

CODON USAGE BIAS ANALYSIS FOR ABIOTIC STRESS GENES IN *ARABIDOPSIS THALIANA* AND *ORYZA SATIVA*

SYED QASIM SHAH, MUHAMMAD DIN, MUHAMMAD YOUNAS KHAN BAROZAI*
AND ABDUL KABIR KHAN ACHAKZAI

Department of Botany, University of Balochistan, Quetta, Pakistan

*Corresponding author: barozaikhan@gmail.com

Abstract

Codon usage bias (CUB) directly affects the expression level of protein. The protein expression also affects by abiotic stresses such as cold, drought, heat, salt, dehydration, heavy metals and oxidative stresses that reduce the productivity of plants. The current research is based to study the relationship of the CUB analysis for abiotic stress responding and housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa*. A total of 159 genes (including 138 abiotic stress resistant and 21 housekeeping genes) were subjected for the CUB analysis. Combinations of various bioinformatics tools were applied for the CUB analysis of abiotic stress resistant and housekeeping genes. Out of 159, the 92 abiotic stress resistant and 16 housekeeping genes showed similarities by more than 50% for the usage of codons in *Arabidopsis thaliana* and *Oryza sativa*. The most preferred codon usage among these stress resistant and housekeeping genes were also calculated. The CUB analysis and most preferred codons findings will be helpful to engineer the abiotic stress resistant crops through adjusting specific codon usage that enhance yield and productivity of significant plants under biotic and abiotic stresses.

Keywords: Abiotic stresses; Codon Usage Bias; plant productivity

Introduction

Codon usage bias (CUB) is the probability of a codons to encode an amino acid over the different codons. Some of the synonymous codons are utilized more frequently, are named as optimal or major codons while less frequently used ones are named as non-optimal or minor codons. Non optimal codons usually relate to less abundant transfer RNA than that of the optimal codons (Akashi, 2001). Code degeneracy support many amino acids to be encoded by multiple codons. The originating of codon preferences may be due to natural selection or mutation (Novoa & Pouplana, 2012). Codon preferences is sustained by the balance between mutational biases and natural selection for translational optimization while it remains unclear whether natural selection acts at the level of the rapidity or accuracy of messenger RNAs translation (Ermolaeva, 2001). Codon usage bias may be in terms of non-synonymous and synonymous mutation that may and may not cause change in the structure of amino acids respectively. Synonymous mutation in plants is thought to be the result of coadaptation between codon usage and transfer RNA abundance that leads to optimize the productivity of protein synthesis (Marais & Laurent, 2000). Codon usage bias varies from species to species, gene to gene in a genome as well as from site to site within a gene. Codons that provide efficient translation of an over-expressed transgene may fluctuate from the effective codons for an endogenous gene (Joshua & Kudla, 2011). Bias in codon utilization improve the reliability of protein synthesis in multicellular individuals (Marais & Laurent, 2000). Codon usage bias is influenced by biotic and abiotic stresses. Abiotic stress is the environmental factor including drought, heat, cold, freezing, water, metal, salt, osmotic and oxidative stress that change the expression level of a gene (Shinwari *et al.*, 1998; Narusaka *et al.* 1999;

Nakashima *et al.*, 2000). Plants show resistance to various stresses through stress resistant genes. Genes that positively affect stress tolerance, are identified as stress tolerant genes (STGs), which may code a transcription factor that control the expression of stress tolerant genes or a protein that give direct response to stress. Plants also response to the abiotic stress by hormones such as abscisic acid, jasmonic acid, ethylene and salicylic acid (Vandenbroucke, 2008). *Arabidopsis thaliana* because of its simple structure, is modal plant in molecular genetics therefore a wide study of stress tolerant genes is conducted (Salt, 2004). About 45% of all STGs are *Arabidopsis* genes, while nearly 50% of all STGs have been characterized using *Arabidopsis* as the transgenic species (Narusaka *et al.*, 2001; Vandenbroucke, 2008).

Codon usage influence the process of gene expression and the process that oversee codons arrangement remain unclear. There are many aspects that influence the selection of codons are codon autocorrelation, sequence of codon, richness of degeneracy of transfer RNAs, and coregulation of gene. Modification of the transfer RNA regulate the translation rate of codons and accordingly the expression levels of specific subgroups of genes. The modification of transfer RNA has an important function in the regulation of genes by promoting the expression ranks of the genes which are involved in cellular response. It shows the complication of matured cellular transfer RNA, which grips potential for the detection of new cellular regulatory tools (Narusaka *et al.*, 2003; Novoa & Pouplana, 2012).

When a plant is exposed to abiotic stress, the expression of many genes is altered to induce protection against the negative effects of the abiotic stress (Masood *et al.* 2005). It has now become clear that increased protection involves a multifaceted governing linkage which intermediates biochemical, physiological,

morphological and molecular changes, it is necessary to know these variations for knowing tolerance to abiotic stress in plants breeding (Kidokoro *et al.*, 2009; Vandenbroucke, 2008).

Current study is aimed to understand the Codon usage bias in terms of synonymous codon usage and their relationship with abiotic stress resistant genes in *Arabidopsis thaliana* and *Oryza sativa*. The main theme of this study is to find the most preferred codon for each Amino acids of 138 abiotic stress genes and 21 housekeeping genes (Sung *et al.*, 2010).

Materials and Methods

The basic steps of this research are as follows;

Identification of abiotic stress resistant and housekeeping genes: The abiotic stress resistant and housekeeping genes in *Arabidopsis thaliana* were found through literature survey. A detail survey is done for the abiotic stress resistant and housekeeping genes in *Arabidopsis thaliana*. A total of 138 abiotic stress genes and 21 housekeeping genes were selected after mining the data.

Sequence retrieval: The Sequences of abiotic stress resistant and housekeeping genes of *Arabidopsis thaliana* were retrieved by using the accession/ *Arabidopsis thaliana* Genome Initiative (AGI) number of the genes through nucleotide database of NCBI (National Center for Biotechnology Information). Using the BLASTn program (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) their homologous were found in *Oryza sativa* (Barozai *et al.*, 2008). The FASTA format of these genes were saved and utilized for more investigation.

Codon usage analysis: For the analysis of codon usage the FASTA sequence of the genes were subjected to codon usage investigation program (Stothard, 2000). The codon usage investigation program were used for analysis of different synonymous codons for the same amino acids that results in the use of synonymous codons with different rate for the same amino acid among which the codon having high value were selected and compared them for each amino acid of *Arabidopsis thaliana* and *Oryza sativa* (Barozai & Din, 2014). It is the key step of research and the results obtained in this step were saved for further analysis.

The relative synonymous codon usage (RSCU) standards for each of the codon in the sequences were calculated by using the following formula.

$RSCU = (\text{no. of codon used} / \text{total no. of codon}) \times \text{no. of amino acid}$

Results and Discussion

Abiotic stress resistant genes: A total of 138 abiotic stress and 21 housekeeping genes were found through literature survey in *Arabidopsis thaliana* and *Oryza sativa*. The detail description of these genes are provided in the Tables 1 and 2.

Codon usage bias (CUB) analysis: The CUB analysis for the 138 stress resistant genes and 21 housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa* revealed remarkable results. The frequently used codons in terms of Relative synonymous codon usage was calculated and summarized for the hundred and thirty eight selected abiotic stress resistant and twenty one housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa* for the stop codons and 20 amino acids.

Out of 138, the 92 abiotic stress resistant genes showed more than 50% similarity in the utilization of codon usage bias in *Arabidopsis thaliana* and *Oryza sativa*. While the 46 abiotic stress resistant genes showed less than 50% similarity in the usage of biased codons in *Arabidopsis thaliana* and *Oryza sativa*. The 16 (out of 21) housekeeping genes showed more than 50% while five showed similarity by less than 50% in the usage of codons in *Arabidopsis thaliana* and *Oryza sativa*. The most preferred codons were found in the 92 abiotic stress and 16 housekeeping genes, as given in Figure 1a and Figure 1b respectively. The comparison of most preferred codons of abiotic stress related and housekeeping genes are given in Table 3. The uniformity were observed in the utilization of ATG codon for methionine, TGG codon for tryptophan amino acid and GAT for asparagine. Mukhopadhyay *et al.*, 2008, described alike codon usage biased studies of housekeeping and tissue specific genes in *Oryza sativa* and *Arabidopsis thaliana*. These research recommend the non-uniform utilization of codon for abiotic stress tolerant genes among the two species of significant plants i.e., dicot, *Arabidopsis thaliana* and monocot, *Oryza sativa*.

More than half number of abiotic stress related and housekeeping genes that shows similarity by greater than 50% in the utilization of codons suggests the direct relationship among the CUB of abiotic stress and CUB of housekeeping genes, irrespective of the plant species. These result represents that plant may be planned for the abiotic stress resistance by the process of optimizing of codons. While the genes that show similarity by less than 50%, proposes the independency of abiotic stress genes and species of plants, suggest their role in stress management. Alike conclusions was also stated for the some other species of plants (Whittle *et al.*, 2007; Wang & Hickey, 2007; Mukhopadhyay *et al.*, 2008)

The abiotic stress related and housekeeping genes association showed a remarkable results for *Oryza sativa* and *Arabidopsis thaliana* (Khan *et al.*, 2011; Aman *et al.*, 2013). The usage of similar codons in terms of CUB is 76% and that of the different codons is 24% for housekeeping and abiotic stress resistant genes in *Arabidopsis thaliana* but in *Oryza sativa*, the ratio of similar and different codons is 62% and 38% respectively in terms of codon usage biased analyses for the housekeeping and abiotic stress tolerant genes.

In the *Arabidopsis thaliana*, similarity of codon usage bias is less than 50% proposes the uniform patterns of codon utilization under normal and any type of stress situations. This type of outcomes plainly show the dependency of codon utilization on plant species during normal and abiotic stress situations. These results reflects

the previously reported research works (Whittle *et al.*, 2007; Wang & Hickey, 2007; Mukhopadhyay *et al.*, 2008). These results also are the platform for the new discoveries by applying bioinformatics (Barozai *et al.*, 2012a; 2012b; Barozai & Wahid, 2012; Barozai, 2012c).

Characterization of stress resistant and housekeeping genes: The 138 abiotic stress and 21 housekeeping genes were also characterized in terms of codon usage bias similarities percentage (Table 4). Out of 138 abiotic stress resistant genes, 4 lies in 1-25% similarity, 42 lies in 26-50%, 81 lies in 51-75% and 11 lies in the 76-100% similarity. There are 92 genes that show similarity by greater than 50% similarities in the utilization of codon

usage bias, same trend of more than 50% similarities is observed for 16 genes out of 21 housekeeping genes while 5 genes show similarity by less than 50%.

Conclusions

This study indicates the multidimensional correlations among the codon usage bias and abiotic stress resistant and housekeeping genes for the significant monocot and dicot plants. In case of housekeeping genes uniform correlation with CUB is observed for monocot and dicot plant. Stress tolerance capacity of the plants can be improved by the phenomena of codon optimization.

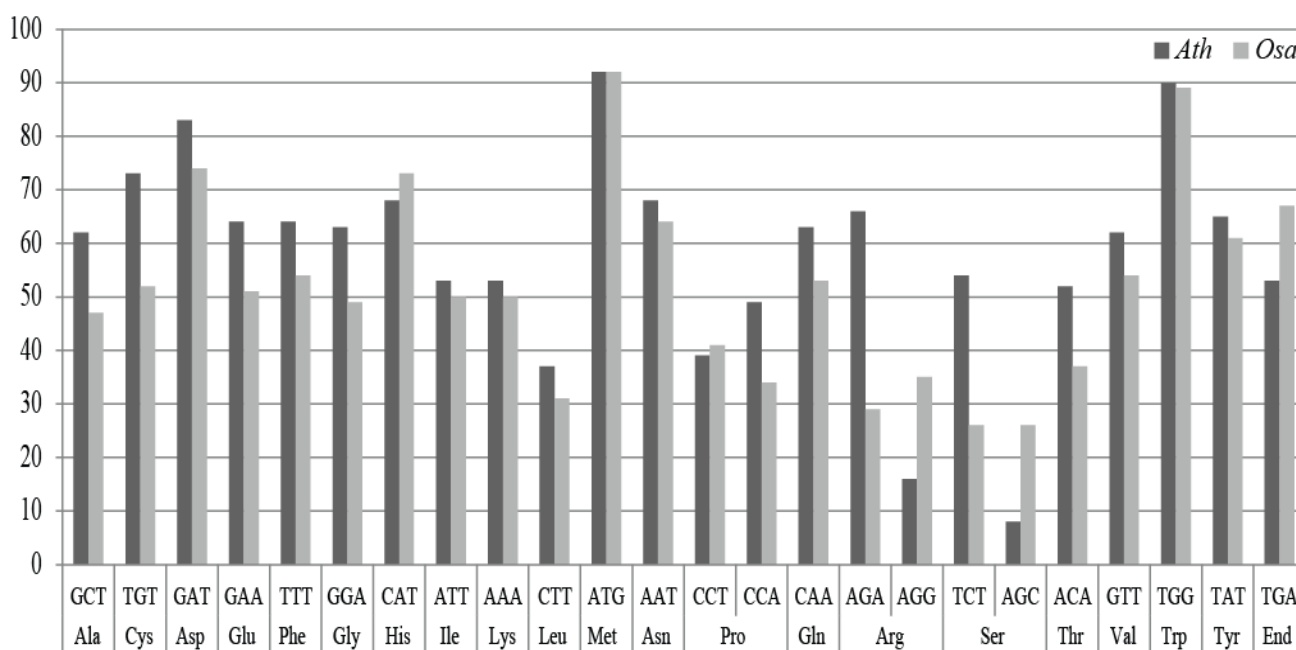


Fig. 1a. Most Preferred Codon for each Amino acid of Abiotic Stress Resistant Genes in *Arabidopsis thaliana* and *Oryza sativa*

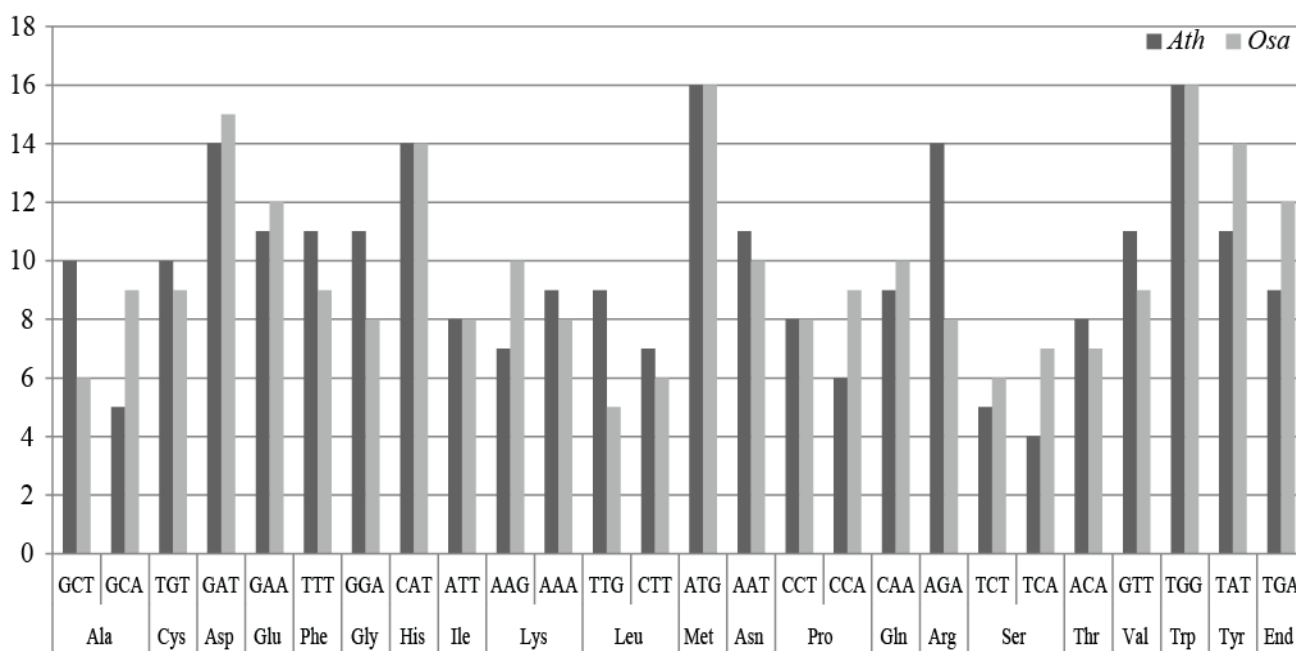


Fig. 1b. Most Preferred Codon for each Amino acid of Housekeeping Genes in *Arabidopsis thaliana* and *Oryza sativa*

Table 1. Abiotic Stress Resistant Genes in *Arabidopsis thaliana* and *Oryza sativa*

S/No	GENES	AGI	Molecular Function	Stress Type
1	MBF1c	AT3G24500	Multi-protein bridging factor 1c	S, O, H
2	MYB60	AT1G08810	transcription factor	D
3	ICE1	AT3G26744	transcription factor	C, F
4	CaMBP25	AT2G41010	calmodulin-binding protein	O, S
5	SRK2C	AT1G78290	SNF1-related protein kinase 2	D, O
6	GOLS2	AT1G56600	galactinol synthase	D
7	TPS1	AT1G78580	trehalose-6-P synthase	D
8	AVP1	AT1G15690	vacuolar H(+)-pyrophosphatase	D, S
9	NHX1	AT5G27150	Na ⁺ /H ⁺ antiporter	S
10	tAPX	AT1G77490	ascorbate peroxidase	Ox, S, O, H
11	AOX1a	AT3G22370	alternative oxidase	C, C/F
12	RCI3	AT1G05260	cell wall peroxidase	D, S
13	HSP101	AT1G74310	heat shock protein 101	H
14	HSP17.6a	AT5G12030	heat shock protein	D, S
15	NCED3/STO1	AT3G14440	Salt Tolerant 1, protein binds to a Myb transcription factor	D, S
16	RZ-1a	AT3g26420	glycine rich RNA binding protein	C, F
17	SRL1	AT5g37370	splicing factor	S
18	SOS1	AT2G01980	plasma membrane Na ⁺ /H ⁺ antiporter	S
19	LEA5	AT4G02380	late embryogenesis abundant 5	Ox
20	GLB1	AT2G16060	non-symbiotic haemoglobin	Ox
21	BCB	AT5G20230	blue copper-binding protein	Ox
22	DREB2A	AT5G05410	drought-responsive element binding protein 2A	H
23	HSFA2	AT2G26150	heat shock factor A2	Ox, H
24	RHL41/ZAT12	AT5G59820	responsive to high light 41	Ox, H
25	CBL1	AT4G17615	calcineurin B-like protein	abiotic stresses
26	PCS1	AT5G44070	Phytochelatin synthesis	M
27	AAO	AT5G21100	Ascorbate oxidase	S, Ox
28	ALDH3I3	AT4G34240	Aldehyde dehydrogenase	D, S, Ox, M
29	ALDH7B4	AT1G54100	Aldehyde dehydrogenase	D, S
30	APX1	AT1G07890	Ascorbate peroxidase	S, O, C/F, H,
31	FRO2	AT1G01580	Ferric chelate reductase responsible	M
32	GI-3	AT1G22770	Gigantean	Ox
33	GPX3	AT2G43350	Glutathione peroxidase	D, O
34	SOD/MSD1	AT3G56350	Mn superoxide dismutase	S
35	VTC1	AT2G39770	Vitamin C defective 1, encodes mannose-1-pyrophosphatase	H
36	VTC2	AT4G26850	Vitamin C defective 2, encodes mannose-1-pyrophosphatase	H
37	SOD/CSD2	AT2G28190	miRNA resistant form of chloroplastic Cu/Zn superoxide dismutases	Ox, M, H
38	PARP2	AT4G02390	Poly(ADP-ribose) polymerase	D, H, Ox
39	UVH6	AT1G03190	UV-sensitive mutant	H
40	PLDalfa1	AT3G15730	Phospholipase Alfa, modulation of COR genes	C/F
41	PLDdelta	AT4G35790	Phospholipase Delta	C/F
42	LTL1	AT3G04290	GDSL-type lipase	S
43	ABA2/GIN	AT1G52340	Cytosolic short-chain dehydrogenase/reductase	S
44	CYP707A3	AT5G45340	ABA 8'-hydroxylase activity	D
45	COR15a	AT2G42540	LEA	C/F
46	DHN (LTI29/ERD10+LTI30)	AT1G20450 and AT3G50970	Dehydrin	C/F

Table 1. (Cont'd)

S/No	GENES	AGI	Molecular Function	Stress Type
47	DHN (RAB18)	AT5G66400	Dehydrin	C/F
48	DHN (COR47)	AT1G20440	Dehydrin	C/F
49	ERD15	AT2G41430	Early responsive to dehydration	D, C/F
50	F9E10.5	AT1G75100	auxilin-like gene	M
51	DJA2	AT5G22060	DNAj domain containing molecular chaperones	H
52	DJA3	AT3G44110	DNAj domain containing molecular chaperones	H
53	ProDH	AT3G30775	Proline dehydrogenase	S, C/F
54	DHS	AT5G05920	Deoxyhypusine synthase, eIF5a activation	D
55	LOS4/CRYOPHITE	AT3G53110	DEAD-Box RNA Helicase and has RNA-dependent ATPase activity	C/F
56	STRS1	AT1G31970	Stress Response Suppressor 1, DEAD-box RNA helicase	S, O, H
57	STRS2	AT5G08620	Stress Response Suppressor 2, DEAD-box RNA helicase	S, O, H
58	GRP2	AT4G13850	Glycine-rich RNA binding protein	S, C/F
59	FTA	AT5G40280	Farnesyltransferase	D
60	SDIR1	AT3G55530	Salt and Drought-Inducible RING finger E3 ligase	D
61	ORE9/MAX2	AT2G42620	the F-box leucine-rich repeat family	Ox
62	PMSR4	AT4G25130	Peptide methionine sulfoxide reductase	Ox
63	ABI1	AT4G26080	Genes Involved in ABA signal transduction.	D
64	HAB1	AT1G72770	Genes Involved in ABA signal transduction.	D
65	GSK1	AT1G06390	Homologue of GSK3/shaggy-like protein kinase	D, S
66	AHK1/ATHK1	AT2G17820	Cytokinin receptor histidine kinase	D
67	MKK2	AT4G29810	MAPKK	S, C/F
68	MKK9	AT1G73500	MAPKKK	S, O
69	NDPK2	AT5G63310	NDP kinases	S, C/F, Ox
70	PP2CA	AT3G11410	Protein phosphatase 2C	C/F
71	RAB7	AT1G49300	Small GTPase, RAB family	S, O
72	RGS1	AT3G26090	Regulation of G-protein signalling	D
73	PP7	AT5G63870	Calmodulin-binding protein phosphatase PP7	H
74	AHK2	AT5G35750	Cytokinin receptor histidine kinase	D, S
75	AHK3	AT1G27320	Cytokinin receptor histidine kinase	D, S
76	TOR	AT1G50030	Target of rapamycin	S
77	CIPK23	AT1G30270	CBL-interacting protein kinase	D
78	CPK23	AT4G04740	Calcium-dependent protein kinase	D, S
79	CBL9		Calcineurin-B-like protein	D
80	EIN2/ORE3	AT5G03280	ethylene mutant	Ox
81	CTR1	AT5G03730	Serine/threonine/tyrosine kinase (Constitutive Transcriptional Response)	S
82	GLI1	AT1G80460	Glycerol kinase	D, S, O, Ox, C/F
83	ABF2/AREB1	AT1G45249	ABA RE binding factor	D
84	ABF3	AT4G34000	ABA RE binding factor	D
85	ABF4	AT3G19290	ABA RE binding factor	S
86	ABI3	AT3G24650	Transcription factor	C/F
87	ABO1/ELO2	AT5G13680	Transcription Elongator complex subunit	D, Ox
88	ANAC002	AT1G01720	Transcription Factor with NAC domain	D
89	ANAC019	AT1G52890	Transcription Factor with NAC domain	D
90	ANAC055/NAC3	AT3G15500	Transcription Factor with NAC domain	D
91	ANAC072/RD26	AT4G27410	Transcription Factor with NAC domain	D
92	CBF1 / DREB1B	AT4G25490	Transcription Factor (Cold binding factor, Drought-Responsive Element Binding protein)	C/F
93	CBF2 / DREB1C	AT4G25470	Transcription Factor (Cold binding factor, Drought-Responsive Element Binding protein)	C/F, D, S

Table 1. (Cont'd)

S/No	GENES	AGI	Molecular Function	Stress Type
94	CBF3 / DREB1A	AT4G25480	Transcription Factor (Cold binding factor, Drought-Responsive Element Binding protein)	D, S, C/F
95	CBF4/DREB1D	AT5G51990	Transcription Factor (Cold binding factor, Drought-Responsive Element Binding protein)	D, C/F
96	CPL1/FRY2	AT4G21670	Transcriptional repressor, C-terminal phosphatase-like	S
97	HD2C	AT5G03740	Histone deacetylase	D, S
98	HSF1	AT4G17750	Transcription Factor, Heat shock factor 1	H
99	HSF3	AT5G16820	Transcription Factor, Heat shock factor 3	H
100	MBF1a	AT2G42680	Multiprotein bridging factor 1a	S
101	MYB2	AT2G47190	Transcription factors	O
102	MYC2	AT1G32640	Transcription factors	O
103	SZF1	??	Transcription factor CCCH-type zinc finger proteins	S
104	SZF2	??	Transcription factor CCCH-type zinc finger proteins	S
105	SHN1	AT1G15360	Transcription factor Shine-clan AP2	D
106	STZ/ ZAT10	AT1G27730	Transcription Factor, Cys2/His2-Type Zinc-Finger Proteins	S, O, H
107	XERICO	AT2G04240	Transcription factor, RING-H2 zinc finger	D
108	ZAT12	AT5G59820	Transcription factor, Zn-finger TF	S, O, H
109	ZAT7	AT3G46090	Transcription factor, Zn-finger TF	Ox
110	HRD/HARDY	AT2G36450	Transcription Factor, AP2/ERF-like	D, S
111	NF-YB1	AT2G38880	Transcription Factor, Plant nuclear factor Y	D
112	DREB2C	??	Transcription Factor (Drought-Responsive Element Binding protein)	H
113	ALS3	Unmapped	Aluminum sensitive 3, ABC transporter	M
114	ATM3	AT5G58270	ABC transporter	M
115	BOR1	AT2G47160	Boron transporter	M
116	MRP5	AT1G04120	ABC transporter	D
117	MTP11	AT2G39450	Golgi-localized manganese transporter that is involved in Mn tolerance	M
118	MTP3	AT3G58810	Zinc transporter (ZAT) family. Contributes to basic cellular Zn tolerance	M
119	PDR12	AT1G15520	ABC transporter	M
120	PDR8	AT1G59870	ABC transporter	M
121	PIP1;4	AT4G00430	Plasma membrane aquaporin	C/F
122	PIP2;5	AT3G54820	Plasma membrane aquaporin	C/F
123	SULTR1;2	AT1G78000	Sulfate transporter	M
124	ZIF1	AT5G13740	Zn sequestration	M
125	SBP1	AT1G45976	Selenium binding protein 1, S-ribonuclease	M
126	ALX8	Unmapped	Altered expression of APX2	D
127	ESK1	AT3G55990	Unknown protein	C/F
128	GPP2	AT5G57440	Haloacid dehalogenase-like hydrolase protein	S, O, Ox
129	ORE1	Unmapped		Ox
130	TRG1	Unmapped	thermoinsinhibition-resistant germination 1	H
131	SPS	Unmapped	Sucrose phosphate synthase	C/F
132	CHYB	AT4G25700	Beta-carotene hydroxylase	
133	CESA8/IRX1/LEW2	AT4G18780	Cellulose synthase	D, S, O
134	HAL3A	AT3G18030	Flavin mononucleotide flavoprotein (phosphopantothenoylecysteine decarboxylase activity)	S, O
135	TMAC2	AT3G02140	Two or more ABREs-containing gene 2	S
136	PCR1	AT1G54560	Myosin like protein	M
137	PHYA (ars4ars5)	AT1G09570	Cytoplasmic red/far-red light photoreceptor	M
138	RCD1	AT1G32230	Radical-induced cell death 1	Ox

Abbreviations; C: cold, F: freezing, D: drought, H: heat stress, O: osmotic stress, Ox: oxidative stress, S: salt stress, C/F: cold / freezing stress, M: metal stress.

Table 2. Housekeeping genes in *arabidopsis thaliana* and *oryza sativa*

S. #.	Gene	Accession no.	Molecular Function
1	PP2AA3	AT1G13320	regulation of phosphorylation
2	helicase	AT1G58050	biological process
3	PPR gene	AT1G62930	5' processing of mitochondrial mRNAs
4	SAND family	AT2G28390	Expressed during 14 growth stages
5	expressed protein	AT2G32170	Growth stages, plant structures,
6	polypyrimidine track-binding protein	AT3G01150	Encodes polypyrimidine tract-binding protein homologs
7	ubiquitin transferase	AT3G53090	Encodes a ubiquitin-protein ligase containing a HECT domain
8	expressed protein	AT4G26410	Biological process
9	UBC9	AT4G27960	Ligase activity and ubiquitin-dependent protein catabolic process
10	expressed protein	AT4G33380	Biosynthesis of plant cell wall
11	TIP41-like	AT4G34270	TIP41-like protein, Elemental activities, such as catalysis or binding
12	mitosis protein YLS8	AT5G08290	Encodes Dim1 homolog
13	expressed protein	AT5G12240	Protein synthesis and degradation, RNA and sugar metabolism,
14	F-box protein	AT5G15710	Growth and Developmental Stage
15	clathrin adaptor complex subunit	AT5G46630	Intracellular protein transport, vesicle-mediated transport
16	UBQ10	AT4G05320	Protein modification, aging, salicylic acid stimulus, growth stages
17	ACT	AT3G18780	Encodes actin protein growth of root hairs
18	TUB	AT5G62690	GTPase activity, response to cadmium ion, salt and cold stress,
19	EF-1 α	AT1G07920	Cd ⁺ binding, translation elongation factor activity, response to Cd ⁺
20	GAPDH	AT1G13340	Regulator of Vps4 activity in the MVB pathway protein
21	UFO	AT1G30950	Whorled pattern, floral meristem, activation of APETALA3 and PISTILLATA, regulate AP3 expression

Table 3. Comparison of most preferred codons calculated in term of RSCU for 138 abiotic stress resistant and 21 housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa*

Amino Acids	<i>Arabidopsis thaliana</i> Abiotic Stress Genes	<i>Arabidopsis thaliana</i> Housekeeping Genes	<i>Oryza sativa</i> Abiotic Stress Genes	<i>Oryza sativa</i> Housekeeping Genes
Ala	GCT	GCT	GCT	GCA
Cys	TGT	TGT	TGT	TGT/TGC
Asp	GAT	GAT	GAT	GAT
Glu	GAA	GAA	GAA	GAA
Phe	TTT	TTT	TTT	TTT/TTC
Gly	GGA	GGA	GGA	GGA
His	CAT	CAT	CAT	CAT
Ile	ATT	AAT	ATT	AAT/ATC
Lys	AAA	AAA	AAA	AAG
Leu	CTT	TTG	CTT	CTT
Met	ATG	ATG	ATG	ATG
Asn	AAT	AAT	AAT	AAT
Pro	CCA/CCA	CCT	CTT	CCA
Gln	CAA	CAA	CAA	CAG
Arg	AGA/AGG	AGA	AGG	AGA
Ser	TCT/AGC	TCT	TCT/AGC	TCA
Thr	ACA	ACA	ACA	ACA
Val	GTT	GTT	GTT	GTT
Trp	TGG	ACA	TGG	ACA
Tyr	TAT	TGG	TAT	TGG
End	TGA	TGA	TGA	TGA

Table 4. Characterization of the abiotic stress resistant and housekeeping genes in terms of similarity in the usage of codons

Similarity % age	No. of Abiotic Stress Genes	Type of Abiotic Stress Resistant Genes	No. of Housekeeping Genes	Type of Housekeeping Genes
1-25	4	CaMBP25, CYP707A3, SZF1, ALS3,	00	None
26-50	42	ICE1, SRK2C, GOLS2, HSP101, BCB, RHL41/ZAT12, PCS1, AAO, ALDH7B4, APX1, VTC2, PLD delta, LTL1, ABA2/GIN, DHN (LTI29/ERD10+LTI30), DHN (COR47), STRS1, STRS2, HAB1, AHK3, ABF2/AREB1, ABF3, ANAC019, ANAC055/NAC3, ANAC072/RD26, CBF1 / DREB1B, CBF2 / DREB1C, CBF4/DREB1D, HD2C, MYC2, SHN1, ZAT7, DREB2C, MRP5, MTP11, MTP3, SULTR1;2, SBP1, ALX8, CHYB, CESA8/IRX1/LEW2, TMAC2,	05	Expressed protein (AT5G12240), F-box protein (AT5G15710), ACT (AT3G18780), TUB (AT5G62690), UFO (AT1G30950)
51-75	81	MBF1c, MYB60, TPS1, AVP1, tAPX, AOX1a, RCI3, HSP17.6a, NCED3/STO1, RZ-1a, SRL1, SOS1, LEA5, GLB1, DREB2A, HSFA2, CBL1, FRO2, GPX3, SOD/MSD1, SOD/CSD2, PARP2, UVH6, PLDalfa1, DHN (RAB18), ERD15, F9E10.5, DJA2, DJA3, ProDH, LOS4/CRYOPHITE, STRS1, GRP2, FTA, SDIR1, ORE9/MAX2, PMSR4, ABI1, GSK1, AHK1/ATHK1, MKK2, MKK9, NDPK2, PP2CA, RAB7, RGS1, PP7, AHK2, CIPK23, CBL9, EIN2/ORE3, CTR1, GLI1, ABF4, ABI3, ANAC002, CBF3 / DREB1A, CPL1/FRY2, HSF1, HSF3, MBF1a, MYB2, SZF2, STZ/ ZAT10, XERICO, ZAT12, HRD/HARDY, NF-YB1, BOR1, PDR8, PIP1;4, PIP2;5, ZIF1, ESK1, GPP2, ORE1, TRG1, SPS, HAL3A, PCR1, RCD1,	11	PP2AA3 (AT1G13320), PPR gene (AT1G62930), SAND family (AT2G28390), polypyrimidine track-binding protein (AT3G01150), ubiquitin transferase (AT3G53090), expressed protein (AT4G26410), UBC9 (AT4G27960), expressed protein (AT4G33380), clathrin adaptor complex subunit (AT5G46630), UBQ10 (AT4G05320), GAPDH (AT1G13340)
76-100	11	NHX1, ALDH3I3, GI-3, VTC1, COR15a, TOR, CPK23, ABO1/ELO2, ATM3, PDR12, PHYA (ars4ars5),	05	Helicase (AT1G58050), expressed protein (AT2G32170), TIP41-like (AT4G34270), mitosis protein YLS8 (AT5G08290), EF-1 α (AT1G07920)

References

- Akashi, H. 2001. Gene expression and molecular evolution. *Curr. Opin. Genet. Dev.*, 11: 660-666.
- Aman, S., M. Iqbal, S. Abbas, S. Banaras, M. Awais, I. Ahmad, Z.K. Shinwari and S.N. Shakeel. 2013. Molecular and comparative analysis of newly isolated beta-tubulin partial gene sequences from selected medicinal plants. *Pak. J. Bot.*, 45(2): 507-512.
- Barozai, M.Y.K and M. Din. 2014. The Relationship between Codon Usage Bias and Cold Resistant Genes. *Pak. J. Bot.*, 46(3): 823-826.
- 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. and H.A. Wahid. 2012. Insilco 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., 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. & Biochem.*, 46(8-9): 739-756.
- Ermolaeva, M.D. 2001. Synonymous Codon Usage in Bacteria. *Curr. Issues. Mol. Bio.*, 3(4): 91-97.
- Joshua, B.P and G. Kudla. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1): 32-42.
- Khan, A.L., M. Hamayun, S.A. Khan, Z.K. Shinwari, M. Kamaran, S.M. Kang, J.G. Kim and I.J. Lee. 2011. Pure

- culture of *Metarhizium anisopliae* LHL07 reprograms soybean to higher growth and mitigates salt stress. *World J. Microb. Biotech.*, 28(4): 1483-1494.
- Kidokoro, S., K. Nakashima, Z.K. Shinwari, K. Shinozaki and K.Y. Shinozaki. 2009. The phytochrome-interacting factor PIF7 negatively regulates *DREB1* expression under circadian control in *Arabidopsis*. *Plant Physiol.*, 151(4): 2046-2057.
- Marais, G. and D. Laurent. 2000. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *J. Mol. Evol.*, 52: 275-280.
- Masood, S., Y. Seiji, Z.K. Shinwari and R. Anwar. 2005. Mapping quantitative trait loci (QTLs) for salt tolerance in rice (*Oryza sativa*) using RFLPs. *Pak. J. Bot.*, 36(4): 825-834.
- 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 Mol. Biol.*, 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 J.*, 34(2): 137-149.
- Narusaka, Y., K. Nakashima, Z.K. Shinwari, K. Shinozaki and K. Shinozaki. 2001. Promoter analysis of *DREB1* genes encoding transcription factors involved in cold responsive gene expression in *Arabidopsis*. *Plant & Cell Physiol.*, 42: 94.
- Narusaka, Y., Z.K. Shinwari, K. Nakashima, K. Yamaguchi-Shinozaki and K. Shinozaki. 1999. The roles of the two CIS-acting elements, DRE and ABRE in the dehydration, high salt and low temperature responsive expression of the rd29A gene in *Arabidopsis thaliana*. *Plant & Cell Physiol.*, 40: 91.
- Novoa, E.M. and L.R. de Poupiana. 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics*, 28(11): 574-581.
- Salt, D. 2004. Update on plant ionomics. *Plant Physiol.*, 136: 2451-2456.
- 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.
- Stothard, P. 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28: 1102-1104.
- Sung, M.H., S.C. Bahn, A. Lyu, H.S. Jung and J.H. Ahn. 2010. Identification and testing of superior reference genes for a starting pool of transcript normalization in *Arabidopsis*. *Plant Cell Physiol.*, 51(10): 1694-1706.
- Vandenbroucke, 2008. Role for hydrogen peroxide during abiotic and biotic stress signaling in plants. (*Doctoral dissertation, Ghent University, Belgium*).
- 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. Genderspecific selection on codon usage in plant genomes. *BMC Genomics*, 8: 169-179.

(Received for publication 15 January 2016)