

曾春涛, 梁俊杰, 仇学文, 等. 豇豆 *GATL* 基因家族密码子偏好性分析[J]. 江苏农业学报, 2024, 40(11): 2131-2139.  
doi:10.3969/j.issn.1000-4440.2024.11.017

## 豇豆 *GATL* 基因家族密码子偏好性分析

曾春涛, 梁俊杰, 仇学文, 张舒悦, 程柳洋, 吴才君  
(江西农业大学农学院, 江西 南昌 330000)

**摘要:** 为明确豇豆类似半乳糖醛酸基转移酶(*GATL*)基因密码子使用偏好及其影响因素,本研究对32条豇豆*GATL*基因家族成员进行分析。结果表明,大部分*GATL*基因家族成员密码子偏好以A/T结尾。豇豆*GATL*基因家族的有效密码子数(*ENc*)为39.29~55.77,平均值为51.06,同义密码子第3位G+C的含量(*GC3*)为34.10%~74.60%,表明豇豆*GATL*基因家族密码子使用偏好性较弱。相对同义密码子使用度(*RSCU*)聚类结果与编码序列(CDS)聚类结果基本一致。通过偏倚分析(PR2-plot分析)、*ENc*-plot分析以及中性绘图分析可以看出,豇豆*GATL*基因家族密码子使用偏好主要受自然选择的影响。本研究结果可为豇豆*GATL*的基因表达和功能验证提供理论依据。

**关键词:** 豇豆; *GATL* 基因家族; 相关性分析; 密码子偏好性

中图分类号: S643.4 文献标识码: A 文章编号: 1000-4440(2024)11-2131-09

## Codon usage bias analysis of cowpea *GATL* gene family

ZENG Chuntao, LIANG Junjie, QIU Xuwen, ZHANG Shuyue, CHENG Liuyang, WU Caijun  
(College of Agriculture, Jiangxi Agricultural University, Nanchang 330000, China)

**Abstract:** To clarify the codon usage bias of galacturonic acid transferase-like (*GATL*) genes in cowpea and influencing factors, the 32 members of the cowpea *GATL* gene family were analyzed in this study. The results showed that most of the codons preferred by the *GATL* gene family ended with A/T. The effective number of codons (*ENc*) of the cowpea *GATL* gene family ranged from 39.29 to 55.77, with an average of 51.06, and the contents of G+C at the third position of synonymous codons (*GC3*) ranged from 34.10% to 74.60%, indicating that the codon usage bias of the cowpea *GATL* gene family was weak. The relative synonymous codon usage (*RSCU*) clustering results were basically consistent with the coding sequence (CDS) clustering results. Results of parity rule 2 (PR2)-plot analysis, *ENc*-plot analysis, and neutral plotting analysis showed that the codon usage bias of cowpea *GATL* gene family was mainly affected by natural selection. The results of this study provide a theoretical basis for the gene expression and functional verification of *GATL* in cowpeas.

**Key words:** cowpea; *GATL* gene family; correlation analysis; codon preference

密码子是 mRNA 上的 3 个相邻核苷酸序列,用于传递生物体内遗传信息,指示蛋白质合成<sup>[1]</sup>。密码子具有简并性<sup>[2]</sup>,即大多数氨基酸可以由多个密

码子编码,而编码同一个氨基酸的不同密码子称为同义密码子<sup>[3]</sup>。各同义密码子在编码蛋白质时使用频率的不均衡,称为密码子的使用偏好(Codon usage bias, CUB)<sup>[4]</sup>。已有研究结果表明,密码子的使用偏好是生物为适应环境在长期进化过程中形成的<sup>[5]</sup>。生物对密码子的使用偏好受到多种因素影响,包括自然选择、碱基突变<sup>[6]</sup>。密码子使用偏好与基因表达紧密关联,因此,研究密码子使用偏好对探索物种遗传和变异具有重要意义。在植物体内,

收稿日期:2024-03-12

基金项目:江西省农作物良种联合攻关项目专项基金[赣财农指(2022)7号]

作者简介:曾春涛(2000-),女,贵州贵阳人,硕士研究生,研究方向为蔬菜生理生态与环境调控。(E-mail) schizobucia@126.com

通讯作者:吴才君,(E-mail) wucj12@126.com

类似半乳糖醛酸基转移酶(*GATL*)基因家族参与细胞壁中的果胶及木聚糖等的生物合成<sup>[7]</sup>。在巨桉中,*GATL*基因表达受低温胁迫诱导,与巨桉的逆境响应相关<sup>[8]</sup>。在水稻中,*GATL*基因表达受到低温、干旱和 ABA 诱导,响应非生物逆境胁迫<sup>[9]</sup>。仇学文等<sup>[10]</sup>对豇豆 *OSCA* 基因的密码子使用偏好进行了研究,而豇豆 *GATL* 基因的密码子使用偏好还未见报道。

豇豆 [*Vigna unguiculata* (L.) Walp] 属豆科 (Fabaceae) 豇豆属 (*Vigna* Savi), 为一年生缠绕或近直立草本植物,喜温、耐热、不耐低温。豇豆在中国栽培面积广阔,栽培方式主要为露地栽培和设施栽培<sup>[11]</sup>。极端天气下,露地栽培的豇豆易落花落荚,根系生长发育不良,最终造成豇豆减产。本研究拟对豇豆 *GATL* 基因家族密码子的偏好性进行分析,为培育抗逆性强的豇豆品种提供理论依据,从而解决在极端天气下露地栽培的豇豆产量降低的问题。

## 1 材料与方法

### 1.1 *GATL* 基因家族成员鉴定

将拟南芥的 10 个 *GATL* 基因家族成员与豇豆的基因组进行 BLAST 比对,获得候选序列,将候选序列上传至 pfam 网站 (<http://pfam.xfam.org/>),验证候选序列,删除不合格的,得到如表 1 所示的 32 个豇豆 *GATL* 基因家族成员,利用 TBtools 获取豇豆 *GATL* 基因家族成员的编码序列(CDS),用于进一步分析。

### 1.2 密码子碱基相关参数的获取

使用 Codon W1.4、MEGA 11 软件以及 CUSP 网站 (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>) 分析豇豆 *GATL* 基因家族成员,获得密码子碱基的组成以及同义密码子相对使用度 (*RSCU*) 等相关参数。

### 1.3 相关性分析

把豇豆 *GATL* 基因家族成员的 *T3* (同义密码子第 3 位碱基 T 的含量)、*C3* (同义密码子第 3 位碱基 C 的含量)、*A3* (同义密码子第 3 位碱基 A 的含量)、*G3* (同义密码子第 3 位碱基 G 的含量)、*GC3* (同义密码子第 3 位 G+C 的含量)、*GC* (基因组中 G+C 含量)、*CAI* (密码子适应指数)、*CBI* (密码子偏性指数)、*Fop* (最优密码子使用频率) 和 *ENc* (有效密码子数) 等相关参数输入 SPSS26 中,进行 Spearman 相关性分析,最后使用 Origin2021 进行图像处理。

表 1 豇豆 *GATL* 基因家族成员

Table 1 Members of the cowpea *GATL* gene family

基因名称	基因 ID	长度 (bp)
<i>VuGATL1</i>	Vigun01g139300	1 887
<i>VuGATL2</i>	Vigun01g195200	1 038
<i>VuGATL3</i>	Vigun02g070000	1 143
<i>VuGATL4</i>	Vigun02g136000	1 071
<i>VuGATL5</i>	Vigun03g253500	1 176
<i>VuGATL6</i>	Vigun03g269400	1 098
<i>VuGATL7</i>	Vigun03g343700	1 611
<i>VuGATL8</i>	Vigun03g429500	1 056
<i>VuGATL9</i>	Vigun03g436400	1 998
<i>VuGATL10</i>	Vigun04g057700	1 605
<i>VuGATL11</i>	Vigun04g067300	1 641
<i>VuGATL12</i>	Vigun04g178400	1 113
<i>VuGATL13</i>	Vigun05g036500	1 608
<i>VuGATL14</i>	Vigun05g077600	1 998
<i>VuGATL15</i>	Vigun05g080500	2 085
<i>VuGATL16</i>	Vigun05g135200	1 671
<i>VuGATL17</i>	Vigun05g203900	1 602
<i>VuGATL18</i>	Vigun05g218700	1 602
<i>VuGATL19</i>	Vigun06g041700	1 587
<i>VuGATL20</i>	Vigun06g147100	1 986
<i>VuGATL21</i>	Vigun07g104800	1 758
<i>VuGATL22</i>	Vigun07g113700	1 056
<i>VuGATL23</i>	Vigun07g122600	1 086
<i>VuGATL24</i>	Vigun08g018500	1 035
<i>VuGATL25</i>	Vigun08g194700	1 584
<i>VuGATL26</i>	Vigun08g214500	1 659
<i>VuGATL27</i>	Vigun09g059500	1 605
<i>VuGATL28</i>	Vigun09g243400	1 095
<i>VuGATL29</i>	Vigun09g278600	2 010
<i>VuGATL30</i>	Vigun10g117000	2 031
<i>VuGATL31</i>	Vigun11g076400	1 602
<i>VuGATL32</i>	Vigun11g109600	1 857

### 1.4 *RSCU* 聚类热图及 *CDS* 序列聚类树状图的绘制

使用 TBtools 分析豇豆 *GATL* 基因家族密码子的 *RSCU*,并绘制 *RSCU* 聚类热图。使用 MEGA11 绘制豇豆 *GATL* 基因的 *CDS* 序列聚类树状图。

### 1.5 PR2-plot 分析、ENc-plot 和中性图分析

以  $G3/(G3+C3)$  为横坐标,  $A3/(A3+T3)$  为纵坐标, 绘制 PR2-plot 图。在 PR2-plot 图上作出参考线  $x=0.5$  和  $y=0.5$ , 用于分析豇豆 *GATL* 基因密码子的偏好性。

使用 Origin2021 软件, 以 *ENc* 值为纵坐标, *GC3* 为横坐标, 绘制 ENc-plot 图。理论 *ENc* 值利用公式  $Enc = 2 + GC3 + 29/[GC3^2 + (1 - GC3)^2]$  计算, 再以理论 *Enc* 值为纵坐标, *GC3* 为横坐标, 在 ENc-plot 图上绘制标准曲线。

以 *GC12* (密码子第 1 位和第 2 位的 G+C 含量的平均值) 为纵坐标, *GC3* 为横坐标, 绘制中性图, 以  $y=x$  直线作为参考, 用于分析影响豇豆 *GATL* 基因密码子使用偏好的因素。利用 Origin2021 对中性图进行线性拟合, 得到相关系数等参数。

## 2 结果与分析

### 2.1 豇豆 *GATL* 基因家族密码子碱基组成分析

采用软件 Codon W1.4 和 MEGA 11 分析豇豆 *GATL* 基因家族密码子偏好性相关参数, 豇豆 *GATL* 基因家族成员 *T3* 为 14.73%~48.02%, 平均值为 35.80%; *C3* 为 20.52%~67.41%, 平均值为 31.92%; *A3* 为 11.28%~40.26%, 平均值为 28.09%; *G3* 为 18.72%~44.40%, 平均值为 30.65%。豇豆 *GATL* 基因家族大部分成员的同义密码子碱基偏好以 T 或 C 结尾, 但 *VuGATL3*、*VuGATL5*、*VuGATL8*、*VuGATL22*、*VuGATL23* 同义密码子偏好以 G 或 C 结尾。*VuGATL4*、*VuGATL5*、*VuGATL6*、*VuGATL28*、*VuGATL31* 同义密码子第 3 位对 A 的偏好略大于对 T 的偏好, 而其他 *GATL* 基因家族成员同义密码子第 3 位对 T 的偏好略大于对 A 的偏好。豇豆 *GATL* 基因家族成员的 *GC* 为 40.60%~58.80%, *GC* 平均值为 46.41% (表 2)。

### 2.2 豇豆 *GATL* 基因家族密码子参数分析

用 Codon W1.4 软件分析豇豆 *GATL* 基因家族密码子相关参数。如表 3 所示, 豇豆 *GATL* 基因家族的有效密码子数 (*ENc*) 为 39.29~55.77, 平均值为 51.06; 最优密码子使用频率 (*Fop*) 为 0.32~0.54, 平均值为 0.41; 密码子适应性指数 (*CAI*) 为 0.17~0.25, 平均值为 0.21; 密码子偏好性指数 (*CBI*) 为 -0.153~0.220, 平均值为 0.001; 同义密码子数 (*L<sub>sym</sub>*) 为 331~670, 平均值为 487; 编码氨基酸数 (*L<sub>aa</sub>*) 为 344~694, 平均值为 509; 编码蛋白质疏水性 (*Gravy*) 为 -0.60~

0.05, 平均值为 -0.24; 编码蛋白质芳香性 (*Aromo*) 为 0.08~0.13, 平均值为 0.11。

表 2 豇豆 *GATL* 基因家族密码子碱基组成

Table 2 Cowpea *GATL* gene family codon base composition

基因	<i>T3</i> (%)	<i>C3</i> (%)	<i>A3</i> (%)	<i>G3</i> (%)	<i>GC3</i> (%)	<i>GC</i> (%)
<i>VuGATL1</i>	46.34	20.73	34.73	26.78	35.80	41.30
<i>VuGATL2</i>	22.90	43.43	13.71	43.35	69.30	55.90
<i>VuGATL3</i>	24.70	42.77	11.28	44.40	69.30	55.60
<i>VuGATL4</i>	15.34	67.41	15.48	18.72	74.60	58.80
<i>VuGATL5</i>	16.08	46.49	20.07	41.41	70.70	56.90
<i>VuGATL6</i>	19.62	49.37	19.76	32.92	67.70	55.20
<i>VuGATL7</i>	41.59	25.00	33.60	27.41	39.80	43.00
<i>VuGATL8</i>	20.34	51.19	14.57	37.72	71.20	55.60
<i>VuGATL9</i>	45.32	21.22	34.43	28.38	37.20	41.80
<i>VuGATL10</i>	44.90	24.72	29.11	28.24	40.10	42.10
<i>VuGATL11</i>	41.61	24.11	26.91	38.40	45.90	42.60
<i>VuGATL12</i>	47.23	20.52	27.03	30.71	38.90	43.20
<i>VuGATL13</i>	43.31	24.04	31.55	28.32	39.50	42.10
<i>VuGATL14</i>	44.70	21.58	35.31	27.49	37.20	41.40
<i>VuGATL15</i>	48.02	21.28	32.09	28.45	37.50	41.60
<i>VuGATL16</i>	42.26	24.71	29.67	34.52	43.80	42.20
<i>VuGATL17</i>	41.18	20.94	37.21	28.13	37.30	41.70
<i>VuGATL18</i>	43.58	23.49	33.15	30.81	40.20	41.10
<i>VuGATL19</i>	40.00	23.29	34.60	26.56	39.10	44.20
<i>VuGATL20</i>	44.27	21.76	32.51	30.11	39.20	42.30
<i>VuGATL21</i>	41.63	25.11	37.62	24.35	37.70	41.40
<i>VuGATL22</i>	25.17	44.14	15.66	40.77	66.60	53.30
<i>VuGATL23</i>	17.25	52.72	15.00	36.21	73.10	58.10
<i>VuGATL24</i>	20.60	50.83	20.17	30.09	66.80	54.00
<i>VuGATL25</i>	38.43	28.01	32.11	27.15	43.20	44.40
<i>VuGATL26</i>	41.07	25.75	30.49	33.06	43.90	43.00
<i>VuGATL27</i>	46.19	20.79	36.34	24.35	34.10	41.20
<i>VuGATL28</i>	14.73	60.50	20.16	25.74	72.20	55.50
<i>VuGATL29</i>	45.66	23.58	36.73	22.78	35.20	40.60
<i>VuGATL30</i>	45.11	21.43	33.82	27.81	37.30	41.40
<i>VuGATL31</i>	37.53	24.01	40.26	25.14	37.80	41.80
<i>VuGATL32</i>	38.95	26.57	33.87	30.48	42.70	41.90
平均值	35.80	31.92	28.09	30.65	48.90	46.41

*T3*: 同义密码子第 3 位碱基 T 的含量; *C3*: 同义密码子第 3 位碱基 C 的含量; *A3*: 同义密码子第 3 位碱基 A 的含量; *G3*: 同义密码子第 3 位碱基 G 的含量; *GC3*: 同义密码子第 3 位 G+C 的含量; *GC*: 基因组中 G+C 含量。

表 3 豇豆 *GATL* 基因家族密码子参数

Table 3 Cowpea *GATL* gene family codon parameters

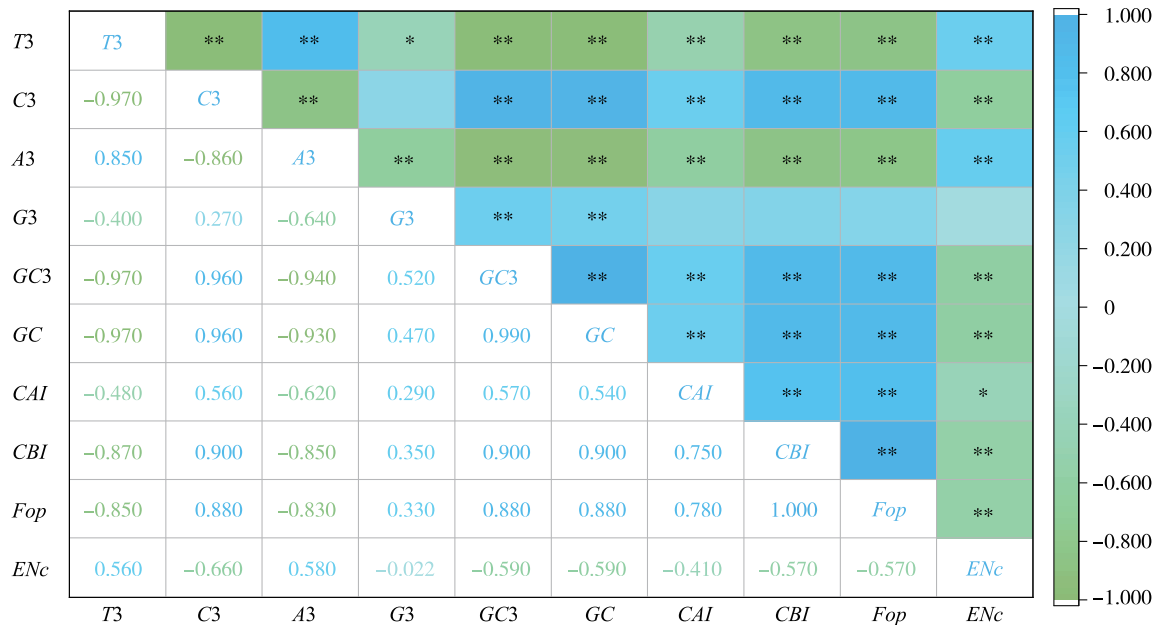
基因	密码子适应性指数	密码子偏好性指数	最优密码子使用频率	有效密码子数	同义密码子数	编码氨基酸数	编码蛋白质疏水性	编码蛋白质芳香性
<i>VuGATL1</i>	0.20	-0.041	0.39	52.82	600	628	-0.48	0.09
<i>VuGATL2</i>	0.22	0.062	0.45	52.07	332	345	0.03	0.13
<i>VuGATL3</i>	0.24	0.182	0.51	49.53	365	380	-0.11	0.12
<i>VuGATL4</i>	0.25	0.220	0.54	39.29	342	356	-0.04	0.11
<i>VuGATL5</i>	0.23	0.187	0.52	50.27	375	391	-0.10	0.13
<i>VuGATL6</i>	0.21	0.091	0.46	52.59	347	365	0.02	0.13
<i>VuGATL7</i>	0.17	-0.153	0.32	55.77	513	536	-0.15	0.11
<i>VuGATL8</i>	0.23	0.167	0.51	46.85	333	351	-0.15	0.11
<i>VuGATL9</i>	0.21	-0.063	0.38	53.67	643	665	-0.48	0.09
<i>VuGATL10</i>	0.21	-0.023	0.40	52.23	511	534	-0.23	0.11
<i>VuGATL11</i>	0.21	-0.064	0.38	52.31	514	546	-0.26	0.11
<i>VuGATL12</i>	0.20	-0.074	0.37	43.68	352	370	-0.12	0.13
<i>VuGATL13</i>	0.21	-0.022	0.40	53.17	511	535	-0.22	0.10
<i>VuGATL14</i>	0.19	-0.077	0.37	52.14	635	665	-0.46	0.10
<i>VuGATL15</i>	0.24	0.003	0.42	46.97	670	694	-0.60	0.08
<i>VuGATL16</i>	0.22	-0.104	0.36	52.02	532	556	-0.37	0.12
<i>VuGATL17</i>	0.18	-0.139	0.33	51.45	509	533	-0.09	0.11
<i>VuGATL18</i>	0.21	-0.049	0.39	54.52	505	533	-0.34	0.12
<i>VuGATL19</i>	0.20	-0.019	0.40	55.74	504	528	-0.14	0.09
<i>VuGATL20</i>	0.21	-0.031	0.40	54.66	640	661	-0.53	0.08
<i>VuGATL21</i>	0.19	-0.088	0.36	53.26	560	585	-0.41	0.10
<i>VuGATL22</i>	0.21	0.054	0.44	47.24	335	351	-0.26	0.12
<i>VuGATL23</i>	0.20	0.138	0.49	47.81	346	361	-0.06	0.12
<i>VuGATL24</i>	0.25	0.176	0.51	49.90	331	344	0.05	0.13
<i>VuGATL25</i>	0.23	0.018	0.42	54.99	507	527	-0.12	0.10
<i>VuGATL26</i>	0.22	-0.108	0.36	52.63	526	552	-0.38	0.11
<i>VuGATL27</i>	0.18	-0.137	0.33	52.60	511	534	-0.16	0.12
<i>VuGATL28</i>	0.22	0.168	0.51	42.80	352	364	-0.12	0.11
<i>VuGATL29</i>	0.21	-0.018	0.41	54.20	640	669	-0.56	0.10
<i>VuGATL30</i>	0.19	-0.073	0.37	53.99	643	676	-0.40	0.10
<i>VuGATL31</i>	0.17	-0.130	0.33	49.68	511	533	-0.14	0.10
<i>VuGATL32</i>	0.21	0.003	0.42	53.22	590	618	-0.31	0.12
平均	0.21	0.001	0.41	51.06	487	509	-0.24	0.11

2.3 豇豆 *GATL* 基因家族密码子参数相关性分析

由图 1 可知, *T3* 与 *C3* 呈显著负相关( $P<0.05$ ), 与 *GC3*、*C3*、*GC*、*CAI*、*CBI*、*Fop* 呈极显著负相关( $P<0.01$ ); *A3* 与 *C3*、*G3*、*GC3*、*GC*、*CAI*、*CBI*、*Fop* 呈极显著负相关( $P<0.01$ )。 *ENc* 与 *T3*、*A3* 呈极显著正相关( $P<0.01$ ); *C3* 与 *GC*、*GC3*、*CAI*、*CBI*、*Fop* 呈极显著正相关( $P<0.01$ ), 与 *ENc* 则呈极显著负相关( $P<$

$0.01$ ); *G3* 与 *GC*、*GC3* 呈极显著正相关( $P<0.01$ ), 与 *CAI*、*CBI*、*Fop*、*Enc* 无显著相关性; *GC3* 与 *GC*、*CAI*、*CBI*、*Fop* 呈极显著正相关( $P<0.01$ ), 与 *ENc* 呈极显著负相关( $P<0.01$ ); *GC* 与 *CAI*、*CBI*、*Fop* 呈极显著正相关( $P<0.01$ ), 与 *ENc* 呈极显著负相关( $P<0.01$ ); *CAI* 与 *CBI*、*Fop* 呈极显著正相关( $P<0.01$ ), 与 *ENc* 呈显著负相关( $P<0.05$ ); *CBI* 与 *Fop* 呈极显著正相关

( $P < 0.01$ ), 与 *ENc* 呈极显著负相关( $P < 0.01$ )。 *Fop* 与 *ENc* 呈极显著负相关( $P < 0.01$ )。



T3:同义密码子第3位碱基T的含量;C3:同义密码子第3位碱基C的含量;A3:同义密码子第3位碱基A的含量;G3:同义密码子第3位碱基G的含量;GC3:同义密码子第3位G+C的含量;GC:基因组中G+C含量;CAI:密码子适应指数;CBI:密码子偏性指数;Fop:最优密码子使用频率;ENc:有效密码子数。\*表示相关性显著( $P < 0.05$ );\*\*表示相关性极显著( $P < 0.01$ )。

图1 豇豆 *GATL* 基因家族密码子参数相关性分析

Fig.1 Correlation analysis of codon parameters of cowpea *GATL* gene family

## 2.4 豇豆 *GATL* 基因家族密码子相对使用度分析

如表4所示,排除3个终止密码子UAA、UGA、UAG,仅编码色氨酸(Trp)的密码子UGG和仅编码甲硫氨酸(Met)的密码子UAG,对剩余的密码子进行分析。 $RSCU > 1.00$ 的密码子共有29个,占总密码子的49.15%。在上述 $RSCU > 1.00$ 的密码子中,以A结尾的密码子有8个,以U结尾的密码子有10个,以C和G结尾的密码子分别有7个和4个。与此相反, $RSCU < 1.00$ 的密码子有27个,其中以C结尾和G结尾的密码子分别有7个和9个,而以U结尾和A结尾的密码子分别有5个和6个。这表明在豇豆 *GATL* 基因家族中,使用度偏高的同义密码子偏好使用U/A结尾,而使用度偏低的同义密码子偏好使用G/C结尾。

## 2.5 豇豆 *GATL* 基因家族密码子RSCU热图聚类分析

如图2所示, *VuGATL1*、*VuGATL9*、*VuGATL10*、*VuGATL11*、*VuGATL12*、*VuGATL13*、*VuGATL15*、*VuGATL17*、*VuGATL20*、*VuGATL21*、*VuGATL29*聚为一类, *VuGATL2*、*VuGATL3*、*VuGATL4*、*VuGATL5*、*VuGATL6*、*VuGATL8*、*VuGATL22*、*VuGATL23*、*VuGATL24*、*VuGATL28*聚为一类, *VuGATL7*、*VuGATL14*、*VuGATL18*、*VuGATL19*、*VuGATL25*、*VuGATL32*聚为一类, *VuGATL16*、*VuGATL26*、*VuGATL27*、*VuGATL30*、*VuGATL31*聚为一类;*GATL*基因成员的CDS进化树如图3所示, *VuGATL7*、*VuGATL14*、*VuGATL18*、*VuGATL32*聚为一类, *VuGATL2*、*VuGATL3*、*VuGATL4*、*VuGATL5*、*VuGATL6*、*VuGATL8*聚为一类。CDS序列聚类结果与RSCU热图聚类结果较为相似,因此,依据RSCU值聚类的结果有一定的参考价值。

*GATL6*、*VuGATL8*、*VuGATL22*、*VuGATL23*、*VuGATL24*、*VuGATL28*聚为一类, *VuGATL7*、*VuGATL14*、*VuGATL18*、*VuGATL19*、*VuGATL25*、*VuGATL32*聚为一类, *VuGATL16*、*VuGATL26*、*VuGATL27*、*VuGATL30*、*VuGATL31*聚为一类;*GATL*基因成员的CDS进化树如图3所示, *VuGATL7*、*VuGATL14*、*VuGATL18*、*VuGATL32*聚为一类, *VuGATL2*、*VuGATL3*、*VuGATL4*、*VuGATL5*、*VuGATL6*、*VuGATL8*聚为一类。CDS序列聚类结果与RSCU热图聚类结果较为相似,因此,依据RSCU值聚类的结果有一定的参考价值。

## 2.6 豇豆 *GATL* 基因家族PR2-plot分析、ENc-plot分析以及中性绘图分析

PR2-plot分析结果如图4所示,各基因所在位置与2条中心线的距离差异较大,4个区域的散点不均匀分布,且多数集中在右下角区域,由此可知,密码子第3位碱基对T的使用频率明显大于A,对G的使用频率大于C。说明豇豆 *GATL* 基因家族密码子使用偏好性同时受到自然选择和基因突变的作用,且自然选择的作用更大。



表 4 豇豆 *GATL* 基因家族密码子相对使用度  
Table 4 Relative codon usage of cowpea *GATL* gene family

氨基酸	密码子	数目(个)	同义密码子 相对使用度	氨基酸	密码子	数目(个)	同义密码子 相对使用度
Phe	UUU	394	0.92	Ser	AGU	182	0.83
	UUC	430	1.08		AGC	122	0.57
	UUA	164	0.51		UCU	302	1.39
	UUG	445	1.45		UCC	270	1.42
	CUU	398	1.26		UCA	287	1.29
Leu	CUC	357	1.31	Pro	UCG	99	0.51
	CUA	161	0.51		CCU	268	1.36
	CUG	291	0.97		CCC	165	0.86
	AUU	356	1.30		CCA	234	1.17
Ile	AUC	239	1.05	Thr	CCG	119	0.61
	AUA	183	0.65		ACU	219	1.02
Met	AUG	363	1.00		ACC	243	1.33
	GUU	411	1.40	Ala	ACA	221	1.03
Val	GUC	205	0.85		ACG	102	0.61
	GUA	119	0.38		GCU	345	1.24
	GUG	358	1.36	Arg	GCC	229	0.91
Tyr	UAU	303	0.95		GCA	341	1.23
	UAC	290	1.05		GCG	144	0.62
Glu	GAA	451	1.03		CGU	96	0.62
	GAG	378	0.97		CGC	115	0.85
His	CAU	299	1.00		CGA	86	0.58
	CAC	276	1.00	Gly	CGG	100	0.71
Gln	CAA	285	1.01		AGA	264	1.76
	CAG	268	0.99		AGG	220	1.49
Asn	AAU	520	1.01		GGU	221	1.16
	AAC	401	1.00		GGC	147	0.84
Lys	AAA	456	0.90		GGA	195	1.04
	AAG	546	1.10		GGG	167	0.96
Asp	GAU	549	1.21	TER	UGA	10	0.94
	GAC	312	0.79		UAA	7	0.66
Cys	UGU	105	0.91		UAG	15	1.41
	UGC	132	1.09	Trp	UGG	338	1.00

Phe: 苯丙氨酸; Leu: 亮氨酸; Ile: 异亮氨酸; Met: 甲硫氨酸; Val: 缬氨酸; Tyr: 酪氨酸; His: 组氨酸; Gln: 谷氨酰胺; Asn: 天冬酰胺; Lys: 赖氨酸; Ser: 丝氨酸; Asp: 天冬氨酸; Glu: 谷氨酸; Pro: 脯氨酸; Thr: 苏氨酸; Ala: 丙氨酸; Cys: 半胱氨酸; Trp: 色氨酸; Arg: 精氨酸; Gly: 甘氨酸; TER: 终止密码子。

对豇豆 *GATL* 基因家族进行 ENc-plot 分析。如图 5 所示,豇豆 *GATL* 基因家族中有 3 个基因与标准曲线的距离较近,表明这 3 个基因更容易受到碱基突变的影响,而其他基因也均分布在标准曲线下方,

表明这些基因的密码子受到碱基突变的影响较小,受到自然选择的影响较大。  
如图 6 所示,豇豆 *GATL* 基因密码子 GC12 为 39.3%~50.3%,GC3 为 36.8%~75.6%。GC12 的范

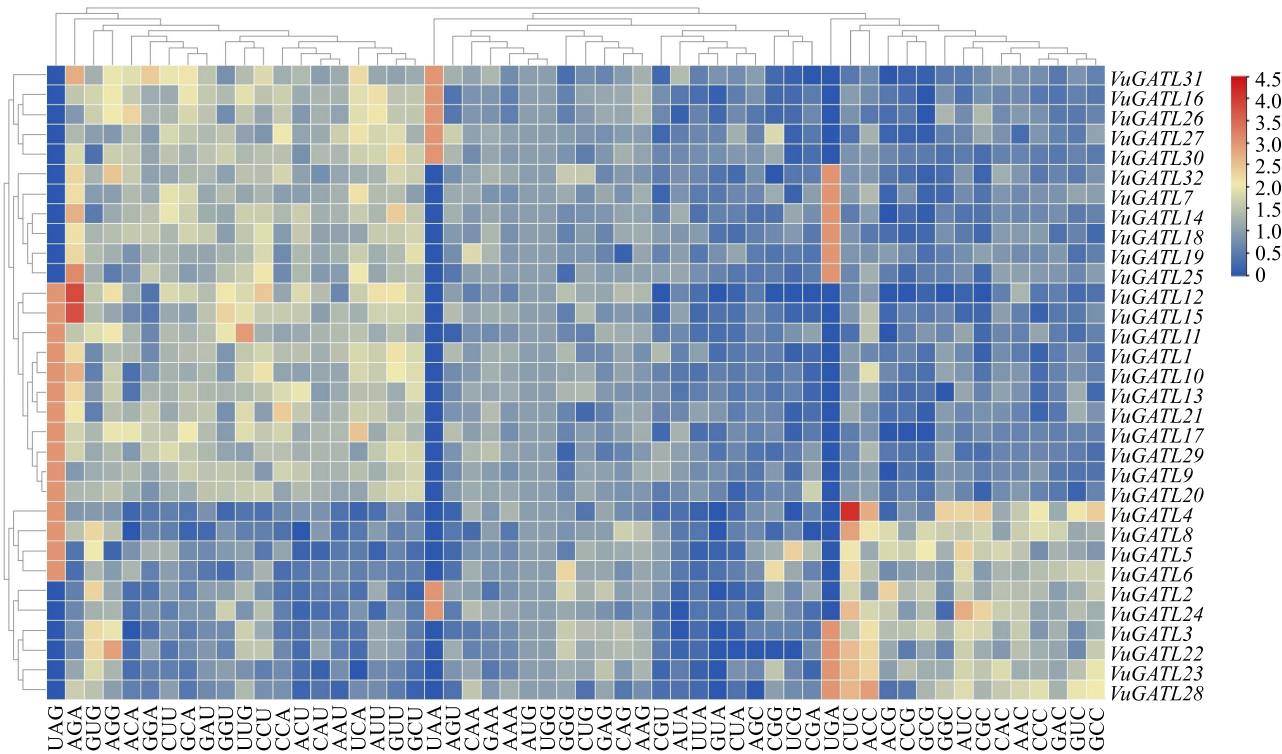


图2 豇豆 *GATL* 基因家族密码子相对同义密码子使用度 (*RSCU*) 聚类热图  
Fig.2 Clustering heatmap of relative synonymous codon usage (*RSCU*) of cowpea *GATL* gene family

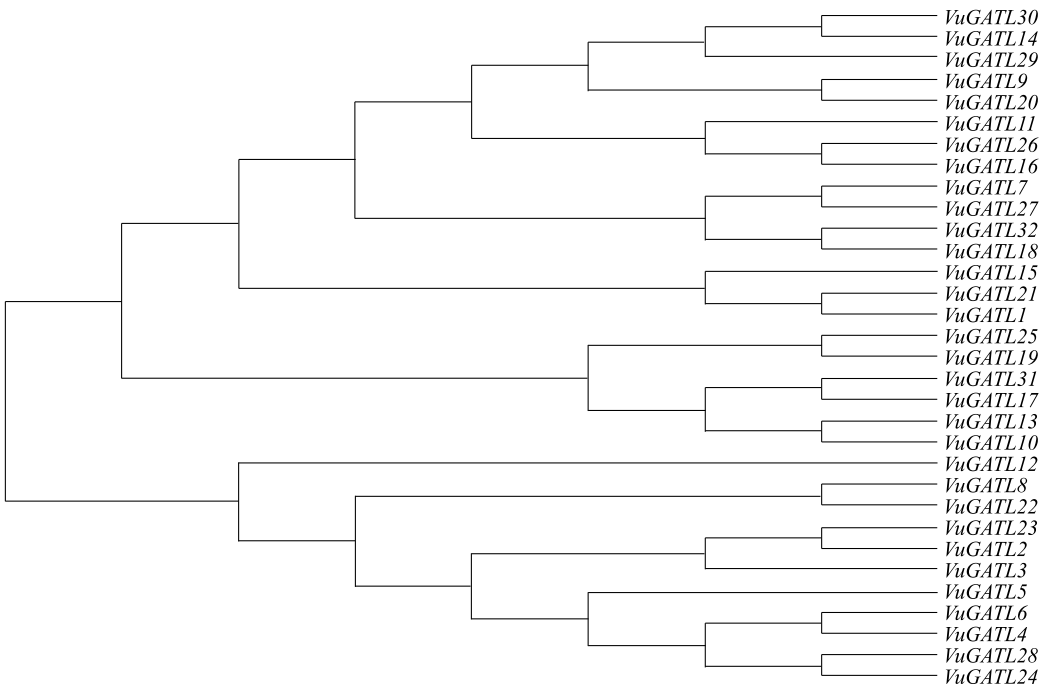
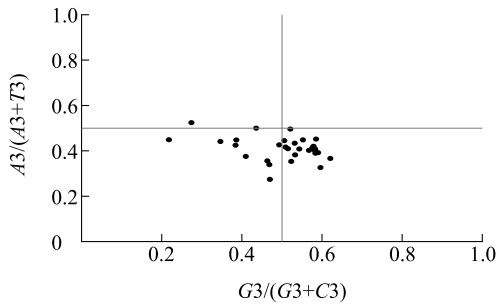


图3 豇豆 *GATL* 基因家族编码序列 (*CDS*) 聚类树状图  
Fig.3 Cluster tree of coding sequence (*CDS*) of cowpea *GATL* gene family

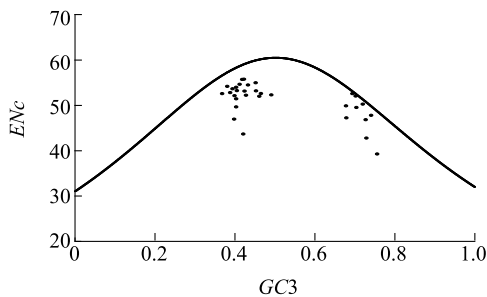
围较窄,致使基因分布在回归线左右,大部分基因处于对角线下方。*GC12* 与 *GC3* 的回归系数为 0.189 74,斜率较小,表明密码子使用偏好受自然选择的影响较大,受基因突变的影响较小。以上结果



T3:同义密码子第 3 位碱基 T 的含量;C3:同义密码子第 3 位碱基 C 的含量;A3:同义密码子第 3 位碱基 A 的含量;G3:同义密码子第 3 位碱基 G 的含量。

图 4 豇豆 *GATL* 基因家族 PR2-plot 分析

Fig.4 PR2-plot analysis of cowpea *GATL* gene family

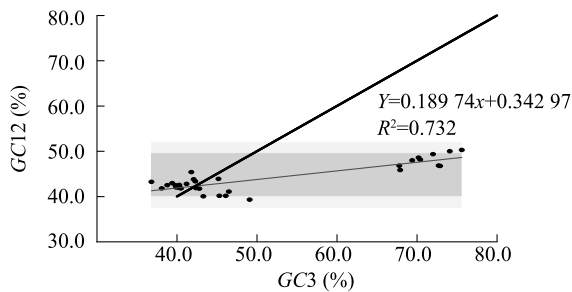


ENc:有效密码子数。GC3:同义密码子第 3 位 G+C 的含量。

图 5 豇豆 *GATL* 基因家族 ENc-plot 分析

Fig.5 ENc-plot analysis of cowpea *GATL* gene family

都表明影响豇豆 *GATL* 基因密码子使用偏好的主要因素是自然选择。



GC3:同义密码子第 3 位 G+C 的含量;GC12:密码子第 1 位和第 2 位的 G+C 含量的平均值。

图 6 豇豆 *GATL* 基因家族中性图分析

Fig.6 Neutral mapping analysis of cowpea *GATL* gene family

### 3 讨论

在遗传信息编码过程中,某一个密码子在编码氨基酸过程中的使用频率不同,这种现象称为密码子使用偏好(Codon usage bias,CUB)<sup>[12]</sup>,密码子使用偏好在自然界中普遍存在,受自然选择、基因突变和遗

传漂变共同影响<sup>[13-15]</sup>。密码子第 1、2 位碱基突变率较低,为非同义突变;而第 3 位碱基的突变为同义突变,不会改变编码的氨基酸,因此研究密码子的第 3 位碱基有利于分析密码子的使用偏好<sup>[16]</sup>。研究密码子的使用偏好有利于揭示生物进化关系,提高基因的表达效率。本研究中豇豆 *GATL* 基因家族成员的 *ENc* 为 39.29~55.77,平均值为 51.06;*CAI* 为 0.17~0.25,平均值为 0.21;*CBI* 为 -0.153~0.220,平均值为 0.001;*GC3* 为 34.10~74.60,平均为 48.90,结果表明豇豆 *GATL* 基因家族密码子的使用偏好性较弱。

本研究中,豇豆 *GATL* 基因家族成员的密码子的第 3 位碱基大多数以 T 或 A 结尾,并且大多数 *GATL* 基因家族成员密码子第 3 位碱基碱基对 T 的偏好略大于对 A 的偏好,这与前人对茶树<sup>[17]</sup>、番茄<sup>[18]</sup>等双子叶植物基因密码子偏好性的研究结果一致。双子叶植物密码子偏好以 T/A 结尾,而单子叶植物偏好以 G/C 结尾<sup>[19]</sup>。*RSCU*>1.00 表明该密码子使用度较高;*RSCU*=1.00 表明密码子随机使用;而 *RSCU*<1.00 表明该密码子使用度较低。本研究中,有 29 个豇豆 *GATL* 基因密码子 *RSCU*>1.00,27 个豇豆 *GATL* 基因密码子 *RSCU*<1.00,5 个豇豆 *GATL* 基因密码子 *RSCU*=1.00。*RSCU*>1.00 的豇豆 *GATL* 基因中有 18 个密码子以 A 或 U 结尾,表明豇豆 *GATL* 基因家族密码子偏好以 A/U(T) 结尾。有研究结果表明,CDS 聚类结果比 *RSCU* 聚类结果更接近基因的真实系统分类结果<sup>[20-22]</sup>,本研究中 CDS 聚类结果与 *RSCU* 聚类结果相似,但 CDS 聚类结果更准确,CDS 聚类结果与 *RSCU* 聚类结果在更小的分类单元中存在差异,这与前人研究结果<sup>[23-24]</sup>一致。*RSCU* 聚类存在一定的局限性,可能因为在 *RSCU* 值筛选过程中,CDS 序列中密码子中低偏好信息被遗漏,因此 *RSCU* 聚类结果只能在一定的范围内作为分类依据<sup>[23-24]</sup>。要将 *RSCU* 聚类结果与 CDS 聚类结果相结合,才能得到更为准确的分类结果,反映物种真实的系统分类,帮助了解生物体的进化历史。

对豇豆 *GATL* 基因家族密码子做相关性分析,发现 *ENc* 与 *GC3* 呈极显著负相关,表明碱基组成对 *Vu-GATL* 基因密码子偏好性影响较小,*GC3* 与 *CAI*、*CBI*、*Fop* 呈极显著正相关,*CAI*、*CBI* 和 *Fop* 两两之间均呈极显著正相关,说明密码子碱基组成对豇豆 *GATL* 基因的表达有较大的影响。进一步对豇豆 *GATL* 基因家族进行 PR2-plot 分析、ENc-plot 分析以及中性绘



图分析。PR2-plot 分析发现,A/T 与 G/C 的使用频率不同,且 A/T 的使用频率高于 G/C;ENc-plot 分析发现,豇豆 *GATL* 基因家族 *ENc* 均分布于标准曲线的下方,而且分布较集中;中性绘图分析发现,*GC12* 与 *GC3* 具有相关性,但不显著。以上结果表明,基因突变对密码子偏好性影响较小,密码子偏好性更多受到自然选择的作用。有研究结果表明辣椒 *PPO* 基因<sup>[25]</sup>、植物 *FER* 基因<sup>[26]</sup>、拟南芥 *HSP20* 基因<sup>[27]</sup> 密码子使用偏好性受自然选择影响,而兰科植物 *ABI3* 基因<sup>[28]</sup>、滇重楼 *WRKY* 基因<sup>[29]</sup>、盐角草 *HKTI* 基因<sup>[30]</sup> 等密码子偏好性主要受到基因突变影响。

## 4 结 论

本研究对豇豆 *GATL* 基因家族密码子使用偏性进行分析,发现大部分豇豆 *GATL* 基因家族密码子偏好以 A/T 结尾。豇豆 *GATL* 基因家族的 *ENc* 为 39.29~55.77,平均值为 51.06,*GC3* 为 34.10%~74.60%,表明豇豆 *GATL* 基因家族密码子使用偏好性较弱。豇豆 *GATL* 基因密码子使用偏好受自然选择影响较大,受基因突变影响较小。本研究结果为豇豆 *GATL* 基因家族的外源表达与功能预测提供了理论依据。

## 参考文献:

- [1] 滕 腾,赵懿琛,赵德刚. 鸟王茶基因密码子偏好性分析[J]. 基因组学与应用生物学,2021,40(2):795-801.
- [2] QUAX T E, CLAASSENS N J, SOLL D, et al. Codon bias as a means to fine-tune gene expression[J]. Molecular Cell,2015,59(2):149-161.
- [3] GIULIO M D. The origin of the genetic code: theories and their relationships, a review[J]. Biosystems,2005,80(2):175-184.
- [4] 吴宪明,吴松峰,任大明,等. 密码子偏性的分析方法及相关研究进展[J]. 遗传,2007,29(4):420-426.
- [5] BOTZMAN M, MARGALIT H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles[J]. Genome Biology,2011,12(10):R109.
- [6] BHATTACHARYYA D, UDDIN A, DAS S, et al. Mutation pressure and natural selection on codon usage in chloroplast genes of two species in *Pisum* L. (Fabaceae: Faboideae)[J]. Mitochondrial DNA, Part A,2019,30(4):664-673.
- [7] BROWN D M, GOUBET F, WONG V W, et al. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis[J]. Plant J,2007,52(6):1154-1168.
- [8] 陆 军,孙丽娟,王晓荣,等. 巨桉糖基转移酶基因 *EgrGATLI* 序列特征及表达分析[J]. 浙江农林大学学报,2018,35(4):604-611.
- [9] LIU J, LUO M, YAN X, et al. Characterization of genes coding for galacturonosyltransferase-like (*GATL*) proteins in rice [J]. Genes & Genomics,2016,38(10):917-929.
- [10] 仇学文,曾春涛,王 楠,等. 豇豆 *OSCA* 基因家族密码子偏好性分析[J/OL]. 分子植物育种,1-12[2024-03-01]. <http://kns.cnki.net/kcms/detail/46.1068.S.20230216.0847.004.html>.
- [11] 吕家龙. 蔬菜栽培学各论[M]. 北京:中国农业出版社,1980.
- [12] HERSHBERG R, PETROV D A. Selection on codon bias[J]. Annual Review of Genetics,2008,42:287-299.
- [13] SHABALINA S A, SPIRIDONOV N A, KASHINA A. Sounds of silence; synonymous nucleotides as a key to biological regulation and complexity[J]. Nucleic Acids Research,2013,41(4):2073-2094.
- [14] SHARP P M, EMERY L R, ZENG K. Forces that influence the evolution of codon bias[J]. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 2010, 365(1544):1203-1212.
- [15] 宋 芸,贾孟君,陈 亮,等. 忍冬 *ICE1* 基因密码子偏好性分析及受体系统选择[J]. 植物生理学报,2020,56(11):2459-2468.
- [16] 孙月琪,李密密,周义峰. 珊瑚菜叶绿体基因组密码子使用偏好性分析[J]. 植物资源与环境学报,2023,32(6):1-10.
- [17] 王占军,吴子琦,王朝霞,等. 3个茶树品种 *WOX* 基因家族的进化及密码子偏好性比较[J]. 南京林业大学学报(自然科学版),2022,46(2):71-80.
- [18] 陆奇丰,黄至欢,骆文华. 番茄 *WRKY* 转录因子密码子偏性分析[J]. 分子植物育种,2020,18(18):5908-5916.
- [19] 刘庆坡,冯 英,董 辉. 20个物种同义密码子偏性的比较分析[J]. 西北农林科技大学学报(自然科学版),2004,32(7):67-71.
- [20] 李 翔,范作义,王井源,等. 红松查尔酮合成酶基因 *CHS* 密码子偏好性分析[J]. 植物研究,2020,40(3):447-457.
- [21] 赖瑞联,林玉玲,钟春水,等. 龙眼生长素受体基因 *TIR1* 密码子偏好性分析[J]. 园艺学报,2016,43(4):771-780.
- [22] 赵春丽,彭丽云,王 晓,等. 苋菜 *AtGAI* 基因密码子偏好性与进化分析[J]. 中国农业大学学报,2019,24(12):10-22.
- [23] 高守舆,李钰莹,杨志青,等. 白羊草叶绿体基因组密码子使用偏好性分析[J]. 草业学报,2023,32(7):85-95.
- [24] 晁岳愚,吴政卿,杨会民,等. 11种植物 *psbA* 基因的密码子偏好性及聚类分析[J]. 核农学报,2011,25(5):927-932.
- [25] 谢雪果,张中荣,吕 慧,等. 辣椒 *PPO* 基因家族密码子偏好性分析[J]. 北方园艺,2022(7):1-8.
- [26] 赵 森,邓力华,陈 芬. 不同植物 *FERONIA* 基因密码子使用偏好性分析[J]. 江苏农业学报,2020,36(5):1073-1081.
- [27] JI H, LIU J, CHEN Y, et al. Bioinformatic analysis of codon usage bias of *HSP20* genes in four cruciferous species[J]. Plants (Basel),2024,13(4):468.
- [28] 龚婷婷,李维仪,陶 媛,等. 兰科 *ABI3* 基因的密码子偏性分析[J]. 热带作物学报,2023,44(2):264-272.
- [29] 常箫月,王晨舒,李维蛟,等. 滇重楼 *WRKY* 基因家族的鉴定及密码子偏好性分析[J]. 分子植物育种,2023.
- [30] 王转霞,赵诗佳,李 婧,等. 盐生植物和甜土植物 *HKTI* 基因密码子偏好性分析[J]. 中国草食动物科学,2023,43(4):12-18.

(责任编辑:成纾寒)