

云南蓝果树叶绿体基因组密码子偏好性分析

原晓龙,康洪梅,王毅*

(云南省林业和草原科学院 云南省森林植物培育与开发利用重点实验室/国家林业和草原局 云南珍稀濒危森林植物保护和繁育重点实验室,云南 昆明 650201)

摘要:以云南蓝果树叶绿体基因组为研究对象,采用 Codon W 1.4.2 等软件,通过分析云南蓝果树叶绿体基因组 52 个基因密码子的偏好性,探究影响密码子使用偏性形成的主要因素。结果表明:1) Codon W 和 CUSP 软件分析显示密码子第 3 位碱基 GC 含量为 28.40%,ENC 值>45 的基因有 39 个,说明密码子偏好以 AT 结尾,且存在较弱偏性。2)中性绘图分析显示,GC₁₂与 GC₃ 的相关系数为 0.161 2,相关性不显著,说明密码子第 1、第 2 位与第 3 位碱基组成存在显著差异。3) ENC 绘图结合 ENC 比值频率表显示,大部分基因距离标准曲线较远,说明密码子偏好性主要受选择的影响。4) PR2-plot 分析显示密码子第 3 位碱基使用频率方面,T>A,G>C,说明密码子偏好性受多重因素影响。5) 确定 UUG、UCA、GCU、AAU、GAU 和 GGA 为最优密码子。综上所述,云南蓝果树叶绿体基因组密码子的使用主要受选择的影响,受其他如突变等因素的影响较弱。

关键词:云南蓝果树;叶绿体基因组;密码子偏好性;选择;最优密码子

中图分类号:Q71 **文献标志码:**A **文章编号:**1001-7461(2020)04-0026-06

Codon Usage Bias Analysis of Chloroplast Genome in *Nyssa yunnanensis*

YUAN Xiao-long, KANG Hong-mei, WANG Yi*

(Yunnan Provincial Key Laboratory of Cultivation and Exploitation of Forest Plants/Conservation of Rare, Endangered & Endemic Forest Plants, Public Key Laboratory of the National Forestry and Grassland Administration, Yunnan Academy of Forestry and Grassland, Kunming 650201, Yunnan, China)

Abstract: The chloroplast genome of *Nyssa yunnanensis* was taken as the research object, the codon usage bias of 52 genes from the chloroplast genome of *N. yunnanensis* was analyzed by the software Codon W 1.4.2, and the main factors affecting the formation of codon usage bias was explored. The results showed as follows 1) According to the analysis results of the software Codon W and CUSP, the GC content of the 3rd base composition of the codons was 28.40%, there were 39 genes when ENC value was greater than 45, therefore, these phenomena illustrated that the codons preferred to end with AT, and had a weaker bias. 2) Neutrality analysis revealed the correlation coefficient of GC₁₂ and GC₃ was 0.161 2, these had no significant correlation, which declared that the 1st two base composition and the 3rd of the codons existed significant variation. 3) The combination of ENC plot with ENC ration table indicated that most genes were far from the standard curve, which suggested that the codon usage bias was principally affected by selection pressure. 4) PR2-plot analysis revealed that the 3rd codon base composition in the frequency of utilization, T > A and G > C, which stated the codon usage bias was affected by multi-factors. 5) The optimal codons were determined as UUG, UCA, GCU, AAU, GAU and GGA. It was summarized that the codon usage bias of the genes from the chloroplast genome of *N. yunnanensis* was mainly affected by selection pressure, oth-

收稿日期:2019-09-10 修回日期:2019-12-18

基金项目:国家自然科学基金(31860177); 云南省应用基础研究计划(2016FD097)。

作者简介:原晓龙。研究方向:林木分子生物技术与分子生物学。E-mail: xiaolony@126.com

* 通信作者:王毅,博士,副研究员。研究方向:分子生物学。E-mail: 22825818@qq.com

er factors, such as mutation presented little effect.

Key words: *Nyssa yunnanensis*; chloroplast genome; codon usage bias; selection; optimal codon

信使 RNA(mRNA)上每相邻的 3 个核苷酸组成一个三联体密码子,共有 64 种密码子,与 20 种氨基酸相对应;除蛋氨酸和色氨酸外,其余 18 种氨基酸均有 2 ~ 6 个密码子,这些编码同一氨基酸的密码子称为同义密码子(synonymous codon)^[1]。在 mRNA 翻译过程中,同义密码子的使用频率并不完全均等,在某些物种的使用频率上偏好某种或某几种同义密码子,这种现象称为密码子偏好性(codon usage bias)^[2]。密码子偏好性广泛存在,是在物种长期进化过程中受密码子碱基组成^[3]、环境选择^[4]、tRNA 丰度^[5]、mRNA 二级结构^[6]、基因表达水平^[7]和碱基突变^[8]等影响形成。密码子偏好性可通过调节基因翻译的准确性和效率进而影响基因的表达水平^[9],编码与光合作用相关蛋白的叶绿体基因组属核外遗传因子,其规模小、拷贝数多^[10]、具相对保守的分子结构等特点^[11];因此从叶绿体基因组中选择最优密码子并设计叶绿体基因表达载体,可迅速提高特定基因在叶绿体中的表达量,并根据已知的密码子使用模式推断未知基因的表达模式,或预测特定未知基因的功能^[12]。亲缘关系较近的物种具相近的密码子使用模式^[1],因此研究叶绿体基因组密码子的使用模式对探讨植物系统进化、遗传多样性分析和提高外源基因表达量等具有重要意义。

云南蓝果树(*Nyssa yunnanensis*)属蓝果树科(Nyssaceae)蓝果树属(*Nyssa*)落叶乔木,是我国热带雨林特有的极度濒危植物^[13-14],分布范围狭窄、野外个体数量极少且存在天然更新困难等难题,目前已处于稳定存活界限,濒临灭绝^[14-15],亟待进行保护,通过人工干预的方式增加云南蓝果树的植株数量、逐步恢复和重建种群的基础上,如何保持其原有的生物多样性和基因资源,本研究期望通过分析云南蓝果树叶绿体基因组密码子偏好性,确定其偏好密码子,为保持云南蓝果树遗传和基因资源多样性、防止种群衰退提供科学依据。

1 材料与方法

1.1 序列的获取与筛选

从云南省林业和草原科学院普文热带林业研究所的种植资源圃中采集新鲜幼嫩、完整的云南蓝果树叶片,置于液氮环境下带回实验室,叶绿体基因组测序由生工生物工程(上海)股份有限公司,再将测序结果提交至美国国立生物技术中心(National Center for Biotechnology Information, NCBI)(登

录编号:MG525000)。根据基因组注释信息,为降低因基因序列较短所带来的样本误差,剔除 87 基因编码序列(coding DNA sequence, CDS)中长度 < 300 bp、序列不完整及内部存在终止密码子的基因序列^[3]共 25 条,得到 52 条 CDS 并将其用于分析云南蓝果树叶绿体基因组密码子使用模式。

1.2 方法

对从云南蓝果树叶绿体基因组中筛选到的 52 条 CDS 进行分析,其中通过在线软件 CUSP(<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>)计算不同基因密码子各位置的 GC 含量(GC₁、GC₂、GC₃)^[16],以 GC₁₂(GC₁、GC₂ 的平均值)为纵坐标、以 GC₃ 位横坐标进行中性绘图分析,以 GC₃ 位横坐标、ENC 值(有效密码子数, effective number of codon, ENC)为纵坐标进行 ENC-plot 分析,以 A₃/(A₃+T₃)为纵坐标、G₃/(G₃+C₃)为横坐标进行 PR2-plot 分析。通过上述分析以明确云南蓝果树叶绿体基因组密码子的使用偏性规律^[11,16]。

结合云南蓝果树叶绿体基因组密码子的高频率密码子和高表达密码子,将 RSCU 值>1 的密码子确定为高频密码子;以 ENC 值为偏好性标准,选择 2 极 10 %的基因构建高低偏性基因库,将 ΔRSCU≥0.08 的密码子定为高表达优越密码子;最终将同时满足高频率密码子和高表达密码子 2 种条件的密码子定为云南蓝果树叶绿体基因组的最优密码子^[1]。

2 结果与分析

2.1 密码子碱基组成

应用 Codon W 1.4.2 软件和在线程序 CUSP 分析云南蓝果树叶绿体基因组中的 52 条 CDS 的碱基组成和 ENC 值。结果显示(表 1),其密码子 3 个碱基的平均 GC 含量为 38.47%,GC₁ 为 47.31%,GC₂ 为 39.71%,GC₃ 为 28.40%。即密码子不同位置的 GC 含量存在较大差异,具体趋势为第 1 位最高,其次第 2 位,第 3 位最低。在云南蓝果树中,其叶绿体基因组的密码子使用频率表现为 AT 结尾的密码子高于 CG 结尾的,说明云南蓝果树叶绿体基因组中富含 AT。

ENC 值的理论范围为 20~61,当 ENC 值为 20 时,代表完全偏倚;当 ENC 值为 61 时,代表完全无偏倚^[17]。云南蓝果树叶绿体基因组中的 ENC 值为 37.46~55.23,平均 46.92,且 ENC 值>45 的基因有 39 条,说明该叶绿体基因组中的密码子存在较弱

偏性。

密码子不同位置的 GC 含量与 ENC、密码子数量 N 的相关性分析(表 2)显示,GC₁ 与 GC₂、GC_{all}呈极显著相关,GC₂ 和 GC_{all} 呈极显著相关,GC₃ 与 GC_{all}、N、ENC 呈极显著相关,而 GC₁、GC₂ 和 GC₃ 均呈现不显著相关。说明在云南蓝果树叶绿体基因组中,其密码子碱基组成第 1 位、第 2 位与第 3 位存在较大差异,而前 2 位碱基组成较为类似。ENC 与

GC₃、N 呈极显著相关,但与 GC₁ 和 GC₂ 均不呈显著相关性,说明密码子第 3 位碱基组成和基因的长度对该基因组密码子偏好性有影响。RSCU 分析(表 3)显示,在 59 个同义密码子(除去 3 个终止密码子、色氨酸(Thr)和甲硫氨酸(Met)的密码子)中,RSCU>1 的密码子有 29 个,其中 16 个以 U 结尾、12 个以 A 结尾、1 个以 G 结尾;说明云南蓝果树的叶绿体基因组偏好 AU 结尾的密码子。

表 1 云南蓝果树叶绿体基因组密码子 3 个位置 GC 含量与 ENC 值

Table 1 The GC content of 3 positions and ENC value of the codons from chloroplast genome in *Nyssa yunnanensis*

基因	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC _{all} /%	ENC	基因	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC _{all} /%	ENC
<i>psbA</i>	50.28	43.79	33.05	42.37	55.23	<i>rps12</i>	51.61	48.39	29.03	43.01	46.40
<i>matK</i>	41.19	32.08	29.11	34.13	50.82	<i>clpP</i>	60.71	37.76	29.59	42.69	43.57
<i>atpA</i>	55.12	40.75	27.17	41.01	51.83	<i>psbB</i>	55.01	46.17	33.01	44.73	44.70
<i>atpF</i>	47.62	35.45	31.75	38.27	49.55	<i>petB</i>	49.07	41.20	30.09	40.12	50.05
<i>atpI</i>	48.79	37.90	27.42	38.04	53.48	<i>petD</i>	50.93	39.13	24.22	38.1	44.56
<i>rps2</i>	44.73	44.30	29.54	39.52	50.25	<i>rpoA</i>	44.38	31.95	26.92	34.42	48.54
<i>rpoC2</i>	46.44	37.87	30.09	38.13	51.97	<i>rps11</i>	54.56	57.55	24.46	45.56	46.22
<i>rpoC1</i>	50.80	37.99	27.51	38.77	53.13	<i>rps8</i>	40.74	41.48	24.44	35.56	49.54
<i>rpoB</i>	50.79	37.63	27.54	38.66	49.52	<i>rpl14</i>	56.10	37.40	26.83	40.11	47.47
<i>psbD</i>	52.26	43.22	31.36	42.28	46.50	<i>rpl16</i>	48.91	51.82	24.82	41.85	48.03
<i>psbC</i>	53.59	46.20	32.49	44.09	49.97	<i>rps3</i>	47.49	33.33	26.94	35.92	47.51
<i>psaB</i>	48.84	42.99	30.75	40.86	48.24	<i>rpl22</i>	44.30	36.08	24.68	35.02	45.81
<i>psaA</i>	52.73	43.54	32.62	42.96	51.37	<i>rpl2</i>	51.09	48.55	32.25	43.96	45.19
<i>ycf3</i>	47.34	38.46	29.59	38.46	50.38	<i>ycf2</i>	41.64	34.50	36.95	37.70	43.54
<i>rps4</i>	51.49	38.12	26.73	38.78	45.73	<i>ndhB</i>	41.49	38.36	31.31	37.05	46.13
<i>ndhJ</i>	51.57	37.74	31.45	40.25	47.34	<i>rps7</i>	53.21	45.51	23.72	40.81	46.21
<i>ndhK</i>	45.58	44.25	25.66	38.50	48.05	<i>ndhF</i>	37.72	36.11	24.70	32.84	39.85
<i>ndhC</i>	44.63	33.88	24.79	34.44	47.35	<i>ccsA</i>	33.85	38.20	25.47	32.51	43.94
<i>atpE</i>	51.49	39.55	28.36	39.8	47.67	<i>ndhD</i>	41.03	37.08	27.61	35.24	40.47
<i>atpB</i>	56.31	42.28	28.66	42.42	45.5	<i>ndhE</i>	38.24	35.29	26.47	33.33	42.52
<i>rbcL</i>	58.19	43.70	30.67	44.19	48.8	<i>ndhG</i>	44.07	33.90	23.73	33.90	38.79
<i>accD</i>	40.39	36.31	31.07	35.92	45.91	<i>ndhI</i>	42.01	37.28	26.63	35.31	41.83
<i>ycf4</i>	42.70	40.54	31.89	38.38	46.72	<i>ndhA</i>	44.51	39.01	23.08	35.53	37.46
<i>cemA</i>	40.43	29.57	30.87	33.62	47.21	<i>ndhH</i>	51.27	37.06	30.20	39.51	37.78
<i>petA</i>	52.34	37.07	31.15	40.19	52.77	<i>ycf1</i>	37.07	29.56	26.65	31.09	40.77
<i>rps18</i>	34.31	43.14	25.49	34.31	48.36	平均值	47.31	39.71	28.40	38.47	46.92
<i>rpl20</i>	38.98	44.07	26.27	36.44	49.27						

表 2 云南蓝果树叶绿体基因组密码子各参数的相关性分析

Table 2 Correlation analysis of the parameters of the codons in chloroplast genome of *N. yunnanensis*

	GC1	GC2	GC3	GCall	N
GC ₂	0.421**				
GC ₃	0.241	0.014			
GC _{all}	0.862**	0.749**	0.433**		
N	-0.152	-0.274*	0.367**	-0.120	
ENC	0.272	-0.081	0.424**	0.239	0.298*

注: ** 在 0.01 水平(双侧)上显著相关, * 在 0.05 水平(双侧)上显著相关。

2.2 中性绘图分析

统计密码子 3 个碱基位置上的 GC 含量(GC₁、GC₂、GC₃),以第 1 位和第 2 位上碱基 GC 含量的平

均值 GC₁₂为纵坐标,以 GC₃ 为横坐标分析其相关性,探究影响密码子偏好性的主要因素。主要判断依据为:如果两者显著相关说明主要受突变的影响,不显著则主要受选择的影响^[18]。GC₁₂ 和 GC₃ 的中性绘图分析(图 1)显示,GC₁₂值在 0.3~0.6,GC₃ 的值在 0.2~0.4;仅 1 个基因位于对角线附近,其余基因均位于对角线上方较远位置分布;GC₁₂与 GC₃ 的相关系数为 0.161 2,回归系数为 0.257 5,二者相关性不显著,第 1、第 2 位与第 3 位碱基组成差异显著,GC₁₂与 GC₃ 两者相关性较弱,云南蓝果树叶绿体基因组中的 GC 含量高度保守,且密码子第 3 位碱基上的 GC 含量最低,说明云南蓝果树叶绿体基因组密码子偏好性主要受选择的影响。

表 3 云南蓝果树叶绿体基因组的 RSCU 分析

Table 3 RSCU analysis in the chloroplast genome of *N. yunnanensis*

AA	密码子	数量	RSCU	AA	密码子	数量	RSCU	AA	密码子	数量	RSCU
Phe	<u>UUU</u>	751	1.29	Pro	<u>CCU</u>	361	1.63	Asp	<u>GAU</u>	714	1.60
	UUC	411	0.71		CCC	159	0.72		GAC	180	0.40
Leu	<u>UUA</u>	716	1.92		<u>CCA</u>	259	1.17	Glu	<u>GAA</u>	861	1.51
	<u>UUG</u>	463	1.24		CCG	109	0.49		GAG	281	0.49
	<u>CUU</u>	475	1.27	Thr	<u>ACU</u>	443	1.64	Cys	<u>UGU</u>	170	1.46
	CUC	148	0.40		ACC	197	0.73		UGC	63	0.54
	CUA	296	0.79		<u>ACA</u>	326	1.21	Arg	<u>CGU</u>	285	1.34
	CUG	140	0.38		ACG	113	0.42		CGC	82	0.39
Val	<u>GUU</u>	422	1.45	Ala	<u>GCU</u>	544	1.85		CGA	279	1.31
	GUC	128	0.44		GCC	185	0.63		CGG	102	0.48
	<u>GUA</u>	446	1.54		<u>GCA</u>	330	1.12		<u>AGA</u>	393	1.85
	GUG	166	0.57		GCG	119	0.40	Gly	<u>AGG</u>	133	0.63
Ser	<u>UCU</u>	467	1.75	Tyr	<u>UAU</u>	649	1.62		<u>GGU</u>	489	1.31
	UCC	250	0.94		UAC	152	0.38		GGC	161	0.43
	<u>UCA</u>	312	1.17	His	<u>CAU</u>	404	1.53		<u>GGA</u>	594	1.59
	UCG	148	0.55		CAC	124	0.47		GGG	251	0.67
	<u>AGU</u>	333	1.25	Gln	<u>CAA</u>	602	1.54		<u>AUU</u>	873	1.43
	AGC	91	0.34		CAG	178	0.46	Ile	AUC	384	0.63
Lys	<u>AAA</u>	816	1.52	Asn	<u>AAU</u>	788	1.56		AUA	574	0.94
	AAC	224	0.44		AAG	260	0.48				

注:下划线标注偏爱密码子。

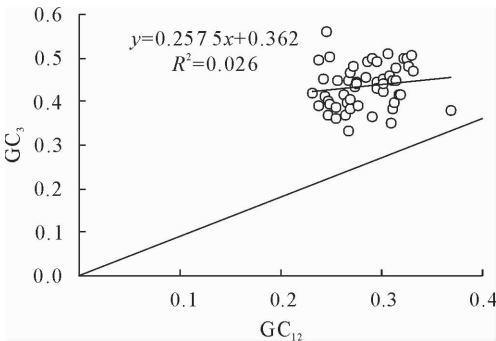


图 1 GC₁₂和 GC₃ 的中性绘图分析

Fig. 1 Neutrality analysis of GC₁₂ and GC₃

2.3 ENC-plot 分析

ENC-plot 分析通过比较各基因与标准曲线的距离可以判断基因密码子的主要影响因素,距离标准曲线较近的基因主要受突变的影响,较远则受选择的影响^[17]。ENC-plot 分析(图 2)显示,大部分基因距标准曲线较远。为了更准确的检验基因与标准曲线的距离,可以采用 ENC 比值频率(表 4),ENC 比值位于-0.05~0.05 的基因有 15 个,即有 15 个基因的实际 ENC 值与预期 ENC 值接近,与标准曲线距离较近;ENC 比值位于-0.05~0.05 外的基因有 37 个,即有 37 个基因与标准曲线距离较远。说明云南蓝果树叶绿体基因密码子偏好性主要受选择的影响,受其他因素的影响较弱。

2.4 PR2-plot 分析

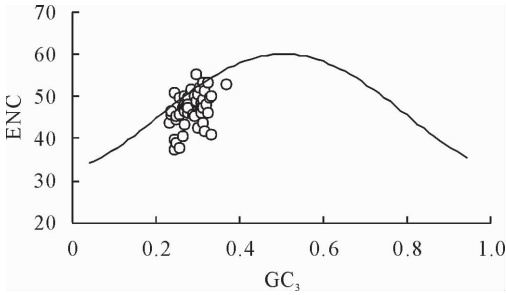
PR2-plot 分析可以推测密码子第 3 位上 4 个碱

基的具体使用情况,能分析各基因部分氨基酸中嘌呤(A 和 G)与嘧啶(T 和 C)的关系,如果各基因密码子偏好性完全受突变影响,则 A 和 T,G 和 C 的使用频率应当相等,即各基因均匀分布在 PR2 平面图的各部分^[19]。PR2-plot 分析显示(图 3),较多基因位于 PR2 平面图的右下方,在各碱基使用频率方面,T>A,G>C,说明云南蓝果树叶绿体基因组密码子偏好性受多重因素的综合影响。

表 4 ENC 比值频率分布

Table 4 Distribution of ENC ratio

组限	中值	频数	频率
-0.15~-0.05	-0.10	2	0.038
-0.05~0.05	0	15	0.288
0.05~0.15	0.10	26	0.500
0.15~0.25	0.20	9	0.173



注:标准曲线方程为 $ENC=2+GC_3+(29/GC_3^2+(1-GC_3)^2)$ 。

图 2 ENC-plot 分析

Fig. 2 ENC-plot analysis

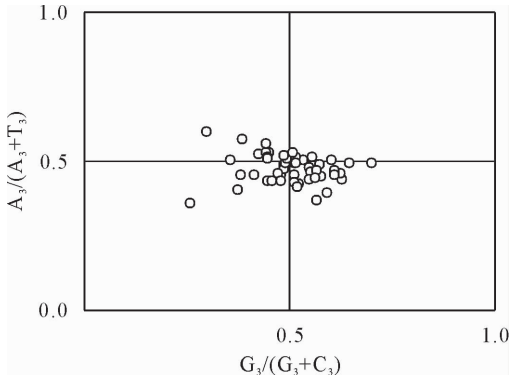


图 3 PR2-plot 分析

Fig. 3 PR2-plot analysis

表 5 云南蓝果树叶绿体基因组最优密码子的确定

Table 5 The determination of preferred codons in chloroplast genome of *N. yunnanensis*

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		数目	RSCU	数目	RSCU	
Phe	UUU	105	1.08	23	1.64	-0.56
	UUC***	89	0.92	5	0.36	0.56
Leu	UUA	66	1.15	23	2.19	-1.04
	UUG*	92	1.60	16	1.52	0.08
	CUU	84	1.46	19	1.81	-0.35
	CUC**	30	0.52	2	0.19	0.33
	CUA***	47	0.82	2	0.19	0.63
	CUG**	26	0.45	1	0.10	0.35
Ser	UCU	77	1.53	12	1.71	-0.18
	UCC**	60	1.20	6	0.86	0.34
	UCA*	56	1.12	7	1.00	0.12
	UCG	39	0.78	7	1.00	-0.22
	AGU	51	1.02	10	1.43	-0.41
	AGC**	18	0.36	0	0	0.36
Pro	CCU	52	1.32	13	1.58	-0.26
	CCC	36	0.91	9	1.09	-0.18
	CCA	46	1.16	9	1.09	0.07
	CCG**	24	0.61	2	0.24	0.37
Thr	ACU	44	1.24	11	1.22	0.02
	ACC	28	0.79	10	1.11	-0.32
	ACA	45	1.27	14	1.56	-0.29
	ACG***	25	0.70	1	0.11	0.59
Ala	GCU*	52	1.64	14	1.51	0.13
	GCC*	26	0.82	5	0.54	0.28
	GCA	35	1.10	13	1.41	-0.31
	GCG	14	0.44	5	0.54	-0.10
Val	GUU	53	1.49	14	1.51	-0.02
	GUC**	19	0.54	2	0.22	0.32
	GUA	43	1.21	17	1.84	-0.63
	GUG**	27	0.76	4	0.43	0.33
Tyr	UAU	105	1.65	15	1.76	-0.11
	UAC*	22	0.35	2	0.24	0.11
His	CAU	63	1.47	6	2.00	-0.53
	CAC***	23	0.53	0	0	0.53
Gln	CAA	90	1.34	16	1.88	-0.54
	CAG***	44	0.66	1	0.12	0.54
Asn	AAU*	154	1.57	24	1.41	0.16
	AAC	42	0.43	10	0.59	-0.16
Lys	AAA	140	1.33	30	1.76	-0.43
	AAG**	71	0.67	4	0.24	0.43
Asp	GAU**	155	1.65	10	1.25	0.40
	GAC	33	0.35	6	0.75	-0.40
Glu	GAA	138	1.37	21	1.62	-0.25
	GAG*	63	0.63	5	0.38	0.25
Cys	UGU	26	1.33	4	2.00	-0.67
	UGC***	13	0.67	0	0	0.67
Arg	CGU	35	0.92	14	1.25	-0.33
	CGC**	14	0.37	0	0	0.37
	CGA	54	1.41	22	1.97	-0.56
	CGG**	21	0.55	2	0.18	0.37
Gly	AGA	73	1.91	24	2.15	-0.24
	AGG**	32	0.84	5	0.45	0.39
	GGU	50	1.03	19	1.73	-0.70
	GGC	18	0.37	4	0.36	0.01
Ile	GGA*	81	1.66	17	1.55	0.11
	GGG***	46	0.94	4	0.36	0.58
	AUU	125	1.30	31	1.45	-0.15
	AUC*	69	0.72	12	0.56	0.16
	AUA	94	0.98	21	0.98	0

注：* 表示 Δ RSCU ≥ 0.08 ，** 表示 Δ RSCU ≥ 0.30 ，*** 表示 Δ RSCU ≥ 0.50 。

3 结论与讨论

云南蓝果树叶绿体基因组密码子存在使用偏性，其使用偏好性受多重因素的共同作用，其中选择是影响该叶绿体基因组密码子偏好性的主要因素，

2.5 最优密码子的确定

将云南蓝果树叶绿体基因组中的 52 条基因整合成一个整体在 Codon W 1.4.2 软件上运行，通过构建高表达和低表达基因库，结果显示(表 5)， Δ RSCU ≥ 0.08 (用 * 表示)的密码子有 28 个，为高表达的优越密码子， Δ RSCU ≥ 0.30 (用 ** 表示)的密码子有 19 个， Δ RSCU ≥ 0.50 (用 *** 表示)的密码子有 7 个；以同时满足高频、高表达优越密码子这 2 个条件，即同时满足 Δ RSCU ≥ 0.08 和 RSCU ≥ 1.00 的密码子作为最优密码子，最优密码子分别为 UUG、UCA、GCU、AAU、GAU 和 GGA。

而受突变等其他因素的影响较弱；中性绘图分析、ENC-plot 分析等均说明其受选择压力的影响最大。根据最优密码子需同时满足高频密码子和高表达优越密码子 2 个条件，最终确定了 UUG、UCA、GCU、AAU、GAU 和 GGA 6 个密码子为最优密码子。

密码子是生物体中联结核酸和蛋白质的钥匙,而密码子的使用频率在不同植物和基因中存在差异,这种密码子的使用偏好性是物种和基因在长期进化及对环境的适应过程中形成的,是多种因素共同作用的结果^[7]。碱基组成对密码子使用偏性影响最普遍,而核苷酸随机突变和回复突变可影响碱基组成,而这些突变能否保留受多因素共同影响^[3-8],其中选择和碱基突变是影响密码子偏好性的主要因素^[8]。叶绿体是植物进行光合作用的场所,也含有相对独立的遗传信息,因此,通过叶绿体基因组可以揭示物种进化、鉴定不同物种间的亲缘关系等,还可以利用叶绿体基因工程安全、表达高效等特点进行基因研究^[20-21]。叶绿体基因组中的密码子使用偏性的研究能够为物种基因组进化关系、基因的异源表达等提供有效材料和基础^[4],GC₃ 是分析密码子偏好性的一个重要指标,因为密码子第 1 位和第 2 位碱基的改变通常会改变编码氨基酸,而第 3 位碱基的改变则对编码氨基酸无影响,因此可以认为,密码子偏好性在某种程度上是对偏好密码子使用和非偏好密码子保留间的一种平衡机制,是物种和基因进化过程中的一种自我保护机制;同时密码子第 3 位碱基具兼并性,面临较小的选择压力,且 GC₃ 含量与密码子使用偏性显著相关^[22-23]。本研究中云南蓝果树叶绿体基因密码子第 3 位碱基的 GC 含量为 28.40%,52 个基因中 ENC 值>45 的有 39 个,偏好以 AU 为结尾的密码子,其使用偏性较弱,且中性绘图分析显示该叶绿体基因组密码子的第 1 位和第 2 位碱基组成与第 3 位存在显著差异,且在变异过程无明显关联;与蒺藜苜蓿(*Medicago truncatula*)的 3 个位置 GC 含量仅在具体数据上存在差别,而趋势基本一致,即蒺藜苜蓿 GC₁(45.5%)>GC₂(36.8%)>GC₃(26.9%)^[1],与蝴蝶兰(*Phalaenopsis aphrodite* subsp. *formosana*)叶绿体基因组密码子受多重因素共同影响的研究结果基本一致^[24]。在密码子第 3 位碱基的使用频率上,云南蓝果树叶绿体基因组密码子以 AT 为主,GC 的使用频率仅占 28.40%,且在具体碱基的使用频率上 T 高于 A、G 高于 C,与蒺藜苜蓿叶绿体基因组密码子第 3 位碱基一致^[1],而与禾本科植物密码子第 3 位碱基 T 高于 A、G、C 使用频率相当的研究结果存在差异^[25]。这些现象说明在不同植物的叶绿体基因组中,其碱基组成具有一定的差异,进而影响密码子偏好性,且各植物中影响密码子偏好性的主要因素不同。同时,密码子的使用偏性与物种和亲缘关系也有一定的关系,且亲缘关系越近,密码子偏好性愈相似^[26];不同植物密码子偏好性均受多因素影响,而主要的影响因素则因植物而

异。在本研究中,云南蓝果树的最优密码子中含有 UUG,大多数高等植物的最优密码子为 NNA 或 NNU 模式^[27];而叶绿体基因组密码子偏好性较为保守,但云南蓝果树叶绿体最优密码子与其他植物存在一定的差异,可能与云南蓝果树适应狭窄的分布环境存在一定的相关性,同时最优密码子的确定能够为云南蓝果树物种进化特性分析、叶绿体基因工程和遗传多样性分析提供参考依据。

参考文献:

[1] 杨国锋,苏昆龙,赵怡然,等.蒺藜苜蓿叶绿体密码子偏好性分析[J].草业学报,2015,24(12):171-179.
YANG G F,SU K L,ZHAO Y R,*et al.* Analysis of codon usage in the chloroplast genome of *Medicago truncatula* [J]. Acta Prataculture Sinica,2015,24(12):171-179. (in Chinese)

[2] DURET L. Evolution of synonymous codon usage in metazoans [J]. Current Opinion in Genetics & Developpment, 2002, 12 (6):640-649.

[3] ROSENBERG M S,SANKAR S,SUDHIR K. Patterns of transi-tional mutation biases within and among mammalian genomes [J]. Molecular Biology & Evolution,2003,20(6):988-993.

[4] 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 Research,2000, 28(10):2084-2090.

[5] MORIYAMA E N,POWELL J R. Codon usage bias and tRNA abundance in *Drosophila* [J]. Journal of Molecular Evolution, 1997,45(5):514-523.

[6] GU W,ZHOU T,MA J,*et al.* The relationship between syn-onymous codon usage and protein structure in *Escherichia coli* and Homo sapiens[J]. Biosystems,2004,73(2):89-97.

[7] HIRAOKA Y,KAWAMATA K,HARAGUCHI T,*et al.* Co-don usage bias is correlated with gene expression levels in the fission yeast schizosaccharomyces pombe[J]. Genes to Cells, 2010,14(4):499-509.

[8] HERSHBERG R,PETROV D A. Selection on codon bias[J]. Annual Review of Genetics,2008,42:287-299.

[9] 王鹏良,杨利平,吴红英,等.普通油茶叶绿体基因组密码子偏好性分析[J].广西植物,2018,38(2):135-144.
WANG P L,YANG L P,WU H Y,*et al.* Codon preference of chloroplast genome in *Camellia oleifera* [J]. Guihaia,2018,38 (2):135-144. (in Chinese)

[10] 牛元,徐琼,王嵩德,等.大花香水月季叶绿体基因组密码子使用偏性分析[J].西北林学院学报,2018,33(3):123-130.
NIU Y,XU Q,WANG Y D,*et al.* An analysis on codon usage bias of chloroplast genome of *Rosa odorata* var. *gigantea* [J]. Journal of Northwest Forestry University,2018,33(3):123-130. (in Chinese)

[11] 吴妙丽,陈世品,陈辉.竹亚科叶绿体基因组的密码子使用偏性分析[J].森林与环境学报,2019,39(01):9-14.
WU M L,CHEN S P,CHEN H. Codon preference of chloro-plast genome of Bambusoideae[J]. Journal of Forest and En-vironment,2019,39(1):9-14. (in Chinese)

[8] 赵源吉,韦雪青,温俊宝,等. 双条杉天牛幼虫声音探测技术初报[J]. 生态科学, 2009,28(3):242-246.
ZHAO Y J, WEI X Q, WEN J B, *et al.* Preliminary study on the acoustic detection of larvae *Semanotus bifasciatus* (Motschulsky)[J]. Ecological Science, 2009,28(3):242-246. (in Chinese)

[9] 韦雪青,温俊宝,赵源吉,等. 害虫声音监测技术研究进展[J]. 林业科学, 2010,46(5):147-154.
WEI X Q, WEN J B, ZHAO Y J, *et al.* Review on monitoring technology of the insect acoustic[J]. Forestry Science, 2010,46(5):147-154. (in Chinese)

[10] 王鸿斌,赵丽稳,罗茜,等. 红脂大小蠹声音信号记录、分析与发声器官电镜观察[J]. 林业科学研究, 2012,25(1):17-22.
WANG H B, ZHAO L W, LUO Q, *et al.* Acoustic signals of *Dendroctonus valens* and structure of its stridulatory apparatus[J]. Forestry Science Research, 2012, 25(1): 17-22. (in Chinese)

[11] 许小芳,周红生,娄定风,等. 木材钻蛀性害虫活动声信号的采集与分析[C]. 西安:上海市声学学会声学学术会议, 2011.

[12] 潘杰,刘翔,邵沛泽,等. 缝锤小蠹与美西部云杉小蠹的声学特征与比较[J]. 环境昆虫学报, 2014,36(3):395-401.
PA J, LIU X, SHAO P Z, *et al.* Acoustic characteristics and their comparison of *Gnathotrichus materiarius* and *Gnathotrichus sulcatus* [J]. Journal of Environmental Entomology, 2014,36(3):395-401. (in Chinese)

[13] 鞠双,李新慈,罗廷芳,等. 马尾松胶合木表面声发射信号各向异性传播规律[J]. 林业工程学报, 2019,4(2):48-53.
JU S, LI X C, LUO T F, *et al.* Anisotropic propagation of acoustic emission signal on surface of *Pinus massoniana* Lamb. glulam[J]. Journal of Forestry Engineering, 2019, 4(2): 48-53. (in Chinese)

[14] 鞠双,李新慈,罗廷芳,等. 应用小波分析法对马尾松胶合木表面声发射信号特征检测[J]. 东北林业大学学报, 2018,46(8):86-92.
JU S, LI X C, LUO T F, *et al.* Detection of acoustic emission signal characteristics on masson pine plywood surface by wavelet analysis[J]. Journal of Northeast Forestry University, 2018,46(8):86-92. (in Chinese)

[15] 李新慈,鞠双,罗廷芳,等. 马尾松胶合木胶层对声发射信号传播特性的影响[J]. 西北林学院学报, 2019,34(3):185-190.
LI X C, JU S, LUO T F, *et al.* Effect of masson pine plywood layer on acoustic emission signal propagation characteristics [J]. Journal of Northwest Forestry University, 2019, 34(3): 185-190. (in Chinese)

[16] 于帅帅,申丽娟,李杨,等. 云南松表面声发射信号采集与特征分析[J]. 西北林学院学报, 2017,32(2):247-251.
YU S S, SHEN L J, LI Y, *et al.* Acquisition and characteristic auaglysis of the sarface of *Pinus yunnanensis* acoustic emission signal [J]. Journal of Northwest Forestry University, 2017,32(2):247-251. (in Chinese)

(上接第 31 页)

[12] 马明呈,张育浩,元王涛,等. 青海五裂茶藨子叶绿体 DNA psbA-trnH 序列的遗传变异研究[J]. 西北林学院学报, 2015,30(4):116-120.
MA M C, ZHANG Y H, YUAN W T, *et al.* Genetic differentiation study of chloroplast DNA psbA-trnH in *Ribes meyeri* from Qinghai[J]. Journal of Northwest University, 2018, 30(4):116-120. (in Chinese)

[13] 傅立国. 中国植物红皮书[M]. 北京:科学出版社, 1991:480.

[14] 陈伟,史富强,杨文忠,等. 云南蓝果树的种群状况及生态习性[J]. 东北林业大学学报, 2011,39(9):17-19.
CHEN W, SHI F Q, YANG W Z, *et al.* Population status and ecological characteristics of *Nyssa yunnanensis* [J]. Journal of Northeast Forestry University, 2011, 39(9): 17-19. (in Chinese)

[15] 吴娅萍,侯昭强,陈中华,等. 云南蓝果树的种群资源及分布现状[J]. 西部林业科学, 2015,44(6):26-30.

[16] 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.

[17] WRIGHT F. The ‘effective number of codons’ used in a gene [J]. Gene, 1990,87(1):23-29.

[18] SUEOKA N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses[J]. Journal of Molecular Evolution, 2001,53(4/5):469-476.

[19] 秦政,郑永杰,桂丽静,等. 樟树叶叶绿体基因组密码子偏好性分析[J]. 广西植物, 2018,38(10):1346-1355.

[20] DANIELL H, KHAN M S, ALLISON L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology[J]. Trends in Plant Science, 2002,7(2):84-91.

[21] DURET L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes[J]. Trends in Genetics, 2000, 16(7): 287-289.

[22] MORTON B R. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA [J]. Journal of Molecular Evolution, 2003, 56(5):616-629.

[23] GU W, ZHOU T, MA J, *et al.* Analysis of synonymous codon usage in SARS *Coronavirus* and other viruses in the *Nidovirales* [J]. Virus Research, 2004,101(2):155-161.

[24] 续晨,贲爱玲,蔡晓宁. 蝴蝶兰叶绿体基因组密码子使用的相关分析[J]. 分子植物育种, 2010,8(5):945-950.
XU C, BEN A L, CAI X N. Analysis of synonymous codon usage in chloroplast genome of *Phalaenopsis aphrodite* subsp. *formosana* [J]. Molecular Plant Breeding, 2010, 8(5): 945-950. (in Chinese)

[25] 张月荣. 禾本科叶绿体基因组密码子使用模式及紫茎泽兰叶绿体 RNA 编辑分析[D]. 杨陵:西北农林科技大学, 2013.

[26] LIU S, XUE D, CHENG R, *et al.* The complete mitogenome of *Apocheima cinerarius* (Lepidoptera: Geometridae: Ennominae) and comparison with that of other lepidopteran insects [J]. Gene, 2014,547(1):136-144.

[27] 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.