

龚意辉,唐诗怡,周桂花,等. 黄荆坪竹根椒叶绿体基因组密码子使用偏好性及影响因素分析[J]. 江苏农业科学,2023,51(20):28-34.  
doi:10.15889/j.issn.1002-1302.2023.20.005

# 黄荆坪竹根椒叶绿体基因组密码子使用偏好性及影响因素分析

龚意辉<sup>1</sup>, 唐诗怡<sup>1</sup>, 周桂花<sup>1</sup>, 陈磊<sup>1</sup>, 张斌<sup>1</sup>, 李鹏<sup>2</sup>

(1. 湖南人文科技学院农业与生物技术学院,湖南娄底 417000; 2. 湘潭市农业科学研究所,湖南湘潭 411134)

**摘要:**为了解黄荆坪竹根椒(*Capsicum annuum* L. cv Huang Jingping)叶绿体基因组密码子使用模式,以其 51 条蛋白编码序列(coding sequence,CDS)为研究对象,采用 Codon W 1.4.2、CUSP、Excel 2010、SPSS 16.0 等软件对黄荆坪竹根椒密码子的 GC 含量、有效密码子值(ENC)、RSCU、中性绘图、ENC - plot、PR2 - plot、最优密码子及其影响因素进行详细的分析。结果表明,黄荆坪竹根椒密码子第 1 位(GC<sub>1</sub>)、第 2 位(GC<sub>2</sub>)、第 3 位(GC<sub>3</sub>)中的 GC 含量分别为 42.68%、36.25%、27.89%,其中 GC<sub>1</sub> > GC<sub>2</sub> > GC<sub>3</sub>,表明密码子第 3 位碱基以 A/U 结尾为主;ENC 值介于 34.10% ~ 54.31% 之间,平均值为 46.55%,表明黄荆坪竹根椒密码子偏好性较弱。相关性分析结果表明:(1)GC<sub>1</sub>与 GC<sub>2</sub>呈极显著正相关,GC<sub>all</sub>与 GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub>均为极显著正相关;(2)ENC 与 GC<sub>1</sub>、GC<sub>2</sub> 均不存在显著相关关系,但与 GC<sub>3</sub>存在极显著相关关系。RSCU 分析结果显示:在黄荆坪竹根椒叶绿体基因组中共有 30 个高频密码子的 RSCU > 1,其中以 U、A、G 结尾的高频密码子数分别为 16、13、1 个,表明黄荆坪竹根椒密码子第 3 位偏好以 A 或 U 结尾。ENC - plot 绘图分析结果表明,黄荆坪竹根椒基因组中大多数基因位于标准曲线下方,有 13 个基因的 ENC 比值介于 -0.05 ~ 0.05 之间,说明选择压力主要影响黄荆坪竹根椒密码子的偏好性;PR2 - plot 分析结果表明,黄荆坪竹根椒叶绿体基因组中大部分基因位于平面图的右下方,即 T > A、G > C,表明碱基突变、自然选择压力等多种因素共同影响黄荆坪竹根椒密码子的偏好性。在黄荆坪竹根椒叶绿体基因组中共筛选出 GCA、GCU、AGA 等 19 个最优密码子并全部以 A/U 结尾。本研究结果可为进一步开展黄荆坪竹根椒性状改良、遗传进化及叶绿体基因工程的应用提供分子理论支撑。

**关键词:**黄荆坪竹根椒;叶绿体基因组;密码子偏好性;最优密码子

中图分类号:S641.301 文献标志码:A 文章编号:1002-1302(2023)20-0028-06

密码子是构成原核生物与真核生物遗传信息的基本单元,在生物体转录、翻译过程中起着非常重要的作用。自然界中 64 种密码子编码 20 种氨基酸和 3 种终止密码子(TGA、TAA 和 TAG),除甲硫氨酸(Met)和酪氨酸(Trp)外,其余 18 种氨基酸均有多个同义密码子与之对应。同义密码子在不同物种叶绿体基因组中,甚至在同种物种同样基因中的使用频率也不是均等的<sup>[1]</sup>。同义密码子使用频率的差异性称之为密码子偏好性(synonymous codon

usage bias,SCUB)。越来越多的研究表明,不同植物中密码子的偏好性不同,例如天山雪莲(*Saussurea involucrata*)<sup>[2]</sup>、云南樟(*Cinnamomum glanduliferum*)<sup>[3]</sup>、橄榄(*Canarium album*)<sup>[4]</sup>。因此,对植物密码子的偏好性进行深入探究,不仅有利于明确植物的分类与鉴定<sup>[5]</sup>,而且还可根据最优密码子设计植物的外源基因,进一步提高外源基因的表达丰度及转化效率,从分子角度全面解析植物的进化过程<sup>[6]</sup>。

叶绿体是植物进行光合作用的重要场所,为植物的正常生长提供源源不断的营养物质。叶绿体拥有一套完整的遗传信息,能进行母系遗传,长度大小一般为 120 ~ 160 kb,具有高度保守的特征<sup>[7]</sup>。因此,叶绿体基因组学在物种遗传分类、鉴定、进化规律、基因工程等方面有重要的研究意义<sup>[8-9]</sup>。植物叶绿体全基因组伴随着现代高通量测序技术的快速发展而不断被美国国家生物技术信息中心(NCBI)数据库收录,已有越来越多的学者在植物中

收稿日期:2022-12-22

基金项目:湖南省教育厅科学研究重点项目(编号:20A281);湖南省自然科学基金面上项目(编号:2021JJ30376);湖南省科技特派员服务乡村振兴项目(编号:2022NK4218);湘潭市农业科学研究所横向项目(编号:380220590112)。

作者简介:龚意辉(1988—),男,湖南涟源人,博士,讲师,主要从事园艺作物分子生物学方向研究。E-mail:gyhzgh@163.com。

通信作者:李鹏,农艺师,主要从事果蔬分子生物学方向研究。E-mail:907653797@qq.com。

开展叶绿体基因组密码子的偏好性研究,例如,萝卜 (*Raphanus sativus* L.)<sup>[10]</sup>、苋菜 (*Amaranthus tricolor* L.)<sup>[11]</sup>、莲藕 (*Nelumbo nucifera* Gaertn.)<sup>[12]</sup>。因此,分析植物密码子的偏好性及其影响因素不仅可以确定该物种的最优密码子,然后根据最优密码子设计基因表达载体,进一步提高外源基因的表达水平,从而有利于阐明植物适应外界环境的分子机制。

黄荆坪竹根椒 (*Capsicum annuum* L. cv Huang Jingping) 是茄科 (Solanaceae) 辣椒属 (*Capsicum*) 一年或有限多年生草本植物,主要分布在湘潭市黄荆坪排头乡。黄荆坪竹根椒因个体细长、色泽鲜艳、皮薄肉多、味道鲜美而深受广大消费者的青睐。近年来,黄荆坪竹根椒规模及产业发展迅速,现已成为湘潭市黄荆坪排头乡实现乡村振兴、农民致富的重要支柱产业。目前有关辣椒的研究主要集中在高产高效技术栽培<sup>[13]</sup>、遗传育种<sup>[14]</sup>、品质评价<sup>[15]</sup>等方面,随着黄荆坪竹根椒全基因组测序和拼接组装工作的完成,目前国内外有关黄荆坪竹根椒密码子偏好性的研究还尚未见报道。本研究对黄荆坪竹根椒密码子的碱基组成及其各相关参数进行分析,首次阐明黄荆坪竹根椒密码子偏好性使用的特征,从而在黄荆坪竹根椒叶绿体基因组中遴选出最优密码子,并明确影响黄荆坪竹根椒密码子偏好性的主要因素,以期为今后进一步开展黄荆坪竹根椒性状改良、进化规律及叶绿体基因工程的应用提供分子理论支撑。

## 1 材料与方法

### 1.1 黄荆坪竹根椒叶绿体全基因序列的获得

2022 年 5 月 15 日采集湘潭市农业科学研究所种苗中心的黄荆坪竹根椒幼嫩叶片,经液氮速冻后送华智生物技术有限公司进行叶绿体基因组测序,采用 Illumina/MGI 测序平台完成了黄荆坪竹根椒叶绿体基因组的测序,全长为 156 817 bp,包括 82 个蛋白编码基因 (CDS), GenBank 登录号为 OP919650.1。在筛选黄荆坪竹根椒蛋白编码基因时,首先删除重复序列,然后选择 ATG 作为起始密码子,TGA、TAA 和 TAG 为终止密码子,且以编码序列大于 300 bp 的叶绿体基因用于密码子偏好性分析,最后共得到 51 个满足条件的 CDS 序列。

### 1.2 中性绘图分析

参照 Liu 等的方法<sup>[16]</sup>,计算黄荆坪竹根椒叶绿

体密码子第 3 位中的 GC 含量 (GC<sub>3</sub>),以 GC<sub>3</sub> 为横坐标,密码子第 1、第 2 位中的平均 GC 含量 (GC<sub>12</sub>) 为纵坐标,制作散点图进行相关性分析。

### 1.3 相对同义密码子使用度分析

参照原晓龙等的方法<sup>[17]</sup>,使用 CodonW 软件分别统计黄荆坪竹根椒各密码子的相对同义密码子使用度 (relative synonymous codon usage, RSCU),将黄荆坪竹根椒密码子的 RSCU > 1 定义为高频密码子。

### 1.4 ENC - plot 绘图分析

参照 Fuglsang 的方法<sup>[18]</sup>,分别以黄荆坪竹根椒的 GC<sub>3s</sub>、有效密码子 ENC (effective number of condon) 为横坐标和纵坐标,绘制散点图,其 ENC 的计算公式为  $ENC = 2 + GC_{3s} + 29 / [GC_{3s}^2 + (1 - GC_{3s})^2]$ 。结合黄荆坪竹根椒的 ENC 比值频数分布对其差异进行定量分析,以 ENC 值为 -0.05 ~ 0.05 区间作为具体的判断标准。

### 1.5 PR2 - plot 绘图分析

参照 Sueoka 的方法<sup>[19]</sup>,分别统计黄荆坪竹根椒密码子第 3 位中碱基组成 (G<sub>3</sub>、C<sub>3</sub>、A<sub>3</sub>、T<sub>3</sub>) 比例,分别以 G<sub>3</sub>/(G<sub>3</sub> + C<sub>3</sub>)、A<sub>3</sub>/(A<sub>3</sub> + T<sub>3</sub>) 为横坐标和纵坐标进行黄荆坪竹根椒的 PR2 偏倚分析。

### 1.6 最优密码子分析

参照原晓龙等的方法<sup>[17]</sup>,最优密码子以黄荆坪竹根椒的 ENC 值作为衡量密码子偏性标准,分别在两端挑选最高和最低的 10% 基因构建黄荆坪竹根椒高低基因表达库,分别计算黄荆坪竹根椒各高低基因的 RSCU 和 ΔRSCU 值,将同时符合高频率密码子 (RSCU > 1) 和高表达密码子 (ΔRSCU ≥ 0.08) 2 个条件的密码子定义为最优密码子。

## 2 结果与分析

### 2.1 黄荆坪竹根椒相关偏性指数分析

利用在线软件 Codon W 1.4.2 和 CUSP 软件对黄荆坪竹根椒叶绿体中的 51 条 CDS 序列进行分析,其各基因不同位置的 GC 含量和 ENC 值如表 1 所示,51 条 CDS 密码子中的总 GC 含量介于 30.73% ~ 45.32% 之间,平均 GC 含量 (GC<sub>all</sub>) 为 35.61%,GC<sub>1</sub> 含量介于 34.39% ~ 58.88% 之间,平均 GC 含量为 42.68%,GC<sub>2</sub> 含量介于 29.13% ~ 56.83% 之间,平均 GC 含量为 36.25%,GC<sub>3</sub> 含量介于 18.52% ~ 37.39% 之间,平均 GC 含量为 27.89%,其结果说明黄荆坪竹根椒密码子不同位置

表 1 黄荆坪竹根椒密码子不同位置的 GC 含量 %

基因	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>	ENC
<i>accD</i>	40.40	36.78	30.25	35.81	47.13
<i>atpA</i>	55.71	39.96	26.77	40.81	46.20
<i>atpB</i>	57.11	41.48	31.06	43.22	51.27
<i>atpE</i>	53.73	39.55	30.60	41.29	49.23
<i>atpF</i>	46.49	35.14	32.43	38.02	43.20
<i>atpI</i>	49.60	37.50	27.42	38.17	44.62
<i>ccsA</i>	34.39	38.22	26.75	33.12	46.75
<i>cemA</i>	38.70	29.13	33.04	33.62	48.58
<i>clpP1</i>	58.88	38.07	32.49	43.15	51.75
<i>matK</i>	37.65	32.75	27.84	32.75	48.99
<i>ndhA</i>	43.96	39.01	24.18	35.71	42.04
<i>ndhB</i>	41.68	38.55	31.51	37.25	46.35
<i>ndhC</i>	46.28	32.23	31.40	36.64	51.33
<i>ndhE</i>	39.22	32.35	24.51	32.03	42.26
<i>ndhF</i>	37.52	36.17	26.32	33.33	42.35
<i>ndhG</i>	42.94	34.46	20.90	32.77	41.72
<i>ndhH</i>	51.02	36.55	29.19	38.92	48.99
<i>ndhI</i>	44.05	35.12	28.57	35.91	52.65
<i>ndhJ</i>	50.94	37.74	35.22	41.30	53.95
<i>ndhK</i>	44.76	43.15	27.02	38.31	49.59
<i>pafl</i>	48.52	37.87	30.77	39.05	54.31
<i>paflI</i>	43.24	40.54	30.81	38.20	51.83
<i>petA</i>	52.34	37.69	33.96	41.33	52.54
<i>petB</i>	49.07	41.67	32.87	41.20	47.95
<i>petD</i>	50.93	38.51	24.22	37.89	41.77
<i>psaA</i>	52.86	43.41	32.49	42.92	47.66
<i>psaB</i>	48.98	43.13	31.56	41.22	46.66
<i>psbA</i>	49.72	43.50	35.31	42.84	41.58
<i>psbB</i>	54.81	46.37	29.86	43.68	44.84
<i>psbD</i>	51.69	43.50	33.33	42.84	45.84
<i>rbcl</i>	58.16	43.31	30.13	43.86	48.15
<i>rpl14</i>	55.28	37.40	24.39	39.02	43.23
<i>rpl16</i>	51.85	52.59	18.52	40.99	36.31
<i>rpl2</i>	50.55	48.36	32.36	43.76	51.51
<i>rpl20</i>	40.88	46.54	35.22	40.88	43.99
<i>rpl22</i>	39.10	39.10	29.49	35.90	48.54
<i>rpoA</i>	43.49	32.25	25.44	33.73	45.70
<i>rpoB</i>	50.89	38.38	27.08	38.78	47.10
<i>rpoC1</i>	51.17	37.68	27.57	38.81	47.58
<i>rpoC2</i>	46.80	37.87	30.74	38.47	48.80
<i>rps11</i>	55.40	56.83	23.74	45.32	48.54
<i>rps12</i>	51.61	48.39	28.23	42.74	43.82
<i>rps14</i>	43.56	46.53	33.66	41.25	36.35
<i>rps18</i>	37.25	41.18	27.45	35.29	45.62
<i>rps2</i>	43.04	43.46	29.54	38.68	48.96
<i>rps3</i>	46.12	34.70	25.11	35.31	48.83
<i>rps4</i>	51.49	37.13	25.25	37.95	52.89
<i>rps7</i>	51.92	44.87	23.08	39.96	42.28
<i>rps8</i>	42.22	40.00	22.96	35.06	34.10
<i>ycf1</i>	36.50	29.37	26.32	30.73	47.02
<i>ycf2</i>	42.20	35.11	37.39	38.23	51.15
平均值	42.68	36.25	27.89	35.61	46.55

注:GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub> 分别表示密码子第 1、2、3 位碱基中的 GC 含量,GC<sub>all</sub>:密码子总 GC 含量。

的 GC 含量存在差异,GC<sub>1</sub> > GC<sub>2</sub> > GC<sub>3</sub>,表明在黄荆坪竹根椒密码子第 3 位多为 A/T 碱基;ENC 值介于 34.10% ~ 54.31% 之间,平均值为 46.55%,其中 35 条 CDS 的 ENC 值高于 45.00%,占总编码基因数的 68.63%;可判断出黄荆坪竹根椒密码子的使用偏好性较弱。

对黄荆坪竹根椒叶绿体密码子各项系数进行相关性分析(表 2),GC<sub>all</sub>与 GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub> 均呈极显著相关( $P < 0.01$ );GC<sub>1</sub> 与 GC<sub>2</sub> 呈极显著相关,但与 GC<sub>3</sub> 相关性不显著,说明黄荆坪竹根椒密码子第 1 位和第 2 位碱基组成存在相似性;ENC 与 GC<sub>3</sub> 呈极显著相关,其相关系数为 0.448,但与 GC<sub>1</sub> 和 GC<sub>2</sub> 并不存在显著相关性,说明 GC<sub>3</sub> 碱基组成主要影响黄荆坪竹根椒密码子的偏好性。

表 2 黄荆坪竹根椒密码子的相关性分析

密码子	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>	0.442 **	1.000	-0.028	0.751 **	-0.237
GC <sub>3</sub>	0.053	-0.028	1.000	0.389 **	0.448 **
GC <sub>all</sub>	0.822 **	0.751 **	0.389 **	1.000	0.139
ENC	0.158	-0.237	0.448 **	0.139	1.000

注:\*\*表示极显著相关( $P < 0.01$ )。

根据黄荆坪竹根椒的 RSCU 分析可知,在黄荆坪竹根椒叶绿体基因组中共有 30 个高频密码子的 RSCU > 1,其中以 U、A、G 结尾的高频密码子数分别为 16、13、1 个(表 3),表明黄荆坪竹根椒密码子第 3 位偏好以 A 或 U 结尾。

2.2 中性绘图分析

中性绘图分析结果(图 1)表明,GC<sub>3</sub> 取值介于 18.52% ~ 37.39% 之间,GC<sub>12</sub>取值介于 31.76% ~ 57.86% 之间,黄荆坪竹根椒全部基因分布在对角线之上,回归系数为 0.022 1, $r^2$  为 0.02,说明黄荆坪竹根椒叶绿体基因组中 GC<sub>12</sub>和 GC<sub>3</sub> 的相关性显著,说明黄荆坪竹根椒密码子第 1 位和第 2 位与第 3 位的碱基组成不同但不存在显著性相关,说明自然选择主要影响黄荆坪竹根椒偏好性的形成。

2.3 ENC - plot 绘图分析

黄荆坪竹根椒叶绿体基因组密码子的 ENC - plot 分析结果(图 2)表明,黄荆坪竹根椒少数基因沿着标准曲线周围分布,说明黄荆坪竹根椒密码子的实际 ENC 值与期望 ENC 值接近,大部分基因则分布在标准曲线下方,并且与标准曲线相距较远,表明自然选择主要影响黄荆坪竹根椒叶绿体基因

表 3 黄荆坪竹根椒叶绿体基因组相对同义密码子使用度分析

氨基酸	密码子	数量	RSCU	氨基酸	密码子	数量	RSCU
Ter	UGA	13	0.76	Met	AUG	456	1.00
	UAA	26	1.53	Asn	AAU	787	1.53
	UAG	12	0.71		AAC	242	0.47
Ala	GCU	489	1.76	Pro	CCC	167	0.77
	GCC	194	0.70		CCU	334	1.54
	GCA	312	1.13		CCG	118	0.55
	GCG	114	0.41		CCA	246	1.14
Cys	UGC	61	0.54	Gln	CAA	572	1.51
	UGU	166	1.46		CAG	188	0.49
Asp	GAU	672	1.59	Arg	CGA	307	1.50
	GAC	171	0.41		CGG	91	0.45
Glu	GAG	287	0.52		CGU	266	1.30
	GAA	826	1.48		AGG	122	0.60
Phe	UUC	381	0.68		CGC	82	0.40
	UUU	737	1.32		AGA	359	1.76
Gly	GGA	535	1.51	Ser	UCG	157	0.61
	GGG	247	0.70		UCA	309	1.19
	GGU	462	1.31		AGU	326	1.26
	GGC	170	0.48		AGC	89	0.34
His	CAC	111	0.46		UCC	230	0.89
	CAU	374	1.54		UCU	445	1.72
Ile	AUA	527	0.92	Thr	ACA	310	1.18
	AUU	865	1.50		ACG	116	0.44
	AUC	335	0.58		ACU	419	1.60
Lys	AAA	825	1.52		ACC	204	0.78
	AAG	258	0.48	Val	GUA	428	1.51
Leu	CUC	142	0.39		GUG	151	0.53
	CUU	456	1.27		GUU	430	1.52
	CUG	143	0.40		GUC	126	0.44
	UUG	448	1.24	Trp	UGG	366	1.00
	CUA	270	0.75	Tyr	UAC	154	0.41
	UUA	701	1.95		UAU	605	1.59

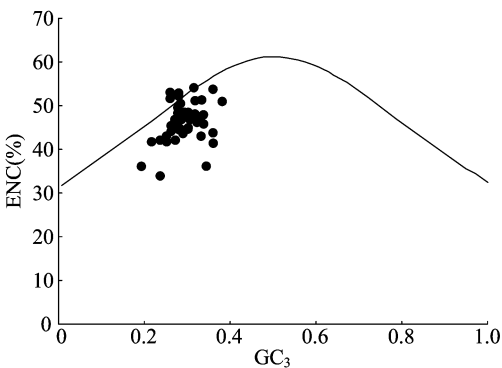


图 2 黄荆坪竹根椒叶绿体基因 ENC-plot 绘图分析

表 4 黄荆坪竹根椒叶绿体基因 ENC 比值频数分布

组段	组中值	组数(个)	组频
-0.15 ~ -0.05	-0.10	1	0.02
-0.05 ~ 0.05	0.00	13	0.25
0.05 ~ 0.15	0.10	30	0.59
0.15 ~ 0.25	0.20	4	0.08
0.25 ~ 0.35	0.30	3	0.06
合计		51	1.00

这些基因与标准曲线相距较远,说明选择压力主要影响黄荆坪竹根椒密码子的偏好性。

2.4 PR2 - plot 分析

采用 PR2 - plot 绘图对黄荆坪竹根椒叶绿体基因氨基酸 A、G 和 T、C 之间的关系进行分析(图 3),在黄荆坪竹根椒叶绿体基因组中大多数基因位于平面图的下半部分,说明 A、G、T、C 4 种碱基的使用频率存在差异,即  $T > A, G > C$ ;这表明碱基突变、自然选择压力等多种因素共同影响黄荆坪竹根椒密码子的偏好性。

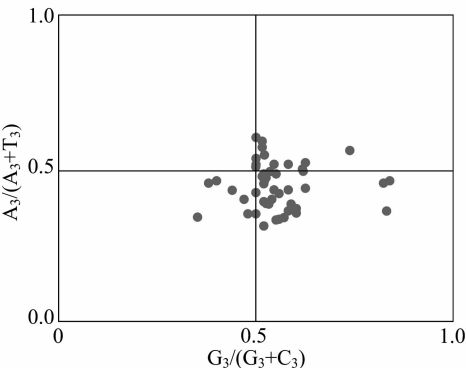


图 3 黄荆坪竹根椒叶绿体基因 PR2-plot 绘图分析

2.5 最优密码子的确定

对黄荆坪竹根椒 51 条 CDS 的 ENC 值进行排序,从两极分别挑选 10% 的基因建立高低表达库,计算各密码子的 RSCU 值、 $\Delta RSCU$  值,将同时符合  $RSCU > 1$  且  $\Delta RSCU \geq 0.08$  这 2 个条件的密码子确

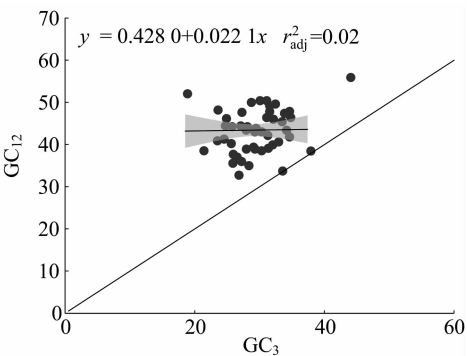


图 1 黄荆坪竹根椒叶绿体基因中性绘图分析

组密码子的偏好性。通过计算黄荆坪竹根椒密码子的 ENC 比值(表 4),ENC 比值介于 -0.05 ~ 0.05 区间的基因共有 13 个,而 ENC 比值位于这个范围之外的基因共有 38 个,且距期望 ENC 值较远,说明

定为黄荆坪竹根椒的最优密码子(表 5),结合表 3 统计的 30 个 RSCU>1 的高频密码子,有 26 个密码子的  $\Delta\text{RSCU}\geq 0.08$ ,最终在黄荆坪竹根椒叶绿体基因组中筛选出 19 个最优密码子,分别为 GCA、GCU、

AGA、CGA、UGU、CAA、GAA、GGU、AUU、CUA、CUU、AAA、CCA、CCU、UCU、ACA、ACU、GUA、GUU,其中 10 个最优密码子以 A 结尾,9 个最优密码子以 U 结尾,说明黄荆坪竹根椒密码子偏好使用 A/U。

表 5 黄荆坪竹根椒最优密码子分析

氨基酸	密码子	高表达基因		低表达基因		$\Delta\text{RSCU}$	氨基酸	密码子	高表达基因		低表达基因		$\Delta\text{RSCU}$
		数目	RSCU	数目	RSCU				数目	RSCU	数目	RSCU	
Ala	GCA *	20	1.21	14	0.95	0.26	Leu	CUA ***	21	1.45	10	0.65	0.80
	GCC	5	0.30	14	0.95	-0.65		CUC	0	0.00	4	0.26	-0.26
	GCG	3	0.18	7	0.47	-0.29		CUG	2	0.14	5	0.32	-0.18
	GCU ***	38	2.30	24	1.63	0.67		CUU *	24	1.66	22	1.42	0.24
Arg	AGA **	16	1.35	11	0.94	0.41	Lys	UUA	24	1.66	27	1.74	-0.08
	AGG *	9	0.76	6	0.51	0.25		UUG	16	1.10	25	1.61	-0.51
	CGA *	24	2.03	21	1.80	0.23		AAA *	20	1.54	38	1.38	0.16
	CGC	3	0.25	7	0.60	-0.35	Met	AAG	6	0.46	17	0.62	-0.16
	CGG *	3	0.25	2	0.17	0.08		AUG	28	1.00	20	1.00	0.00
	CGU	16	1.35	23	1.97	-0.62	Phe	UUC *	24	1.00	14	0.72	0.28
Asn	AAC ***	19	0.88	8	0.31	0.57		UUU	24	1.00	25	1.28	-0.28
	AAU	24	1.12	43	1.69	-0.57	Pro	CCA **	12	1.37	14	1.00	0.37
Asp	GAC ***	7	0.67	2	0.10	0.57		CCC	5	0.57	16	1.14	-0.57
	GAU	14	1.33	40	1.90	-0.57		CCG	4	0.46	6	0.43	0.03
Cys	UGC	0	0.00	6	0.80	-0.80		CCU *	14	1.60	20	1.43	0.17
	UGU ***	7	2.00	9	1.20	0.80	Ser	AGC	5	0.42	4	0.41	0.01
Gln	CAA *	17	1.55	37	1.45	0.10		AGU	15	1.25	12	1.22	0.03
	CAG	5	0.45	14	0.55	-0.10		UCA	13	1.08	10	1.02	0.06
Glu	GAA *	34	1.55	47	1.47	0.08	Thr	UCC	10	0.83	11	1.12	-0.29
	GAG	10	0.45	17	0.53	-0.08		UCG	7	0.58	6	0.61	-0.03
Gly	GGA	25	1.35	22	1.38	-0.03		UCU *	22	1.83	16	1.63	0.20
	GGC	7	0.38	10	0.62	-0.24	Trp	ACA *	14	1.30	11	1.02	0.28
	GGG	4	0.22	13	0.81	-0.59		ACC *	9	0.84	8	0.74	0.10
	GGU ***	38	2.05	19	1.19	0.86		ACG	2	0.19	8	0.74	-0.55
Ile	AUA	26	0.85	32	1.08	-0.23		ACU *	18	1.67	16	1.49	0.18
	AUC	18	0.59	17	0.57	0.02	Tyr	UGG	19	1.00	12	1.00	0.00
	AUU *	48	1.57	40	1.35	0.22		UAC *	11	0.71	13	0.49	0.22
Val	GUA **	21	1.79	19	1.31	0.48	His	UAU	20	1.29	40	1.51	-0.22
	GUC	1	0.09	8	0.55	-0.46		CAC	6	0.57	6	0.57	0.00
	GUG	4	0.34	9	0.62	-0.28		CAU	15	1.43	15	1.43	0.00
	GUU *	21	1.79	22	1.52	0.27							

注: \*、\*\*、\*\*\* 分别表示  $\Delta\text{RSCU}\geq 0.08$ 、 $\Delta\text{RSCU}\geq 0.3$ 、 $\Delta\text{RSCU}\geq 0.5$ 。

3 结论与讨论

密码子偏好性是有效分析原核生物和真核生物进化特征的一个重要手段<sup>[20]</sup>。大量研究表明,密码子偏好性受多种生物因素的影响,例如基因表达丰度、基因片段大小、tRNA 丰度、碱基组成及突变等因素均可以影响物种密码子的使用模式<sup>[21]</sup>。然而,已有研究报道碱基突变和自然选择是决定物种密码子使用偏性的重要原因,已被广泛用于阐明物种基因组密码子使用的特异性<sup>[17]</sup>。据报道,在植

物中,核苷酸组成在核基因组密码子偏性形成中发挥重要作用,但自然选择则在植物线粒体基因组和叶绿体基因组密码子偏好性形成中发挥重要作用<sup>[22-23]</sup>。本研究系统分析了影响黄荆坪竹根椒密码子偏好性的主要因素,中性分析表明,黄荆坪竹根椒基因组中 GC<sub>12</sub>和 GC<sub>3</sub>不存在显著相关性,且回归方程斜率较低,说明自然选择可能比突变更能影响黄荆坪竹根椒密码子偏好性的形成。

本研究发现黄荆坪竹根椒密码子 GC<sub>1</sub>(42.68%)>GC<sub>2</sub>(36.25%)>GC<sub>3</sub>(27.89%),GC<sub>3</sub>

与 GC<sub>1</sub>、GC<sub>2</sub> 均不存在显著相关性,但 GC<sub>1</sub> 与 GC<sub>2</sub> 呈极显著相关( $P < 0.01$ ),说明黄荆坪竹根椒密码子第 3 位偏好使用 A 或 U 结尾,使用模式为 NNA 或 NNU,其偏好性主要受自然选择因素的影响。这与茄科同属植物番茄(*Solanum lycopersicum*)<sup>[24]</sup>、枸杞(*Lycium barbarum*)<sup>[25]</sup>等研究结果相似,表明黄荆坪竹根椒与番茄和枸杞的密码子使用模式相似。

本研究发现黄荆坪竹根椒密码子 ENC 平均值为 46.55 > 45,其中 35 条 CDS 的 ENC 值高于 45,可推测这些基因密码子的使用偏性较弱。对黄荆坪竹根椒密码子的中性绘图分析表明,密码子第 3 位(GC<sub>3</sub>)与密码子第 1、2 位(GC<sub>12</sub>)中的 A、T、C、G 组成存在较大差异,GC<sub>3</sub> 含量明显低于 GC<sub>12</sub>,进一步阐明自然选择对黄荆坪竹根椒密码子偏好性的影响较大,这与秋茄(*Kandelia obovata*)<sup>[26]</sup>的研究结果一致。利用 ENC - plot 分析了影响黄荆坪竹根椒密码子偏性因素,自然选择相比于突变因素更加能影响黄荆坪竹根椒密码子偏好性的形成。黄荆坪竹根椒与迎春樱桃(*Prunus discoidea*)<sup>[27]</sup>、酸枣(*Ziziphus jujuba* var. *spinosa*)<sup>[28]</sup>等植物密码子偏好性研究结果相似,而突变是影响籽粒苋(*Amaranthus hypochondriacus*)<sup>[29]</sup>密码子偏好性形成的主要原因,突变和选择在芒果(*Mangifera indica* L.)<sup>[30]</sup>密码子偏好性形成中起同等重要作用。可见,不同植物在自然进化过程中影响密码子偏好性的因素可能存在一定的差异,但具体影响密码子偏好性的机制还有待深入研究。

在黄荆坪竹根椒叶绿体基因组中共筛选出 GCA、GCU、AGA 等 19 个最优密码子,且全部以 A/U 结尾,再次证明了黄荆坪竹根椒密码子使用模式为 NNA 或 NNU,其密码子的使用模式与菠萝(*Ananas comosus*)<sup>[31]</sup>、金花茶(*Camellia nitidissima*)<sup>[32]</sup>等植物的最优密码子 NNA、NNU 的模式相同。黄荆坪竹根椒、菠萝和金花茶等植物密码子偏好使用 NNA 或 NNU 模式,可能是由于这些植物中密码子第 3 位 A、U 含量偏高所造成的。最优密码子分析表明,在利用黄荆坪竹根椒进行基因工程研究设计外源基因时,优先选择 A 或 U 结尾的密码子,在一定程度上可提高其外源基因的表达转化效果,为今后开展其性状改良、种质资源保存和进化机制研究提供理论支撑。

#### 参考文献:

[1] Wei L, He J, Jia X, et al. Analysis of codon usage bias of

- mitochondrial genome in *Bombyx mori* and its relation to evolution [J]. BMC Evolutionary Biology, 2014, 14(1): 262.
- [2] 何亚玲,彭业军,李 锦,等. 天山雪莲叶绿体基因组密码子偏好性分析[J]. 石河子大学学报(自然科学版), 2022, 40(1): 84 - 92.
- [3] 肖明昆,严 炜,熊贤坤,等. 云南樟叶绿体基因组密码子偏好性分析[J]. 中南林业科技大学学报, 2022, 42(6): 127 - 134.
- [4] 赖瑞联,陈 瑾,冯 新,等. 橄榄叶绿体基因组密码子偏好性特征[J]. 福建农林大学学报(自然科学版), 2022, 51(4): 502 - 509.
- [5] Leffler E M, Bullaughey K, Matute D R, et al. Revisiting an old riddle: what determines genetic diversity levels within species? [J]. PLoS Biology, 2012, 10(9): e1001388.
- [6] Christofi M, Pavlou A, Lantzouraki D Z, et al. Profiling carotenoid and phenolic compounds in fresh and canned fruit of peach cultivars: impact of genotype and canning on their concentration[J]. Journal of Food Composition and Analysis, 2022, 114: 104734.
- [7] Niu Y F, Gao C W, Liu J. Complete mitochondrial genomes of three *Mangifera* species, their genomic structure and gene transfer from chloroplast genomes[J]. BMC Genomics, 2022, 23(1): 147.
- [8] Han C Y, Ding R, Zong X Y, et al. Structural characterization of *Platanthera ussuriensis* chloroplast genome and comparative analyses with other species of Orchidaceae [J]. BMC Genomics, 2022, 23(1): 84.
- [9] Lian C L, Yang H, Lan J X, et al. Comparative analysis of chloroplast genomes reveals phylogenetic relationships and intraspecific variation in the medicinal plant *Isodon rubescens* [J]. PLoS One, 2022, 17(4): e0266546.
- [10] 孙继峰,韩太利,徐立功,等. 萝卜 ICE1 密码子使用偏性分析 [J]. 核农学报, 2018, 32(3): 477 - 485.
- [11] 肖 昉,陈 何,陈家兰,等. 莧菜 *AtrADH2* 基因密码子偏好性与进化分析[J]. 分子植物育种, 2022, 20(9): 2821 - 2830.
- [12] 惠小涵,程婷婷,柯卫东,等. 莲藕 *PPO* 基因密码子偏好性特征分析[J]. 江苏农业学报, 2020, 36(2): 438 - 446.
- [13] 闫文涛,张建金,张 婧,等. 生物有机肥用量对日光温室辣椒生长、产量及品质的影响[J]. 中国瓜菜, 2022, 35(8): 86 - 91.
- [14] Barik S, Ponnann N, Reddy A C, et al. Breeding peppers for industrial uses: progress and prospects [J]. Industrial Crops and Products, 2022, 178: 114626.
- [15] Chatterjee D, Dutta S K, Kikon Z J, et al. Recycling of agricultural wastes to vermicomposts: characterization and application for clean and quality production of green bell pepper (*Capsicum annuum* L.) [J]. Journal of Cleaner Production, 2021, 315(4): 128115.
- [16] Liu H B, Lu Y Z, Lan B L, et al. Codon usage by chloroplast gene is bias in *Hemiptelea davidii* [J]. Journal of Genetics, 2020, 99(1): 8.
- [17] 原晓龙,刘 音,康洪梅,等. 蒜头果叶绿体基因组密码子偏好性分析[J]. 西南林业大学学报(自然科学), 2021, 41(3): 15 - 22.
- [18] Fuglsang A. Impact of bias discrepancy and amino acid usage on estimates of the effective number of codons used in a gene, and a test for selection on codon usage [J]. Gene, 2008, 410(1): 82 - 88.
- [19] Sueoka N. Near homogeneity of PR2 - bias fingerprints in the human genome and their implications in phylogenetic analyses [J]. Journal of Molecular Evolution, 2001, 53(4): 469 - 476.

张路阳,张有建,饶聪颖,等. 烟草 *DUF538* 基因家族鉴定及根结线虫胁迫下表达分析[J]. 江苏农业科学,2023,51(20):34-42.  
doi:10.15889/j.issn.1002-1302.2023.20.006

# 烟草 *DUF538* 基因家族鉴定及根结线虫胁迫下表达分析

张路阳<sup>1</sup>, 张有建<sup>2</sup>, 饶聪颖<sup>2</sup>, 许燕彪<sup>2</sup>, 谢 可<sup>2</sup>, 李 伟<sup>2</sup>, 郑 聪<sup>2</sup>

(1. 河南农业大学烟草学院,河南郑州 450002; 2. 福建省南平市烟草公司,福建南平 353000)

**摘要:***DUF538* 基因是没有功能注释的蛋白,在根结线虫胁迫的响应过程中,含有 *DUF538* 结构域的蛋白质的基因表达发生变化。鉴定烟草 *DUF538* 基因家族,为烟草 *DUF538* 基因功能研究提供理论依据,有利于筛选防御根结线虫的有效基因。基于烟草全基因组数据,通过 BLAST 对烟草 *DUF538* 基因家族进行鉴定,利用生物信息学的方法分析了该家族成员的理化性质、基因结构、蛋白质结构域、染色体分布、系统进化及启动子,并利用 qRT-PCR 验证了 *DUF538* 基因家族在 K326 各个组织的表达情况及响应南方根结线虫侵染的表达情况。结果表明,烟草中共鉴定出 8 个 *NtDUF538* 基因,将其分为 3 类;已知的 *NtDUF538* 基因位于不同的染色体上且分布不均匀,所有 *NtDUF538* 基因均具有保守的结构域;*NtDUF538* 基因家族启动子区域显著富集生长发育和激素响应的顺式元件,部分 *NtDUF538* 基因还具有低温胁迫的作用元件;7 个 *NtDUF538* 基因在 K326 不同组织中存在不同程度的表达,*NtDUF538-7* 基因只在叶和花中表达,其中 *NtDUF538-5* 和 *NtDUF538-6* 在遭受根结线虫侵染后表达量增加。研究结果揭示了烟草 *NtDUF538* 基因家族成员在根结线虫侵染过程中起到重要作用,为探究烟草 *NtDUF538* 基因家族的生物学功能提供了理论基础。

**关键词:**烟草;*DUF538* 基因;全基因组鉴定;生物信息学;表达分析;根结线虫

**中图分类号:**S572.01    **文献标志码:**A    **文章编号:**1002-1302(2023)20-0034-09

烟草(*Nicotiana tabacum* L.)是我国重要的经济作物,病虫害是制约其生产的主要因素之一,每年

收稿日期:2023-01-06

基金项目:浙江中烟工业有限责任公司项目(编号:NYK2022-04-03)。

作者简介:张路阳(1997—),男,河南郑州人,硕士研究生,主要从事烟草育种研究。E-mail:zhangluyang97@126.com。

通信作者:郑 聪,硕士,农艺师。E-mail:npyzc@163.com。

因为病虫害造成的经济损失到达 15% 以上,其中烟草根结线虫病是烟草上危害程度深、范围广、防治难度大的一类病害<sup>[1-3]</sup>。目前,有关烟草根结线虫病防治的研究和报道主要在化学防治和生物防治方面,但是烟草根结线虫病并没有得到彻底解决。烟草抗南方根结线虫基因的筛选为南方根结线虫抗性育种提供了新的基因资源,具有重要的理论及

[20] Wang B, Yuan J, Liu J, et al. Codon usage bias and determining forces in green plant mitochondrial genomes [J]. Journal of Integrative Plant Biology, 2011, 53(4): 324-334.

[21] Song Y F, Yang Q H, Yi X G, et al. Comparative analysis of Codon usage patterns in chloroplast genomes of cherries [J]. Forests, 2022, 13(11): 1891.

[22] Morton B R, Wright S I. Selective constraints on Codon usage of nuclear genes from *Arabidopsis thaliana* [J]. Molecular Biology and Evolution, 2007, 24(1): 122-129.

[23] Liu Q P, Xue Q Z. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species [J]. Journal of Genetics, 2005, 84(1): 55-62.

[24] 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.

[25] 段淋渊,戴国礼,焦恩宁,等. 枸杞自交不亲和基因 S-RNase 密码子偏性分析[J]. 西南林业大学学报(自然科学), 2020, 40

(2): 44-52.

[26] 赵 森,邓力华,陈 芬. 秋茄叶绿体基因组密码子使用偏好性分析[J]. 森林与环境学报, 2020, 40(5): 534-541.

[27] 叶 琦,宋炎峰,李 蒙,等. 迎春樱桃叶绿体基因组特征及其密码子使用偏好性分析[J]. 分子植物育种, 2022, 20(14): 4576-4585.

[28] 胡晓艳,许艳秋,韩有志,等. 酸枣叶绿体基因组密码子使用偏好性分析[J]. 森林与环境学报, 2019, 39(6): 621-628.

[29] 喻 凤,韩 明. 紫花苜蓿叶绿体基因组密码子偏好性分析[J]. 广西植物, 2021, 41(12): 2069-2076.

[30] 唐玉娟,赵 英,黄国弟,等. 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.

[31] 杨祥燕,蔡元保,谭秦亮,等. 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.

[32] 耿晓姗,贾 魏,陈佳宁,等. 金花茶叶绿体基因组密码子偏好性分析[J]. 分子植物育种, 2022, 20(7): 2196-2203.