甘草	GC2	0.576**					
	GC12	0.905**	0.869**				
	GC3	0.188	0.207	0.222			
	GCall	0.867**	0.841**	0.962**	0.478**		
	ENC	0.189	-0.019	0.105	0.352*	0.192	
	N	-0.177	-0.273	-0.249	0.199	-0.169	0.111
刺果甘草	GC2	0.623**					
	GC12	0.915**	0.886**				
	GC3	0.168	0.224	0.215			
	GCall	0.872**	0.861**	0.962**	0.475**		
	ENC	0.063	-0.680	0.002	0.292*	0.084	
	N	-0.180	-0.279	-0.250	0.163	-0.180	0.056
圆果甘草	GC2	0.594**					
	GC12	0.910**	0.874**				
	GC3	0.218	0.230	0.250			
	GCall	0.873**	0.845**	0.963**	0.503**		
	ENC	0.205	0.097	0.174	0.290*	0.224	
	N	-0.183	-0.273	-0.251	0.151	-0.182	0.056
甘草	GC2	0.574**					
	GC12	0.904**	0.869**				
	GC3	0.191	0.187	0.213			
	GCall	0.869**	0.837**	0.962**	0.472**		
	ENC	0.190	-0.140	0.107	0.346*	0.194	
	N	-0.180	-0.269	-0.249	0.194	-0.171	0.115
云南甘草	GC2	0.585**					
	GC12	0.907**	0.872**				
	GC3	0.212	0.184	0.223			
	GCall	0.875**	0.835**	0.961**	0.483**		
	ENC	0.167	-0.057	0.072	0.315*	0.153	
	N	-0.188	-0.270	-0.253	0.175	-0.178	0.076

** $P < 0.01$ 表示极显著相关; * $P < 0.05$ 表示显著相关

** $P < 0.01$ indicates a very significant correlation; * $P < 0.05$ indicate a significant correlation

RSCU 是衡量密码子偏好性的重要参数，当 RSCU 值大于 1 时，表明该密码子受基因组偏爱，为高频密码子，反之则为低频密码子。在研究中发现这 7 种甘草属物种的 RSCU 值接近（图 2），RSCU 值大于 1 的密码子有 29 个，其中有一个密码子 UUA 的 RSCU 值大于 2，仅有一个密码子

UUG 是以 G 结尾。其它高频密码子均以 A/U 结尾，密码子第 3 位为 A 的有 13 个，密码子第 3 位为 U 的有 15 个。在 RSCU 小于 1 的 30 个低频密码子中，仅有 2 个密码子以 A/U 结尾，表明 7 种甘草属植物叶绿体基因组偏好使用 A/U 结尾的同义密码子。

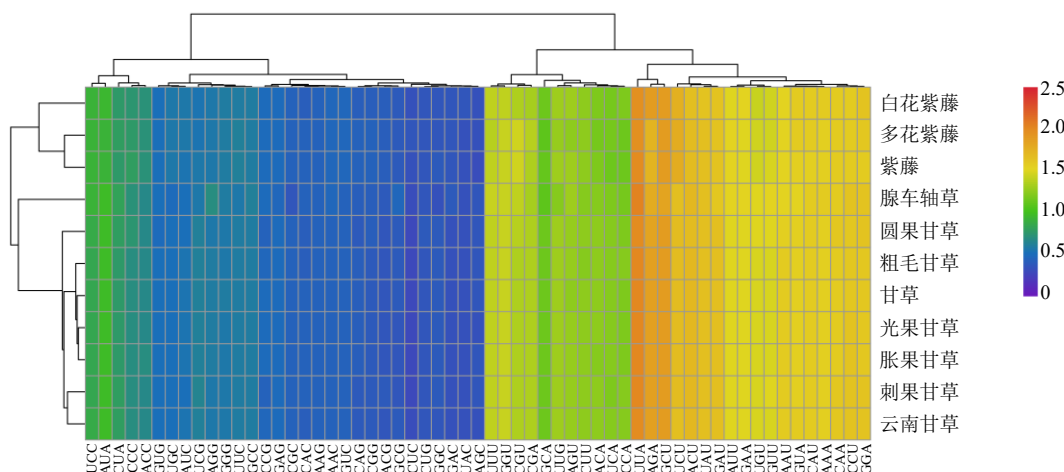


图 2 各物种叶绿体基因组 RSCU 值

Fig. 2 Chloroplast genome RSCU values of each specie

3.2 最优密码子

将 ENC 值按大小排序，选取 ENC 值最低的 *rps18*、*rpl16*、*atpI*、*ndhF*、*psbA* 作为高表达基因库，ENC 值最高的 *ycf3*、*rpl20*、*ycf2*、*clpP*、*rpl2*（圆果甘草为 *rps11*）作为低表达基因库，计算 2 个基因库每个密码子 RSCU 差值 (Δ RSCU)，满足 RSCU 大

于 1、 Δ RSCU 小于 0.08 的最优密码子有 17 个（表 3），其中有 14 个最优密码子为 7 个甘草属物种共有，其中有 5 个共有最优密码子的 Δ RSCU 大于 0.5，分别为 CCU、ACU、CUU、CGA、GCU，所有最优密码子均以 A/U 结尾，以 A 结尾的有 5 个，以 U 结尾的有 12 个。

表 3 甘草属植物叶绿体基因组最优密码子分析

Table 3 Optimal codons analysis of chloroplast genome in *Glycyrrhiza* species

氨基酸	密码子	Δ RSCU						
		粗毛甘草	光果甘草	胀果甘草	刺果甘草	圆果甘草	甘草	云南甘草
Phe	UUU	-0.452	-0.450	-0.450	-0.472	-0.254	-0.452	-0.462
	UUC	0.454**	0.452**	0.452**	0.474**	0.256*	0.454**	0.464**
Leu	UUA	0.508***	0.490**	0.600***	0.544***	0.222*	0.488**	0.570***
	UUG	-0.490	-0.472	-0.580	-0.602	-0.110	-0.474	-0.432
	CUU	0.654***	0.650***	0.650***	0.654***	0.760***	0.654***	0.670***
	CUC	-0.230	-0.230	-0.230	-0.188	-0.070	-0.230	-0.240
	CUA	0.012	0.014	0.014	0.026	-0.098	0.012	-0.108
	CUG	-0.462	-0.462	-0.462	-0.436	-0.708	-0.462	-0.468
Ser	UCU	0.408**	0.414**	0.416**	0.456**	0.526***	0.390**	0.416**
	UCC	0.390**	0.390**	0.392**	0.288*	0.210*	0.382**	0.332**
	UCA	-0.012	-0.106	-0.116	-0.130	-0.138	-0.120	-0.110
	UCG	-0.272	-0.282	-0.276	-0.166	-0.366	-0.280	-0.282
Tyr	UAU	0.310**	0.330**	0.330**	0.286*	0.138*	0.336**	0.290*
	UAC	-0.310	-0.330	-0.330	-0.286	-0.138	-0.336	-0.290

续表 3

氨基酸	密码子	Δ RSCU						
		粗毛甘草	光果甘草	胀果甘草	刺果甘草	圆果甘草	甘草	云南甘草
Cys	UGU	-0.394	-0.394	-0.394	-0.172	-0.506	-0.394	-0.406
	UGC	-0.006	-0.006	-0.006	-0.228	0.106*	-0.006	0.006
Pro	CCU	0.558***	0.568***	0.568***	0.568***	0.576***	0.600***	0.552***
	CCC	0.100*	0.088*	0.088*	0.094*	0.306**	0.066	0.100*
	CCA	0.164*	0.176*	0.176*	0.140*	-0.126	0.164*	0.168*
	CCG	-0.016	-0.026	-0.026	0.002	0.046	-0.026	-0.014
His	CAU	-0.308	-0.308	-0.308	-0.108	-0.394	-0.308	-0.318
	CAC	-0.092	-0.092	-0.092	-0.292	-0.006	-0.092	-0.082
Gin	CAA	0.402**	0.398**	0.400**	0.332**	0.282*	0.398**	0.390**
	CAG	-0.402	-0.398	-0.400	-0.332	-0.282	-0.398	-0.390
Arg	CGU	0.394**	0.394**	0.394**	0.398**	0.358**	0.394**	0.412**
	CGC	-0.142	-0.142	-0.142	0.052	-0.108	-0.142	-0.144
	CGA	1.016***	1.018***	1.018***	0.940***	0.978***	1.016***	0.958***
	CGG	-0.092	-0.092	-0.092	-0.208	-0.278	-0.092	-0.104
Lle	AUU	0.186*	0.186*	0.212*	0.206*	0.030	0.186*	0.184*
	AUC	-0.106	-0.106	-0.132	-0.182	0.078	-0.106	-0.106
	AUA	-0.080	-0.080	-0.080	-0.018	-0.106	-0.080	-0.078
Thr	ACU	0.752***	0.756***	0.756***	0.652***	0.718***	0.756***	0.756***
	ACC	-0.152	-0.160	-0.160	-0.070	-0.148	-0.160	-0.160
	ACA	-0.166	-0.164	-0.164	-0.154	-0.230	-0.164	-0.184
	ACG	-0.436	-0.434	-0.434	-0.428	-0.340	-0.434	-0.412
Asn	AAU	0.208*	0.204*	0.206*	0.222*	0.188*	0.206*	0.208*
	AAC	-0.208	-0.204	-0.206	-0.222	-0.188	-0.206	-0.208
Lys	AAA	-0.270	-0.302	-0.300	-0.292	-0.152	-0.302	-0.356
	AAG	-0.128	-0.096	-0.098	-0.108	-0.246	-0.096	-0.106
Ser	AGU	-0.620	-0.620	-0.618	-0.654	-0.510	-0.568	-0.622
	AGC	0.198*	0.198*	0.198*	0.198*	0.274*	0.194*	0.198*
Arg	AGA	-0.642	-0.638	-0.638	-0.780	-0.398	-0.642	-0.610
	AGG	-0.542	-0.550	-0.550	-0.410	-0.556	-0.542	-0.526
Val	GUU	0.658***	0.592***	0.592***	0.568***	0.418**	0.592***	0.594***
	GUC	0.424**	0.422**	0.422**	-0.580	0.364**	0.422**	0.418**
	GUA	-0.626	-0.618	-0.618	-0.468	-0.250	-0.618	-0.624
	GUG	-0.398	-0.400	-0.400	-0.520	-0.534	-0.400	-0.392
Ala	GCU	1.098***	1.110***	1.110***	1.120***	0.898***	1.110***	1.110***
	GCC	-0.826	-0.824	-0.824	-0.764	-0.598	-0.824	-0.824
	GCA	-0.224	-0.238	-0.238	-0.326	-0.272	-0.238	-0.216
	GCG	-0.046	-0.046	-0.046	-0.028	-0.028	-0.046	-0.066
Asp	GAU	0.248*	0.248*	0.246	-0.134	-0.332	0.246*	-0.154
	GAC	-0.248	-0.248	-0.246	0.134*	0.332**	-0.246	0.154*
Glu	GAA	0.456**	0.462**	0.460**	0.416**	0.314**	0.464**	0.260*
	GAG	-0.458	-0.464	-0.462	-0.416	-0.316	-0.466	-0.262
Gly	GGU	0.388**	0.388**	0.388**	0.396**	0.232*	0.468**	0.394**
	GGC	-0.266	-0.266	-0.266	-0.254	-0.208	-0.266	-0.268
	GGA	0.066	0.066	0.066	-0.138	-0.122	0.066	-0.060
	GGG	-0.076	-0.186	-0.186	-0.002	0.094*	-0.266	-0.068

表示 Δ RSCU>0.5; **表示 Δ RSCU>0.3; *表示 Δ RSCU>0.08 means Δ RSCU>0.5; ** means Δ RSCU>0.3; * means Δ RSCU>0.08

3.3 PR2-plot 分析

PR2-plot 分析结果显示(图 3), 7 种甘草属物种的叶绿体编码基因不均匀的分布于 4 个区域内, 即密码子第 3 位 4 种碱基未被均衡使用。大部分基因分布

于下方区域, 分布右下区域的基因数略多于左下区域, 说明密码子第 3 位碱基使用频率为 T 大于 A, G 大于 C, 7 种甘草属植物叶绿体密码子偏好性不仅仅受突变的影响, 还受自然选择等因素的影响。

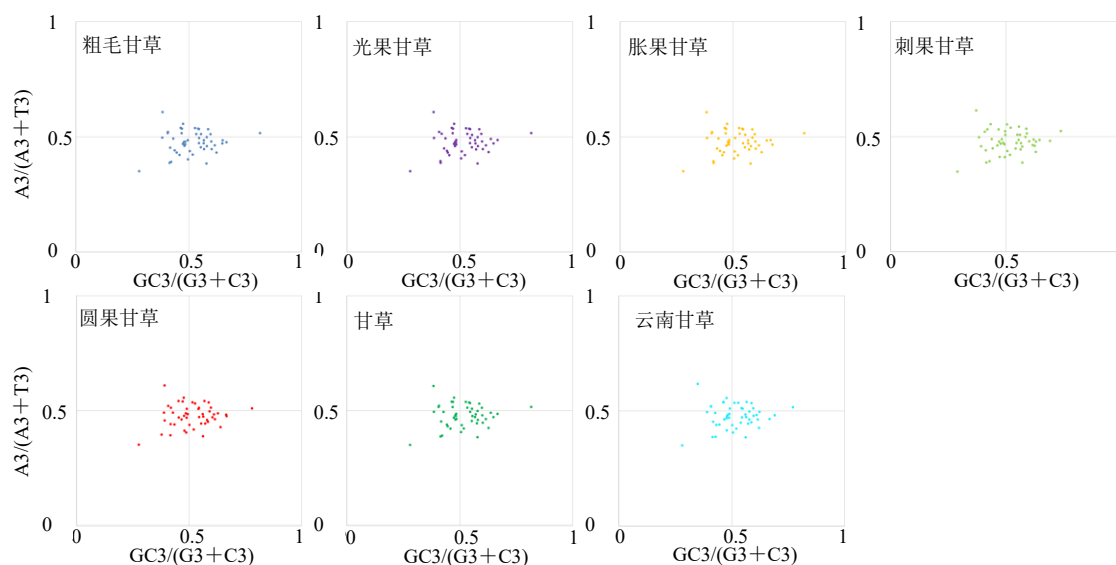


图 3 PR2-plot 分析

Fig. 3 Analysis of PR2-plot

3.4 ENC-plot 分析

如图 4 所示, 7 种甘草属物种每个基因的 ENC 均大于 35, 不同的基因受到自然选择和突变压力的影响程度不同, 大多数基因偏离标准曲线分布于曲线下, 少数基因接近或分布在标准曲线上, 图 5 结果显示, 大多数基因分布于 $-0.05 \sim 0.05$ 之外, 和分布于该区间以内的基因相比, 该部分基因实际 ENC 值与 ENC 期望值相差较大, 受突变的影响程

度较小。总体上, 7 种甘草属植物叶绿体密码子偏好性主要受自然选择的影响。

3.5 中性绘图分析

如图 6 所示, 7 种甘草属物种所有叶绿体编码基因均位于对角线以上的区域, 除 *cemA* 与 *ycf2* 基因接近对角线, 其余叶绿体编码基因分布于对角线以上较远的位置。圆果甘草的回归系数和相关系数最大, 其突变压力对密码子偏好性的贡献占 39.84%,

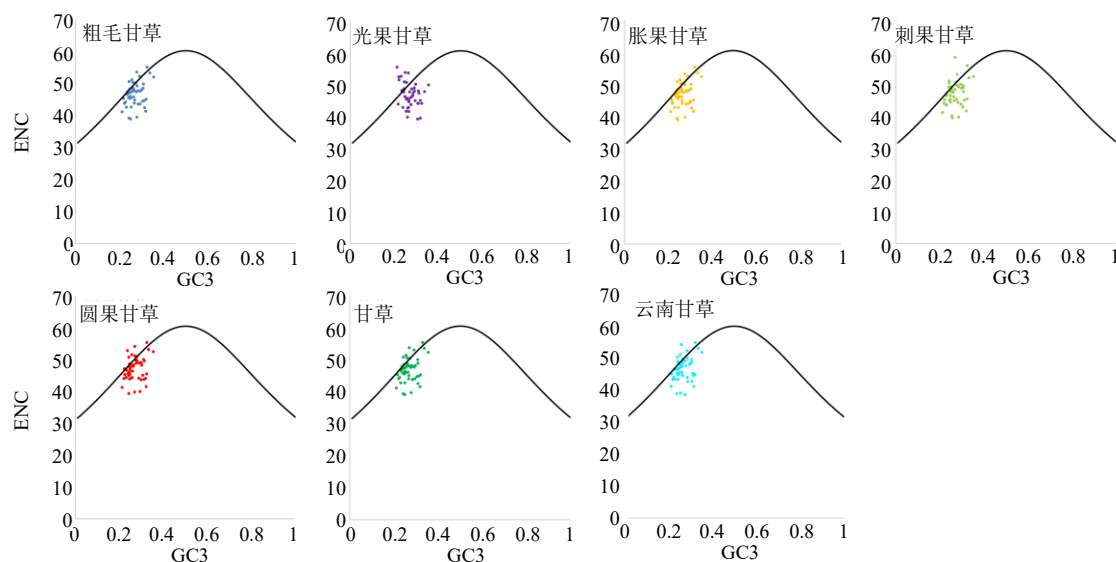


图 4 ENC-plot 分析

Fig. 4 Analysis of ENC-plot

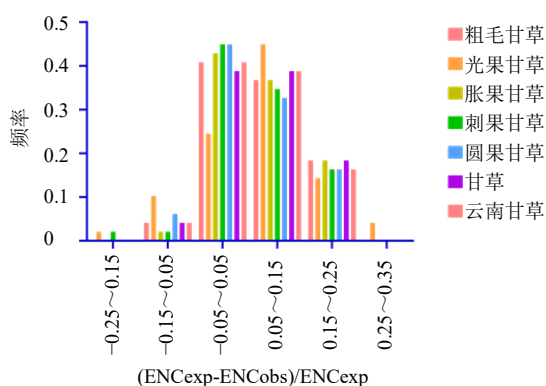


图 5 ENC 比值频率分布

Fig. 5 Frequency distribution of ENC ratio

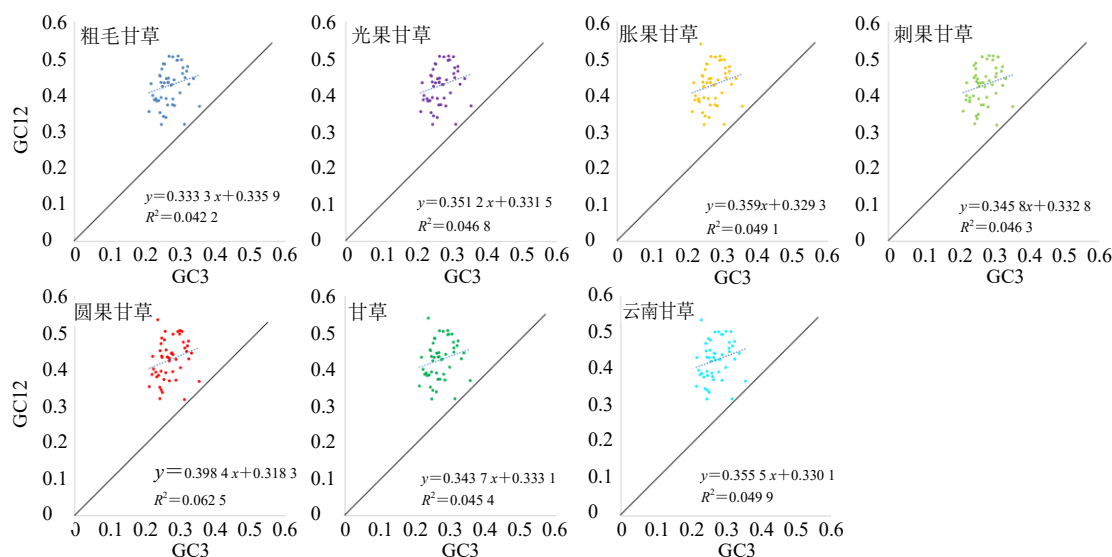


图 6 中性绘图分析

Fig. 6 Analysis of neutrality plot

个紫藤属物种、1 个车轴草属物种的 CDS 系统发育树具有相似的拓扑结构，此处将两树合并以 NJ 树展示（图 7），在基于叶绿体 CDS 的 NJ 系统发育树中，除粗毛甘草与甘草单独以 83% 支持率聚为一支外，其他分支支持率均为 100%，甘草属物种与紫藤属物种分别聚为 2 大枝；胀果甘草与光果甘草单独聚为一枝，粗毛甘草与甘草 2 个物种聚成一支，然后这两支再聚为一大支；刺果甘草与圆果甘草单独聚为一支后与云南甘草聚为一大支；多花紫藤与白花紫藤单独聚为一枝后与紫藤形成并支。ML 树与 NJ 树拓扑结构的差异在于粗毛甘草与甘草并未聚为一支，多花紫藤与白花紫藤以 84% 支持率聚为一支。RSCU 聚类分析结果为甘草属与紫藤属物种分别聚为 2 大类，白花紫藤与紫藤聚为一类后与多

是 7 个甘草属物种中密码子偏好性受突变压力影响最大和自然选择影响最小的甘草属物种。与之相反，粗毛甘草是 7 个甘草属物种中密码子偏好性受突变压力影响最小和自然选择影响最大的甘草属物种。7 种甘草属物种的回归系数分布范围为 0.333 3~0.398 4，GC3 与 GC12 相关性不显著，表明 GC3 与 GC12 进化模式存在较大差异，自然选择和突变压力均为影响 7 种甘草属植物叶绿体密码子偏好性的重要因素，但自然选择在 7 种甘草属植物叶绿体密码子偏好性形成中的贡献更大。

3.6 系统发育分析

基于 NJ 法和 ML 法构建的 7 个甘草属物种、3

花紫藤构成紫藤属分类群。在甘草属分类群中，圆果甘草单独作为一类，光果甘草、胀果甘草、甘草、粗毛甘草聚为一类后与云南甘草和刺果甘草聚为一大类。甘草属各物种 RSCU 聚类分析与叶绿体 CDS 的系统发育树拓扑结果具有一定的相似性。

4 讨论

研究表明，7 种甘草属物种叶绿体编码基因的密码子不同位置的 GC 含量为 $GC3 < GC2 < GC1$ ，密码子总 GC 含量和密码子不同位置的 GC 含量均小于 AU 含量，RSCU 分析结果也显示高频密码子多为 A/U 结尾的密码子，表明甘草属叶绿体编码基因倾向使用 A/U 结尾的密码子，在菊科^[21]、大戟科^[13]、草蓼属^[22]、酸枣^[23]等双子叶植物和兰科^[24]、石斛属^[25]、豆蔻属^[14]、稻属^[16]等单子叶植物的叶绿

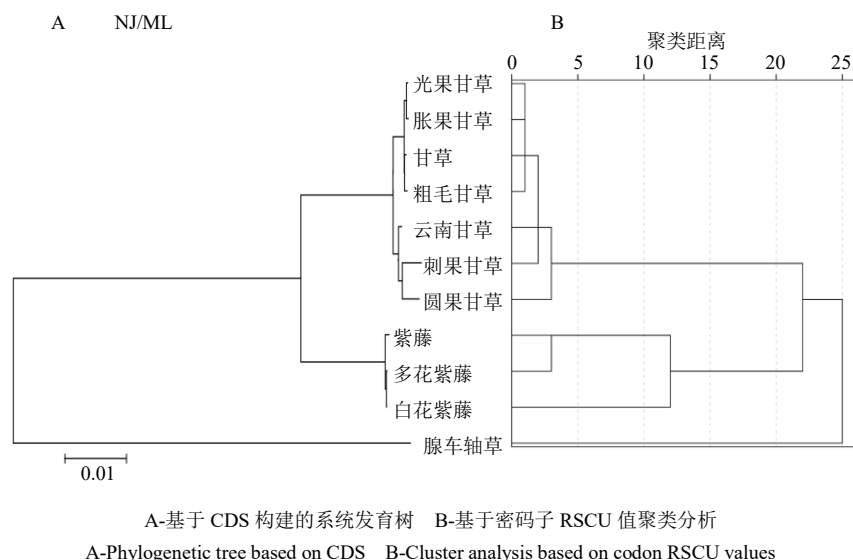


图 7 甘草属物种系统发育分析

Fig. 7 Phylogenetic analysis of Glycyrrhiza species

体基因组密码子偏好性研究中也观察到这一现象。密码子第 3 位碱基发生改变一般不影响其所编码的氨基酸种类，因此密码子第 3 位碱基在进化过程中受到的选择压力较小，往往被视为密码子偏好性研究的重要指标^[26-27]。相较于高等植物叶绿体基因组密码子多表现为密码子第 3 位碱基为 A/U 的倾向^[28]，核基因组在密码子第 3 位碱基使用的倾向上存在差异，双子叶植物核基因组倾向于使用 A/U 结尾的密码子，而单子叶植物核基因组倾向于 G/C 结尾的密码子^[29]，这一现象的出现可能是由于叶绿体基因组和核基因组分属不同的遗传系统，其进化模式存在差异，叶绿体基因组的保守性更强。

每个甘草属物种叶绿体基因组有 29 个高频密码子，但仅有编码亮氨酸的 UUA 密码子 RSCU 值大于 2，每个甘草属物种叶绿体编码基因的 ENC 值均大于 35，平均 ENC 值均大于 45，表明 7 种甘草属植物叶绿体基因组的密码子偏好性相对较弱。研究得到的 29 个高频密码子和 17 个最优密码子均表现为强烈的密码子第 3 位碱基使用 A/U 的倾向，但在高频密码子和最优密码子中均观察到第 3 位碱基使用 U 的密码子数量大于使用 A 的密码子数量，说明相较于使用以 A 结尾的密码子，7 种甘草属植物叶绿体基因组密码子更倾向于使用以 U 结尾的密码子，后续的 PR2-plot 分析也证明了这一点，在草莓属^[22]、绞股蓝属^[17]、杠果^[30]等植物中也表现为密码子第 3 位碱基倾向于使用 U 的倾向。

生物体在进化过程中密码子偏好性的形成受多

种因素的影响，在本研究中，甘草属各物种叶绿体编码基因的密码子数与 ENC 相关性不显著，表明编码序列长度对 7 种甘草属植物叶绿体基因组的密码子偏好性影响较弱。生物界普遍认为自然选择和突变压力是密码子偏好性形成过程中的主要影响因素，二者对密码子偏好性形成的贡献程度是目前的研究热点。PR2-plot 分析中密码子第三位碱基的不均衡使用表明突变并非是 7 种甘草属物种叶绿体基因组密码子偏好性形成的唯一驱动力，ENC-plot 分析中大多数甘草属编码基因偏离标准曲线和中性绘图分析中回归系数均小于 0.5 且相关性不显著，表明自然选择和突变压力均为影响 7 种甘草属植物叶绿体基因组密码子偏好性的重要因素，但二者对不同甘草属物种不同基因的密码子偏好性形成的作用强度存在差异，总体上，自然选择在 7 种甘草属植物叶绿体基因密码子偏好性的形成中起主导作用，在绞股蓝属^[17]、草莓属^[22]、豆蔻属^[14]、降香黄檀^[8]等植物的研究中也得到类似结果。

进化关系较近的物种具有相似的密码子偏好性，有研究认为，密码子使用模式可用于表征物种间基因组的进化距离^[31]。在本研究中，基于 RSCU 构建的聚类分析谱系图与基于叶绿体编码序列构建的系统进化树图均可将同为蝶形花亚科的三个属区分开，二者的拓扑结构有部分相似，表明甘草属叶绿体基因组密码子偏好性与物种的系统发育关系存在一定的联系。属内聚类结果的差异可能是由于基于 RSCU 聚类分析缺乏了部分遗传信息，如无偏好

密码子和位点突变特征等^[18]。

本研究首次探讨了7种甘草属植物的叶绿体基因组密码子使用模式及影响因素,有利于从分子水平研究甘草属植物的进化机制,同时也为后续通过密码子优化提高特定基因在甘草叶绿体基因组的表达以培育具有优良性状的栽培甘草新种提供了参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志 [M]. 北京: 科学出版社, 1998: 56.
- [2] Duan L, Harris A J, Su C, *et al.* Chloroplast phylogenomics reveals the intercontinental biogeographic history of the liquorice genus (Leguminosae: *Glycyrrhiza*) [J]. *Front Plant Sci*, 2020, 11: 793.
- [3] 中国药典 [S]. 一部. 2020: 26.
- [4] 李想, 李冀. 甘草提取物活性成分药理作用研究进展 [J]. *江苏中医药*, 2019, 51(5): 81-86.
- [5] 孙娟, 庄亦心, 胡锋清, 等. 甘草酸作为甜味剂的研究进展 [J]. *中国食品添加剂*, 2020, 31(11): 144-148.
- [6] Hanqing, Wang, . Identification wild and cultivated licorice by multidimensional analysis [J]. *Food Chem*, 2021, 339: 128111.
- [7] 魏胜利, 王文全, 王继永, 等. 我国不同产区野生与栽培甘草的甘草酸含量及其影响因素的初步研究 [J]. *中国中药杂志*, 2012, 37(10): 1341-1345.
- [8] 原晓龙, 李云琴, 张劲峰, 等. 降香黄檀叶绿体基因组密码子偏好性分析 [J]. *广西植物*, 2021, 41(4): 622-630.
- [9] Qi Y Y, Xu W J, Xing T, *et al.* Synonymous Codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity [J]. *Evol Bioinform Online*, 2015, 11: EBO.S22566.
- [10] 寇莹莹, 宋英今, 杨少辉, 等. 植酸酶phyA基因的密码子优化及其在大豆中的表达 [J]. *作物学报*, 2016, 42(12): 1798-1804.
- [11] Li Y T, Dong Y, Liu Y C, *et al.* Comparative analyses of *Euonymus* chloroplast genomes: Genetic structure, screening for loci with suitable polymorphism, positive selection genes, and phylogenetic relationships within Celastraceae [J]. *Front Plant Sci*, 2021, 11: 593984.
- [12] Huang R, Xie X N, Chen A M, *et al.* The chloroplast genomes of four *Bupleurum* (Apiaceae) species endemic to Southwestern China, a diversity center of the genus, as well as their evolutionary implications and phylogenetic inferences [J]. *BMC Genomics*, 2021, 22(1): 714.
- [13] Wang Z J, Xu B B, Li B, *et al.* Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species [J]. *PeerJ*, 2020, 8: e8251.
- [14] 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好性分析 [J]. *中草药*, 2021, 52(12): 3661-3670.
- [15] Sheng J J, She X, Liu X Y, *et al.* Comparative analysis of codon usage patterns in chloroplast genomes of five *Miscanthus* species and related species [J]. *Peer J*, 2021, 9: e12173.
- [16] Chakraborty S, Yengkhom S, Uddin A. Analysis of Codon usage bias of chloroplast genes in *Oryza* species: Codon usage of chloroplast genes in *Oryza* species [J]. *Planta*, 2020, 252(4): 67.
- [17] 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.
- [18] 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.
- [19] 李若愚, 张小丹, 马昕怡, 等. 桃基因密码子使用模式及其偏好性分析 [J]. *分子植物育种*, 2021, 19(3): 799-807.
- [20] 赵森, 邓力华, 陈芬. 秋茄叶绿体基因组密码子使用偏好性分析 [J]. *森林与环境学报*, 2020, 40(5): 534-541.
- [21] Nie X J, Deng P C, Feng K W, *et al.* Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family [J]. *Plant Mol Biol Rep*, 2014, 32(4): 828-840.
- [22] 刘华波, 熊融, 倪杨, 等. 草莓属植物叶绿体基因组密码子偏好性模式的比较分析 [J/OL]. *分子植物育种*, 2021, <https://kns.cnki.net/kcms/detail/46.1068.S.20211118.2344.010.html>.
- [23] 胡晓艳, 许艳秋, 韩有志, 等. 酸枣叶绿体基因组密码子使用偏好性分析 [J]. *森林与环境学报*, 2019, 39(6): 621-628.
- [24] 季凯凯. 兰科植物基因组密码子使用特征分析及其系统进化关系 [D]. 海口: 海南大学, 2020.
- [25] 王晓双, 杨芳, 罗茜, 等. 石斛属叶绿体基因组密码子使用偏好性及系统发育分析 [J/OL]. *分子植物育种*, 2021 [2021-03-03]. <https://kns.cnki.net/kcms/detail/46.1068.S.20210303.1340.012.html>.
- [26] 丁锐, 胡兵, 宗小雁, 等. 构兰叶绿体基因组密码子偏好性分析 [J]. *林业科学研究*, 2021, 34(5): 177-185.
- [27] Shen Z N, Gan Z M, Zhang F, *et al.* Analysis of codon usage patterns in citrus based on coding sequence data [J]. *BMC Genomics*, 2020, 21(Suppl 5): 234.
- [28] Sablok G, Nayak K C, Vazquez F, *et al.* Synonymous Codon usage, GC3, and evolutionary patterns across plastomes of three pooid model species: Emerging grass genome models for monocots [J]. *Mol Biotechnol*, 2011, 49(2): 116-128.
- [29] Zhang Y R, Nie X J, Jia X O, *et al.* Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family [J]. *Aust J Bot*, 2012, 60(5): 461.
- [30] 辛雅萱, 黎若竹, 李鑫, 等. 芒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.
- [31] Li G, Zhang L, Xue P. Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species [J]. *Gene*, 2021, 802: 145866.

[责任编辑 时圣明]