

THE RELATIONSHIP BETWEEN CODON USAGE BIAS AND SALT RESISTANT GENES IN *ARABIDOPSIS THALIANA* AND *ORYZA SATIVA*

Muhammad Younas Khan Barozai,^{a*} Abdul Ghani Kakar,^a Muhammad Din^a

^a Department of Botany, University of Baluchistan, Sariab Road Quetta Pakistan.

*Corresponding author: barozaikhan@gmail.com

Abstract

This research is based on synonymous codon usage and its relation with the salt resistant genes. The codon usage directly affects the level or even direction of changes in protein expression in responses to environmental stimuli. Soil salinity is a major abiotic stress for plant agriculture. Sodium ions in saline soils are toxic to plants because of their adverse effects on K⁺ nutrition, cytosolic enzyme activities, photosynthesis, and metabolism. In the recent study codon usage bias (CUB) has been studied in *Arabidopsis thaliana* salt resistant and housekeeping genes and their homologs in rice (*Oryza sativa*) to understand the salt stress and housekeeping genes relation with CUB. Total five salt stress tolerant genes; Salt Overly sensitive (SOS1), basic lucien-Zipper Protein (bZIP), Calcinurin B like binding 10 (CBL10), Mitogen Activated Protein Kinase-3 (MPK3) and Histidine Kinase-1 (HK1) were identified in *Arabidopsis thaliana*. For the comparative study and removing the organism bias the orthologues of the five salt stress tolerant genes were also identified in rice. Briefly, five salinity resistant and three housekeeping genes in *Arabidopsis thaliana* and their orthologues in rice, were subjected to CUB analysis. The four salt resistant genes (SOS1, CBL10, bZIP and HK1) showed more than 50% (71%, 57%, 67% and 57% respectively) similar codon usage bias for *Arabidopsis thaliana* and rice. On the other hand fifth salt resistant gene (MPK3) showed less than 50% (29%) 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 will help us to engineer the salt resistant crops by adjusting the codon usage.

Keywords: *Arabidopsis thaliana*; codon usage bias; housekeeping genes; rice; salt stress

Introduction

Codon Usage Bias (CUB) is defined as the non-uniform use of synonymous codons within a gene [1]. CUB has been extensively documented across a wide range of organisms and varies greatly both within and between genomes [2, 3].

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. Specifically, the most highly expressed genes use codons that are complementary to the most abundant tRNA anticodons. For multicellular eukaryotes, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, there is also some evidence that codon bias might be caused by selection for translational efficiency. For the majority of multicellular organisms, however, it has been difficult to explain codon usage variation within a genome in terms of natural selection. This would mean that selection for gene high expression is the primary factor determining the codon usage bias in this case. Synonymous codon usage has long been known as a factor that affects average expression level of protein in an organism. 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 [1, 2]. 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.

Soil salinity is a major abiotic stress for plant agriculture. Sodium ions in saline soils are toxic to plants because of their adverse effects on K⁺ nutrition, cytosolic enzyme activities, photosynthesis, and metabolism [4]. Salinity limits both yield and the expansion of agricultural plants. In the last 20 years our basic understanding of the mechanisms underlying plant tolerance and adaptation to saline environments has greatly improved owing to active development of advanced tools in molecular, genomics, and bioinformatics analyses. However, the full potential of investigative power has not been fully exploited. The changes in expression of genes in response to salt stress 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 salt stress. 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 [5, 6].

Bioinformatics' approaches are the well-known way of research for nucleic acid (DNA and RNA) sequence analyses to find new findings [7-9]. In this particular study, five salt stress tolerant genes Salt Overly sensitive (SOS1), basic lucien-Zipper Protein (bZIP), Calcinurin B like binding 10 (CBL10), Mitogen Activated Protein Kinase-3 (MPK3) and Histidine Kinase-1 (HK1) and three housekeeping genes (Actin, Tubulin and Ubiquitin) were selected from *Arabidopsis thaliana* and rice (*Oryza sativa*) and subjected to CUB analysis using bioinformatic's tools. The aim of the study is to find co-relation between CUB and salt stress along with the housekeeping genes in *Arabidopsis thaliana* and rice (*Oryza sativa*).

Materials and Methods

Identification of Candidate Genes

The five salt resistant genes (SOS1, bZIP, CBL10, MPK3 and HK1) and three housekeeping genes (Actin, Tubulin and Ubiquitin) in *Arabidopsis thaliana* and their homologs in *Oryza sativa* were identified through literature survey.

Sequence Retrieval

The five salt resistant genes (SOS1, bZIP, CBL10, MPK3 and HK1) and three housekeeping genes (Actin, Tubulin and Ubiquitin) in *Arabidopsis thaliana* and *Oryza sativa* were retrieved from the NCBI nucleotide data base. The FASTA formats of these sequences were saved and used for further analysis.

Prediction of 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 *Arabidopsis thaliana* and *Oryza sativa* gene in FASTA format were subjected to NCBI, ORF Finder and longest frame is selected for downstream analysis and study.

Codon Usage Analysis

For codon usage analysis the FASTA format sequences were subjected to 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:

$$RSCU = (\text{No. of codon used} / \text{Total No. of codon}) \times \text{No. of Amino Acids.}$$

Results and Discussion

Salt Resistant Genes and their ORF (Open Reading Frame)

The five genes, SOS1, bZIP, CBL10, MPK3 and HK1 were identified as salt resistant genes in *A. thaliana* and *Oryza sativa*. A number of researchers reported these genes as salt resistant genes [10-12]. The ORF analysis (Table-1) reveals that the *A. thaliana* (Ath)-SOS1 prefers the frame +1, while the *O. sativa* SOS1 showed preference to the frame +3. The frame length in base pair and Amino acid in *A. thaliana* and *O. sativa* is almost same. HK1 preferred the frame +1 in both, but the frame length in base pair and amino acid in *A. thaliana* is larger than the frame length in *O. sativa*. Similarly the Ath-CBL10 prefer the frame +1 in *A. thaliana* and *O. sativa* CBL10 also showed preference to the same frame that is +1 respectively. The frame length in base pair is larger in *A. thaliana* than *O. sativa*, while frame length in Amino acid for CBL10 in *A. thaliana* is shorter than the *O. sativa*. The Ath-MPK3 prefer the frame +2 in *A. thaliana* while, *O. sativa* showed preference to the frame +1. The frame length in base pair and amino acid for MPK3 was observed in the same length for *A. thaliana* and *O. sativa*. For the housekeeping genes Actin and Tubulin showed almost same frame length in base pair and amino acids. For ubiquitin there is a significant difference 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 [13, 14].

Table-1: The ORF features of the selected Cold Resistant genes in *A. thaliana* (Ath) and their homologs in *Oryza sativa* (Osa).

Gene name	Nucleotide base pairs	Frame	Length	Amino acids
SOS1 Ath	3682	+1	3441	1146
SOS1 Osa	3660	+3	3447	1148
CBL10 Ath	1101	+1	616	204
CBL10 Osa	801	+1	801	266
AtHK1 Ath	4040	+1	3624	1207
HK1 Osa	1467	+1	1317	438
AtbZIP 17 Ath	2467	+3	2166	721
OsbZIP Osa	1083	+1	1082	360
AtMPK3 Ath	1483	+2	1113	370
MPK3 Osa	1110	+1	1110	369

Codon Usage Bias (CUB) Analysis

The CUB analysis for the five salt stress tolerant genes Salt Overly sensitive (SOS1), basic lucien-Zipper Protein (bZIP), Calcinurin B like binding 10 (CBL10), Mitogen Activated Protein Kinase-3 (MPK3) and Histidine Kinase-1 (HK1) and

Table-2: The RSCU Features of the selected Salt Resistant Genes in *A. thaliana* (ath) and their homologs in *Oryza sativa* (Osa).

Amino acids	SOS1		CBL10		BZIP		HK1		MPK3	
	Ath	Os	Ath	Os	Ath	Os	Ath	Os	Ath	Os
Ala	GCA (1.23%)	GCA (1.88%)	GCA (1.42%)	GCC (1.28%)	GCA (2.18%)	GCT (1.52%)	GCT (1.54%)	GCA (1.86%)	GCT (1.68%)	GCG (2.26%)
Cys	TGA (1.1%)	TGC (1.37%)	TGT (1.2%)	TGC (1%)	TGT (1.14%)	TGC (1.33%)	TGT (1%)	TGC (1.11%)	TGT (1.33%)	TGC (1%)
Asp	GAT (1.16%)	GAT (1.1%)	GAT (1.2%)	GAT (1%)	GAT (1.33%)	GAT (1.55%)	GAT (1.45%)	GAT (1.14%)	GAT (1.17%)	GAC (1.76%)
Glu	GAA (1.25%)	GAA (1.06%)	GAG (1.6%)	GAG (1.21%)	GAG (1%)	GAA (1.25%)	GAA (1.09%)	GAA (1.25%)	GAA (1.57%)	GAG (1.77%)
Phe	TTT (1.10%)	TTT (1.02%)	TTT (1.14%)	TTT (1.05%)	TTT (1.03%)	TTT (1.11%)	TTT (1.29%)	TTT (1%)	TTT (1.30%)	TTC (1.88%)
Gly	GGA (1.33%)	GGC (1.05%)	GGA (2.28%)	GGG (1.5%)	GGA (2.34%)	GGA (1.23%)	GGA (2.16%)	GGC (1.22%)	GGA (2.93%)	GGC (2%)
His	CAC (1.2%)	CAT (1.05%)	CAT (1.63%)	CAT (0.5%)	CAT (1%)	CAT (1.6%)	CAC (1.31%)	CAT (1.06%)	CAT (1%)	CAC (1.69%)
Ile	ATC (1.28%)	ATC (1.36%)	ATT (1.12%)	ATT (1.61%)	ATT (1.13%)	ATT (1.25%)	ATT (1.3%)	ATA (1.09%)	ATC (1.2%)	ATC (2.63%)
Lys	AAA (1.43%)	AAA (1.93%)	AAG (1.2%)	AAG (1.41%)	AAG (1.31%)	AAG (1.2%)	AAG (1.08%)	AAG (1.41%)	AAA (1.22%)	AAG (2%)
Leu	CTC (1.53%)	TTG (1.93%)	TTG (2.03%)	CTG (2.54%)	CTA (1.16)	CTA (1.44%)	CTT (1.81%)	CTT (1.73%)	CTC (1.56%)	CTC (2.86%)
Met	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)	ATG (1%)
Asn	AAT (1.52%)	AAT (1.2%)	AAT (1.06%)	AAT (1.14%)	AAT (1.13%)	AAT (1%)	AAT (1.01%)	AAT (1.18%)	AAT (1.6%)	AAC (2%)
Pro	CCA (1.44%)	CCA (1.5%)	CCG (1.6%)	CCG (2%)	CCA (1.6%)	CCA (1.77%)	CCA (2.13%)	CCA (1.9%)	CCA (1.6%)	CCG (2.5%)
Gln	CAA (1.44%)	CAA (1.33%)	CAA (1.33%)	CAA (1.42%)	CAG (1%)	CAG (1.9%)	CAA (1.17%)	CAG (1.12%)	CAG (1%)	CAG (1.69%)
Arg	AGA (1.83%)	AGG (2.28%)	AGA (2.60%)	AGA (3.2%)	AGA (2.38%)	AGG (2.57%)	AGA (2.75%)	AGA (1.86%)	AGA (2.4%)	CGG (2%)
Ser	TCT (1.61%)	TCA (1.29%)	TAT (1.46%)	TCC (1.8%)	TCA (1.05%)	AGC (2%)	TCT (1.59%)	AGC (1.6%)	TCA (1.68%)	TCC (2.76%)
Thr	ACA (1.76%)	ACA (2.13%)	ACT (1.33%)	ACC (1.42%)	ACG (1.3%)	ACT (1.83%)	ACT (1.25%)	ACA (1.36%)	ACT (1.53%)	ACG (2.5%)
Val	GTT (1.63%)	GTG (2.10%)	GTT (1.33%)	GTG (1.55%)	GTG (1.89%)	GTG (1.5%)	GTT (1.63%)	GTG (1.75%)	GTT (1.84%)	GTG (2.4%)
Trp	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)	TGG (1%)
Tyr	TAT (1.2%)	TAT (1.6%)	TAT (1.6%)	TAC (1.33%)	TAT (1.33%)	TAT (1.2%)	TAC (1.02%)	TAT (1.42%)	TAT (1.03%)	TAC (0.83%)
End	TGA (1.08%)	TGA (1.32%)	TGA (1.85%)	TAG (1%)	TGA (1.41%)	TAA (3%)	TGA (1.66%)	TGA (3%)	TGA (1.75%)	TAG (3%)

three housekeeping genes (Actin, Tubulin and Ubiquitin) in *A. thaliana* and rice (*Oryza sativa*) showed very interesting result. The Table-2 summarized the most used codon usage bias (CUB) calculated in term of RSCU (Relative synonymous codon usage) for the five selected salt resistant genes in *Arabidopsis thaliana* and rice (*Oryza sativa*) for the 20 amino acids and stop codon.

The four salt resistant genes (SOS1, CBL10, bZIP and HK1) showed more than 50% (71%, 57%, 67% and 57% respectively) similar codon usage bias for *Arabidopsis thaliana* and rice. On the other hand fifth salt resistant gene (MPK3) showed less than 50% (29%) 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.

Mukhopadhyay et al., [15] reported similar CUB analyses for tissue specific and housekeeping genes in rice and *A. thaliana*. These findings suggest that there is a clear unimodal distribution of codon usage patterns for the salt resistant genes between the two species of the dicot (*A. thaliana*) and monocot (rice) plants. The 80% salt resistant and 100% housekeeping genes showing >50% similarity, suggests the straight co-relation between the CUB: salt 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 20% showing <50% similarity suggests the independence of the salt stress and plant species. Similar findings were reported for other plant species [15-17]. These findings also confirm that applying bioinformatics'

tools novel interesting findings can be made [18-22].

Conclusions

From the current study it is concluded that there is unimodal co-relation between the codon usage bias (CUB) and salt 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.

Reference

- Ikemura T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 2: 13–34.
- Matassi G, Montero LM, Salinas J, Bernardi G (1989). The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. *Nucleic Acids Res.* 17: 5273–5290.
- Sharp PM, Li WH (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281–1295.
- Cardinale F, Meskiene I, Ouaked F, Hirt H (2002). Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *Plant Cell* 14: 703–711.
- Huazhong S, Ishitani M, Kim C, Zhu JK (2000). The *Arabidopsis thaliana* salt tolerance gene SOS1 encodes a putative Na⁺/H⁺ antiporter. *Proc Natl Acad Sci USA.* 12: 6896–6901.
- Kim BG, Waadt R, Cheong YH, Pandey GK, Dominguez-Solis JR, Schultke S, Lee SC, Kudla J, Luan S (2007). The calcium sensor CBL10 mediates salt tolerance by regulating ion homeostasis in *Arabidopsis*. *The Plant Journal.* 52: 473–484.
- Barozai MYK, Irfan M, Yousaf R, Ali I, Qaisar U, Maqbool A, Zahoor M, Rashid B, Hussnain T, Riazuddin S (2008). Identification of micro-RNAs in cotton. *Plant Physiol and Biochem.* 46 (8-9): 739-51.
- Barozai MYK (2012). Identification and characterization of the microRNAs and their targets in *Salmo salar*. *Gene.* 499(1): 163-168.
- Barozai MYK, Baloch IA, Din M (2012). Identification of MicroRNAs and their targets in *Helianthus*. *Mol Bio Rep.* 39(3): 2523-32.
- Tran LSP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, Shinozaki KY (2007). Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 104(51): 20623–20628.
- Liu JX, Srivastava R, Che P, Howell SH (2008). Salt stress signaling in *Arabidopsis thaliana* involves a membrane-bound transcription factor AtbZIP17 as a signal transducer. *Plant Signal Behav.* 3(1): 56-57.
- Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K (2000). Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases ATMPK4 and ATMPK6. *Plant. J.* 24: 655–665.
- Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au WC, et al (2006). Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* 16: 365–373.
- McCoy MW, Allen AP, Gillooly JF (2009). The random nature of genome architecture: predicting open reading frame distributions. *PLoS One.* 4 (7): e6456. doi: 10.1371/journal.pone.0006456.
- Mukhopadhyay P, Basak S, Ghosh TC (2008). Differential selective constraints shaping codon usage pattern of housekeeping and tissue-specific homologous genes of rice and *Arabidopsis*. *DNA Res.* 15: 347-356.
- Whittle CA, Malik MR, Krochko JE (2007). Gender-specific selection on codon usage in plant genomes. *BMC Genomics* 8: 169–179.
- Wang HC, Hickey DA (2007). Rapid divergence of codon usage patterns within the rice genome. *BMC Evol Biol.* 7(Suppl 1):S6.
- Barozai MYK, Din M, Baloch IA (2011). Identification of microRNAs in ecological model plant *Mimulus*. *J of Biophy Chem.* 2(3): 322-331.
- Barozai MYK, Husnain T (2012). Identification of biotic and abiotic stress up-regulated ESTs in *Gossypium arboreum*. *Mol Biol Rep.* 39(2): 1011-1018.
- Barozai MYK (2012). The novel 172 sheep (*Ovis aries*) microRNAs and their targets. *Mol Biol Rep.* 39(5):6259–6266.
- Barozai, MYK (2012). The microRNAs and their targets in the Channel Catfish (*Ictalurus punctatus*). *Mol Biol Rep.*, 39: 8867–8872.
- Barozai MYK, Wahid HA (2012). Insilico identification and characterization of cumulative abiotic stress responding genes in Potato (*Solanum tuberosum* L.). *Pak. J. Bot.*, 44: 57-96.