The Journal of Animal & Plant Sciences, 28(5): 2018, Page: 1532-1536
ISSN: 1018-7081

Short Communication

BIOINFORMATICS ANALYSIS OF CODON USAGE BIAS AND RNA SECONDARY STRUCTURES FOR SALT STRESS RESPONSIVE GENES IN *ARABIDOPSIS THALIANA* AND *SOLANUM LYCOPERSICUM*

B. Sadia[1*], W. N. Hamzazai[1], I. A. Baloch[2,3], S. Kazmi[1], Asyia[1], P. S. Jan[1]and S. Irfan[1]

[1]Department of Botany, SBK Women's University Quetta, Pakistan; [2]Department of Botany, University of Balochistan, Quetta, Pakistan; [3]Govt. Degree College Khanozai, Balochistan
[*]Corresponding author's email:iftekharbaloch@gmail.com

ABSTRACT

Salt stress is one of the most important abiotic stresses for agricultural crops and other plants. Salinity highly affects physical, physiological and metabolic responses of plants. Codon usage bias (CUB) and RNA secondary structures (RSS) were analyzed for salt stress responsive (SSR) genes of *Arabidopsis thaliana* and *Solanum lycopersicum*. Four salt stress resistant genes SOS1 (salt overly sensitive 1), BZIP (basic lucien –zipper protein), HVA22-like protein and CDPK1 (calcium-dependent protein kinase 1) of *Arabidopsis thaliana* as well as *Solanum lycopersicum* were compared for CUB and RSS. All genes expressed similarity for codon usage, as both plants showed to prefer more than 50% (65%, 50%, 60%, and 80%) similar codon usage for all four selected salt stress responsive genes. RSS analysis revealed that three SSR genes had significant difference in minimum free energy (MFE) while two differed in number of loops and stems depicting overall mix behavior. These findings can further help to engineer the salt resistant genes by adjusting the codon usage through genetic screening and by using bioinformatics tools.

Key words: Codon usage bias; Genetic code; Salinity responsive genes; Secondary structure.

INTRODUCTION

Proteins are coded by twenty different types of amino acids with all amino acids having more than two codons, except Methionine and Tryptophan which have single codon. The 64 codons for 20 amino acids suggests that synonymous codons for single type of amino acid are not used with the same frequency leading to the choosing situation for codon usage, termed as codon usage bias (CUB) (Ikemura, 1985). CUB is non-uniform usage of synonymous codons and exists due to some type of natural selection for required protein production. CUB can be different among different genomes (Barozai *et al*., 2014). With maximum codon usage appeared to be conserved in both the dicots and monocots. Most of the codons used by the plants for same type of amino acid usually have cytosine (C) or the guanine (G) bases in the end (Wang and Roossinck, 2006).

Codon usage bias has been studied in both prokaryotes and eukaryotes; however, the biasness is different among genes of a species (Comeron and Aguadé, 1998). Codon usage bias has also been investigated in higher plants, and monocots have been found to retain more GC content as compared to dicots. Thus it explains variable mutational conditions and different genomic level organization (Kawabe and Miyashita, 2003). So codon usage studies support in understanding evolutionary processes of plants along with the sequences of genes in different plants preferring specific codons (Campbell and Gowri, 1990). Interaction of base pairing within a single molecule as well as in a set of molecules is referred as secondary structure for RNA (Krieger *et al*., 2004). For the prediction of RNA secondary structure mfold (Zuker, 2003) is the most commonly used tool.

Salt, drought, heat and cold stresses are important environmental factors for plants regarding agricultural productivity. Salt stress is considered to be main damaging factor for irrigated lands in many regions of world (Zhu, 2002) and salt tolerance can be acquired by the reprogramming of genes. The changes in gene expressions due to stress conditions are preceded by the unusual increase in the number of metabolite contents which have protective capacity. The gene expression reprogramming in stress conditions not only results in formation of resistant proteins but also the number of metabolites having protection capabilities (Barozai and Wahid, 2012). For agriculture, one of the most important and serious abiotic factors is the salt stress as it lowers the overall yield of agricultural plants (Barozai *et al*., 2012). In saline soil, the presence of excessive concentrations of sodium ions (Na^+) has toxic effects on plants physiological and metabolic functions such as photosynthesis, enzymatic cytoplasmic activities and plant nutrition. The aim behind conducting this research was to examine and compare the codon usage bias and RNA secondary structures for salinity stress related genes in *Arabidopsis thaliana* and *Solanum lycopersicum.*

MATERIALS AND METHODS

Candidate genes identification: Through literature survey, following four genes were identified as salt stress responsive genes in *Arabidopsis thaliana* and *Solanum lycopersicum*; SOS1, BZIP, HVA22-like protein and CDPK1.

Sequence retrieval: The sequences of four salinity responsive genes (SOS1, BZIP, HVA22-like protein& CDPK1) of *Arabidopsis thaliana* and *Solanum lycopersicum* were retrieved from NCBI. The FASTA formats of all these genes of two plants were saved for further analysis and studies.

Prediction of ORF: The FASTA format of four salt stress responsive genes of *Arabidopsis thaliana* and *Solanum lycopersicum* were subjected to ORF finder which conducts graphical analysis to study and know all possible open reading frames. From the results, the longest possible frame was selected for analysis of codon usage bias.

Analysis of codon usage bias: For codon usage bias study, the FASTA format sequences (the sequence retrieved from ORF) of four salinity responsive genes (SOS1, BZIP, HVA22-like protein& CDPK1***)*** of *Arabidopsis thaliana* and *Solanum lycopersicum* were subjected to publically available website *(http://genome.urv.es/CAIcal/*) to check the codon usage pattern. Finally the analysis of codon usage bias in four salt stress responsive genes was conducted by calculating the Relative synonymous codon usage (RSCU) values for various amino acids of all the genes.

Secondary structure analysis: Publically available Mfold web server at*http://unafold.rna.albany.edu/?q=mfold* was used for the prediction of secondary structures. The secondary structures were analyzed by counting the number of loops, number of stems and minimum free energy.

RESULTS AND DISCUSSION

Results of ORF revealed that all four salt resistant genes in both plants prefer frame +1 or +2. They share same frames for two genes while different frames were observed in other two genes compared. While the frame lengths were almost same for three genes, a major difference in frame length was observed for one of the genes. The ORF results for genes SOS1 and CDPK1 suggested that both *Arabidopsis thaliana* and *Solanum lycopersicum* preferred the frame +2, while the lengths of frames were almost same in both plants for these two genes. For BZIP, *Arabidopsis thaliana* preferred frame +2 whereas *Solanum lycopersicum* preferred +1, while the length for *Arabidopsis thaliana* was 1104 base pairs, compared to *Solanum lycopersicum* which was 492 base pairs long. For the gene HVA22-like protein, *Arabidopsis thaliana* and *Solanum lycopersicum* preferred different frames +1 and +2, respectively, however, the length of frames was almost same for both plants (Table # 1).

Analysis of codon usage bias: As per codon usage results, the four genes (SOS1, BZIP, HVA22-like protein & CDPK1) showed very much similarity and were found to prefer more than 50% (65%, 50%, 60%, and 80% respectively) similar codons in salt stress conditions (Table # 2). These results were in agreement with that of Barozai *et al*., (2012) where same results were obtained by comparing salt stress responsive genes of *Arabidopsis thaliana* and *Oryza sativa.* Singha *et al*., (2014) compared the codon usage pattern of mitogen activated protein kinase (MAPK) genes in *Arabidopsis thaliana, Glycine max* and *Oryza sativa* and concluded that MAPK genes (responsible for different biotic and abiotic stress responses) also prefer specific codons in stress conditions.

Secondary structure analysis: RNA Secondary Structures analysis revealed that except SOS1 all three salt stress resistance genes had significant difference in minimum free energy (MFE) while the two genes BZIP and HVA22-like protein exhibited significant difference in number of loops and stems (Table # 3). Over all mixed results were obtained for RNA secondary structures in terms of MFE, loops and stem, confirming the results reported by Barozai *et al*., (2014) while working on virus resistant genes in *Arabidopsis thaliana* and *Oryza sativa*.

Table 1. ORF features of selected four salt stress responsive genes of *Arabidopsis thaliana* and *Solanum lycopersicum.*

Gene names	No. of Nucleotides	ORF	ORF Lengths	Amino acids
SOS1 *Arabidopsis thaliana*	3682	+1	3441	1146
SOS1 *Solanum lycopersicum*	3822	+1	3456	1151
bZIP *Arabidopsis thaliana*	1473	+2	1104	367
bZIP *Solanum lycopersicum*	1256	+1	492	163
HVA22-like protein *Arabidopsis thaliana*	802	+1	603	200
HVA22-like protein *Solanum lycopersicum*	1113	+2	564	187
CDPK1 *Arabidopsis thaliana*	2568	+2	1638	545
CDPK1 *Solanum lycopersicum*	2147	+2	1566	521

Conclusion: There exists a strong correlation among plants for codon usage during salt stress condition. This inference indicates that two plants belonging to different families prefer to use same codons for amino acids against salt stress situation. Through present study it has been confirmed that plants prefer to use similar codons in stress conditions. Genetic engineering endeavors can be facilitated by the use of these bioinformatics tools for crop improvement and to produce resistant varieties of plants through introduction of preferable salt stress genes in foreground.

REFERENCES

Barozai, M. Y. K, and A. H. Wahid (2012). Insilico identification and characterization of cumulative abiotic stress responding genes in Potato (Solanum tuberosum L.). Pakistan J. Bot44: 57-69.

Barozai, M. Y. K., A. G. Kakar, and M. Din (2012). 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., S. Q. Shah., M. Din, and R. Muhammad (2014). Codon usage bias and RNA secondary structures analysis for virus resistant genes in *Arabidopsis thaliana* and *Oryza sativa*. Pure Appl Biol.3: 81-91.

Barozai, M. Y. K., S.Q. Shah., M. Din, and R. Muhammad (2014). Codon usage bias and RNA secondary structures analysis for virus resistant genes in Arabidopsis thaliana and Oryza sativa. Pure ApplBiol3: 81-91.

Campbell, W. H, and G. Gowri (1990). Codon usage in higher plants, green algae, and cyanobacteria. Plant Physiology.92(1): 1-11.

Comeron, J. M, and M. Aguadé (1998). An evaluation of measures of synonymous codon usage bias. J. Molecular Evolution. 47(3): 268-274.

Ikemura, T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Molecular biology and evolution.2(1): 13-34.

Kawabe, A, and N. T. Miyashita (2003). Patterns of codon usage bias in three dicot and four monocot plant species. Genes & genetic systems.78(5): 343-352

Krieger, M., M. P. Scott., P. T. Matsudaira., H. F. Lodish., J. E. Darnell., Z. Lawrence., C. Kaiser, and A. Berk (2004). Structure of Nucleic Acids. Molecular cell biology. New York: W. H. Freeman *& CO*, ISBN 0-7167-4366-3.

Singha, H. S., S. Chakraborty, and H. Deka (2014). Stress induced MAPK genes show distinct pattern of codon usage in *Arabidopsis thaliana*, Glycine max and *Oryza sativa*. Bioinformation 10(7): 436.

Wang, L, and M. J. Roossinck (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. Plant molecular biology.61(4-5): 699-710.

Zhu, J. K (2002). Salt and drought stress signal transduction in plants. Annual review of plant biology53: 247.

Zuker, M (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research 31(13): 3406-3415.

Table 2. Most used codons for each amino acid of Salt Stress Responsive genes in *Arabidopsis thaliana* (Ath) and *Solanum lycopersicum* (Sly) and their RSCU values (in parenthesis).

Amino acids	SOS1		BZIP		HVA22-like protein		CDPK1	
	Ath	Sly	Ath	Sly	Ath	Sly	Ath	Sly
Ala	GCT (1.72)	GCT (2.03)	GCT (2.55)	GCA (1.78)	GCA (2.00)	GCA (2.15)	GCT (2.53)	GCT (2.36)
Cys	TCT (1.20)	TGT (I.43)	TGC (1.33)	TGT/TGC (0.00)	TGC (2.00)	TGT/TGC (1.00)	TGT (1.71)	TGT (1.78)
Asp	GAT (1.35)	GAT (1.63)	GAT (1.37)	GAT (1.56)	GAT (1.25)	GAT (1.56)	GAT (1.37)	GAT (1.40)
Glu	GAA (1.44)	GAA (1.07)	GAG (1.31)	GAA (1.60)	GAA (1.23)	GAG (1.09)	GAG (1.15)	GAA (1.37)
Phe	TTT (1.10)	TTT (1.42)	TTC (1.17)	TTT (1.20)	TTC (1.40)	TTT (1.33)	TTC (1.04)	TTT/TTC (1.00)
Gly	GGA (1.47)	GGT (1.59)	GGA (1.80)	GGT (2.40)	GGA (1.75)	GGT/GGC/GGG (1.33)	GGA (2.06)	GGT (1.60)
His	CAT (1.36)	CAT (1.41)	CAT (1.50)	CAT (1.33)	CAC (1.50)	CAC (1.33)	CAT (1.09)	CAT (1.20)
Ile	ATT (1.46)	ATT (1.46)	ATT (2.00)	ATT (1.50)	ATT (1.29)	ATT (2.40)	ATT (1.16)	ATT (1.76)
Lys	AAA (1.22)	TTT (0.91)	AAA (1.14)	AAA (1.60)	AAG (1.43)	AAG (1.43)	AAA/AAG (1.00)	AAG (1.15)
Leu	CTT (1.87)	CTT (1.86)	CTC (1.82)	CTT (2.25)	CTT (2.53)	CTT (2.10)	TTG (1.56)	TTG/CTT (1.46)
Asn	ATT (1.24)	AAT (1.56)	AAC (1.33)	AAT (1.53)	AAT/AAC (1.00)	AAC (1.33)	AAT/AAC (1.00)	AAT (1.78)
Pro	CCA (1.66)	CCT (2.10)	CCT (1.91)	CCT/CCC (2.00)	CCA (2.00)	CCT (2.29)	CCA (1.66)	CCA (2.36)
Gln	CAA (1.32)	CAA (1.26)	CAA (1.21)	CAA/CAG (1.00)	CAA (1.33)	CAA (1.60)	CAG (1.07)	CAA (1.29)
Arg	AGA (1.70)	AGA (2.32)	AGG (1.83)	AGA (3.00)	CGT/CGA (2.00)	AGA (2.31)	AGA (1.88)	AGA (2.57)
Ser	TCT (1.56)	TCA (1.67)	TCT (1.81)	TCA (2.18)	TCC/TCA(1.64)	TCT (2.77)	TCA (1.69)	AGT (2.29)
Thr	ACT (1.36)	ACT (1.52)	ACT (1.56)	ACA (2.29)	ACC/ACA (1.33)	ACT (2.00)	ACT (2.40)	ACT (2.00)
Val	GTT (2.05)	GTT (1.51)	GTC (2.40)	GTT (1.71)	GTG (1.33)	GTT (2.50)	GTT (1.90)	GTT (1.73)
Tyr	TAT/TAC (1.00)	TAT (1.26)	TAT (1.33)	TAT (1.50)	TAC (1.14)	TAT (1.78)	TAT (1.23)	TAT (1.58)

Table 3. Comparison table for RNA secondary structures analysis in terms of Number of loops, Number of stems and Minimum Free Energy (MFE) for salt stress responsive genes in *Arabidopsis thaliana* and *Solanum lycopersicum*

1: SOS1

Plant name	Accession #	MFE (kcal/mol)	Number of loops	Number of stems
Arabidopsis thaliana	NM_126259.3	-907.89	145	143
Solanum lycopersicum	NM_001247769.3	-933.92	141	140

2: BZIP

Plant name	Accession #	MFE (kcal/mol)	Number of loops	Number of stems
Arabidopsis thaliana	NM_129624.4	-293.16	56	56
Solanum lycopersicum	NM_001246936.2	-93.96	22	22

3: HVA22-like protein

Plant name	Accession #	MFE (kcal/mol)	Number of loops	Number of stems
Arabidopsis thaliana	NM_119836.2	-159.56	21	21
Solanum lycopersicum	XM_004248672.2	-135.73	26	25

4: CDPK1

Plant name	Accession #	MFE (kcal/mol)	Number of loops	Number of stems
Arabidopsis thaliana	NM_101746.4	-375.65	76	76
Solanum lycopersicum	NM_001247653.2	-417.97	75	76