

金刚, 王丽萍, 龙凌云, 吴凤, 唐玉娟, 覃剑峰, 危丹妮, 黄秋伟, 苏文潘. 普通野生稻线粒体蛋白质编码基因密码子使用偏好性的分析[J].

植物科学学报, 2019, 37(2): 188–197

Jin G, Wang LP, Long LY, Wu F, Tang YJ, Qin JF, Wei DN, Huang QW, Su WP. Analysis of codon usage bias in the mitochondrial protein-coding genes of *Oryza rufipogon*[J]. *Plant Science Journal*, 2019, 37(2): 188–197

普通野生稻线粒体蛋白质编码基因密码子使用偏好性的分析

金 刚*, 王丽萍, 龙凌云, 吴 凤, 唐玉娟, 覃剑峰,
危丹妮, 黄秋伟, 苏文潘

(广西壮族自治区亚热带作物研究所, 南宁 530001)

摘 要: 以普通野生稻 (*Oryza rufipogon* Griff.) 线粒体基因组为对象, 分析其蛋白质编码基因的密码子使用特征及与亚洲栽培稻 (*O. sativa* L.) 的差异, 探讨其密码子偏性形成的影响因素和进化过程。结果显示: 普通野生稻线粒体基因组编码序列第 1、第 2 和第 3 位碱基的 GC 含量依次为 49.18%、42.67% 和 40.86%; 有效密码子数 (Nc) 分布于 45.32 ~ 61.00 之间, 其密码子偏性较弱; Nc 值仅与 GC₃ 呈显著相关, 密码子第 3 位的碱基组成对密码子偏性影响较大; 第 1 向量轴上显示 9.91% 的差异, 其与 GC_{3s}、Nc、密码子偏好指数 (CBI) 和最优密码子使用频率 (Fop) 的相关性均达到显著水平; 而 GC₃ 和 GC₁₂ 的相关性未达到显著水平。因此, 普通野生稻线粒体基因组密码子的使用偏性主要受自然选择压力影响而形成。本研究确定了 21 个普通野生稻线粒体基因组的最优密码子, 大多以 A 或 T 结尾, 与叶绿体密码子具有趋同进化, 但是与核基因组具有不同的偏好性。同义密码子相对使用度 (RSCU)、PR2 偏倚分析和中性绘图分析显示, 普通野生稻线粒体基因功能及其密码子使用密切相关, 且线粒体密码子使用在普通野生稻、梗稻 (*O. sativa* L. subsp. *japonica* Kato) 和籼稻 (*O. sativa* L. subsp. *indica* Kato) 内具有同质性。

关键词: 普通野生稻; 线粒体基因组; 密码子偏好性; 最优密码子

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837 (2019)02-0188-10

DOI: 10.11913/PSJ.2095-0837.2019.20188

开放科学 (资源服务) 标识码 (OSID):



Analysis of codon usage bias in the mitochondrial protein-coding genes of *Oryza rufipogon*

Jin Gang*, Wang Li-Ping, Long Ling-Yun, Wu Feng, Tang Yu-Juan, Qin Jian-Feng,
Wei Dan-Ni, Huang Qiu-Wei, Su Wen-Pan

(Guangxi Subtropical Crops Research Institute, Nanning 530001, China)

Abstract: The mitochondrial genome of *Oryza rufipogon* Griff. was used to analyze the codon usage characteristics of protein-coding genes and the differences with Asian cultivated rice (*O. sativa* L.), and to explore the influencing factors of codon usage bias and codon evolution. Results showed that the effective number of codons (Nc) ranged from 45.32 to 61.00, indicating that codon bias was weak. GC content at the three codon positions was 49.18%, 42.67%, and 40.86%. The Nc value was significantly correlated with GC₃ but not with GC₁ or GC₂, suggesting that base composition at the third codon position had a greater

收稿日期: 2018-08-26, 退修日期: 2018-11-01。

基金项目: 广西自然科学基金项目 (2016GXNSFAA380093, 2018GXNSFBA281109)。

This work was supported by grants from the Natural Science Foundation of Guangxi (2016GXNSFAA380093, 2018GXNSFBA281109)。

作者简介: 金刚 (1981–), 男, 博士, 高级实验师, 研究方向为作物分子育种 (E-mail: Jing8118@163.com)。

* 通讯作者 (Author for correspondence. E-mail: Jing8118@163.com)。

effect on codon bias. In the corresponding analysis, the first axis showed 9.91% variation and was significantly correlated with GC_{3s}, Nc, CBI, and Fop. Furthermore, GC₁₂ showed non-significant correlation with GC₃. Codon bias in the mitogenome of *O. rufipogon* was mainly affected by natural selection. In addition, we identified 21 optimal codons, with most of the preferred synonymous codons ending in A or T. The mitochondrial codons showed convergent evolution with *Oryza* chloroplast codons, but different preferences with the nuclear genome. Based on neutrality plot analysis, PR2-plot analysis, and RSCU analysis, we found no significant differences among the three *Oryza* species. Our results also confirmed homogeneity in mitochondrial codon usage among the three *Oryza* species.

Key words: *Oryza rufipogon* Griff.; Mitogenome; Codon usage bias; Optimal codon

同义密码子使用偏性 (Synonymous codon usage bias, SCUB) 指在不同物种之间或同一物种不同基因之间的同义密码子出现频率的差异^[1, 2], 是生物界普遍存在的现象, 在原核生物、动物和植物中都得到了验证^[3]。遗传密码的功能完整性主要通过同义密码子维持。就通用三联体遗传密码而言, 密码子的简并性多数发生在第3位置碱基, 偶尔也会发生在第2位。在不改变氨基酸序列的前提下, 遗传密码的简并性保证了蛋白质翻译效率和准确性。虽然许多同义密码子参与翻译的过程, 但只有特定密码子是优选的。优选密码子的使用频率可能会因为突变的偏好性或因为在蛋白质合成阶段最佳翻译的自然选择而发生变化。自然选择和突变压力是导致同义密码子使用偏性的两个主要进化力量^[4]。同义密码子使用偏性揭示了自然选择和突变压力之间的平衡, 以及随机遗传漂变的影响^[5, 6]。分子进化中性理论认为分子进化速率与选择动力呈反比关系。蛋白质编码基因的同义替代速度慢于假基因, 暗示了选择动力对同义密码子进化的速率有影响^[7]。影响同义密码子使用偏性模式的其他因素有基因长度^[8]、CpG 岛^[9]、tRNA 丰度^[10]和蛋白质折叠等^[11]。自然选择的因素涉及翻译选择、基因长度和基因功能, 而突变偏性的因素包括 GC 含量和碱基的突变位置。同义密码子使用的研究可揭示基因的分子进化信息, 用于鉴定遗传信息的水平转移事件^[12, 13]。

线粒体是真核细胞内具有双层膜结构的重要细胞器, 可为细胞的一系列生命活动提供场所, 如能量转换、细胞信息传递、细胞分化、细胞凋亡和细胞生长的调控等^[14]。在高等植物中, 线粒体 DNA (mtDNA) 作为核基因组外的重要遗传物质, 具有

母系遗传的特征。与核基因组相比, 线粒体基因组较小, 其所含几十个基因分别编码核糖体 RNA、转运 RNA 和能量代谢相关的功能蛋白^[14]。且 mtDNA 易扩增、高度保守, 具有较高的进化速率和无组织特异性等特征^[15]。因此, mtDNA 被广泛用于开发作物种鉴定、群体遗传学、系统发育和进化研究的分子标记^[16]。植物线粒体密码子使用偏性的分析对研究其基因表达和物种进化具有重要意义。在过去的十年中随着高通量测序技术的发展, 对植物线粒体基因组测序及其后基因组的分析日益深入。但关于植物线粒体基因组密码子使用偏性的相关研究却鲜有报道。

普通野生稻 (*Oryza rufipogon* Griff.) 是亚洲栽培稻 (*O. sativa* L.) 的近缘祖先, 它们均具有 AA 型染色体组, 而亚洲栽培稻又进一步分化为粳稻 (*O. sativa* L. subsp. *japonica* Kato) 和籼稻 (*O. sativa* L. subsp. *indica* Kato) 2 个亚种。普通野生稻线粒体基因组含有细胞质雄性不育相关的基因 (如 *CW-orf307*)^[17], 以其作为母本, 并以细胞核基因组携带无功能型恢复基因的栽培稻材料为父本, 通过饱和核置换回交可创制出在农业生产上有重要应用价值的水稻细胞质雄性不育系^[18]。因此, 普通野生稻的 mtDNA 在水稻的三系杂交育种中有着极其重要的作用。

刘庆坡等^[19, 20]在全基因组水平上研究了籼稻品种‘9311’同义密码子的使用偏性, 并确定了对水稻转基因具有重要意义的 25 个最优密码子; 通过对品种‘日本晴’叶绿体基因组的分析, 明确了粳稻叶绿体基因组的密码子用法。但目前对稻属植物线粒体编码基因密码子使用的系统研究较少, 其线粒体基因组密码子使用偏好性与进化关系的原因

尚不清楚。本研究以普通野生稻线粒体基因组为对象,首次系统地对其密码子使用偏性进行分析,明确影响其密码子使用模式的主要因素,确定最优密码子,并比较其与亚洲栽培稻线粒体基因组密码子间的使用差异,以期为水稻线粒体基因表达和进化遗传的深入研究提供理论依据。

1 材料与方法

1.1 实验材料

在 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中检索梗稻、籼稻和普通野生稻的线粒体基因组相关信息,其 GenBank 登录号分别为 NC_011033.1^[17]、NC_007886.1^[21] 和 NC_013816.1^[22]。从中分别提取、下载这 3 个稻属植物的全部编码序列,剔除重复序列和长度小于 300 bp 的序列,余下序列经开放阅读框研判后用于后续分析。

1.2 方法

1.2.1 密码子偏好性分析

应用 CodonW 软件 (<http://codonw.sourceforge.net>) 分析计算普通野生稻、梗稻和籼稻 3 个稻属植物各线粒体编码基因的同义密码子第 3 位碱基 GC 含量 (GC_{3s})、有效密码子数 (Effective number of codons, Nc)、密码子适应度 (Codon adaptation index, CAI)、最优密码子使用频率 (Frequency of optimal codons, Fop)、氨基酸总数 (Length amino acids, L_aa)、总平均亲水性 (Grand average of hydropathy, Gravy) 等指数。计算各编码基因的同义密码子相对使用度 (Relative synonymous codon usage, RSCU)、GC 总含量 (GC_{all}) 以及密码子第 1、第 2、第 3 位的 GC 含量 (GC_1 、 GC_2 和 GC_3)^[23]。采用 SPSS 19.0 软件开展各指数之间的相关性分析和显著性检验。

1.2.2 PR2-plot 分析

选择具有 4 种密码子的氨基酸,即丝氨酸、亮氨酸、脯氨酸、精氨酸、苏氨酸、缬氨酸、丙氨酸和甘氨酸,计算每个基因的 $G_3/(G_3 + C_3)$ 和 $A_3/(A_3 + T_3)$,以此分别作为横坐标和纵坐标,以平面图显示各基因的碱基含量关系^[24]。若坐标点在平面图中均匀分布,即第 3 位密码子 A 与 T 或 G 与 C 使用频率相同,则密码子使用偏好可能完全是由突变引起的。

1.2.3 中性绘图分析

为分析三联体密码子 3 个位置碱基组成间的相关性,在分别统计编码基因密码子 3 个位置 GC 含量的基础上,以 GC_{12} (GC_2 和 GC_1 的平均值) 为纵坐标,以 GC_3 为横坐标进行中性作图^[25]。若 GC_{12} 与 GC_3 不相关,则说明密码子第 1、第 2 位与第 3 位碱基组成差异较大,同时回归系数接近 0,密码子偏性受选择压力影响较大;反之,若 GC_{12} 与 GC_3 显著相关,同时回归系数接近 1,说明密码子第 1、第 2 位与第 3 位碱基组成相似,密码子偏性的形成受外部选择压力的影响较小。

1.2.4 对应分析

对应分析广泛应用于密码子使用模式的多元统计分析^[26]。利用 CodonW 程序,根据各编码基因 59 个同义密码子的 RSCU 值进行对应性分析 (Corresponding analysis, COA)。基于 RSCU 值将普通野生稻所有线粒体编码基因按密码子的使用频率分布于 1 个 59 维的向量空间,自由度为 41。计算 59 个同义密码子相对使用度的欧氏平方距离,以此作为基因间的距离。

1.2.5 最优密码子分析

以 Nc 值为偏性标准,将线粒体基因组编码基因中 Nc 值居于最低和最高的 10% 基因分别作为高、低表达组,分别计算 2 组的 RSCU 值, $\Delta RSCU \geq 0.08$ 的密码子定义为高表达密码子。将 RSCU 值大于 1 的密码子确定为高频率密码子。同时满足上述 2 种条件的密码子定义为最优密码子^[27]。

2 结果与分析

2.1 密码子使用偏性分析

研究结果显示,普通野生稻 39 个线粒体编码基因的总 GC 含量为 44.24%,而其整个线粒体基因组 GC 含量为 44.04%,两者基本一致。GC 含量在密码子 3 个位置上有差异, GC_1 、 GC_2 和 GC_3 分别为 49.18%、42.67% 和 40.86%,其分布趋势依次为: $GC_1 > GC_2 > GC_3$ 。这与密码子第 1、第 2、第 3 位的简并性趋势相反,第 1、第 2 位受到的选择压力比第 3 位更大,即在普通野生稻的线粒体基因组中,以 G 或 C 碱基结尾的密码子使用频率低于 A 或 T 碱基结尾的密码子。Nc 的理论取值范围为 20 ~ 61,其值越小,同义密码子的偏性

则越强。普通野生稻线粒体基因组各编码序列的 Nc 值为 45.32 ~ 61.00(表 1)。包括细胞质雄性不育相关基因 *CW-orf307* 在内的所有基因 Nc 值均大于 45, 可以判断普通野生稻线粒体基因密码子偏性很弱。其中 *orf153*、*orf176*、*orf152a*、*rps7* 4 个基因的 Nc 值为 61, 表明其同义密码子的使用没有偏好性。*rps13* 基因的 Nc 值最小, 其数值为 45.32。细胞质雄性不育相关基因 *CW-orf307*

表 1 普通野生稻 39 个线粒体基因的 GC 含量及 Nc 值
Table 1 GC content and Nc value of 39 gene codons from *Oryza rufipogon* mitochondrial genome

基因 Gene	GC (%)				Nc
	GC ₁	GC ₂	GC ₃	GC _{all}	
<i>orfX</i>	43.33	41.85	44.07	43.09	55.07
<i>orf187</i>	47.34	42.55	45.21	45.04	54.06
<i>orf194</i>	49.23	50.26	47.69	49.06	46.81
<i>rps3</i>	45.45	41.29	43.75	43.50	58.83
<i>nad3</i>	44.23	41.03	37.18	40.81	56.64
<i>rps12</i>	54.76	47.62	26.98	43.12	49.66
<i>rps2</i>	46.90	40.50	39.46	42.29	57.05
<i>atp6</i>	45.56	34.62	35.80	38.66	52.48
<i>orf153</i>	41.56	46.10	41.56	43.07	61.00
<i>rps13</i>	51.28	41.88	29.06	40.74	45.32
<i>rps4</i>	40.79	40.79	37.11	39.57	52.71
<i>orf176</i>	57.63	51.98	42.94	50.85	61.00
<i>atp9</i>	47.66	48.44	39.84	45.31	57.40
<i>cox3</i>	52.67	45.55	36.30	44.84	55.27
<i>orf25</i>	45.96	40.91	37.88	41.58	59.99
<i>orf152a</i>	49.67	45.75	53.59	49.67	61.00
<i>rps7</i>	50.34	40.94	36.24	42.51	61.00
<i>orf490</i>	51.32	41.34	35.03	42.57	53.28
<i>orfB</i>	33.97	37.18	44.87	38.68	58.20
<i>orf181</i>	42.86	40.66	40.66	41.39	55.78
<i>nad6</i>	41.75	36.89	40.78	39.81	53.62
<i>ccmC</i>	46.89	47.72	32.78	42.46	47.15
<i>orf183</i>	54.89	40.76	35.87	43.84	54.63
<i>orf173</i>	45.98	47.13	54.60	49.23	51.91
<i>nad9</i>	50.79	41.88	32.98	41.88	54.34
<i>atp1</i>	56.67	41.37	34.90	44.31	52.71
<i>CW-orf307</i>	54.22	44.16	43.18	47.19	58.47
<i>orf160</i>	52.80	34.16	42.24	43.06	54.07
<i>cox2</i>	51.34	39.08	32.95	41.12	49.14
<i>rpl5</i>	47.62	31.75	42.86	40.74	53.57
<i>cox1</i>	46.13	41.90	37.41	41.81	56.11
<i>orf284</i>	53.68	35.44	44.91	44.68	58.31
<i>orf165</i>	43.37	49.40	45.78	46.18	52.54
<i>CcmFn</i>	50.42	49.24	41.51	47.06	57.48
<i>rps1</i>	49.71	38.73	39.88	42.77	50.53
<i>mat-r</i>	52.81	43.85	57.51	51.39	56.99
<i>Cob</i>	50.75	41.21	33.67	41.88	56.63
<i>rf241</i>	49.17	40.08	43.39	44.21	51.06
<i>rpl2</i>	54.85	51.88	44.16	50.30	57.44

的 GC₁、GC₂ 和 GC₃ 值分别为 54.22%、44.16% 和 43.18%, 均略大于普通野生稻 39 个线粒体编码基因的第 1、第 2、第 3 位 GC 含量。

GC_{all} 与 GC₁、GC₂ 及其与 GC₃ 均呈极显著相关性(表 2)。GC₁、GC₂ 和 GC₃ 相互间的相关性水平均不显著, 表明密码子第 1、第 2、第 3 位的碱基组成有差异。Nc 值仅与 GC₃ 呈显著相关, 与 GC₁ 和 GC₂ 的相关性均不显著。表明密码子第 3 位的碱基组成对密码子偏性影响较大, 而第 1、第 2 位的碱基组成对其影响较弱。

表 2 不同位置密码子 GC 含量的相关性分析
Table 2 Correlation analysis of codon GC content at different positions

参数 Parameter	GC ₃	GC ₂	GC ₁	GC _{all}
GC ₂	0.161			
GC ₁	-0.154	0.235		
GC _{all}	0.642 **	0.701 **	0.508 **	
Nc	0.339 *	0.051	-0.054	0.215

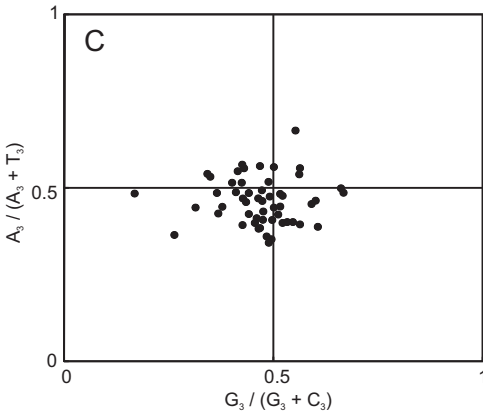
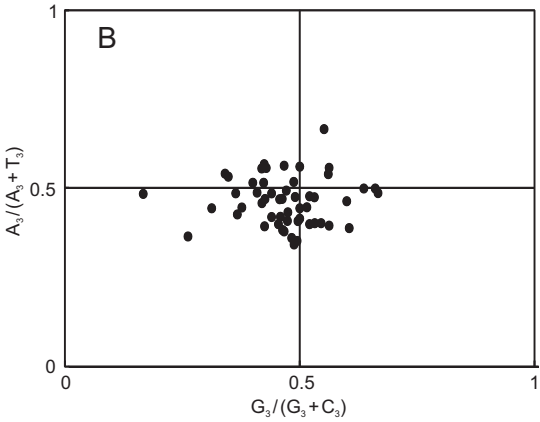
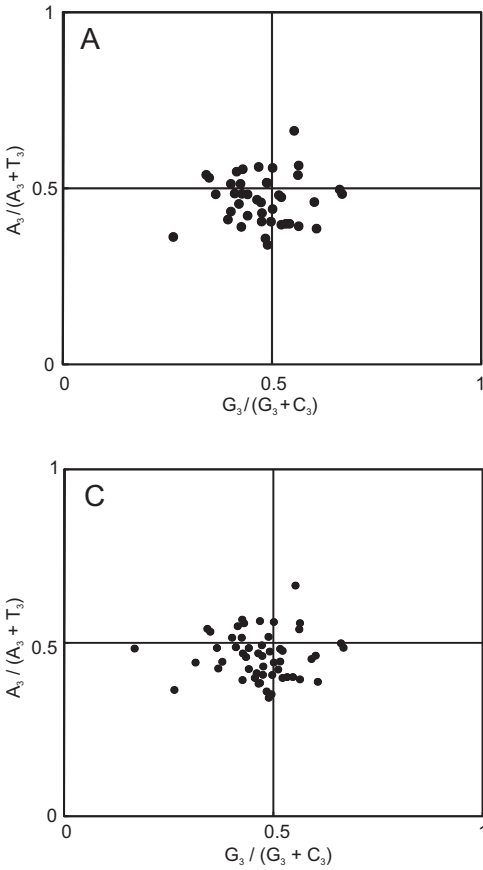
注: * 表示在 0.05 水平(双侧)上显著相关; ** 表示在 0.01 水平(双侧)上显著相关。下同。
Notes: * significant difference at $P < 0.05$; ** significant difference at $P < 0.01$. Same below.

2.2 PR2-plot 分析

本研究采用 PR2-plot 分析, 比较了普通野生稻和亚洲栽培稻(梗稻、籼稻)线粒体基因组各编码基因中由 4 种碱基组成的三联体密码子第 3 位嘧啶(T 和 C)以及嘌呤(A 和 G)之间的关系(图 1)。结果表明, 普通野生稻和亚洲栽培稻在 PR2-plot 分析图中的坐标点分布格局一致, 不均匀地分布于平面图的 4 个区域, 且包括细胞质雄性不育相关基因 *CW-orf307* 在内的多数基因分布于左下方, 说明其在密码子第 3 位嘌呤中 T 的使用频率高于 A, 嘧啶中 C 的使用频率高于 G, 且嘧啶的使用频率高于嘌呤。若密码子使用偏好性完全受突变压力影响时, 则理论上 A/T 及 G/C 的使用频率相等。本研究中 4 种碱基的使用不平衡, 暗示 3 种稻属植物线粒体密码子的使用偏好性在受突变因素影响的同时, 还受到自然选择压力等其他因素的影响。

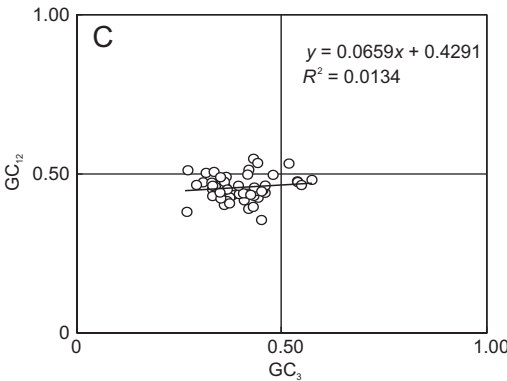
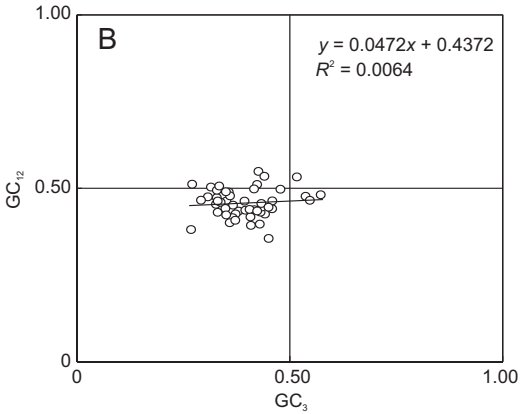
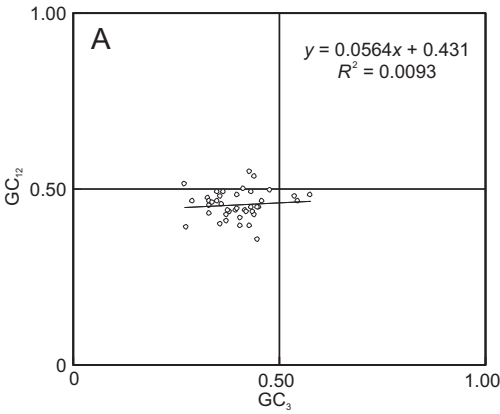
2.3 中性绘图分析

为明确碱基组成对密码子偏性形成的影响, 本研究进行中性绘图分析(图 2)。在平面图中, 3 个稻属植物的坐标点分布格局一致, 回归系数(斜率)分别为 0.0564、0.0472 和 0.0659。



A: 普通野生稻; B: 籼稻; C: 粳稻。
A: *Oryza rufipogon*; B: *O. sativa* subsp. *indica*; C: *O. sativa* subsp. *japonica*.

图 1 线粒体基因中三联体密码子第 3 位嘧啶及其与嘌呤之间的相关性 PR2-plot 分析
Fig. 1 Parity rule 2 bias plot analysis of the correlation between pyrimidine and purine at the third codon position in the mitochondrial gene



A: 普通野生稻; B: 籼稻; C: 粳稻。
A: *Oryza rufipogon*; B: *O. sativa* subsp. *indica*; C: *O. sativa* subsp. *japonica*.

图 2 碱基组成对密码子偏性形成影响的中性绘图分析

Fig. 2 Neutrality plot analysis of the effect of base composition on codon bias

在普通野生稻中，GC₁₂ 分布于 0.3558 ~ 0.5480 之间，GC₃分布于 0.2698 ~ 0.5751 之间。相关性分析显示 GC₃和 GC₁₂的相关系数为 0.003，双尾检测显示其相关性未达显著水平，表明其线粒体基因组编码基因第 1、第 2 和第 3 位碱基组成不同。因此，推断突变可能在形成稻属线粒体基因密码子偏性形成中起次要作用，而自然选择压力可能起主导作用。

2.4 对应性分析

本研究对野生稻线粒体基因组编码基因开展了基于 RSCU 的对应性分析(表 3)，去除了氨基酸组成对密码子使用的影响。结果表明，第 1 向至第 4 向量轴 (Axis) 可以分别解释密码子总变异的 9.91%、9.25%、7.94%和 7.31%，前 4 轴累计解释的差异为 34.41%，但第 1 ~ 第 4 轴的差异水平均小于 10%。第 1 轴对密码子偏性的贡献率最大。GC_{3s} 与 Nc 呈显著相关($r = 0.349$, $P < 0.05$)，与 Axis1 呈极显著负相关($r = -0.879$, $P < 0.01$)，提示了碱基组成可能对密码子偏性产生影响。Axis1

与 Nc 和 GC_{3s} 的负相关关系均达到极显著水平，相关系数分别为 -0.443 和 -0.879 ；同时，Axis1 与 Fop 呈极显著负相关($r = -0.432$, $P < 0.01$)，与 CBI 的负相关性也达到显著水平($r = -0.405$, $P < 0.05$)。表明碱基组成和基因表达对普通野生稻 mtDNA 密码子偏性的形成均具有显著的影响。第 1~ 第 4 向量轴与 GRAVY(普通平均蛋白质亲水性)均不存在显著相关性，提示编码疏水性氨基酸的密码子所承受的选择压力小。L_aa 与各向量轴也均无显著相关性，表明基因长度对密码子偏性没有明显影响。

本研究基于基因功能将普通野生稻线粒体基因分为 7 类，并以 Axis1 为横坐标、Axis2 为纵坐标作图。结果表明，呼吸系统基因及遗传系统基因分布相对集中，说明此类基因的密码子使用模式相对一致(图 3)。核糖体蛋白基因集中于 Axis1 的左端，而呼吸链复合体 I ~ V 基因集中于 Axis1 的右端。由此推断，普通野生稻线粒体基因功能和其密码子使用密切相关。

表 3 普通野生稻线粒体基因密码子偏性相关参数的相关性
Table 3 Correlation among main parameters of gene codons in *O. rufipogon* mtDNA

参数 Parameter	Axis2	Axis3	Axis4	Nc	GC _{3s}	L_aa	GC _{all}	Gravy	CAI	CBI	Fop
Axis1	0.008	0.001	-0.005	-0.443 **	-0.879 **	-0.119	-0.588 **	0.049	-0.090	-0.405 *	-0.432 **
Axis2		-0.005	-0.003	-0.353 *	0.061	-0.127	0.242	0.107	0.129	0.228	0.233
Axis3			-0.003	0.096	-0.166	0.138	0.030	0.190	0.228	0.048	-0.071
Axis4				-0.163	-0.205	-0.042	-0.174	0.122	0.054	-0.386 *	-0.354 *
Nc					0.349 *	0.160	0.220	0.090	0.047	0.172	0.195
GC _{3s}						0.141	0.661 **	-0.073	-0.072	0.558 **	-0.548 **
L_aa							0.203	0.074	-0.105	0.049	0.046
GC _{all}								-0.258	-0.026	0.390 *	0.431 **
Gravy									0.211	-0.013	-0.197
CAI										0.270	0.331 *
CBI											0.949 **

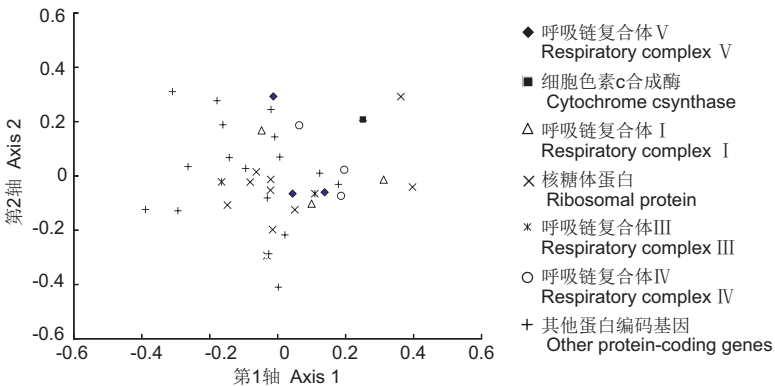


图 3 基于普通野生稻线粒体基因 RSCU 值的对应性分析
Fig. 3 Corresponding analysis based on RSCU in *O. rufipogon* mtDNA

2.5 最优密码子分析

本研究将普通野生稻、粳稻和籼稻的线粒体编码基因的同义密码子相对使用度和使用数目进行统计。在 3 个稻属植物之间，其 59 个同义密码子的 RSCU 值均差异不显著(表 4)。

同时，根据普通野生稻线粒体基因组各编码基因的 Nc 值排序，建立了由 *orf153*、*orf176*、*orf152a*、*rps7* 组成的低表达组和由 *orf194*、*rps13*、*ccmC*、

cox2 组成的高表达组。分别计算 2 组基因同义密码子的 RSCU 值，并依据对高频率密码子和高表达密码子的定义，最终确定 TTA、TTG、ATT 等 21 个密码子为最优密码子。其中有 8 个最优密码子以 A 结尾，11 个以 T 结尾，以 C 和 G 结尾的各有 1 个。表明普通野生稻线粒体基因组编码基因偏好于使用 A 或 T 结尾的密码子，这与该物种线粒体基因组密码子第 3 位富含 A/T (59.14%) 的结果相一致。

表 4 稻属植物线粒体编码基因的相对同义密码子使用度
Table 4 Relative synonymous codon usage of mitochondrial genes in *Oryza*

氨基酸 Amino acid	密码子 Codon	同义密码子相对使用度/使用次数 RSCU/Number				
		粳稻	籼稻	普通野生稻 <i>O. rufipogon</i>		
		<i>O. sativa</i> subsp. <i>japonica</i>	<i>O. sativa</i> subsp. <i>indica</i>	基因组 Genome	高表达组 High expression	低表达组 Low expression
Leu/L	TTA *	1.2777/342	1.2898/342	1.2598/236	1.6154	0.8000
	TTG *	1.2403/332	1.2370/328	1.2544/235	1.4615	1.2000
	CTT	1.2814/343	1.3011/345	1.2758/239	1.2308	2.0000
	CTC	0.721/193	0.7014/186	0.7473/140	0.6923	0.4000
	CTA	0.8219/220	0.8297/220	0.7527/141	0.3077	0.9333
	CTG	0.6575/176	0.6411/170	0.7100/133	0.6923	0.6667
Ile/I	ATT *	1.2731/460	1.2778/460	1.2090/322	1.4143	0.8571
	ATC	0.8413/304	0.8361/301	0.8711/232	1.0286	1.0286
	ATA	0.8856/320	0.8861/319	0.9199/245	0.5571	1.1143
Val/V	GTT *	1.2711/279	1.2754/279	1.2798/207	1.3023	1.0667
	GTC	0.7608/167	0.7543/165	0.7357/119	0.8372	0.8889
	GTA *	1.0387/228	1.0514/230	1.0572/171	1.3023	1.0667
	GTG	0.9294/204	0.9189/201	0.9274/150	0.5581	0.9778
Ser/S	TCT *	1.3283/292	1.3085/287	1.3642/211	1.5584	1.4737
	TCC *	1.0690/235	1.0714/235	1.0086/156	1.0909	0.6316
	TCA	1.1645/256	1.1763/258	1.1703/181	1.4805	1.4737
	TCG	0.7915/174	0.7979/175	0.8276/128	0.7013	0.6316
	AGT	0.9780/215	0.9939/218	0.9763/151	1.0130	1.0526
	AGC	0.6687/147	0.6520/143	0.6530/101	0.1558	0.7368
Pro/P	CCT *	1.2293/252	1.2338/252	1.1150/160	1.2364	1.0476
	CCC	0.9512/195	0.9547/195	0.9617/138	1.0182	0.9524
	CCA *	1.2098/248	1.2093/247	1.2753/183	1.4545	1.1429
	CCG	0.6098/125	0.6022/123	0.6481/93	0.2909	0.8571
Thr/T	ACT *	1.4374/267	1.4290/264	1.3933/186	1.5686	1.2857
	ACC	1.0229/190	1.0230/189	0.9963/133	0.7059	1.5714
	ACA *	0.9637/179	0.9581/177	1.0337/138	1.2549	0.5714
	ACG	0.5760/107	0.5900/109	0.5768/77	0.4706	0.5714
Ala/A	GCT *	1.5238/376	1.5234/374	1.4642/261	2.1600	1.7000
	GCC	0.8430/208	0.8473/208	0.8247/147	0.7200	1.0000
	GCA	1.0213/252	1.0143/249	1.0603/189	0.4800	0.8000
	GCG	0.6120/151	0.6151/151	0.6508/116	0.6400	0.5000

续表 4

氨基酸 Amino acid	密码子 Codon	同义密码子相对使用度/使用次数 RSCU/Number				
		籼稻	粳稻	普通野生稻 <i>O. rufipogon</i>		
		<i>O. sativa</i> subsp. <i>japonica</i>	<i>O. sativa</i> subsp. <i>indica</i>	基因组 Genome	高表达组 High expression	低表达组 Low expression
Tyr/Y	TAT *	1.426/313	1.4492/321	1.3657/211	1.5789	1.1111
	TAC	0.574/126	0.5508/122	0.6343/98	0.4211	0.8889
His/H	CAT *	1.4763/265	1.4734/263	1.4015/185	1.9000	0.8889
	CAC	0.5237/94	0.5266/94	0.5985/79	0.1000	1.1111
Gln/Q	CAA *	1.4218/332	1.4366/334	1.3539/241	1.6471	1.2381
	CAG	0.5782/135	0.5634/131	0.6461/115	0.3529	0.7619
Asn/N	AAT	1.3068/328	1.3158/325	1.2732/254	0.9167	1.0303
	AAC	0.6932/174	0.6842/169	0.7268/145	1.0833	0.9697
Lys/K	AAA *	1.1605/376	1.1730/383	1.1217/295	1.1333	1.1667
	AAG	0.8395/272	0.8270/270	0.8783/231	0.8667	0.8333
Asp/D	GAT	1.3829/372	1.3870/362	1.3468/266	1.1429	1.1818
	GAC	0.6171/166	0.6130/160	0.6532/129	0.8571	0.8182
Glu/E	GAA *	1.3142/435	1.3272/436	1.2972/323	1.5610	1.0909
	GAG	0.6858/227	0.6728/221	0.7028/175	0.4390	0.9091
Cys/C	TGT *	1.1528/132	1.1342/131	1.1266/89	1.5000	0.9231
	TGC	0.8472/97	0.8658/100	0.8734/69	0.5000	1.0769
Arg/R	CGT *	1.0941/186	1.1209/190	1.0978/144	1.3548	1.1613
	CGC	0.5765/98	0.5664/96	0.6557/86	0.4839	0.6774
	CGA	1.1647/198	1.1858/201	1.1436/150	1.0645	1.2581
	CGG	0.7647/130	0.7316/124	0.7166/94	1.1613	0.5806
	AGA	1.5353/261	1.5398/261	1.4943/196	1.1613	1.3548
	AGG	0.8647/147	0.8555/145	0.8920/117	0.7742	0.9677
	GGT	1.1446/289	1.1535/293	1.0841/203	1.0667	1.2308
	GGC	0.6059/153	0.5984/152	0.6195/116	0.0889	0.5128
Gly/G	GGA *	1.3941/352	1.4016/356	1.4366/269	2.4000	1.4359
	GGG	0.8554/216	0.8465/215	0.8598/161	0.4444	0.8205
Phe/F	TTT *	1.1638/547	1.1645/545	1.1514/365	1.2941	1.1429
	TTC	0.8362/393	0.8355/391	0.8486/269	0.7059	0.8571
Met/M	ATG	1.0000/380	1.0000/377	1.0000/242		
Trp/w	TGG	1.0000/247	1.0000/235	1.0000/181		

注：* 表示 $\Delta\text{RSCU} \geq 0.08$ 且 RSCU 值 ≥ 1 。
Note: * indicates $\Delta\text{RSCU} \geq 0.08$ and RSCU value ≥ 1 .

3 讨论

研究证实，同义的三联体密码子通常不被随机使用，产生这种偏好性的主要动力是自然选择和突变^[6]。密码子使用模式及影响其多样性的各种因素的研究，是了解基因组进化和生物体适应生境的基础^[28-30]。植物线粒体基因组密码子偏好性分析

对于研究其 mtDNA 的遗传规律、系统发育关系和进化等具有重要意义，以往对植物密码子使用偏性的研究主要集中于核基因组和叶绿体基因组水平^[19, 20, 27]。本研究对普通野生稻线粒体基因组的密码子使用偏性进行分析，综合碱基组成、PR2-plot 分析、中性绘图和对应性分析的结果，推断自然选择和突变共同影响其密码子使用偏性的形成。

本研究结果与 Zhou 等^[31] 对小麦 (*Triticum aestivum* L.)、玉米 (*Zea mays* L.)、拟南芥 (*Arabidopsis thaliana* (L.) Heynh.)、烟草 (*Nicotiana tabacum* L.)、小立碗藓 (*Physcomitrella patens* (Hedw.) Bruch & Schimp.) 和地钱 (*Marchantia polymorpha* L.) 等 6 种植物线粒体基因组密码子偏性的特征分析结果一致。RSCU 值、PR2-plot 分析和中性绘图分析结果在普通野生稻、粳稻和籼稻之间并无显著差异。表明这 3 种稻属植物线粒体基因组密码子使用偏好性差异较小，其在密码子使用模式上并无显著差别。这一结果也暗示了稻属不同物种线粒体基因组密码子使用的同质性。

刘庆坡等^[20] 通过对日本晴叶绿体全基因组的分析，确定了 TGT 和 GAA 等粳稻叶绿体基因组的 12 个最优密码子，本研究所确定的普通野生稻线粒体基因组最优密码子中有 9 个与之相同。稻属植物线粒体基因组和叶绿体基因组最优密码子的一致性，可能与其趋同进化有关。而刘庆坡等^[19] 在籼稻 9311 全基因组中确定的 25 个最优密码子中，仅有 6 个同为普通野生稻线粒体基因组最优密码子，提示水稻的核基因组和线粒体基因组具有不同的密码子使用偏好性。本研究未发现普通野生稻细胞质雄性不育相关基因 *CW-orf307* 在其密码子使用偏性上的特殊性。就密码子偏性而言，表明 *CW-orf307* 具有与其线粒体基因组共同进化的特征。

本研究确定的野生稻线粒体基因组 21 个最优密码子多以 A 或 T 结尾。值得注意的是，小麦、玉米、拟南芥、烟草、小立碗藓和地钱等物种的最优密码子也更倾向于以 A 或 T 结尾^[31, 32]。水稻叶绿体基因组最优密码子大多也均以 A 或 T 结尾^[19, 28]，这似乎也是稻属植物核外遗传信息表达系统趋同进化的表现。普通野生稻线粒体基因组的这一特征也可能为其内共生进化方式提供了一定的佐证。这些密码子偏好性与翻译水平显著相关，对从分子水平研究植物进化机制时的简并引物设计、点突变的引入等具有重要的参考意义。本研究也为通过比较水稻核基因、线粒体基因和叶绿体基因密码子使用模式的差异来预测未知蛋白的亚细胞定位提供了参考信息。

参考文献：

[1] Hanson G, Collier J. Codon optimality, bias and usage in

translation and mRNA decay [J]. *Nat Rev Mol Cell Biol*, 2018, 19(1): 20–30.

[2] Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species [J]. *Genome Biol Evol*, 2011, 3: 868–880.

[3] Supek F. The code of silence: widespread associations between synonymous codon biases and gene function [J]. *J Mol Evol*, 2016, 82(1): 65–73.

[4] Duret L. Evolution of synonymous codon usage in metazoans [J]. *Curr Opin Genet Dev*, 2002, 12(6): 640–649.

[5] Sharp MP, Li W. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons [J]. *Nucleic Acids Res*, 1986, 14(19): 7737–7749.

[6] Bulmer M. The selection-mutation-drift theory of synonymous codon usage [J]. *Genetics*, 1991, 129(3): 897–907.

[7] Li WH, Gojobori T, Nei M. Pseudogenes as a paradigm of neutral evolution [J]. *Nature*, 1981, 292 (5820): 237–239.

[8] Song H, Gao H, Liu J, Tian P, Nan Z. Comprehensive analysis of correlations among codon usage bias, gene expression, and substitution rate in *Arachis duranensis* and *Arachis ipaënsis* orthologs [J]. *Sci Rep*, 2017, 7(1): 14853.

[9] Clément Y, Sarah G, Holtz Y, Homa F, Pointet S, et al. Evolutionary forces affecting synonymous variations in plant genomes [J]. *PLoS genetics*, 2017, 13 (5): e1006799.

[10] Olejniczak M, Uhlenbeck OC. tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition [J]. *Biochimie*, 2006, 88(8): 943–950.

[11] Gu W, Zhou T, Ma J, Sun X. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens* [J]. *Biosystems*, 2004, 73 (2): 89–97.

[12] Fickett JW. Recognition of protein coding regions in DNA sequences [J]. *Nucleic Acids Res*, 1982, 10: 5303–5318.

[13] Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation [J]. *Annu Rev Plant Biol*, 2017, 68: 225–252.

[14] Liberatore KL, Dukowicz-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance [J]. *Free Radic Biol Med*, 2016, 100: 238–256.

[15] Curole JP, Kocher TD. Mitogenomics: digging deeper with complete mitochondrial genomes [J]. *Trends Ecol Evol*, 1999, 14(10): 394–398.

- [16] Liang X, Tian X, Liu W, Wei T, Wang W, *et al.* Comparative analysis of the mitochondrial genomes of *Colletotrichum gloeosporioides* sensu lato: insights into the evolution of a fungal species complex interacting with diverse plants[J]. *BMC Genomics*, 2017, 18(1): 171.
- [17] Fujii S, Kazama T, Yamada M, Toriyama K. Discovery of global genomic reorganization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes[J]. *BMC Genomics*, 2010, 11: 209.
- [18] 陈乐天, 刘耀光. 水稻野败型细胞质雄性不育的发现利用与分子机理[J]. 科学通报, 2016, 61(35): 3804–3812.
Chen LT, Liu YG. Discovery, utilization and molecular mechanisms of CMS-WA in rice[J]. *Chinese Science Bulletin*, 2016, 61(35): 3804–3812.
- [19] 刘庆坡, 谭军, 薛庆中. 籼稻品种 93-11 同义密码子的使用偏性[J]. 遗传学报, 2003, 30(4): 335–340.
Liu QB, Tan J, Xue QZ. Synonymous codon usage bias in the rice cultivar 93-11 (*Oryza sativa* L. ssp. *indica*) [J]. *Acta Genetica Sinica*, 2003, 30(4): 335–340.
- [20] 刘庆坡, 薛庆中. 粳稻叶绿体基因组的密码子用法[J]. 作物学报, 2004, 30(12): 1220–1224.
Liu QP, Xue QZ. Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*) [J]. *Acta agronomica sinica*, 2004, 30(12): 1220–1224.
- [21] Tian X, Zheng J, Hu S, Yu J. The rice mitochondrial genomes and their variations[J]. *Plant Physiol*, 2006, 140(2): 401–410.
- [22] Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, *et al.* The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants[J]. *Mol Genet Genomics*, 2002, 268(4): 434–445.
- [23] Puigbo P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation[J]. *Biol Direct*, 2008, 3: 38.
- [24] Sueoka N. Two aspects of DNA base composition; G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C[J]. *J Mol Evol*, 1999, 49(1): 49–62.
- [25] Sueoka N. Directional mutation pressure and neutral molecular evolution[J]. *Proc Natl Acad Sci USA*, 1988, 85(8): 2653–2657.
- [26] Maria D, Ermolaeva. Synonymous codon usage in bacteria[J]. *Curr Issues Mol Biol*, 2001, 3(4): 91–97.
- [27] Wang L, Xing H, Yuan Y, Wang X, Saeed M, *et al.* Genome-wide analysis of codon usage bias in four sequenced cotton species[J]. *PLoS One*, 2018, 13(3): e0194372.
- [28] Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Ann Harikrishna J. Codon usage and codon pair patterns in non-grass monocot genomes[J]. *Ann Bot*, 2017, 120(6): 893–909.
- [29] Nair RR, Nandhini MB, Sethuraman T, Doss G. Mutational pressure dictates synonymous codon usage in freshwater unicellular α -cyanobacterial descendant *Paulinella chromatophora* and β -cyanobacterium *Synechococcus elongatus* PCC6301[J]. *Springer Plus*, 2013, 2: 492.
- [30] Kannaujiya VK, Rastogi RP, Sinha RP. GC constituents and relative codon expressed amino acid composition in cyanobacterial hycobiliproteins[J]. *Gene*, 2014, 546(2): 162–171.
- [31] Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes[J]. *Mol Biol Rep*, 2009, 36(8): 2039–2046.
- [32] Zhang WJ, Zhou J, Li ZF, Wang L, Gu X, Zhong Y. Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. [J]. *J Integr Plant Biol*, 2007, 49(2): 246–254.

(责任编辑：周 媛)