

须弥葛叶绿体基因组密码子使用偏好性分析

万美暄¹, 黄显俊², 李雪¹, 徐立^{1,2*}

(1. 大理大学基础医学院, 云南大理 671000; 2. 大理大学药学院, 云南省高校滇西道地药材资源开发重点实验室, 云南大理 671000)

摘要 [目的]揭示须弥葛叶绿体基因组中影响其密码子使用偏好和模式的因素。[方法]利用 CUSP 在线程序、Codon W 1.4.2、Microsoft Excel、SPSS 24.0 等软件,计算并分析须弥葛叶绿体基因组中 52 条蛋白质编码基因序列(coding sequences, CDS)的密码子使用偏好性指标和碱基组成。[结果]须弥葛密码子更偏好 A/U 结尾的密码子。52 条 CDS 的有效密码子数(effective number of codon, ENC)以及各项偏性指标分析结果均提示密码子偏好性较弱。相关性分析显示,ENC 与 GC₁、GC₂ 的相关性均不显著,但与 GC₃ 相关性极显著,表明须弥葛密码子第 3 位碱基组成对其密码子偏好性影响较大。通过 ENC-plot、中性绘图、PR2-plot 分析发现,须弥葛叶绿体基因组同义密码子使用偏好性的主要影响因素是自然选择;最后筛选出 10 个最优密码子 UUG、CUU、UCC、UCA、CCA、ACA、GCA、UAU、CAU、AGA。[结论]该结果可为须弥葛叶绿体基因组密码子优化、遗传修饰及系统进化的相关研究提供理论依据。

关键词 须弥葛;密码子偏好;最优密码子;自然选择

中图分类号 Q943 文献标识码 A

文章编号 0517-6611(2022)21-0102-06

doi:10.3969/j.issn.0517-6611.2022.21.024



开放科学(资源服务)标识码(OSID):

Analysis of Codon Bias in the Chloroplast Genome of *Haymondia wallihii*WAN Mei-xuan¹, HUANG Xian-jun², LI Xue¹ et al (1. College of Basic Medicine, Dali University, Dali, Yunnan 671000; 2. College of Pharmaceutical Science, Dali University, Key Laboratory of Yunnan Provincial Higher Education Institutions for Development of Yunnan Dao-di Medicinal Materials Resources, Dali, Yunnan 671000)

Abstract [Objective] To reveal the factors that affect the preference and usage patterns of synonymous codons in the *Haymondia wallihii* chloroplast genome. [Method] The codon usage bias indexes and base composition of 52 coding sequences (CDS) were calculated and analyzed using the CUSP online program, Codon W 1.4.2, Microsoft Excel, and SPSS 24.0. [Result] The results showed that synonymous codons had a greater preference for using codons ending with A/T. The effective number of codon (ENC) values of the 52 CDS and each bias indicator analysis indicated a weak codon preference. The correlation analysis showed that ENC was strongly correlated with GC₃, but neither were correlated with GC₁ and GC₂. GC₃ could affect codon usage bias. ENC-plot, Neutrality plot, PR2-plot revealed that natural selection might play an important role in the synonymous codon usage bias of the *Haymondia wallihii* chloroplast genome. After filtering, ten optimal codons UUG, CUU, UCC, UCA, CCA, ACA, GCA, UAU, CAU and AGA were obtained. [Conclusion] The results of this study can provide a theoretical basis for the related research on codon optimization, genetic modification and phylogenetic evolution of the *Haymondia wallihii* chloroplast genome.

Key words *Haymondia wallihii*; Codon preference; Optimal codon; Natural selection

叶绿体(chloroplast, Cp)是半自主性细胞器,在植物光合作用、代谢反应、脂肪酸、氨基酸、维生素等生物合成中至关重要^[1]。叶绿体 DNA 一般以双链环状存在于高等植物叶绿体中,长度在 120~180 kb,具有典型的四分体结构^[2],且由于其母系遗传和适度的进化速度,在序列和结构上高度保守^[3]。因此,Cp 基因组测序备受关注,NCBI 数据库也提供了越来越多的叶绿体基因组,这为种族遗传学和系统发育方面的研究提供了方便。

密码子(condon)是信使 RNA(mRNA)翻译的基本信息单位,61 个密码子编码 20 种氨基酸,除色氨酸和蛋氨酸,其他氨基酸均有 1 种以上的密码子编码^[4]。然而,在蛋白质合成中同义密码子(Synonymous codon)不一定以相同的频率被使用,这种特定密码子使用频率高于其他同义密码子的现象被称为“同义密码子使用偏好性(codon usage bias, CUB)”^[5-6]。CUB 广泛存在于生物体中,反映了不同物种或品种之间基因的起源、进化趋势和突变模式^[7]。有研究表

明,CUB 与基因表达水平之间存在相关性^[8]。此外,密码子使用模式可能会影响 mRNA 生物合成、蛋白质折叠、翻译延伸率及其他下游表达的生物学功能^[9]。因此,对 CUB 分析可以更好地了解物种的遗传规律、蛋白质的机制及预测新基因。

须弥葛(*Haymondia wallihii*),俗称“马鹿花”,为豆科蝶形花亚科(Papilionaceae)须弥葛属(*Haymondia*)植物^[10]。在我国主要分布于广州、四川、广西、福建、云南等省区及云南省内的思茅、红河、临沧、保山、德宏、大理、楚雄、怒江等州市^[11]。须弥葛花较大,生长速度快,可做观赏绿化,在大理白族自治州常被采集花朵,腌制野菜^[12]。有研究发现,须弥葛中含有大豆素、大豆苷、葛根素、豆甾醇等化合物^[12],具有抗心血管疾病的生理活性,还能缓解高血压患者的头痛,解决平滑肌痉挛,镇咳及抑制癌细胞生长等药性^[13]。笔者基于须弥葛植物叶绿体基因组高通量测序结果,挑选出长度>300 bp 的 52 条蛋白编码序列,对其进行整合分析,探讨须弥葛密码子使用偏性的影响因素,并最终确定最优密码子,以为须弥葛的进化关系研究、品种改良及叶绿体基因组的应用提供参考。

1 材料与方法

1.1 试验材料的获取 须弥葛植株的新鲜幼嫩叶片于 2019 年 12 月采自云南省丽江永胜县(25°53'73.23"E,

基金项目 云南省科学研究基金项目(2022Y814);云南省地方本科高校(部分)联合专项[2019FH001(-040)]。**作者简介** 万美暄(1994—),女,宁夏石嘴山人,硕士,从事葛属植物分类及分子进化与适应研究。*通信作者,教授,博士,从事葛属植物分类及分子进化与适应研究。**收稿日期** 2022-06-06

100°25'48.40"N), 经云南省高校滇西道地药材资源开发重点实验室的夏从龙教授鉴定为豆科须弥葛属植物须弥葛 (*Haymndia wallichii*)。用无菌水将采集到的健康须弥葛叶片冲洗多次, 装入有变色硅胶的自封袋中脱水干燥, 委托上海生物科技有限公司的 Illumina No-vaSeq 6000 平台测序, 最后通过组装和注释获得须弥葛叶绿体全基因组序列号 (NCBI 数据库, 登录号为 NC052910.1)。

1.2 叶绿体基因组蛋白质编码序列的获取 从 NCBI 数据库下载须弥葛叶绿体基因组蛋白质编码序列 (CDS) 后, 为加强数据可靠性及减少数据冗杂, 剔除重复基因和长度 <300 bp 的基因^[14]。最后将符合条件的 52 条 CDS 整合成一个 fasta 文件, 用于后续分析。

1.3 密码子使用偏好性分析 将整理的 fasta 文件通过 CodonW1.4.2 软件处理^[15], 获取到须弥葛叶绿体基因组蛋白质编码基因序列的相对同义密码子使用度 (relative synonymous codon usage, RSCU)、编码氨基酸的密码子数量、有效密码子数 (effective number of codon, ENC)、最优密码子使用频率 (frequency of optical codons, Fop)、密码子适应指数 (codon adaptation index, CAI)、密码子偏性指数 (codon bias index, CBI)。然后, 将 52 条 CDS 依次提取整理成 52 个 fasta 文件, 使用在线程序 CUSP^[16] 计算每个基因不同位置的密码子 GC 含量 (GC_{all}, GC₁, GC₂, GC₃), 将所得数据使用 SPSS 24.0 和 Microsoft Excel 软件进行统计及 Person 相关性分析。

1.4 PR2 plot 绘图分析 PR2 plot 图是以 C=G 和 A=U 为矢量发出中心点, 用 $A_3/(A_3+T_3)$ 为纵坐标, $G_3/(G_3+C_3)$ 为横坐标绘制而成, 表明影响密码子偏好性的主要因素^[17]。

1.5 中性绘图分析 将 GC₃ 作为横坐标, GG₁ 和 GC₂ 的平均值 (GC₁₂) 为纵坐标, 作二维散点图, 再次绘制回归曲线, 然后对 GC₃ 和 GC₁₂ 进行相关性分析。

1.6 ENC-plot 绘图分析 首先利用 Excel 绘制 GC₃ 和 ENC 的二维散点图, 并根据公式 $ENC = 2 + GC_3 + 29/[GC_{32} + (1 - GC_3)2]$ 绘制标准曲线, 以此判断基因点在图中分布情况。然后利用 ENC 公式求得 ENC 期望, 再通过 $ENC \text{ 比值} = (ENC \text{ 期望} - ENC \text{ 实际}) / ENC \text{ 期望}$, 算出 ENC 比值频数分布, 从而确定影响密码子偏好性的主要因素。

1.7 最优密码子的筛选 以 52 条 CDS 分别得到的 ENC 值为基础, 用 Excel 将 ENC 值依次排序, 选出 ENC 值最高的 5 个基因和最低的 5 个基因建立高低表达基因库, 然后将这些基因序列整理成不同的 fasta 文件, 运行 Codon W 1.4.2 软件得出 2 个库中编码各氨基酸密码子的 RSCU 值, 计算 $\Delta RSCU$ 值 (RSCU 高表达, -RSCU 低表达)。以高表达组 RSCU > 1, 且 $\Delta RSCU \geq 0.08$ 的密码子为最优密码子^[18]。

1.8 统计分析 使用 SPSS 24.0 和 Excel 软件, 进行 Pearson 相关分析, 显著性水平为 $P < 0.05$ 或 $P < 0.01$ 。

2 结果与分析

2.1 须弥葛叶绿体基因组密码子偏性分析

2.1.1 相对同义密码子使用度。 须弥葛叶绿体基因组中的 52 条 CDS 序列的相对同义密码子使用度 (relative synony-

mous codon usage, RSCU) 结果显示 (表 1), 31 个高频密码子 (RSCU > 1) 中以 U 结尾的有 16 个, 以 A 结尾的有 14 个, 以 G 结尾的有 1 个, 表明须弥葛叶绿体基因组密码子更倾向 A/U 结尾。所有高频密码子 RSCU 值均在 1~2, 除亮氨酸 (Leu) 密码子 UUA > 2, 提示须弥葛叶绿体基因组密码子使用偏好性较弱。

表 1 须弥葛叶绿体基因组中各氨基酸密码子的 RSCU 值

Table 1 RSCU analysis of each amino acid codon in the chloroplast genome of *Haymndia wallichii*

氨基酸 Amino acid	密码子 Codon	RSCU 值 RSCU value	氨基酸 Amino acid	密码子 Codon	RSCU 值 RSCU value	
Phe	UUU *	1.34	Thr	ACU *	1.62	
	UUC	0.66		ACC	0.71	
Ser	UCU *	1.67	Asn	ACA *	1.25	
	UCC	0.94		ACG	0.41	
	UCA *	1.28	AAU *	1.59		
	AGU *	1.32	AAC	0.41		
Leu	AGC	0.28	Lys	AAA *	1.49	
	UCG	0.52		AAG	0.51	
	UUA *	2.11	Val	GUU *	1.49	
	UUG *	1.14		GUC	0.47	
Pro	CUU *	1.26	Ala	GUA *	1.50	
	CUC	0.37		GUG	0.54	
	CUA	0.78	Glu	GCU *	1.75	
	CUG	0.34		GCC	0.62	
	CCU *	1.53	Asp	GCA *	1.18	
	CCC	0.87		GCG	0.45	
	Tyr	CCA *	1.15	Gly	GAU *	1.60
		CCG	0.45		GAC	0.40
UAU *		1.65	Gln	GAA *	1.51	
UAC		0.35		GAG	0.49	
Cys	UGU *	1.48	His	GGU *	1.28	
	UGC	0.52		GGC	0.45	
	Arg	CGU *	1.40	Gln	GGA *	1.59
		CGC	0.38		GGG	0.69
CGA *		1.31	TER	CAU *	1.59	
AGA *		1.89		CAC	0.41	
Ile	AGG	0.52	Gln	CAA *	1.51	
	CGG	0.50		CAG	0.49	
	AUU *	1.43		UAA *	1.59	
	AUC	0.57	UAG	0.71		
	AUA *	1.01	UGA	0.71		

注: * 表示 RSCU > 1 的高频密码子

Note: * indicates codons usage with high frequency and the value of RSCU > 1

2.1.2 基因组密码子碱基组成及偏性。 利用软件 CodonW1.4.2 和 CUSP 分析符合条件的 52 条 CDS 序列 (表 2), 结果表明, 52 条 CDS 序列的总 GC 含量 (GC_{all}) 为 38.10% (范围为 29.27%~46.56%), 第 1 位 (GC₁) 为 46.89% (范围为 31.99%~59.31%), 第 2 位 (GC₂) 为 39.48% (范围为 27.62%~54.68%), 第 3 位 (GC₃) 为 27.87% 范围 (为 21.15%~36.88%)。GC₂ 与 GC_{all} 相近, GC₁ 和 GC₃ 与 GC_{all} 相差较大, 表现为 GC₁ > GC₂ > GC₃。进一步说明须弥葛叶绿体基因组密码子偏好 A/U 结尾的密码子。

表 2 密码子的碱基组成及偏性相关指标

Table 2 Condon base composition and bias-related indicators

基因 Gene	GC _{all} //%	GC ₁ //%	GC ₂ //%	GC ₃ //%	ENC	CAI	CBI	FOP
<i>accD</i>	32.99	36.89	35.66	26.43	43.96	0.181	-0.212	0.320
<i>atpA</i>	39.96	56.69	39.57	23.62	44.76	0.210	-0.028	0.398
<i>atpB</i>	42.08	55.71	41.28	29.26	47.41	0.186	-0.049	0.381
<i>atpE</i>	38.97	50.00	38.24	28.68	45.83	0.179	-0.032	0.389
<i>atpF</i>	36.22	45.41	34.05	29.19	40.36	0.127	-0.193	0.300
<i>atpI</i>	37.10	49.19	36.69	25.40	46.20	0.164	-0.082	0.345
<i>ccsA</i>	31.88	31.99	38.51	25.16	46.57	0.145	-0.185	0.303
<i>clpP</i>	41.83	59.31	35.29	30.88	58.63	0.194	-0.032	0.387
<i>matK</i>	30.47	38.40	29.63	23.39	46.47	0.162	-0.176	0.310
<i>ndhA</i>	34.16	42.31	39.01	21.15	43.83	0.120	-0.151	0.298
<i>ndhB</i>	37.51	42.27	39.53	30.72	46.71	0.163	-0.104	0.344
<i>ndhC</i>	37.19	47.93	34.71	28.93	47.73	0.188	-0.142	0.306
<i>ndhD</i>	35.40	39.92	37.52	28.74	46.89	0.130	-0.152	0.305
<i>ndhE</i>	31.05	39.22	30.39	23.53	43.74	0.166	-0.160	0.309
<i>ndhF</i>	31.76	34.86	36.47	23.96	44.22	0.129	-0.210	0.283
<i>ndhG</i>	33.52	43.82	34.83	21.91	44.52	0.136	-0.197	0.269
<i>ndhH</i>	38.66	50.51	36.29	29.91	49.35	0.141	-0.120	0.332
<i>ndhI</i>	33.70	40.56	36.11	24.44	46.29	0.192	-0.127	0.349
<i>ndhJ</i>	38.16	47.17	37.11	30.19	49.55	0.151	-0.155	0.322
<i>ndhK</i>	37.42	42.15	42.15	27.97	50.44	0.170	-0.150	0.332
<i>petA</i>	39.15	53.89	35.51	28.04	49.94	0.169	-0.128	0.328
<i>petB</i>	40.12	48.61	41.20	30.56	44.98	0.208	-0.064	0.360
<i>petD</i>	37.55	50.00	37.35	25.30	40.52	0.155	-0.124	0.304
<i>psaA</i>	42.61	52.06	43.14	32.62	51.22	0.190	-0.115	0.348
<i>psaB</i>	41.22	48.30	42.86	32.52	50.03	0.177	-0.122	0.346
<i>psbA</i>	41.24	48.59	43.50	31.64	42.05	0.294	0.175	0.520
<i>psbB</i>	44.07	53.63	46.56	32.02	48.11	0.185	-0.046	0.389
<i>psbC</i>	44.84	53.60	46.19	34.75	47.43	0.175	-0.045	0.381
<i>psbD</i>	42.94	51.41	43.50	33.90	45.06	0.242	0.047	0.444
<i>rbcL</i>	43.31	57.38	43.04	29.52	48.97	0.267	0.053	0.456
<i>rpl14</i>	39.30	56.91	37.40	23.58	48.52	0.166	-0.013	0.390
<i>rpl16</i>	45.99	54.74	53.28	29.93	38.66	0.142	-0.035	0.397
<i>rpl2</i>	44.16	51.82	50.00	30.66	52.70	0.166	-0.090	0.362
<i>rpl20</i>	36.72	38.98	43.22	27.97	45.57	0.110	-0.189	0.301
<i>rpl22</i>	34.10	40.46	40.46	21.37	45.98	0.157	-0.033	0.395
<i>rpoA</i>	35.32	44.64	34.52	26.79	50.72	0.156	-0.135	0.333
<i>rpoB</i>	38.60	50.23	37.98	27.60	48.69	0.152	-0.132	0.336
<i>rpoC1</i>	39.24	49.64	38.03	30.04	50.68	0.150	-0.133	0.330
<i>rpoC2</i>	36.19	44.38	36.62	27.56	48.97	0.144	-0.170	0.314
<i>rps11</i>	46.56	56.83	54.68	25.18	49.78	0.151	-0.070	0.368
<i>rps12</i>	41.67	51.61	46.77	26.61	42.97	0.142	-0.082	0.352
<i>rps14</i>	40.92	47.52	44.55	30.69	49.88	0.132	-0.083	0.354
<i>rps18</i>	35.62	38.24	42.16	26.47	44.17	0.114	-0.203	0.283
<i>rps2</i>	37.69	41.77	40.93	30.38	50.56	0.181	-0.120	0.348
<i>rps3</i>	34.09	47.03	33.33	21.92	45.07	0.158	-0.147	0.340
<i>rps4</i>	37.13	48.02	37.62	25.74	49.27	0.149	-0.056	0.365
<i>rps7</i>	40.60	52.56	45.51	23.72	48.31	0.183	-0.079	0.373
<i>rps8</i>	34.84	42.11	38.35	24.06	49.63	0.126	-0.052	0.378
<i>ycf1</i>	29.27	35.29	27.62	24.90	47.35	0.166	-0.153	0.341
<i>ycf2</i>	37.18	40.72	33.94	36.88	52.95	0.159	-0.144	0.336
<i>ycf3</i>	39.57	47.37	39.18	32.16	59.33	0.158	-0.179	0.335
<i>ycf4</i>	39.10	45.41	41.08	30.81	50.51	0.156	-0.103	0.345
平均 Mean	38.10	46.89	39.48	27.87	47.54	0.166	-0.104	0.349

注:密码子的第1、2、3位及总GC含量:GC₁、GC₂、GC₃和GC_{all};有效密码子数:ENC;密码子适应指数:CAI;密码子偏性指数:CBI;最优密码子使用频率:FOP

Note:GC₁,GC₂,GC₃ and GC_{all} indicate the first, second,third positions and total GC content;Number of effective condond;ENC;Codon adaptation index;CAI;Codon bias index;CBI;Optimal codon usage frequency:FOP

52条CDS的CAI(codon adaptation index)为0.166(范围-0.212~0.175),FOP(frequency of optical codons)为0.349(范围0.110~0.294),CBI(codon bias index)为-0.104(范围0.269~0.520),ENC为47.54(范围38.66~

59.33)。以上指标进一步提示须弥葛叶绿体基因组密码子的使用偏性和基因表达均较弱。

ENC 值与各位置 GC 含量的相关性分析显示(表 3), GC_{all} 与 GC_1 、 GC_2 和 GC_3 含量的相关性均达到极显著相关水平($P < 0.01$); GC_1 与 GC_2 含量的相关性达到极显著水平($P < 0.01$); GC_3 与 GC_1 及 GC_2 含量的相关性达到显著水平($P < 0.05$); ENC 与 GC_1 、 GC_2 含量的相关性均不显著($P > 0.05$), 但与 GC_3 含量的相关性极显著($P < 0.01$), 说明第 3 位 GC 含量对密码子的使用偏性有一定程度的影响。

表 3 各基因参数的相关性分析

Table 3 Correlation analysis of relative indexes of gene

指标 Index	GC_1	GC_2	GC_3	GC_{all}
GC_2	0.473**			
GC_3	0.296*	0.316*		
GC_{all}	0.845**	0.805**	0.595**	
ENC	0.201	-0.024	0.378**	0.214

注: * 表示在 0.05 水平(双侧)上显著相关; ** 表示在 0.01 水平(双侧)上显著相关

Note: *. Significant correlation at the 0.05 level (both side); **. Significant correlation at the 0.01 level (both side)

2.2 影响密码子偏性的因素分析

2.2.1 PR2-plot 分析.

PR2 绘图分析(图 1), 所有点分布在 $G_3/(G_3+C_3)$ 上的 0.26~0.67 和 $A_3/(A_3+T_3)$ 上的 0.38~0.57, 但各基因在 4 个象限中明显分布不均匀。若密码子使用偏好完全受碱基突变的影响, 大多数点应位于 0.5 水平中心线或附近, 但该结果却不同。因此, 就碱基频率而言, $G > C, T(U) > A$ 。由此推测, 自然选择和突变都是影响须弥葛叶绿体基因组密码子使用偏性的因素。

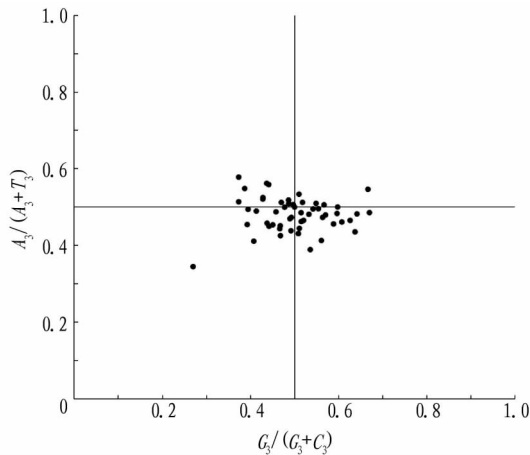


图 1 须弥葛叶绿体基因组 PR2-plot 分析

Fig. 1 PR2-plot analysis of the *Haymondia wallihii* genome

2.2.2 中性绘图分析.

以中性绘图分析 GC_{12} 和 GC_3 的关系(图 2), GC_{12} 的取值 0.340 1~0.557 5, GC_3 的取值 0.211 5~0.368 8, 所有基因位于对角线以上(除 *Ycf2* 基因)。Pearson 相关分析发现, GC_{12} 和 GC_3 之间的相关性较弱($R^2 = 0.125 8, r = 0.355 0$), 相关系数 r 值距离 1 较远。该结果提示自然选择是须弥葛叶绿体基因组密码子使用偏性的主要影响因素。

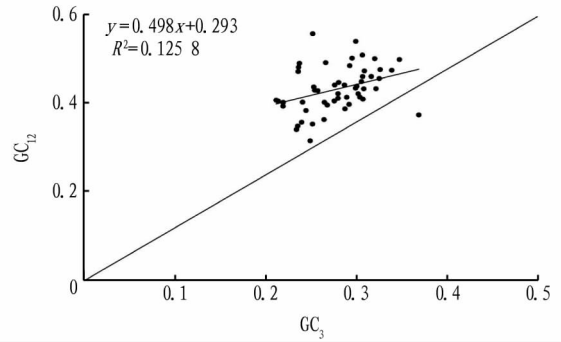


图 2 须弥葛叶绿体基因组中性绘图

Fig. 2 Neutrality plot analysis of the chloroplast genome of *Puerariawallichii*

2.2.3 ENC-plot 分析.

ENC-plot 绘图分析(图 3), 大部分基因散乱地分布在标准曲线下方, 只有少数基因分布在曲线附近或曲线上。另外, 52 条 CDS 的 ENC 比值在 -0.1~0.3 (表 4), 其中, 25 个基因的 ENC 值和预期接近, 分布在 -0.050~0.050; 而其他 27 个基因的 ENC 值不在该范围内。若碱基突变是影响密码子的主要因素, 大部分基因应在曲线上方或曲线附近, 且 ENC 比值在 -0.05~0.05, 但结果却相反。提示碱基突变虽是影响须弥葛密码子偏好性的因素, 但受自然选择更多, 这与中性绘图分析结果相一致。

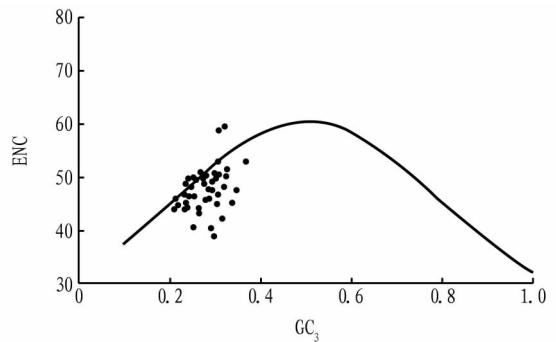


图 3 须弥葛叶绿体基因组 ENC-plot 分析

Fig. 3 ENC-plot analysis of the chloroplast genome of *Puerariawallichii*

表 4 ENC 比值频率分布

Table 4 Distribution of ENC frequency ratio

组限 Class boundary	组中值 Class midvalue	频数 Frequency number	频率 Frequency
-0.150~-0.050	-0.1	2	0.038
-0.050~0.050	0	25	0.481
0.050~0.150	0.1	20	0.385
0.150~0.250	0.2	4	0.077
0.250~0.350	0.3	1	0.019
合计 Total	—	52	1

2.3 最优密码子确定

该研究将 ENC 值最高的 5 个基因(*ycf3*、*clpP*、*ycf2*、*rpl2*、*rpoA*)和 ENC 值最低的 5 个基因(*rpl6*、*aptF*、*petD*、*psbA*、*rps12*)分别作为高低基因表达组, 建立高低表达基因库(表 5), 计算得到 $\Delta RSCU$ 。 $RSCU > 1$ 为高频密码子, $\Delta RSCU \geq 0.08$ 的密码子为高表达优越密码子, 选取二者

共有的密码子为最优密码子,据此符合条件的有 10 个密码子,其中 8 个以 A/U 结尾。

表 5 须弥葛叶绿体基因组最优密码子分析

Table 5 Optimal codon analysis in the chloroplast genome of *Haymondia wallihii*

氨基酸 Amino acid	密码子 Codon	RSCU		Δ RSCU	氨基酸 Amino acid	密码子 Codon	RSCU		Δ RSCU
		高表达基因 Highly expressed genes	低表达基因 Lowly expressed genes				高表达基因 Highly expressed genes	低表达基因 Lowly expressed genes	
Phe	UUU	1.04	1.00	0.04	Ala	GCU	1.52	2.11	-0.59
	UUC	0.96	1.00	-0.04		GCC	0.73	0.63	0.10
Leu	UUA	1.32	2.37	-1.05	Tyr	GCA*	1.26	0.84	0.42
	UUG*	1.37	1.25	0.12		GCG	0.49	0.42	0.07
	CUU*	1.49	1.19	0.30		UAU*	1.54	1.38	0.16
	CUC	0.48	0.26	0.22		UAC	0.46	0.62	-0.16
	CUA	0.81	0.92	-0.11		CAU*	1.50	1.41	0.09
Ile	CUG	0.53	0	0.53	His	CAC	0.50	0.59	-0.09
	AUU	1.31	1.58	-0.27	Gln	CAA	1.30	1.45	-0.15
	AUC	0.71	0.59	0.12		CAG	0.70	0.55	0.15
Val	AUA	0.99	0.83	0.16	Asn	AAU	1.50	1.45	0.05
	GUU	1.41	1.41	0	AAC	0.50	0.55	-0.05	
	GUC	0.59	0.29	0.30	Lys	AAA	1.26	1.62	-0.36
	GUA	1.44	1.88	-0.44	AAG	0.74	0.38	0.36	
Ser	GUG	0.56	0.41	0.15	Asp	GAU	1.65	1.73	-0.08
	UCU	1.51	2.31	-0.80		GAC	0.35	0.27	0.08
	UCC*	1.13	0.92	0.21	Glu	GAA	1.32	1.48	-0.16
	UCA*	1.26	0.58	0.68		GAG	0.68	0.52	0.16
	UCG	0.64	0.12	0.52	Cys	UGU	1.35	1.75	-0.40
	AGU	1.19	1.73	-0.54		UGC	0.65	0.25	0.40
Pro	AGC	0.26	0.35	-0.09	Arg	CGU	0.90	2.29	-1.39
	CCU	1.50	1.49	0.01		CGC	0.38	0.35	0.03
	CCC	0.76	1.18	-0.42	CGA	1.22	1.24	-0.02	
	CCA*	1.29	1.02	0.27	CGG	0.68	0.18	0.50	
	CCG	0.45	0.31	0.14	AGA*	2.04	1.59	0.45	
Thr	ACU	1.38	1.75	-0.37	AGG	0.79	0.35	0.44	
	ACC	0.77	1.02	-0.25	Gly	GGU	1.06	1.54	-0.48
	ACA*	1.21	1.02	0.19		GGC	0.59	0.58	0.01
	ACG	0.64	0.22	0.42					

注: * 表示最优密码子

Note: * indicated the optimal codons

3 讨论

密码子使用偏好广泛存在于多种生物体内,可能受 GC 含量、基因长度、基因功能、选择、突变等多种因素影响,是一种复杂而又不可避免的现象^[19]。以往大多利用全基因组信息研究物种的密码子偏好^[20]。随着近年来新一代测序技术的飞速发展,出现了大量的基因组和转录组数据,Machado 等^[21]通过使用深度基因组群体测序推断黑腹果蝇同义位点的选择。Chu 等^[22]利用 Ribo-seq 和 RNA-seq 方法研究了密码子使用偏好如何影响拟南芥的翻译模式。Guan 等^[23]通过对非牛蛭密码子使用模式的分析,发现基因进化的基本要素是突变压力和选择。该研究通过 PR2 绘图、中性绘图、ENC-plot 等对影响须弥葛密码子偏好性的因素进行了分析整理,并最终确定最优密码子。

由于密码子的简并性,氨基酸的类型虽不会因 GC₃ 同义突变所改变,但其仍被认为是影响氨基酸类型的一个重要因素^[24]。该研究发现,须弥葛叶绿体基因组总 GC 含量(GC_{all})是 38.10%,且 GC₁>GC₂>GC₃,表明密码子不同位置 GC 含量不同,密码子偏好使用以 A/U 结尾的密码子,这与文心花、中华绒螯蟹、豆蔻等植物偏好性一致^[16,25-26]。RSCU 是确定密码子使用偏好性的主要指标,若 RSCU<1,则表明该密码子

使用频率偏低;RSCU>1,则说明该密码子使用频率偏高;当 RSCU=1 时,该密码子使用无偏好性^[27]。该研究发现,须弥葛叶绿体基因组高频密码子中 RSCU>1 的有 18 个,大部分是以 U 和 A 结尾,G 和 C 结尾的仅 2 个。最终确定最优密码子是 10 个,且 8 个是以 A/U 结尾,这与 GC 含量分析结果相同。

ENC 通常表示密码子偏性的强弱,其 ENC>45 表示密码子偏好性较弱^[28]。须弥葛 ENC 平均值为 47.54,则说明密码子使用偏好性较弱。PR2-plot、中性绘图、ENC-plot 分析结果显示,自然选择和突变在须弥葛叶绿体密码子偏好形成中都起作用,但前者作用更大。这与“玉铃铛”枣、翠雀叶绿体基因组中的密码子偏性影响结果一致^[29],但与蒙古韭不同^[30]。由此可知,造成植物密码子偏好的因素存在差异,不同因素在不同物种间发挥作用也不同。

该研究通过生物信息学的方法介绍了影响须弥葛叶绿体基因组密码子使用偏好和模式的因素,并筛选出最优密码子,这对须弥葛的进一步开发运用有着重要意义。

参考文献

- [1] DANIELL H, LIN C S, YU M, et al. Chloroplast genomes; Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17 (1): 1-29.

- [2] PARK I, KIM W J, YEO S M, et al. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species [J]. *Molecules*, 2017, 22(6): 1-15.
- [3] SUN J, SUN R, LIU H B, et al. Complete chloroplast genome sequencing of ten wild *Fragaria* species in China provides evidence for phylogenetic evolution of *Fragaria* [J]. *Genomics*, 2021, 113(3): 1170-1179.
- [4] FRUMKIN I, LAJOIE M J, GREGG C J, et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency [J]. *Proc Natl Acad Sci USA*, 2018, 115(21): E4940-E4949.
- [5] LONG S Y, YAO H P, WU Q, et al. Analysis of compositional bias and codon usage pattern of the coding sequence in Banna virus genome [J]. *Virus Res*, 2018, 258: 68-72.
- [6] WANG H J, MENG T, WEI W Q. Analysis of synonymous codon usage bias in helicase gene from *Autographa californica* multiple nucleopolyhedrovirus [J]. *Genes Genomics*, 2018, 40(7): 767-780.
- [7] HASSAN S, MAHALINGAM V, KUMAR V. Synonymous codon usage analysis of thirty two mycobacteriophage genomes [J]. *Adv Bioinformatics*, 2009, 2009: 1-12.
- [8] YI S K, LI Y H, WANG W M. Selection shapes the patterns of codon usage in three closely related species of genus *Misgurnus* [J]. *Genomics*, 2018, 110(2): 134-142.
- [9] HUANG X, XU J, CHEN L, et al. Analysis of transcriptome data reveals multifactor constraint on codon usage in *Taenia multiceps* [J]. *BMC Genomics*, 2017, 18(1): 1-13.
- [10] 谷勇, 潘学正. 瓦氏葛藤造林技术 [J]. *云南林业科技*, 1999, 28(1): 80-83.
- [11] 谷勇, 邹恒芳, 孙汝林, 等. 马鹿花受冻害与恢复能力的研究 [J]. *林业科学研究*, 2001, 14(3): 328-331.
- [12] 张建云, 安银岭, 李正红, 等. 马鹿花化学成分研究初探 [J]. *林产化学与工业*, 2001, 21(2): 67-70.
- [13] 张晓娟, 周海纯. 葛根化学成分, 现代药理及临床应用研究进展 [J]. *中医药信息*, 2017, 34(1): 124-126.
- [14] 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.
- [15] WRIGHT F. The 'effective number of codons' used in a gene [J]. *Gene*, 1990, 87(1): 23-29.
- [16] 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.
- [17] SUEOKA N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses [J]. *J Mol Evol*, 2001, 53(4/5): 469-476.
- [18] ZHANG H, LI J, WANG R X, et al. Comparative analysis of expansin gene codon usage patterns among eight plant species [J]. *J Biomol Struct Dyn*, 2019, 37(4): 910-917.
- [19] WEI L, HE J, JIA X, et al. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution [J]. *BMC Evol Biol*, 2014, 14: 1-12.
- [20] TROTTE E. Selection on codon bias in yeast: A transcriptional hypothesis [J]. *Nucleic Acids Res*, 2013, 41(20): 9382-9395.
- [21] MACHADO H E, LAWRIE D S, PETROV D A. Pervasive strong selection at the level of codon usage bias in *Drosophila melanogaster* [J]. *Genetics*, 2020, 214(2): 511-528.
- [22] CHU D, WEI L. Characterizing the heat response of *Arabidopsis thaliana* from the perspective of codon usage bias and translational regulation [J]. *J Plant Physiol*, 2019, 240: 1-9.
- [23] GUAN D L, MA L B, KHAN M S, et al. Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints [J]. *BMC Genomics*, 2018, 19(1): 1-14.
- [24] INGVARSSON P K. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula* [J]. *Mol Biol Evol*, 2007, 24(3): 836-844.
- [25] XU C, CAI X N, CHEN Q Z, et al. Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium Gower Ramsey* [J]. *Evol Bioinform Online*, 2011, 7: 271-278.
- [26] 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.
- [27] NIE L P, CUI Y X, CHEN X L, et al. Complete chloroplast genome sequence of the medicinal plant *Arctium lappa* [J]. *Genome*, 2020, 63(1): 53-60.
- [28] 王晓双, 王雁青, 李仕洪, 等. 化香属植物叶绿体基因组密码子偏好性分析 [J]. *分子植物育种*, 2021, 19(16): 5352-5362.
- [29] DUAN H R, ZHANG Q, WANG C M, et al. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints [J]. *PeerJ*, 2021, 9: 1-20.
- [30] 王媛媛, 杨美青. 蒙古韭叶绿体基因组密码子使用偏好性分析 [J]. *分子植物育种*, 2021, 19(4): 1084-1092.

(上接第 60 页)

促进种子萌发。该研究创新性地将固氮菌与乳酸菌构建合成微生物菌落,并应用于黄瓜种子萌发,使黄瓜种子的萌发率得到了较大的提升。

种子萌发是一个复杂的过程,用植物促生菌处理黄瓜种子,对其萌发率进行综合评价,对后续的黄瓜生产具有重要意义,也为植物促生菌的开发利用提供了新思路。植物促生菌与植物之间的互作关系复杂,植物促生菌与植物内部的酶活性和激素之间如何互作来调控种子萌发的分子机制,还有待于进一步深入研究。

参考文献

- [1] BASHAN Y, DE-BASHAN L E. BACTERIA/plant growth-promoting [M]//HILLEL D. *Encyclopedia of soils in the environment*. Amsterdam: Elsevier/Academic Press, 2005: 103-115.
- [2] SINGH N, SINGH D. Effect of plant growth promoting bacteria on seed germination, seedling vigor and growth *Lagenaria siceraria* (Molina) standl [J]. *International journal of current microbiology and applied sciences*, 2020, 9(8): 1161-1168.
- [3] MATSUMOTO H, FAN X Y, WANG Y, et al. Bacterial seed endophyte shapes disease resistance in rice [J]. *Nature plants*, 2021, 7(1): 60-72.
- [4] HADAR Y, HARMAN G E, TAYLOR A G, et al. Effects of pregermination of pea and cucumber seeds and of seed treatment with *Enterobacter cloacae* on rots caused by *Pythium* spp. [J]. *Phytopathology*, 1983, 73(9): 1322-1325.
- [5] 卢玉君, 王孝先, 赵伟进, 等. 砂生槐根瘤内生菌对青稞种子萌发及幼苗的促生作用 [J]. *广西植物*, 2021, 41(2): 206-215.
- [6] 麦靖雯, 黎瑞君, 张巨明. 植物根际促生菌研究综述 [J]. *现代农业科技*, 2018(12): 179-180, 183.
- [7] TODD-BOCKARIE A H, DURYE A M L. Seed pretreatment methods to improve germination of the multipurpose West African forest species *Dialium guineense* [J]. *Forest ecology and management*, 1993, 57(1/2/3/4): 257-273.
- [8] TOKLU F, BALOCH F S, KARAKTÖY T, et al. Effects of different priming applications on seed germination and some agromorphological characteristics of bread wheat (*Triticum aestivum* L.) [J]. *Turkish journal of agriculture and forestry*, 2015, 39(6): 1005-1013.
- [9] 李琳琳, 张会玲, 宋宇静. 几种常见的种子处理方法 [J]. *新农民*, 2019(1): 33.
- [10] 张蕊, 田叶韩, 李超, 等. 球毛壳菌 ND35 对黄瓜种子萌发和胚根生长的影响 [J]. *北方园艺*, 2020(23): 1-9.
- [11] 陈沫先, 韦中, 田亮, 等. 合成微生物群落的构建与应用 [J]. *科学通报*, 2021, 66(3): 273-283.
- [12] 曲泽鹏, 陈沫先, 曹朝辉, 等. 合成微生物群落研究进展 [J]. *合成生物学*, 2020, 1(6): 621-634.
- [13] 曹毅, 丘明祺, 温海洋, 等. 耐氮固氮菌对刺黄瓜生长、产量及品质的影响 [J]. *湖北农业科学*, 2004, 43(1): 73-75.
- [14] LIMANSKA N, IVANYTSIA T, BASIUL O, et al. Effect of *Lactobacillus plantarum* on germination and growth of tomato seedlings [J]. *Acta physiologiae plantarum*, 2013, 35(5): 1587-1595.