Pak. J. Bot., 46(3): 823-826, 2014.

THE RELATIONSHIP BETWEEN CODON USAGE BIAS AND COLD RESISTANT GENES

MUHAMMAD YOUNAS KHAN BAROZAI* AND MUHAMMAD DIN

Department of Botany, University of Baluchistan, Sariab Road Quetta, Pakistan
**Corresponding author e-mail: barozaikhan@gmail.com*

Abstract

This research is based on synonymous codon usage which has been well-known as a feature that affects typical expression level of protein in an organism. Different organisms prefer different codons for same amino acid and this is called Codon Usage Bias (CUB). The codon usage directly affects the level or even direction of changes in protein expression in responses to environmental stimuli. Cold stress is a major abiotic factor that limits the agricultural productivity of plants. In the recent study CUB has been studied in *Arabidopsis thaliana* cold resistant and housekeeping genes and their homologs in rice (*Oryza sativa*) to understand the cold stress and housekeeping genes relation with CUB. Six cold resistant and three housekeeping genes in *Arabidopsis thaliana* and their homologs in rice, were subjected to CUB analysis. The three cold resistant genes (DREB1B, RCI and MYB15) showed more than 50% (52%, 61% and 66% respectively) similar codon usage bias for *Arabidopsis thaliana* and rice. On the other hand three cold resistant genes (MPK3, ICE1 and ZAT12) showed less than 50% (38%, 38% and 47% respectively) similar codon usage bias for *Arabidopsis thaliana* and rice. The three housekeeping genes (Actin, Tubulin and Ubiquitin) showed 76% similar codon usage bias for *Arabidopsis thaliana* and rice. This study will help to manage the plant gene expression through codon optimization under the cold stress.

Introduction

The average expression level of protein in an organism is affects by synonymous codon usage. The vast majority of prokaryotes and eukaryotes species have non-random codon usage. The major factor in codon choice in many unicellular and multicellular organism is Darwinian selection between synonymous; high expressed gene using a restricted set of codon. Most synonymous codons differ by only one base at their 3’ end, the so–called third position base. Analysis of codon usage has been used to identify highly expressed gene (Ikemura, 1985; Matassi *et al*., 1989).

Codon usage bias (CUB) refers to differences in the frequency of occurrence of synonymous codons in coding DNA. CUB is defined as the non-uniform use of synonymous codons within a gene. CUB invokes some type of natural selection. Synonymous codon usage patterns can vary significantly among genomes. In addition, one can also observe differences in synonymous codon usage among different genes within a single genome. For prokaryotes and unicellular eukaryotes such as yeast, the variation in codon usage within a genome is thought to be due to natural selection acting to optimize protein production (Ikemura, 1985; Sharp & Li, 1987).

Cold stress is a major environmental factor that limits the agricultural productivity of plants. Low temperature has a huge impact on the survival and productivity of plants (Shinwari *et al*., 1998). Cold stress often affects plant growth and crop productivity, which causes significant crop losses. Plants differ in their tolerance to chilling (0-15ºC) and freezing (<0ºC) temperatures. Plants are capable of developing freezing tolerance when they are exposed to low non-freezing temperatures (Xin & Browse, 2000; Mohamed *et al*., 2013; Nakashima *et al*., 2000).

Acquired freezing tolerance involves extensive reprogramming of gene expression and metabolism. The changes in expression of genes in response to cold temperatures are followed by increases in the levels of hundreds of metabolites, some of which are known to have protective effects against the damaging effects of cold stress (Barozai & Husnain, 2011; Narusaka *et al*., 2003). The reprogramming of gene expression results in the accumulation not only of protective proteins but also of hundreds or more of metabolites, some of which are known to have protective effects. Global transcript profiling analyses indicate that >10 of genes in the *Arabidopsis* genome are regulated during cold acclimation (Fowler & Thomashow, 2002; Kidokoro *et al*., 2009).

Bioinformatics’ approaches are the well-known way of research for nucleic acid (DNA and RNA) sequence analyses to find new findings (Barozai *et al*., 2008; Barozai, 2012a; 2012b; Ghani *et al.,* 2013). In this particular study, six cold resistant genes (zinc-finger protein 12 (ZAT12), Inducer of CBF expression (ICE), Rare Cold Inducible Gene (RCI), C-repeated Binding Factor)/Drought Responsive Element Binding 1 (CBF1/DREB1B), Myeloblastosis (MYB15) and Mitogen –activated protein kinase (MPK3) and three housekeeping genes (Actin, Tubulin and Ubiquitin) were selected from *Arabidopsis thaliana* and rice (*Oryza sativa*) and subjected to CUB analysis using bioinformatics tools. The aim of the study is to find co-relation between CUB and cold stress along with the housekeeping genes in *Arabidopsis thaliana* and rice (*Oryza sativa*).

Materials and Methods

Identification of candidate genes: The 6 cold resistant genes (ZAT12, ICE1, RCI, CBF1/DREB1B, MYB15 and AtMPK3) and three housekeeping genes (Actin, Tubulin and Ubiquitin) in *A. thaliana* and their homologs in *O. sativa* were identified through literature survey.

Sequence retrieval: The 6 cold resistant genes (ZAT12, ICE1, RCI, CBF1/DREB1B, MYB15 and AtMPK3) and three housekeeping genes (Actin, Tubulin and Ubiquitin) in *A. thaliana* and *O. sativa* were retrieved from the NCBI nucleotide data base. The FASTA formats of these sequences were saved and used for further analysis.

ORF (Open Reading Frame): The ORF is a graphical analysis tool which finds all open reading frames. The ORF finder is helpful in preparing complete protein sequence for a nucleotide sequence. The *A. thaliana* and *O. sativa* gene in FASTA format were subjected to NCBI, ORF Finder and longest frame is selected for analysis and study.

Codon usage analysis: For codon usage analysis the FASTA format sequences were pasted in genetic codon usage analysis program publically available at www.geneinfinity.org. The codon usage analysis program was used for analysis of amino acids showing different codons. The results were saved.

The RSCU (Relative Synonymous Codon Usage) values for each codon in the sequences were determined with the help of following formula:

$$\text{RSCU} = \frac{\text{No. of codon used}}{\text{Total No. of codon}} \text{ x No. of amino acids}$$

Results and Discussion

Cold resistant genes and their ORF (Open Reading Frame): The six genes, ZAT12, ICE1, RCI, CBF1/DREB1B, MYB15 and MPK 3 were retrieved by literature survey as cold resistant genes in *A. thaliana* and *O. sativa*. A number of researchers reported these genes as cold resistant genes (Tsuyoshi *et al*., 1996; Chinnusamy *et al*., 2003; Sholpan *et al*., 2005). The ORF analysis (Table 1) revealed that the *A. thaliana* MYB15 preferred the frame +1, while the *O. sativa* MYB15 showed preference to the frame +3. The frame length in base pair and amino acid for *A. thaliana* MYB15 is shorter than *O. sativa*. MPK3 preferred the frame +1 and observed with almost same frame length in base pair and amino acid in both plants *A. thaliana* and *O. sativa*. Similarly, ICE1 also preferred the frame +1 in *A. thaliana* and *O. sativa*. The ICE1 frame lengths in base pair and amino acid are found little shorter in *A. thaliana* than *O. sativa*. The *A. thaliana* ZAT12 was observed with the frame +1 while in *O. sativa* ZAT12 showed preference to +3 frame. The frame lengths in base pair and amino acid for ZAT12 were observed shorter in *A. thaliana* than the *O. sativa*. CBF1/DREB1 preferred the frame +2 in *A. thaliana* while, in *O. sativa* showed preference to the frame +3, whereas, almost the same frame lengths for base pair and amino acid were observed in both plants. RCI preferred the frame +2 in *A. thaliana* while, in *O. sativa* it showed preference to the frame +3. The RCI frame lengths in base pair and amino acid were observed shorter in *A. thaliana* than the *O. sativa*. The housekeeping genes Actin and Tubulin showed almost same frame lengths in base pair and amino acids. For ubiquitin there is a significant difference observed in the both frame lengths. The tubulin observed with the same frame but actin and ubiquitin adopted different frames in *A. thaliana* and *O. sativa*. The ORF role in gene expression is reported for different organisms (Kastenmayer *et al*., 2006; McCoy *et al*., 2009).

Table 1. The ORF features of the selected cold resistant genes in *A. thaliana* and their homologs in *Oryza sativa*.

Gene name	Plant	Frame	Frame length in bp	Frame length in AA
MPK3	Ath	+1	1113	370
	Osa	+1	1110	369
ICE1	Ath	+1	1485	494
	Osa	+1	1575	524
ZAT12	Ath	+1	489	162
	Osa	+3	1257	418
CBF1/DREB1	Ath	+2	642	213
	Osa	+3	687	228
MYB15	Ath	+1	822	273
	Osa	+3	1500	499
RCI	Ath	+2	981	326
	Osa	+3	2769	922
Actin	Ath	+1	1134	377
	Osa	+2	1134	377
Tubulin	Ath	+3	1335	444
	Osa	+3	1344	447
Ubiquitin	Ath	+3	921	306
	Osa	+2	474	157

where, Ath (*A. thaliana*), Osa (*O. sativa*), bp (base pair) and AA (Amino Acids)

Codon usage bias (CUB) analysis: The CUB analysis for the six cold stress resistant genes (zinc-finger protein 12 (ZAT12), Inducer of CBF expression (ICE), Rare Cold Inducible Gene (RCI), C-repeated Binding Factor)/Drought Responsive Element Binding 1 (CBF1/DREB1B), Myeloblastosis (MYB15) and Mitogen –activated protein kinase (MPK3) and three housekeeping genes (Actin, Tubulin and Ubiquitin) in *A. thaliana* and rice (*O. sativa*) showed very interesting results. The Table 2 summarized the most used codon usage bias (CUB) calculated in term of RSCU (Relative synonymous codon usage) for the six selected cold resistant and three housekeeping genes in *A. thaliana* and *O. sativa* for the 20 amino acids and stop codons.

The three cold resistant genes (DREB1B, RCI and MYB15) showed more than 50% (52%, 61% and 66% respectively) similar codon usage bias in *A. thaliana* and *O. sativa*. On the other hand three cold resistant genes (MPK3, ICE1 and ZAT12) showed less than 50% (38%, 38% and 47% respectively) similar codon usage bias in *A. thaliana* and *O. sativa*. The three housekeeping genes (Actin, Tubulin and Ubiquitin) showed 76% similar codon usage bias in *A. thaliana* and *O. sativa*. Mukhopadhyay *et al*., (2008) reported similar CUB analyses for tissue specific and housekeeping genes in rice and *A. thaliana*. These findings suggest that there is non-unimodal distribution of codon usage patterns for the cold resistant genes between the two species; one from the dicot (*A. thaliana*) and other from monocot (rice) plants. The half set of the cold resistant and housekeeping genes showing >50% similarity suggests the straight co-relation between the CUB: cold stress and CUB: housekeeping, irrespective of the plant species. This finding indicating that plant can be engineered for the cold stress resistant through codon optimization. On the other hand the half set showing <50% similarity suggests the independence of the cold stress and plant species. Similar findings were reported for other plant species (Whittle *et al*., 2007; Wang *et al*., 2007; Mukhopadhyay *et al*., 2008).

Table 2. The most used codon usage bias (CUB) calculated in term of RSCU (relative synonymous codon usage) for the six selected cold resistant and three housekeeping genes in *A. thaliana* and rice (*O. sativa*).

Amino acids	The most used codon usage bias (CUB)			
	A. thaliana **cold stress genes**	***A. thaliana*** **housekeeping genes**	***O. sativa*** **cold stress genes**	***O. sativa*** **housekeeping genes**
Ala	GCT	GCT	GCG	GCT
Cys	TGT	TGT	TGC	TGT
Asp	GAT	GAT	GAC	GAT
Glu	GAG	GAG/GAA	GAG	GAA
Phe	TTT	TTC	TTC	TTC
Gly	GGT	GGA	GGC	GGA
His	CAT	CAC	CAT	CAC
Ile	ATT/ATC	ATT	ATC	ATT
Lys	AAA	AAA	AAG/AAC	AAG
Leu	TTG	CTG	CTC	CTT
Met	ATG	ATG	ATG	ATG
Asn	AAT/AAC	AAT	AAT/AAC	AAT
Pro	CCT	CCT	CCG	CCA
Gln	CAA	CAA	CAG	CAA
Arg	AGG	AGA	CGC	AGG
Ser	TCT	TCA	TCT	TCC
Thr	ACT	ACA	ACC/ACA	ACC/ACA
Val	GTT	GTG	GTG	GTT/GTG
Trp	TGG	TGG	TGG	TGG
Tyr	TAT	TAT	TAC	TAT
Stop-codon	TAG	TGA	TGA/TAG	TGA

The housekeeping and cold stress resistant genes comparison showed very interesting results for rice and *A. thaliana*. The 57% codon showed similar and 43% codon showed different CUB for housekeeping and cold stress resistant genes in *A. thaliana*. In rice, the 37% codons were observed with same and 63% codons were observed with different CUB analyses for housekeeping and cold stress resistant genes. The CUB similarity <50% for rice, housekeeping and cold stress resistant genes suggests separate codon optimization program for the stress handling. In the *A. thaliana*, the CUB similarity >50% suggests unimodal codon usage patterns under normal and cold stress conditions. Such findings clearly indicate the dependence of the codon usage on the plant species under normal and stress conditions. These are in agreements with the previous reported works (Whittle *et al*., 2007; Wang *et al*., 2007; Mukhopadhyay *et al*., 2008). These findings also confirm that applying bioinformatics' new and interesting discoveries can be made (Barozai *et al*., 2012a; 2012b; Barozai & Wahid, 2012; Barozai, 2012c).

Conclusions

From the current study it is concluded that there is multimodal co-relation between the codon usage bias (CUB) and cold stress resistant genes among the plant species. For the housekeeping genes unimodal co-relation with CUB is observed in dicot and monocot plants. The plant stress resistant can be improved by optimizing the codon usage.

References

Barozai, M.Y.K. 2012a. The novel 172 sheep (*Ovis aries*) microRNAs and their targets. *Mol Biol Rep*., 39(5):6259-6266.

Barozai, M.Y.K. 2012b. Identification and characterization of the microRNAs and their targets in Salmo salar. *Gene*., 499(1): 163-168.

Barozai, M.Y.K. 2012c. Insilco identification of microRNAs and their targets in fiber and oil producing plant Flax (*Linum usitatissimum* L.). *Pak. J. Bot*., 44: 1357-1362.

Barozai, M.Y.K., A.G. Kakar and M. Din. 2012a. The relationship between codon usage bias and salt resistant genes in *Arabidopsis thaliana* and *Oryza sativa*. *Pure. Appl. Biol*., 1(2):48-51.

Barozai, M.Y.K. and H.A. Wahid. 2012. Insilico identification and characterization of cumulative abiotic stress responding Genes in Potato (*Solanum tuberosum* L.). *Pak. J. Bot.*, 44(SI): 57-69.

Barozai, M.Y.K. and T. Hussnain. 2012. Identification of biotic and abiotic stress up-regulated ESTs in *Gossypium arboreum*. *Mol. Biol. Rep*., 39(2): 1011-1018.

Barozai, M.Y.K., I.A. Baloch and M. Din. 2012b. Identification of MicroRNAs and their targets in Helianthus. *Mol Biol Rep*., 39(3): 2523-2532.

Barozai, M.Y.K., M. Irfan, R. Yousaf, I. Ali, U. Qaisar, A. Maqbool, M. Zahoor, B. Rashid, T. Hussnain and S. Riazuddin. 2008. Identification of micro-RNAs in cotton. *Plant Physiol. and Biochem.*, 46 (8-9): 739-51.

Chinnusamy V., M. Ohta, S. Kanrar, B. Lee, X. Hong, M. Agarwal and J. Zhu. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes and Dev*., 17(8): 1043-1054.

Fowler, S. and M.F. Thomashow. 2002. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*., 14: 1675-1690.

Ghani, A., M. Din., I.A. Baloch and M.Y.K. Barozai. 2013. Identification of microRNAs in 12 plant species of Fabaceae. *Pure. Appl. Biol*., 2(3): 104-115.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*., 2: 13-34.

Kastenmayer, J.P., L. Ni, A. Chu, L.E. Kitchen and W.C. Au. 2006. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res*., 16: 365-373.

Kidokoro, S.K. Nakashima, Z.K. Shinwari, K. Shinozaki and K. Yamaguchi-Shinozaki. 2009. The phytochrome-interacting factor PIF7 negatively regulates *DREB1* expression under circadian control in *Arabidopsis. Plant Physiology*, 151(4): 2046-2057.

Matassi, G., L.M. Montero, J. Salinas and G. Bernardi. 1989. The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. *Nucleic Acids Res*., 17: 5273-5290.

Mccoy, M.W., A.P. Allen and J.F. Gillooly. 2009. The random nature of genome architecture: predicting open reading frame distributions. *PLoS One*, 4(7): (2009) e6456. doi: 10.1371/journal.pone.0006456.

Mohamed, M.S., M.M. Hanafi, S.N. Akmal, A. Naghmeh, N. Mohd, Y. Rafii and P. Azizi. 2013. Extraction of total RNA and identification of different genes in mangrove plant involved in its adaptability to the variety of stresses. *Pak. J. Agri. Sci*., 50: 529-536.

Mukhopadhyay, P., S. Basak and T.C. Ghosh. 2008. Differential selective constraints shaping codon usage pattern of housekeeping and tissue-specific homologous genes of rice and Arabidopsis. *DNA Res*., 15: 347-356.

Nakashima, K., Z.K. Shinwari, S. Miura, Y. Sakuma, M. Seki, K. Yamaguchi-Shinozaki and K. Shinozaki. 2000. Structural organization, expression and promoter activity of an Arabidopsis gene family encoding DRE/CRT binding proteins involved in dehydration- and high salinity-responsive gene expression. *Plant Molecular Biology*, 42(4): 657-665.

Narusaka, Y., K. Nakashima, Z.K. Shinwari, Y. Sakuma, T. Furihata, H. Abe, M. Narusaka, K. Shinozaki and K.Y. Shinozaki. 2003. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high salinity stresses. *The Plant Journal*, 34(2): 137-149.

Sharp, P.M. and W.H. Li. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*., 15: 1281-1295.

Shinwari, Z.K., K. Nakashima, S. Miura, M. Kasuga, M. Seki, K. Yamaguchi-Shinozaki and K. Shinozaki. 1998. An Arabidopsis gene family encoding DRE binding protein involved in low temperature-responsive gene expression. *Biochemical Biophysical Research Communications*, 250: 161-170.

Sholpan, D., S. Karen, C. Jesse and M. Ron. 2005. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Plant Physiol*., 139(2): 847-856.

Tsuyoshi, M., I. Kenji, H. Takashi, H. Nobuaki, Y.S. Kazuko, M. Kunihiro and S. Kazuo. 1996. A gene encoding a Mitogen-activated protein- in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*., 93: 765-769.

Wang, H.C. and D.A. Hickey. 2007. Rapid divergence of codon usage patterns within the rice genome. *BMC Evol. Biol*., (2007): 7(Suppl 1): S6.

Whittle, C.A., M.R. Malik and J.E. Krochko. 2007. Gender-specific selection on codon usage in plant genomes. *BMC Genomics*, 8: 169-179.

Xin, Z. and J. Browse. 2000. Cold comfort farm: the acclimation of plants to freezing temperatures. *Plant Cell and Env*., 23(9): 893-902.

(Received for publication 4 June 2012)