

枸杞叶绿体基因组密码子偏好性分析

杜雨¹, 李效雄², 贾西贝¹, 胡晓桐¹, 刘筠¹, 马彦军^{1*}

1. 甘肃农业大学, 甘肃 兰州 730070

2. 兰州资源环境职业技术大学, 甘肃 兰州 730030

摘要: 目的 探究枸杞密码子的使用模式及影响因素。方法 基于中国枸杞 *Lycium chinense*、黄果枸杞 *L. barbarum* 和白果枸杞 *L. ruthenicum* 的叶绿体基因组, 利用 CodonW 1.4.2 软件、Python 软件和 Excel 软件等分析枸杞密码子使用偏性。结果 3 个枸杞叶绿体基因组具有相似的密码子使用模式, 密码子第 3 位碱基平均 GC 含量为 25.68%~25.77%; 3 个枸杞的 ENC 值均在 35 以上, 表明枸杞叶绿体基因组密码子偏性较弱。ENC-plot、PR2-plot 和中性分析结果显示枸杞叶绿体基因组密码子偏性主要受自然选择影响。基于相对同义密码子分析, 枸杞植物中共有最优密码子 13 个。结论 枸杞叶绿体基因组密码子第 3 位碱基主要以 A/T 结尾, 且密码子偏性主要受自然选择的影响。通过对枸杞植物的聚类分析, 揭示了枸杞植物的遗传关系以及枸杞植物密码子使用模式的影响因素, 为后续探究枸杞植物的系统发育提供参考。

关键词: 枸杞; 中国枸杞; 黄果枸杞; 白果枸杞; 叶绿体基因组; 密码子偏好性; 聚类分析

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)04-1316-10

DOI: 10.7501/j.issn.0253-2670.2024.04.024

Codon preference analysis of chloroplast genome of Chinese wolfberry

DU Yu¹, LI Xiaoxiong², JIA Xibei¹, HU Xiaotong¹, LIU Yun¹, MA Yanjun¹

1. Gansu Agricultural University, Lanzhou 730070, China

2. Lanzhou Resources & Environment Voc-tech University, Lanzhou 730030, China

Abstract: Objective To explore the use pattern and influencing factors of codons in Chinese wolfberry. **Methods** Based on chloroplast genomes of *L. chinese*, *L. barbarum* and *L. ruthenicum*, CodonW 1.4.2 software, Python software and Excel software were used to analyze the codon usage bias of Chinese wolfberry. **Results** The three Chinese wolfberry chloroplast genomes had similar codon usage patterns, and the average GC content of the third base of codon was 25.68%—25.77%. The ENC values were all above 35, indicating that the codon bias of Chinese wolfberry chloroplast genomes was weak. The results of ENC-plot, PR2-plot and neutral analysis showed that the codon bias of Chinese wolfberry chloroplast genomes was mainly affected by natural selection. Based on the analysis of relative synonymous codons, there were 13 optimal codons in Chinese wolfberry plants. **Conclusion** The third base of codons of the *L. chinese* chloroplast genomes mainly ends with A/T, and the codon bias is mainly affected by natural selection. Through the cluster analysis of *L. chinese* plants, the genetic relationship and the influencing factors of codon usage patterns of Chinese wolfberry plants were revealed, which provided a reference for the subsequent exploration of the phylogenetic development of Chinese wolfberry plants.

Key words: Chinese wolfberry; *Lycium chinense* Mill.; *Lycium barbarum* Linn. var. *auranticarpum* K. F. Ching; *Lycium ruthenicum* Murray; chloroplast genome; codon bias; cluster analysis

叶绿体是一种存在于高等植物和藻类中含有光合色素的质体, 同时叶绿体基因组也为植物系统发育研究提供了重要作用^[1-2]。与核基因组相比, 叶绿体基因组具有序列高度保守和进化速率适中等优势, 可用于研究物种的起源和进化等方向^[3-7]。目前, 因为高通量测序具有比较高的准确性, 因此可以较

为准确的研究生物体的系统发育和进化等^[8-10]。通常来说, 发育关系越密切的物种, 其密码子使用模式越相近^[11]。如白羊草 *Bothriochloa ischaemum* L.^[12]、刺柏属 *Juniperus* L.^[13]、烟草 *Nicotiana tabacum* L.^[14] 及其近缘种利用叶绿体基因组同义密码子相对使用度 (relative synonymous codon usage, RSCU) 构建

收稿日期: 2023-10-02

基金项目: 兰州资源环境职业技术大学重点研究项目 (X2022ZD-05); 兰州市科技计划项目资助 (2022-2-22)

作者简介: 杜雨 (1994—), 女, 汉族, 河南周口人, 硕士研究生, 主要从事植物种质资源调查收集、保存与研究。

E-mail: Duyu0126@126.com

*通信作者: 马彦军, 教授, 研究方向为植物种质资源调查收集、保存与研究。E-mail: lxsys01@126.com

聚类分析图, 分析其与近缘种的系统发育关系。

mRNA 上相邻的 3 个碱基构成了密码子, 且密码子是连接核酸与蛋白质的纽带, 在生物遗传信息传递时具有重要作用^[12]。密码子的简并性是指在编码过程中一种氨基酸可以由 2 个或 2 个以上的密码子编码, 而这些密码子则被称为同义密码子。在自然界中存在的氨基酸有 20 种, 其中甲硫氨酸和色氨酸由一个密码子编码外, 剩余氨基酸则由 2~6 个同义密码子编码^[15]。由于同义密码子的使用频率有所差异, 形成了密码子使用的偏好性 (codon usage bias, CUB)。而密码子使用的偏好性在不同物种或同一物种中也有所差异^[16], 如杜仲 *Eucommia ulmoides* Oliv. 密码子主要以 G/C 结尾^[17], 唐古特白刺 *Nitraria tangutorum* Bobr. 密码子则主要使用 A/U 结尾^[18]等。影响 CUB 的因素主要有突变和自然选择等。目前对密码子偏好性的研究主要集中在分子进化、翻译调控等方面^[19-20]。研究表明, 密码子偏好性可以通过改变蛋白质结构、核苷酸序列等方式影响外源基因的表达^[21-22]。因此, 研究密码子的偏好性不仅可帮助了解物种的分子进化、环境适应和基因组特征, 还可以用来判断未知基因的表达以及预测一些未知功能基因^[23-24]。

枸杞属 *Lycium* L. 植物为多年生灌木, 隶属于茄科 (Solanaceae)。枸杞属植物约有 80 种, 以离散型方式随机分布在各地, 其中南美洲和北美洲分布较广。我国枸杞属植物分为 7 个种和 3 个变种^[25]。在野外调查时发现枸杞果实颜色较为丰富, 主要有红果、黄果、黑果和白果等。其中, 中国枸杞具有长椭圆形或卵状披针形的叶片, 并且果实呈红橘黄色, 其直径在 5~10 mm, 叶片单叶互生或 2~4 枚簇生。中国枸杞的色素主要是类胡萝卜素。黄果枸杞的叶片较为狭窄, 呈条形或狭披针形, 果实呈橙黄色, 其直径为 4~8 mm, 是宁夏枸杞的 1 个变种, 被称为“黄金枸杞”, 种植面积较小。白果枸杞的条形或条状披针形的叶片在短枝上 2~6 枚簇生, 果实为球状浆果, 呈白色或表皮略带紫色斑点, 其直径在 4~9 mm。枸杞的花期均为 5~10 月^[26]。目前, 在野外调查时发现存在中国枸杞、黑果枸杞、黄果枸杞和白果枸杞等, 其中中国枸杞、黑果枸杞和黄果枸杞的分类学地位已经清晰, 并且也应用于研究中, 而对白果枸杞的分类地位尚不清楚。因此, 本研究以中国枸杞、黄果枸杞和白果枸杞的叶绿体基因组为研究对象, 分析枸杞的密码子使用的特征及其影

响因素, 阐明枸杞叶绿体基因组密码子使用模式, 为枸杞属植物的系统发育关系提供理论依据。

1 材料与仪器

1.1 材料

进行枸杞种质资源调查时在青海采集中国枸杞、黄果枸杞和白果枸杞的植株并栽植到甘肃农业大学校内实习基地 (地理坐标: 38°28'N, 106°16'E)。经甘肃农业大学马彦军教授鉴定为茄科中国枸杞 *Lycium chinense* Mill.、黄果枸杞 *L. barbarum* Linn. var. *auranticarpum* K. F. Ching. 和白果枸杞 *L. ruthenicum* Murray。

1.2 仪器

S220 型超声波 DNA 破碎仪 (美国 Covaris 公司), Thermo Qubit 4.0 荧光定量仪 (Q33226), 台式高速低温离心机 (Thermo Scientific Sorvall LegendMicro 21R), DE13805054 型 Agilent 2100 Bioanalyzer (美国 Agilent 公司)。

2 方法

2.1 样品处理

通过 Illumina 高通量测序 (上海生工生物公司测) 获得高质量数据 (clean data), 然后进行叶绿体全基因组序列拼接、组装与注释, 最终汇总整理为完整的注释结果。注释完成后, 提交到 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/genbank/>)。其中国枸杞、黄果枸杞和白果枸杞的叶绿体基因组登录号为 OP866962、OP846044 和 OP846050。中国枸杞、黄果枸杞和白果枸杞的叶绿体基因组大小分别为 155 655、154 978、154 976 bp, 其中中国枸杞和黄果枸杞含有 83 个蛋白编码基因 (protein-coding genes, CDS), 白果枸杞含有 81 个蛋白编码基因 (protein-coding genes, CDS)。根据注释信息剔除长度小于 300 bp 和重复的蛋白质编码序列, 筛选起始密码子为 ATG, 且终止密码子为 TAA、TGA 或 TAG 的序列。

2.2 数据分析

2.2.1 密码子组成分析 利用 CodonW 1.4.2 软件 (<http://codonw.sourceforge.net>) 对中国枸杞、黄果枸杞和白果枸杞的有效密码子数 (effective number of codons, ENC)、同义密码子相对使用 (relative synonymous codon usage, RSCU) 以及密码子 GC 值进行分析, 并利用 SPSS24.0 和 Chiplot (<https://www.chiplot.online/#Bubble-plot>) 进行显著性分析。其中 GC 值包括第 1、2、3 位碱基以及三碱基中的平均含量。

2.2.2 ENC-plot 绘图分析 ENC 值是了解密码子

偏好性的重要参数,可以判断内源基因的表达量,ENC 越小,密码子偏好性强,为高表达基因。使用 Python 软件绘制以 ENC 值为纵坐标,GC3 含量(密码子第 3 位碱基的 GC 含量)为横坐标的曲线图(终止密码子、甲硫氨酸和色氨酸除外)。同时,根据 ENC 的比值频数可以纠正 ENC-plot 分析可能存在的错误,按照以下公式计算 ENC。

$$ENC=2+GC3+29/[(GC3)_2+(1-GC3)_2]$$

2.2.3 PR2-plot 和中性绘图分析 以 $G3/(G3+C3)$ 为 X 轴, $A3/(A3+T3)$ 为 Y 轴绘制 PR2-plot 散点图。其中 PR2-plot 图的中心点位置表明碱基含量为 $A=T$ 、 $G=C$,即密码子不具有偏好性,其余点与中心点的矢量距离则代表其偏倚程度和方向。

以第 1、2 位碱基上的 GC 平均含量为纵坐标,GC3 为横坐标绘制中性分析(neutrality plots)散点图。通过直线拟合分析图进行分析。利用二者的相关性,推断影响密码子偏好性的因素。当密码子受突变影响时,直线斜率接近 1,二者呈显著相关;反之,则受自然选择影响,无显著相关,GC 含量高度保守。

2.2.4 最优密码子的确定及聚类分析 利用 ENC 值和 CodonW 软件构建高、低表达基因库和同义密码子使用度(relative synonymous codon usage, RSCU)值和 $\Delta RSCU$ 值($\Delta RSCU=RSCU$ 高表达— $RSCU$ 低表达)的计算。最优密码子的确定则是需要同时满足高表达密码子和高频密码子的条件。

利用 NCBI 数据库下载茄属秘鲁茄 *Solanum peruvianum* L.、智利番茄 *S. chilense* (Dunal) Reiche.、多毛番茄 *S. habrochaites* S. Knapp & D. M. Spooner、契斯曼尼番茄 *S. cheesmanii*、栽培番茄 *S. lycopersicum*

L.、黑果枸杞 *Lycium ruthenicum* Murray. 的叶绿体基因组序列并计算密码子的 RSCU 值,通过聚类分析验证 3 种枸杞叶绿体基因组密码子偏好性。

3 结果与分析

3.1 密码子偏性分析

分析结果(表 1)显示,3 种枸杞植物的 RSCU 比较接近,其中有 30 个密码子的 RSCU 值大于 1,且以 U、A 和 G 结尾的密码子分别有 16、13 和 1;而 3 种植物 RSCU 值小于 1 的密码子中只有 3 个密码子以 A 结尾,表明 3 种枸杞植物的密码子中主要以 A/U 结尾。中国枸杞(OP866962)、黄果枸杞(OP846044)和白果枸杞(OP846050)的密码子 GC 含量、ENC 值分析结果显示(表 2、3):在 3 种枸杞植物的 GC1 平均含量均为 49.80%,GC2 平均含量为 39.83%~39.85%,GC3 平均含量为 25.68~25.77%,GC 总含量(GC_{all_s})平均值为 38.44~38.47%,ENC 平均值为 48.05~48.06,表明 3 种枸杞植物的 GC 值和 ENC 无明显差异。同时 3 种枸杞植物密码子的 GC 含量均表现为 $GC1>GC2>GC_{all_s}>GC3$,且 GC1、GC2、GC3 平均含量都小于 50%,表明 3 种枸杞不同位置的密码子 GC 含量有所差异且更偏好使用以 A/U 结尾的密码子。

枸杞的相关分析表明 GC_{all_s} 和 GC1、GC2、GC3 具有极显著($P<0.01$)相关性(0.361~0.848),与 ENC 不相关;GC3 与 GC1、GC2 不具有相关性,表明密码子的第 3 位为碱基组成与第 1 位和第 2 位有所差异,而 GC3 与 GC_{all_s} (0.361~0.362)、ENC (0.431~0.438)有显著($P<0.05$)相关性,表明枸杞密码子的使用主要是受 GC3 含量变化的影响(图 1)。

表 1 枸杞叶绿体基因组中各氨基酸的 RSCU 分析

Table 1 RSCU analysis of amino acids in chloroplast genomes of *L. chinense*

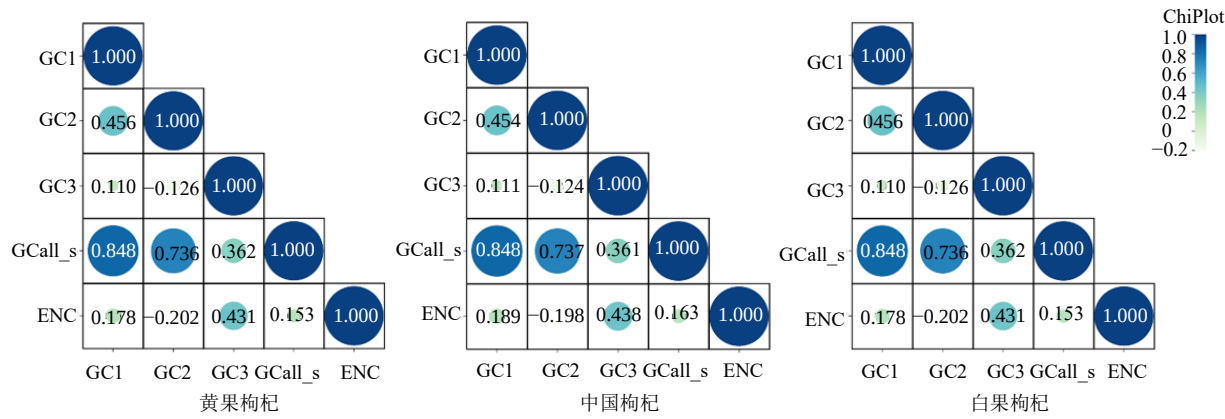
氨基酸	中国枸杞				白果枸杞				黄果枸杞				tRNA
	密码子	数量	同义密码使用度		密码子	数量	同义密码使用度		密码子	数量	同义密码使用度		
Phe	UUU	737	1.31		UUU	740	1.32		UUU	740	1.32		trnF-GAA
	UUC	384	0.69		UUC	384	0.68		UUC	384	0.68		
Ser	UCU	442	1.73		UCU	440	1.72		UCU	440	1.72		trnS-GGA
	UCC	243	0.95		UCC	244	0.96		UCC	244	0.96		
	UCA	316	1.23		UCA	312	1.22		UCA	312	1.22		
	UCG	144	0.56		UCG	144	0.56		UCG	144	0.56		
	AGU	300	1.17		AGU	302	1.18		AGU	302	1.18		
Tyr	AGC	91	0.36		AGC	89	0.35		AGC	89	0.35		trnS-GCU
	UAU	597	1.62		UAU	597	1.62		UAU	597	1.62		
	UAC	141	0.38		UAC	141	0.38		UAC	141	0.38		
Cys	UGU	158	1.44		UGU	157	1.44		UGU	157	1.44		trnC-GCA
	UGC	61	0.56		UGC	61	0.56		UGC	61	0.56		
Leu	UUA	681	1.90		UUA	681	1.90		UUA	681	1.90		trnL-UAA
	UUG	443	1.23		UUG	443	1.23		UUG	443	1.23		

表 1 (续)

氨基酸	中国枸杞			白果枸杞			黄果枸杞			tRNA
	密码子	数量	同义密码使	密码子	数量	同义密码使用度	密码子	数量	同义密码使用度	
Leu	CUU	462	1.29	CUU	462	1.29	CUU	462	1.29	tmL-UAG
	CUC	147	0.41	CUC	145	0.40	CUC	145	0.40	
	CUA	278	0.77	CUA	281	0.78	CUA	281	0.78	
	CUG	144	0.40	CUG	143	0.40	CUG	143	0.40	
Trp	UGG	380	1.00	UGG	380	1.00	UGG	380	1.00	tmW-CCA
Pro	CCU	347	1.61	CCU	346	1.61	CCU	346	1.61	tmP-UGG
	CCC	156	0.72	CCC	155	0.72	CCC	155	0.72	
	CCA	247	1.14	CCA	245	1.14	CCA	245	1.14	
	CCG	114	0.53	CCG	116	0.54	CCG	116	0.54	
His	CAU	372	1.54	CAU	370	1.54	CAU	370	1.54	tmH-GUG
	CAC	111	0.46	CAC	111	0.46	CAC	111	0.46	
Arg	CGU	277	1.34	CGU	280	1.36	CGU	280	1.36	tmR-ACG
	CGC	80	0.39	CGC	79	0.38	CGC	79	0.38	
	CGA	307	1.49	CGA	307	1.49	CGA	307	1.49	
	CGG	88	0.43	CGG	87	0.42	CGG	87	0.42	
	AGA	355	1.72	AGA	356	1.73	AGA	356	1.73	
	AGG	129	0.63	AGG	129	0.63	AGG	129	0.63	
Gln	CAA	566	1.50	CAA	568	1.50	CAA	568	1.50	tmQ-UUG
	CAG	188	0.50	CAG	188	0.50	CAG	188	0.50	
Ile	AUU	843	1.47	AUU	843	1.48	AUU	843	1.48	tmI-GAU
	AUC	343	0.60	AUC	342	0.60	AUC	342	0.60	
	AUA	529	0.93	AUA	526	0.92	AUA	526	0.92	
Thr	ACU	406	1.57	ACU	403	1.56	ACU	403	1.56	tmT-GGU
	ACC	213	0.82	ACC	212	0.82	ACC	212	0.82	
	ACA	310	1.20	ACA	311	1.21	ACA	311	1.21	
	ACG	104	0.40	ACG	105	0.41	ACG	105	0.41	
Asn	AAU	756	1.53	AAU	759	1.53	AAU	759	1.53	tmN-GUU
	AAC	234	0.47	AAC	233	0.47	AAC	233	0.47	
Lys	AAA	808	1.53	AAA	810	1.54	AAA	810	1.54	tmM-CAU
	AAG	247	0.47	AAG	245	0.46	AAG	245	0.46	
Met	AUG	454	1.00	AUG	455	1.00	AUG	455	1.00	tmV-GAC
Val	GUU	423	1.49	GUU	422	1.48	GUU	422	1.48	
	GUC	131	0.46	GUC	133	0.47	GUC	133	0.47	
	GUA	432	1.52	GUA	432	1.52	GUA	432	1.52	
	GUG	153	0.54	GUG	152	0.53	GUG	152	0.53	
Ala	GCU	506	1.78	GCU	507	1.78	GCU	507	1.78	tmA-UGC
	GCC	198	0.70	GCC	199	0.70	GCC	199	0.70	
	GCA	317	1.11	GCA	321	1.12	GCA	321	1.12	
	GCG	118	0.41	GCG	115	0.40	GCG	115	0.40	
Asp	GAU	657	1.59	GAU	655	1.59	GAU	655	1.59	tmD-GUC
	GAC	168	0.41	GAC	171	0.41	GAC	171	0.41	
Gly	GGU	461	1.29	GGU	461	1.29	GGU	461	1.29	tmG-GCC
	GGC	168	0.47	GGC	167	0.47	GGC	167	0.47	
	GGA	551	1.54	GGA	550	1.54	GGA	550	1.54	
	GGG	252	0.70	GGG	253	0.71	GGG	253	0.71	
Glu	GAA	833	1.50	GAA	834	1.51	GAA	834	1.51	tmE-UUC
	GAG	274	0.50	GAG	271	0.49	GAG	271	0.49	
Stop	UAA	27	1.59	UAA	27	1.59	UAA	27	1.59	
	UGA	14	0.82	UGA	14	0.82	UGA	14	0.82	
	UAG	10	0.59	UAG	10	0.59	UAG	10	0.59	

表 2 枸杞植物叶绿体基因组
Table 2 Chloroplast genomes of *L. Chinese* plants

物种	长度/bp	GC1/%	GC2/%	GC3/%	GCall_s/%	ENC
中国枸杞	155 655	49.80	39.85	25.77	38.47	48.06
黄果枸杞	154 978	49.80	39.83	25.68	38.44	48.05
白果枸杞	154 976	49.80	39.83	25.68	38.44	48.05



* $P < 0.05$; ** $P < 0.01$.

图 1 枸杞叶绿体基因组密码子参数的相关性分析

Fig. 1 Correlation analysis of codon parameters in chloroplast genomes of *L. Chinese*

3.2 ENC-plot 分析

根据 ENC 值和 GC3 的关联分析以及标准曲线的距离来判断对密码子偏好性起主要作用的因素。当突变起主要作用时，数值主要分布于曲线附近，反之选择压力则起主导作用。根据 ENC-plot 分析图（图 2），发现 3 种枸杞植物的 ENC

值均大于 30，且多数数值分布于曲线附近。结合 ENC 比值频数分布表（表 3）显示，3 种枸杞叶绿体基因组分别有 20、21、21 个基因位于标准曲线附近，分别有 31、30、30 个基因远离标准曲线，表明影响大部分密码子偏好性的要素是自然选择。

表 3 ENC 比值频率分布

Table 3 Frequency distribution of ENC ratio

组段	中国枸杞		黄果枸杞		白果枸杞	
	频数	频率/%	频数	频率/%	频数	频率/%
(-0.2, -0.1]	2	0.04	2	0.04	2	0.04
(-0.1, 0]	20	0.39	21	0.41	21	0.41
(0, 0.1]	22	0.43	21	0.41	21	0.41
(0.1, 0.2]	6	0.12	6	0.12	6	0.12
(0.2, 0.3]	1	0.02	1	0.02	1	0.02
合计	51		51		51	

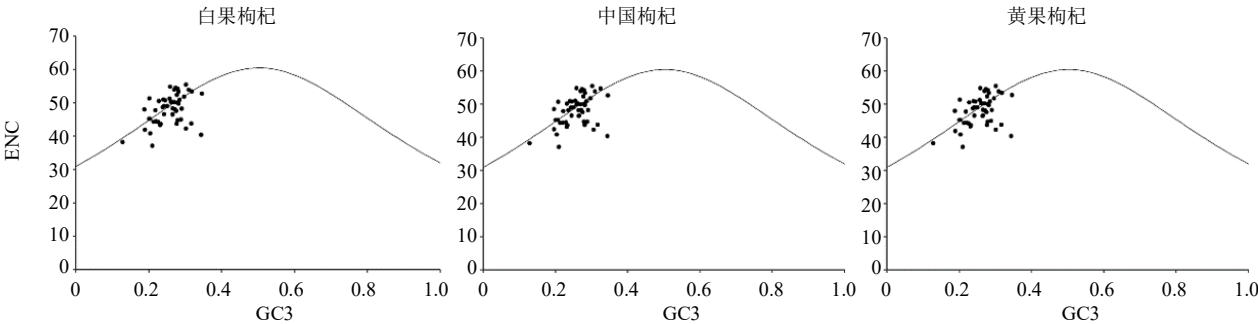


图 2 ENC-plot 分析图
Fig. 2 Analysis of ENC-plot

3.3 PR2-plot 和中性绘图分析

当影响密码子偏好性的因素仅为突变时,因突变具有随机性的特点会使第 3 位碱基 A/T (U) 或 C/G 具有相等的概率,而自然选择则会使 A/T (U) 或 G/C 使用不均等。由图 2 可知,4 个区域内的点分布不均匀:在垂直方向上多数基因在中线的下方;在水平方向上基因稍微偏向于中线右侧,垂直方向分布差异较大。(图 3)。因此,枸杞植物的密码子第 3 位碱基 C 频率低于 G, T (U) 频率大于 A,表明了枸杞植物多数基因受到自然选择的影响。

密码子第 3 位碱基的改变通常不会引起编码氨

基酸的改变,对密码子偏好性的研究有重要意义。可根据枸杞植物的中性绘图分析判断影响密码子偏好性的因素。当基因沿对角线分布且斜率接近 1 时,碱基在 3 个位置的差异不明显,说明影响密码子偏好性的主要因素是突变,反之则受自然选择。如图 4 所示,回归系数在 $-0.003\ 97\sim-0.001\ 74$,两者呈负相关关系。其中黄果枸杞的回归系数为 $-0.003\ 97$,受自然选择影响最小;中国枸杞的回归系数为 $-0.001\ 74$,自然选择的影响较大。中性绘图分析表明影响枸杞密码子偏好性的主要因素是自然选择,与 ENC-plot 的结果相同。

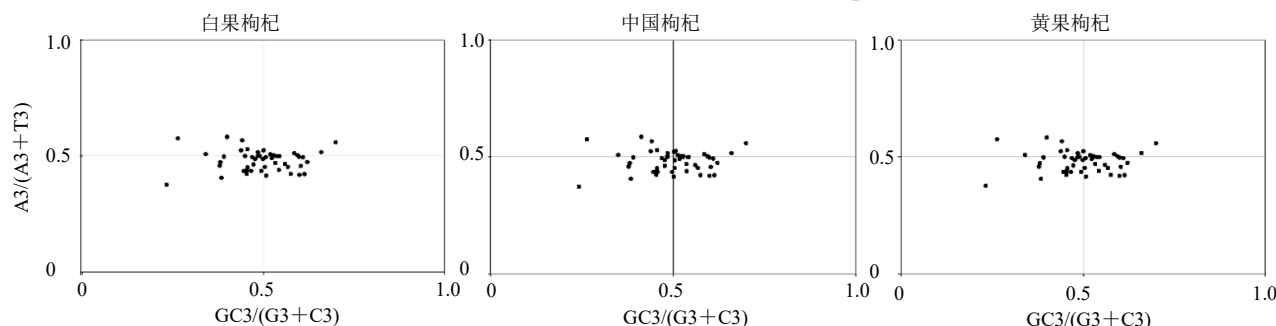


图 3 PR2-plot 分析

Fig. 3 Analysis of PR2-plot

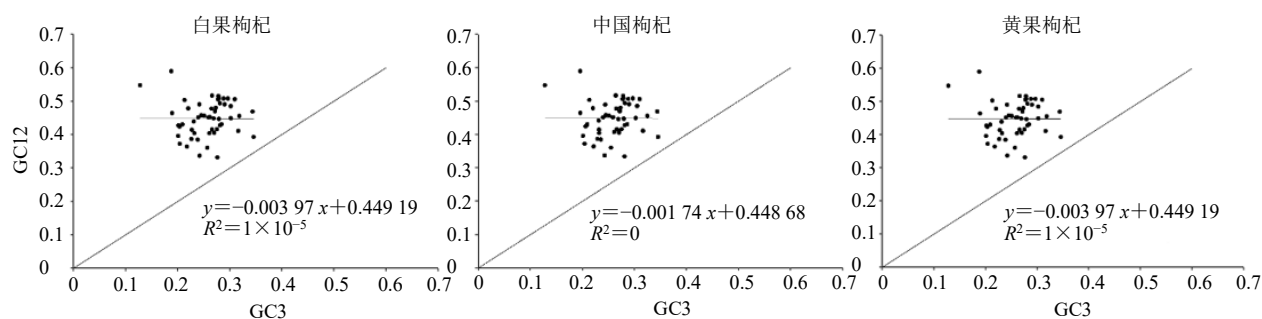


图 4 中性绘图分析

Fig. 4 Analysis of neutrality plot

3.4 最优密码子的确定及聚类分析

同时满足 $RSCU > 1$ 且 $\Delta RSCU$ 值 > 0.08 的密码子即为枸杞的最优密码子。结果显示(表 4):枸杞叶绿体基因组中绝大多数密码子以 A 和 U 结尾,且以 U 结尾的密码子较多。其中中国枸杞有 17 个最优密码子,黄果枸杞和白果枸杞具有 18 个相同的最优密码子。此外,3 种枸杞植物的共同最优密码子有 13 个,分别是 UGU、UAA、UUG、CCU、CGA、AUU 等。

对 3 种枸杞植物叶绿体基因组密码子 RSCU 进行聚类分析,使用 SPSS 24.0 进行聚类分析(图

5)。在欧氏距离为 5 时,9 个物种分为 3 大类:黑果枸杞、黄果枸杞和白果枸杞聚为一类,栽培番茄单独聚为一类,中国枸杞与其余近缘物种聚为一类;在欧氏距离为 10 时,黑果枸杞、黄果枸杞和白果枸杞与栽培番茄聚为一类,中国枸杞与其余近缘物种聚为一类;在欧氏距离为 25 时,所有茄科植物聚为一类。由此可知,中国枸杞未与白果枸杞、黄果枸杞聚为一类,这表明在密码子的使用模式上它们存在一定的差异。同时从枸杞植物的 RSCU 聚类结果来看白果枸杞与黄果枸杞的遗传关系较近。

表 4 枸杞植物叶绿体基因组密码子 RSCU 值
Table 4 RSCU values of codons in chloroplast genomes of *L. Chinese* plants

氨基酸	密码子	中国枸杞			黄果枸杞			白果枸杞		
		高表达基因 RSCU	低表达基因 RSCU	Δ RSCU	高表达基因 RSCU	低表达基因 RSCU	Δ RSCU	高表达基因 RSCU	低表达基因 RSCU	Δ RSCU
Phe	UUU	0.88	1.33	-0.45	1.52	1.47	0.05	1.52	1.47	0.05
	UUC	1.12	0.67	0.45	0.48	0.53	-0.05	0.48	0.53	-0.05
Ser	<u>UCU</u> (1)	2.00	1.66	0.34	1.54	1.66	-0.12	1.54	1.66	-0.12
	UCC	0.95	1.03	-0.08	1.08	1.02	0.06	1.08	1.02	0.06
	UCA(2,3)	0.95	1.03	-0.08	1.69	1.15	0.54	1.69	1.15	0.54
	UCG	0.29	0.52	-0.23	0.31	0.51	-0.20	0.31	0.51	-0.20
	AGU	1.33	1.55	-0.22	1.23	1.53	-0.30	1.23	1.53	-0.30
	AGC	0.48	0.21	0.27	0.15	0.13	0.02	0.15	0.13	0.02
Tyr	<u>UAU</u> (2,3)	1.38	1.59	-0.21	1.89	1.59	0.30	1.89	1.59	0.30
	UAC	0.62	0.41	0.21	0.11	0.41	-0.30	0.11	0.41	-0.30
Cys	<u>UGU</u> (1,2,3)	2.00	1.00	1.00	2.00	1.00	1.00	2.00	1.00	1.00
	UGC	0.00	1.00	-1.00	0.00	1.00	-1.00	0.00	1.00	-1.00
Leu	UUA	1.32	1.71	-0.39	1.79	1.75	0.04	1.79	1.75	0.04
	<u>UUG</u> (1,2,3)	1.56	1.45	0.11	1.70	1.37	0.33	1.70	1.37	0.33
	CUU(2,3)	1.48	1.45	0.03	1.52	1.22	0.30	1.52	1.22	0.30
	CUC	0.08	0.33	-0.25	0.09	0.30	-0.21	0.09	0.30	-0.21
	CUA	1.32	0.66	0.66	0.72	0.84	-0.12	0.72	0.84	-0.12
	CUG	0.25	0.40	-0.15	0.18	0.53	-0.35	0.18	0.53	-0.35
TER	UAA(1,2,3)	2.40	1.80	0.60	2.40	1.80	0.60	2.40	1.80	0.60
	UGA	0.00	0.60	-0.60	0.00	0.60	-0.60	0.00	0.60	-0.60
	UAG	0.60	0.60	0.00	0.60	0.60	0.00	0.60	0.60	0.00
	UGG	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
Trp	<u>CCU</u> (1,2,3)	1.94	1.43	0.51	1.58	1.18	0.40	1.58	1.18	0.40
	CCC	0.39	0.98	-0.59	0.73	0.82	-0.09	0.73	0.82	-0.09
	CCA	1.16	1.13	0.03	1.09	1.29	-0.20	1.09	1.29	-0.20
	CCG	0.52	0.45	0.07	0.61	0.71	-0.10	0.61	0.71	-0.10
His	CAU(2,3)	1.37	1.67	-0.30	1.80	1.65	0.15	1.80	1.65	0.15
	CAC	0.63	0.33	0.30	0.20	0.35	-0.15	0.20	0.35	-0.15
Arg	CGU	1.52	1.68	-0.16	1.22	1.66	-0.44	1.22	1.66	-0.44
	CGC	0.24	0.62	-0.38	0.19	0.62	-0.43	0.19	0.62	-0.43
	<u>CGA</u> (1,2,3)	2.08	1.59	0.49	2.25	1.66	0.59	2.25	1.66	0.59
	CGG	0.08	0.35	-0.27	0.09	0.31	-0.22	0.09	0.31	-0.22
	<u>AGA</u> (1,2,3)	1.36	1.24	0.12	1.50	1.34	0.16	1.50	1.34	0.16
	AGG	0.72	0.53	0.19	0.75	0.41	0.34	0.75	0.41	0.34
Gln	CAA(1)	1.56	1.37	0.19	1.57	1.54	0.03	1.57	1.54	0.03
	CAG	0.44	0.63	-0.19	0.43	0.46	-0.03	0.43	0.46	-0.03
Ile	<u>AUU</u> (1,2,3)	1.50	1.40	0.10	1.55	1.35	0.20	1.55	1.35	0.20
	AUC	0.73	0.56	0.17	0.40	0.59	-0.19	0.40	0.59	-0.19
	AUA	0.77	1.05	-0.28	1.05	1.06	-0.01	1.05	1.06	-0.01
Thr	ACU(1)	1.62	1.54	0.08	1.20	1.43	-0.23	1.20	1.43	-0.23
	ACC	0.97	0.82	0.15	0.80	0.86	-0.06	0.80	0.86	-0.06
	<u>ACA</u> (1,2,3)	1.30	1.03	0.27	1.87	1.00	0.87	1.87	1.00	0.87
	ACG	0.11	0.62	-0.51	0.13	0.71	-0.58	0.13	0.71	-0.58
Asn	AAU	1.21	1.65	-0.44	1.65	1.71	-0.06	1.65	1.71	-0.06
	AAC	0.79	0.35	0.44	0.35	0.29	0.06	0.35	0.29	0.06
Lys	<u>AAA</u> (1,2,3)	1.54	1.35	0.19	1.64	1.41	0.23	1.64	1.41	0.23
	AAG	0.46	0.65	-0.19	0.36	0.59	-0.23	0.36	0.59	-0.23
Met	AUG	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
Val	<u>GUU</u> (1,2,3)	2.10	1.42	0.68	1.84	1.33	0.51	1.84	1.33	0.51
	GUC	0.00	0.47	-0.47	0.22	0.53	-0.31	0.22	0.53	-0.31
	<u>GUA</u> (1,2,3)	1.71	1.22	0.49	1.62	1.24	0.38	1.62	1.24	0.38
	GUG	0.19	0.88	-0.69	0.32	0.89	-0.57	0.32	0.89	-0.57
Ala	GCU(1)	2.12	1.88	0.24	1.21	1.89	-0.68	1.21	1.89	-0.68
	GCC	0.48	0.82	-0.34	0.65	0.91	-0.26	0.65	0.91	-0.26
	GCA	1.21	0.71	0.50	1.86	0.70	1.16	1.86	0.70	1.16
	GCG	0.18	0.59	-0.41	0.28	0.49	-0.21	0.28	0.49	-0.21
Asp	GAU	1.00	1.80	-0.80	1.13	1.82	-0.69	1.13	1.82	-0.69
	GAC	1.00	0.20	0.80	0.88	0.18	0.70	0.88	0.18	0.70
Gly	<u>GGU</u> (1,2,3)	2.15	1.11	1.04	1.87	1.33	0.54	1.87	1.33	0.54
	GGC	0.37	0.62	-0.25	0.18	0.44	-0.26	0.18	0.44	-0.26
	GGA(2,3)	1.29	1.35	-0.06	1.69	1.33	0.36	1.69	1.33	0.36
	GGG	0.18	0.92	-0.74	0.27	0.89	-0.62	0.27	0.89	-0.62
Glu	GAA(1,2,3)	1.58	1.37	0.21	1.57	1.45	0.12	1.57	1.45	0.12
	GAG	0.42	0.63	-0.21	0.43	0.55	-0.12	0.43	0.55	-0.12

1-中国枸杞的最优密码子 2-黄果枸杞的最优密码子 3-白果枸杞的最优密码子；下划线代表 3 种枸杞相同的最优密码子。

1-The optimal codon for *L. chinense*. 2-The optimal codon for *L. barbarum* 3-The optimal codon for *L. dasystemum* Pojark. var. *rubricaulum*; The underline represents the same optimal codon for three types of *L. chinense*.

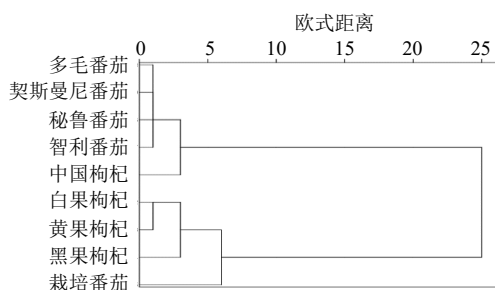


图 5 基于 RSCU 值的树状聚类

Fig. 5 Cluster tree based on RSCU values

4 讨论

由于物种和基因的演化，造成了密码子的使用偏性现象。并且密码子使用偏性在不同物种或同一物种的不同基因中有明显的差异^[27]。目前由于物种基因组测序的广泛使用，使人们对密码子使用模式的研究产生了兴趣^[27]。本研究通过对 3 种枸杞植物的密码子使用偏性和影响因素进行分析，为枸杞种质资源的开发和遗传育种提供参考。

在进化时遗传关系较近的物种，它们的第 3 位密码子碱基组成比较相同。如大多数双子叶植物的密码子偏好使用以 A/T (U) 碱基结尾，单子叶植物则以 G/C 碱基结尾^[28]。通常密码子第 3 位碱基受到较小的选择压力，对氨基酸的对应关系影响较小，因此密码子偏好性的研究对遗传育种具有重要意义^[29]。密码子偏性和氨基酸的组成受 GC 含量影响。目前研究发现，大部分植物的密码子使用偏性会随着物种的进化而降低，并且单细胞藻类的 GC 含量最高，双子叶植物的 GC 含量最低^[27]。3 种枸杞植物的密码子 GC 平均含量为 38.44%~38.47%，且主要以 A/T (U) 碱基结尾，这与大部分被子植物的密码子偏好性相同^[9, 29-30]。此外，本研究中 3 种枸杞植物的密码子偏好性均有较弱的表达 (ENC 值>48)，推测与叶绿体基因的保守性有关。

遗传密码子变异的主要进化力量包括 2 个因素，即碱基突变和自然选择。在具有极高核苷酸含量的基因组中，影响同义密码子使用模式的一个重要因素是突变^[31]。通过对枸杞 ENC-plot、中性绘图和 PR2-plot 分析发现，GC3 和 GC1、GC2 表现为不显著相关性，且 GC3<GC1、GC2，说明第 3 位碱基与第 1、2 位碱基存在差异性，且 A/T (U) 碱基有较高的使用频率，与双子叶植物的密码子偏好使用结果一致^[32-33]，这表明对枸杞密码子偏性影响较小的是碱基突变，可能更多地取决于物种进化和人

工选择作用。

在高表达基因组中，最优密码子的使用可以提高翻译效率，而除最优密码子外，低表达基因可以降低翻译效率^[34]。宿主细胞通过密码子碱基的组成影响外源基因的表达，也可通过优化密码子来提高基因表达的水平^[8]。在不同植物中，最优密码子的种类也存在巨大差异。在单子叶植物中，如水稻 *Oryza sativa* L.、玉米 *Zea mays* L.等通常以 G/C 结尾，而拟南芥、烟草等主要以 A/T 结尾，也有一些以 G/C 结尾的植物，南方菟丝子的高表达基因主要以 G/C 结尾^[31]。在物种的进化过程中，遗传密码子的偏性逐渐减少，低等植物需要更多的密码子优化，而高等植物的密码子优化较少^[30]，其主要原因是高等植物的基因表达受到顺式作用元件等多种因素的调控^[34]。枸杞的 ENC 平均值为 48.05~48.06，说明大多数基因在序列水平上表达潜力较低，但也有部分 ENC 值高的基因有相对高的组织表达，表明高等植物的基因表达受多种因素调控，由此可知，仅通过 ENC 值来预测基因表达水平是不准确的。从 3 种枸杞密码子中筛选出 71 个最优密码子，其中 13 个为共有最优密码子，大部分最优密码子以 A 或 U 结尾，尤其以 A 结尾，这与部分双子叶植物一致，说明高等植物叶绿体基因组的进化相对保守。

研究表明，根据 RSCU 值构建聚类分析图可能与其系统发育树不相同。前者只能在一定程度上支持系统分类，而不能真实反映物种的系统发育位置，这种现象一般是因为筛选的 RSCU 值缺少 CDS 序列的中低偏好密码子信息造成的；而后者包含位点突变和基因组非编码区序列等相关遗传信息，使种间关系更真实可靠。由于叶绿体基因在进化时受到的外界影响较小，因此可用于研究物种进化的进程^[8]。物种间的遗传关系越近，密码子的偏性越类似，欧式距离则越小^[35]。聚类图可以反映出某些物种的遗传关系在一定程度上与其密码子偏好性有关。聚类结果显示，黄果枸杞和白果枸杞遗传关系较近，单独聚为一类，而中国枸杞则与近缘物种聚为一类，与木兰科 (Magnoliaceae)^[36]、玉米^[37]、番茄^[38]等研究结果相似，进一步说明了 3 种果色枸杞的遗传关系，也为异源基因表达载体的设计和基因工程的应用和研究提供数据基础。

综上所述，3 种枸杞植物的叶绿体基因组的密码子碱基组成位置特征显示 GC1>GC2>GC3，3 种枸杞植物的叶绿体基因组密码子的使用

模式主要以 A/T (U) 碱基结尾; 密码子的使用偏好性 ENC-plot、PR2-plot 和中性分析显示枸杞植物的密码子使用模式主要受自然选择的影响; 最优密码子分析显示枸杞材料共有密码子有 13 个; 聚类分析表明 3 种枸杞的密码子使用存在一些差异, 其中黄果枸杞和白果枸杞无明显差异, 说明其遗传关系较近。3 种枸杞植物的叶绿体基因组特征为改良枸杞植物重要性状和探究枸杞的系统发育提供基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 檀鹏辉, 袁丽丽, 樊波, 等. 日本结缕草滞绿基因 ZjSGR 对烟草的转化及功能分析 [J]. 草业学报, 2017, 26(5): 155-162.
- [2] Zhang W, Zhao Y L, Yang G Y, *et al.* Determination of the evolutionary pressure on *Camellia oleifera* on Hainan Island using the complete chloroplast genome sequence [J]. *PeerJ*, 2019, 7: e7210.
- [3] 张悦. 红豆杉科叶绿体基因组结构多样化研究 [D]. 北京: 北京林业大学, 2019.
- [4] 任婷. 十种报春花属植物的叶绿体基因组研究 [D]. 西安: 西北大学, 2018.
- [5] Li S F, Su T, Cheng G Q, *et al.* Chromosome evolution in connection with repetitive sequences and epigenetics in plants [J]. *Genes*, 2017, 8(10): 290.
- [6] 姜汶君, 郭梦月, 庞晓慧. 叶绿体基因组在药用植物鉴定及系统进化研究中的应用 [J]. 世界中医药, 2020, 15(5): 702-708.
- [7] 丁彦强, 方扬, 靳艳玲, 等. 基于叶绿体基因组的浮萍亚科系统进化 [J]. 应用与环境生物学报, 2017, 23(2): 215-219.
- [8] 马孟莉, 张薇, 孟衡玲, 等. 草果叶绿体基因组特征及系统发育分析 [J]. 中草药, 2021, 52(19): 6023-6031.
- [9] 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.
- [10] Belser C, Istace B, Denis E, *et al.* Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps [J]. *Nat Plants*, 2018, 4(11): 879-887.
- [11] Romero H, Zavala A, Musto H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces [J]. *Nucleic Acids Res*, 2000, 28(10): 2084-2090.
- [12] 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/OL]. 草业学报, 2023. [2023-03-18]. <https://kns.cnki.net/kcms/detail/62.1105.S.20230316.1737.012.html>.
- [13] 舒军霞, 杨林, 周涛, 等. 刺柏属 4 种药用植物叶绿体基因组密码子偏好性分析 [J]. 中草药, 2022, 53(23): 7507-7515.
- [14] 杨惠娟, 刘国顺, 张松涛, 等. 烟草叶绿体密码子的偏好性及聚类分析 [J]. 中国烟草学报, 2012, 18(2): 37-43.
- [15] 段祥光, 于涛, 曹萌, 等. 樟叶槭叶绿体基因组序列密码子偏性分析 [J]. 西北林学院学报, 2023, 38(1): 102-107.
- [16] 郑长远, 苏旭, 刘玉萍, 等. 葵花大蓟叶绿体基因组结构特征及系统发育分析 [J]. 药学报, 2022, 57(7): 2224-2233.
- [17] 王艳, 赵懿琛, 赵德刚. 杜仲基因密码子使用模式分析 [J]. 广西植物, 2021, 41(2): 274-282.
- [18] 张得芳, 夏涛. 基于转录组的唐古特白刺密码子偏好性分析 [J]. 干旱区资源与环境, 2021, 35(12): 163-170.
- [19] Yu C H, Dang Y K, Zhou Z P, *et al.* Codon usage influences the local rate of translation elongation to regulate co-translational protein folding [J]. *Mol Cell*, 2015, 59(5): 744-754.
- [20] 吴宪明, 吴松锋, 任大明, 等. 密码子偏性的分析方法及相关研究进展 [J]. 遗传, 2007, 29(4): 420-426.
- [21] 唐晓芬, 陈莉, 马玉韬. 密码子使用偏性量化方法研究综述 [J]. 基因组学与应用生物学, 2013, 32(5): 660-666.
- [22] 杨云彭, 马晓焉, 霍毅欣. 密码子优化策略在异源蛋白表达中的应用 [J]. 生物工程学报, 2019, 35(12): 2227-2237.
- [23] Zhang R Z, Zhang L, Wang W, *et al.* Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *Solanum* species [J]. *Int J Mol Sci*, 2018, 19(10): 3142.
- [24] Tang L, Shah S, Chung L, *et al.* Cloning and heterologous expression of the epothilone gene cluster [J]. *Science*, 2000, 287(5453): 640-642.
- [25] 王益民, 张宝琳. 我国枸杞属物种资源及发展对策 [J]. 世界林业研究, 2021, 34(3): 107-111.
- [26] 刘俭, 张波, 秦垦, 等. 不同枸杞种质间品质比较与分析 [J]. 江西农业学报, 2015, 27(1): 53-56.
- [27] Haig D, Westoby M. Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis [J]. *Philos Trans R Soc Lond B*, 1991, 333(1266): 1-13.
- [28] Li N, Li Y Y, Zheng C C, *et al.* Genome-wide comparative analysis of the codon usage patterns in plants [J]. *Genes Genom*, 2016, 38(8): 723-731.
- [29] 王晓双, 杨芳, 罗茜, 等. 石斛属叶绿体基因组密码子

- 使用偏性及系统发育分析 [J]. 分子植物育种, 2022, 20(23): 7759-7769.
- [30] Tang D F, Wei F, Cai Z Q, *et al.* Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth [J]. *Dev Genes Evol*, 2021, 231(1/2): 1-9.
- [31] 刘潮, 韩利红, 吴丽芳, 等. 辣椒基因组密码子使用偏性分析 [J]. 江苏农业科学, 2022, 50(5): 16-22.
- [32] 关云会, 姚文燕, 杨青淑, 等. 西藏凹乳芹叶绿体基因组密码子偏好性分析 [J]. 分子植物育种, 2021, 19(23): 7771-7781.
- [33] 沈莲文, 田金红, 王玉昌, 等. 2 种玉兰属植物叶绿体基因组密码子偏好性分析 [J]. 西南林业大学学报: 自然科学, 2023, 43(2): 44-53.
- [34] Liu X Y, Li Y, Ji K K, *et al.* Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in *Cuscuta australis* [J]. *Genomics*, 2020, 112(4): 2695-2702.
- [35] Wang Z J, Cai Q W, Wang Y, *et al.* Comparative analysis of codon bias in the chloroplast genomes of Theaceae species [J]. *Front Genet*, 2022, 13: 824610.
- [36] 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.
- [37] 晁岳恩, 常阳, 王美芳, 等. 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, 2012, 27(4): 60-64.
- [38] Zhang R, Zhang L, Wang W, *et al.* Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild solanum species [J]. *Intern J Mol Sci*, 2018, 19(10): 3142.

[责任编辑 时圣明]