

11种植物 *psbA* 基因的密码子偏好性及聚类分析

晁岳恩 吴政卿 杨会民 何宁 杨攀

(河南省农科院小麦研究中心/河南省小麦生物学重点实验室,河南 郑州 450002)

摘要:植物叶绿体 *psbA* 基因的启动子是叶绿体基因工程中常用的启动子,研究该基因的编码特点对完善叶绿体基因工程的研究设计、提高外源基因在受体物种中高效、稳定的表达具有重要作用。本研究综合运用了多种分析软件,对 11 种植物的叶绿体 *psbA* 基因进行了分析。结果表明,11 种植物 *psbA* 基因的 ENC (Effective Number of Codons) 值都小于 40,显示出了明显的密码子偏好性,即在碱基组成上偏爱以 C 结尾的密码子。RSCU (Relative Synonymous Codon Usage) 值表明共有 20 个密码子在编码使用上具有偏好性,其中有 8 个表现出较强的偏好性;另有 12 个密码子在 *psbA* 中出现率极低或没有出现。在聚类分析中,基于密码子偏好性参数 RSCU 的聚类不能正确反映物种间的进化关系,而基于基因序列的聚类更适合作为系统发育分析的参考。

关键词:*psbA* 基因;密码子偏好性;聚类分析

CLUSTER ANALYSIS AND CODON USAGE BIAS STUDIES ON *psbA* GENES FROM 11 PLANT SPECIES

CHAO Yue-en WU Zheng-qing YANG Hui-min HE Ning YANG Pan

(Henan Key Laboratory of Wheat Biology, Wheat Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002)

Abstract: The promoter of *psbA* gene was frequently used in genetic engineering as the promoter expressed in the chloroplast. Characterization of the codons usage of *psbA* genes is important to the perfect investigations of genetic engineering in chloroplast. In this study, codon usages of *psbA* genes from 11 plant species were analyzed by several softwares. The results showed that all the effective number of codons (ENC) were less than 40 in 11 plant species, suggesting that codon usage bias existed in the *psbA* genes. For example, The codons ending with C were frequently used. Relative synonymous codon usage analyses showed that 20 codons were preferable in the codons of coding DNA sequences (CDS) of *psbA* gene, whereas 12 codons were seldom used. The cluster tree based on the relative synonymous codon usage (RSCU) could not reveal the evolutionary relations among the 11 plant species, whereas that based on CDS of *psbA* gene could reveal their evolutionary relationship much better.

Key words: *psbA* genes; codon usage bias; cluster analysis

叶绿体是植物体内的半自主细胞器,在进化上源于内共生细菌,拥有自己的一套基因组。叶绿体基因在遗传上属细胞质遗传,花粉中的精细胞不含叶绿体基因;因此,以叶绿体基因组作为受体的转基因(叶绿体基因工程),可以避免核转基因系统中因花粉逃逸

而带来的生物安全性问题,另外,还具有目的基因表达量高、定点整合等优点,成为植物基因工程新的研究热点^[1]。

1988年,Boynton等利用基因枪转化体系,成功地衣藻叶绿体进行了遗传转化,标志着叶绿体基因工

收稿日期:2011-01-03 接受日期:2011-05-25

基金项目:农业部转基因生物新品种培育科技重大专项(2009ZX08002-002B和2009ZX08002-014B)

作者简介:晁岳恩(1974-),男,河南濮阳人,副研究员,主要从事作物分子生物学和生物信息学研究。E-mail: chaoyueen@sohu.com

程的开始^[2],1989年,Svab Z等人首次在高等植物中实现了叶绿体转化^[3]。到目前为止,已在烟草、油菜、大豆、拟南芥及棉花等多种植物中成功地对叶绿体进行了遗传转化^[3~7]。

转基因研究中经常要进行基因的异源表达,在翻译过程中,受体物种对外源基因密码子的翻译效率对表达有非常大的制约。因此,利用相应的生物信息学数据库及软件对目标序列进行受体物种的密码子偏好性分析,把需转化的基因进行修饰,以利于转化基因的稳定表达,具有重要的意义。*psbA*基因编码光合系统Ⅱ反应中心的D1蛋白,是叶绿体基因组中一个重要的光调控基因,在光合作用中具有极其重要的作用^[8,9]。在叶绿体基因工程中,*psbA*基因的启动子(T_{psbA}),经常作为目的基因或筛选标记基因的启动子使用。在前人的研究中,多注重该基因和启动子的克隆与功能分析研究,未发现关于*psbA*基因密码子偏好性的探讨^[10~13]。本文以11种植物的叶绿体*psbA*基因为研究对象,分析该基因的碱基组成特点以及密码子的偏好性,以期对叶绿体的基因工程研究提供理论参考。

1 材料与方法

1.1 序列数据

本研究所采用的*psbA*基因的全长编码区序列(coding DNA sequence, CDS)均来源于GenBank,不同植物的*psbA*基因在NCBI(<http://www.ncbi.nlm.nih.gov>)数据库的登录号见表1。

表1 *psbA*基因序列登录号

Table 1 The accession number of *psbA* genes

序列登录号 accession number	物种 species
NC_001320	水稻 <i>Oryza sativa</i> Japonica Group
NC_002762	小麦 <i>Triticum aestivum</i>
NC_001666	玉米 <i>Zea mays</i>
NC_008602	高粱 <i>Sorghum bicolor</i>
NC_007944	陆地棉 <i>Gossypium hirsutum</i>
DQ317523	大豆 <i>Lycine max</i> cultivar PI 437654
NC_001879	烟草 <i>Nicotiana tabacum</i>
DQ231562	马铃薯 <i>Solanum tuberosum</i>
DQ347959	番茄 <i>Lycopersicon esculentum</i> cultivar LA3023
NC_007144	黄瓜 <i>Cucumis sativus</i>
NC_002202	菠菜 <i>Spinacia oleracea</i>

1.2 方法

1.2.1 同义密码子的偏好性分析 下载各*psbA*基因

的全长CDS,包括起始密码子和终止密码子,运用相关分析软件计算出反映密码子偏好性的参数:(1)有效密码子数(effective number of codons, ENC),该值是一个基因的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,代表某个基因的密码子偏好程度^[14];ENC值的范围为20(每个氨基酸只使用1个密码子的极端情况)到61(各个密码子均被平均使用),各基因的ENC值使用在线软件分析获得(<http://vm-bioinfo.toulouse.inra.fr/emboss>)。(2)全长基因中各种碱基的含量及密码子第3位上的各种碱基含量,在蛋白质的合成过程中同义密码子的使用概率并不相同,而同义密码子的主要差别体现在第三位碱基上;该部分研究是利用DNAS_{tar}软件中的Gene Quest模块功能对各基因的CDS进行分析,导出、汇总并计算各CDS的密码子相关数据。(3)同义密码子相对使用度(relative synonymous codon usage, RSCU),是指对于某一特定的密码子在编码对应氨基酸的同义密码子间的相对概率,它去除了氨基酸组成对密码子使用的影响。该值的计算方法为某一密码子所使用的频率与其在无偏使用时预期频率之间的比值,如果密码子的使用没有偏好性,该密码子的RSCU值等于1,当某一密码子的RSCU值大于1时,代表该密码子为使用相对较多的密码子,反之亦然^[15]。

1.2.2 RSCU及CDS的聚类分析 利用SPSS 16.0对11种植物*psbA*基因的RSCU进行聚类分析,聚类方法参照刘汉梅等的介绍^[16,17]。利用ClustalW v1.83软件对*psbA*基因的CDS序列进行联配和保守性分析。在DNAS_{tar}的MegAlign模块中,利用Jotun Hein方法构建系统进化树。

2 结果与分析

2.1 *psbA*基因的密码子偏好性

各物种*psbA*基因的ENC值、各种碱基的总量以及密码子第3位的碱基含量计算结果见表2。结果显示,所分析的11种植物*psbA*基因的ENC值均小于40,表明它们具有明显的密码子使用偏爱性。4种禾本科植物(小麦、玉米、高粱和水稻)的ENC值差别不大,在37.262~38.103之间;而双子叶植物之间的ENC值差异范围明显加大,为34.985~39.244。进一步分析*psbA*基因的碱基组成发现,A含量最高,其次为T,A+T含量在56%~59%之间,而G、C含量相对较低。由于同义密码子之间的最大差别主要在第3位碱基上,因此在表2中分别列出了各种碱基在密码子

第 3 位的含量和百分比,从表中可以看出:以 C 结尾的密码子最多,在 33% 左右,占总 C 含量的 50% 以上,但密码子第 3 位的 GC 含量并不高(50% 以下)。

表 2 *psbA* 基因总碱基构成以及密码子第 3 位碱基的构成(总量及百分比)

Table 2 Base composition of CDS and the third base of codon in *psbA* genes(contents and percent)

物种 species	ENC	总碱基构成 base composition of CDS(%)								密码子末位碱基构成 the third base of codons(%)							
		A	C	G	T	A	C	G	T	A	C	G	T				
水稻 <i>O. sativa</i>	38.103	372	35.0	232	21.8	217	20.4	241	22.7	85	24.0	116	32.8	59	16.7	94	26.6
小麦 <i>T. aestivum</i>	37.386	377	35.5	230	21.7	213	20.1	242	22.8	86	24.3	117	33.1	56	15.8	95	26.8
玉米 <i>Z. mays</i>	38.063	375	35.3	232	21.8	215	20.2	240	22.6	84	23.7	116	32.8	60	16.9	94	26.6
高粱 <i>S. bicolor</i>	37.262	374	35.2	232	21.8	216	20.3	240	22.6	85	24.0	116	32.8	59	16.7	94	26.6
陆地棉 <i>G. hirsutum</i>	36.223	367	34.6	226	21.3	220	20.7	249	23.4	85	24.0	117	33.1	57	16.1	95	26.8
大豆 <i>L. max</i>	36.692	382	36.0	230	21.7	202	19.0	248	23.0	85	24.0	117	33.1	57	16.1	95	26.8
烟草 <i>N. tabacum</i>	37.958	355	33.4	227	21.4	227	21.4	253	23.8	82	23.2	117	33.1	60	16.9	95	26.8
马铃薯 <i>S. tuberosum</i>	39.027	354	33.3	229	21.6	227	21.4	252	23.7	83	23.4	117	33.1	59	16.7	95	26.8
番茄 <i>L. esculentum</i>	39.244	352	33.1	228	21.5	229	21.6	253	23.8	82	23.2	117	33.1	60	16.9	95	26.8
黄瓜 <i>C. sativus</i>	34.985	367	34.6	226	21.3	220	20.7	249	23.4	83	23.4	118	33.3	59	16.7	94	26.6
菠菜 <i>S. oleracea</i>	37.709	372	35.0	228	21.5	210	19.8	252	23.7	84	23.7	117	33.1	58	16.4	95	26.8

注:ENC:有效密码子数;CDS:编码 DNA 序列。在碱基构成中,前面数据为碱基的数量,后面的百分数是其比例。

Note: ENC: effective number of codons; CDS: coding DNA sequence. In the datas of base composition, prior data is contents of a certain base and posterior data is ratio of this base.

2.2 *psbA* 基因密码子的使用特性分析

RSCU 值作为衡量密码子偏好性的标准被该领域研究者普遍接受并广泛应用。从各植物的 *psbA* 基因密码子偏好性的 RSCU 值(表 3)可以看出,*psbA* 基因的密码子有很强的偏好性,以脯氨酸为例:编码脯氨酸的密码子共有 4 个(CCA、CCC、CCG 和 CCT),其中 CCA 在 11 个植物中的 RSCU 值全部在 3 以上,而 CCG 则全部为 0,表明在 *psbA* 基因中,编码脯氨酸的密码子主要是 CCA,而基本不用 CCG。综合分析这 11 个 *psbA* 基因,发现共有 20 个密码子在编码使用上具有偏好性(在表 3 中以 * 表示),其中偏好性较强的有 GAT、GAA、CTC、CCA、AGA、AGG、AGC 和 ACC。

2.3 *psbA* 基因密码子偏好性及 CDS 的聚类分析

在对基因密码子使用概率进行聚类分析的过程中,将每一条基因作为一个对象,将密码子的 RSCU 值作为变量,根据表 3 中各密码子的相对使用度(RSCU),利用 SPSS17.0 进行聚类分析,聚类结果见图 1。由图 1 可以看出:禾本科的 4 种植物密码子偏好性比较接近,其中高粱和玉米比较相似,而水稻和小麦较相似;双子叶植物之间的密码子偏好性差异比较大。另外从图 1 中也可以看出,基于 RSCU 的聚类并不能反映出植物之间的亲缘关系,比如禾本科的 4 种植物就没法单独聚成一类,而是和番茄、大豆、烟草、马铃薯和菠菜聚成了一个大大类,陆地棉和黄瓜则单独聚

成了一类。这一聚类结果与基于 *psbA* 基因的 CDS 构建的系统进化树(图 2,在 DNSstar 的 MegAlign 模块中,用 Jotun Hein 方法构建的系统进化树)不一致,基于 CDS 的系统进化树更接近这 11 种植物的真实系统分类,并且按照亲缘关系将茄科的 3 种植物聚在了一起,表明 *psbA* 基因 CDS 的差异大小与物种的亲缘关系远近有一定关系。

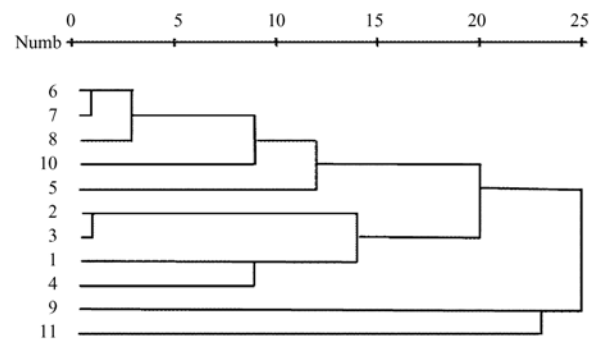


图 1 基于 *psbA* 基因密码子偏好性的聚类分析

Fig. 1 Cluster tree based on the RSCU values of *psbA* genes

1: 小麦; 2: 玉米; 3: 高粱; 4: 水稻; 5: 菠菜; 6: 马铃薯; 7: 番茄; 8: 烟草; 9: 陆地棉; 10: 大豆; 11: 黄瓜

1: *Triticum aestivum*; 2: *Zea mays*; 3: *Sorghum bicolor*; 4: *Oryza sativa*; 5: *Spinacia oleracea*; 6: *Solanum tuberosum*; 7: *Lycopersicon esculentum*; 8: *Nicotiana tabacum*; 9: *Gossypium hirsutum*; 10: *Lycine max*; 11: *Cucumis sativus*

表 3 各植物 *psbA* 基因同义密码子的相对使用度Table 3 The relative synonymous codon usage of *psbA* genes

氨基酸 amino acid	密码子 codon	水稻 <i>O. sativa</i>	小麦 <i>T. aestivum</i>	玉米 <i>Z. mays</i>	高粱 <i>S. bicolor</i>	陆地棉 <i>G. hirsutum</i>	大豆 <i>L. max</i>	烟草 <i>N. tabacum</i>	马铃薯 <i>S. tuberosum</i>	番茄 <i>L. esculentum</i>	黄瓜 <i>C. sativus</i>	菠菜 <i>S. oleracea</i>
Ala	GCA	0.44	0.44	0.40	0.40	0.00	0.33	0.00	0.00	0.00	0.00	0.00
	GCC	0.89	0.89	0.80	0.80	1.45	1.33	2.00	1.67	1.67	1.33	0.89
	GCG	1.33	1.33	1.60	1.6	1.09	0.67	0.67	0.67	0.67	0.44	1.33
Cys	GCT*	1.33	1.33	1.20	1.20	1.45	1.67	1.33	1.67	1.67	2.22	1.78
	TGC*	1.71	1.71	1.75	1.75	1.50	1.71	1.56	1.56	1.56	1.75	1.50
Asp	TGT	0.29	0.29	0.25	0.25	0.50	0.29	0.44	0.44	0.44	0.25	0.50
	GAC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Glu	GAT*	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
	GAA*	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Phe	GAG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	TTC*	2.00	1.85	1.83	1.83	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Gly	TTT	0.00	0.15	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	GGA	0.73	1.00	1.33	1.33	0.86	0.92	1.45	1.45	1.45	1.33	2.00
	GGC	1.45	1.50	0.44	0.44	0.57	0.92	1.09	1.09	1.09	0.33	0.80
His	GGG	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.00
	GGT*	1.82	1.50	2.22	2.22	2.29	2.15	1.45	1.45	1.45	2.33	1.20
	CAC	0.15	0.15	0.15	0.15	0.00	0.00	0.00	0.00	0.00	0.15	0.00
Ile	CAT*	0.85	1.85	1.85	1.85	2.00	2.00	2.00	2.00	2.00	1.85	2.00
	ATA	0.82	0.86	0.84	0.87	0.82	0.83	0.83	0.88	0.88	0.75	0.84
Lys	ATC	0.82	0.75	0.72	0.75	0.82	0.72	0.67	0.71	0.71	0.75	0.72
	ATT*	1.36	1.38	1.44	1.38	1.36	1.45	1.50	1.41	1.41	1.50	1.44
	AAA	1.00	1.00	0.80	0.77	1.08	1.33	1.08	1.08	1.08	0.91	1.20
Leu	AAG	1.00	1.00	1.20	1.23	0.92	0.67	0.92	0.92	0.92	1.09	0.80
	CTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CTC*	3.20	3.20	3.38	3.38	2.18	2.77	2.50	2.50	2.18	2.31	2.50
Asn	CTG	0.40	0.40	0.75	0.75	0.55	0.92	0.50	1.00	1.09	0.46	0.50
	CTT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00
	TTA	0.40	0.40	0.38	0.38	0.55	0.46	0.50	0.50	0.55	0.46	0.50
Pro	TTG*	2.00	2.00	1.50	1.50	2.73	1.85	2.50	2.00	2.18	2.31	2.50
	AAC	0.74	0.81	0.74	0.71	0.69	0.52	0.69	0.72	0.75	0.71	0.67
	AAT*	1.26	1.19	1.26	1.29	1.31	1.48	1.31	1.28	1.25	1.29	1.33
Gln	CCA*	3.33	3.64	3.08	3.33	3.64	3.33	3.33	3.08	3.08	3.64	3.33
	CCC	0.33	0.00	0.61	0.33	0.00	0.33	0.33	0.62	0.62	0.00	0.33
	CCG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Arg	CCT	0.33	0.36	0.31	0.33	0.36	0.33	0.33	0.31	0.31	0.36	0.33
	CAA*	1.40	1.56	1.50	1.56	0.25	2.00	1.71	1.71	1.71	1.75	2.00
	CAG	0.60	0.44	0.50	0.44	0.75	0.00	0.29	0.29	0.29	0.25	0.00
Ser	AGA*	3.11	3.11	2.77	3.00	3.12	2.54	2.61	2.61	2.61	3.39	2.61
	AGG*	2.44	2.44	2.77	2.54	2.16	2.77	2.61	2.61	2.61	2.61	2.35
	CGA	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.26	0.26	0.00	0.52
Thr	CGC	0.44	0.44	0.46	0.46	0.24	0.69	0.26	0.26	0.26	0.00	0.52
	CGG	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.00
	CGT	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.26	0.26	0.00	0.00
Val	AGC*	3.07	3.07	3.43	3.35	3.71	3.37	3.35	3.35	3.35	3.55	3.26
	AGT*	1.53	1.73	1.57	1.53	1.00	1.17	1.40	1.40	1.40	1.09	1.57
	TCA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tyr	TCC	0.98	0.93	0.71	0.84	0.86	0.88	0.70	0.70	0.70	0.82	0.78
	TCG	0.14	0.13	0.14	0.14	0.14	0.29	0.28	0.28	0.28	0.27	0.13
	TCT	0.28	0.13	0.14	0.14	0.29	0.29	0.28	0.28	0.28	0.27	0.26
Tyr	ACA	0.10	0.10	0.10	0.10	0.21	0.11	0.22	0.22	0.22	0.21	0.20
	ACC*	3.36	2.40	2.46	2.46	2.36	2.44	2.27	2.33	2.33	2.46	2.40
	ACG	0.82	0.80	0.72	0.72	0.82	0.89	0.86	0.89	0.89	0.92	0.80
Tyr	ACT	0.72	0.70	0.72	0.72	0.62	0.56	0.65	0.56	0.56	0.41	0.60
	GTA	0.96	0.87	1.00	0.83	1.04	1.00	0.97	0.97	0.97	1.04	0.91
	GTC	0.16	0.17	0.17	0.17	0.17	0.25	0.41	0.41	0.41	0.30	0.18
Tyr	GTT*	0.96	1.04	1.17	1.17	0.70	0.75	0.83	0.69	0.69	0.89	1.09
	TAC*	1.92	1.91	1.67	1.83	2.09	2.00	1.79	1.93	1.93	1.78	1.82
Tyr	TAT	1.69	1.83	1.57	1.69	1.71	1.73	1.71	1.71	1.71	1.83	1.60
	TAT	0.31	0.17	0.43	0.31	0.29	0.27	0.29	0.29	0.29	0.17	0.40

注: * 表示 RSCU 值大于 1 的密码子。

Noet: * means the RSCU value is more than 1.

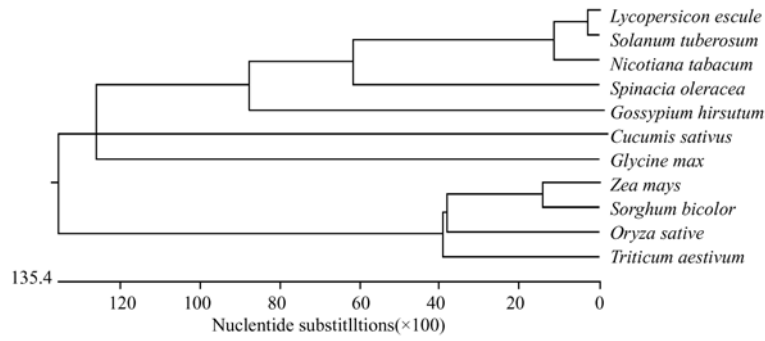


图 2 基于 *psbA* 基因 CDS 的聚类图

Fig. 2 A dendrogram of CDS sequences of *psbA* genes

为分析 *psbA* 基因 CDS 的序列保守性,利用 ClustalW v1.83 软件对 11 种植物的 *psbA* 基因的 CDS 进行了序列联配(图 3),发现了 2 个在 20 个碱基以上的序列保守区(图 3 黑方框内序列),位置分别在 218

~237bp 处和 1037~1062bp 处,利用这 2 个位置的序列设计引物,有望在叶绿体基因组未测序的植物中克隆出 *psbA* 基因,进而为叶绿体基因工程研究奠定基础。

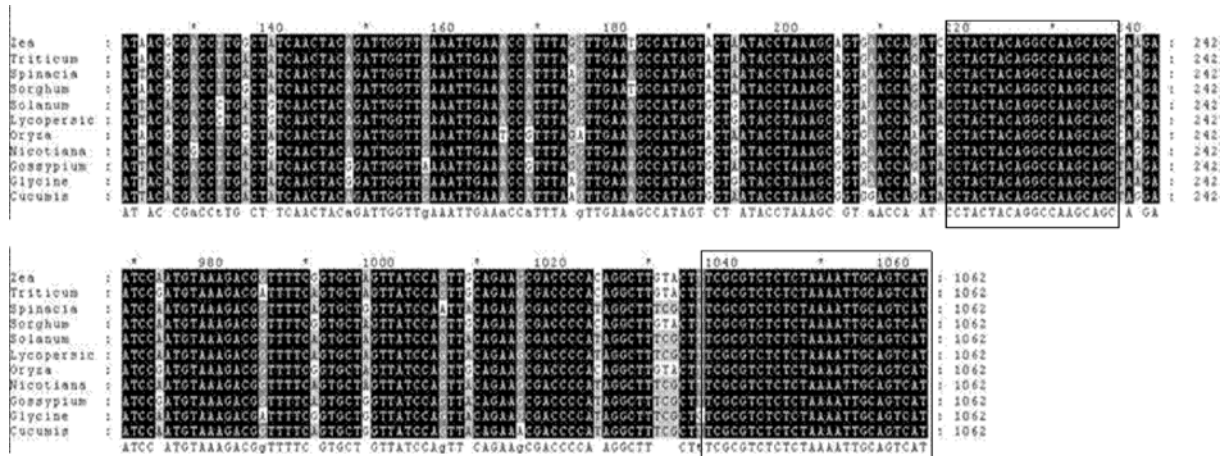


图 3 11 种植物的 *psbA* 基因编码区序列的联配(部分展示)

Fig. 3 Alignment of CDS sequences of *psbA* genes from 11 plant species (partial)

联配软件为 ClustalW v1.83,参数未做任何调整,保守的碱基展示在联配图的最下面一行。

Sequences were aligned using ClustalW v1.83 without any manual adjustment. The consensus sequence is shown below.

3 讨论

植物基因中不同密码子的出现频率明显不同,同一种氨基酸的不同密码子比率也有差异。密码子偏好性是在复杂因子作用下形成的,是基因突变和选择的结果^[18, 19],还与基因编码结构和功能及基因表达有密切的联系^[20],受进化过程中多种因素的影响。

从 *psbA* 基因的 ENC 值和 CDS 的碱基组成上看(表 2),11 个植物的 ENC 值都小于 40,表明这些植物在密码子使用上都有明显的偏好性;在碱基组成上,总 C 含量并不高(21.3% ~ 21.8%),但密码子末位 C 含

量却高达 33% (32.8% ~ 33.3%),表明 *psbA* 基因的编码区偏爱以 C 结尾的密码子。RSCU 值也是衡量密码子偏好性的重要指标,表 3 中 RSCU 值显示出共有 20 个密码子在编码使用上具有明显的偏好性,其中有 8 个表现出较强的偏好性,在这 8 个强偏好性密码子中有 3 个是以 C 结尾;这些分析均表明 *psbA* 基因在密码子的编码使用上具有以 C 结尾的偏好性。

在这 11 种植物的 *psbA* 基因中,有 5 个密码子的 RSCU 值全部显示为 0 (GAC、GAG、CTA、CCG 和 TCA),另外还有 7 个密码子只在几种植物中有少量应用(RSCU 值在多数植物中为 0,少数植物在 0.5 以下:GCA、TTT、GGG、CTT、GGA、CGG 和 GGT);虽然这些

密码子的偏好性使用情况在叶绿体基因工程研究中需加以考虑,但尽量选用 20 个有偏好性的密码子,避开 12 个使用率非常低的密码子。

虽然这 11 种植物的 *psbA* 基因都具有一定的密码子使用偏好性,但差异并不大,表明叶绿体基因在密码子使用上具有保守性,这可能与其做为核外遗传物质的特性有关,相对于核基因组来说,叶绿体基因组不容易接受外来遗传物质的影响,在进化过程中变化不大。

基于基因密码子用法的聚类分析与基于 CDS 的进化分析结果可能存在一定差异,这种现象在前人的研究中也出现过^[16]。在本研究中,图 1 和图 2 的 2 个聚类图也不完全一致:图 1 未能将禾本科的 4 种植物和其他的双子叶植物区分开,而图 2 则较好地反映了植物之间的进化关系,说明基因序列比密码子偏好性更适应于做进化分析使用。分析导致该现象的可能原因是选取的基因数量太少,如果多增加一些基因参与聚类分析,图 1 的聚类结果就有可能更接近真实的系统分类;也有可能和叶绿体基因的遗传特性有关,核外遗传物质在进化上非常保守,虽然叶绿体基因在植物的长期进化过程中,通过与核基因组之间的物质与信息交流,受核基因组的影响而表现出一定的种族特异性,但通过单基因的密码子偏性聚类仍然难以反映出正确的亲缘关系;当然还有可能是其他原因影响了密码子偏好性参数的聚类效果,如本文仅选用了 RSCU 值进行聚类,如果选择别的密码子偏好性参数,或多个参数联用,效果可能会更好一些。

在叶绿体转基因研究中,外源基因是通过同源序列重组的方法来实现定点插入的,因此,获得叶绿体基因的部分序列是进行叶绿体转基因的前提。本研究在对 11 个 *psbA* 基因进行联配时发现,有 2 个保守区的序列长度在 20 个碱基以上(图 3 黑色方框内序列),利用这 2 个位置的序列设计引物,有望在叶绿体基因组未测序的植物中克隆出 *psbA* 基因,进而为叶绿体基因工程研究奠定基础。

参考文献:

- [1] Daniell H, Khan M, Allison L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology [J]. Trends Plant Sci, 2002, 7(2): 84-91
- [2] Boynton J E, Gillham N W, Harris E H, Hosler J P, Johnson A M, Jones A R, Randolph-Anderson B L, Robertson D, Klein T M, Shark K B, Sanford J C. Chloroplast transformation in chlamydomonas with high velocity microprojectiles [J]. Science, 1988, 240: 1534-1538
- [3] Svab Z, Hajdukiewicz P, Maliga P. Stable transformation of plastids in higher plants [J]. Proc Natl Acad Sci USA, 1990, 87(21): 8526-8530
- [4] Hou B K, Zhou Y H, Wan L H, Zhang L Z, Shen G F, Chen Z H, Hu Z M. Chloroplast transformation in oilseed rape [J]. Transgenic Res, 2003, 12(1): 111-114
- [5] Zhang X H, Portis A R Jr, Wildholm J M. Plastid transformation of soybean suspension cultures [J]. J Plant Biotechnol, 2001, 3: 39-44
- [6] Sikdar S R, Serino G, Chaudhuri S, Pal M. Plastid transformation in *Arabidopsis thaliana* [J]. Plant Cell Rep, 1998, 18(1): 20-24
- [7] Kumar S, Dhingra A, Daniell H. Stable transformation of the cotton plastid genome and maternal inheritance of transgenes [J]. Plant Mol Biol, 2004, 56(2): 203-216
- [8] Ellis R J. Chloroplast proteins: synthesis, transport, and assembly [J]. Annu Rev Plant Physiol, 1981, 32: 111-137
- [9] Wu N H, Cote C J. Structure of the chloroplast *psbA* gene encoding the Q_B protein from *Oryza sativa* [J]. Develop Genet, 1987, 8: 339-350
- [10] 吴乃虎,方晓华,施晓梅,张晓武,周立,黄美娟,孙凯鸣. 高粱叶绿体 *psbA* 基因的结构特征及其 5' 非编码区的调控效应[J]. 中国科学 C 辑, 1999, 29(4): 397-406
- [11] Zeidner G, Preston C M, Delong E F, Massana R, Post A F, Scanlan D L, Béjà O. Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses [J]. Environ Microbiol, 2003(5): 212-216
- [12] 吴晓微,孙雪,陆开彤,汪一冰,张晓龙. 小球藻 pM 基因的克隆与序列分析[J]. 水产科学, 2008, 27(7): 360-362
- [13] 袁进成,刘颖慧. 基于叶绿体 *psbA* 基因初步探讨植物系统发育的关系[J]. 江苏农业科学, 2009, 4: 46-49
- [14] Wright F. The "effective number of codons" used in a gene [J]. Gene, 1990, 87(1): 23-29
- [15] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. J Mol Evol, 1986, 24: 28-38
- [16] 刘汉梅,赵耀,顾勇,张怀渝,黄玉碧. 几种植物 waxy 基因的密码子用法特性分析[J]. 核农学报, 2010, 24(3): 476-481
- [17] 刘汉梅,何瑞,赵耀,张怀渝,黄玉碧. 玉米密码子用法分析[J]. 核农学报, 2008, 22(2): 141-147
- [18] 周晓明,赵慧芳,陆嘉良,潘浩,徐眉,赵根明,姜庆五,汪华,俞顺章. 甲型流感病毒基因变异与生存选择压力相关性分析[J]. 中华传染病杂志, 2005, 23(4): 221-223
- [19] Wong G K S, Wang J, Tao L, Tan J, Zhang J G, Douglas A P, Yu J. Compositional gradients in gramineae genes [J]. Genome Res, 2002, 12: 851-856
- [20] Hélène C, Frédérique L, Michel C, Alain H. Codon usage and gene function are related in sequences of *Arabidopsis thaliana* [J]. Gene, 1998, 209: GC1-GC38

(责任编辑 王媛媛)