RESEARCH PAPER

Comparative analysis of synonymous codon usage patterns among six chloroplast genomes of wide taxonomic range species

Babak Latif[1], Rasoul Mohammadi Najafabadi[1], Mokhtar Jalali Javaran[1*], Hamid Rajabi Memari[2], Houshang Alizadeh[3]

[1]*Department of Plant Breeding and Biotechnology, College of Agriculture, Tarbiat Modares University, Tehran, Iran*

[2]*Department of Agronomy and Plant Breeding, College of Agriculture, Shahid Chamran University, Ahwaz, Iran*

[3]*Department of Agronomy and Plant Breeding, College of Agriculture and Natural Resources, University of Tehran, Tehran, Iran*

Abstract

It is important to understand general patterns of codon usage in chloroplast genomes. In the present study, codon usage patterns in chloroplast genomes of wide taxonomic range organisms namely *Adiantum capillus-veneris*, *Arabidopsis thaliana*, *Chlorella vulgaris*, *Glycine max*, *Pinus koraiensis,* and *Triticum aestivum* were analyzed using multivariate statistical analysis. The results indicated that the GC content at the third position of codons was low in the six investigated chloroplast genomes. There was a significant positive correlation between gene expression level, assessed by CAI values, and the position of genes along axis 1. On the other hand, there was a negative correlation between CAI and GC3s in the chloroplast genomes, which indicated that the gene expression level was the main factor in shaping codon usage, whereas mutational bias played a minor role in determining codon usage bias. Furthermore, gene length played a crucial role in codon usage bias of chloroplast genome in *C. vulgaris* and *G. max* and hydrophobicity of each protein played an important role in *A. thaliana, C. vulgaris,* and *P. koraiensis*. In addition, optimal codons of the chloroplast genomes were defined which may be usable to introduce point mutation, design of degenerate primers, investigate molecular evolution, and heterologous gene expression in the chloroplast genome.

*Corresponding Author: Mokhtar Jalali Javaran ✉ m_jalali@modares.ac.ir

Introduction

The genetic code determines 61 codons or triplets which codes 20 different amino acids. All amino acids are encoded by two to six different codons excluding methionine and tryptophan. Different codons that encode the same amino acid are known as synonymous codons. Even though synonymous codons encode the same amino acids, it has been found in a wide range of biological systems, from prokaryotes to eukaryotes, that different synonymous codons are used with different frequencies. This phenomenon has been called codon usage bias. Synonymous codon usage patterns vary not only among different organisms but also among genes in the same genome. Many previous analyses have suggested that synonymous codon usage bias is correlated with many factors, but explanations for the existence of codon bias fall into two general classes (Shields and Sharp, 1987; Shields *et al.*, 1988; Bulmer, 1991). Both mutational bias and selective constraints seem to affect codon usage bias. In unicellular eukaryotes such as *Saccharomyces cerevisiae* and some prokaryotes such as *Escherichia coli*, codon usage is thought to be determined by the balance between natural selection and compositional constraints (Sharp and Li, 1986; Sharp *et al.*, 1993). In multicellular eukaryotes, such as *Caenorhabditis elegans* and *Drosophila melanogaster*, codon usage bias might be mainly caused by selection for translational efficiency (Shields *et al.*, 1988; Stenico *et al.*, 1994). However, in human and some prokaryotes with extremely high AT or GC contents, mutational bias is the most important factor accounting for the variation in codon usage (Sharp *et al.*, 1993; Karlin and Mrázek, 1996).

Codon usage pattern studies could provide information about the biased usage of synonymous codons and selecting appropriate host expression systems to improve the expression of target genes *in vivo* and *in vitro* (Powell and Moriyama, 1997; Zhao *et al.*, 2007b). In addition, codon usage profiles provide data to improve gene recognition algorithms that detect open reading frames (ORFs) in uncharacterized genomic DNA and improve the accuracy of gene prediction from genomic sequences (Rogic *et al.*, 2001), as well as protein functional classification (Lin *et al.*, 2002).

Chloroplasts have their own genome and a specific machinery to decode its genetic information. The chloroplast has long been studied in plant molecular evolution because of its relatively stable, small size, and high copy number genome and plays a crucial role in the process of photosynthesis. The translation machinery in chloroplast is recognized to be structurally similar to those in prokaryotes, suggesting that the translation mechanisms and patterns of codon usage in this organelle might be similar to that in *E. coli* (Sugiura, 1992).

For the codon usage pattern in the chloroplast, Morton (Morton, 1999) suggested that the asymmetry of two strands of the chloroplast genome from *Euglena gracilis* played an essential role contributing to codon usage. It was considered that mutational pressures played some roles in shaping codon usage of grass chloroplast genomes (Morton, 2003), whereas in certain chloroplast genes, the codon usage was influenced by selection (Morton, 1998).

Here we are interested in understanding the patterns of synonymous codon usage in plant chloroplast genomes across a wide taxonomic range. Awareness of the codon usage patterns in different plant chloroplast genomes can improve the understanding of the mechanisms governing codon distribution and factors shaping codon usage bias in plant chloroplast genomes. In this study, analysis of codon usage bias in six different organisms including *Adiantum capillus-veneris*, *Arabidopsis thaliana*, *Chlorella vulgaris*, *Glycine max*, *Pinus koraiensis*, *Triticum aestivum* was reported using multivariate statistical analysis methods and correlation analysis. In addition, optimal codons were also determined.

Materials and Methods

Sequence data

Complete chloroplast genome sequences of six organisms were obtained from the GenBank (http://www.ncbi.nlm.nih.gov). To minimize sampling errors (Wright, 1990), coding sequences (CDSs) which were more than 100 nucleotides in length and had correct initial and termination codons were included in codon usage calculations. Totally, 59 CDSs for *A. capillus-veneris*, 59 CDSs for *A. thaliana*, 72 CDSs for *C. vulgaris*, 57 CDSs for *G. max*, 46 CDSs for *P. koraiensis*, and 55 CDSs for *T. aestivum* were analyzed in this study (Supplementary data).

Indices of codon usage bias

Relative synonymous codon usage (RSCU) values were computed for the 59 informative codons, excluding methionine and tryptophan, to normalize codon usage within datasets of differing amino acid compositions (Sharp and Li, 1986). RSCU values were estimated as the ratio of the observed codon usage by that expected when all codons for the same amino acid are used equally. The effective number of codons (ENC) is a general measure of bias from equal codon usage in a gene, yielding values ranging 20 from the strongest bias (where only one codon is used for each amino acid) to 61 for no bias (where all synonymous codons are used equally) (Wright, 1990). The codon adaptation index (CAI) was used to estimate the magnitude of bias toward codons that were known to be preferred in highly expressed genes (Sharp and Li, 1987). The maximum possible value for CAI is 1, when only optimal codons are used, and a higher value means a likely stronger codon usage bias and potentially higher expression level. In this study, a set of sequences, including the genes coding for ribosomal proteins, was used to calculate CAI values (Peixoto *et al.*, 2003; Gupta *et al.*, 2004). GC content of entire gene (GCall), at the first, second and third position were calculated with PERL script. Hydrophobicity scale is the general average hydrophobicity (GRAVY) score for the conceptually translated gene. It was calculated as the arithmetic mean of hydrophobic indices of each amino acid.

Corresponding analysis

Corresponding analysis (COA) has been successfully used to investigate the codon usage patterns (Liu *et al.*, 2004; Wang and Hickey, 2007). Since there is a total of 59 synonymous codons (61 sense codons, except for the unique methionine and tryptophan codons), this analysis partitions the variation along 59-dimensional hyperspace (orthogonal axes) according to their usage of 59 informative codons. The first axis accounts for most of the variation of codon usage, with subsequent axes explaining a diminishing amount of the variance.

Statistical analysis

The correlation between codon usage indices and the first two axes of COA analysis was based on the nonparametric Spearman's rank correlation. In addition, to identify the putative preferred codons, we used 10% of the total genes at two extremes of the principal axis (arbitrary axis orientation) which were regarded as the high and low expression gene datasets. A Chi-square test was used to examine the significance of codon usage difference between two datasets. Codons which frequencies of usage were significantly higher ($P<0.05$) in highly expressed genes than that in genes with low expression level, would be defined as the optimal codons (Liu, 2006).

Analysis tools

The indices of codon usage were obtained using the CodonW v.1.4 program (http://codonw.sourceforge.net). SPSS v.20.0 software (http://www.spss.com/statistics/) was used for statistical analysis.

Results

Nucleotide content of the chloroplast genomes

The average GC percent for the first, second, and third positions of codons in the six investigated chloroplast genomes are represented in table 1. The results indicate that the GC content at the third position (GC3) is low, which is consistent with the total low GC content in the chloroplast genomes. In all studied chloroplast genomes, GC3s are the lowest among three positions of codons and variations in GC content are greatest at the third position, followed by the first and second positions.

Table 1. Summary of codon usage in the six investigated chloroplast genomes.

The total number of used coding sequences (CDS), total GC content of genome (GC), GC content at the first, second, and third position of codons (GC1, GC2, and GC3), effective number of codons (ENC), total length of coding sequences in base pair (L_CDS) are shown.

Organism	Classifi-cation	GenBank accession	CDS	GC	GC1	GC2	GC3	ENC	L_CDS
Adiantum capillus-veneris	Ferns	AY178864	87	0.42	0.5	0.4	0.35	55.11	24892
Arabidopsis thaliana	Dicots	AP000423	87	0.37	0.47	0.37	0.26	49.19	23692
Chlorella vulgaris	Algae	AB001684	174	0.36	0.5	0.38	0.19	43.01	24313
Glycine max	Dicots	DQ317523	83	0.36	0.46	0.37	0.25	48.22	24347
Pinus koraiensis	White pine	AY228468	164	0.4	0.52	0.4	0.27	49.63	15694
Triticum aestivum	Monocots	AB042240	83	0.39	0.5	0.39	0.27	49.81	18241

Nc-plot

We investigate the codon usage variation among the genes by the plot of ENC against GC3s (Nc-plot). Wright (Wright, 1990) suggests that the Nc-plot could be used to display codon usage patterns. He argues that genes, whose codon usage is only constrained by a GC mutation bias, will lie on or just below the curve of the predicted values, but genes lying well below the expected curve, are influenced by some factors other than compositional constraints. Fig. 1 indicates the distribution plot of ENC against GC3s for six chloroplast genomes. The continuous curve is the expected position of genes which codon usage is only determined by the GC3s. The six investigated chloroplast genomes show the same patterns of Nc-plot. CDSs of chloroplast genomes appear to cluster around the expected ENC of 20-40% GC3s. The result indicates that although a few genes lie on the expected curve, a majority of the points is lying well below the continuous line, suggesting that some other factors influence the codon usage variation among the genes, which are independent of compositional constraints.
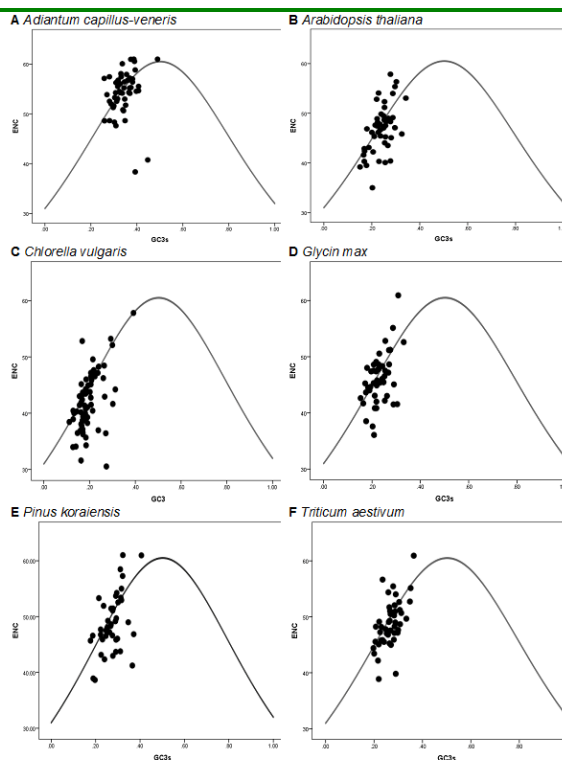

Fig. 1. ENC versus GC3s plot of the six investigated chloroplast genomes (A-F). Continues curve represents the expected ENC value if the codon bias is only due to GC3s.

Correspondence analysis

Correspondence analysis is employed to investigate the variation of synonymous codon usage among genes in chloroplast genomes. The first axis account for 13.05, 10.95, 13.96, 10.09, 21.6, and 9.06 percent of total codon usage variation among genes in *A. capillus-veneris*, *A. thaliana*, *C. vulgaris*, *G. max*, *P. koraiensis,* and *T. aestivum* respectively, and each

subsequent axes explain a diminishing amount of the variance. The position of each gene on the first two COA axes is displayed in Fig. 2, for the six chloroplast genomes. The origin of coordinates indicates the average RSCU for all genes, regarding to the first two axes. The distance between genes on the plot is a reflection of their dissimilarity in RSCU. As the first axis explain only a partial amount of variation of codon usage among genes in the chloroplast genomes, it is postulated that there are several major factors in shaping codon usage of chloroplast genes.

Effect of gene expression level on synonymous codon usage bias

Codon adaptation index (CAI) has been extensively used to estimate the expresses of genes and is now considered as a measure of gene expression level (Naya *et al.*, 2001; Gupta *et al.*, 2004). The reference sequences used for computing CAI values in the present study are those genes coding ribosomal proteins. To assess the effect of gene expression level on codon usage bias, we calculate the correlation coefficients between CAI and the position of genes along the first axis and GC3s for the six investigated chloroplast genomes. According to results, there are significant positive correlations between gene expression level, assessed by CAI values, and the position of genes along axis 1 in *A. capillus-veneris* (r=0.45**), *A. thaliana* (r=0.53**), *C. vulgaris* (r=0.60**), *G. max* (r=0.40**), *P. koraiensis* (r=0.44**), and *T. aestivum* (r=0.56**) chloroplast genomes; at the same time, there are significant negative correlations between CAI values and GC3s in *A. capillus-veneris* (r=-0.70**), *A. thaliana* (r=-0.92**), *C. vulgaris* (r=-0.93**), *G. max* (r=-0.94**), *P. koraiensis* (r=-0.89**), and *T. aestivum* (r=-0.56**) chloroplast genomes. The results suggest that the gene expression level is the main factor in shaping codon usage, whereas mutational bias plays a minor role in determining the variation of codon usage in the chloroplast genome.

Effect of the other factors on codon usage

It has been considered that CDS length and hydrophobicity of encoded protein are the main factors in shaping codon usage variation of some species (Liu *et al.*, 2004; Liu, 2006; Zhao *et al.*, 2007a; Liu *et al.*, 2010). In the present study, the results of correlation analysis between CDSs length and gene positions on axis 1, GC3s, CAI, and ENC values indicate that only the correlation coefficients between CDSs and axis 1 in *C. vulgaris* (r=0.39**) and *G. max* (r=0.29**) chloroplast genomes are significantly correlated. The findings show that more bias genes with longer CDSs length in these two species are distributed on the left side of the axis 1. These results suggest that the CDSs length seems to play an important role in shaping codon usage in *C. vulgaris* and *G. max* chloroplast genomes.

In addition, the correlation between the hydrophobicity of each protein and synonymous codon usage is investigated in the present study. The results of the correlation analysis between the hydrophobicity of each protein and axis 1, CAI values, CG3s, and ENC values indicate that three correlation coefficients between hydrophobicity and axis 1 in *A. thaliana* (r=0.39**), *C. vulgaris* (r=0.39**), and *P. koraiensis* (r=0.43**) chloroplast genomes are significantly correlated. The results indicate that in these three species, genes encoding more hydrophobic protein show a strong codon bias.

Translational optimal codons

Table 2 shows the putative optimal codons in the six investigated chloroplast genomes. According to the results, 11 codons in *A. capillus-veneris*, 6 codons in *A. thaliana*, 15 codons in *C. vulgaris*, 7 codons in *G. max*, 7 codons in *P. koraiensis,* and 8 codons in *T. aestivum* chloroplast genome are identified by χ² test. Furthermore, in all chloroplast genomes, preferred codons tend to have T or C in the third position of codons, except A in *C. vulgaris*.
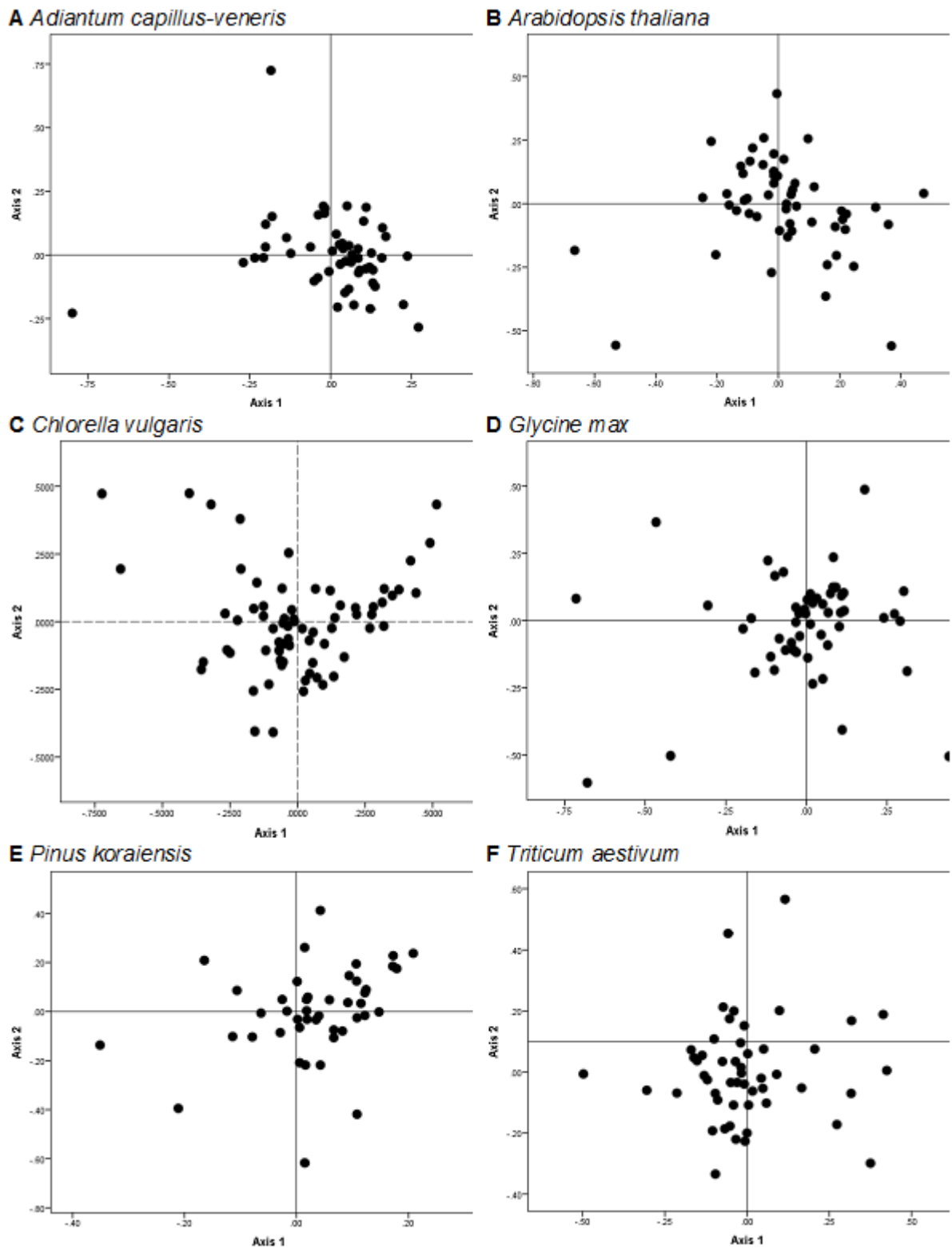
Fig. 2. Correspondence analysis of RSCU for the six referred chloroplast genomes (A-F). Each point on the plot corresponds to the coordinates on the first and second principal axes produced by the correspondence analysis.

Table 2. Putative optimal codons defined by Chi-square contingency test. Those codons that occur significantly more often (*P*<0.01) in the highly biased dataset relative to the low biased dataset are considered optimal, and are represented with an asterisk (*). Species name: AC, *Adiantum capillus-veneris*; AT, *Arabidopsis thaliana*; CV, *Chlorella vulgaris*; GM, *Glycine max*; PK, *Pinus koraiensis*; TA, *Triticum aestivum*.

A.A.	Codon	AC	AT	CV	GM	PK	TA
Phe	TTT	-	-	-	-	-	-
	TTC	*	-	*	*	-	-
Leu	TTA	-	-	*	-	-	-
	TTG	-	-	-	-	-	-
	CTT	-	-	-	-	-	-
	CTC	-	-	-	-	-	-
	CTA	-	-	-	-	-	-
	CTG	-	-	-	-	-	-
Ile	ATT	-	*	-	*	-	-
	ATC	*	-	*	-	*	-
	ATA	-	-	-	-	-	-
Met	ATG	-	-	-	-	-	-
Val	GTT	-	-	*	-	-	-
	GTC	-	-	-	-	-	-
	GTA	-	-	-	-	-	-
	GTG	-	-	-	-	-	-

A.A.	Codon	AC	AT	CV	GM	PK	TA
Ser	TCT	*	-	-	*	*	*
	TCC	-	-	-	-	-	-
	TCA	-	*	*	-	-	-
	TCG	-	-	-	-	-	-
Pro	CCT	-	*	-	-	-	*
	CCC	-	-	-	-	-	-
	CCA	-	-	*	-	-	-
	CCG	-	-	-	-	-	-
Thr	ACT	*	-	-	*	-	-
	ACC	-	-	-	-	*	-
	ACA	-	-	-	-	-	-
	ACG	-	-	-	-	-	-
Ala	GCT	-	-	*	-	-	-
	GCC	-	-	-	-	*	-
	GCA	-	-	-	-	-	-
	GCG	-	-	-	-	-	-

A.A.	Codon	AC	AT	CV	GM	PK	TA
Tyr	TAT	-	-	-	-	-	-
	TAC	*	*	*	-	-	*
TER	TAA	-	-	-	-	-	-
	TAG	-	-	-	-	-	-
His	CAT	-	-	-	-	-	-
	CAC	-	-	-	-	-	-
Gln	CAA	-	-	*	-	-	-
	CAG	-	-	-	-	-	-
Asn	AAT	-	-	-	-	-	-
	AAC	*	-	*	*	-	*
Lys	AAA	-	-	*	-	-	*
	AAG	-	-	-	-	-	-
Asp	GAT	-	-	*	-	-	-
	GAC	*	-	-	-	-	-
Glu	GAA	-	-	*	-	-	-
	GAG	-	-	-	-	-	-

A.A.	Codon	AC	AT	CV	GM	PK	TA
Cys	TGT	-	-	*	-	-	-
	TGC	*	-	-	-	-	-
TER	TGA	-	-	-	-	-	-
Trp	TGG	-	-	-	-	-	-
Arg	CGT	*	*	-	-	*	*
	CGC	*	*	-	-	-	-
	CGA	-	-	-	-	-	-
	CGG	-	-	-	-	-	-
Ser	AGT	-	-	-	-	*	*
	AGC	-	-	-	-	-	-
Arg	AGA	-	-	-	-	-	-
	AGG	-	-	-	-	-	-
Gly	GGT	*	*	*	*	*	*
	GGC	-	-	-	*	-	-
	GGA	-	-	-	-	-	-
	GGG	-	-	-	-	-	-

Discussion

Many previous studies have been widely investigated the codon usage bias in eukaryotes and prokaryotes and it has been found that the codon bias is affected by many factors such as GC composition, mutational bias (Karlin and Mrázek, 1996; Hou and Yang, 2003), gene expression level, natural selection (Sharp and Li, 1986), gene length (Moriyama and Powell, 1998; Zhang *et al.*, 2012), tRNA levels (Percudani *et al.*, 1997), mRNA secondary structure (Gu *et al.*, 2004), codon-anticodon interaction (Shi *et al.*, 2001) and, protein hydropathy (Romero *et al.*, 2000). The codon usage of organelle genomes differs from their host cells' in both their directional and strength of codon usage bias (Pfitzinger *et al.*, 1987). However, mutational bias and natural selection for translational optimization are more important factors that have been widely proposed to account for both interspecies and intra-genomic codon usage variation (Shields and Sharp, 1987; Shields *et al.*, 1988; Bulmer, 1991). In the present study, the results show that the location of genes on the primary axis has a strong negative correlation with GC and GC3s, and positively correlated with CAI values indicating that the nucleotide composition and gene expression level are the major factors shaping codon usage bias in the chloroplast

genomes. It has been found that natural selection has a large contribution to codon usage bias of O. sativa, Zea mays, and T. aestivum chloroplast genes and all of them have A and T ending codons as their optimal codons (Kawabe and Miyashita, 2003; Liu and Xue, 2005). In this study, the preferred codons in the six studied chloroplast genomes tend to have T or C in the third position except A in C. vulgaris. The findings suggest that the codon usage patterns of the chloroplast genomes are conserved in GC content and affected by natural selection acting at the level of mRNA translation. Codon bias contributes to the translational accuracy and efficiency of protein expression, thus highly expressed genes tend to use optimal codons. In addition, the Nc-plot shows that although a small number of genes lying on the expected curve, a majority of points with low ENC values are lying below the continuous curve in all investigated chloroplast genomes. COA of RSCU in the chloroplast genomes yields a first axis that explains only a partial amount of codon bias variation among the genes. So we conclude that not only the natural selection, but also the other factors may be involved in determining the variation of codon usage in the chloroplast genomes. Apart from the mutation bias and natural selection, the gene length plays a crucial role in codon bias of chloroplast genomes in C. vulgaris and G. max and hydrophobicity of each protein has an important role in A. thaliana, C. vulgaris, and P. koraiensis. It was reported that codon usage bias was influenced by gene length in O. sativa, Deinococcus raimondii, S. cerevisiae, D. melanogaster, and E. coli (Moriyama and Powell, 1998; Liu et al., 2004; Liu, 2006), but the opposite results were found in adenoviruses (Das et al., 2006). That evidence suggests that different genomes had different gene length and there were no universal rules regarding the relationship between codon usage and gene length in all genomes. In this study, CDS length in C. vulgaris and G. max chloroplast genomes are closely related

to position of genes in axis 1 and CAI values. It suggests that the longer genes have a higher expression level and higher codon usage bias in those chloroplast genomes. It may be due to selecting to avoid missense errors during translation. Since the energy used for producing a protein is proportional to its length, selection in favor of optimal codons which increase translation efficiency and accuracy should be stronger in longer genes, thus longer genes should therefore have a higher codon bias. In the present study, codon usage has significantly positive correlation with the hydrophobicity of each protein in A. thaliana, C. vulgaris, and P. koraiensis chloroplast genomes, suggesting that there are no general rules about the relationship between codon usage and hydrophobicity of each protein in all chloroplast genomes. The correlation between hydrophobicity and codon usage in A. thaliana, C. vulgaris, and P. koraiensis chloroplast genomes may be caused by the fact that many of the highly expressed genes are hydrophilic, because they accomplish their function in the aqueous media of the chloroplast.

Codon optimization has been widely accepted to enhance protein expression in heterogeneous systems. In this study, we define the preferred codons of the six chloroplast genomes. The frequency of the optimal codons in a gene appears to be associated with the level of gene expression and may be valuable indicator in the case of genes which expression level are unknown, and also will be used to introduce point mutation, design of degenerate primers, investigate molecular evolution, and heterologous gene expression in the chloroplast genome.

References

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129**, 897-907.

Das S and Paul S, Dutta C. 2006. Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus research **117**, 227-236.

Gu W, Zhou T, Ma J, Sun X and Lu Z. 2004. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. BioSystems **73**, 89-97.

Gupta S, Bhattacharyya T and Ghosh T. 2004. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. Journal of biomolecular structure & dynamics **21**, 527-536.

Hou ZC and Yang N. 2003. Factors affecting codon usage in *Yersinia pestis*. Sheng wu hua xue yu sheng wu wu li xue bao Acta biochimica et biophysica Sinica **35**, 580-586.

Karlin S and Mrázek J. 1996. What drives codon choices in human genes? Journal of Molecular Biology **262**, 459-472.

Kawabe A and Miyashita NT. 2003. Patterns of codon usage bias in three dicot and four monocot plant species. Genes & genetic systems **78**, 343-352.

Lin K, Kuang Y, Joseph JS and Kolatkar PR. 2002. Conserved codon composition of ribosomal protein coding genes in *Escherichia coli*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae*: lessons from supervised machine learning in functional genomics. Nucleic Acids Research **30**, 2599-2607.

Liu H, He R, Zhang H, Huang Y, Tian M and Zhang J. 2010. Analysis of synonymous codon usage in *Zea mays*. Molecular Biology Reports **37**, 677-684.

Liu Q. 2006. Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*. Biosystems **85**, 99-106.

Liu Q, Feng Y, Zhao X, Dong H and Xue Q. 2004. Synonymous codon usage bias in *Oryza sativa*. Plant Science **167**, 101-105.

Liu Q and Xue Q. 2005. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. Journal of genetics **84**, 55-62.

Moriyama EN and Powell JR. 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Research **26**, 3188-3193.

Morton BR. 1998. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. Journal of molecular evolution **46**, 449-459.

Morton BR. 1999. Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. Proceedings of the National Academy of Sciences, USA **96**, 5123-5128.

Morton BR. 2003. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. Journal of molecular evolution **56**, 616-629.

Naya H, Romero H, Carels N, Zavala A and Musto H. 2001. Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. FEBS letters **501**, 127-130.

Peixoto L, Zavala A, Romero H and Musto H. 2003. The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*. Gene **320**, 109-116.

Percudani R, Pavesi A and Ottonello S. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. Journal of Molecular Biology **268**, 322-330.

Pfitzinger H, Guillemaut P, Weil JH and Pillay D. 1987. Adjustment of the tRNA population to the codon usage in chloroplasts. Nucleic Acids Research **15**, 1377-1386.

Powell JR and Moriyama EN. 1997. Evolution of codon usage bias in *Drosophila*. Proceedings of the National Academy of Sciences, USA **94**, 7784-7790.

Rogic S, Mackworth AK and Ouellette FBF. 2001. Evaluation of gene-finding programs on mammalian sequences. Genome Research **11**, 817-832.

Romero H, Zavala A and Musto H. 2000. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Research **28**, 2084-2090.

Sharp PM and Li WH. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. Journal of molecular evolution **24**, 28-38.

Sharp PM and Li WH. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research **15**, 1281-1295.

Sharp PM, Stenico M, Peden JF and Lloyd AT. 1993. Codon usage: mutational bias, translational selection, or both? Biochemical Society Transactions **21**, 835-841.

Shi X, Huang J, Liang C, Liu S, Xie J and Liu C. 2001. Is there a close relationship between synonymous codon bias and codon-anticodon binding strength in human genes? Chinese Science Bulletin **46**, 1015-1019.

Shields DC and Sharp PM. 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Research **15**, 8023-8040.

Shields DC, Sharp PM, Higgins DG and Wright F. 1988. " Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Molecular Biology and Evolution **5**, 704-716.

Stenico M, Lloyd AT and Sharp PM. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Research **22**, 2437-2446.

Sugiura M. 1992. The chloroplast genome. Plant Molecular Biology **19**, 149-168.

Wang HC and Hickey D. 2007. Rapid divergence of codon usage patterns within the rice genome. BMC evolutionary biology **7**, S6.

Wright F. 1990. The 'effective number of codons' used in a gene. Gene **87**, 23-29.

Zhang Y, Nie X, Jia X, Zhao C, Biradar SS, Wang L, Du X and Weining S. 2012. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family. Australian Journal of Botany **60**, 461-470.

Zhao S, Zhang Q, Chen Z, Zhao Y and Zhong J. 2007a. The factors shaping synonymous codon usage in the genome of *Burkholderia mallei*. Journal of Genetics and Genomics **34**, 362-372.

Zhao WM, Wang H and *et al.*. 2007b. Codon usage bias in *Chlamydia trachomatis* and the effect of codon modification in the MOMP gene on immune responses to vaccination. Biochemistry and cell biology **85**, 218-226.