

兰科 *ABI3* 基因的密码子偏性分析

龚婷婷, 李维仪, 陶媛, 李少凡, 王嘉文, 杨柏云, 罗火林*

南昌大学生命科学学院, 江西南昌 330031

摘要: 密码子偏性是影响外源基因表达的重要因素, 也是反映基因家族进化的关键线索之一。兰科植物是植物保护中的“旗舰”类群, 且多数物种具有药用和观赏价值, 但种子不易萌发, 而 *ABI3* (abscisic acid insensitive 3) 基因是 ABA 信号转导的中心调控因子, 是调控植物种子萌发的关键因子。因此, 深入研究兰科 *ABI3* 基因密码子偏性, 可为解析该基因功能、优化密码子组成奠定基础。本研究以拟南芥 *ABI3* 基因为探针, 从公共数据库 (NCBI) 筛选出 45 条兰科植物 *ABI3* 基因, 通过 CodonW、SPSS、MEGA 等生物信息学程序对其进行密码子使用模式和偏好性分析。相关参数分析结果表明, 该基因的 ENC (effective number of codon) 值范围分布为 40.84~58.46, 其中大于 52 的占 82.2%, CAI (codon adaptation index) 平均值为 0.203, 远小于 1, 表明 *ABI3* 基因的密码子偏性相对较弱; 中性绘图分析显示, GC_{12} 与 GC_3 的回归曲线的斜率为 0.6103, $R^2=0.7928$, $P<0.05$, 有极高的相关性, 这暗示着碱基组成对密码子偏性具有重要的影响; ENC-plot 分析中, 所有基因 ENC 值都低于期望值, 但相差不大, 说明兰科 *ABI3* 基因密码子主要受碱基突变的影响, 同时也受选择压力等其他因素影响。奇偶偏好性分析显示, 大多数基因分布在平面图的右侧区域, 密码子第 3 位 A/T (U) 的使用频率高于 G/C。通过 RSCU 和 $\Delta RSCU$ 分析确定了 26 个最优密码子, 其中以 A/T (U) 结尾的密码子有 14 个, 以 T (U) 结尾的密码子有 10 个, 说明兰科植物 *ABI3* 基因相对偏好使用 A/U 结尾的密码子, 尤其偏好使用 T (U) 结尾的密码子, 这与大多数单子叶植物密码子使用偏好不同。本研究可为兰科植物 *ABI3* 基因的系统发育、功能解析、提高 *ABI3* 基因表达效率提供理论基础, 同时可为促进植物密码子生物学研究及外源基因改良提供参考。

关键词: 兰科; *ABI3* 基因; 密码子偏性

中图分类号: Q943 文献标识码: A

Analysis of Codon Usage Bias of *ABI3* Gene in Orchidaceae

GONG Tingting, LI Weiyi, TAO Yuan, LI Shaofan, WANG Jiawen, YANG Boyun, LUO Huolin*

School of Life Science, Nanchang University, Nanchang, Jiangxi 330031, China

Abstract: Codon bias is an important factor affecting gene expression, and it is also one of the key clues to reflect gene evolution. Orchidaceae plants are the “flagship” group in plant protection, and most species have medicinal and ornamental values, but seeds are not easy to germinate. *ABI3* (abscisic acid insensitive 3) gene is the central regulator of ABA signal transduction, involved in seed development and dormancy processes, and is the key factor to regulate plant seed germination. Therefore, in-depth study of codon bias of orchid *ABI3* gene can lay a foundation for analyzing the function of the gene and optimizing the codon composition. Using *Arabidopsis thaliana ABI3* gene as a probe, 45 Orchidaceae *ABI3* genes were screened from the public database (NCBI), and the codon use patterns and preferences were analyzed by bioinformatics programs such as CodonW, SPSS and MEGA. The ENC (effective number of codon) value of the genes ranged from 40.84 to 58.46, of which 82.2% was greater than 52. The average value of CAI (codon adaptation index) was 0.203, far less than 1, indicating that the codon usage bias of *ABI3* gene was relatively weak. Neutral plot analysis showed that the slope of the regression curve between GC_{12} and GC_3 was 0.6103, $R^2=0.7928$ ($P<0.05$), indicating that the base composition has an important influence on codon bias. In ENC plot analysis, the ENC value of all genes was lower

收稿日期 2022-04-24; 修回日期 2022-05-26

基金项目 国家自然科学基金项目 (No. 32160720); 兰科植物保护与利用国家林业和草原局重点实验室开放基金项目 (No. OC202101)。

作者简介 龚婷婷 (2002—), 女, 本科生, 研究方向: 兰科植物生物信息分析。*通信作者 (Corresponding author): 罗火林 (LUO Huolin), E-mail: 572460991@163.com。

than the expected value, but the difference was small, indicating that the codon of orchid *ABI3* gene is mainly affected by base mutation and other factors such as selection pressure. Parity preference analysis showed that most genes were distributed in the right region of the plan, and the use frequency of A/T (U) at codon 3 was higher than that of G/C. Through RSCU and Δ RSCU analysis, 26 optimal codons were identified, including 14 codons ending in A/T (U) and 10 codons ending in T (U), indicating that *ABI3* gene of orchid plants relatively prefers to use codons ending in A/T (U), especially codons ending in T (U), which is different from that of most monocotyledonous plants. This study may provide a theoretical basis for the phylogenetic and functional elucidation of *ABI3* genes in Orchidaceae and for improving *ABI3* gene expression efficiency, as well as facilitating codon biology in plants and the improvement of exogenous genes.

Keywords: Orchidaceae; *ABI3* gene; codon usage bias

DOI: 10.3969/j.issn.1000-2561.2023.02.005

编码同一氨基酸的不同密码子称同义密码子, 在不存在外界压力或者其他干扰因素条件下, 同义密码子的使用频率是相同的。但是实际情况下, 基因在编码氨基酸时同义密码子的使用频率存在一定的倾向, 这种现象称为密码子偏好性^[1]。生物在长时间的进化过程中由于基因突变和外界环境的影响会形成自身偏好的密码子系统, 当某一密码子在编码氨基酸受阻时, 它的同义密码子能够代替其行使功能从而保证蛋白质合成的准确性^[2]。早在 20 世纪 70 年代, 在对基因的异源表达进行研究时, 人们就意识到密码子偏性的重要性^[3]。密码子偏好性广泛存在于生物界中, 不仅可以用来描述基因进化规律、基因表达过程中的调控机制, 也可用于预测外源基因的最适宿主, 通过密码子优化提高宿主中外源基因的表达量以及评估基因的表达水平等^[4-5]。研究表明, 原核生物和真核生物中均存在密码子使用偏好性。目前已有对辣椒^[6]、芒果^[7]、陆地棉^[8]等高等植物进行了密码子偏性分析。分析生物对同义密码子的使用情况特别是偏好性可为该生物的外源基因插入及表达提供一定的理论依据^[9]。

兰科 (Orchidaceae) 植物作为被子植物的大科之一, 是植物保护中的“旗舰”类群^[10]。此外, 很多兰科植物具有极高的观赏价值和药用作用^[11], 受世人的喜爱与追捧。然而其种子细小无胚乳, 自然条件下需与适宜的菌根真菌共生才能萌发, 萌发率低且萌发时间漫长^[12]。为了做好兰科植物资源的保护与利用工作, 针对其种子萌发的研究也引起越来越多的关注。*ABI3* 是 ABA 信号转导的中心调控因子, 调节胚胎成熟和幼苗早期发育之间的过渡, 是种子成熟所必需的^[13]。目前国内对外对兰科萌发基因 *ABI3* 基因研究较少。本研究通过 CodonW、MEGA 等程序对兰科 *ABI3* 基因密码子进行偏性分析, 以期优化密码子来提高外源基

因的表达量, 然后对外源基因进行密码子改造, 提高兰科种子萌发率, 为兰科植物种子萌发的分子育种提供理论依据。

1 材料与amp;方法

1.1 材料

以拟南芥 *ABI3* 的基因序列 (登录号为 NC_003074.8) 为探针, 在 NCBI 的数据库中, 通过 blastn 软件查找同源的序列, 获得的序列经 CD-HIT (<http://www.bioinformatics.org/cd-hit/>) 在线软件去冗余后, 选择包含 300 bp 以上 CDS (coding DNA sequence) 的序列用于本研究 (表 1)。

1.2 方法

1.2.1 中性绘图分析 以 GC_3 为横坐标, GC_{12} (GC_1 和 GC_2 的平均值) 为纵坐标绘图, 解析密码子 3 个位置碱基组成的相关性。若 GC_{12} 与 GC_3 值无显著性差异, 则所有点将沿对角线分布, 说明主要来自碱基突变, 外界选择影响弱或者无; 反之, 则说明选择效应是基因最主要的进化约束力。

1.2.2 有效密码子数 (effective number of codons, ENC) 绘图分析 ENC 值反映同义密码子的使用偏好, ENC 值范围为 20~61, 当 ENC=20 时表示密码子偏性最强; ENC=61 表示同义密码子被平均使用, 不存在密码子偏性。ENC 绘图分析主要反映 ENC 与 GC_3 之间的分布关系, 标准曲线 $ENC=2+GC_3+29/(2\times GC_3^2-2\times GC_3+1)^{[14-15]}$ 。当无自然选择压力影响时, 密码子偏性仅由碱基突变产生, 散点则位于曲线之上, 如散点远离曲线时, 则表示自然选择压力更多地参与了同义密码子的使用频率^[16]。

1.2.3 PR2-plot 分析 (parity rule 2, PR2) 密码子奇偶偏好分析主要分析密码子第三位的奇偶偏性。理论上 DNA 的 2 条互补链, 如没有发生突变则 $A=T$, $G=C$ 。分别以基因 $A3/(A3+T3)$,

表 1 兰科植物 *ABI3* 基因序列信息
Tab. 1 Sequence information of *ABI3* genes
in Orchidaceae

序号 No.	物种 Species	基因登录号 Gene Accession
1	<i>Dendrobium catenatum</i> Lindl.	XM_028692103.1
2	<i>Phalaenopsis equestris</i> (Schauer) Rchb. f.	XM_020721254.1
3	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020721253.1
4	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020721314.1
5	<i>D. catenatum</i> Lindl.	XM_028692102.1
6	<i>P. equestris</i> (Schauer) Rchb. f.	XR 002293779.1
7	<i>D. catenatum</i> Lindl.	XM 020846725.1
8	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020736089.1
9	<i>D. catenatum</i> Lindl.	XM_020845421.2
10	<i>D. catenatum</i> Lindl.	XM_020825921.2
11	<i>D. catenatum</i> Lindl.	XM_020825920.2
12	<i>D. catenatum</i> Lindl.	XM 020827036.1
13	<i>P. equestris</i> (Schauer) Rchb. f.	XM 020721402.1
14	<i>P. equestris</i> (Schauer) Rchb. f.	XM 020721400.1
15	<i>P. equestris</i> (Schauer) Rchb. f.	XM 020721401.1
16	<i>D. catenatum</i> Lindl.	XM 020827037.1
17	<i>D. catenatum</i> Lindl.	XM_020827035.2
18	<i>D. catenatum</i> Lindl.	XM_020827032.2
19	<i>D. catenatum</i> Lindl.	XM_020827036.2
20	<i>D. catenatum</i> Lindl.	XM_020827030.2
21	<i>D. catenatum</i> Lindl.	XM_020827033.2
22	<i>P. equestris</i> (Schauer) Rchb. f.	XR 002295213.1
23	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740565.1
24	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740564.1
25	<i>P. equestris</i> (Schauer) Rchb. f.	XM 020716201.1
26	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740563.1
27	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740560.1
28	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740562.1
29	<i>D. catenatum</i> Lindl.	XM_020848014.2
30	<i>D. catenatum</i> Lindl.	XM_020848005.2
31	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020740561.1
32	<i>D. catenatum</i> Lindl.	XM_020841375.2
33	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020723903.1
34	<i>D. catenatum</i> Lindl.	KX130300.1
35	<i>D. catenatum</i> Lindl.	XM_020830439.2
36	<i>P. equestris</i> (Schauer) Rchb. f.	KX215852.1
37	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020716458.1
38	<i>D. catenatum</i> Lindl.	XM_020821301.2
39	<i>Apostasia odorata</i> Bl.	KX156926.1
40	<i>P. aphrodite</i> subsp. <i>formosana</i>	AJ888299.1
41	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020722242.1
42	<i>P. equestris</i> (Schauer) Rchb. f.	XM_020722239.1
43	<i>D. catenatum</i> Lindl.	XM 020828654.1
44	<i>P. equestris</i> (Schauer) Rchb. f.	XM 020733863.1
45	<i>D. catenatum</i> Lindl.	XM 020830815.1

G3/(G3+C3)为横纵坐标做散点图, 图中中心 A=T 且 C=G 横纵坐标均为 0.5, 从该中心出发的矢量表示奇偶性的方向与程度^[17]。

1.2.4 最优密码子分析 相对同义密码子使用度 (relative synonymous codon usage, RSCU) 是指对于某一特定的密码子在编码对应氨基酸的同义密码子间的相对概率, 在某一密码子与其编码相同氨基酸的多种同义密码子之间 (同一基因序列上)^[18], 是某一个密码子使用的频率与平均使用期望值的比值, 它去除了氨基酸内部组成对密码子使用的影响^[19]。如果密码子的使用没有偏好性, 该密码子的 RSCU=1; 若 RSCU>1, 表示密码子偏性较强, 使用频率相对较高; 若 RSCU<1, 表示基因偏性较弱, 基因低表达。由此可以确定基因的最优密码子, 并进一步判断基因密码子偏好 A/T 或 C/G 结尾。

将 ENC 值由大到小排序, 分别将前 10% 和后 10% 的基因设为高表达基因库和低表达基因库, 计算 Δ RSCU 值。最优密码子确定: 高表达基因库中 RSCU>1; 低表达基因库中 RSCU<1; Δ RSCU ≥ 0.2 ^[20-21]。

1.3 数据处理

将获得的序列整理后, 用 CodonW 软件对 45 条序列进行密码子组成和偏好性参数分析; 在 Excel 软件中对运行获得密码子第 1、2、3 位上的 GC 含量, 各碱基 ENC、CAI 及 RSCU 等值进行整理; 运用 SPSS13.0、MEGA5.1 软件进行数据分析和绘图分析。

2 结果与分析

2.1 兰科 *ABI3* 基因的碱基组成及相关性分析

利用 CodonW、MEGA 软件获得并计算 *ABI3* 基因编码区序列不同位置的 GC 含量、ENC 值和 CAI 值(表 2)。总 GC(GC_{all})含量为 42.8%~65.9%, 平均值为 49.45%, GC₁ 为 43.8%~68.7%, 平均值为 50.58%, GC₂ 为 41.8%~61.4%, 平均值为 48.48%, GC₃ 为 41.1%~67.6%, 平均值为 49.33%, 接近 50%, 说明其在编码氨基酸碱基选择上对于 G/C 和 A/T 没有明显的偏好性。同样地, ENC 值范围在 40.84~58.46, 平均值为 49.81; CAI 平均值为 0.203, 远小于 1, 这些证据都表明这类基因的密码子偏性较弱。

表 2 兰科植物 *ABI3* 基因编码区密码子不同位置 GC 含量
Tab. 2 GC content of different positions of codon in *ABI3* genes of Orchidaceae

基因序号 Gene No.	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC _{all} /%	ENC	CAI
1	47.6	46.9	46.8	47.1	51.18	0.198
2	51.1	47.9	50.4	49.8	52.94	0.196
3	51.3	47.9	50.2	49.7	52.85	0.196
4	47.8	50.2	49.7	49.3	57.01	0.223
5	52.3	48.4	49.5	50.1	53.03	0.208
6	51.0	48.1	52.1	50.4	52.99	0.200
7	52.9	49.3	50.8	51.0	53.06	0.211
8	50.7	54.9	46.6	50.8	55.02	0.167
9	46.4	50.5	43.7	46.9	54.46	0.172
10	44.4	46.1	46.2	45.6	52.15	0.203
11	44.4	46.2	46.2	45.6	52.18	0.204
12	44.3	45.9	42.6	44.2	52.96	0.186
13	51.5	50.8	45.8	49.3	58.40	0.197
14	51.1	50.8	45.7	49.2	58.43	0.197
15	51.6	50.7	45.7	49.4	58.46	0.198
16	45.1	45.6	40.8	43.9	52.56	0.186
17	44.7	46.1	41.9	44.2	53.28	0.189
18	44.7	46.4	41.9	44.3	52.81	0.186
19	44.9	45.7	41.1	43.9	52.16	0.182
20	44.5	46.2	42.1	44.3	52.87	0.186
21	44.7	46.0	42.6	44.4	53.16	0.185
22	43.8	43.2	41.4	42.8	52.56	0.185
23	43.8	43.2	41.4	42.8	52.56	0.185
24	44.0	43.4	41.6	43.1	52.92	0.182
25	45.3	45.4	44.4	45.0	53.80	0.207
26	44.2	42.8	41.6	42.9	53.17	0.188
27	43.7	42.6	41.1	42.5	52.87	0.190
28	43.9	42.6	41.4	42.6	52.78	0.182
29	46.5	50.0	44.6	47.0	53.92	0.187
30	46.1	50.1	44.4	46.8	53.99	0.187
31	43.8	41.8	41.4	42.3	52.97	0.193
32	60.9	56.1	59.3	58.7	40.84	0.231
33	60.7	54.0	59.0	57.9	44.88	0.198
34	61.4	52.1	59.1	57.5	52.12	0.210
35	61.1	51.8	59.1	57.3	52.31	0.210
36	63.9	54.8	62.1	60.2	54.67	0.197
37	63.9	54.8	62.1	60.2	54.67	0.197
38	50.2	43.4	53.5	49.0	53.88	0.221
39	68.7	61.4	67.6	65.9	41.94	0.243
40	53.0	49.6	59.7	54.0	43.46	0.253
41	52.4	48.2	58.2	52.9	46.71	0.252
42	51.6	48.4	56.7	52.2	47.64	0.257
43	57.6	50.1	63.2	57.0	44.60	0.242
44	60.8	49.3	56.8	55.6	54.51	0.208
45	57.6	52.1	57.8	55.8	54.67	0.266
平均	50.58	48.48	49.33	49.45	49.81	0.203

2.2 中性绘图分析

ABI3 基因的中性绘图分析结果显示(图 1), 45 个 *ABI3* 基因坐标点几乎均沿着图中对角线分布, 有些甚至位于对角线上, 表明 *ABI3* 基因主要受碱基突变的影响, 来自外界的选择压力较小。 GC_{12} 的范围为 43.50%~65.05%, GC_{12} 与 GC_3 的回归曲线的斜率为 0.6103, $R^2=0.7928$, $P<0.05$, 有极高的相关性, 说明碱基组成对密码子偏性在一定程度上受碱基组成的影响。

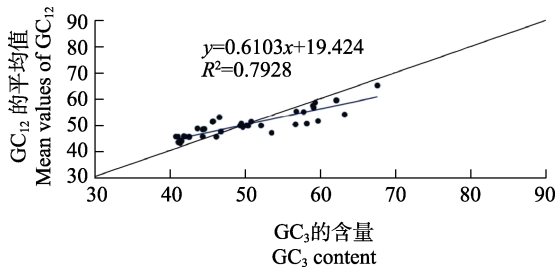


图 1 中性绘图分析
Fig. 1 Neutrality plot analysis

2.3 基因密码子使用参数之间的相关性分析

对基因 ENC 值以及密码子不同位置碱基组成情况进行相关性分析, 分析结果如表 3 所示, GC_1 、 GC_2 、 GC_3 任意二者之间均存在极显著正相关性 ($P<0.01$), 说明各基因密码子三位碱基的组成之间密切相关, 受随机突变或者有一定方向突变压力的影响较大; 这暗示着 3 个位点的 GC 值 (GC_1 、 GC_2 、 GC_3) 都可能对密码子偏性产生一定程度的影响; 而 ENC 值与 GC_1 存在显著负相关性 ($P<0.05$), 与 GC_2 、 GC_3 和基因总的 GC 含量 (GC_{all}) 都存在极显著负相关性 ($P<0.01$), 与基因编码氨基酸的长度 (Laa) 存在极显著正相关性。说明在基因密码子中, 第 3 位碱基组成对密码子使用模式存在较大的影响。由参数间的相关性可知, 同义密码子碱基所在位置及含量对基因密码子使用偏好性的影响从大到小排序: $GC_3>GC_{all}>GC_1>GC_2$ (密码子使用偏好性越强, ENC 值越小)。

2.4 ENC 与 GC_3 的关联分析

以各基因 ENC 值为纵坐标, GC_3 值为横坐标做散点图(图 2)。*ABI3* 基因 GC_3 值 41.1%~67.6% 之间, 平均值为 49.33%, 表明密码子第 3 位碱基 GC 变化频率较大。ENC 值在 40.84~58.46 之间, 说明不同基因密码子偏性相差较大。部分分布在标准曲线周围的 *ABI3* 基因密码子偏好性主要由

表 3 各基因相关参数的相关性分析
Tab. 3 Correlation analysis of each gene's related parameters

项目 Item	GC_1	GC_2	GC_3	GC_{all}	ENC
GC_2	0.839**				
GC_3	0.927**	0.739**			
GC_{all}	0.983**	0.877**	0.962**		
ENC	-0.371*	-0.288*	-0.534**	-0.441**	
Laa	-0.681**	-0.477**	-0.744**	-0.696**	0.449**

注: *表示显著相关 ($P<0.05$); **表示极显著相关 ($P<0.01$)。
Notes: * indicates significant correlation ($P<0.05$); ** indicates extremely significant correlation ($P<0.01$).

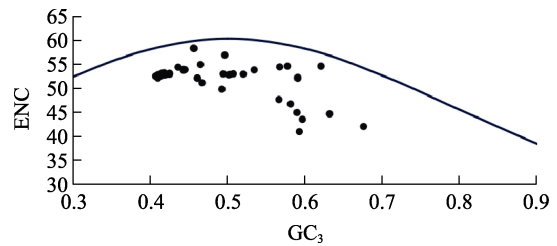


图 2 ENC 与 GC_3 关联分析
Fig. 2 Analysis of ENC and GC_3 relationship

突变造成; 而在偏离期望曲线位点基因密码子偏好性受自然选择影响大; 没有一个点落在曲线上, 说明不存在完全由碱基突变的密码子偏性情况。

2.5 奇偶偏好分析

密码子奇偶偏好性主要用于分析密码子第 3 位发生奇偶偏性, 图中中心横纵坐标均是 0.5, $A=C$ 且 $G=C$, 表示 DNA 两条互补链之前不存在突变。由 PR2 分析图(图 3)可知, 大多数基因 $A_3/(A_3+T_3)$ 或 $G_3/(G_3+C_3)$ 值都偏离了 0.5, 只有个别基因接近 0.5, 表明该基因密码子偏性受基因突变影响大, 其他因素的影响依然存在, 例如自然选择等。从横坐标来看, 除个别基因小于 0.5, 大多数大于 0.5, 表明第 3 位上 A 的比例大于 T (U), 密码子在 A 和 T (U) 之间具有一定的偏

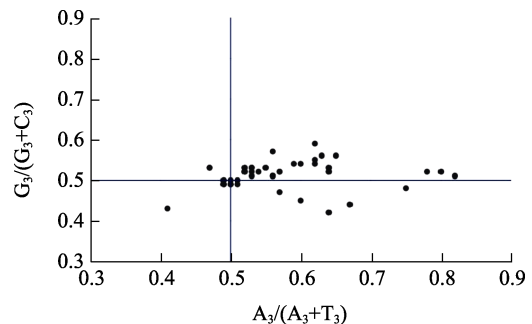


图 3 PR2 bias-plot 分析图
Fig. 3 Analysis of PR2 bias-plot

好性；从纵坐标来看，大多数大于 0.5，表明第 3 位上 G 的比例大于 C。综上所述进一步说明了 *ABI3* 基因密码子偏性除了主要受基因突变影响以外还受其他因素的影响。

2.6 *ABI3* 基因相对同义密码子使用度分析 (RSCU)

通过 CodonW 软件计算 45 条 *ABI3* 基因序列的 RSCU 值 (表 4)。结果显示，RSCU=1 的密

码子有 4 个 (AAC、UGG、GCC、GCA)，表明这几个同义密码子无偏性被随机使用；RSCU>1 的密码子有 28 个 (占全部密码子的 46.67%)，其中以 T (U) 结尾的有 12 个，A 结尾的有 7 个，C 结尾的有 5 个，G 结尾的有 4 个，以 A/T (U) 结尾的占比 67.86%，表明兰科 *ABI3* 基因偏爱 A/T (U) 结尾的密码子，显著偏爱 T (U) 结尾的密码子。

表 4 兰科植物 *ABI3* 基因编码序列相对同义密码子使用度
Tab. 4 RSCU analysis of *ABI3* gene coding region in Orchidaceae

氨基酸 AA	密码子 Codon	RSCU	氨基酸 AA	密码子 Codon	RSCU	氨基酸 AA	密码子 Codon	RSCU
Phe	UUU	<u>1.01</u>	Tyr	UAU	<u>1.11</u>	Ser	UCU	<u>1.48</u>
	UUC	0.99		UAC	0.89		UCC	<u>1.15</u>
Leu	UUA	0.68	His	CAU	<u>1.19</u>	UCA	<u>1.24</u>	
	UUG	<u>1.04</u>		CAC	0.81	UCG	0.62	
	CUU	<u>1.53</u>	Pro	CCU	<u>1.45</u>	AGU	0.59	
	CUC	<u>1.40</u>		CCC	0.80	AGC	0.92	
	CUA	0.61		CCA	<u>1.10</u>	Cys	UGU	0.72
CUG	0.75	CCG	0.64	UGC	<u>1.23</u>			
Ile	AUU	<u>1.42</u>	Thr	ACU	<u>1.26</u>	Arg	CGU	0.86
	AUC	0.87		ACC	0.99		CGC	<u>1.37</u>
	AUA	0.71		ACA	<u>1.20</u>		CGA	0.63
Val	GUU	<u>1.42</u>	Asn	ACG	0.55	Gln	CGG	0.75
	GUC	0.96		AAU	<u>1.01</u>		CAA	<u>1.07</u>
	GUA	0.42	AAC	1.00	CAG	0.93		
	GUG	<u>1.21</u>	Lys	AAA	0.99	Arg	AGA	<u>1.25</u>
Trp	UGG	1.00		AAG	<u>1.13</u>		AGG	<u>1.14</u>
	Ala	GCU	<u>1.30</u>	Asp	GAU	<u>1.23</u>	Gly	GGU
GCC		1.00	GAC		0.77	GGC		<u>1.07</u>
GCA		1.00	Glu	GAA	<u>1.03</u>	GGA		<u>1.41</u>
GCG	0.60	GAG		0.97	GGG	0.7		

注：下划线的数据表示 RSCU>1，该密码子使用频率相对较高。

Note: The codon with underline means RSCU>1 and shows that the codon is used relatively frequently.

2.7 最优密码子分析

通过 ENC 值大小排序，取前后 10% 值作为高低表达基因库，计算 Δ RSCU 值 (表 5)，确定最优密码子。最终确定了 *ABI3* 基因编码 12 个氨基酸的 26 个最优密码子，其中有编码 12 个氨基酸的 22 个密码子 Δ RSCU \geq 0.5，G/C 结尾的有 12 个，以 A/T (U) 结尾的有 14 个，其中以 U 结尾的密码子有 10 个，表明兰科植物 *ABI3* 基因偏好以 A/T (U) 结尾的最优密码子，尤其偏爱以 T (U) 结尾的密码子。

3 讨论

密码子是基因传递信息中的基本结构单位，

自然界生物基因组中密码子偏性使用现象广泛存在^[22]。不同物种在自然选择中生存下来，密码子存在一定差异并且有其偏好性，因此对密码子偏性研究可反映一定的进化特征^[23]。作为生物碱基组成重要指标之一，GC 含量可反映密码子使用偏性突变的整体趋势^[24]，密码子第 3 位 GC₃ 含量通常可作为分析密码子偏性的主要衡量标准之一。ENC 值可表示偏性的强弱，兰科植物 *ABI3* 基因的 ENC 平均值为 49.81，表明在兰科植物中大部分 *ABI3* 基因密码子的偏性较弱。通过 RSCU 和 Δ RSCU 分析，发现 *ABI3* 基因主要偏爱使用以 A/T (U) 结尾的密码子，与 KUMAR 等^[25]研究马流感病毒，陈哲等^[26]研究菠萝和 WU 等^[27]研究

表 5 兰科植物 *ABI3* 基因最优密码子
Tab. 5 Preferred codons in *ABI3* genes of Orchidaceae

氨基酸 AA	密码子 Codon	Δ RSCU	氨基酸 AA	密码子 Codon	Δ RSCU	氨基酸 AA	密码子 Codon	Δ RSCU
Leu	UUG ^{***}	0.890	Gly	GGA ^{***}	0.550	Ser	UCU ^{***}	0.544
Leu	CUU ^{***}	0.650	Pro	CCC ^{***}	1.124	Ser	UCC ^{***}	1.768
Leu	CUC ^{***}	2.202	Pro	CCA ^{***}	0.502	Ser	UCG ^{**}	0.482
Ile	AUU ^{***}	0.688	Pro	CCG ^{**}	0.414	Arg	CGC ^{***}	2.214
Val	GUU ^{**}	0.322	Pro	CCU [*]	0.296	Arg	AGA ^{***}	1.054
Val	GUC ^{***}	0.808	Thr	ACU ^{***}	0.552	Arg	AGG ^{**}	0.498
Ala	GCU ^{***}	0.836	Thr	ACA ^{***}	0.702	His	CAU ^{***}	0.826
Tyr	UAC ^{***}	0.952	Asp	GAU ^{***}	1.024	His	CAC ^{***}	0.826
Tyr	UAU ^{***}	0.954	Asp	GAC ^{***}	1.024			

Note: * indicates Δ RSCU ≥ 0.2 ; ** indicates Δ RSCU ≥ 0.3 ; *** indicates Δ RSCU ≥ 0.5 .

的芍药密码子偏性使用一致,尤其偏爱以 T(U) 结尾的密码子;与其他多数单子叶植物主要偏爱 C/G 结尾的密码子不同,如小麦 *CBL4* 基因密码子^[28],香蕉 *Aux/IAA* 基因家族密码子^[29]等。在对多种植物基因组密码子进行偏好性分析中发现,双子叶植物的 GC 含量最低,而单细胞藻类 GC 含量最高,说明密码子使用偏性的程度可能随物种进化的程度而降低^[30]。

影响密码子偏性的因素主要包括:碱基组成差异、自然选择、tRNA 丰富度、翻译效率、蛋白质功能,选择突变漂变学说等^[31]。通过中性绘图分析、ENC 与 GC₃ 关联分析和奇偶偏好分析,说明兰科植物 *ABI3* 基因受碱基突变影响较大,同时也受到其他因素的影响,如选择压力。植物基因的密码子偏性形成过程中,一般受由碱基突变和自然选择等多重影响,如杨祥燕等^[32]研究的菠萝叶绿体基因密码子、ZHOU 等^[33]研究的一些种子植物叶绿体基因密码子、梁娥等^[34]研究的竹节参转录组密码子、刘慧等^[35]研究的糜子叶绿体基因组密码子;有时也会其中一个占主导作用,如马盛超等^[18]研究的蓼科大黄属植物 *CHS* 基因,主要受自然选择影响。另外,对基因 ENC 值以及密码子不同位置碱基组成情况进行相关性分析发现,GC₁、GC₂、GC₃ 显著相关,说明同义密码子 3 个位置上碱基含量接近,受随机突变或者有一定方向突变压力的影响较大;*ABI3* 基因同义密码子碱基所在位置及含量均对基因密码子使用偏好性产生一定影响。

由于兰科植物自然生境的破坏和资源的过度开采,野生兰科植物极度濒危。濒危兰科植物的再生和合理利用有赖于种苗繁殖技术的蓬勃发

展,而该技术的核心问题之一便是种子萌发。*ABI3* 是种子成熟所必需的,它调节胚胎成熟和幼苗早期发育之间的过渡,被视为假定的种子特异性转录激活剂^[36]。基因功能研究表明该基因可有效提高种子萌发效率。因此,通过遗传转化技术,提高兰科植物 *ABI3* 基因的表达水平,是解决兰科植物种子萌发障碍的重要途径之一。本研究通过分析 *ABI3* 密码子偏性并筛选出 26 个最优密码子,为优化育种提供理论依据,在今后 *ABI3* 基因转化兰科植物时,可以对外源基因序列进行适当改造,然后转化至兰科植物中,有望提高其表达效率,更有利于培育出种子易萌发的兰科花卉。

参考文献

- [1] 胡桂兵,张上隆,徐昌杰,林顺权. 不同种类柑橘的密码子用法分析[J]. 华南农业大学学报, 2006, 27(1): 13-16.
HU G B, ZHANG S L, XU C J, LIN S Q. Analysis of codon usage between different *Citrus* species[J]. Journal of South China Agricultural University, 2006, 27(1): 13-16. (in Chinese)
- [2] YENGGHONM S, UDDIN A, CHAKRABORTY S. Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. *assamica* and *Camellia sinensis* var. *sinensis* in comparison to *Camellia pubicosta*[J]. Journal of Integrative Agriculture, 2019, 18(12): 2771-2785.
- [3] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression[J]. Trends in Biotechnology, 2004, 22(7): 346-353.
- [4] WANG X X, WU L, ZHOU P, ZHU S F, AN W, CHEN Y, ZHAO L. Analysis of synonymous codon usage patterns in the genus *Rhizobium*[J]. World Journal of Microbiology and

- Biotechnology, 2013, 29(11): 2015-2024.
- [5] 胡振民, 万青, 李欢, 李荣林, 王枫, 杨亦扬. 茶树 *CsNRT1.1* 基因密码子使用特性分析[J]. 江苏农业学报, 2019, 35(4): 896-903.
HU Z M, WAN Q, LI H, LI R L, WANG F, YANG Y Y. Analysis of codon usage features of *CsNRT1.1* gene in *Camellia sinensis*[J]. Jiangsu Journal of Agricultural Sciences, 2019, 35(4): 896-903. (in Chinese)
- [6] 刘潮, 韩利红, 吴丽芳, 代小波, 刘婕. 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.
LIU C, HAN L H, WU L F, DAI X B, LIU J. Genome-wide codon usage bias analysis of *Capsicum annuum*[J]. Jiangsu Agricultural Sciences, 2022, 50(5): 16-22. (in Chinese)
- [7] 唐玉娟, 赵英, 黄国弟, 付海天, 宋恩亮, 李日旺, 金刚. 芒果叶绿体基因组密码子使用偏好性分析[J]. 热带作物学报, 2021, 42(8): 2143-2150.
TANG Y J, ZHAO Y, HUANG G D, FU H T, SONG E L, LI R W, JIN G. Analysis on codon usage bias of chloroplast genes from mango[J]. Chinese Journal of Tropical Crops, 2021, 42(8): 2143-2150. (in Chinese)
- [8] 尚明照, 刘方, 华金平, 王坤波. 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.
SHANG M Z, LIU F, HUA J P, WANG K B. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*[J]. Scientia Agricultura Sinica, 2011, 44(2): 245-253. (in Chinese)
- [9] 雷慧, 李鸽, 王娜玉. 金莲花叶绿体基因组密码子偏好性分析[J]. 山西农业科学, 2019, 47(8):1300-1305, 1445.
LEI H, LI G, WANG N Y. Analysis of codon usage bias in the chloroplast genome of *Trollius chinensis* Bunge[J]. Journal of Shanxi Agricultural Sciences, 2019, 47(8): 130-1305, 1445. (in Chinese)
- [10] 罗毅波, 贾建生, 王春玲. 中国兰科植物保育的现状和展望[J]. 生物多样性, 2003, (1): 70-77.
LUO Y B, JIA J S, WANG C L. A general review of the conservation status of Chinese orchids[J]. Biodiversity Science, 2003, (1): 70-77. (in Chinese)
- [11] 李南岍, 吴记贵, 蒋万杰, 沐先运, 程瑾. 北京松山国家级自然保护区兰科植物多样性及其保护评价[J]. 植物科学学报, 2013, 31(5): 510-516.
LI N Q, WU J G, JIANG W J, MU X Y, CHENG J. Biodiversity and conservation of orchids in Songshan National Nature Reserve, Beijing[J]. Plant Science Journal, 2013, 31(5): 510-516. (in Chinese)
- [12] 刘思思, 陈娟, 郭顺星. 兰科植物种子萌发的研究进展[J]. 种子, 2015, 34(6): 43-50.
LIU S S, CHEN J, GUO S X. Review on germination of orchid seeds[J]. Seed, 2015, 34(6): 43-50. (in Chinese)
- [13] 冯园. 外源 GA、ABA 对小麦种子活力的影响[D]. 泰安: 山东农业大学, 2015.
FENG Y. Effects of exogenous GA and ABA on wheat seed vigor[D]. Tai'an: Shangdong Agricultural University, 2015. (in Chinese)
- [14] 李蓉. CRISPR/Cas9 介导的文心兰 *Fd* 和 *FNR* 基因编辑研究[D]. 福州: 福建农林大学, 2019.
LI R. Study on CRISPR/Cas9 mediated *Fd* and *FNR* gene editing in *Oncidium hybridum*[D]. Fuzhou: Fujian Agriculture and Forestry University, 2019. (in Chinese)
- [15] WRIGHT F. The effective number of codons used in a gene[J]. Gene, 1990, 87(1): 23-29.
- [16] LU H, ZHAO W M, ZHENG Y, WANG H, QI M, YU X P. Analysis of synonymous codon usage bias in *Chlamydia*[J]. Acta Biochimica et Biophysica Sinica, 2005, 37(1): 1-10.
- [17] SUEOKA N. Translation-coupled violation of parity rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position[J]. Gene, 1999, 238(1): 53-58.
- [18] 马盛超, 余鑫梅, 姚慧鹏. 蓼科大黄属植物 *CHS* 基因密码子偏爱性分析[J]. 分子植物育种, 2017, 15(6): 2171-2178.
MA S C, YU X M, YAO H P. The codon usage analysis in *CHS* gene within the *Polygonaceae rheum*[J]. Molecular Plant Breeding, 2017, 15(6): 2171-2178. (in Chinese)
- [19] 李滢, 匡雪君, 朱孝轩, 朱英杰, 孙超. 长春花密码子使用偏好性分析[J]. 中国中药杂志, 2016, 41(22): 4165-4168.
LI Y, KUANG X J, ZHU X X, ZHU Y J, SUN C. Codon usage bias of *Catharanthus roseus*[J]. China Journal of Chinese Materia Medica, 2016, 41(22): 4165-4168. (in Chinese)
- [20] 方辉, 曲俊杰, 孙嘉曼, 尹玲, 黄羽. 香蕉全基因组 *NBS* 抗病基因密码子使用偏好性分析[J]. 分子植物育种, 2017, 15(3): 883-889.
FANG H, QU J J, SUN J M, YIN L, HUANG Y. The codon usage bias of *NBS* disease resistance genes in whole genome of banana[J]. Molecular Plant Breeding, 2017, 15(3): 883-889. (in Chinese)
- [21] NIE X J, DENG P C, FENG K W, LIU P X, DU X H, FRANK M Y, SONG W N. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family[J]. Plant Molecular Biology Reporter, 2014, 32(4): 828-840.
- [22] 刘潮, 韩利红, 宋培兵, 王德琴, 王海波, 唐利洲. 桑树 *WRKY* 转录因子的全基因组鉴定及生物信息学分析[J]. 南方农业学报, 2017, 48(9): 1691-1699.
LIU C, HAN L H, SONG P B, WANG D Q, WANG H B, TANG L Z. Genome-wide identification and bioinformatics analysis for mulberry *WRKY* transcription factors[J]. Journal

- of Southern Agriculture, 2017, 48(9): 1691-1699. (in Chinese)
- [23] 时慧, 王玉, 杨路成, 丁兆堂. 茶树抗寒调控转录因子 *ICE1* 密码子偏性分析[J]. 园艺学报, 2012, 39(7): 1341-1352.
SHI H, WANG Y, YANG L C, DING Z T. Analysis of codon bias of the cold regulated transcription factor *ICE1* in tea plant[J]. Acta Horticulturae Sinica, 2012, 39(7): 1341-1352. (in Chinese)
- [24] ZHOU M, LONG W, LI X. Analysis of synonymous codon usage in chloroplast genome of *Populus alba*[J]. Journal of Forestry Research, 2008, 19(4): 293-297.
- [25] KUMAR N, BERA B C, GREENBAUM B D, BHATIA S, SOOD R, SELVARAJ P, ANAND T, TRIPATHI B N, VIRMANI N. Revelation of influencing factors in overall codon usage bias of *Equine influenza viruses*[J]. PLoS One, 2016, 11(4): e0154376.
- [26] 陈哲, 胡福初, 王祥和, 范鸿雁, 张治礼. 菠萝密码子使用偏好性分析[J]. 果树学报, 2017, 34(8): 946-955.
CHEN Z, HU F C, WANG X H, FAN H Y, ZHANG Z L. Analysis of codon usage bias of *Ananas comosus* with genome sequencing data[J]. Journal of Fruit Science, 2017, 34(8): 946-955. (in Chinese)
- [27] WU Y Q, ZHAO D Q, TAO J. Analysis of codon usage patterns in herbaceous peony (*Paeonia lactiflora* Pall.) based on transcriptome data[J]. Genes, 2015, 6(4): 1125-1139.
- [28] 陈文辉, 焦义然, 杨军峰, 刘永伟, 杨帆, 董福双, 王海波, 杜进民, 周硕. 小麦 *CBL4* 基因密码子使用特征分析[J]. 河南农业科学, 2018, 47(8): 11-16.
CHEN W H, JIAO Y R, YANG J F, LIU Y W, YANG F, DONG F S, WANG H B, DU J M, ZHOU S. Analysis of codon bias of *CBL4* gene in *Triticum aestivum*[J]. Journal of Henan Agricultural Sciences, 2018, 47(8): 11-16. (in Chinese)
- [29] 孙雪丽, 郝向阳, 李蓉, 项蕾蕾, 彭丽云, 刘范, 田娜, 程春振, 赖钟雄, 王天池. 香蕉 *Aux/IAA* 基因家族密码子偏好性分析[J]. 分子植物育种, 2019, 17(8): 2467-2475.
SUN X L, HAO X Y, LI R, XIANG L L, PENG L Y, LIU F, TIAN N, CHENG C Z, LAI Z X, WANG T C. Codon bias analysis of *Aux/IAA* gene family in banana[J]. Molecular Plant Breeding, 2019, 17(8): 2467-2475. (in Chinese)
- [30] LI N, LI Y Y, ZHENG C C, HUNAG J G, ZHANG S Z. Genome-wide comparative analysis of the codon usage patterns in plants[J]. Genes & Genomics, 2016, 38(8): 723-731.
- [31] LIU G M, WU J Y, YANG H M, BAO Q Y. Codon usage patterns in *Corynebacterium glutamicum*: mutational bias, natural selection and amino acid conservation[J]. Comp Funct Genomics, 2010 (2010): 343569.
- [32] 杨祥燕, 蔡元保, 谭秦亮, 覃旭, 黄显雅, 吴密. 菠萝叶绿体基因组密码子偏好性分析[J]. 热带作物学报, 2022, 43(3): 439-446.
YANG X Y, CAI Y B, TAN Q L, QIN X, HAUNG X Y, WU M. Analysis of codon usage bias in the chloroplast genome of *Ananas comosus*[J]. Chinese Journal of Tropical Crops, 2022, 43(3): 439-446. (in Chinese)
- [33] ZHOU M, LONG W, LI X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants[J]. Forest Study of China, 2018, 11(4): 235-242.
- [34] 梁娥, 齐敏杰, 丁延庆, 张来. 竹节参转录组使用密码子偏好性分析[J]. 江苏农业科学, 2019, 47(2): 67-71.
LIANG E, QI M J, DING Y Q, ZHANG L. Analysis of codon bias in *Panax japonicus* transcriptome[J]. Jiangsu Agricultural Sciences, 2019, 47(2): 67-71. (in Chinese)
- [35] 刘慧, 王梦醒, 岳文杰, 邢光伟, 葛玲巧, 聂小军, 宋卫宁. 糜子叶绿体基因组密码子使用偏性的分析[J]. 植物科学学报, 2017, 35(3): 362-371.
LIU H, WANG M X, QUE W J, XING G W, GE L Q, NIE X J, SONG W N. Analysis of codon usage in the chloroplast genome of broomcorn millet (*Panicum miliaceum* L.)[J]. Plant Science Journal, 2017, 35(3): 362-371. (in Chinese)
- [36] NAMBARA E, MCCOURT P, NAITO S. A regulatory role for the *ABI3* gene in the establishment of embryo maturation in *Arabidopsis thaliana*[J]. Development, 1995, 121(3): 629-636.