

Analysis of synonymous codon usage in *FAD7* genes from different plant species

Q.P. Ma, C. Li, J. Wang, Y. Wang and Z.T. Ding

Tea Research Institute, Qingdao Agricultural University, Qingdao, Shandong, China

Corresponding author: Z.T. Ding
E-mail: dztea@163.com

Genet. Mol. Res. 14 (1): 1414-1422 (2015)
Received January 6, 2014
Accepted March 13, 2014
Published February 13, 2015
DOI <http://dx.doi.org/10.4238/2015.February.13.20>

ABSTRACT. In this study, the codon bias of the *FAD7* genes among 10 different plant species was analyzed to identify general patterns of codon usage in the *FAD7* genes. Our results showed that U-ended or A-ended codons were preferentially used in *FAD7* for dicots, whereas G-ended or C-ended codons were preferentially used in *FAD7* for monocots. An ENC-plot showed that some other factors may influence the codon usage of *FAD7*, except mutation bias in plant species. A correlation analysis between the codon adaptation index and GC or GC3s contents demonstrated that the codon usage bias of the *FAD7* gene in plant species could be influenced by the gene expression level. The cluster analysis of relative synonymous codon usage values and phylogenetic trees of protein sequences for *FAD7* genes confirm that the codon preference of *FAD7* is influenced by genetic relationships. Moreover, *Arabidopsis thaliana* and *Nicotiana tabacum* were predicted to be the most appropriate expression hosts for the *FAD7* genes from dicots, and *Zea mays* may be suitable for the expression of the *FAD7* genes from monocots. Our results provide useful insights into the evolutionary relationships of plant species.

Key words: *FAD7*; ENC-plot; Correlation analysis; Cluster analysis; Codon usage frequency

INTRODUCTION

The genetic code includes 61 codons that represent 20 amino acids, which are each encoded by 2-6 synonymous codons, except Met and Trp. The codon usage bias exists in a wide range of biological systems, including prokaryotes and eukaryotes. Genes from different species or within the same species exhibit various patterns of codon usage bias, and codon usage analysis is crucial for understanding the evolutionary or environmental adaptations of living species (Grantham et al., 1980; Nakamura et al., 1991; Ghosh et al., 2000; Carbone et al., 2003; Mougél et al., 2004; Angellotti et al., 2007; Subramanian, 2008; Bragg et al., 2012; Meganathan et al., 2012).

In previous studies, factors such as mutational bias, selection, intron splicing, gene conversion, protein secondary structures, and DNA replication were shown to be strongly related to synonymous codon usage biases (Birdsell, 2002; Kahali et al., 2007; Warnecke and Hurst, 2007; Drummond and Wilke, 2008; Warnecke et al., 2008). Codon usage is primarily determined by the balance between mutational bias and selection in prokaryotes or unicellular eukaryotes (Gouy and Gautier, 1982; Stenico et al., 1994; Sharp et al., 2005; Bragg et al., 2012). In multicellular eukaryotes, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, codon biases are mainly determined by the selection of translational efficiency (Shields et al., 1988; Stenico et al., 1994; Vicario et al., 2007). In viruses, such as Torque teno sus virus 1 (TTSuV1) or Parvoviridae, the synonymous codon usage patterns of genomes are determined by the interaction between mutation pressure and natural selection (Zhou et al., 2012; Shi et al., 2013; Zhang et al., 2013). In plant species, studies on codon usage in *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* revealed that codon usage bias among 3 plant species was determined by factors such as the base composition of genes, gene expression levels, and complete coding sequence (CDS) lengths (Chiapello et al., 1998; Morton and Wright, 2007; Liu et al., 2004, 2010).

The *FAD7* gene encodes an ω -3 fatty acid desaturase, which catalyses the production of trienoic fatty acids in plant chloroplasts. It plays an important role in the cold resistance of plants. In the present study, sequences of *FAD7* obtained from 10 plant species were used to determine the codon usage bias of *FAD7*. Some indices of codon usage bias were studied in order to identify the codon usage patterns of *FAD7* genes. Our results provide insight into the study of the synonymous codon usage patterns of *FAD7* in plant species, and they also provide a theoretical basis for the transformation of codons and improvements in the expression level of *FAD7*.

MATERIAL AND METHODS

Coding sequence data

Complementary DNA (cDNA) sequences that contain CDSs and protein sequences of *FAD7* for 10 plant species, including *Camellia sinensis* (*CsFAD7*), *Musa basjoo* (*MuFAD7*), *Glycine max* (*GmFAD7*), *Solanum lycopersicum* (*LeFAD7*), *Z. mays* (*ZmFAD7*), *Olea europaea* (*OeFAD7*), *A. thaliana* (*AtFAD7*), *O. sativa* (*OsFAD7*), *Helianthus annuus* (*HaFAD7*), and *Betula pendula* (*BpFAD7*), were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>).

Indices of codon usage bias

Relative synonymous codon usage (RSCU) was calculated for the 59 informative codons (excluding Met, Trp, and the 3 termination codons) for the purpose of exploring codon usage preference in the translation of amino acids. RSCU values close to 1.0 indicated codons for the same amino acids that were used equally, and RSCU values >1.0 indicated a strong bias for the corresponding codons.

The effective number of codons (ENC) was used to measure the degree of codon bias for a single gene. Values of ENC range from 20 (for a gene with extreme bias using only one codon per amino acid) to 61 (for a gene with no bias using synonymous codons equally).

The codon adaptation index (CAI) was used to estimate the extent of bias toward preferred codons in highly expressed genes. The CAI value ranged between 0 and 1.0, and a higher value likely indicates a stronger codon usage bias and higher expression level (Liu et al., 2010).

The frequency of GC, GC3s (at the third synonymously variable coding position), and the single base at the third positions (such as A3, U3, C3, and G3) were calculated for all of the informative codons (excluding Met, Trp, and termination codons). GC3s was used as a good indicator of the extent of base composition bias.

Analysis tools

Indices of codon usage bias were calculated (excluding Met, Trp, and the termination codons) using the CodonW program (<http://codonw.sourceforge.net>). A phylogenetic tree was created from the alignments of the protein sequences of the 10 species using the MEGA 5.0 program. The plot of ENC against GC3s, clustering tree for RSCU, and correlation analysis were carried out by using the multi-analysis SPSS v18.0 software. The RSCU values of codons among the 10 species were clustered using a K-means clustering method (Spearman rank correlation) and implemented in the Gene Cluster 3.0 software. The clusters were viewed by the TreeView program (<http://www.eisenlab.org/eisen/>). The frequencies of the *FAD7* genes were calculated using the CUSP program (<http://mobyline.pasteur.fr/cgi-bin/portal.py?#forms::cusp>); the frequencies of the 5 genomes, including *A. thaliana*, *Nicotiana tabacum*, *Saccharomyces cerevisiae*, *Escherichia coli*, and *Z. mays*, were obtained from the Codon Usage Database (<http://www.kazusa.or.jp/codon/>).

RESULTS

Base composition of *FAD7* among the plant species

The nucleotide content of A3, C3, U3, G3, GC, and GC3s% for *FAD7*-coding region sequences among 10 plant species are listed in Table 1. GC% content fluctuated from 41.7 to 62.1, with a mean value of 47.95 and an SD of 7.47.

U3% was the highest in the dicots when compared to the values of A3, U3, C3, and G3%. The G3, C3, GC, and GC3s% in the monocots such as *MuFAD7*, *ZmFAD7*, and *OsFAD7* were all higher than the values in the dicots, which indicated that the U-ended or A-ended synonymous codons were preferentially used in the *FAD7*-coding region, whereas G-ended or C-ended synonymous codons were preferentially used in *FAD7* for the monocots. The GC% ranged from 41.7 to 62.1, with a mean value of 47.95. The mean value of GC3s% was similar

to that of GC%, and suggested that GC3s% was influenced by GC%.

Table 1. Summary of codon usage in the *FAD7* genes for 10 species.

Species	Genbank accession No.	A3 (%)	U3 (%)	C3 (%)	G3 (%)	GC (%)	GC3s (%)	CAI	ENC
<i>C. sinensis</i>	JX943516	29.3	43.3	25.9	26.6	45.1	40.7	0.207	53.89
<i>M. basjoo</i>	JX911316	20.6	32.1	32.9	36.8	53.9	55.8	0.173	54.78
<i>G. max</i>	NM_001250432	27.6	50.7	18.6	26.5	43.0	35.0	0.191	42.61
<i>S. lycopersicum</i>	NM_001247663	35.5	48.6	16.5	24.6	41.8	31.1	0.169	48.34
<i>Z. mays</i>	NM_001111833	14.1	15.5	51.6	39.1	62.1	75.2	0.241	46.93
<i>O. europaea</i>	HQ889832	33.1	44.9	20.5	26.8	43.6	36.4	0.180	49.53
<i>A. thaliana</i>	NM_111953	35.7	42.7	23.1	23.6	43.2	36.3	0.195	47.04
<i>O. sativa</i>	AB232382	20.7	18.7	42.2	38.8	58.6	66.9	0.196	51.98
<i>H. annuus</i>	AY254858	41.6	43.5	21.1	18.9	41.7	31.2	0.175	51.23
<i>B. pendula</i>	AY135565	31.7	39.3	27.3	24.1	46.5	41.3	0.181	52.60

ENC-plot

The ENC values of the *FAD7* sequences fluctuated from 42.61 to 54.78, with a mean value of 49.89. A plot of ENC against GC3s has been widely used to investigate patterns of codon usage of genes among various species. The distribution plot of ENC against GC3s for *FAD7* among 10 species indicated that *AtFAD7*, *GmFAD7*, and *ZmFAD7* were below the curve, and the remaining genes were above the curve. Moreover, *MuFAD7* and *OsFAD7* were close to the expected curve, and only *OeFAD7* was on the curve. The results are shown in Figure 1.

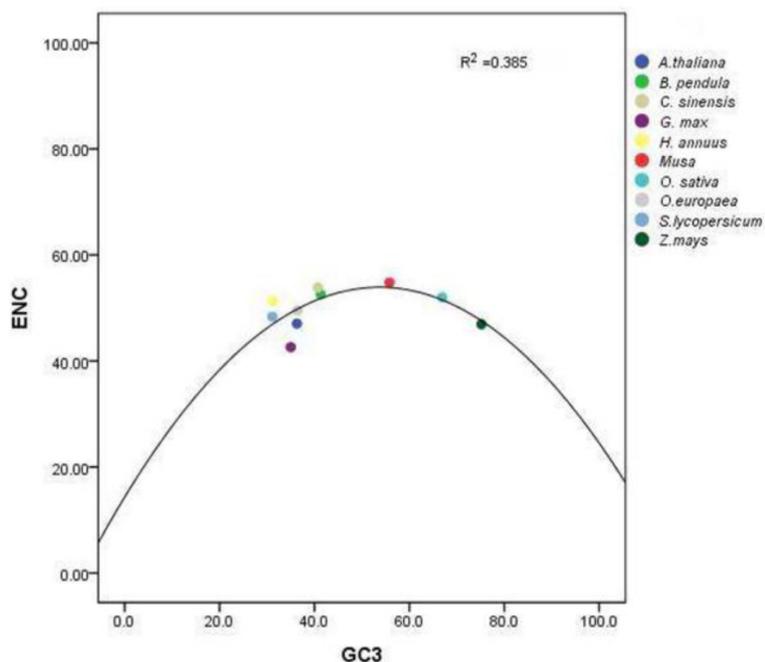


Figure 1. Distribution of the effective number of codons (ENC) and GC content at the synonymous third codon position (GC3) of the *FAD7* genes for 10 species. The curve indicates the expected codon usage if the GC composition that is constrained alone accounts for codon usage bias.

Effect of gene expression level on codon usage bias

The CAI value was extensively used as a measure of gene expression level and has been proven to be the best gene expression theory value (Naya et al., 2001; Gupta et al., 2004). The CAI values of all species were between 0.169 and 0.241, with a mean value of 0.1908 and an SD of 0.0213. To explore whether there was a correlation between the codon usage bias and the gene expression level, the correlation coefficients between CAI and ENC, and GC and GC3s were calculated. We found that there were 2 positive correlations between the CAI value and GC3s or GC contents ($r = 0.665$ and 0.619 , respectively; $P < 0.05$).

Relative synonymous codon usage of *FAD7* genes

To investigate the codon usage pattern of *FAD7* genes, the RSCU values of various codons from the coding sequences of 10 plant species were calculated (excluding Met, Trp, and the termination codons) (Table S1). In order to understand the differences in codon usage of the *FAD7* genes for plant species, a hierarchical clustering analysis of RSCU for the *FAD7* genes from different species was constructed using the K-means clustering method (Figure 2). The results showed that there were several synonymous codons that maintain a strong discrepancy in codon usage among species. For monocots, *ZmFAD7* and *OsFAD7* used G-ended or C-ended codons preferentially, except Gln, which preferred CAA (i.e., A-ended). For most dicots, optimal codons ended with U or A, except UUG for Leu, AAG for Lys, and AGG for Arg.

Thus, by highlighting the significant differences in codon usage between monocots and dicots, a clustering tree (Figure 3) based on the RSCU values of the *FAD7* genes was constructed. The clustering tree showed similar results to those of the phylogenetic tree (Figure 4), which was constructed based on deducing the amino acid sequences of the *FAD7* genes; monocots and dicots belonged to different classes, and only a few species were located in different positions between the 2 trees. For example, *CsFAD7*, *OeFAD7*, and *BpFAD7* were clustered into one group in Figure 3; however, in Figure 4, *CsFAD7* was grouped together with *LeFAD7*, and they were separated from another group that contained *OeFAD7* and *BpFAD7*.

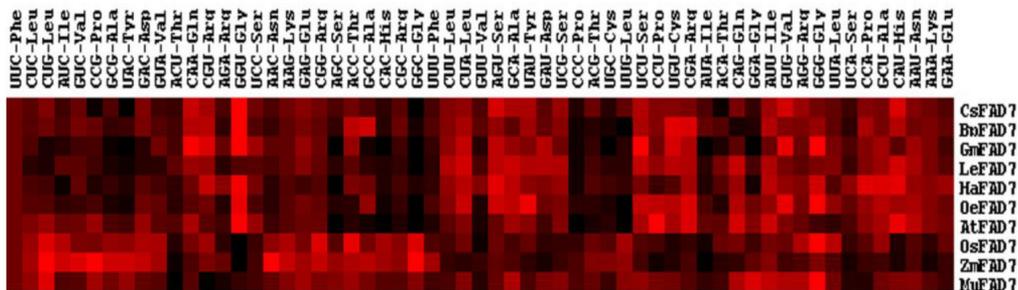


Figure 2. Hierarchical cluster analysis of relative synonymous codon usage (RSCU) for the *FAD7* genes. Each square on the self-organizing map represents the RSCU value of a codon (shown in columns) corresponding to the species (shown in rows). Color coding varies from black to red, with low to high values of the RSCU, respectively.

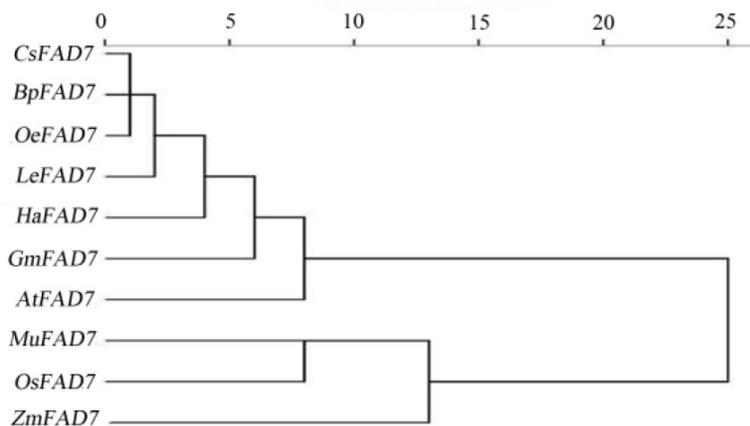


Figure 3. Cluster tree of RSCU for the *FAD7* genes. The ruler represents the coefficient of Euclidean distance.

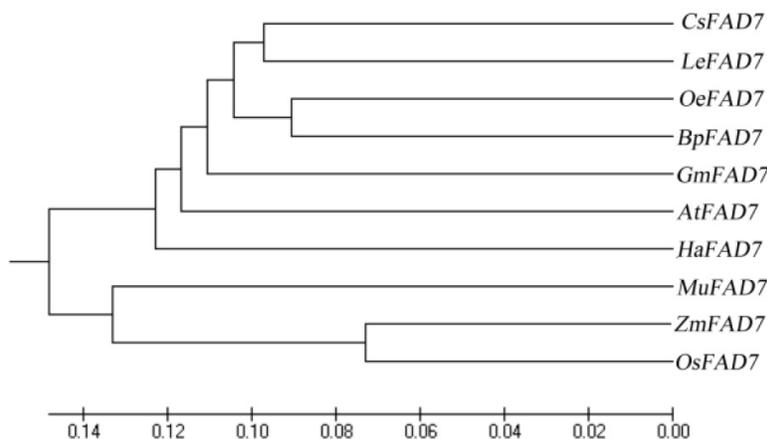


Figure 4. Phylogenetic trees of protein sequences encoded by the *FAD7* genes. The tree was constructed using UPGMA by MEGA 5.0.

Codon usage frequency of *FAD7* genes

In order to choose the appropriate expression hosts for the *FAD7* genes, the correlation of codon usage frequency between the *FAD7* genes and several species genomes were calculated (Table 2). A large correlation coefficient (CC) indicated a notable correlation. Thus, the results showed that discrepancy also existed in monocots and dicots. The frequencies of the *FAD7* genes from the dicots showed a positive correlation ($P < 0.01$) with that of *A. thaliana*, *N. tabacum*, and *S. cerevisiae*; they were most relevant to *N. tabacum* but showed very small CCs with *E. coli* and *Z. mays*. Conversely, frequencies of the *FAD7* genes for monocots showed a significant correlation with *Z. mays*.

Table 2. Correlation of codon usage frequency between *FAD7* genes and several species genomes.

Name	Correlation coefficient				
	<i>A. thaliana</i>	<i>N. tabacum</i>	<i>S. cerevisiae</i>	<i>E. coli</i>	<i>Z. mays</i>
<i>CsFAD7</i>	0.63**	0.65**	0.51**	0.14	0.24*
<i>GmFAD7</i>	0.64**	0.69**	0.56**	0.07	0.17
<i>LeFAD7</i>	0.62**	0.67**	0.54**	0.1	0.08
<i>OeFAD7</i>	0.62**	0.66**	0.53**	0.13	0.18
<i>AtFAD7</i>	0.63**	0.61**	0.53**	0.13	0.2
<i>HaFAD7</i>	0.63**	0.67**	0.62**	0.17	0.05
<i>BpFAD7</i>	0.58**	0.57**	0.42**	0.11	0.23
<i>OsFAD7</i>	0.07	0.04	-0.11	0.24	0.60**
<i>MuFAD7</i>	0.41**	0.37**	0.12	0.2	0.53**
<i>ZmFAD7</i>	-0.03	-0.11	-0.18	0.33**	0.66**

* $P < 0.05$; ** $P < 0.01$.

DISCUSSION

We observed that many different factors were involved in codon usage bias, such as mutational bias, gene expression level, natural selection, gene length, or GC composition (Sueoka and Kawanishi, 2000; Blake et al., 2003; Lü et al., 2005; Behura and Severson, 2013). In the present study, the codon usage of the *FAD7* genes among 10 plant species was analyzed. We found that the codon usage between monocots and dicots differed significantly. U-ended or A-ended synonymous codons were used preferentially in the *FAD7* genes for dicots, whereas monocot species preferentially used G-ended or C-ended synonymous codons. Meanwhile, the GC and GC3s contents in the *FAD7* genes in monocots were higher than the contents in the dicots, which is consistent with the codon usage bias of the genomes in some other monocots and dicots (Wang and Roossinck, 2006; Liu et al., 2010). The results suggested that the *FAD7* gene codon usage bias might primarily be influenced by base compositions in the monocots and dicots of group specificity.

We also observed that directional mutation pressure and natural selection were the 2 major factors manipulating gene translation for interspecific or intragenomic codon usage variation (Lü et al., 2005). It has also been demonstrated that in highly expressed genes, the selection of a codon to improve translation efficiency dominates the codon bias, whereas in lowly expressed genes, the mutation determines codon usage (Shields and Sharp, 1987; Sueoka and Kawanishi, 2000; Mitreva et al., 2006; Mukhopadhyay et al., 2007). Interestingly, all of the *FAD7* genes analyzed in this study were predicted to have low expression levels.

Wright (1990) suggested the use of an ENC-plot as part of a general strategy to investigate patterns of synonymous codon usage. Genes, whose codon preferences were constrained only by a GC mutation bias, will lie on or just below the curve of the predicted values. In this study, the ENC-plot showed that the *FAD7* genes were distributed irregularly; only *AtFAD7*, *GmFAD7*, *ZmFAD7*, and *OeFAD7* lied on or just below the curve, suggesting that the codon usage of these genes was determined by the GC mutation bias. However, some other factors may affect codon usage variation among the remaining *FAD7* genes, which did not rely solely on compositional constraints.

The RSCU clustering analysis for the coding region of the *FAD7* genes from 7 dicots and 3 monocots showed minimal variation in the codon usage among dicots only or monocots only but significant variation between monocots and dicots. This result was consistent with the phylogenetic tree, which was conducted using protein sequences of the *FAD7* genes, suggest-

ing that the codon usage of the *FAD7* genes is influenced by genetic relationships. A further relationship indicates a larger difference in codon preference and vice versa.

Codon optimization has been widely used to enhance protein and DNA expression in heterologous systems (Frelin et al., 2004; Ko et al., 2005; Peng et al., 2006). It is important to select an appropriate expression host for transformation of the *FAD7* genes. In this study, the correlation of codon usage frequency between the *FAD7* genes and several species genomes showed that *A. thaliana* and *N. tabacum* were the most appropriate expression hosts for the *FAD7* genes of dicots, and *Z. mays* may be suitable for the expression of the *FAD7* genes from monocots.

ACKNOWLEDGMENTS

Research supported by a Sci-Tech Development Project of Shandong Province (#2011GNC11010) and the National Science and Technology Support Program (#2011BAD01B01).

[Supplementary material](#)

REFERENCES

- Angellotti MC, Bhuiyan SB, Chen G and Wan XF (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* 35: 132-136.
- Behura SK and Severson DW (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev. Camb. Philos. Soc.* 88: 49-61.
- Birdsell JA (2002). Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* 19: 1181-1197.
- Blake WJ, Kaern M, Cantor CR and Collins JJ (2003). Noise in eukaryotic gene expression. *Nature* 422: 633-638.
- Bragg JG, Quigg A, Raven JA and Wagner A (2012). Protein elemental sparing and codon usage bias are correlated among bacteria. *Mol. Ecol.* 21: 2480-2487
- Carbone A, Zinovyev A and Képès F (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 19: 2005-2015.
- Chiappello H, Lisacek F, Caboche M and Hénaut A (1998). Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* 209: GC1-GC38.
- Drummond DA and Wilke CO (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134: 341-352.
- Frelin L, Ahlén G, Alheim M, Weiland O, et al. (2004). Codon optimization and mRNA amplification effectively enhances the immunogenicity of the hepatitis C virus nonstructural 3/4 A gene. *Gene Ther.* 11: 522-533.
- Ghosh TC, Gupta SK and Majumdar S (2000). Studies on codon usage in *Entamoeba histolytica*. *Int. J. Parasitol.* 30: 715-722.
- Gouy M and Gautier C (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10: 7055-7074.
- Grantham R, Gautier C and Gouy M (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* 8: 1893-1912.
- Gupta SK, Bhattacharyya TK and Ghosh TC (2004). Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J. Biomol. Struct. Dyn.* 21: 527-536.
- Kahali B, Basak S and Ghosh TC (2007). Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: A new insight from protein secondary structure analysis. *Biochem. Biophys. Res. Commun.* 354: 693-699.
- Ko HJ, Ko SY, Kim YJ, Lee EG, et al. (2005). Optimization of codon usage enhances the immunogenicity of a DNA vaccine encoding mycobacterial antigen Ag85B. *Infect. Immun.* 73: 5666-5674.
- Liu H, He R, Zhang H, Huang Y, et al. (2010). Analysis of synonymous codon usage in *Zea mays*. *Mol. Biol. Rep.* 37: 677-684.
- Liu QP, Feng Y, Zhao XA, Dong H, et al. (2004). Synonymous codon usage bias in *Oryza sativa*. *Plant Sci.* 167: 101-105.
- Lü H, Zhao WM, Zheng Y, Wang H, et al. (2005). Analysis of synonymous codon usage bias in *Chlamydia*. *Acta Biochim. Biophys. Sin.* 37: 1-10.

- Meganathan PR, Pagan HJT, McCulloch ES, Stevens RD, et al. (2012). Complete mitochondrial genome sequences of three bats species and whole genome mitochondrial analyses reveal patterns of codon bias and lend support to a basal split in Chiroptera. *Gene* 492: 121-129.
- Mitrevva M, Wendl M, Martin J, Wylie T, et al. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7: 1-19.
- Morton BR and Wright SI (2007). Selective constraints on codon usage of nuclear genes from *Arabidopsis thaliana*. *Mol. Biol. Evol.* 24: 122-129.
- Mougel F, Manichanh C, N'guyen GD and Termier M (2004). Genomic choice of codons in 16 microbial species. *J. Biomol. Struct. Dyn.* 22: 315-329.
- Mukhopadhyay P, Basak S and Ghosh TC (2007). Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes. *Gene* 400: 71-81
- Nakamura T, Suyama A and Wada A (1991). Two types of linkage between codon usage and gene-expression levels. *FEBS Lett.* 289: 123-125.
- Naya H, Romero H, Carels N, Zavala A, et al. (2001). Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. *FEBS Lett.* 501: 127-130.
- Peng RH, Yao QH, Xiong AS, Cheng ZM, et al. (2006). Codon-modifications and an endoplasmic reticulum-targeting sequence additively enhance expression of an *Aspergillus* phytase gene in transgenic canola. *Plant Cell Rep.* 25: 124-132.
- Sharp PM, Bailes E, Grocock RJ, Peden JF, et al. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33: 1141-1153.
- Shi SL, Jiang YR, Liu YQ, Xia RX, et al. (2013). Selective pressure dominates the synonymous codon usage in Parvoviridae. *Virus Genes* 46: 10-19.
- Shields DC and Sharp PM (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res.* 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. *Mol. Biol. Evol.* 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 Res.* 22: 2437-2446.
- Subramanian S (2008). Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. *Genetics* 178: 2429-2432.
- Sueoka N and Kawanishi Y (2000). DNA G+C content of the third codon position and codon usage biases of human genes. *Gene* 261: 53-62.
- Vicario S, Moriyama EN and Powell JR (2007). Codon usage in twelve species of *Drosophila*. *BMC Evol. Biol.* 7: 226-226.
- Wang L and Roossinck MJ (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Mol. Biol.* 61: 699-710.
- Warnecke T and Hurst LD (2007). Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. *Mol. Biol. Evol.* 24: 2755-2762.
- Warnecke T, Batada NN and Hurst LD (2008). The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS Genet.* 4: e1000250-e1000250.
- Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
- Zhang Z, Dai W, Wang Y, Lu C, et al. (2013). Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.* 158: 145-154.
- Zhou JH, Gao ZL, Zhang J, Chen HT, et al. (2012). Comparative the codon usage between the three main viruses in pestivirus genus and their natural susceptible livestock. *Virus Genes* 44: 475-481.