

# 7 种作物叶绿体基因的密码子偏好性及聚类分析

晁岳恩 常 阳 王美芳 何盛莲 赵献林 雷振生

(河南省农业科学院 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:** 研究作物叶绿体蛋白编码基因的编码特点,对指导农作物的叶绿体基因工程研究设计,促进外源基因在受体作物中的高效、稳定表达具有重要作用。为此,综合运用了多种分析软件,对 7 种大田作物的叶绿体蛋白编码基因进行分析。结果表明:叶绿体蛋白编码基因的总碱基构成在 7 种作物中差异不大,都是 A 含量最高,而 G、C 的含量较低;但在密码子第 3 位的碱基构成上却有明显差别,3 种双子叶作物偏爱以 C、T 结尾的密码子,而禾本科的 4 种作物则偏爱以 T、C 结尾的密码子;RSCU(同义密码子的相对使用度)值显示出禾本科作物有 26 个密码子具有偏好性,而双子叶作物有 25 个密码子具有偏好性,但二者之间有 22 个相同的偏好性密码子。基于 RSCU 值的聚类结果显示,植物的密码子偏好性具有种族特征,表明基于叶绿体蛋白编码基因的密码子使用特性的聚类结果可以作为系统发育分析的重要补充。

**关键词:** 作物;叶绿体基因;密码子偏好性;聚类分析

中图分类号:Q349+.55 文献标识码:A 文章编号:1000-7091(2012)04-0060-05

## Codon Usage Bias and Cluster Analysis on Chloroplastic Genes from Seven Crop Species

CHAO Yue-en, CHANG Yang, WANG Mei-fang, HE Sheng-lian, ZHAO Xian-lin, LEI Zhen-sheng

(Henan Key Laboratory of Wheat Biology, Wheat Research Center, Henan Academy of

Agricultural Sciences, Zhengzhou 450002, China)

**Abstract:** The analysis of coded characters of crop chloroplast genes is benefit to the genetic engineering design of chloroplast and the efficient and stable expression of exotic genes in recipient crops. In this study, the codon usage bias of genes encoded by chloroplast DNA in 7 field crops were analyzed by means of several softwares. Results showed that there was little difference among the characters of chloroplast-coded genes among the 7 crops, which held a higher content of A and lower content of G and C. However, there were significant differences in the base composition of the third codon. The codons in the 3 dicotyledon crops were preferred to end with C, and second with T, while the other 4 gramineous crops preferred the codons ended with prior T, and second with C. There were 26 preferable codons used in the 4 gramineous crops and 25 in the 3 dicotyledon plants, among which 22 codons were totally same between the 2 kinds of crops. The cluster analysis based on RSCU value showed that the codon usage bias in the genes encoded by chloroplast was related to the plant species, indicating that the cluster analysis of RSCU in chloroplast genes could be used as a complementary method for the systematic developmental analysis.

**Key words:** Crop; Chloroplastic gene; Codon usage bias; Cluster analysis

叶绿体普遍存在于植物、藻类和部分原生生物中,是把光能转化为化学能的能量转换器。叶绿体是质体的一种,进化上源于内共生细菌,拥有自身的一套基因组,是植物体内的半自主细胞器,在遗传特性上属细胞质遗传,花粉中的精细胞不含叶绿体的基因。因此,以叶绿体基因组作为受体进行的转基因

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,叶绿体基因还具有拷贝数多的特点,一个细胞内的叶绿体基因组可以达到上万个,一旦目的基因成功转化到叶绿体基因组内,再通过选择达到纯合后,就能得到远高于常规转基因表达量的转基因后代<sup>[1]</sup>。

收稿日期:2011-12-01

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;2009ZX08002-014B)

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

自首例转基因烟草问世以来,植物核基因组转化技术已日趋完善,成为现代生物学研究的重要技术之一。但常规核基因组转化方法也存在一系列问题,如目的基因表达量低且在后代中不稳定,目的基因易随花粉漂移产生环境污染问题等,这促使生物研究者开展新的转基因技术研究。1988 年,Boynton 等<sup>[2]</sup>利用基因枪转化体系,成功地对衣藻叶绿体进行了遗传转化,标志着叶绿体基因工程的开始;之后 Svab 等<sup>[3]</sup>于 1989 年首次在高等植物中实现了叶绿体转化。到目前为止,已有烟草、油菜、大豆、拟南芥及棉花等多种植物叶绿体转化成功的报道<sup>[3-7]</sup>。

叶绿体遗传转化是基于基因序列同源重组的转化技术,研究叶绿体基因组的序列特征、密码子碱基的组成规律,不仅有助于转化效率的提高,也有利于目的基因的高效、稳定表达。目前,对物种基因组特点及密码子使用特性的相关研究,大多是针对细胞核基因<sup>[8-13]</sup>,专门对叶绿体基因组的研究极少。为此,本研究以 7 种大田作物为研究对象,分析它们之间基因的碱基组成特点以及密码子的偏好性差异,期望能对这几种作物的叶绿体基因工程研究提供一些理论上的参考。除此之外,还探讨了利用叶绿体蛋白编码基因的密码子偏好性进行物种间分类的可能性。

表 1 叶绿体基因序列登录号

Tab. 1 The accession number of chloroplastic genes

| 序列登录号<br>Accession number | 物种<br>Species                         |
|---------------------------|---------------------------------------|
| NC_002762                 | 小麦 <i>Triticum aestivum</i>           |
| NC_001320                 | 水稻 <i>Oryza sativa Japonica Group</i> |
| NC_001666                 | 玉米 <i>Zea mays</i>                    |
| NC_008602                 | 高粱 <i>Sorghum bicolor</i>             |
| DQ231562                  | 马铃薯 <i>Solanum tuberosum</i>          |
| NC_001879                 | 烟草 <i>Nicotiana tabacum</i>           |
| NC_007944                 | 陆地棉 <i>Gossypium hirsutum</i>         |

## 1 材料和方法

### 1.1 序列数据

对于本研究中的 7 种植物,分别从叶绿体基因组中选取约 10 070 个碱基(序列来自 atpA、atpB、psaA、psaB、psbA、psbB 共 6 个蛋白编码基因),大约占总叶绿体基因中蛋白编码序列的 12.8%。不同作物的叶绿体基因序列在 NCBI(<http://www.ncbi.nlm.nih.gov>) 数据库上的登录号见表 1。

### 1.2 方法

1.2.1 同义密码子的偏好性分析 下载所研究的各种植物叶绿体的蛋白编码序列,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。因为在蛋白质的合成过程中,同义密码子的使用概率并不相同,主要差别体现在第 3 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 (Coding DNA sequence) 进行分析,导出、汇总并计算各 CDS 的密码子相关数据。②同义密码子相对使用度(Relative synonymous codon usage, RSCU)。它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,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<sup>[14]</sup>,计算公式为:

$$RSCU_{ij} = \frac{x_{ij}}{x} = \frac{x_{ij}}{\frac{1}{n} \sum_{j=1}^n x_{ij}}$$

公式中  $x_{ij}$  表示编码第  $i$  个氨基酸的第  $j$  个密码子的出现次数,  $n_i$  表示编码第  $i$  个氨基酸的同义密码子的数量(其值为 1~6)。

1.2.2 RSCU 及 CDS 的聚类分析 利用 SPSS 16.0 对 7 种作物叶绿体基因的 RSCU 进行聚类分析,聚类方法参照相关的文献介绍<sup>[15-16]</sup>。

## 2 结果与分析

### 2.1 叶绿体基因的密码子偏好性

参与分析的各物种 6 个蛋白编码基因的 4 种碱基总量以及密码子第 3 位的碱基含量计算结果见表 2 和表 3。从表 2 可以看出,所分析的 7 种作物的叶绿体蛋白编码基因中,总碱基构成以 A 最多,占总含量的 30% 左右, A + T 的含量在 57% 以上,而 G、C 的含量较低。从表 3 可以看出,4 种禾本科作物与 3 种双子叶作物之间有明显差别,在禾本科作物中,以 T 结尾的密码子最多,在 29% 以上,其次是 C;而在马铃薯、烟草和棉花中,则是以 C 结尾的密码子最多,在 30% 以上,而排在第 2 位的是 T。

### 2.2 叶绿体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。从各作物叶绿体蛋白编码基因的 RSCU 值(表 4)可以看出,叶绿体蛋白编码基因在密码子的使用上有一定的偏好性。以亮氨酸(Leu)和丝氨酸(Ser)为例,编码亮氨酸的密

码子共有 6 个(CTA、CTC、CTG、CTT、TTA 和 TTG) , 其中 TTG 在 7 种作物中的 RSCU 值均在 2 以上,说明在叶绿体中编码亮氨酸时偏好使用 TTG 密码子,而尽可能避开 CTA( RSCU 值全部在 0.3 以下);对于丝氨酸也是有 6 个密码子编码( AGC、AGT、TCA、TCC、TCG 和 TCT) ,但在编码氨基酸时主要使用 AGC、AGT 和 TCC,尤其是 AGC 密码子的使用率最高,使用率最低的是 TCA。综合分析这 7 种作物的叶绿体蛋白编码基因,共发现有 22 个密码子在编码使用上具有共同的偏好性( RSCU 值在 1 以上,在表

4 中以\* 表示)。除此以外,还有一些密码子的偏好性在物种上有所差异,主要表现在禾本科的 4 种作物与双子叶的 3 种作物之间:在禾本科的 4 种作物中,另外有 4 个密码子( 3 个 T 结尾和 1 个 A 结尾: ATT、CTT、ACT 与 GTA) 的 RSCU 值在 1 以上;而在 3 种双子叶作物中,则另有 3 个以 C 结尾的密码子( GGC、ATC 和 TAC) 的 RSCU 值在 1 以上。这些差异正好与表 3 的结果相对应,表 3 显示,禾本科的作物偏好以 T、C 结尾的密码子,而双子叶的作物主要偏好以 C、T 结尾的密码子。

表 2 叶绿体蛋白编码基因的碱基构成

Tab. 2 Base composition of chloroplastic genes

%

| 物种<br>Species           | 总碱基构成 Base composition of CDS |       |       |       |       |       |
|-------------------------|-------------------------------|-------|-------|-------|-------|-------|
|                         | A                             | C     | G     | T     | A + T | G + C |
| 小麦 <i>T. aestivum</i>   | 30.80                         | 21.02 | 21.28 | 26.90 | 57.70 | 42.30 |
| 水稻 <i>O. sativa</i>     | 30.50                         | 21.14 | 21.54 | 26.82 | 57.32 | 42.68 |
| 玉米 <i>Z. mays</i>       | 30.50                         | 21.10 | 21.58 | 26.82 | 57.32 | 42.68 |
| 高粱 <i>S. bicolor</i>    | 30.48                         | 21.12 | 21.65 | 26.75 | 57.23 | 42.77 |
| 马铃薯 <i>S. tuberosum</i> | 30.17                         | 21.70 | 20.83 | 27.30 | 57.47 | 42.53 |
| 烟草 <i>N. tabacum</i>    | 30.22                         | 21.68 | 20.74 | 27.36 | 57.69 | 42.31 |
| 陆地棉 <i>G. hirsutum</i>  | 29.81                         | 21.74 | 20.98 | 27.47 | 57.28 | 42.72 |

表 3 叶绿体蛋白编码基因密码子第 3 位碱基的构成

Tab. 3 The third base of codons in chloroplastic genes

%

| 物种<br>Species           | 密码子末位碱基构成 The third base of codons |       |       |       |       |       |
|-------------------------|------------------------------------|-------|-------|-------|-------|-------|
|                         | A                                  | C     | G     | T     | A + T | G + C |
| 小麦 <i>T. aestivum</i>   | 26.03                              | 26.16 | 18.49 | 29.32 | 55.35 | 44.65 |
| 水稻 <i>O. sativa</i>     | 25.89                              | 26.22 | 18.77 | 29.11 | 55.00 | 45.00 |
| 玉米 <i>Z. mays</i>       | 25.98                              | 26.21 | 18.56 | 29.25 | 55.23 | 44.77 |
| 高粱 <i>S. bicolor</i>    | 26.07                              | 26.25 | 18.59 | 29.09 | 55.16 | 44.84 |
| 马铃薯 <i>S. tuberosum</i> | 23.06                              | 30.27 | 19.13 | 27.53 | 50.59 | 49.41 |
| 烟草 <i>N. tabacum</i>    | 23.10                              | 30.29 | 19.14 | 27.47 | 50.57 | 49.43 |
| 陆地棉 <i>G. hirsutum</i>  | 23.09                              | 30.13 | 18.95 | 27.83 | 50.92 | 49.18 |

表 4 7 种植物叶绿体基因同义密码子的相对使用度

Tab. 4 The relative synonymous codon usage of chloroplastic genes from 7 plant species

| 氨基酸<br>Amino acid | 密码子<br>Codon | 小麦<br><i>T. aestivum</i> | 水稻<br><i>O. sativa</i> | 玉米<br><i>Z. mays</i> | 高粱<br><i>S. bicolor</i> | 马铃薯<br><i>S. tuberosum</i> | 烟草<br><i>N. tabacum</i> | 陆地棉<br><i>G. hirsutum</i> |
|-------------------|--------------|--------------------------|------------------------|----------------------|-------------------------|----------------------------|-------------------------|---------------------------|
| Ala               | GCA          | 0.67                     | 0.70                   | 0.63                 | 0.63                    | 0.52                       | 0.51                    | 0.42                      |
|                   | GCC*         | 1.25                     | 1.13                   | 1.25                 | 1.23                    | 1.61                       | 1.57                    | 1.45                      |
|                   | GCG          | 0.60                     | 0.62                   | 0.58                 | 0.60                    | 0.56                       | 0.63                    | 0.69                      |
| Cys               | GCT*         | 1.48                     | 1.56                   | 1.55                 | 1.55                    | 1.31                       | 1.29                    | 1.45                      |
|                   | TGC*         | 1.23                     | 1.15                   | 1.14                 | 1.12                    | 1.30                       | 1.33                    | 1.17                      |
|                   | TGT          | 0.77                     | 0.85                   | 0.86                 | 0.88                    | 0.70                       | 0.67                    | 0.83                      |
| Asp               | GAC          | 0.42                     | 0.38                   | 0.35                 | 0.34                    | 0.33                       | 0.31                    | 0.48                      |
|                   | GAT*         | 1.58                     | 1.62                   | 1.65                 | 1.66                    | 1.67                       | 1.69                    | 1.52                      |
| Glu               | GAA*         | 1.61                     | 1.64                   | 1.65                 | 1.66                    | 1.65                       | 1.66                    | 1.66                      |
|                   | GAG          | 0.39                     | 0.36                   | 0.35                 | 0.34                    | 0.35                       | 0.34                    | 0.34                      |
| Phe               | TTC*         | 1.12                     | 1.08                   | 1.13                 | 1.13                    | 1.14                       | 1.13                    | 1.13                      |
|                   | TTT          | 0.88                     | 0.92                   | 0.87                 | 0.87                    | 0.86                       | 0.87                    | 0.87                      |
| Gly               | GGA*         | 1.22                     | 1.18                   | 1.16                 | 1.17                    | 1.14                       | 1.13                    | 1.18                      |
|                   | GGC          | 0.78                     | 0.84                   | 0.81                 | 0.83                    | 1.14                       | 1.13                    | 1.02                      |
|                   | GGG          | 0.82                     | 0.93                   | 0.85                 | 0.86                    | 0.49                       | 0.51                    | 0.50                      |
|                   | GGT*         | 1.18                     | 1.05                   | 1.18                 | 1.15                    | 1.24                       | 1.23                    | 1.29                      |
| His               | CAC          | 0.44                     | 0.45                   | 0.52                 | 0.50                    | 0.47                       | 0.44                    | 0.52                      |
|                   | CAT*         | 1.55                     | 1.55                   | 1.48                 | 1.50                    | 1.53                       | 1.56                    | 1.48                      |
| Ile               | ATA          | 0.84                     | 0.83                   | 0.85                 | 0.86                    | 0.92                       | 0.92                    | 0.90                      |

续表 4:

| 氨基酸        | 密码子   | 小麦                 | 水稻               | 玉米             | 高粱                | 马铃薯                 | 烟草                | 陆地棉                |
|------------|-------|--------------------|------------------|----------------|-------------------|---------------------|-------------------|--------------------|
| Amino acid | Codon | <i>T. aestivum</i> | <i>O. sativa</i> | <i>Z. mays</i> | <i>S. bicolor</i> | <i>S. tuberosum</i> | <i>N. tabacum</i> | <i>G. hirsutum</i> |
| Lys        | ATC   | 0.93               | 0.95             | 0.92           | 0.93              | 1.09                | 1.06              | 1.01               |
|            | ATT   | 1.18               | 1.22             | 1.24           | 1.21              | 0.99                | 1.02              | 1.08               |
|            | AAA*  | 1.31               | 1.29             | 1.28           | 1.28              | 1.23                | 1.25              | 1.22               |
|            | AAG   | 0.69               | 0.71             | 0.72           | 0.72              | 0.77                | 0.75              | 0.78               |
| Leu        | CTA   | 0.24               | 0.22             | 0.29           | 0.29              | 0.14                | 0.14              | 0.08               |
|            | CTC   | 0.93               | 1.02             | 0.98           | 0.96              | 0.96                | 0.98              | 0.95               |
|            | CTG   | 0.75               | 0.67             | 0.71           | 0.72              | 0.99                | 0.98              | 1.06               |
|            | CTT   | 1.01               | 1.00             | 1.06           | 1.04              | 0.55                | 0.52              | 0.64               |
| Asn        | TTA   | 0.88               | 0.89             | 0.87           | 0.88              | 0.46                | 0.46              | 0.50               |
|            | TTG*  | 2.19               | 2.21             | 2.09           | 2.11              | 2.90                | 2.91              | 2.76               |
|            | AAC   | 0.54               | 0.57             | 0.55           | 0.56              | 0.64                | 0.66              | 0.56               |
|            | AAT*  | 1.46               | 1.43             | 1.45           | 1.44              | 1.36                | 1.33              | 1.44               |
| Pro        | CCA*  | 1.96               | 1.91             | 1.93           | 1.98              | 2.11                | 2.11              | 1.91               |
|            | CCC*  | 1.12               | 1.11             | 1.14           | 1.11              | 1.16                | 1.12              | 1.24               |
|            | CCG   | 0.27               | 0.31             | 0.27           | 0.25              | 0.21                | 0.24              | 0.31               |
|            | CCT   | 0.65               | 0.68             | 0.67           | 0.67              | 0.53                | 0.53              | 0.54               |
| Gln        | CAA*  | 1.52               | 1.57             | 1.51           | 1.51              | 1.51                | 1.47              | 1.68               |
|            | CAG   | 0.48               | 0.43             | 0.49           | 0.49              | 0.49                | 0.53              | 0.32               |
| Arg        | AGA*  | 1.73               | 1.87             | 0.82           | 1.81              | 1.95                | 1.90              | 1.80               |
|            | AGG*  | 1.79               | 1.61             | 1.67           | 1.60              | 1.46                | 1.49              | 1.43               |
|            | CGA   | 0.61               | 0.60             | 0.72           | 0.69              | 0.38                | 0.44              | 0.37               |
|            | CGC   | 0.75               | 0.80             | 0.66           | 0.69              | 0.70                | 0.63              | 1.08               |
| Ser        | CGG   | 0.31               | 0.40             | 0.39           | 0.45              | 0.65                | 0.69              | 0.56               |
|            | CGT   | 0.81               | 0.72             | 0.75           | 0.75              | 0.86                | 0.85              | 0.77               |
|            | AGC*  | 1.92               | 1.87             | 1.91           | 1.93              | 2.12                | 2.15              | 2.14               |
|            | AGT*  | 1.42               | 1.34             | 1.28           | 1.26              | 1.36                | 1.31              | 1.29               |
| Thr        | TCA   | 0.17               | 0.16             | 0.24           | 0.24              | 0.08                | 0.08              | 0.13               |
|            | TCC*  | 1.42               | 1.53             | 1.52           | 1.52              | 1.34                | 1.36              | 1.48               |
|            | TCG   | 0.32               | 0.40             | 0.41           | 0.41              | 0.46                | 0.44              | 0.36               |
|            | TCT   | 0.75               | 0.71             | 0.63           | 0.63              | 0.64                | 0.67              | 0.60               |
| Val        | ACA   | 0.48               | 0.39             | 0.40           | 0.42              | 0.28                | 0.29              | 0.30               |
|            | ACC*  | 1.67               | 1.70             | 1.73           | 1.72              | 2.00                | 2.03              | 2.03               |
|            | ACG   | 0.66               | 0.68             | 0.70           | 0.70              | 0.83                | 0.76              | 0.80               |
|            | ACT   | 1.19               | 1.24             | 1.17           | 1.16              | 0.88                | 0.92              | 0.87               |
| Tyr        | GTA   | 1.19               | 1.21             | 1.18           | 1.13              | 0.78                | 0.80              | 0.81               |
|            | GTC   | 0.63               | 0.64             | 0.67           | 0.67              | 0.89                | 0.86              | 0.97               |
|            | GTG   | 0.88               | 0.87             | 0.99           | 1.03              | 0.91                | 0.94              | 0.94               |
|            | GTT*  | 1.30               | 1.28             | 1.16           | 1.17              | 1.42                | 1.41              | 1.29               |
| Ala        | TAC   | 1.08               | 1.10             | 0.97           | 1.03              | 1.25                | 1.25              | 1.18               |
|            | GCA   | 0.67               | 0.70             | 0.63           | 0.63              | 0.52                | 0.51              | 0.42               |

注: \* 表示 RSCU 值大于 1 的密码子。  
Note: \* . Means the RSCU value is more than 1.

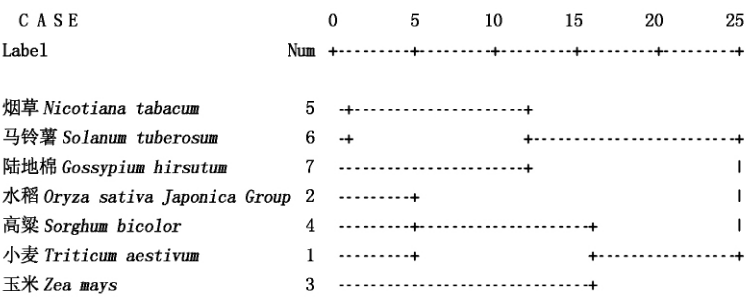


图 1 基于叶绿体基因密码子偏好性的聚类分析

Fig. 1 Cluster tree based on the RSCU values of chloroplastic genes from 7 plant species

### 2.3 叶绿体基因密码子偏好性的聚类分析

在对基因密码子的使用概率进行聚类分析过程中,将每个基因作为一个对象,将密码子的 RSCU 值(表 4)作为变量,利用 SPSS 17.0 进行聚类分析,结果见图 1。从图 1 可以看出,利用叶绿体蛋白编码基因的 RSCU 值得出的聚类图较好地反映了植物间的进化关系,禾本科的 4 种作物与双子叶的 3 种作物分成了两类;而在双子叶的 3 种作物中,烟草与马铃薯同属茄科,所以又单独聚成了一个子类,在禾本科的 4 种作物中,水稻、高粱与小麦 3 种作物聚成了一个子类,而玉米单独成为一个子类。

### 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<sup>[17]</sup>,还与基因编码的蛋白结构和功能及基因表达有密切的联系,受进化过程中多种因素的影响<sup>[18]</sup>。

在植物叶绿体基因工程研究中,首先要分析受体植物的叶绿体蛋白编码基因的特点,进而对需转化的基因进行修饰,以利于转化基因的稳定表达。本研究选取了 6 个基因参与分析,虽然这 6 个基因的总长度仅占总叶绿体基因中蛋白编码序列的 12.8%,但它们都是序列较长的基因(短序列易变异进而会影响到分析结果,去除短基因影响是生物信息学研究中的常见做法),并且也是表达量较大的基因,例如 *psbA* 基因就是高表达基因之一,其启动子序列也是叶绿体基因工程中常用的启动子。

从碱基组成上看,这 7 种作物之间差异不大,都是 A 含量最高,在 30% 左右, C 含量最低,在 21% 左右;但在密码子第 3 位的碱基上却差异较大,其中,双子叶的 3 种作物以 C 最多,其次是 T,而在禾本科的 4 种作物则是以 T 最多, C 次之,表明密码子的偏好性与物种的亲缘关系之间有一定的相关性。RSCU 值也是衡量密码子偏好性的重要指标,禾本科作物有 26 个密码子具有偏好性,而双子叶作物有 25 个密码子具有偏好性,但二者之间有 22 个相同的偏好性密码子。若对这 7 种作物进行叶绿体基因工程研究,可参考表 4 对异源表达的基因进行修饰。

在基于基因密码子用法的聚类分析中,有可能产生与传统分类不符的现象,这在前人的研究中经常出现<sup>[15]</sup>。根据本研究结果看,导致该现象的可能原因是样本数量太少,也有可能是由于核基因易受外界花粉影响容易产生变异的缘故。而本研究由于同时选用了 6 个叶绿体蛋白编码基因参与分析,因

而,利用 RSCU 值进行聚类的结果能够较好地反映物种之间的进化关系。在植物的遗传体系中,叶绿体基因与核基因不同,叶绿体基因属于细胞质遗传,在进化过程中受外界影响较小,可能较适合作为各种系统发育分析方法的重要补充,用于物种进化关系的研究。

### 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 *et al.* 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 *et al.* Chloroplast transformation in oilseed rape [J]. Transgenic Res 2003, 12(1): 111-114.
- [5] Zhang X H, Portis A R J, Wildholm J M. Plastid transformation of soybean suspension cultures [J]. J Plant Biotechnol 2001, 3(1): 39-44.
- [6] Sikdar S R, Serino G, Chaudhuri S *et al.* 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] 徐利娟, 钟金城, 杨万远, 等. 牦牛 mtDNA 编码蛋白质的基因密码子偏好性研究及聚类分析 [J]. 西北农业学报, 2010, 19(6): 13-17.
- [9] 强承魁, 王胜永, 凤舞剑, 等. 不同昆虫泛素延伸蛋白基因的生物信息学分析 [J]. 河南农业科学, 2011, 40(3): 98-101.
- [10] 华慧颖, 王芳, 常重杰, 等. 鲤鱼 *Sox9* 基因 HMG 保守区的克隆及序列分析 [J]. 河南农业科学, 2011, 40(7): 153-157.
- [11] Zhou M, Tong C F, Shi J S. A preliminary analysis of synonymous codon usage in poplar species [J]. Journal of Plant Physiology and Molecular Biology, 2007, 33(4): 285-293.
- [12] 刘伟兰, 李祥龙, 周荣艳, 等. 不同物种 *TYRP2* 基因完整编码区生物信息学分析 [J]. 河南农业科学, 2011, 40(10): 150-154.
- [13] 赵薇, 陈必链, 王明兹, 等. 藻类  $\Delta 6$  脂肪酸脱饱和酶基因密码子偏好性分析 [J]. 热带海洋学报, 2010, 29(3): 126-134.
- [14] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. J Mol Evol, 1986, 24(1-2): 28-38.
- [15] 刘汉梅, 赵耀, 顾勇, 等. 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.
- [16] 周晓明, 赵慧芳, 陆嘉良, 等. 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.
- [17] Wong G K S, Wang J, Tao L *et al.* Compositional gradients in gramineae genes [J]. Genome Res 2002, 12(6): 851-856.
- [18] Hélène C, Frédérique L, Michel C *et al.* Codon usage and gene function are related in sequences of *Arabidopsis thaliana* [J]. Gene, 1998, 209(1-2): GC1-GC38.