

· 研究报告 ·

籽粒苋丙酮酸磷酸二激酶(PPDK)基因的密码子偏好性

聂江婷¹, 白云凤^{2,3*}, 贺飞燕¹, 闫建俊², 张维锋²

¹山西大学生物工程学院, 太原 030006; ²山西省农业科学院作物科学研究所, 太原 030031

³山西省农业科学院农作物品种资源研究所, 农业部黄土高原作物基因资源与种质创制重点实验室, 太原 030031

摘要 运用CHIPS、CUSP和CodonW等程序分析了双子叶C₄植物籽粒苋(*Amaranthus hypochondriacus*)丙酮酸磷酸二激酶(PPDK)基因的密码子偏好性, 并与马铃薯(*Solanum tuberosum*)和苜蓿(*Medicago truncatula*)等双子叶植物及水稻(*Oryza sativa*)和玉米(*Zea mays*)等单子叶植物进行了比较, 建立了聚类树状图, 以期在作物高光效基因工程中为籽粒苋PPDK基因选择合适的受体植物提供依据。研究结果表明, 籽粒苋PPDK基因偏好于以A或T结尾的密码子, 与其它几种被比较的双子叶作物的PPDK基因密码子偏好性趋势一致, 而玉米和水稻等单子叶植物更偏好使用以G或C结尾的密码子。PPDK基因密码子使用偏好的系统聚类分析表明, 籽粒苋与马铃薯和苜蓿等双子叶植物聚为一类, 而稗草(*Echinochloa crusgalli*)、玉米和高粱(*Sorghum bicolor*)等单子叶植物聚为一类, 与系统进化地位一致。但单子叶植物水稻的密码子偏好性与籽粒苋较为接近, 与玉米和高粱相差较远。为了选择合适的蛋白质表达系统, 比较并分析了籽粒苋PPDK基因的密码子偏好性与大肠杆菌(*Escherichia coli*)及酵母菌的异同, 发现其与酵母菌的差异小于大肠杆菌, 表明选择酵母菌表达系统更为合适。

关键词 籽粒苋, 密码子偏好性, PPDK基因

聂江婷, 白云凤, 贺飞燕, 闫建俊, 张维锋 (2014). 籽粒苋丙酮酸磷酸二激酶(PPDK)基因的密码子偏好性. 植物学报 49, 672–681.

丙酮酸磷酸二激酶(pyruvate orthophosphate di-kinase, PPDK)存在于各类生物体(动物除外)中。细菌和原生动物中的PPDK起丙酮酸激酶的作用, 催化磷酸烯醇式丙酮酸(phosphoenolpyruvate, PEP)生成ATP和丙酮酸(Pocalyko et al., 1990)。植物体中的PPDK有两种类型, 即细胞质型和C₄叶绿体型, 二者有很高的同源性, 但细胞质型PPDK与叶绿体型PPDK相比在N端缺少一段叶绿体转运肽(Matsuoka, 1995)。细胞质型PPDK活性较低, 不参与光合作用, 但可能与维持细胞的pH值及应对生物(如病毒感染)(Ryšlavá et al., 2003)或非生物胁迫(如干旱、低温、高盐和UV-B)有关(Moons et al., 1998; Yamakawa et al., 2007)。叶绿体型PPDK主要存在于C₄植物叶肉细胞的叶绿体中, 催化ATP和丙酮酸(pyruvate)生成PEP(李卫华等, 1999)。PEP作为CO₂的初级受体, 在位于叶肉细胞质中的磷酸烯醇式丙酮酸羧化酶(phosphoenolpyruvate carboxylase, PEPC)的

催化下, 生成四碳酸草酰乙酸(OAA)。OAA被NADP-苹果酸脱氢酶(NADP-MDH)还原为苹果酸或通过天冬氨酸转氨酶的转氨作用转变成天冬氨酸。苹果酸或天冬氨酸被转运到维管束鞘细胞(Bsc), 在核酮糖-1,5-二磷酸羧化酶/加氧酶附近脱羧释放出CO₂, 同时产生还原力和丙酮酸。丙酮酸转回到叶肉细胞的叶绿体中, 在PPDK的催化下生成PEP, 完成1个C₄循环。释放的CO₂提高了Rubisco周围的CO₂浓度, 抑制了Rubisco的加氧活性, 增强了羧化性能, 使C₄植物比C₃植物具更高的光合效率。迄今已有诸多将C₄光合型PPDK基因转入C₃植物的报道(魏松涛等, 2008; 杜西河等, 2013), 但所转基因大多来自单子叶植物玉米(*Zea mays*)或高粱(*Sorghum bicolor*), 其它来源的很少。籽粒苋(*Amaranthus hypochondriacus*)是双子叶C₄植物, 生长快且光合效率高。本实验室前期克隆了籽粒苋C₄光合型PPDK基因(GenBank登录号为: JF907701)。进一步探明该基因的特征特性将为其用

收稿日期: 2014-04-24; 接受日期: 2014-08-04

基金项目: 国家自然科学基金(No.30971838)、山西省科技攻关项目(No.201303110015-1)和山西省回国留学人员基金(No.2013-146)

* 通讯作者。E-mail: byfok@126.com

于植物高光效基因工程提供依据。

生物体中, 蛋白质的氨基酸组成和排列顺序取决于编码基因的核苷酸序列, 即每3个连续的核苷酸序列(三联体密码子)对应1个特定的氨基酸。三联体密码子有64个, 而氨基酸仅有20种, 使得每种氨基酸可由1~6个密码子编码。编码同一氨基酸的密码子称为同义密码子(synonymous codon)。在长期的进化过程中, 不同物种的基因在密码子使用上存在着明显的偏好性。在转基因研究中, 如果外源基因存在较多的宿主表达系统的稀有密码子, 其高效表达将会受到影响(Li and Xue, 2005)。分析密码子偏好性, 并与不同受体表达系统进行比较, 可明确合适的表达系统, 或明确需要改造的密码子(吴宪明等, 2007; 李平等, 2011; 郭秀丽等, 2012)。本研究采用几种较为常用的密码子分析方法, 对粟粒苋*PPDK*基因的密码子偏好性进行了分析, 并与马铃薯(*Solanum tuberosum*)、苜蓿(*Medicago truncatula*)和蓖麻(*Ricinus communis*)等双子叶植物及水稻(*Oryza sativa*)、玉米和高粱等单子叶植物进行比较, 同时基于*PPDK*基因的密码子偏好性进行了聚类分析, 以期为粟粒苋*PPDK*基因在作物高光效基因工程中选择合适的受体植物提供依据。此外, 比较并分析了粟粒苋*PPDK*基因的密码子偏好性与大肠杆菌及酵母菌的异同, 以期为选择合适的蛋白质表达系统奠定基础。

1 材料与方法

1.1 序列来源和分析软件

粟粒苋(*Amaranthus hypochondriacus* L.)*PPDK*基因由本实验室自主克隆(GenBank登录号: JF907701), ORF长度为2 871 bp, 编码957个氨基酸。蓖麻(*Ricinus communis* L.)、马铃薯(*Solanum tuberosum* L.)、水稻(*Oryza sativa* L.)和玉米(*Zea mays* L.)等植物的*PPDK*基因序列来源于GenBank, 序列登录号见表1。利用在线数据库Codon Usage Database(<http://www.kazusa.or.jp/codon/>)查询大肠杆菌(*Escherichia coli*)和酵母菌基因组的密码子偏好性。用Mobyle@Pasteur(<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms>)在线程序中的EMBOSS6.3.1: chips(<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::chips>)、EMBOSS6.3.1:cusp(<http://mobyle.past>

eur.fr/cgi-bin/portal.py?#forms::cusp)以及codon W1.4.4(<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::CodonW>)分析*PPDK*基因的密码子偏好性。

表1 *PPDK*基因的完整编码区序列来源

Table 1 Sources of coding sequence of *PPDK* genes

Accession number	Species
JF907701	<i>Amaranthus hypochondriacus</i>
NM_001125515	<i>Arabidopsis thaliana</i>
XM_003609490	<i>Medicago truncatula</i>
X78347	<i>Mesembryanthemum crystallinum</i>
XM_002516781	<i>Ricinus communis</i>
XR_415309	<i>Solanum lycopersicum</i>
XM_006365216	<i>S. tuberosum</i>
DQ083757	<i>Echinochloa crusgalli</i>
AY268138	<i>Sorghum bicolor</i>
AJ004965	<i>Oryza sativa</i>
FJ517555	<i>Zea mays</i>

1.2 方法

用Mobyle@Pasteur(<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms>)网站的CHIPS在线程序分析*PPDK*基因的有效密码子数(effective number of codons, ENC)。用CUSP在线程序分析密码子的使用频率。用CodonW在线程序分析基因编码区的GC含量、三联体密码子中第3位碱基的GC含量(GC3s)以及相对同义密码子的使用度(relatively synonymous codon usage, RSCU)。用SPSS 19.0软件对不同植物基于*PPDK*基因的密码子使用的偏好性进行聚类。用DNAMAN软件进行多序列比对并建立各个物种*PPDK*基因的核苷酸序列同源树。最佳比对采用完全比对方式, 按照输入顺序输出。两序列比对采用快速比对进行运算, 参数为空位罚分(7)、顶部对角线数(4)、K-tuple(2)和窗口大小(4), 空位罚分是在序列对齐时插入空位并进行罚分以控制空位插入的合理性, 设置较小的K-tuple值可以提高精确度, 当序列较长时, 一般设置较大的K-tuple值(DNA序列: K-tuple值可选范围为2~6)。多重比对参数包括空位开放罚分(10)、空位延伸罚分(5)、延迟趋异序列%(40)和蛋白质加权矩阵(BLOSUM)。蛋白质空位参数使用亲水性罚分和残基特异罚分, 其参数为亲水性残基(GPSN-

DQEKR)和空位分隔距离(8)。

2 结果与讨论

2.1 有效密码子数和GC含量分析

基因有效密码子数(ENc)是某一个密码子使用频率与同义密码子平均使用频率偏差的量化值。如果某一特定基因中,各个同义密码子出现的频率相同,即对密码子没有偏好性,则ENc值为61;如果只使用同义密码子中的1个,即20种氨基酸只使用20种密码子,表明密码子偏好性最强,ENc值为20。用CHIPS和CUSP在线程序计算结果表明,籽粒苋PPDK基因的ENc值为53.518,表明该基因各密码子出现的频率有一定的差异,即具有密码子偏好性。PPDK基因的GC含量为0.435,表明该基因的GC含量低于AT;其GC3s(密码子第3位碱基中G+C的含量在第3位碱基总量中所占的比率)为0.387,比GC含量更低,说明该基因偏好使用以A和T结尾的密码子。

2.2 密码子偏好性分析

某一密码子出现的频率与编码该氨基酸的所有同义密码子出现频率的比值(Fraction)(所有同义密码子的比值总和为1)和该密码子在基因的总密码子中出现的频率(Frequency,即在1 000个密码子中出现的次数)可从不同角度反映密码子的偏好性。相对同义密码子使用度是衡量密码子偏好的另一重要指标,该指标指某一特定密码子在编码对应氨基酸的同义密码子中的相对概率,鉴于去除了氨基酸组成对密码子使用的影响,其结果更加直观可靠(Wang and Roossinck, 2006)。如果密码子的使用没有偏好性,则该密码子的RSCU=1;如果某一密码子使用频率高,则RSCU>1;如使用频率低,则小于1。CUSP和CodonW在线分析表明,在籽粒苋PPDK基因中,RSCU值大于1.2的密码子有16个,最高的为2.36(CCT),是该基因偏好的密码子。在这些偏好的密码子中,除了AGG外,其余15个均以A或T碱基结尾,并且这些密码子的Fraction和Frequency值也较高,验证了籽粒苋PPDK基因偏好以A或T结尾的密码子(表2)。

比较籽粒苋PPDK基因各个密码子在编码基因总密码子中出现的频率与Codon Usage Database籽粒

苋21个CDS(包含8 280个codon)中相同密码子出现的频率,发现二者之间有较高的一致性,相关系数高达0.900 4,二者的比值介于0.47~1.74之间,小于0.5的只有2个(CCC和TGG),大于1.5的有3个(GCA、CTA和CTC),表明籽粒苋PPDK基因的偏好性与其基因组总的偏好性较为一致。

2.3 不同植物PPDK基因的密码子偏好性比较

2.3.1 有效密码子数和GC含量的比较

籽粒苋和其它10种植物PPDK基因的ENc值及GC和GC3s含量计算结果见表3。Sharp和Li(1986)的研究表明,ENc值越小该基因的表达水平越高。本研究中,拟南芥(*Arabidopsis thaliana*)和水稻的ENc值大于55,玉米和苜蓿的ENc值小于50,籽粒苋及其它6种植物的ENc值介于50~55之间。GC和GC3s含量在单子叶和双子叶植物之间差异显著,籽粒苋、马铃薯和苜蓿等7种双子叶植物的CG和GC3s均低于0.5,而水稻、稗草(*Echinochloa crusgalli*)、玉米和高粱4种单子叶植物均高于0.5。将双子叶植物组和单子叶植物组进行t检验:双样本等方差假设得出, $t_{ENc}=0.046$, 小于 $t_{0.05}=2.262$, 表明双子叶与单子叶植物的ENc值无显著差异。 $t_{GC}=-6.733$, $t_{GC3s}=-6.957$, 两者的绝对值均大于 $t_{0.01}=3.250$, 差异达极显著水平,表明双子叶植物密码子比单子叶植物对A或T的偏好性更强。

2.3.2 基于密码子使用偏性的系统聚类

为进一步比较籽粒苋与其它植物PPDK基因密码子的使用特性,以每个基因作为1个对象,以59个密码子(剔除色氨酸(Trp)和甲硫氨酸(Met)以及3个终止密码子)的相对使用度为变量,计算不同植物PPDK基因之间的欧氏平方距离(Sharp et al., 1988; Sau et al., 2005)。对于基因a与基因b,其密码子使用偏性的欧式距离系数 D_{ab} 的计算公式为:

$$D_{ab} = \sqrt{\sum_{i=1}^{59} (RSCU_{ai} - RSCU_{bi})^2}$$

不同植物PPDK基因的欧式平方距离计算结果见表4。

对于PPDK基因类与类之间的距离,依据离差平方和(Ward)法计算,以使归为同类的PPDK基因之间的离差平方和最小,而不同类的PPDK基因之间的离

表2 粽粒苋PPDK基因的密码子偏好性**Table 2** Codon bias of *PPDK* gene

Codon	Amino acid	Fraction	Frequency (f‰)	Number	RSCU
GCA	A (Ala)	0.318	29.258	28	<u>1.27</u>
GCC	A	0.159	14.629	14	0.64
GCG	A	0.091	8.359	8	0.36
GCT	A	0.432	39.707	38	<u>1.73</u>
TGC	C (Cys)	0.571	8.359	8	1.14
TGT	C	0.429	6.270	6	0.86
GAC	D (Asp)	0.375	18.809	18	0.75
GAT	D	0.625	31.348	30	<u>1.25</u>
GAA	E (Glu)	0.605	48.067	46	<u>1.21</u>
GAG	E	0.395	31.348	30	0.79
TTC	F (Phe)	0.500	17.764	17	1.00
TTT	F	0.500	17.764	17	1.00
GGA	G (Gly)	0.393	34.483	33	<u>1.57</u>
GGC	G	0.143	12.539	12	0.57
GGG	G	0.179	15.674	15	0.71
GGT	G	0.286	25.078	24	1.14
CAC	H (His)	0.429	6.270	6	0.86
CAT	H	0.571	8.359	8	1.14
ATA	I (Ile)	0.204	10.449	10	0.61
ATC	I	0.224	11.494	11	0.67
ATT	I	0.571	29.258	28	<u>1.71</u>
AAA	K (Lys)	0.482	28.213	27	0.96
AAG	K	0.518	30.303	29	1.04
CTA	L (Leu)	0.133	12.539	12	0.80
CTC	L	0.089	8.359	8	0.53
CTG	L	0.133	12.539	12	0.80
CTT	L	0.322	30.303	29	<u>1.93</u>
TTA	L	0.133	12.539	12	0.80
TTG	L	0.189	17.764	17	1.13
ATG	M (Met)	1.000	35.528	34	1.00
AAC	N (Asn)	0.500	17.764	17	1.00
AAT	N	0.500	17.764	17	1.00
CCA	P (Pro)	0.250	11.494	11	1.00
CCC	P	0.091	4.180	4	0.36
CCG	P	0.068	3.135	3	0.27
CCT	P	0.591	27.168	26	<u>2.36</u>
CAA	Q (Gln)	0.514	19.854	19	1.03
CAG	Q	0.486	18.809	18	0.97
AGA	R (Arg)	0.277	13.584	13	<u>1.66</u>
AGG	R	0.298	14.629	14	<u>1.79</u>
CGA	R	0.149	7.315	7	0.89
CGC	R	0.064	3.135	3	0.38
CGG	R	0.085	4.180	4	0.51
CGT	R	0.128	6.270	6	0.77
AGC	S (Ser)	0.145	9.404	9	0.87
AGT	S	0.145	9.404	9	0.87
TCA	S	0.258	16.719	16	<u>1.55</u>

表2 (续) Table 2 (continued)

Codon	Amino acid	Fraction	Frequency (f%)	Number	RSCU
TCC	S	0.145	9.404	9	0.87
TCG	S	0.032	2.090	2	0.19
TCT	S	0.274	17.764	17	<u>1.65</u>
ACA	T (Thr)	0.333	16.719	16	<u>1.33</u>
ACC	T	0.125	6.270	6	0.50
ACG	T	0.042	2.090	2	0.17
ACT	T	0.500	25.078	24	<u>2.00</u>
GTA	V (Val)	0.125	9.404	9	0.50
GTC	V	0.236	17.764	17	0.94
GTG	V	0.264	19.854	19	1.06
GTT	V	0.375	28.213	27	<u>1.50</u>
TGG	W (Trp)	1.000	7.315	7	1.00
TAC	Y (Tyr)	0.444	8.359	8	0.89
TAT	Y	0.556	10.449	10	1.11
TAA	*	0.000	0.000	0	*
TAG	*	0.000	0.000	0	*
TGA	*	1.000	1.045	1	*

*终止密码子; RSCU: 相对同义密码子使用度。大于1.2的RSCU值用下划线表示。

*Termination codon; RSCU: Relative synonymous codon usage. The data with underline mean that the value of RSCU>1.2.

表3 不同物种PPDK的有效密码子数及GC和GC3s含量**Table 3 The ENc values and contents of GC and GC3s for PPDK in different species**

Species	ENc	GC	GC3s
<i>Amaranthus hypochondriacus</i>	53.518	0.435	0.387
<i>Arabidopsis thaliana</i>	56.611	0.479	0.451
<i>Mesembryanthemum crystallinum</i>	51.534	0.453	0.395
<i>Medicago truncatula</i>	47.517	0.417	0.326
<i>Solanum lycopersicum</i>	52.694	0.456	0.411
<i>Solanum tuberosum</i>	52.932	0.455	0.404
<i>Ricinus communis</i>	52.327	0.440	0.371
<i>Echinochloa crusgalli</i>	53.706	0.546	0.584
<i>Oryza sativa</i>	56.626	0.510	0.520
<i>Zea mays</i>	48.425	0.584	0.700
<i>Sorghum bicolor</i>	50.687	0.555	0.611

ENc: 有效密码子数; GC: 密码子的GC含量; GC3s: 密码子第3位的GC含量

ENc: Effective number of codons; GC: GC content of codons; GC3s: The third GC content of codon

差平方和最大(顾万君等, 2002)。分析结果表明, 粽粒苋与马铃薯等6种双子叶植物聚为一类, 玉米、高粱和稗草3种单子叶植物聚为一类, 而水稻所处的地位较为复杂(图1)。通过分析表4数据, 可知水稻与稗草的欧氏平方距离系数最小, 与粽粒苋次之, 与玉米则较远。一般认为, 基于密码子使用频率的聚类树在某种程度上优于基于基因组的系统进化树(Zhou et al., 2005)。为此, 利用DNAMAN软件建立了PPDK基因的核苷酸序列同源树(图2)。水稻与稗草、高粱和玉米

3种单子叶植物聚为一类, 同源性分别为79.2%、79.1%和77.7%, 与粽粒苋的同源性为69.4%(数据未显示), 这与系统分类结果相一致。该结果也表明了水稻PPDK基因的密码子偏好性较为特殊。

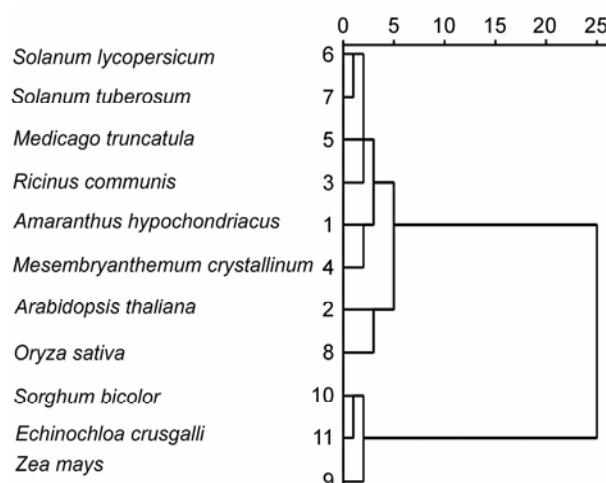
2.4 粽粒苋PPDK基因与大肠杆菌和酵母菌基因组的密码子偏好性比较

密码子偏好性会影响基因在宿主中的高效表达。某一基因与宿主基因组的密码子使用频率的比值可作为

表4 各物种PPDK基因间密码子使用偏性的欧氏平方距离系数**Table 4** Coefficient of absolute squared euclidean distance of codon usage bias among the sample of PPDK

Species	11	1	2	3	4	5	6	7	8	9
1	7.460									
2	7.345	9.114								
3	4.206	10.247	7.259							
4	7.232	13.053	6.166	7.718						
5	5.111	7.667	6.138	4.977	4.810					
6	5.154	8.208	5.720	5.851	4.502	0.552				
7	8.596	9.009	13.272	10.120	16.698	8.845	8.993			
8	32.075	29.650	43.040	36.779	53.140	34.663	34.975	17.228		
9	20.050	19.848	29.505	23.505	37.012	22.755	22.483	9.129	5.921	
10	16.328	17.792	22.771	19.734	31.012	18.607	18.269	7.065	6.668	2.868

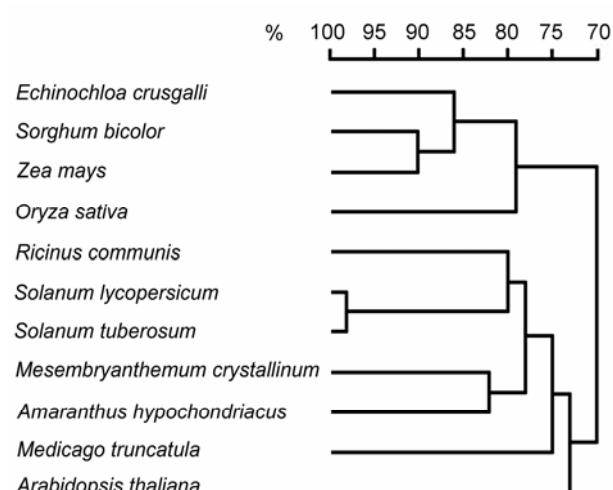
1: *Arabidopsis thaliana*; 2: *Ricinus communis*; 3: *Mesembryanthemum crystallinum*; 4: *Medicago truncatula*; 5: *Solanum lycopersicum*; 6: *Solanum tuberosum*; 7: *Oryza sativa*; 8: *Zea mays*; 9: *Sorghum bicolor*; 10: *Echinochloa crusgalli*; 11: *Amaranthus hypochondriacus*

**图1** 基于PPDK基因相对同义密码子使用度的聚类树状图**Figure 1** Cluster analysis dendrogram of RSCU values of PPDK gene

衡量密码子使用偏好性差异的指标之一。本研究结果表明, 粟粒苋PPDK基因与大肠杆菌和酵母菌基因组密码子使用频率比值大于2或小于0.5的分别有22个和7个(表5), 说明粟粒苋PPDK基因在酵母菌中更易表达。若选择大肠杆菌表达系统, 需对粟粒苋PPDK基因的部分密码子进行优化。

2.5 讨论

不同物种或同一物种的不同基因对密码子的偏好性

**图2** 基于PPDK核苷酸序列建立的植物PPDK基因同源树**Figure 2** Homology trees of plants based on the complete nucleotide sequence of PPDK gene

有所不同(赵耀等, 2008; 晁岳恩等, 2011)。本研究结果表明, 粟粒苋PPDK基因偏好使用以A或T结尾的密码子, 符合双子叶植物密码子的使用特性。ENc值可反映密码子的偏好程度, 与基因的表达水平相关。一般情况下ENc值越大, 基因的表达水平越低。本研究中, 粟粒苋、玉米、高粱和稗草的PPDK基因均为C₄叶绿体型, 是C₄光合途径的关键酶基因, 应为高表达基因, 但粟粒苋PPDK基因的ENc值略低于稗草, 高于玉米和高粱, 甚至高于苜蓿和马铃薯等C₃植物, 其

表5 粱粒苋PPDK基因与大肠杆菌和酵母菌基因组的密码子出现频率(f/‰)

Table 5 Codon frequency of *Amaranthus hypochondriacus* PPDK gene (*AhPPDK*), *Escherichia coli* genome and yeast genome (f/‰)

Codon	Amino acid	<i>AhPPDK</i>	<i>Escherichia coli</i> genome	Yeast genome	<i>AhPPDK/E. coli</i> genome	<i>AhPPDK/Yeast</i> genome
GCA	A (Ala)	9.258	20.6	16.2	0.90	1.15
GCC	A	14.629	25.1	12.6	0.53	1.05
GCG	A	8.359	31.7	6.2	<u>0.17</u>	0.87
GCT	A	39.707	15.6	21.2	<u>2.58</u>	1.9
TGC	C (Cys)	8.359	6.9	4.8	1.09	1.56
TGT	C	6.270	5.5	8.1	1.18	0.8
GAC	D (Asp)	18.809	18.6	20.2	0.81	0.75
GAT	D	31.348	32.1	37.6	1.18	1.01
GAA	E (Glu)	48.067	38.2	45.6	1.06	0.89
GAG	E	31.348	17.7	19.2	1.58	1.46
TTC	F (Phe)	17.764	16.9	18.4	1.03	0.95
TTT	F	17.764	23.2	26.1	1.13	1.00
GGA	G (Gly)	34.483	9	10.9	<u>2.83</u>	<u>2.34</u>
GGC	G	12.539	27.9	9.8	<u>0.32</u>	0.90
GGG	G	15.674	11.3	6	1.19	<u>2.23</u>
GGT	G	25.078	24.4	23.9	0.90	0.92
CAC	H (His)	6.270	9.8	7.8	0.66	0.83
CAT	H	8.359	13.6	13.6	0.94	0.94
ATA	I (Ile)	10.449	5.4	17.8	<u>2.04</u>	0.62
ATC	I	11.494	24.2	17.2	0.63	0.89
ATT	I	29.258	29.8	30.1	1.02	1.01
AAA	K (Lys)	28.213	33.2	41.9	0.90	0.72
AAG	K	30.303	10.7	30.8	<u>3.17</u>	1.10
CTA	L (Leu)	12.539	4.0	13.4	1.80	0.54
CTC	L	8.359	11.0	5.4	<u>0.45</u>	0.93
CTG	L	12.539	50.9	10.5	<u>0.19</u>	0.93
CTT	L	30.303	11.7	12.3	<u>2.62</u>	<u>2.49</u>
TTA	L	12.539	13.9	26.2	0.93	<u>0.49</u>
TTG	L	17.764	14.0	27.2	1.71	0.88
ATG	M (Met)	35.528	27.0	20.9	0.99	1.28
AAC	N (Asn)	17.764	21.4	24.8	0.69	0.59
AAT	N	17.764	18.8	35.7	1.48	0.78
CCA	P (Pro)	11.494	8.5	18.3	<u>2.20</u>	1.02
CCC	P	4.180	5.8	6.8	1.47	1.25
CCG	P	3.135	21.8	5.3	<u>0.17</u>	0.70
CCT	P	27.168	7.3	13.5	<u>2.84</u>	1.53
CAA	Q (Gln)	19.854	15.0	27.3	1.95	1.07
CAG	Q	18.809	29.5	12.1	<u>0.38</u>	0.92
AGA	R (Arg)	13.584	2.9	21.3	<u>5.03</u>	0.69
AGG	R	14.629	1.9	9.2	<u>6.21</u>	1.28
CGA	R	7.315	3.9	3.0	1.08	1.40
CGC	R	3.135	21.0	2.6	<u>0.22</u>	1.81
CGG	R	4.180	6.3	1.7	<u>0.40</u>	1.47
CGT	R	6.270	20.3	6.4	0.54	1.72
AGC	S (Ser)	9.404	16.0	9.8	<u>0.50</u>	0.82

表5 (续) Table 5 (continued)

Codon	Amino acid	AhPPDK	<i>Escherichia coli</i> genome	Yeast genome	AhPPDK/ <i>E. coli</i> genome	AhPPDK/Yeast genome
AGT	S	9.404	9.5	14.2	1.21	0.81
TCA	S	16.719	7.8	18.7	1.79	0.75
TCC	S	9.404	8.9	14.2	1.30	0.82
TCG	S	2.090	8.7	8.6	<u>0.48</u>	<u>0.49</u>
TCT	S	17.764	8.7	23.5	<u>2.25</u>	0.83
ACA	T (Thr)	16.719	8.2	17.8	1.68	0.78
ACC	T	6.270	22.8	12.7	0.53	0.94
ACG	T	2.090	14.8	8.0	<u>0.24</u>	<u>0.45</u>
ACT	T	25.078	9.1	20.3	<u>2.31</u>	1.03
GTA	V (Val)	9.404	11.1	11.8	0.89	0.84
GTC	V	17.764	15.1	11.8	0.73	0.93
G TG	V	19.854	25.5	10.8	0.66	1.56
GTT	V	28.213	18.5	22.1	1.49	1.24
TGG	W (Trp)	7.315	15.2	10.4	1.02	1.49
TAC	Y (Tyr)	8.359	12.1	14.8	1.36	1.11
TAT	Y	10.449	16.5	18.8	0.95	0.83
TAA	*	0.000	2.00	1.1	0.80	1.45
TAG	*	0.000	0.30	0.5	0.67	<u>0.40</u>
TGA	*	1.045	1.10	0.7	0.64	1.00

具下划线的数据表示密码子偏好性具明显偏差($\leq 0.5, \geq 2$)的分值。*终止密码子

The underlined data refer obvious differences of values ($\leq 0.5, \geq 2$) about the codons bias. *The termination codon

原因尚待进一步研究。

密码子偏好性是生物进化过程中逐步形成的, 是对宿主基因组环境的一种适应。粽粒苋 $PPDK$ 基因的密码子偏好性与马铃薯和苜蓿等双子叶植物较为相近, 表明其适宜导入这些植物中。研究中还发现, 水稻 $PPDK$ 基因的密码子偏好性与粽粒苋较近, 与玉米(或高粱)较远, 但基于核苷酸序列的同源性分析, 发现水稻 $PPDK$ 基因与玉米和高粱更为接近, 表明水稻 $PPDK$ 基因的密码子偏好性较为特殊。近年来, 已有将玉米 $PPDK$ 基因导入水稻的研究(张边江等, 2008)。将粽粒苋 $PPDK$ 基因导入水稻, 与转玉米 $PPDK$ 基因水稻比较, 将有助于进一步理解基因的核苷酸序列及密码子偏好性对转基因表达的影响。

外源基因在大肠杆菌或酵母菌中的高效表达将为鉴定基因功能和制备抗体奠定基础。粽粒苋 $PPDK$ 基因的密码子偏好性与酵母基因组较为接近, 与大肠杆菌差异较大。酵母菌表达系统作为后起的外源蛋白表达系统, 与大肠杆菌表达系统相比, 具有翻译后修饰功能, 为进一步鉴定粽粒苋 $PPDK$ 基因功能提供了较为合适的蛋白质表达系统。

参考文献

- 晁岳恩, 吴政卿, 杨会民, 何宁, 杨攀 (2011). 11种植物 $psbA$ 基因的密码子偏好性及聚类分析. 核农学报 **25**, 927–932.
- 杜西河, 许为钢, 胡琳, 张磊, 李艳, 齐学礼, 王会伟, 王玉民 (2013). 转 $ZmPEPC$ 与 $ZmPPDK$ 基因拟南芥对干旱胁迫的反应. 分子植物育种 **11**, 477–484.
- 顾万君, 马建民, 周童, 孙啸, 陆祖宏 (2002). 不同结构的蛋白编码基因的密码子偏性研究. 生物物理学报 **18**, 81–86.
- 郭秀丽, 王玉, 杨路成, 丁兆堂 (2012). 茶树 $CBF1$ 基因密码子使用特性分析. 遗传 **34**, 1614–1623.
- 李平, 白云凤, 冯瑞云, 王原媛, 张维峰 (2011). 粽粒苋苹果酸酶(NAD-ME)基因密码子偏好性分析. 应用与环境生物学报 **17**, 12–17.
- 李华卫, 郝乃斌, 戈巧英, 张其德 (1999). C_3 植物中 C_4 途径的研究进展. 植物学通报 **16**, 97–106.
- 魏松涛, 迟伟, 张立新 (2008). 高等植物碳循环基因工程研究进展. 植物学通报 **25**, 516–525.
- 吴宪明, 吴松锋, 任大明, 朱云平, 贺福初 (2007). 密码子偏性的分析方法及相关研究进展. 遗传 **29**, 420–426.

- 张边江, 华春, 周峰, 周泉澄, 陈全战, 王荣富, 焦德茂 (2008). 转 $PEPC+PPDK$ 双基因水稻的光合特性. 中国农业科学 **41**, 3008–3014.
- 赵耀, 刘汉梅, 顾勇, 黄玉碧 (2008). 玉米waxy基因密码子偏好性分析. 玉米科学 **16**(2), 16–21.
- Li J, Xue QZ (2005). Comparison of MADS transcriptional factor on codon bias in *Arabidopsis* and rice. *J Zhejiang Univ Agric Life Sci* **31**, 513–517.
- Matsuoka M (1995). The gene for pyruvate, orthophosphate dikinase in C₄ plants: structure, regulation and evolution. *Plant Cell Physiol* **36**, 937–943.
- Moons A, Valcke R, Van Montagu M (1998). Low-oxygen stress and water deficit induce cytosolic pyruvate orthophosphate dikinase (PPDK) expression in roots of rice, a C₃ plant. *Plant J* **15**, 89–98.
- Pocalyko DJ, Carroll LJ, Martin BM, Babbitt PC, Dunaway-Mariano D (1990). Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, enzyme I of the bacterial phosphoenolpyruvate: sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. *Biochemistry* **29**, 10757–10765.
- Ryšlavá H, Müller K, Semorádová Š, Synková H, Čeřovská N (2003). Photosynthesis and activity of phosphoenolpyruvate carboxylase in *Nicotiana tabacum* L. leaves infected by *Potato virus A* and *Potato virus Y*. *Photosynthetica* **41**, 357–363.
- Sau K, Gupta SK, Sau S, Ghosh TC (2005). Synonymous codon usage bias in 16 *staphylococcus aureus* phages: implication in phage therapy. *Virus Res* **113**, 123–131.
- Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. *Nucleic Acids Res* **16**, 8207–8211.
- Sharp PM, Li WH (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* **24**, 28–38.
- Wang LJ, Roossinck MJ (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Mol Biol* **61**, 699–710.
- Yamakawa H, Hirose T, Kuroda M, Yamaguchi T (2007). Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant Physiol* **144**, 258–277.
- Zhou H, Wang H, Huang LF, Naylor M, Clifford P (2005). Heterogeneity in codon usages of sobemovirus genes. *Arch Virol* **150**, 1591–1605.

Codon Bias of Pyruvate Orthophosphate Dikinase Gene in *Amaranthus hypochondriacus*

Jiangting Nie¹, Yunfeng Bai^{2,3*}, Feiyan He¹, Jianjun Yan², Weifeng Zhang²

¹College of Biological Engineering, Shanxi University, Taiyuan 030006, China

²Institute of Crop Sciences, Shanxi Academy of Agricultural Sciences, Taiyuan 030031, China

³Key Laboratory of Loess Plateau Crop Gene Resources & Germplasm Creation of Agriculture Ministry, Institute of Variety Resources, Shanxi Academy of Agricultural Sciences, Taiyuan 030031, China

Abstract We used CHIPS, CUSP and Codon W online programs to confirm the codon bias of the complete coding sequence of *PPDK* in the grain *Amaranthus hypochondriacus*, then compared the codon bias with *PPDK* genes from different plant species. We constructed a phylogenetic tree based on codon bias to provide a basis for selecting appropriate *PPDK* receptors for *A. hypochondriacus* by high-photosynthesis gene engineering. *PPDK* in *A. hypochondriacus* and several other dicotyledonous crops was biased toward synonymous codons with A and T at the third codon position, but *Zea mays*, *Oryza sativa* and other monocotyledons use the end of the G or C codon. Phylogenetic analysis suggested that *PPDK* of *A. hypochondriacus* was evolutionarily closer to dicotyledons such as *Solanum tuberosum* and *Medicago truncatula* than the monocotyledons *Z. mays*, *Sorghum bicolor* and *Echinochloa crusgalli*. However, *O. sativa* is an exception in that the codon bias of *PPDK* is closer to that of dicotyledons *A. hypochondriacus* than the monocotyledons *Z. mays* or *S. bicolor*. To select the appropriate protein expression system, we compared the *PPDK* codon bias expression in *A. hypochondriacus* and *Escherichia coli* or yeast. Differences between *A. hypochondriacus* and yeast expression were less than between *A. hypochondriacus* and *E. coli*. The yeast system may be more suitable for expressing *PPDK* of *A. hypochondriacus*.

Key words *Amaranthus hypochondriacus*, codon bias, *PPDK* gene

Nie JT, Bai YF, He FY, Yan JJ, Zhang WF (2014). Codon bias of pyruvate orthophosphate dikinase gene in *Amaranthus hypochondriacus*. *Chin Bull Bot* **49**, 672–681.

* Author for correspondence. E-mail: byfok@126.com

(责任编辑: 孙冬花)