

Distinct Structural and Functional Characteristics of Stress-Related Genes of Different Plants Revealed by *In-silico* Analysis

Ashraful Islam Bhuiya¹, and Abul B.M.M.K. Islam^{2*}

¹Institute of Nuclear Medicine & Allied Sciences, Dhaka Medical College Hospital Campus, Bangladesh Atomic Energy Commission, Dhaka-1000, Bangladesh

²Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh

***Corresponding author:** Dr. Abul Bashar Mir Md. Khademul Islam, Assistant Professor, Department of Genetic Engineering & Biotechnology, University of Dhaka, Science Complex Building, Dhaka 1000, Bangladesh, Tel: +880-2-9661900 Extn. 7825; E-mail: khademul@du.ac.bd

Received Date: 7th April 2014

Accepted Date: 20th August 2014

Published Date: 25th August 2014

Citation: Bhuiya AI, Islam ABMMK (2014) Distinct Structural and Functional Characteristics of Stress-Related Genes of Different Plants Revealed by In-silico Analysis. Enliven: Bioinform 1(2): 004.

Copyright: © 2014 Dr. Abul Bashar Mir Md. Khademul Islam. This is an Open Access article published and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

Plants always have to fight against various environmental stress conditions like cold, drought, salinity, submergence, etc. The prime target of recent research in plant biology is to unveil the intricate series of events in responses and adaptation to different stress conditions. Sufficient *in-silico* computational studies are yet to be done to distinguish the stress related genes from the non-stress related ones. As common mechanisms of stress responses exist among different plants, we sought to identify the general structural and functional features that may be hidden in stress related genes of different plant species. We assumed that these features in stress-related genes might be different from non stress related genes. One hundred and sixty stress-responsive genes from five different plant species were studied. Computational and bioinformatics studies were done to determine several structural properties like length of gene, exon, intron, UTRs as well as to identify overrepresented sequence motif and enrichment of gene ontology (GO) functions. The UTRs of stress related genes were found to be significantly different from non-stress related genes and a “G-C” rich small sequence motif was found to be associated significantly with stress genes. Key biological processes like small GTPase mediated signal transduction, cellular components like thylakoid and molecular functions like oxidoreductase activity are significantly enriched for stress related genes. Further studies are required to identify more stress specific features of plant stress genes which may help to establish a computational model for detecting stress related genes from various gene lists.

Keywords

Plant stress; Bioinformatics; UTRs; Motif; Gene ontology; Enrichment

Introduction

As plant cannot migrate from one place to another, harsh environmental conditions can be an important cause of mortality for plants. Environmental stresses can be biotic caused by different plant pathogens or abiotic such as cold, drought, salinity, submergence, heavy metals, radiation, etc. These stresses have great influence on the evolution of plant species and also have detrimental effects on plant growth and agricultural productivity [1,2]. Nowadays, due to the abiotic stresses the estimated gap between the attainable and actual yields of crops is 40 – 50% [http://www.isaaa.org/]. Thus, introduction of crop varieties with enhanced tolerance to environmental stresses and sustainable growth rate under suboptimal conditions are the crucial objectives in modern agriculture.

Using *Arabidopsis thaliana* and *Oryza sativa* (rice) as model systems, various genes were over expressed in these which led to the identification of stress tolerant genes and transcription factors [3,4]. Many other studies have also been done with the model plant *Arabidopsis thaliana* to elucidate the biochemical pathways of stress perception, signal transduction and adaptive responses [5-8].

Functional basis of stress tolerance should be explained on the basis of molecular mechanism and energetic vacillation. The mechanistic viewpoint of stress tolerance focuses on the similarities between cellular responses to different types of stresses [9]. To perceive environmental stresses and to respond to them, plants have evolved the mechanisms of complex signaling crosstalk such as interactive and antagonistic actions of different phytohormones like salicylic acid (SA), jasmonic acid (JA), ethylene (ET), abscisic acid (ABA), etc [10]. They regulate the prophylactic responses against both biotic and abiotic stresses. Again the generation of reactive oxygen species (ROS) has been proposed to be a common response in different stress conditions [10].

Plants produce energetic resources to activate the mechanisms of stress tolerance and survival. These metabolic shifts of energy reallocation represents a common response occurred under different adverse conditions [1,11,12]. Recently it was shown that modulation of cellular energy homeostasis and increased pool of NAD⁺ and NADH may play role to improve the yield of crop in environmental stress conditions [13,14].

Stress resistance traits that are functionally correlated with different stress mechanisms have been identified by quantitative genetic studies [11,15,16]. The complex mechanisms of stress perception, signal transduction and intonation of gene expression in stress environment have partially been uncovered by functional genomics study [17]. It was found that in transgenic rice (*Oryza sativa*), stress-responsive transcription factor SNAC1 over expression enhance drought resistance significantly [4]. Stress response mechanisms bring great changes in global gene expression, manner of protein modification and compositions of different metabolites [18]. Recently non-coding RNA has been found to be involved in stress response mechanisms of plants [19]. In last decade, it was revealed that the expression of different but overlapping gene suits are regulated by both biotic and abiotic stresses [20]. Some heat-shock proteins are generally stimulated as a common response to various stress environments [21,22]. Again, DREB transcription factors and phytochrome abscisic acid (ABA) have been identified as shared components in drought, salinity and unusual temperature responsive pathways in *Arabidopsis* model system [23,24].

The existence of some genes associated with general stress responsive mechanisms has been discovered by extensive study from the viewpoint of the cell physiology [25,26], evolutionary biology [1,12] and most importantly biotechnology [27-29]. The elucidation of the complex biochemical networks and structural properties of these stress responsive genes may provide targets that lead to the production of engineered stress resistant plant species.

Sufficient computational studies have not yet been done to identify significant characteristics of stress responsive genes that can differentiate them from non-stress related genes. Only a few studies were carried out to discover the stress responsive DNA regulatory motifs in *Arabidopsis thaliana* [30,31]. Due to the insufficient data on general structural properties of plant stress related genes, no computational method could be devised to predict stress related genes. For these reasons, laborious and cumbersome wet lab analyses have to be done to identify even a single stress related gene. In this study bioinformatics and computational analyses were performed with stress and non-stress related genes from five different plant species (*Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Solanum lycopersicum* and *Glycine max*) to identify generalized structural properties (gene length, exon length, motifs, length of 5' and 3' untranslated regions) of stress responsive gene that will distinguish them from non-stress related genes. This effort may be helpful to develop tools to identify stress related genes *in silico*. We have examined the Gene Ontology (GO) annotations in the group of stress related genes of these plant species to delineate the trends in the biology of stress responses.

Materials and Methods

Datasets

List of stress related genes were obtained from Plant Stress Gene Database (<http://cbb.jnu.ac.in/stressgenes/frontpage.html>) [32]. A total of 160 stress related gene sequences from five plant species (*Arabidopsis thaliana*, *Glycine max*, *Oryza sativa*, *Solanum lycopersicum* and *Zea mays*) (out of available 259 stress related genes from 11 plant species), were used in this study. The gene sequences, both stress and non-stress, were downloaded through the Biomart portal of Ensembl Plants, release - 19 (<http://plants.ensembl.org/biomart/martview>). The datasets used in this study are *Arabidopsis thaliana* (TAIR10 GCA_000001735.1 2010-09-TAIR), *Glycine max* (V1.0 GCA_000004515.1 2012-07-JGI), *Oryza sativa* (MSU6 GCA_000005425.2 2009-01-MSU), *Solanum lycopersicum* (SL2.40 GCA_000188115.1 2011-04-ITAG), and *Zea mays* (AGPv3 2010-01-MaizeSequence). The stress related genes were subtracted from entire genome dataset and the remaining data were used as negative dataset. There are 160 stress related genes (for detail list see supplementary Table S1) are included in positive dataset of which 33 from *Arabidopsis thaliana*, 55 from *Glycine max*, 9 from *Oryza sativa*, 26 from *Solanum lycopersicum* and 37 from *Zea mays*; whereas 256672 genes were included in negative dataset, of which 34259 from *Arabidopsis thaliana*, 56709 from *Glycine max*, 65518 from *Oryza sativa*, 34689 from *Solanum lycopersicum* and 65497 from *Zea mays*.

Length Analysis

Genomic location information of features like genes, exons and UTRs for both positive and negative dataset was obtained from the Ensembl Plants databases using Biomart portal. Their lengths were calculated using in house Perl script. For statistical significance, 1000 random sets for each feature, consisting of 160 members, were produced from negative dataset. The Z score and p-value of significance were calculated from these datasets using R statistical programming [33,34]. Distribution of both 5' and 3' UTR length are represented in Box-Whisker plot. Significance of difference of UTR length between stress related genes and non-stress related genes was calculated using Wilcoxon test.

Motif Analysis

MEME (Multiple Em for Motif Elicitation) [35,36] package (version 4.9.1) was used to identify the significantly over-represented motif in stress related genes (positive dataset). Thousands of random datasets consisting of 160 gene sequences each were produced from negative dataset and searched for similar motif using the position weight matrix (PWM) of MEME and STORM program [37] specifying the p-value cut off 0.00001. Z-score and p-value of significance were calculated using the random dataset ("expected") and the positive dataset ("observed") with R statistical programming [33,34].

Functional Enrichment Analysis

Functional annotation of plant stress-related genes is based on Gene Ontology (GO) (Consortium, 2000; <http://www.geneontology.org>) [38] extracted from Ensembl Plant (release - 19). Accordingly, all genes are classified into three ontology categories (i) biological process (BP), (ii) cellular component, (CC) and (iii) molecular function (MF) and pathways when possible. We considered only those GO pathway categories that have at least 9 genes annotated. We used Gitools [39] for enrichment analysis using non-stress related genes as background, and for heatmap generation. Resulting p-values were adjusted for multiple testing using the Benjamin and Hochberg's method of False Discovery Rate (FDR).

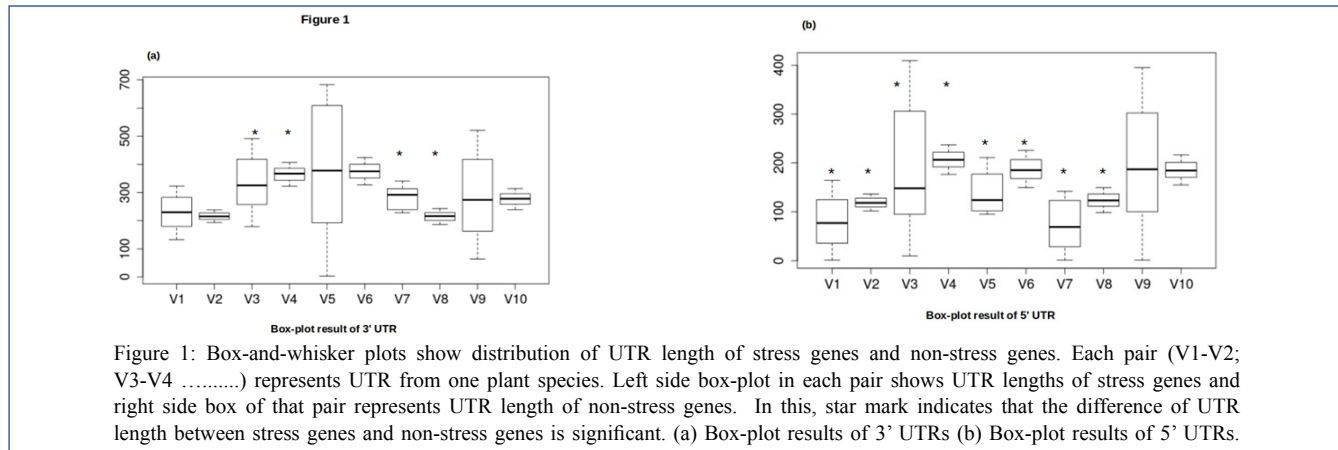
Results and Discussion

Results

The length of UTRs of stress related genes differ significantly from non-stress related genes in different plant species

Genes encode proteins which are the key functional components in different cellular mechanisms. Upstream and downstream sequences of genes play role in the regulation of gene expression. We calculated the length of stress related genes, their exons, 5' and 3' UTRs, and compared those to non-stress related genes to determine if any significant difference exists. For gene length and exon length, no statistically significant difference was obtained between stress and non-stress related genes. But z-score and p-value confirmed the significant difference in 5' and 3' UTR lengths when

compared between species. It was further confirmed by Wilcoxon test (Figure 1). In case of 3' UTRs, significant difference was obtained for *Glycine max* and *Solanum lycopersicum* with p-value 1.078e-05 and 3.145e-06, respectively. In case of 5' UTRs significant difference was obtained for *Arabidopsis thaliana*, *Glycine max*, *Oryza sativa* and *Solanum lycopersicum* with p-value of 6.54e-09, 0.0001262, 6.084e-06 and 6.315e-06, respectively. These distinct UTRs of stress related genes may play important role in more stringent regulation of gene expression in stress conditions.



Stress related genes contain specific and significantly over-represented small motif

At this stage we sought to know if there is any common short sequence motif that significantly prevails in stress-related genes. We used MEME package [35,36] to search for overrepresented motifs in 160 stress-related genes. One significantly overrepresented G-C rich motifs was obtained with the e-value 1.1e-007. This motif was selected for further analysis (Figure 2). This is a 12 nucleotides long motif with pattern GGC[GT]GC[TG]GC[GTA]GC. In non-stress related genes the frequency of occurrence of this motif was searched by STORM program [37]. From non-stress related gene dataset 1000 sets, each consisting of randomly selected 160 genes, were constructed to use as control for this purpose. To find the significant difference in the frequency of occurrence of this motif, z-score was calculated. The occurrence of this motif in non-stress related genes was not significant, which implies for its highly specific association with stress related genes.

Important biological processes, cellular components and molecular functions are enriched for stress related genes

Functional enrichment analysis is performed to assign biological meaning to genes. It is performed to assess if a gene or a group of genes show any significant over-representation of any biological characteristics. In this study Gtools [39] was used for enrichment analysis using Gene Ontology (GO) [38] database. GO database centralizes and disseminates the prior knowledge of known gene which allows researchers to assign attributes to their experimentally identified genes. Stress related genes of five plants were analyzed in a background of non-stress related genes, to determine the enrichment of particular biological processes (BP), molecular functions (MF) and cellular components (CC) as termed by GO. Detail

statistical results are presented in supplementary tables (S2 to S12). The stress related genes of *Arabidopsis thaliana* were enriched for removal of superoxide radicals, generation of precursor metabolites and energy, thylakoid, metal ion binding, copper ion binding, chromatin binding, etc (Figure 3 and Table 1). *Glycine max* stress related genes were enriched for transferase activity, oxidoreductase activity, etc (Figure 4 and Table 2). In *Oryza sativa* these were enriched for plasma membrane localized proteins (Figure 5 and Table 3). The stress related genes of *Solanum lycopersicum* were enriched for oxidoreductase activity, organelle, etc (Figure 6 and Table 4). In *Zea mays* they were enriched for small GTPase mediated signal transduction, oxidation-reduction process, carbohydrate metabolic process, biosynthetic process, plasma membrane, cytoplasm, organelle, chloroplast, GTP binding, oxidoreductase activity, nucleotide binding, etc. (Figure 7 and Table 5). In this analysis oxidoreductase activity was found to be a common mechanism to stress response in almost all plants. Some findings from this analysis are seemed to be specifically significant like thylakoid and chromatin binding of *A. thaliana* and small GTPase activity of *Zea mays*. From literature mining enough supporting data regarding the significant association of these components with stress were obtained. In *Arabidopsis*, ascorbate peroxidase bound to thylakoid contributes in scavenging reactive oxygen species produced in different stress conditions [40]. *Arabidopsis* TAAC (Thylakoid ATP/ADP Carrier) gene is highly up-regulated in leaves under different stress environments [41]. Gene activation in dehydration stress responses depends on a specific pattern of histone modification and chromatin structure [42]. H3K4me3 (H3 Lys4 trimethylation) has a function as epigenetic marker of stressed memory [43]. Epigenetic regulations mediated by the modification of histone proteins are conserved in plant [44]. Modifications on the sites of H3K4 and H3K9 are correlated with the activities

of abiotic stress responsive genes in Arabisopsis [45]. In Arabidopsis, Rop GTPase signalling influences the mechanisms of alcohol dehydrogenase activity at low O₂ condition [46]. Monomeric RopGTPases regulate the production of H₂O₂, responses to hormones, programmed cell death, etc [47]. As small GTPase activity was significantly enriched in Zea mays, it can

be deciphered that they may play similar role in this plant as in Arabidopsis. From this enrichment analysis, it can be concluded that components enriched in different plants are highly co-related with different stress conditions and can be considered as distinct features of plant stress related genes.



Figure 2: Selected motif obtained from MEME analysis. This is represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. There are stacks of letters at each position in the motif. The total height of the stack is the “information content” of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack.

Figure 3

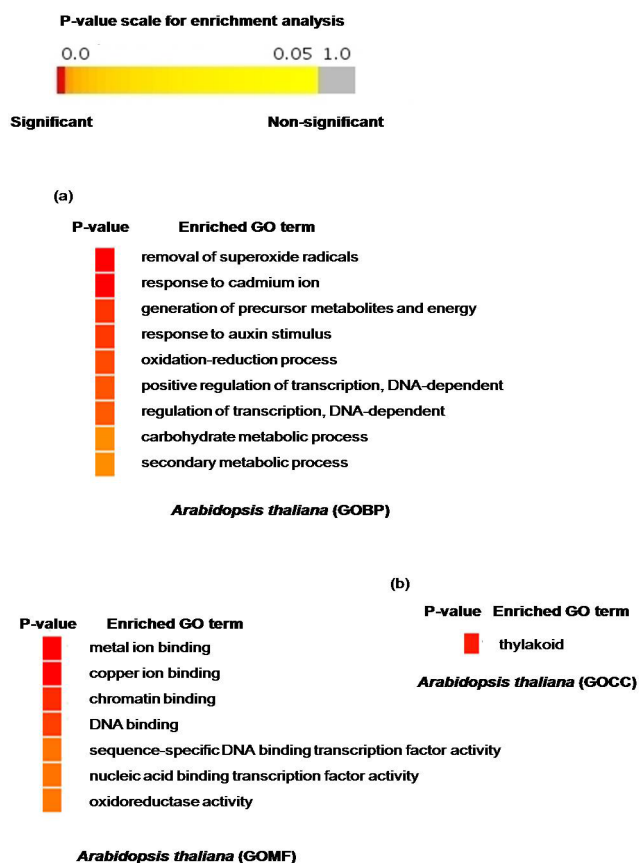





Figure 3: Significantly enriched Gene Ontology terms for the stress related genes of Arabidopsis thaliana. The enrichment of the terms with significant p-value is denoted by red color coding. The deep red color indicates to the highly significant p-value. The lighter the red color the lesser the significance. Gray color indicates the insignificant p-value. (a) Enriched GO terms of Biological Process. (b) Enriched GO terms of Cellular Component. (c) Enriched GO terms of Molecular Function.

GO term for Biological Process	Total studied Gene	Observed	Expected mean	Corrected right-p-value
Removal of superoxide redicals	33	7	0.0117853635	5.81E-16
Response to cadmium ion	33	8	0.4586470641	3.66E-6
Generation of precursor metabolites and energy	33	6	0.5823933811	0.0015
Response to auxin stimulus	33	5	0.3594535877	0.0015
Oxidation reduction process	33	8	1.3287997381	0.0023
Positive regulation of transcription, DNA dependent	33	5	0.4537364959	0.0033
Regulation of transcription, DNA dependent	33	9	1.8404809381	0.0035
Carbohydrate metabolic process	33	8	1.909228892	0.0157
Secondary metabolic process	33	5	0.7081039255	0.0155
GO term for Cellular Component				
Thylakoid	33	6	0.4910568137	0.0019
GO term for Molecular Function				
Metal ion binding	33	7	0.2632064522	4.77E-6
Copper ion binding	33	5	0.2386536115	6.31E-4
Chromatin binding	33	5	0.3280259516	0.0021
DNA binding	33	10	2.0182435047	0.0036
Sequence specific DNA binding transcription factor	33	8	1.6509330079	0.0168
Nucleic acid binding transcription factor activity	33	8	1.6519151216	0.0147
Oxidoreductase activity	33	8	1.6764679623	0.0144

Table 1: Enriched Gene Ontology terms for stress related genes of *Arabidopsis thaliana*

P-value **Enriched GO term**

 transferase activity, transferring alkyl or aryl (other than methyl) groups
 oxidoreductase activity
 ion binding


Glycine max

Figure 4: Significantly enriched GO terms for the stress related genes of *Glycine max*. The red color coding indicates the significance of the p-value as described in Figure 3.

GO term	Total Studied Genes	Observed	expected-mean	Corrected right-p-value
Transferase activity, transferring alkyl or aryl (other than methyl) groups	55	22	0.12182966	9.28E-41
Oxidoreductase activity	55	15	2.9300033226	1.89E-5
Ion binding	55	14	5.437663824	0.025

Table 2: Enriched Gene Ontology terms for stress related genes of *Glycine max*

P-value **Enriched GO term**

 plasma membrane

Oryza sativa

Figure 5: Significantly enriched GO terms for the stress related genes of *Oryza sativa*. The red color coding indicates the significance of the p-value as described in Figure 3.

GO term	Total Studied genes	Observed	Expected-mean	Corrected right-p-value
Plasma membrane	10	4	0.3035335430	0.0385

Table 3: Enriched Gene Ontology term for stress related genes of *Oryza sativa*

P-value **Enriched GO term**

- oxidoreductase activity
- organelle
- cytoplasm
- intracellular
- ion binding

Solanum lycopersicum

Figure 6: Significantly enriched GO terms for the stress related genes of *Solanum lycopersicum*. The red color coding indicates the significance of the p-value as described in Figure 3.

GO term	Total Studied Genes	Observed	Expected-mean	Corrected right-p-value
Oxidoreductase activity	26	8	1.2154928765	0.0021359861
Organelle	26	8	1.2589836765	0.0018170424
Cytoplasm	26	6	0.9020591798	0.0061442772
Intracellular	26	9	2.1580434908	0.0062167642
Ion binding	26	7	2.0980561804	0.0478418769

Table 4: Enriched Gene Ontology terms for stress related genes of *Solanum lycopersicum*

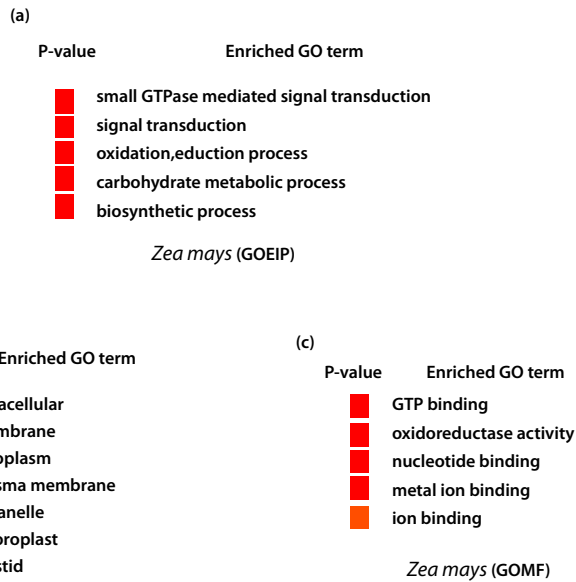


Figure 7: Significantly enriched GO terms for the stress related genes of *Zea mays*. (a) Enriched GO terms of Biological Process. (b) Enriched GO terms of Cellular Components. (c) Enriched GO terms of Molecular Function. The red color coding indicates the significance of the p-value as described in Figure 3.

GO term for Biological Process	Total Observed Genes	Observed	Expected-mean	Corrected right-p-value
Small GTPase mediated signal transduction	41	8	0.0506382897	8.78E-14
Signal transduction	41	9	0.2282392476	2.19E-10
Oxidation – reduction process	41	9	0.6502249947	8.91E-7
Carbohydrate metabolic process	41	7	0.4109774233	4.93E-6
Biosynthetic process	41	10	1.2270611638	9.41E-6
GO term for Cellular Component				
Intracellular	41	22	2.8695030804	3.05E-11
Membrane	41	13	0.9826763746	1.15E-9
Cytoplasm	41	16	1.8493984048	2.85E-9
Plasma membrane	41	10	0.5753683926	9.67E-9
Organelle	41	13	2.3095463412	1.04E-5
Chloroplast	41	7	0.5438111976	1.59E-5
Plastid	41	7	0.6186677997	3.36E-5
GO term for Molecular Function				
GTP binding	41	8	0.1343015508	4.51E-10
Oxidoreductase activity	41	9	0.7192104907	5.41E-6
Nucleotide binding	41	9	0.7705826686	7.23E-6
Metal ion binding	41	7	0.5739006161	6.78E-5
Ion binding	41	8	1.3518221673	0.0016

Table 5: Enriched Gene Ontology terms for stress related genes of *Zea mays*

Discussion

Environmental stresses are limiting factors for plant growth. Extensive studies are being done on *Arabidopsis* and some other plants regarding their responses to different environmental stresses. But not enough data are available on mechanisms of stress responses in most of the plants. Even all the stress responsive genes are not identified yet and the mechanisms of stress responses are not completely known. Some bioinformatics studies have been done to identify specific features of regulatory regions of stress related genes [30,31] but not in the region of genes themselves. In this study bioinformatics and computational analyses were performed with plant stress related genes listed from plant stress gene database [32].

Each gene is flanked by short 5' and 3' untranslated regions (UTRs) followed by gene start site and gene end site [48]. Computational analyses identified significant differences in the length of 5' and 3' UTRs between stress related and non-stress related genes (Figure 1). The significant differences calculated as z-score and p-value by Perl scripts, were further confirmed by the Wilcoxon test. It is well established that the regulation of gene expression become highly stringent in different stress conditions [49-52]. UTRs play vital roles in the regulation of gene expression [53-59]. Therefore, distinct UTRs of stress related genes may play pivotal roles to ensure tight regulation of gene expression in stress environments. Further analysis should be propounded to discover the exact role of distinct UTRs in stress conditions.

Distinct motifs were discovered in the regulatory regions of plant stress related genes in previous studies [30,31]. But in this study, by MEME analysis with the clause that the motif should exist in 50% or more genes, a small G-C rich motif (Figure 2) was obtained in the regions of genes. The p-value justified that the motif was significantly overrepresented in stress related genes. The frequency of occurrence of these motifs in non-stress related genes calculated by STORM program with PWM was accountably low. This result corroborates the possibility of occurrence of such distinct motifs in stress related genes. Occurrence of such motif insinuates that it may have important role in up-regulation, down-regulation or epigenetic regulation of genes in stress condition. Further extensive studies including more plant species and newly discovered genes are necessary to discover more such motifs and their plausible roles.

In past decade, each gene product was studied individually to assign its role in biological process but now tools exist to make this process automated. Gitools is such a tool used in this study. Primarily a group of genes are clustered based on some common properties. Enrichment analysis is performed to assess if a group of genes shows any significant over-representation of any biological characteristics. In this study, after detecting the over-represented biological characteristics of stress related genes, data mining was performed manually to explain their possible role in stress responsive mechanisms of plants.

Oxidative stress in plants is a common scenario in different stress conditions like cold, submergence, drought, salinity etc. [60-64]. In our enrichment analysis oxidoreductase activity was significantly enriched in stress related genes of almost all plants included in the study (Figure 3, 4, 6, 7 & Table 1, 2, 4, 5), which is coherent with this scenario. Therefore oxidoreductase activity can be considered as a specific feature of plant stress related genes.

Small GTPase mediated signal transduction was significantly enriched in *Zea mays*. The G proteins have important role in signal transduction. They mediate the signal transduction to downstream effectors [65]. In rice, a small GTPase, Rac1, regulate the death of hypersensitive cells in innate immune response while heterotrimeric G protein regulates the Rac1 [66,67]. Low O₂ regulates the ADH (alcohol dehydrogenase) activity that depends on RopGTPase signaling in *Arabidopsis* [68]. Chromatin binding was significantly enriched as molecular function in *Arabidopsis thaliana* which insinuates toward the epigenetic correlation with stress conditions. Though it is not well understood whether chromatin mediated regulation has positive effects on stress tolerance, it is obvious that there are correlations between epigenetic modifications and plant stress responses [43]. It was observed that linker histones and HMGB (High Mobility Group) proteins play role in abiotic stress responses [69]. Promoter specific histone modification H3K4me3 plays an important role in dehydration and ABA stress responses [70]. In drought response, some lysine modification states on histone H3 N tail are altered which revealed that upon gene activation in stress responses histone modification states changes [42]. From these data it can lucidly be told that small GTPase mediated signal transduction and chromatin binding are the specific phenomena in different stress conditions.

Again, in *Arabidopsis*, only the cellular component, thylakoid was enriched. This result indicates toward the unique role of thylakoid in stress responses. It has been discovered that stresses have significant effects on the different components of thylakoid [40]. The transcript level of OsCYP20-2 gene in thylakoid lumen of rice is highly regulated under abiotic stress conditions and CYP20-2 gene is also found to be well conserved in some photosynthetic plants [71]. TLP18.3 gene is up regulated in dehydration stress and thylakoid protease Deg2 consorts in stress related degradation of Lhc6, light harvesting protein of photosystem II, in *Arabidopsis thaliana* [72,73]. Therefore, thylakoid is a very important cellular component that may have more crucial role in stress responses than other organelles in *Arabidopsis*. Some other molecular functions, biological processes and cellular components (shown in Figure 3 - 7 and Table 1 - 5) were significantly enriched in stress related genes. All these findings from enrichment analysis can be considered as significant and specific features of stress related genes.

Conclusions

In this study, structural and functional analyses have been done with plant stress related genes with a view to identify hidden features that can discriminate them from non-stress related genes. Extensive computational and bioinformatics analysis were performed and differential outcomes gave an overall idea that discriminating features between stress related and non-stress related genes exist at every level of biological hierarchy. The different UTRs length, existence of distinct G-C rich motifs and selectively enriched some biological phenomena and constituents like small GTPase mediated signal transduction, chromatin binding, oxidoreductase activity and thylakoid identified as stress specific features prove this decipherment.

This analysis proved that there are specific features hidden in stress related genes which are different from non-stress related genes. It can be suggested that further studies should be done by including updated and classified data of plant stress to identify more common and specific features. If enough features can be identified which are highly specific for stress related genes and also discriminating from non-stress related genes, a computational model can be devised that can discern stress related genes from the stockpile of genes.

Acknowledgments

We acknowledge Professor Dr. Haseena Khan, Chairperson, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh for her support at the beginning of the project. We also acknowledge Dr. Abu A Sajib, Assistant Professor, Department of Genetic Engineering & Biotechnology, University of Dhaka for critical reading of the manuscript.

Reference

- Hoffmann AA, Parsons PA (1997) *Extreme Environmental Change and Evolution*. Cambridge University Press, Cambridge, UK.
- Khan NA and Singh S (2008) *Abiotic stress and plant responses*. I.K. International Pub. House.
- Bohnert HJ, Gong Q, Li P, Ma S (2006) Unraveling abiotic stress tolerance mechanisms – getting genomics going. *Curr Opin Plant Biol* 9: 180-188.
- Hu H, Dai M, Yao J, Xiao B, Li X, et al. (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc Natl Acad Sci U S A* 103: 12987-12992.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, et al. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31: 279-292.
- Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, et al. (2002) Transcriptomic changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol* 130: 2129-2141.
- Takahashi S, Seki M, Ishida J, Satou M, Sakurai T, et al. (2004) Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in Arabidopsis cell culture using a full-length cDNA microarray. *Plant Mol Biol* 56: 29-55.
- Swindell WR (2006) The Association Among Gene Expression Responses to Nine Abiotic Stress Treatments in Arabidopsis thaliana. *Genetics* 174: 1811-1824.
- Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J* 45: 523-539.
- Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, et al. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr Opin Plant Biol* 9: 436-442.
- Hoffmann AA, Parsons PA (1989) An integrated approach to environmental stress tolerance and life-history variation: desiccation tolerance in *Drosophila*. *Biol J Linn Soc* 37: 117-136.
- Hoffmann AA, Parsons PA (1991) *Evolutionary Genetics and Environmental Stress*. Oxford University Press, Oxford.
- Dutilleul C, Garmier M, Noctor G, Mathieu C, Chetrit P, et al. (2003) Leaf mitochondria modulate whole cell redox homeostasis, set antioxidant capacity, and determine stress resistance through altered signaling and diurnal regulation. *Plant Cell* 15: 1212-1226.
- Noctor G, Queval G, Gakiere B (2006) NAD(P) synthesis and pyridine nucleotide cycling in plants and their potential importance in stress conditions. *J Exp Bot* 57: 1603-1620.
- Hoffmann AA, Parsons PA (1989) Selection for increased desiccation resistance in *Drosophila melanogaster*: additive genetic control and correlated responses for other stresses. *Genetics* 122: 837-845.
- Agrawal AA, Connerand JK, Stinchcombe JR (2004) Evolution of plant resistance and tolerance to frost damage. *Ecol Lett* 7: 1199-1208.
- Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, et al. (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using timing array. *Plant Cell Physiol* 49:1135-1149.
- Urano K, Kurihara Y, Seki M, Shinozaki K (2010) ‘Omics’ analyses of regulatory networks in plant abiotic stress responses. *Curr Opin Plant Biol* 13: 132-138.
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123: 1279-12791.
- Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, et al. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr Opin Plant Biol* 9: 436-442.
- Feder ME (1999) Organismal, ecological, and evolutionary aspects of heat-shock proteins and the stress response: established conclusions and unresolved issues. *Am Zool* 39: 857-864.
- Vierling E (1991) The roles of heat shock proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* 42: 579-620.
- Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, et al. (1998) Two transcription factors DREB1 and DREB2, with an EREBP/AP2 DNA binding protein, separate two cellular signal transduction pathways in drought and low-temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* 10: 1391-1406.
- Kim JB, Kang JY, Kim SY (2004) Over-expression of a transcription factor regulating ABA-responsive gene expression confers multiple stress tolerance. *Plant Biotechnol J* 2: 459-466.
- Chen WJ, Zhu T (2004) Networks of transcription factors with roles in environmental stress response. *Trends Plant Sci* 9: 591-596.
- Mahalingam R, Gomez-Buitrago A, Eckardt N, Shah N, Guevara-Garcia A, et al. (2003) Characterizing the stress/defense transcriptome of Arabidopsis. *Genome Biol* 4: R20.
- Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* 17: 287-291.
- Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity, and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218: 1-14.

29. Denby K, Gehring C (2005) Engineering drought and salinity tolerance in plants: lessons from genome-wide expression profiling in Arabidopsis. *Trends Biotechnol* 23: 547-552.
30. Geisler M, Kleczkowski LA, Karpinski S (2006) A universal algorithm for genome-wide in silico identification of biologically significant gene promoter putative cis-regulatory-elements; identification of new elements for reactive oxygen species and sucrose signaling in Arabidopsis. *Plant J* 45: 384-398.
31. Ma S, Bachan S, Porto M, Bohnert HJ, Snyder M, et al. (2012) Discovery of Stress Responsive DNA Regulatory Motifs in Arabidopsis. *PLoS One* 7: e43198.
32. Prabha R, Ghosh I, Singh DP (2011) Plant Stress Gene Database: A Collection of Plant Genes Responding to Stress Condition. *ARPN Journal of Science and Technology* 1: 28-31.
33. R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
34. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
35. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 28-36, AAAI Press, Menlo Park, California.
36. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34: W369-W373.
37. Schones DE, Smith AD, Zhang MQ (2007) Statistical significance of cis-regulatory modules. *BMC Bioinformatics* 8: 19.
38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25: 25-29.
39. Perez-Llamas C, Lopez-Bigas N (2011) Gitoools: Analysis and Visualisation of Genomic Data Using Interactive Heat-Maps. *PLoS One* 6: e19541.
40. Kangasjärvi S, Lepistö A, Hännikäinen K, Piippo M, Luomala EM, et al. (2008) Diverse roles for chloroplast stromal and thylakoid bound ascorbate peroxidases in plant stress responses. *Biochem J* 412: 275-85.
41. Thuswaldner S, Lagerstedt JO, Rojas-Stütz M, Bouhidel K, Der C, et al. (2007) Identification, expression and functional analyses of a thylakoid ATP/ADP carrier from Arabidopsis. *J Biol Chem* 282: 8848-59.
42. Kim JM, To TK, Ishida J, Morosawa T, Kawashima M, et al. (2008) Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in Arabidopsis thaliana. *Plant Cell Physiol* 49: 1580-1588.
43. Kim JM, To TK, Ishida J, Matsui A, Kimura H, et al. (2012) Transition of Chromatin Status During the Process of Recovery from Drought Stress in Arabidopsis thaliana. *Plant Cell Physiol* 53: 847-856.
44. Pfluger J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol* 10: 645-652.
45. Kim JM, To TK, Nishioka T, Seki M (2010) Chromatin regulation functions in plant abiotic stress responses. *Plant Cell Environ* 33: 604-611.
46. Baxter-Burrell A, Yang Z, Springer PS, Bailey-Serres J (2002) RopGAP4-dependent Rop GTPase rheostat control of Arabidopsis oxygen deprivation tolerance. *Science* 296: 2026-2028.
47. Yang Z (2002) Small GTPases: versatile signaling switches in plants. *Plant Cell* 14: 375-388.
48. Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. *Genome Biol* 3: REVIEWS0004.
49. Bilgin DD, Zavala JA, Zhu J, Clough SJ, Ort DR, et al. (2010) Biotic stress globally down regulates photosynthesis genes. *Plant, Cell and Environment* 33: 1597-1613.
50. Liu X, Vance Baird Wm (2003) Differential Expression of Genes Regulated in Response to Drought or Salinity Stress in Sunflower. *Crop Sci* 43: 678-687.
51. Givens RM, Lin MH, Taylor DJ, Mechold U, Berry JO, et al. (2004) Inducible Expression, Enzymatic Activity, and Origin of Higher Plant Homologues of Bacterial RelA/SpoT Stress Proteins in *Nicotiana tabacum*. *J Biol Chem* 279: 7495-7504.
52. Dabrowska G, Prusińska J, Goc A (2006) Plant mechanism of an adaptive stress response homologous to bacterial stringent response. *Postepy Biochem* 52: 94-100.
53. Merritt C, Rasoloson D, Ko D, Seydoux G (2008) 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr Biol* 18: 1476-1482.
54. Hughes TA (2006) Regulation of gene expression by alternative untranslated regions. *Trends Genet* 22: 119-22.
55. Kim SH, Samal SK (2010) Role of Untranslated Regions in Regulation of Gene Expression, Replication, and Pathogenicity of Newcastle Disease Virus Expressing Green Fluorescent Protein. *J Virol* 84: 2629-2634.
56. Ng DW, Chandrasekharan MB, Hall TC (2004) The 5' UTR negatively regulates quantitative and spatial expression from the ABI3 promoter. *Plant Mol Biol* 54: 25-38.
57. Kang MK, Han SJ (2011) Post-transcriptional and post-translational regulation during mouse oocyte maturation. *BMB Rep* 44: 147-157.
58. Yang J, Schuster G, Stern DB (1996) CSP41, a sequence-specific chloroplast mRNA binding protein, is an endoribonuclease. *Plant Cell* 8: 1409-1420.
59. Yang J, Stern DB (1997) The spinach chloroplast endoribonuclease CSP41 cleaves the 3' untranslated region of the petD mRNA primarily within its terminal stem-loop structure. *J Biol Chem* 272: 12874-12880.
60. Pang CH, Wang BS (2008) Oxidative Stress and Salt Tolerance in Plants. *Progress in Botany* 69: 231-245.
61. Hernández JA, Corpas FJ, Gómez M, del Río LA, Sevilla F (2006) Salt-induced oxidative stress mediated by activated oxygen species in pea leaf mitochondria. *Plant Cell Physiol* 47: 103-110.
62. Sharma P, Dubey RS (2005) Drought induces oxidative stress and enhances the activities of antioxidant enzymes in growing rice seedlings. *Plant Growth Regulation* 46: 209-221.
63. Borsani O, Di'az P, Agius MF, Monza VVJ (2001) Water stress generates an oxidative stress through the induction of a specific Cu/Zn superoxide dismutase in *Lotus corniculatus* leaves. *Plant Science* 161: 757-763.

64. Prasad TK, Anderson MD, Martin BA, Steward CR (1994) Evidence for chilling-induced oxidative stress in maize seedlings and a regulatory role for hydrogen peroxide. *Plant Cell* 6: 65-74.
65. Tuteja N, Sopory SK (2008) Plant signaling in stress. *Plant Signal Behav* 3: 79-86.
66. Nakashima A, Chen L, Thao NP, Fujiwara M, Wong HL, et al. (2008) RACK1 functions in rice innate immunity by interacting with the Rac1 immune complex. *Plant Cell* 20: 2265-2279.
67. Suharsono U, Fujisawa Y, Kawasaki T, Iwasaki Y, Satoh H, et al. (2002) The heterotrimeric G protein α subunit acts upstream of the small GTPase Rac in disease resistance in rice. *Proc Natl Acad Sci U S A* 99: 13307-13312.
68. Baxter-Burrell A, Yang Z, Springer PS, Bailey-Serres J (2002) RopGAP4- dependent Rop GTPase rheostat control of Arabidopsis oxygen deprivation tolerance. *Science* 296: 2026-2028.
69. Antosch M, Mortensen SA, Grasser KD (2012) Plant Proteins Containing High Mobility Group Box DNA-Binding Domains Modulate Different Nuclear Processes. *Plant Physiol* 159: 875-883.
70. van Dijk K, Ding Y, Malkaram S, Riethoven JJ, Liu R, et al. (2010) Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana. *BMC Plant Biol* 10: 238.
71. Kim SK, You YN, Park JC, Young Y, Kim BG, et al. (2012) The rice thylakoid luminal cyclophilin OsCYP20-2 confers enhanced environmental stress tolerance in tobacco and Arabidopsis. *Plant Cell Rep* 31: 417-426.
72. Ansari MI, Lin TP (2011) Arabidopsis thaliana thylakoid lumen 18.3 protein gene is up-regulated during dehydration. *J Physiol Pathophysiol* 2: 18-24.
73. Luciński R, Misztal L, Samardakiewicz S, Jackowski G (2011) The thylakoid protease Deg2 is involved in stress-related degradation of the photosystem II light-harvesting protein Lhcb6 in Arabidopsis thaliana. *New Phytologist* 192: 74-86.

Submit your manuscript at

<http://enlivenarchive.org/submit-manuscript.php>

New initiative of Enliven Archive

Apart from providing HTML, PDF versions; we also provide **video version** and deposit the videos in about 15 freely accessible social network sites that promote videos which in turn will aid in rapid circulation of articles published with us.