

刘 秦, 蔡雨静, 李锦兰, 等. 木薯扩展蛋白基因家族密码子使用偏好性分析及影响因素[J]. 江苏农业学报, 2025, 41(3): 468-476.  
doi: 10.3969/j.issn.1000-4440.2025.03.005

# 木薯扩展蛋白基因家族密码子使用偏好性分析及影响因素

刘 秦<sup>1</sup>, 蔡雨静<sup>1</sup>, 李锦兰<sup>1</sup>, 耿晓珊<sup>2</sup>, 卢恩科<sup>1</sup>

(1. 玉林师范学院广西农产资源化学与生物技术重点实验室, 广西 玉林 537000; 2. 广西农业职业技术大学农业工程学院, 广西 南宁 530000)

**摘要:** 木薯(*Manihot esculenta*)是全球一种重要的粮食作物。扩展蛋白在植物生长发育中起关键作用, 具有改良作物性状的潜力。本研究利用 CodonW、R 语言等软件对木薯基因组中的 42 个扩展蛋白基因进行密码子使用偏好(CUB)的系统分析。结果显示, 木薯扩展蛋白基因的平均有效密码子数(ENC)为 53.74, 表明其密码子使用存在一定偏好性, 但偏好程度较轻。ENC 值与密码子第一、第三位置的 G+C 含量极显著相关, 与氨基酸长度显著相关。相对同义密码子使用频率(RSCU)分析结果表明, 木薯扩展蛋白基因倾向于选择以 A/U 结尾的密码子。中性绘图和 ENC-plot 分析结果显示, 自然选择是影响木薯扩展蛋白基因 CUB 形成的主导力量, 突变压力作用较小。PR2 绘图进一步验证了自然选择在密码子使用偏好中的重要作用。通过比较高偏好基因和低偏好基因的 RSCU 差异, 筛选出 23 个最优密码子。本研究结果揭示了木薯扩展蛋白基因 CUB 形成的分子机制, 为利用密码子优化策略提高木薯产量和抗逆性提供了理论基础。

**关键词:** 木薯; 扩展蛋白基因; 密码子偏好性; 最优密码子

**中图分类号:** S889+.5 **文献标识码:** A **文章编号:** 1000-4440(2025)03-0468-09

## Codon usage bias analysis and influencing factors of the cassava expansin protein gene family

LIU Qin<sup>1</sup>, CAI Yujing<sup>1</sup>, LI Jinlan<sup>1</sup>, GENG Xiaoshan<sup>2</sup>, LU Enke<sup>1</sup>

(1. Guangxi Key Laboratory of Agricultural Resources Chemistry and Biotechnology, Yulin Normal University, Yulin 537000, China; 2. College of Agricultural Engineering, Guangxi Vocational University of Agriculture, Nanning 530000, China)

**Abstract:** Cassava (*Manihot esculenta*) is a globally important food crop. Expansins (EXPs) play a crucial role in plant growth and development, holding the potential to improve crop traits. This study aimed to systematically analyze the

收稿日期: 2024-06-05

基金项目: 国家自然科学基金项目(32101691); 广西自然科学基金项目(2021GXNSFBA196077); 国家级大学生创新创业训练计划项目(202310606031)

作者简介: 刘 秦(1986-), 男, 湖北广水人, 博士, 副研究员, 主要从事木薯块根发育研究。(Tel) 18176600018; (E-mail) liuqin@ylnu.edu.cn

通讯作者: 耿晓珊, (E-mail) gengxs@ylnu.edu.cn; 卢恩科, (E-mail) luek2823348@ylnu.edu.cn

codon usage bias (CUB) of 42 expansin genes in the cassava genome using softwares such as CodonW and R language. The results revealed that the average effective number of codons (ENC) for *M. esculenta* EXP genes (*MeEXPs*) was 53.74, indicating a moderate level of codon usage bias. The ENC values showed highly significant correlations with the G+C contents at the first and third codon positions, and showed significant correlation with amino acid length. Relative synonymous codon usage (RSCU)

analysis demonstrated that *MeEXPs* tended to utilize codons ending with A/U. The results of neutral plot and *ENC*-plot analyses suggested that natural selection was the predominant force influencing the CUB shaping of *MeEXPs*, while mutational pressure played a relatively minor role. The PR2 plot further corroborated the pivotal role of natural selection in codon usage bias. By comparing the *RSCU* differences between high and low bias genes, a total of 23 optimal codons were identified. This study unraveled the molecular mechanisms underlying the formation of CUB in *MeEXPs*, and provided a theoretical foundation for using codon optimization strategies to enhance cassava yield and stress tolerance.

**Key words:** cassava; expansin genes; codon usage bias; optimal codon

木薯作为全球性的重要粮食作物,不仅在动物饲料和工业原料领域扮演着关键角色,而且在非洲和亚洲部分地区的日常饮食中占据重要地位<sup>[1]</sup>。面对全球人口增长和资源有限的挑战,开发和利用木薯等作物资源对于增强全球粮食安全具有重大意义<sup>[2]</sup>。为了提高作物的产量和适应性,科研人员正不断探索影响植物生长和发育的关键因素。

植物扩展蛋白(Expansins),是一类位于植物细胞壁胞外区域的蛋白质,对植物生长和发育具有至关重要的作用<sup>[3-4]</sup>。它们的主要功能是促进细胞壁的伸展与松弛,从而增强细胞壁的柔韧性并支持细胞的伸展过程。这一机制对植物的多个关键生物过程至关重要,包括根毛的形成、叶片的发育、花粉管的生长以及果实的成熟<sup>[4]</sup>。此外,扩展蛋白在植物应对环境变化,如干旱、高温和病原体攻击时的响应机制中也扮演着重要角色,它们通过影响根系的发育和植物形态来增强植物的适应性<sup>[5]</sup>。鉴于扩展蛋白在植物生理过程中的多功能性,它们被认为是改良作物性状的重要潜在靶点,有望通过基因工程手段提高作物的产量和抗逆性<sup>[6]</sup>。

在基因编码系统中,三联体密码子是构成信使RNA序列的基本单元,它们将DNA中的遗传信息转化为蛋白质<sup>[7]</sup>。尽管64个密码子中有61个密码子编码20种氨基酸(其余3个为终止密码子),但编码同一氨基酸的不同密码子(同义密码子)使用频率不一,形成密码子使用偏倚(CUB)。CUB是生物适应环境和进化压力的结果,受基因组碱基组成、基因表达水平、基因长度、转运RNA丰度及其相互作用、自然选择和遗传漂变等因素影响<sup>[8]</sup>。深入研究CUB有助于理解植物基因功能和进化机制,对物种适应性和进化研究具有重要意义。

鉴于CUB对基因表达和生物适应性的重要影响,尽管在植物基因组CUB分析方面已取得一定进展,但其在基因家族层面的作用 and 意义仍需进一步

探讨。特别是针对特定基因家族(如木薯中的扩展蛋白基因)的CUB研究,可以为作物改良提供更为精确的策略。本研究专注于木薯扩展蛋白基因的密码子偏好性分析,旨在探明其偏好性的形成机制,并筛选最优密码子。此研究将深化对木薯扩展蛋白基因功能和进化的理解,通过密码子优化策略,改造木薯相关基因,并为提高其产量和抗逆性提供理论基础。

## 1 材料与方法

### 1.1 数据来源

本研究以拟南芥扩展蛋白序列为探针,在Phytozome数据库(<https://phytozome-next.jgi.doe.gov/>)中进行BLASTP比对,设定 $e$ -value阈值小于 $10^{-5}$ 。初步筛选得到候选蛋白质序列后,下载Pfam数据库(<https://pfam.xfam.org/>)中的扩展蛋白家族蛋白PF01357(Pollen\_allerg\_1)和PF03330(DPBB\_1)的保守结构域序列。随后,应用HMMER 3.3.2软件包含的HMMSEARCH工具( $e$ -value阈值 $<10^{-5}$ ),基于隐马尔可夫模型对前期筛选的木薯扩展蛋白候选氨基酸序列进行深入筛查。

### 1.2 密码子使用偏倚参数计算

使用CodonW(1.4.2版本)软件,计算有效密码子数(*ENC*)、相对同义密码子使用频率(*RSCU*)、基因整体的G+C含量( $GC_{all}$ )、同义密码子第三位的G+C含量( $GC_{3s}$ )以及同义密码子第三位上的T、C、A和G含量( $T_{3s}$ 、 $C_{3s}$ 、 $A_{3s}$ 、 $G_{3s}$ )。通过EMBOSS套件中的CUSP统计程序(<https://www.bioinformatics.nl/cgi-bin/emboss/cusp>)计算第一、第二、第三密码子位的G+C含量( $GC_1$ 、 $GC_2$ 、 $GC_3$ ),以及第一和第二密码子位的平均G+C含量( $GC_{12}$ )。采用R语言中的cor()函数,执行皮尔逊相关系数计算,以探究密码子使用各指标间的相关性。

### 1.3 中性绘图分析

中性绘图分析旨在区分影响密码子使用偏好的

主要因素,判定是自然选择还是突变压力所致<sup>[9]</sup>。在中性绘图分析中,纵轴代表  $GC_{12}$  值,横轴代表  $GC_3$  值。若  $GC_3$  的相关系数在统计上显著且接近 1,则认为突变压力是影响密码子使用的主要力量。斜率的降低表明突变压力对密码子使用影响的减弱,斜率为 0 则意味着密码子使用偏倚完全由自然选择引起<sup>[8]</sup>。利用 R 语言中的  $\text{lm}()$  函数拟合线性模型,估计  $GC_3$  与  $GC_{12}$  之间的线性关系。

#### 1.4 ENC-plot 分析

ENC-plot 分析通过衡量同义密码子使用差异性的 ENC 值,揭示特定密码子的选择倾向。ENC 值范围为 20 至 61,与密码子选择倾向呈负相关。ENC 值较低的基因表现出对特定密码子的明显偏好,ENC 值超过 35 则偏好性较弱,接近 61 时密码子选择几乎不受偏好性影响<sup>[10]</sup>。 $GC_{3s}$  作为核苷酸组成偏差的关键指标,是通过 ENC 图表分析密码子选择偏好的主要驱动因素。ENC 的理论曲线根据公式  $ENC_{期望} = 2 + GC_{3s} + 29/[GC_{3s}^2 + (1 - GC_{3s})^2]$  绘制<sup>[10]</sup>。为精确衡量期望 ENC 值与实际 ENC 值的差异,计算  $ENC_{比率} = (ENC_{期望} - ENC_{实际})/ENC_{期望}$ ,并通过直方图展示  $ENC_{比率}$  分布。

#### 1.5 PR2 图分析

PR2 图用于分析评估突变和自然选择对 CUB 的影响,专注于分析同义密码子第 3 位的 4 种碱基组成。PR2 图通过在 y 轴上绘制  $A_{3s}/(A_{3s} + T_{3s})$  值,在 x 轴上绘制  $G_{3s}/(G_{3s} + C_{3s})$  值,展示同义密码子第 3 位碱基优选值在平衡态和偏移态之间的分布情况<sup>[11]</sup>。图中点的分布可用于分析碱基偏移的进化力和突变压力对密码子偏好性的影响。点均匀分布在中心点(碱基中 A 与 T 比例相等,G 与 C 比例相等)周围表明突变压力为 CUB 主要因素,这意味着密码子第 3 位的碱基中 A 与 T 比例相等,G 与 C 比例相等,密码子组成没有明显偏向;数据点偏离原点则表明自然选择等其他因素影响密码子使用偏好。数据点偏离方向和程度揭示碱基组成偏离模式和强度。

#### 1.6 最优密码子的筛选

RSCU 作为评估密码子使用频率的指标,反映特定密码子组合的相对使用频率。RSCU 值等于 1 时密码子使用频率均匀,大于 1 表明密码子使用频率超出平均值,小于 1 则表明使用频率低于平均值。筛选最优密码子过程中,首先计算所有基因的

ENC,依据 ENC 值排序。选取 ENC 值最低和最高的 10% 基因,形成低偏好和高偏好基因群,分析 2 组基因中 RSCU 值的差异( $\Delta RSCU$ ),辨识高表达基因中更受青睐的密码子。具体而言, $\Delta RSCU$  等于或大于 0.08,且高偏好基因中  $RSCU > 1$ ,低偏好基因中  $RSCU < 1$  时,该密码子被认定为最优密码子<sup>[12]</sup>。

## 2 结果与分析

### 2.1 密码子碱基组成分析

木薯扩展蛋白基因的  $GC_1$ 、 $GC_2$ 、 $GC_3$  以及整体 G+C 含量( $GC_{all}$ )的百分比分布数据显示出一定的变化趋势,反映了基因序列中密码子使用的特征。表 1 显示, $GC_1$  的分布介于 41.13% 至 54.21%,平均值约为 47.97%。 $GC_2$  的分布范围介于 44.28% 至 54.23%,平均值约为 49.33%。 $GC_3$  的分布范围则相对较宽,从 34.63% 延伸至 55.00%,平均值约为 43.63%,显示出在密码子的第 3 位 G+C 碱基的使用变化较大,可能存在较强的选择压力。 $GC_{all}$  的分布范围为 43.06% 至 51.67%,平均值约为 46.98%。

这些数据表明,在木薯扩展蛋白基因中,密码子的使用在一定程度上呈现出 G+C 含量的偏好性,在碱基选择上偏好 A/T,尤其是在密码子的第 3 位。有效密码子数(ENC)的分布范围为 44.44 至 60.25,平均值为 53.74,超过了阈值(35)。这表明虽然木薯扩展蛋白基因的密码子使用存在一定的偏好性,但这种偏好并不显著,暗示密码子使用偏倚程度较轻。

### 2.2 密码子使用偏倚指数的相关性分析

Pearson 相关性分析揭示了木薯扩展蛋白基因中有效密码子数(ENC)与密码子碱基组成之间存在极显著的正相关关系(表 2)。具体而言,ENC 与  $GC_1$  ( $r = 0.44, P < 0.01$ )、 $GC_3$  ( $r = 0.43, P < 0.01$ ) 以及整体  $GC_{all}$  ( $r = 0.42, P < 0.01$ ) 的相关系数均超过 0.40,且 P 值均小于 0.01,这些相关性在统计学上具有高度显著性。该结果表明,密码子使用偏倚与密码子碱基组成之间,尤其是密码子的第一和第三位的碱基组成之间,存在密切的联系。进一步的分析结果表明,ENC 与氨基酸长度( $L_{aa}, r = 0.37, P < 0.05$ ) 呈显著正相关。尽管其相关系数略低于 0.40,但 P 值小于 0.05,表明基因长度与密码子使用偏倚之间存在显著的相关性。

表 1 木薯扩展蛋白基因密码子在不同基因中的 G+C 含量及有效密码子数

Table 1 Codon G+C content and effective number of codons for *Manihot esculenta* expansins

基因	GC <sub>1</sub> (%)	GC <sub>2</sub> (%)	GC <sub>3</sub> (%)	GC <sub>all</sub> (%)	ENC(个)
MeEXP1	49.00	51.79	47.01	49.27	55.74
MeEXP2	48.78	44.72	45.53	46.34	58.00
MeEXP3	47.84	50.59	43.92	47.45	52.29
MeEXP4	48.08	54.23	49.62	50.64	56.81
MeEXP5	46.85	47.64	48.82	47.77	54.27
MeEXP6	53.18	50.94	43.07	49.06	60.25
MeEXP7	41.13	51.70	46.04	46.29	47.52
MeEXP8	48.79	44.35	43.55	45.56	52.49
MeEXP9	48.63	51.37	41.96	47.32	52.71
MeEXP10	47.69	52.31	55.00	51.67	58.02
MeEXP11	46.43	44.84	39.68	43.65	51.68
MeEXP12	46.59	48.59	42.57	45.92	48.69
MeEXP13	46.22	49.80	38.25	44.75	52.77
MeEXP14	49.62	46.15	38.46	44.74	54.87
MeEXP15	46.67	49.41	49.02	48.37	53.28
MeEXP16	54.21	49.82	39.93	47.99	58.67
MeEXP17	49.60	52.38	50.00	50.66	52.34
MeEXP18	44.88	51.18	40.55	45.54	44.44
MeEXP19	44.44	47.83	49.76	47.34	57.14
MeEXP20	50.00	49.20	42.80	47.33	56.89
MeEXP21	48.06	49.12	47.70	48.29	58.96
MeEXP22	47.10	52.12	50.97	50.06	50.67
MeEXP23	47.43	50.99	42.29	46.90	49.69
MeEXP24	49.46	47.31	36.92	44.56	58.97
MeEXP25	49.09	50.91	51.27	50.42	60.05
MeEXP26	46.37	50.40	36.29	44.35	47.66
MeEXP27	46.30	50.48	45.34	47.37	59.28
MeEXP28	47.50	45.36	41.07	44.64	54.98
MeEXP29	47.50	45.00	41.07	44.52	55.29
MeEXP30	46.86	44.28	42.44	44.53	54.67
MeEXP31	48.25	52.53	50.97	50.58	57.74
MeEXP32	48.39	52.02	35.08	45.16	48.99
MeEXP33	50.20	50.20	37.55	45.98	48.48
MeEXP34	47.66	52.34	42.19	47.40	53.48
MeEXP35	44.71	48.46	45.39	46.19	55.37
MeEXP36	48.16	44.90	39.18	44.08	54.49
MeEXP37	53.96	49.43	49.43	50.94	55.98
MeEXP38	50.60	50.60	43.37	48.19	51.48
MeEXP39	45.85	46.64	43.08	45.19	51.76
MeEXP40	48.25	46.30	34.63	43.06	50.02
MeEXP41	44.75	51.75	38.91	45.14	48.00
MeEXP42	49.60	51.98	41.67	47.75	52.17

GC<sub>1</sub>:密码子第 1 位的 G+C 碱基含量;GC<sub>2</sub>:密码子第 2 位的 G+C 碱基含量;GC<sub>3</sub>:密码子第 3 位的 G+C 碱基含量;GC<sub>all</sub>:单个基因总 G+C 碱基含量;ENC:有效密码子数。

表 2 木薯扩展蛋白基因密码子相关指标的相关性

Table 2 Correlation coefficient between codon-related indicators in the genes of *Manihot esculenta* expansins

指标	相关系数				
	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>	ENC
GC <sub>2</sub>	0.04				
GC <sub>3</sub>	-0.04	0.32 *			
GC <sub>all</sub>	0.35 *	0.65 **	0.84 **		
ENC	0.44 **	-0.12	0.43 **	0.42 **	
L <sub>aa</sub>	0.07	-0.02	0.03	0.04	0.37 *

GC<sub>1</sub>:密码子第 1 位的 G+C 碱基含量;GC<sub>2</sub>:密码子第 2 位的 G+C 碱基含量;GC<sub>3</sub>:密码子第 3 位的 G+C 碱基含量;GC<sub>all</sub>:单个基因总 G+C 碱基含量;ENC:有效密码子数;L<sub>aa</sub>:氨基酸长度。\* 表示 0.05 水平上相关性显著;\*\* 表示 0.01 水平上相关性显著。

2.3 相对同义密码子使用频率(RSCU)分析

通过 CodonW 软件对木薯扩展蛋白基因的相对同义密码子使用频率进行分析,结果(图 1)表明,有 29 个密码子的 RSCU 值超过了 1.00,这些密码子被归类为 MeEXPs 基因的高频密码子。进一步分析这些高频密码子,发现其中 8 个以腺嘌呤(A)结尾,4 个以胞嘧啶(C)结尾,5 个以鸟嘌呤(G)结尾,12 个以尿嘧啶(U)结尾。这种分布模式揭示了木薯扩展蛋白基因中倾向于选择以 A 或 U 结尾的密码子。这种碱基选择的非随机性反映了自然选择对 Me-EXPs 基因密码子使用的影响。

2.4 中性绘图分析

本研究采用中性绘图(图 2)探究 GC<sub>3</sub> 与 GC<sub>12</sub> 之间的关联。GC<sub>3</sub> 取值范围为 35%至 55%,GC<sub>12</sub> 取值范围为 46%至 52%,两者的平均值分别为 44%和 49%。回归分析确定的直线斜率(回归系数)约为 0.08,表明突变压力对密码子选择的贡献约为 8%,其余 92%则由其他因素解释,例如自然选择。此外,统计分析结果表明 GC<sub>12</sub> 与 GC<sub>3</sub> 之间无显著相关性(R<sup>2</sup> = 0.043 77, P = 0.183 60),提示木薯扩展蛋白基因家族中基因前两个密码子碱基与第三个密码子碱基在组成上存在显著差异。综合分析结果表明,自然选择和突变压力共同决定了木薯扩展蛋白基因密码子的选择偏好,其中自然选择起主导作用。

2.5 ENC-plot 分析结果

图 3A 显示,木薯扩展蛋白基因的 ENC 值分布呈离散状,只有少量的基因完全落在或靠近 ENC 标准曲线,而大部分的基因离标准曲线较远。从 ENC<sub>比率</sub> 统计分布图(图 3B)也可以看出,在靠近 ENC 标准曲线的区间(-0.05~0.05)内,共分布 14



个基因,剩余 28 个基因分布在离 *ENC* 标准曲线较远的区间。以上结果表明编码木薯扩展蛋白的大部

分基因的密码子使用偏倚受到了自然选择的作用。

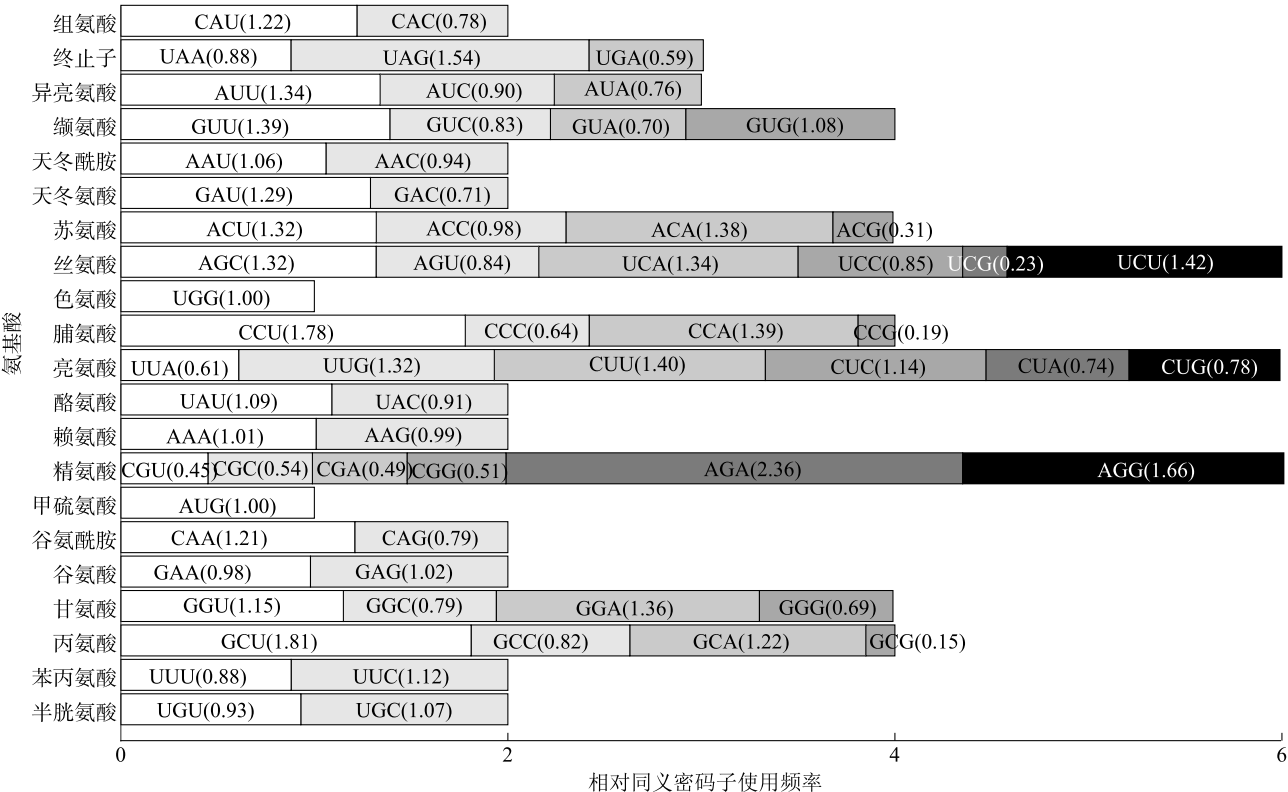


图 1 木薯扩展蛋白基因同义密码子相对使用频率分析  
Fig.1 Relative synonymous codon usage analysis of *Manihot esculenta* expansins

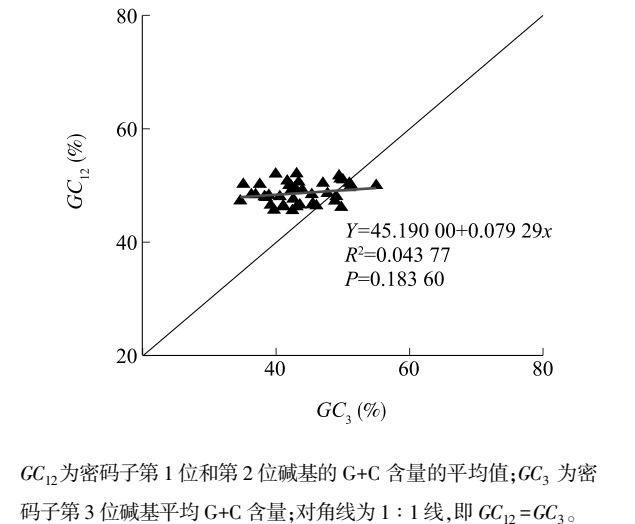


图 2 木薯扩展蛋白基因密码子中性绘图分析  
Fig.2 Neutrality plot analysis of *Manihot esculenta* expansin codons

2.6 PR2-plot 分析结果

图 4 显示,大部分基因沿纵轴分布在左下区域,

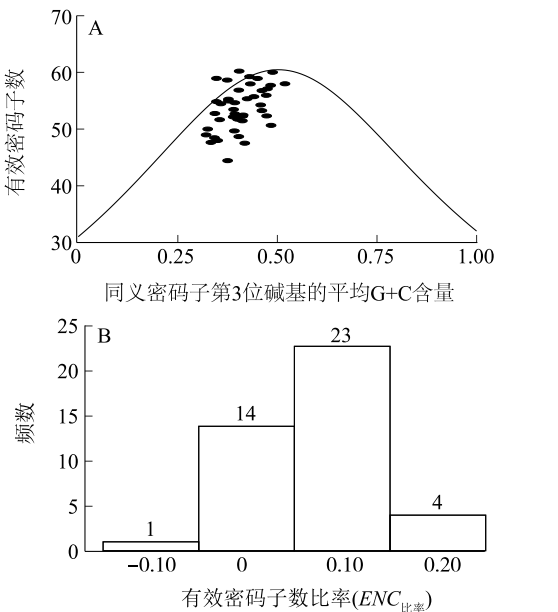


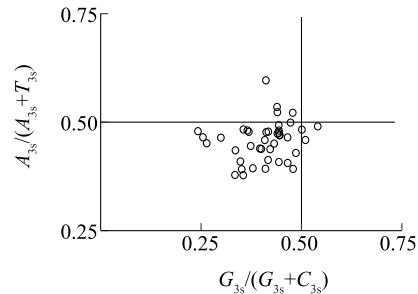
图 3 木薯扩展蛋白基因密码子的 *ENC*-plot 分析 (A) 和 *ENC* 比率分布 (B)  
Fig.3 *ENC*-plot analysis and *ENC* ratio distribution of codons in *Manihot esculenta* expansins

即这些基因的数据点在纵轴  $G_{3s}/(G_{3s}+C_{3s})$  和横轴  $A_{3s}/(A_{3s}+T_{3s})$  的数值都低于 0.50, 反映了这些基因的同义密码子第 3 位碱基中, 碱基 A 的使用频率低于碱基 T, 碱基 G 的使用频率低于碱基 C。  $A_{3s}/(A_{3s}+T_{3s})$  和  $G_{3s}/(G_{3s}+C_{3s})$  平均值分别为 0.46 和 0.41。这些结果揭示了 A+T 和 G+C 在同义密码子的第 3 位碱基存在使用不平衡的现象, 表明木薯扩展蛋白基因同义密码子的使用模式不仅受突变的影响, 还受自然选择等因素的影响。

## 2.7 最优密码子的确定

经过筛选, 确定了 *MeEXPs* 基因的 23 个最优密码子(表 3)。在这些密码子中, 有 6 个以 A 结尾, 包括 UCA、ACA、AAA、UGA、AGA 和 GGA; 3 个以 C 结尾, 包括 UUC、AUC 和 AGC; 4 个以 G 结尾, 包括 UUG、GUG、GAG 和 AGG。值得注意的是, 以 U 结尾

的密码子数量最多, 共有 10 个, 分别为 CUU、GUU、UCU、CCU、ACU、GCU、UAU、CAU、GAU 和 UGU。该结果揭示了 *MeEXPs* 基因的最优密码子倾向于以 A 或 U 结尾。



$T_{3s}$ 、 $C_{3s}$ 、 $A_{3s}$ 、 $G_{3s}$  分别为同义第三位密码子上的 T、C、A 和 G 含量。

图 4 木薯扩展蛋白基因密码子的 PR2-plot 绘图分析

Fig.4 PR2-plot analysis of codons in *Manihot esculenta* expansins

表 3 木薯扩展蛋白基因最优密码子筛选

Table 3 Selection of optimal codons for *Manihot esculenta* expansins

氨基酸	密码子	高表达基因的 RSCU	低表达基因的 RSCU	$\Delta RSCU$	氨基酸	密码子	高表达基因的 RSCU	低表达基因的 RSCU	$\Delta RSCU$
Phe	UUU	0.68	0.88	-0.20	Tyr	UAU	1.25	1.03	0.22
	UUC	1.32	1.12	0.20		UAC	0.75	0.97	-0.22
Leu	UUA	0.67	0.70	-0.03	TER	UAA	0.75	0.75	0
	UUG	1.71	0.98	0.73		UAG	0.75	1.50	-0.75
	CUU	1.52	1.19	0.33	His	CAU	1.53	1.16	0.37
	CUC	0.86	1.19	-0.33		CAC	0.47	0.84	-0.37
	CUA	0.86	0.84	0.02	Gln	CAA	1.30	1.23	0.07
Ile	CUG	0.38	1.12	-0.74		CAG	0.70	0.77	-0.07
	AUU	1.21	1.29	-0.08	Asn	AAU	1.10	1.21	-0.11
	AUC	1.04	0.86	0.18		AAC	0.90	0.79	0.11
	AUA	0.75	0.86	-0.11	Lys	AAA	1.21	0.98	0.23
Met	AUG	1.00	1.00	0		AAG	0.79	1.02	-0.23
Val	GUU	1.53	1.21	0.32	Asp	GAU	1.42	1.30	0.12
	GUC	0.95	1.15	-0.20		GAC	0.58	0.70	-0.12
	GUA	0.44	0.88	-0.44	Glu	GAA	0.55	1.23	-0.68
	GUG	1.09	0.77	0.32		GAG	1.45	0.77	0.68
Ser	UCU	1.79	1.56	0.23	Cys	UGU	1.06	0.81	0.25
	UCC	0.71	0.87	-0.16		UGC	0.94	1.19	-0.25
	UCA	1.43	1.27	0.16	TER	UGA	1.50	0.75	0.75
	UCG	0.07	0.35	-0.28	Trp	UGG	1.00	1.00	0
Pro	CCU	1.64	1.25	0.39	Arg	CGU	0.19	0.91	-0.72
	CCC	0.64	0.75	-0.11		CGC	0	1.04	-1.04

续表3 Continued3

氨基酸	密码子	高表达基因的 RSCU	低表达基因的 RSCU	$\Delta RSCU$	氨基酸	密码子	高表达基因的 RSCU	低表达基因的 RSCU	$\Delta RSCU$
Thr	CCA	1.64	1.69	-0.05	Ser	CGA	0	0.65	-0.65
	CCG	0.09	0.31	-0.22		CGG	0.19	0.52	-0.33
	<u>ACU</u>	1.50	1.12	0.38		AGU	0.57	0.81	-0.24
	ACC	0.85	1.06	-0.21		<u>AGC</u>	1.43	1.15	0.28
	<u>ACA</u>	1.45	1.29	0.16	Arg	<u>AGA</u>	4.12	1.70	2.42
Ala	ACG	0.20	0.53	-0.33		<u>AGG</u>	1.50	1.17	0.33
	<u>GCU</u>	2.36	1.54	0.82	Gly	GGU	0.77	1.09	-0.32
	GCC	0.46	0.84	-0.38		GGC	0.73	0.74	-0.01
	GCA	1.09	1.32	-0.23		<u>GGA</u>	1.70	1.44	0.26
	GCG	0.08	0.31	-0.23		GGG	0.80	0.74	0.06

Phe: 苯丙氨酸; Leu: 亮氨酸; Ile: 异亮氨酸; Met: 甲硫氨酸; Val: 缬氨酸; Ser: 丝氨酸; Pro: 脯氨酸; Thr: 苏氨酸; Ala: 丙氨酸; Tyr: 酪氨酸; His: 组氨酸; Gln: 谷氨酰胺; Asn: 天冬酰胺; Lys: 赖氨酸; Asp: 天冬氨酸; Glu: 谷氨酸; Cys: 半胱氨酸; Trp: 色氨酸; Arg: 精氨酸; Gly: 甘氨酸; TER: 终止密码子。加下划线的密码子表示最优密码子; RSCU: 相对同义密码子使用频率;  $\Delta RSCU$ : 基因间相对同义密码子使用频率值的差异。

3 讨论

CUB 广泛存在于物种遗传信息的翻译表达过程中,其产生是突变和自然选择共同作用的结果<sup>[13-18]</sup>。根据突变机制,CUB 的形成是由点突变和修复速率共同作用的结果,这解释了 CUB 在不同物种间的变异;而从自然选择机制的角度看,影响生物体适应性的同义突变在进化过程中可能被促进或抑制,从而导致整个基因组或特定基因的密码子使用发生变化<sup>[19]</sup>。深入理解植物的密码子使用模式有助于揭示植物物种的环境适应和生物多样性的分子机制<sup>[13]</sup>。此外,高表达的蛋白质通常由使用最优密码子的基因编码<sup>[20]</sup>。通过对基因进行密码子优化,可以提高基因的蛋白质表达水平<sup>[21]</sup>。鉴于扩展蛋白基因作为作物改良的重要候选基因<sup>[6]</sup>,本研究针对木薯这一典型的块根作物,从基因组层面对其扩展蛋白基因成员进行了密码子偏好性分析,探讨了 CUB 的强弱、成因、最优密码子的筛选等方面的内容。

突变引起的碱基组成变化是影响 CUB 的关键因素之一<sup>[22]</sup>。G+C 含量,尤其是 GC<sub>3</sub>,是密码子使用模式的直接反映指标<sup>[13]</sup>。木薯扩展蛋白基因的 GC<sub>1</sub> 和 GC<sub>3</sub> 与有效密码子数(ENC)紧密相关,这进一步证实了碱基组成是评估 CUB 的关键指标。在不同植物类群中,密码子使用偏好存在显著差异;双子叶植物倾向于使用以 A/U 结尾的密码子,而单子叶植物则偏好以 G/C 结尾的密码子<sup>[23]</sup>。MeEXPs

的密码子整体 G+C 含量为 46.98%,GC<sub>3</sub> 为 43.63%,表明木薯扩展蛋白基因的密码子倾向于以 A/U 结尾,与双子叶植物的普遍趋势一致。此外,向日葵中的 WRKY 基因<sup>[24]</sup>、十字花科中的 HSP20 基因<sup>[25]</sup> 以及苜蓿中的 MsPIPs 基因<sup>[26]</sup> 均表现出对以 A/U 结尾的密码子的使用偏好,进一步证实了双子叶植物中这一趋势的普遍性。单子叶植物中的 FAD7 基因和水稻的 WRKY 基因表现出对以 G/C 结尾的密码子的使用偏好<sup>[27]</sup>,这表明在基因家族层面,单子叶植物与双子叶植物在密码子使用模式上存在显著差异。

在研究 CUB 的成因时,中性绘图是重要的分析工具,它能够揭示自然选择与突变压力对 CUB 的相对影响<sup>[28]</sup>。MeEXPs 的密码子中性绘图结果显示,GC<sub>3</sub> 与 GC<sub>12</sub> 之间存在相关性,但这种相关性并不显著( $R^2=0.043\ 77, P=0.183\ 60$ ),表明 MeEXPs 密码子的第三位碱基与前两位相比存在较大差异,密码子的偏好受到突变压力的作用。进一步的分析结果显示,回归直线的斜率(回归系数)约为 0.08,说明突变压力对整体 CUB 的贡献约为 8%,与突变压力相比,自然选择在密码子组成成分上发挥了更重要的作用。基于同义密码子第 3 位的 G+C 含量(GC<sub>3s</sub>)绘制了 ENC-plot 图,分析结果表明,与大部分基因(28 个)相比,较少数量的基因(14 个)接近预期曲线,这暗示突变压力相对于自然选择而言,对同义密码子偏好性的影响较小。中性绘图和 ENC-plot 分析结果均显示自然选择是影响 MeEXPs 密码

子 *CUB* 的主要因素。基因家族 *CUB* 成因主要受自然选择影响的现象在辣椒的 *PPO*<sup>[29]</sup>、*PIF*<sup>[30]</sup> 及苜蓿的 *PIP* 家族基因<sup>[26]</sup> 中均有观察到,然而,在单子叶植物中,例如香蕉的 *Aux/IAA* 基因家族<sup>[31]</sup> 和大蒜的 *WOX* 基因家族<sup>[32]</sup>,*CUB* 的成因主要受突变压力影响,这些差异可能与物种的进化历史、生物学特性及其面临的环境压力有关。木薯 *MeEXPs* 的 *CUB* 受到突变压力和自然选择共同作用的观点,在 *PR2*-plot 绘图中得到进一步支持,在 *PR2*-plot 分析中,极少量的基因位于中心坐标(0.5,0.5),而大部分基因位于第三象限,这表明同义位点上不同碱基的使用频率有所不同。这也表明,在形成 *MeEXPs* 密码子 *CUB* 时,自然选择的影响可能大于突变压力。

高表达基因通常具有更高频率的“最优”密码子使用<sup>[20,33]</sup>。通过 *CUB* 分析筛选的最优密码子,显著提高了翻译效率和蛋白质表达水平。利用植物首选密码子合成的 *cry2AX1* 基因在水稻和棉花中的表达,对多种鳞翅目昆虫显示出显著的抗性<sup>[34-35]</sup>。通过 *CUB* 分析,共筛选出 *MeEXPs* 基因的 23 个最优密码子,这些密码子为提高 *MeEXPs* 基因在木薯细胞中的表达水平和提升重组 *MeEXPs* 蛋白产量提供了理论基础,将来经过试验验证有可能增强木薯扩展蛋白基因的生物功能及适应性。

## 4 结 论

本研究系统分析了木薯扩展蛋白基因的密码子使用偏好性,揭示了自然选择在形成 *CUB* 中的主导作用,鉴定出 23 个最优密码子,这些密码子整体偏好以 A/U 结尾。本研究结果为将来在木薯中实施密码子优化提供了理论基础,并为提升木薯的产量和抗逆性奠定了基础,对作物遗传改良具有一定的指导意义。

## 参考文献:

- [1] COCK J H. Cassava: a basic energy source in the tropics[J]. Science, 1982, 218(4574): 755-762.
- [2] MAVROEIDIS A, ROUSSIS I, KAKABOUKI I. The role of alternative crops in an upcoming global food crisis: a concise review[J]. Foods, 2022, 11(22): 3584.
- [3] MCQUEEN-MASON S J. Expansins and cell wall expansion[J]. Journal of Experimental Botany, 1995, 46(11): 1639-1650.
- [4] 徐 筱, 徐 倩, 张 曙, 等. 植物扩展蛋白基因的研究进展[J]. 北京林业大学学报, 2010, 32(5): 154-162.
- [5] CHOI D, CHO H T, LEE Y. Expansins: expanding importance in plant growth and development[J]. Physiologia Plantarum, 2006, 126(4): 511-518.
- [6] MAROWA P, DING A M, KONG Y Z. Expansins: roles in plant growth and potential applications in crop improvement[J]. Plant Cell Reports, 2016, 35(5): 949-965.
- [7] CRICK F H C. Codon—anticodon pairing: the wobble hypothesis[J]. Journal of Molecular Biology, 1966, 19(2): 548-555.
- [8] HERSHBERG R, PETROV D A. Selection on codon bias[J]. Annual Review of Genetics, 2008, 42: 287-299.
- [9] SUEOKA N. Directional mutation pressure and neutral molecular evolution[J]. Proceedings of the National Academy of Sciences of the United States of America, 1988, 85(8): 2653-2657.
- [10] WRIGHT F. The ‘effective number of codons’ used in a gene[J]. Gene, 1990, 87(1): 23-29.
- [11] 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.
- [12] GENG X S, HUANG N, ZHU Y L, et al. Codon usage bias analysis of the chloroplast genome of cassava[J]. South African Journal of Botany, 2022, 151: 970-975.
- [13] MAZUMDAR P, BINTI OTHMAN R, MEBUS K, et al. Codon usage and codon pair patterns in non-grass monocot genomes[J]. Annals of Botany, 2017, 120(6): 893-909.
- [14] 龚明贵, 白 娜, 李 影, 等. 硬头黄竹叶叶绿体基因组密码子偏好性分析[J]. 江苏农业科学, 2024, 52(3): 67-72.
- [15] 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.
- [16] 包国媛, 李文辛, 杨鑫光, 等. 海甜菜线粒体和叶绿体基因组密码子使用偏好性分析[J]. 江苏农业学报, 2023, 39(9): 1804-1817.
- [17] 李亚麒, 严 炜, 姜予强, 等. 高产小粒咖啡叶绿体基因组密码子偏好性分析[J]. 南方农业学报, 2023, 54(8): 2330-2339.
- [18] 秦斗文, 徐庭亮, 闫京艳, 等. 柔毛郁金香叶绿体基因组密码子偏好性分析[J]. 江苏农业科学, 2023, 51(22): 41-47.
- [19] PLOTKIN J B, KUDLA G. Synonymous but not the same: the causes and consequences of codon bias[J]. Nature Reviews Genetics, 2011, 12(1): 32-42.
- [20] FRUMKIN I, LAJOIE M J, GREGG C J, et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency[J]. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115(21): E4940-E4949.
- [21] KWON K C, CHAN H T, LEÓN I R, et al. Codon optimization to enhance expression yields insights into chloroplast translation[J]. Plant Physiology, 2016, 172(1): 62-77.
- [22] SALIM H M W, CAVALCANTI A R O. Factors influencing codon usage bias in genomes[J]. Journal of the Brazilian Chemical Society, 2008, 19(2): 257-262.
- [23] FENNOY S L, BAILEY-SERRES J. Synonymous codon usage in *Zea mays* L. nuclear genes is varied by levels of C and G-ending



- codons[J]. *Nucleic Acids Research*, 1993, 21(23): 5294-5300.
- [24] GAO Y, LU Y, SONG Y, et al. Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus* [J]. *BMC Genomic Data*, 2022, 23(1): 46.
- [25] JI H Y, LIU J N, CHEN Y N, et al. Bioinformatic analysis of codon usage bias of *HSP20* genes in four cruciferous species [J]. *Plants*, 2024, 13(4): 468.
- [26] 张海霞, 王玉道, 许雪妮. 苜蓿质膜内在蛋白编码基因 *MsPIPs* 家族的密码子偏好性分析 [J]. *江苏农业学报*, 2021, 37(6): 1393-1399.
- [27] MA Q P, LI C, WANG J, et al. Analysis of synonymous codon usage in *FAD7* genes from different plant species [J]. *Genetics and Molecular Research*, 2015, 14(1): 1414-1422.
- [28] LI T T, MA Z, DING T M, et al. Codon usage bias and phylogenetic analysis of chloroplast genome in 36 Gracilariaceae species [J]. *Functional & Integrative Genomics*, 2024, 24(2): 45.
- [29] 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.
- [30] 陈 悦, 仇学文, 李 钰, 等. 辣椒和番茄 *PIF* 基因家族密码子偏好性分析 [J]. *分子植物育种*, 2023, 21(17): 5602-5610.
- [31] 孙雪丽, 郝向阳, 李 蓉, 等. 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.
- [32] 董新玉, 邓 浪, 张亚春, 等. 大蒜全基因组 *WOX* 基因家族鉴定及密码子使用偏性分析 [J]. *西南农业学报*, 2022, 35(2): 285-293.
- [33] ZHOU Z P, DANG Y K, ZHOU M, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(41): E6117-E6125.
- [34] MANIKANDAN R, BALAKRISHNAN N, SUDHAKAR D, et al. Transgenic rice plants expressing synthetic *cry2AXI* gene exhibits resistance to rice leaf folder (*Cnaphalocrosis medinalis*) [J]. *3 Biotech*, 2016, 6(1): 10.
- [35] CHAKRABORTY M, REDDY P S, MUSTAFA G, et al. Transgenic rice expressing the *cry2AXI* gene confers resistance to multiple lepidopteran pests [J]. *Transgenic Research*, 2016, 25(5): 665-678.

(责任编辑:陈海霞)