

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.
doi: 10.3969/j.issn.1000-4440.2021.06.005

苜蓿质膜内在蛋白编码基因 *MsPIPs* 家族的密码子偏好性分析

张海霞, 王玉道, 许雪妮

(甘肃省民勤县林业技术推广站, 甘肃 民勤 733300)

摘要: 利用 CodonW 软件分析苜蓿质膜内在蛋白编码基因 (*MsPIPs*) 的密码子组成, 发现它们的密码子在碱基使用中存在偏倚, 偏好使用以 T/A 结尾的密码子, 其中 T 结尾的密码子出现频率为 47%。*MsPIPs* 密码子的平均有效密码子数 (*ENC*) 为 45.47 个, 偏好性较弱。通过主成分分析 (PCA)、对应性分析 (COA)、偏倚分析 (PR2)、有效密码子数分析 (*ENC* plot) 和中性分析等多重分析方法对苜蓿 *MsPIPs* 密码子偏好性的成因进行分析, 发现苜蓿 *MsPIPs* 的偏好性主要受自然选择的影响。

关键词: 质膜内在蛋白; 密码子偏好性; 自然选择; 突变

中图分类号: S551⁺.7 **文献标识码:** A **文章编号:** 1000-4440(2021)06-1393-07

Codon usage bias analysis of *MsPIPs* gene family which encode plasma membrane intrinsic proteins in alfalfa

ZHANG Hai-xia, WANG Yu-dao, XU Xue-ni

(Forestry Technology Popularization Station of Minqin County in Gansu Province, Minqin 733300, China)

Abstract: By using software CodonW to analyze the codon composition of plasma intrinsic protein encoded by *MsPIPs* gene in alfalfa, it was found that the codons showed bias in nucleobase use, the codons of *MsPIPs* preferred T/A endings, and the frequency of occurrence of T-ending codons was 47%. The average effective number of codons (*ENC*) of *MsPIPs* codons was 45.47, which had a weak codon bias. Through multivariate analytical methods combined by principal component analysis (PCA), corresponding analysis (COA), parity rule 2 plot mapping analysis (PR2), *ENC* plot analysis and neutrality plot analysis, the reason of the codon bias of the *MsPIPs* gene in alfalfa was analyzed, and it was found that the bias of *MsPIPs* in alfalfa might be driven by natural selection.

Key words: plasma membrane intrinsic proteins; codon usage bias; natural selection; mutation

密码子在遗传信息传递过程中起着至关重要的作用^[1]。编码不同氨基酸的密码子的准确识别是保证基因信息准确表达的关键^[2]。除了甲硫氨酸 (Met) 和色氨酸 (Trp) 由唯一的密码子编码外, 大部分氨基酸均由 2~6 个同义密码子编码, 同义密码子在不同植物体内的使用频率有差异, 形成了密码子使用偏性^[3]。密码子使用偏性受群体中突变压力、选择压

力和基因漂变的影响, 对研究基因功能、物种进化等具有重要科学意义^[4]。

质膜内在蛋白 (Plasma membrane intrinsic proteins, PIPs) 定位在质膜上, 是水通道蛋白 (Aquaporins, AQP) 的亚家族成员, 属于主要内在膜蛋白 (Major intrinsic protein, MIP) 家族^[5]。目前, 研究者们已经从拟南芥、玉米和番茄等中分别鉴定到 13 个、13 个和 18 个 PIPs 基因^[6]。在植物生长发育过程中, PIPs 主要负责植物体内水分的吸收和运输, 也参与植物体内营养及元素的运输及细胞生长分化过程^[7]。在干旱和盐等非生物胁迫下, 通过活性和表达情况的改变, PIPs 在植物对非生物胁迫的应答

收稿日期: 2021-03-29

作者简介: 张海霞 (1975-), 女, 甘肃民勤人, 本科, 高级工程师, 研究方向为林草有害生物防治。(E-mail) 1098149863@qq.com

通讯作者: 王玉道, (E-mail) 1602239412@qq.com

过程中发挥着重要作用^[8]。对 *PIPs* 基因家族密码子使用偏性的分析可为该家族编码蛋白结构和功能研究提供基础数据。

2020年,苜蓿基因组全序列的公布^[9]使得利用生物信息学手段在全基因组水平上研究苜蓿成为可能,但是苜蓿 *PIPs* 基因家族尚未见报道,且 *PIPs* 的密码子使用偏性规律也不清楚。本研究以从苜蓿基因组中鉴定到的 *PIPs* 基因家族为基础,进行密码子偏好性分析,明确该基因家族的密码子使用规律,为进一步展开功能研究和苜蓿抗性改良提供理论依据。

1 材料和方法

1.1 数据来源

在苜蓿基因组的 Swissprot 数据库^[9]中搜索包含“*PIP*”和编码序列(CDS)长度大于300 bp的序列,并在NCBI网站上进行这些序列的BLAST搜索,进一步鉴定 *PIP* 序列。筛选出的26个 *MsPIPs* 基因序列按照 *MsPIP1*~*MsPIP26* 进行命名(表1)。

1.2 苜蓿 *PIPs* 基因家族密码子组成分析

采用 CondonW 1.3 软件(<https://sourceforge.net/projects/codonw/>)分析计算苜蓿 *PIPs* 基因家族基因的密码子组成,包括同义密码子相对使用度(Relative synonymous codon usage value, RSCU)值、有效密码子数(Effective number of codons, ENC)、G+C含量、第3位碱基组成(A_{3S} 、 T_{3S} 、 G_{3S} 和 C_{3S})等指标^[10]。第1位碱基的G+C含量(GC_1)、第2位碱基的G+C含量(GC_2)、第3位碱基的G+C含量(GC_3)及第1和2位碱基的G+C含量(GC_{12})通过EMBOSS(http://imed.med.ucm.es/cgi-bin/emboss.pl?_action=input&_app=cusp)的Cusp功能计算得到^[11]。

1.3 最优密码子的确定

以ENC值为标准对上述筛选出的 *PIPs* 序列进行排序,从排序结果的两端各选取5%的基因建立高低偏性库,计算高低偏性库的 $\Delta RSCU$ 值,计算公式为: $\Delta RSCU = RSCU_{高偏性} - RSCU_{低偏性}$ 。具有最高RSCU值的密码子被称为高频密码子,具有最大 $\Delta RSCU$ 值的密码子被称为高表达密码子,同时满足最高RSCU值和最大 $\Delta RSCU$ 值的密码子被确定为最优密码子^[12]。

1.4 苜蓿 *MsPIPs* 基因家族密码子的偏好性分析

按照Wei等^[13]的方法进行主成分分析(Principal component analysis, PCA),将密码子的使用频率分布在

一个59维的向量空间中。利用对应性分析(Corresponding analysis, COA),比较2个或更多类别的变量数据,为密码子使用趋势和基因的主要变化研究提供直观结果。利用偏倚分析(Parity rule 2 plot mapping analysis, PR2)分析各个密码子第3位碱基组成(A_{3S} 、 T_{3S} 、 G_{3S} 和 C_{3S})的情况。利用ENC-plot分析展示ENC和 GC_3 的函数关系,使用中性分析(Neutrality plot mapping analysis)展示 GC_{12} 和 GC_3 的函数关系。

1.5 数据统计分析

使用Excel 2016制图。

2 结果与分析

2.1 苜蓿 *MsPIPs* 家族基因 CDS 序列的密码子组成

从苜蓿基因组中筛选到26条 *MsPIPs* 基因,总GC含量为38%~50%,平均值为45%(表1)。*MsPIPs* 的密码子第3位碱基上A、T、C、G出现的平均频率分别为28%、47%、26%和17%,表明苜蓿 *MsPIPs* 偏好使用以T/A结尾的密码子。*MsPIPs* 的ENC为38.87~49.54个,平均值为45.47个,表明苜蓿 *MsPIPs* 对密码子使用的偏好性较弱。

2.2 苜蓿 *MsPIPs* 的最优密码子

将苜蓿 *MsPIPs* 各成员氨基酸对应密码子的RSCU值整理求平均值后,得到 *MsPIPs* 家族密码子的RSCU分析结果。各氨基酸中具有最高RSCU值的20个密码子被确定为高频密码子(表2)。27个密码子被确定为高表达密码子($\Delta RSCU > 0.08$)(表3)。10个密码子被确定为最优密码子,分别是UUU、GUU、CCA、ACU、GCU、UAC、CAA、AAA、UGU和GGU。

2.3 苜蓿 *MsPIPs* 密码子的偏好性

基于RSCU值的PCA结果显示,26个 *MsPIPs* 的密码子使用频率分布在25个轴上,向量轴1~4分别解释了20.65%、16.18%、14.57%和13.81%的轴贡献,累计贡献率为65.21%,表明向量轴1~4对密码子偏好性有重要影响(图1)。以第1和第2向量轴将 *MsPIPs* 分布在图上,可以看出大部分成员分布在向量轴1的右半轴上(图2)。进一步通过COA分别分析了G+C含量(GC)、ENC值、 GC_{3S} 和向量轴1~4的相关性。结果表明, GC_{3S} 与向量轴1和轴3呈极显著负相关,ENC值与向量轴2和轴4呈显著或极显著正相关,G+C含量与向量轴3也呈显著负相关(表4)。

表 1 苜蓿 *MsPIP*s 基因序列密码子组成及使用参数Table 1 Codon composition and usage parameters of *MsPIP*s sequences in alfalfa

基因	编号	$T_{3S}(\%)$	$C_{3S}(\%)$	$A_{3S}(\%)$	$G_{3S}(\%)$	$GC_{3S}(\%)$	$GC(\%)$	$ENC(\text{个})$
<i>MsPIP1</i>	MS.gene08634.t1	44	29	30	16	37	46	49.54
<i>MsPIP2</i>	MS.gene08872.t1	45	29	28	16	37	46	49.36
<i>MsPIP3</i>	MS.gene20063.t1	41	32	29	16	40	47	47.86
<i>MsPIP4</i>	MS.gene22434.t1	41	32	29	16	40	47	47.86
<i>MsPIP5</i>	MS.gene31293.t1	47	26	28	14	34	44	46.81
<i>MsPIP6</i>	MS.gene33486.t1	52	17	34	17	27	42	47.64
<i>MsPIP7</i>	MS.gene34032.t1	56	21	24	13	29	44	43.02
<i>MsPIP8</i>	MS.gene54792.t1	56	21	24	13	29	44	43.02
<i>MsPIP9</i>	MS.gene56672.t1	47	22	30	20	34	45	42.42
<i>MsPIP10</i>	MS.gene69248.t1	48	26	24	20	38	45	38.87
<i>MsPIP11</i>	MS.gene80664.t1	49	25	25	16	35	46	45.06
<i>MsPIP12</i>	MS.gene80665.t1	48	24	25	19	35	46	41.19
<i>MsPIP13</i>	MS.gene007432.t1	41	32	28	16	41	48	47.90
<i>MsPIP14</i>	MS.gene007515.t1	41	32	29	16	40	47	47.90
<i>MsPIP15</i>	MS.gene007660.t1	44	29	31	15	37	46	47.50
<i>MsPIP16</i>	MS.gene014206.t1	37	40	22	16	49	50	46.84
<i>MsPIP17</i>	MS.gene024637.t1	48	18	28	26	35	45	46.08
<i>MsPIP18</i>	MS.gene028722.t1	48	25	30	15	33	44	45.40
<i>MsPIP19</i>	MS.gene030381.t1	52	21	24	19	33	44	48.16
<i>MsPIP20</i>	MS.gene031996.t1	47	27	25	20	38	45	39.05
<i>MsPIP21</i>	MS.gene037437.t1	51	16	36	17	26	42	44.78
<i>MsPIP22</i>	MS.gene048817.t1	46	26	27	22	38	45	47.54
<i>MsPIP23</i>	MS.gene050133.t1	45	29	29	16	37	46	48.22
<i>MsPIP24</i>	MS.gene057322.t1	46	27	24	21	39	46	39.18
<i>MsPIP25</i>	MS.gene069771.t1	54	20	36	09	24	38	45.70
<i>MsPIP26</i>	MS.gene074338.t1	48	25	31	14	32	44	45.20

T_{3S} : 第 3 位碱基上 T 出现的频率; C_{3S} : 第 3 位碱基上 G 出现的频率; A_{3S} : 第 3 位碱基上 A 出现的频率; G_{3S} : 第 3 位碱基上 G 出现的频率; GC_{3S} : 第 3 位碱基上 G+C 出现的频率; GC : G+C 含量; ENC : 有效密码子数。

表 2 苜蓿 *MsPIP*s 同义密码子相对使用度 ($RSCU$) 分析Table 2 Analysis on relative synonymous codon usage value ($RSCU$) of the synonymous codons from *MsPIP*s in alfalfa

氨基酸	密码子	数量(个)	$RSCU$ 值	氨基酸	密码子	数量(个)	$RSCU$ 值
Phe	UUU	241	1.01	Tyr	UAU	64	0.46
	UUC	235	0.99		UAC	214	1.54
Leu	UUA	65	0.66	TER	UAA	16	1.85
	UUG	179	1.95	UAG	1	0.12	
	CUU	183	2.02	His	CAU	75	0.98
	CUC	59	0.58		CAC	77	1.02
	CUA	41	0.45	Gln	CAA	108	1.44
CUG	34	0.35	CAG		44	0.56	
Ile	AUU	315	1.66	Asn	AAU	80	0.67
	AUC	179	0.94		AAC	144	1.33
	AUA	76	0.40	Lys	AAA	156	1.10
Met	AUG	116	1.00		AAG	119	0.90

续表2 Continued2

氨基酸	密码子	数量(个)	RSCU 值	氨基酸	密码子	数量(个)	RSCU 值
Val	GUU	316	2.15	Asp	GAU	158	1.16
	GUC	82	0.57		GAC	111	0.84
	GUA	48	0.33	Glu	GAA	83	1.03
	GUG	133	0.95		GAG	104	0.97
Ser	UCU	117	1.92	Cys	UGU	79	1.61
	UCC	52	0.83		UGC	20	0.39
	UCA	86	1.44	TER	UGA	9	1.04
	UCG	1	0.02	Trp	UGG	119	1.00
	AGU	68	1.10	Arg	CGU	18	0.56
	AGC	43	0.70		CGC	5	0.11
	Pro	CCU	111		1.30	CGA	5
CCC		32	0.39	CGG	5	0.13	
CCA		186	2.10	AGA	129	3.17	
CCG		19	0.22	AGG	77	1.93	
Thr	ACU	201	1.98	Gly	GGU	409	2.04
	ACC	122	1.20		GGC	93	0.45
	ACA	76	0.76		GGA	243	1.21
	ACG	6	0.06		GGG	61	0.30
Ala	GCU	473	2.28				
	GCC	148	0.69				
	GCA	206	0.97				
	GCG	15	0.06				

TER:终止子;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;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;Pro:脯氨酸;Thr:苏氨酸;Ala:丙氨酸;Trp:色氨酸;Arg:精氨酸;Gly:甘氨酸。

表3 苜蓿 *MsPIPs* 高、低表达基因的密码子使用统计Table 3 Statistics of codons in high and low expression genes of *MsPIPs* in alfalfa

氨基酸	密码子	高表达基因		低表达基因		$\Delta RSCU$	氨基酸	密码子	高表达基因		低表达基因		$\Delta RSCU$
		出现次数(次)	RSCU 值	出现次数(次)	RSCU 值				出现次数(次)	RSCU 值	出现次数(次)	RSCU 值	
Phe	UUU*	6	1.00	7	0.82	0.18	Ser	UCU	6	1.71	6	2.25	-0.54
	UUC	6	1.00	10	1.18	-0.18		UCC	2	0.57	4	1.50	-0.93
Leu	UUA*	5	1.25	2	0.46	0.79	UCA*	4	1.14	1	0.38	0.76	
	UUG	6	1.50	7	1.62	-0.12	UCG	0	0	0	0	0	
	CUU	5	1.25	6	1.38	-0.13	AGU*	6	1.71	1	0.38	1.33	
	CUC	2	0.50	9	2.08	-1.58	AGC	3	0.86	4	1.50	-0.64	
Ile	CUA*	6	1.50	2	0.46	1.04	Pro	CCU*	6	1.85	7	1.65	0.20
	CUG	0	0	0	0	0		CCC	1	0.31	3	0.71	-0.40
	AUU*	6	1.38	6	0.90	0.48		CCA*	6	1.85	6	1.41	0.44
	AUC	5	1.15	10	1.50	-0.35	CCG	0	0	1	0.24	-0.24	
Met	AUA	2	0.46	4	0.60	-0.14	Thr	ACU*	9	2.25	6	1.85	0.40
	AUG	3	1.00	5	1.00	0		ACC	5	1.25	6	1.85	-0.60
Val	GUU*	9	2.25	9	1.44	0.81		ACA*	2	0.50	1	0.31	0.19
	GUC	2	0.50	8	1.28	-0.78	ACG	0	0	0	0	0	
	GUA*	4	1.00	3	0.48	0.52	Ala	GCU*	8	2.91	19	2.17	0.74

续表3 Continued3

氨基酸	密码子	高表达基因		低表达基因		$\Delta RSCU$	氨基酸	密码子	高表达基因		低表达基因		$\Delta RSCU$
		出现次数(次)	RSCU值	出现次数(次)	RSCU值				出现次数(次)	RSCU值			
Tyr	GUG	1	0.25	5	0.80	-0.55	Cys	GCC	0	0	4	0.46	-0.46
	UAU*	5	1.67	0	0	1.67		GCA*	3	1.09	7	0.80	0.29
	UAC	1	0.33	10	2.00	-1.67		GCG	0	0	5	0.57	-0.57
TER	UAA	0	0	1	3.00	-3.00	UGU*	5	2.00	3	1.50	0.50	
	UAG*	1	3.00	0	0	3.00	UGC	0	0	1	0.50	-0.50	
His	CAU*	3	2.00	4	0.73	1.27	UGA	0	0	0	0	0	
	CAC	0	0	7	1.27	-1.27	Trp	UGG	1	1.00	5	1	0
Gln	CAA*	5	2.00	2	1.33	0.67	Arg	CGU*	1	3.00	1	0.75	2.25
	CAG	0	0	1	0.67	-0.67	CGC	0	0	0	0	0	
Asn	AAU*	8	1.00	1	0.25	0.75	CGA	0	0	0	0	0	
	AAC	8	1.00	7	1.75	-0.75	CGG	0	0	1	0.75	-0.75	
Lys	AAA*	7	1.40	7	1.17	0.23	AGA	0	0	5	3.75	-3.75	
	AAG	3	0.60	5	0.83	-0.23	AGG*	1	3	1	0.75	2.25	
Asp	GAU*	9	1.64	1	0.25	1.39	Gly	GGU*	12	2.09	14	1.93	0.16
	GAC	2	0.36	7	1.75	-1.39	GGC	1	0.17	10	1.38	-1.21	
Glu	GAA	4	1.14	4	1.33	-0.19	GGA*	10	1.74	3	0.41	1.33	
	GAG*	3	0.86	2	0.67	0.19	GGG	0	0	2	0.28	-0.28	

各氨基酸符号对应的中文名称见表2注。标注*的为高表达密码子。RSCU:同义密码子相对使用度。

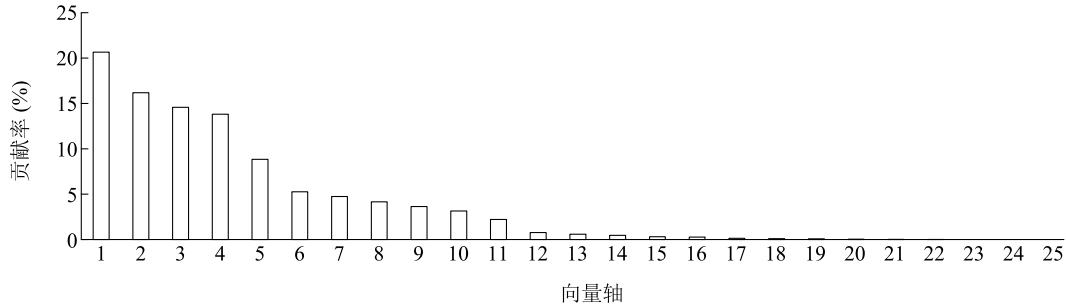


图1 苜蓿 *MsPIPs* 密码子使用频率的主成分分析

Fig.1 Principal component analysis of codon usage frequency in *MsPIPs* of alfalfa

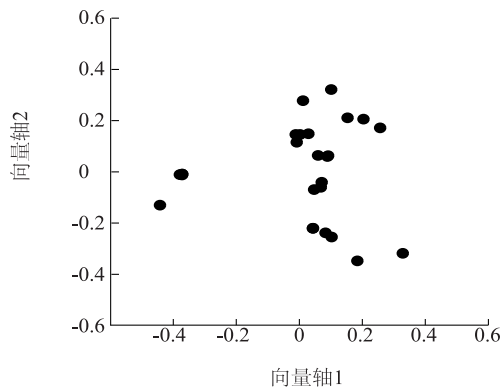


图2 苜蓿 *MsPIPs* 密码子基于 RSCU 值的对应性分析

Fig.2 Corresponding analysis based on the RSCU values of *MsPIPs* codons in alfalfa

表4 影响苜蓿 *MsPIPs* 密码子偏好性重要指标的相关性分析

Table 4 Correlation analysis of the indices influencing codon bias of *MsPIPs* in alfalfa

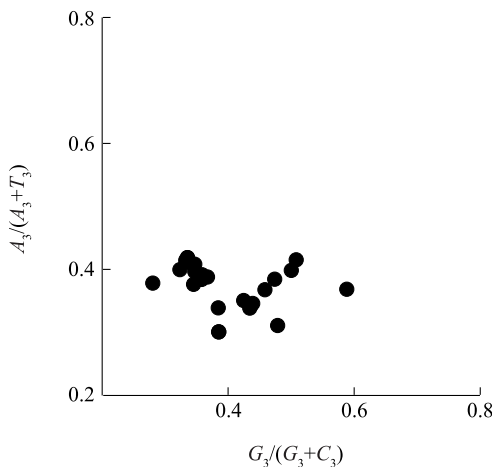
指标	GC	ENC	GC _{3S}	向量轴1	向量轴2	向量轴3	向量轴4
GC	1.000						
ENC	0.149	1.000					
GC _{3S}	0.945**	0.113	1.000				
向量轴1	-0.832	-0.334	-0.798**	1.000			
向量轴2	0.022	0.515**	0.061	0	1.000		
向量轴3	-0.495*	0.305	-0.517**	0	0	1.000	
向量轴4	-0.104	0.443*	-0.061	0	0	0	1.000

GC:G+C含量;ENC:有效密码子数;GC_{3S}:第3位碱基上G、C的组成;* :显著相关(P<0.05);** :极显著相关(P<0.01)。

通过 PR2 分析组成 *MsPIP*s 氨基酸的第 3 位密码子中 A/G 和 T/C 的关系(图 3)。除 *MsPIP17* 和 *MsPIP21* 分布在第 4 象限, *MsPIP6* 分布在中轴线上, 其余 *MsPIP*s 均分布在第 3 象限, 表明 *MsPIP*s 的密码子组成存在偏倚, 且大部分 *MsPIP*s 结尾 T 的使用频率高于 A, C 的使用频率高于 G。

如果基因密码子的组成完全受突变的影响, 则 *ENC* 的分析图中, 所有基因应该落在标准曲线上。但是从 *ENC* 的分析结果可见, *MsPIP*s 的所有基因成员均分布在预期 *ENC* 值的下方(图 4), 表明 *MsPIP*s 的密码子偏好受其他因素影响。此外, 由表 4 可见, *ENC* 和 GC_{3s} 之间并无显著相关性 ($r = 0.113$), 表明第 3 位碱基上的 G/C 含量对密码子的使用模式影响较小。

中性分析图将 GC_{12} 和 GC_3 的关系可视化, 并可进一步帮助评估密码子组成受突变和选择压力影响的程度。由图 5 可见, *MsPIP*s 的所有基因成员均分布在对角线上方。



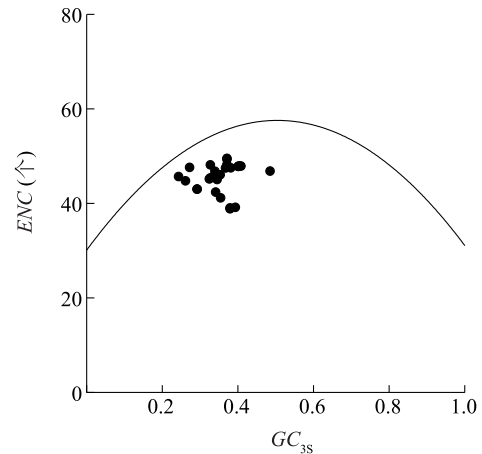
$A_3/(A_3+T_3)$: 第 3 位碱基上 A 的组成占第 3 位碱基上 A、T 组成的百分比; $G_3/(G_3+C_3)$: 第 3 位碱基上 G 的组成占第 3 位碱基上 G、C 组成的百分比。

图 3 利用 PR2-plot 绘图分析各个密码子第 3 位碱基组成

Fig.3 Composition of the third base of each codon by using PR2-plot mapping analysis

3 讨论

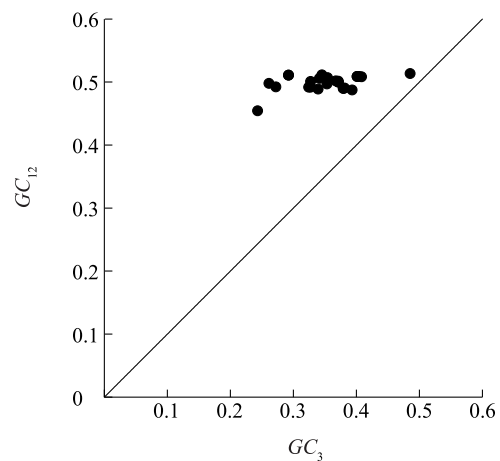
密码子偏好性是多因素共同作用的结果, 受自然选择、碱基组成、基因表达水平、蛋白质二级结构等因素的影响^[14-16], 其中自然选择和突变是主导密码子偏好性的 2 类主要因素。杨国锋等^[17]发现莨



ENC: 有效密码子数; GC_{3s} : 第 3 位碱基上 G、C 的组成。

图 4 利用 *ENC* 绘图分析 *ENC* 和 GC_3 的函数关系

Fig.4 Functional relation of effective number of codons (*ENC*) and GC_3 by using *ENC* mapping analysis



GC_{12} : 第 1 和 2 位碱基的 G+C 含量; GC_3 : 第 3 位碱基的 G+C 含量。

图 5 利用中性绘图分析 GC_{12} 和 GC_3 的函数关系

Fig.5 Functional relation of GC_{12} and GC_3 by using neutrality plot analysis

藜苜蓿叶绿体基因组的密码子偏好性主要受到突变的影响。屈蒙蒙等^[18]对香蕉质膜内在蛋白 (PIPs) 的密码子偏好性分析结果表明, 该家族基因受自然选择的影响较大。可见, 不同物种、不同基因类型的同义密码子偏好性不同。

本研究结果表明, 苜蓿 *MsPIP*s 家族基因的密码子多以 T/A 结尾。在得到的 10 个最优密码子中, 有 9 个也是以 T/A 结尾。通过 PR2 分析, 进一步证实密码子的第 3 位碱基组成存在偏倚。本研究结果与其他学者在翠雀^[1]、沙枣^[12]、蒺藜苜蓿^[17]等植物中的研究结果一致。*ENC* 值直接反映密码子偏好

性,以 35 为界,ENC 值 ≤ 35 表示密码子偏好性强,ENC 值 > 35 则表示密码子偏好性弱^[19]。苜蓿 *MsPIP*s 密码子的平均 ENC 值为 45.47,表明 *MsPIP*s 存在密码子偏好性,只是程度较弱。

通过 PCA、COA、ENC-plot 和中性分析,进一步分析造成 *MsPIP*s 密码子偏好性的主要原因。从 PCA 和 COA 的结果可见,*MsPIP*s 的成员主要分布在轴 1 的右半轴上,而轴 1 与 GC_{3s} 呈显著负相关关系,这表明第 3 位碱基组成(主要是 A/T 结尾)对 *MsPIP*s 的密码子偏好性有影响。轴 1 仅解释 20.65%的密码子使用频率差异,可见除碱基组成外,还有其他因素对 *MsPIP*s 的密码子偏好性产生影响。ENC-plot 分析能通过基因和标准曲线的分布关系直接反映自然选择和突变因素中的主导因素,如果基因落在标准曲线上,则表明基因的密码子偏好性完全受突变的影响,反之则主要受自然选择的影响^[20]。本研究结果表明,*MsPIP*s 所有成员均落在标准曲线下方,表明它们可能主要受到自然选择的影响。这一推测也被中性分析进一步证实。中性分析能有效地比较自然选择和突变对密码子偏好性的影响程度,当 GC_{12} 和 GC_3 数值相同时,基因会分布在回归曲线上,且该基因不同位置的碱基组成无差异,密码子使用更多地受突变影响,反之则主要受自然选择的影响^[21]。*MsPIP*s 的中性分析中,所有基因均分布在回归曲线的上方,表明所有成员主要受到自然选择的影响。

4 结论

苜蓿质膜水通道蛋白家族编码基因 (*MsPIP*s) 的密码子偏好性较弱,碱基使用存在偏倚,偏好使用以 A/T 结尾的密码子,且偏好性主要受自然选择的影响。本研究结果可为通过密码子改造 *MsPIP*s 提供理论依据。

参考文献:

- [1] DUAN H R, ZHANG Q, WANG C M, et al. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints [J]. PeerJ, 2021, 9(6): e10787.
- [2] SAU K, GUPTA S K, SAU S, et al. Factors influencing synonymous codon and amino acid usage biases in Mimivirus [J]. Biosystems, 2006, 85: 107-113.
- [3] GUAN D L, MA L B, KHAN M S, et al. Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints [J]. BMC Genomics, 2018, 19: 542.
- [4] BULMER M. The selection mutation drift theory of synonymous codon usage [J]. Genetics, 1991, 129: 897-907.
- [5] AYADI M, BRINI F, MASMOUDI K. Overexpression of a wheat aquaporin gene, TdPIP2;1, enhances salt and drought tolerance in transgenic durum wheat cv. Maali [J]. International Journal of Molecular Sciences, 2019, 20(10): 2389.
- [6] 孙天旭, 李玉花, 张旸. 逆境条件下水孔蛋白 PIPs 作用的研究进展[J]. 植物生理学报, 2014, 50(6): 749-757.
- [7] 刘海莉. 苹果质膜水通道蛋白基因 PIP2;1 的耐旱及耐盐性功能分析[D]. 杨凌:西北农林科技大学, 2019.
- [8] 王星, 张纪龙, 冯秀秀, 等. 植物质膜水通道蛋白转运及逆境胁迫响应的分子调控机制[J]. 遗传, 2017, 39(4): 293-301.
- [9] CHEN H T, ZENG Y, YANG Y Z, et al. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa [J]. Nature Communications, 2020, 11: 2494.
- [10] ZHANG Y Y, SHI E, YANG Z P, et al. Development and application of genomic resources in an endangered palaeoendemic tree, *Parrotia subaequalis* (Hamamelidaceae) from eastern China [J]. Frontiers in Plant Science, 2018, 9: 246.
- [11] CUI G X, WANG C M, WEI X X, et al. Analysis of synonymous codon usage of the complete chloroplast genome in *Phleum pratense* cv. Minshan [J]. International Journal of Agriculture & Biology, 2020, 24(2): 352-358.
- [12] 王婧, 王天翼, 王罗云, 等. 沙枣叶绿体全基因组序列及其使用密码子偏性分析[J]. 西北植物学报, 2019, 39(9): 1559-1572.
- [13] 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: 262.
- [14] 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.
- [15] 张玉波, 周正湘, 吴小玉, 等. 基于转录组的大头金蝇密码子的偏好性分析[J]. 江苏农业科学, 2019, 47(11): 78-81.
- [16] 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.
- [17] 杨国锋, 苏昆龙, 赵怡然, 等. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. 草业学报, 2015, 24(12): 171-179.
- [18] 屈蒙蒙, 孙雪丽, 郝向阳, 等. 香蕉水通道蛋白基因家族的密码子偏好性分析[J]. 园艺与种苗, 2018(5): 4-10.
- [19] SONG H, LIU J, CHEN T, et al. Synonymous codon usage pattern in model legume *Medicago truncatula* [J]. Journal of Integrative Agriculture, 2018, 17: 2074-2081.
- [20] WRIGHT F. The effective number of codons used in a gene [J]. Gene, 1990, 87: 23-29.
- [21] ZHANG D S, HU P, LIU T G, et al. GC bias lead to increased small amino acids and random coils of proteins in coldwater fishes [J]. BMC Genomics, 2018, 19: 315.

(责任编辑:张震林)