Insights from the protein sequence and structure analysis of PgHsc70 and OsHsp70 genes

Gugulothu Baloji2, Sandhya Jagtap2, Ashwini Talakayala2, Meghana Kolli2, Lali Lingfa2, Mallikarjuna Garladinne1 & Srinivas Ankanagari2*

1Plant Molecular Biology Laboratory, Agri Biotech Foundation, Rajendra Nagar, Hyderabad (T.S) 500 030, India; 2Department of Genetics, Osmania University, Hyderabad - 50007 (T.S) India; *Corresponding author - Srinivas Ankanagari

Author contacts:
Gugulothu Baloji – E-mail: baloji2020@gmail.com; Sandhya Jagtap – E-mail: sandhya.chavaan@gmail.com; Ashwini Talakayala – E-mail: talakayalaashwini@gmail.com; Meghana Kolli – E-mail: kollimeghana08@gmail.com; Lali Lingfa – E-mail: llingfa@gmail.com; Mallikarjuna Garladinne – E-mail:garladinnemarjun@gmail.com; Srinivas Ankanagari - E-mail: asrinivas@osmania.ac.in;

Abstract:
Heat shock proteins are induced in a wide range of abiotic and biotic stresses. They are well known for cellular chaperone activities and play an important role in protecting plants through regulation of homeostasis and survival. A comprehensive characterization and comparative analysis of the Hsp70 family members within the closely related plant species helps in better interpretation of these proteins’ contribution to cell function and response to specific environmental stresses. Therefore, it is of interest to glean insights from the protein sequence analysis of PgHsc 70 and OsHsp70 genes. Thus, we document data from the sequence and structure analysis of PgHsc 70 and OsHsp 70 gene a.
Background:
Plant defense mechanisms are induced rapidly and plants adapt at morphological, molecular and physiological levels [1, 2, 3]. The sensing of abiotic stresses initiates complex signaling pathways controlling the stress and tolerance responses. The signal transmission of stress and subsequent induction of stress receptive pathways involves expression of genes and proteins related to tolerance that have been studied extensively at the molecular level [4, 5]. The molecular mechanisms have identified a large number of genes induced with abiotic stress factors and characterized using approaches such as subtractive cDNA libraries [6], microarrays [7] and NGS based RNA sequencing [8]. In the cell, protein aggregation due to environmental stresses is a major effect resulting in their dysfunction. Understanding the protective mechanisms to abiotic stresses is indispensable for developing crops with increased stress tolerance [9].

In the environmental stress conditions, the cell survival and sustenance is dependent upon protein native conformation and preventing protein aggregation. In the event of environment stress conditions, chaperone proteins, assist to fold cellular proteins into three-dimensional conformation and avoid abnormal folding and aggregation [10, 11]. Heat shock proteins (HSPs) are the central chaperone proteins, involved in the maintenance of homeostasis, nascent protein folding, denatured proteins refolding, aggregation prevention and aiding protein transport across the membranes [12,13]. Besides these processes, HSP gene family members also protects the cells from the damage caused in the events of extreme temperatures, salinity, dehydration, oxidative stress, heavy metal toxicity, high intense irradiation and wounding [14,7]. In the cells under higher temperature environments, the HSPs are immediately synthesized and expressed, and on the other hand, most of other proteins' synthesis is detained. Thus, heat shock proteins are performing key role in protecting plants through cellular homeostasis regulation to stress conditions [15, 16]. Furthermore, based on stress signal, research has shown that high HSPs expression and accumulation is involved in different stress signaling pathways. In plants, the HSP induction, synthesis and increase in thermo tolerance are well documented [17, 18, 19].

Based on the molecular weights, five families of heat shock proteins are identified. The major HSPs families are chaperonin (Hsp60/GroEL), 70-kDa Hsp (Hsp70/DnaK), Hsp90, Hsp100/CipB, and the small heat shock proteins (sHsp) [20]. Of these, the HSP70 family is of greatest interest. It is evolutionarily conserved, present in archaeabacteria, plants and humans [21]. The four Hsp70 gene subgroup family members are localized in the sub-cellular compartments: plastids, mitochondria, endoplasmic reticulum and cytosol [22]. Furthermore, Hsp70 family includes genes that are constitutive (housekeeping) and predominantly associated with physiological functioning such as heat shock cognate (hsc) 70 gene, or stress-induced such as hsp70. In general, the newly synthesized proteins are folded by constitutive expressed members whereas protein translocation into the organelles, involve stress-induced members that re-fold and degrade mis-folded proteins in adverse environmental stress conditions [23, 24]. Both these proteins have modular structure playing role in cell growth with conserved ATPase domain and hydrophobic pocket with lid-like structure of substrate-binding domain (SBD) at N-terminal end and variable C-terminal domain but conserved [23]. Because of diverse subcellular localizations, Hsp70 plays critical role in development or specific protein communications [25]. The activity of Hsp70 is also modulated by post-translational modifications and by interaction with other co-chaperones [26].

The yield improvement of crops in the unfavorable abiotic conditions is a challenge. In particular the role played by Poaceae crops in food demand is well known, contributing high amount of calories in the human diet [27, 28]. Wheat, rice, maize, millet, sorghum, barley and rye starchy grains serving as important food sources for the world’s majority population [27, 28]. The Hsp70 family members comprehensive characterization in plant species is needed to know how these members contribute towards cell function and protect in adverse abiotic stresses [24]. Therefore, it is of interest to document the protein sequence analysis data of PgHsc 70 and OsHsp 70 genes to glean useful information.

Materials and Methods:
Data source:
The complete nucleotide sequence and the CDS of Pennisetum glaucum heat shock cognate 70 kDa protein (PgHSC70) and Oryza sativa hsp70 gene for heat shock protein 70 (OsHSP70) registered in GenBank and their protein sequences are obtained in FASTA format from National Center for Biotechnology information (NCBI). The gene and protein structural and functional analysis of sequences was done using in-silico tools.

Gene structure analysis:
The gene structure of PgHSC70 and OsHSP70 were predicted based on the genome and coding sequences using the Gene Structure Display Server. This server analyses exon/intron organization of PgHSC70 and OsHsp70 genes. In the gene structure analysis exons/introns and intronic phase distribution (phase 0, 1, 2) were identified and marked. Based on position relative to reading frame three intron phases exist: insertion between two codons (phase 0), insertion after first base codon (phase 1) or after second base codon (phase 2).

Conserved motif analysis:
Using MEME suite sequence motifs were scanned over the nucleotide sequences in PgHSC70 and OsHSP70 genes. Input of cDNA sequences of PgHSC70 and OsHSP70 genes were given to the MEME suite. In a default setting, it can help in predicting up to three motifs and selected for finding distribution of motifs with three different parameters. Parameters were optimized in MEME suite set to 10 as maximum number and 1 as minimum occurrence.
of motif site per sequence. All the other parameters were kept in default value. Predicting the width and the occurrence number of motifs, in order to minimize the E-value, was automatically done with MEME suite.

**Cis-Acting regulatory elements:**
Using Plant CARE database cis-acting regulatory elements (CREs) were scanned in PgHSC70 and OsHsp70 genes. The PgHSC70 and OsHsp70 genes cDNA sequences were uploaded and evaluated for cis-regulatory response elements presence in promoter regions to predict computationally the regulatory elements. For the response elements shown, a matrix value of ≥5 was considered for acceptance on the sense strand. The obtained cis-elements were compared with each other.

**miRNA Target sites prediction:**
Plant small RNA-targeted gene prediction was performed on sequences PgHSC70 and OsHsp70 genes using psRNATarget server. The miRNA target sites were analyzed using default parameters.

**Multiple sequence alignment (MSA) and phylogenetic analysis:**
For identification of similarity of PgHSC70 and OsHsp70, the homology search of each protein was performed by BLAST using blastp algorithm respectively. Using multiple sequence alignment (MSA) tool ClustalW2, the protein sequences of all the identified homologues of PgHSC70 and OsHsp70 were aligned. Using the BLOSUM 62 substitution matrix evolutionary alignment was inferred with progressive method. The phylogenetic tree was constructed for the identified PgHSC70 and OsHsp70 proteins using ClustalW2.

**Physicochemical characterization:**
Physicochemical characterization of the target protein sequences of PgHSC70 and OsHsp70 such as mol. wt, aa composition, isoelectric point (pI), instability index (II), the total negative and positive residues, extinction coefficient (EC), grand average of hydropathicity (GRAVY) and aliphatic index (AI) were analyzed using Expasy’s ProtParam prediction server.

**Secondary structure prediction:**
SOPMA tool was used to predict PgHSC70 and OsHsp70 proteins secondary structure for assigning positional possibility of various regions of α-helix, β-strands, turns as well as random coils likely to fold. The method makes use of predicting consensus from multiple alignments of the relative frequencies of each amino acid anchored in the X-ray crystallographic solved protein templates.

**Prediction of Subcellular localization and SignalP:**
PgHSC70 and OsHsp70 subcellular localization was predicted using CELLO v.2.5 which is a multiclass support vector machine classification system. SignalP was used to verify the presence of signal peptide cleavage sites and their locations in both proteins, which works on the basis of a combination of several neural networks, namely artificial neural network (ANN) and Hidden Markov Model (HMM).

**Protein-Protein Interaction network:**
Using STRING (Search Tool for Retrieval of Interacting Genes) v 9.1 protein-protein interactions (PPIs) analysis was done. The STRING repository consisted of PPIs concerning stable protein complexes, functional and regulatory interactions. The PPIs of PgHSC70 and OsHsp70 were searched individually by submitting a protein query sequence in the search box of STRING. Determining the protein–protein interaction network would empower study of signaling pathways.

**Disulfide-Bonding in protein:**
In protein folding and formation a functional and stable confirmation is determined by disulfide bonds among its cysteine residues. To predict cysteine bonds (disulfide bonds) presence and absence and their bonding patterns CYS_REC tool was used.

**Post-translational modification sites**
The targets in PgHSC70 and OsHsp70 were predicted for putative acetylation, methylation, phosphorylation, ubiquitination, and N-glycosylation sites. Phosphorylation at serine, threonine and tyrosine residues in the PgHSC70 and OsHsp70 proteins were predicted using NetPhos 3.1 was used. To complete this task it ensembles neural networks and residues having scores >0.5 threshold as phosphorylated. The N-glycosylation sites of the target proteins were predicted with NetGlyc 1.0 server, with threshold value of >0.5. By default the predictions are done only on the Asn-Xaa-Ser/Thr sequons.

**Protein disorder analysis:**
The estimation of intrinsic disordered regions (IDRs) of PgHSC70 and OsHsp70 was made by DisEMBL tool.

**Homology modeling and docking:**
The protein structures of PgHSC70 and OsHsp70 was modeled using the bovine HSC70 (PDB ID: 1YUW) as template sequence exhibiting the highest similarity identified by BLAST against the PDB database. DS was used to design a homology model of both proteins and the each protein model with less geometric function was selected and energy minimized in CHARMm force field using DS minimization algorithms. For structural validation, the obtained final models are further subjected to PROCHECK for Psi/Phi Ramachandran plots analysis. Protein binding/catalytic sites are identified using DS Analyze Binding Site tool.

**Molecular docking:**
The homology models of PgHSC70 and OsHSP70 were analyzed for docking with brassinolide. Using DS LibDock docking simulation was performed. The obtained confirmations were then summarized and analyzed for interactions. The interactions showing highest scores and docking energy were considered best for protein-ligand complex structure.
Figure 1: Gene structural organization of the PgHsc70 and OsHSP70 genes.

Figure 2: MEME identified consensus sequence motifs in PgHSC70 and OsHsp70 cds. Ten motifs were obtained using MEME software. The motifs are indicated with filled boxes of different colors and numbered in order from 1 to 10.

Figure 3: Phylogenetic tree showing relationships of HSP70s from pearl millet and rice.
Results & Discussion:
Comparative alignment of genomic and cDNA nucleotide sequences of \textit{PgHsc70} and \textit{OsHSP70} genes is shown in Figure 1. The intron phases consisted of phase 0, 1 and 2. Exon count and intronic phase distribution was similar in both the genes. Two exons were present in both the genes and their distribution was found at phase 0 (exon 1) and phase 2 (exon 2). Phase 1 intronic phase distribution was found in both the genes. The intron length varied with \textit{OsHSP70} intron of 1935 bp while \textit{PgHSC70} intron is 141 bp. The MEME tool identified 10 significant conserved motifs as shown in Figure 2. The length of conserved motifs varied from 21 to 50 amino acids. The consensus sequence motifs identified are given in Table 1.

In silico analysis of cis-regulatory elements (CREs) in the CDS sequence of \textit{PgHSC70} and \textit{OsHsp70} genes revealed different elements in the upstream region (Table 2). In \textit{PgHSC70} a total of 33 CAREs whereas, in \textit{OsHSP70}, 29 CAREs were identified and a few were uniquely present in each gene. Cis-elements were found responsive to light, meristem-specific activation, abscisic acid, methyl jasmonate, gibberellins, low temperature, seed-specific regulation, root-specific expression, anoxia, and circadian regulation. The CCGTCC motif, ABRE, W box, CCAAT-box, GC motif, G-Box, MYB recognition site STRE, MYB, TGA-element, plant_AP-2-like, WRE3, and unnamed-1 are the common CREs found in both the genes. All of them were present on sense strand with matrix value ≥ 5.

The plant small RNA target analysis server (psRNATarget) was used to predict miRNA target sites. The miRNAs comprising target sites in \textit{PgHSC70} and \textit{OsHSP70} genes were identified with expectation score lower than 4.0 (Table 3). MSA using ClustalW was constructed by aligning seven \textit{PgHSC70} protein sequences along with twelve \textit{OsHsp70} protein sequences. Phylogenetic tree based on MSA is shown in Figure 3. Comparative phylogenetic analysis of \textit{PgHSC70} and \textit{OsHsp70} revealed major groups of \textit{HSP70} genes with paralogous as well as orthologous genes. Each group contained both \textit{PgHSC70} and \textit{OsHsp70}.

![Secondary structure analysis of PgHsc70 using SOPMA](image1)

Figure 4: Secondary structure analysis of PgHsc70 using SOPMA (h-helix, e-extended strand, c-random coil and t-beta turn)
Table 1: Consensus sequence motifs in PgHsc70 and OsHsp70 CDS

| motif | E-value | Sites | Width | Consensus sequence |
|-------|---------|-------|-------|--------------------|
| 1     | 1.0e-032| 10    | 50    | GAGAAGCTTCCTCGAGCCGAGTCCTCACAGTCTCACACAGAAGT |
| 2     | 1.4e-025| 9     | 50    | TCGAATCTCAGGACCGGCTCGAAGATCCGAGAATTCGAGGAGAGCTGCA |
| 3     | 1.2e-019| 10    | 50    | TCGAATCCACATCGAGGCTCGGAGGAGAGCTGCAATCT |
| 4     | 5.8e-010| 8     | 36    | TCGAATCCACATCGAGGCTCGGAGGAGAGCTGCAATCT |
| 5     | 3.6e-009| 10    | 39    | GTGACACAGCTCAAGGAAGAGACTGCAGGAGAGAGCTGCAATCT |
| 6     | 1.4e-004| 9     | 21    | TCGAATCCACATCGAGGAGGAGGCAGGAGCT |
| 7     | 1.4e-001| 6     | 38    | TGAGAAGATCGAGACCTGCAAGCTGCAATCT |
| 8     | 3.3e+002| 8     | 21    | TGAGAAGATCGAGACCTGCAAGCTGCAATCTG |
| 9     | 2.0e+004| 2     | 27    | CAATGGTGAGGACCTCGAGAGGAGGAG |
| 10    | 6.8e+004| 2     | 21    | CAATGGTGAGGACCTCGAGAGGAGGAG |

Figure 5: Secondary structure analysis of OsHsp70 with SOPMA (h-helix; c-random coil; e-extended strand and t-beta turn).
Figure 6: Phosphorylation plot obtained from netphosK 3.1 server of protein (a) PgHsc70 and (b) OsHsp70.

Table 2: The cis-acting regulatory elements in PgHSC70 and OsHSP70 gene promoter regions

| Site Name                  | Matrix score | Sequence                  | Function                                      |
|----------------------------|--------------|---------------------------|-----------------------------------------------|
| ABRE                       | 9            | GCCCGGTCGGC               | abscisic acid responsiveness                  |
|                            | 5            | ACGTG                    |                                               |
| CCAAT-box                  | 6            | CAACCGG                  | MYBHv1 binding site                           |
| CCGTCC motif               | 6            | CCGTCC                   |                                               |
| G-Box                      | 6            | CAACGTG                  | light responsiveness                          |
| GC-motif                   | 6            | CCCCGG                   | enhancer-like element involved in anoxic specific inducibility |
| MYB                        | 6            | CAACAG                   |                                               |
| MYB recognition site       | 6            | CCGTG                    |                                               |
| STRE                       | 5            | AGCGG                    |                                               |
| TGA-element                | 6            | AAGCGAC                  | auxin-responsive element                      |
| W box                      | 6            | TTGACC                   |                                               |
| WRE3                       | 6            | CACCT                    |                                               |
| plant_AP-2-like            | 8            | CGACCAGG                 |                                               |
| Unnamed_1                  | 5            | CGTG                     |                                               |
|                            |              | CAREs present only in Pg |                                               |
| A-box                      | 6            | CCGTCC                   | cis-acting regulatory element                  |
| AC1                        | 8.5          | (T/C)(T/C)(C/T)ACC(T/C)ACC |                                               |
| DRE core                   | 6            | GCCGAC                   |                                               |
| Myb-binding site           | 6            | CAACAG                   | MeJA-responsiveness                           |
| CGTCA-motif                | 5            | CGTCA                    |                                               |
| GATA-motif                 | 9            | AAGGATAAGG               | part of a light responsive                     |
| Element                      | Sequence                  | Description                                                                 |
|-----------------------------|---------------------------|-----------------------------------------------------------------------------|
| MYB-like sequence           | TAACCA                    | Cis-element related to meristem specific activation                        |
| NON-box                     | AGATCGACG                 | Light responsive element                                                   |
| Sp1                         | GGCCGG                    | cis-element involved in the MeJA-responsiveness                            |
| TGACG-motif                 | TGACG                     | MeJA-responsiveness                                                        |
| Unnamed_1_6                 | GCTGCCCGCTC               | CAREs present only in Os                                                    |
| as-1                        | TGACG                     | MYB binding site involved in drought-inducibility                          |
| DREI                        | ACCGAGA                   |                                                                             |
| MBS                         | CAACTG                    |                                                                             |
| Myb                         | CAACCTG                   |                                                                             |
| Unnamed_2                   | CCCCCGG                   |                                                                             |
| box S                       | AGCCACC                   |                                                                             |

Figure 7: N-glycosylation plot obtained from NetNglyc server of protein (a) PgHsc70 and (b) OsHsp70.
Figure 8: STRING network representing the predicted functional partners of the protein (a) PgHsc70 and (b) OsHsp70

Table 3: miRNAs predicted by psRNATarget

| miRNA     | Target       | Cleavage |
|-----------|--------------|----------|
| gma-miR1533 | X67711.2  | 2.0     | 12.416 | 1 |
| bdi-miR5059 | X67711.2  | 2.0     | 18.103 | 1 |
| osa-miR1852 | X67711.2  | 2.5     | 17.488 | 1 |
| gma-miR4374a | X67711.2 | 2.5     | 24.309 | 1 |
| mtr-miR2673a | X67711.2  | 3.0     | 11.695 | 1 |
| mtr-miR2673b | X67711.2  | 3.0     | 11.695 | 1 |
| osa-miR2919 | X67711.2  | 3.0     | 6.097  | 2 |
| ath-miR5658 | X67711.2  | 3.0     | 12.416 | 1 |
| bra-miR5720 | X67711.2  | 3.0     | 14.377 | 2 |
Table 4: Amino acid composition of PgHsc70 and OsHsp70

| Amino acid | PgHsc70 N | % | Os-Hsp70 N | % | Amino acid | PgHsc70 N | % | Os-Hsp70 N | % |
|------------|-----------|---|------------|---|------------|-----------|---|------------|---|
| Alanine    | 58        | 8.9 | 58          | 9.0 | Lysine     | 52        | 8.0 | 52          | 8.0 |
| Arginine   | 30        | 4.6 | 30          | 4.6 | Methionine | 17        | 2.6 | 17          | 2.6 |
| Asparagine | 34        | 5.2 | 33          | 5.1 | Phenylalanine | 22       | 3.4 | 21          | 3.2 |
| Aspartic acid | 48  | 7.4 | 47          | 7.3 | Proline    | 26        | 4.0 | 25          | 3.9 |
| Cysteine   | 7         | 1.1 | 7           | 1.1 | Serine     | 35        | 5.4 | 29          | 4.5 |
| Glutamine  | 23        | 3.5 | 25          | 3.9 | Threonine  | 39        | 6.0 | 40          | 6.2 |
| Glutamic acid | 52  | 8.0 | 51          | 7.9 | Tryptophan | 3         | 0.5 | 3           | 0.5 |
| Glycine    | 52        | 8.0 | 55          | 8.5 | Tyrosine   | 14        | 2.2 | 15          | 2.3 |
| Histidine  | 5         | 0.8 | 5           | 0.8 | Valine     | 42        | 6.5 | 43          | 6.6 |
| Isoleucine | 44        | 6.8 | 48          | 7.4 | Pyrolysine | 0         | 0.0 | 0           | 0.0 |
| Leucine    | 46        | 7.1 | 44          | 6.8 | Selenocysteine | 0       | 0.0 | 0           | 0.0 |

N represents total number and % the numeric percentage of each amino acid.
Figure 9: Intrinsic disordered regions of (a) PgHsc70 and (b) OsHsp70 calculated using DisEMBL server. Disordered residues are colored and ordered residues are black.

Figure 10. 3D structure of modeled PgHSC70 and OsHsp70
Figure 11: Brassinolide hydrogen bonding interactions in (a) PgHsc70 and (b) OsHsp70 proteins. Dotted green lines represent hydrogen bonds.

Table 5: Physicochemical properties of PgHSC70 and OsHsp70 predicted proteins

| Parameter                                      | Value          |
|------------------------------------------------|----------------|
| Molecular Weight                               |                |
| PgHsc70                                        | 71105.48       |
| OsHsp70                                        | 70997.56       |
| Amino Acid Length                              |                |
| PgHsc70                                        | 649            |
| OsHsp70                                        | 648            |
| pI                                             |                |
| PgHsc70                                        | 5.10           |
| OsHsp70                                        | 5.16           |
| Total negatively charged residues (Asp + Glu)  |                |
| PgHsc70                                        | 100            |
| OsHsp70                                        | 98             |
| Total positively charged residues (Arg + Lys)   |                |
| PgHsc70                                        | 82             |
| OsHsp70                                        | 82             |
| Extinction coefficients                        |                |
| all pairs of Cys residues form cystines        | 37735          |
| all Cys residues are reduced                   | 37360          |
| Instability Index (II)                         |                |
| PgHsc70                                        | 34.52          |
| OsHsp70                                        | 33.69          |
| Aliphatic Index (AI)                           |                |
| PgHsc70                                        | 81.79          |
| OsHsp70                                        | 83.56          |
| Half-life                                      |                |
| Mammalian reticulocytes                        | 30 hrs         |
| Yeast                                          | >20 hrs        |
| E. coli                                        | >10 hours      |
| GRAVY                                          | -0.427         |

The Expasy’s ProtParam tool determined the length (649 and 648 aa), mol. wt. (71105.48 and 70997.56 Da), isoelectric point (pI = 5.10 and 5.16), the total negative and positive charged residues of PgHSC70 and OsHsp70 proteins (100 and 98 and 82 and 82, respectively). Proteins are found be acidic (pI < 7). The amino acid composition analysis revealed that Ala (58) was the most abundant amino acid in both the sequences, while Trp (3) was the least abundant (Table 4). PgHsc 70 had the atomic composition C<sub>2777</sub>H<sub>4999</sub>N<sub>861</sub>O<sub>995</sub>S<sub>24</sub>, while OsHsp70 had the atomic composition C<sub>3108</sub>H<sub>5007</sub>N<sub>861</sub>O<sub>987</sub>S<sub>24</sub>.

The calculated extinction coefficient (EC) of PgHsc70 and OsHsp70 proteins was found to correlate with the cysteine, tryptophan and tyrosine content at 280 nm as 37735 and 39225 (assuming all pairs of cysteine residues form cysteines) and 37360 and 38850 (assuming all cysteine residues are reduced) M<sup>-1</sup>cm<sup>-1</sup>, respectively. The instability index (II) values were 34.52(PgHsc70) and 33.69(OsHsp70) classifying both proteins as stable. Similarly, the aliphatic index (AI) had high value of 81.79 (PgHsc70) and 83.56 (OsHsp70), indicating proteins are thermo stable.

The estimated half-lives for both proteins were 30 h in mammalian reticulocytes (in vitro), >20 h in yeast (in vivo) and >10 h in Escherichia coli (in vivo). The GRAVY was -0.427 (PgHsc70) and -0.399 (OsHsp70), respectively. Both proteins are highly water soluble. Table 5 shows the predicted physicochemical properties of the PgHSC70 and OsHsp70. In Figures 4 and 5, the secondary structure of protein sequences of PgHSC70 and OsHsp70, predicted using SOPMA server are shown. The evaluated percentage α-helices, β-turn, extended strand, and random coils with output

The calculated extinction coefficient (EC) of PgHsc70 and OsHsp70 proteins was found to correlate with the cysteine, tryptophan and tyrosine content at 280 nm as 37735 and 39225 (assuming all pairs of cysteine residues form cysteines) and 37360 and 38850 (assuming all cysteine residues are reduced) M<sup>-1</sup>cm<sup>-1</sup>, respectively. The instability index (II) values were 34.52(PgHsc70) and 33.69(OsHsp70) classifying both proteins as stable. Similarly, the aliphatic index (AI) had high value of 81.79 (PgHsc70) and 83.56 (OsHsp70), indicating proteins are thermo stable.

The estimated half-lives for both proteins were 30 h in mammalian reticulocytes (in vitro), >20 h in yeast (in vivo) and >10 h in Escherichia coli (in vivo). The GRAVY was -0.427 (PgHsc70) and -0.399 (OsHsp70), respectively. Both proteins are highly water soluble. Table 5 shows the predicted physicochemical properties of the PgHSC70 and OsHsp70. In Figures 4 and 5, the secondary structure of protein sequences of PgHSC70 and OsHsp70, predicted using SOPMA server are shown. The evaluated percentage α-helices, β-turn, extended strand, and random coils with output
that cysteine residues were present in the position 20, 272, 318, 325, of flexible glycine and proline amino acids in the proteins. Interactions. Also, high coiled structural content might be because of flexible glycine and proline amino acids in the proteins. The high percentage of random coils indicates protein flexibility and more interactions. Also, high coiled structural content might be because of flexible glycine and proline amino acids in the proteins.

Using String, the interacting partners predicted in both PgHsc70 and OsHsp70 is shown in the Figure 8. From the analysis, the functional partners observed in the string network of PgHsc70 protein were HSFA2, HSFI, HSP101, HSP90.1, Hsp81.4, Hsp3, Hsp81-3, LOS1, J3 and Hop1. The functional partners observed in the string network of OsHsp70 are Osj-17347, OSIIT0703900-01, Osj_11911, HSP81-2, Osj_12871, DJA6, and OsS04T0107900-02. These interactions give some insights into understanding the functioning of these proteins in response to heat stress and tolerance. Using DisEMBL the predicted intrinsic disorder regions (IDRs) of PgHSC70 and OsHsp70 is shown in Figure 9.

Homology modeling was done to predict the 3-D structures of PgHSC70 and OsHsp70 based on the template structure HSC70 (PDB ID: 1YUW) from bovine, at a resolution of 2.6 Å deposited in PDB. The template protein had identity of 81% (PgHsc70) and 80.18 % (OsHsp70). The initial models of PgHsc70 and OsHsp70 proteins were built using the crystal coordinates information of the template 1YUW. The models generated by DS, were scored with discrete optimized protein energy (DOPE) geometric function, and the model with the lowest DOPE score was taken as final model as shown in Figure 10. After the proteins were energy minimized, the final models were validated using PROCHECK. Corresponding to core regions most favorable Psi/Phi value combinations are present in the darkest areas in Ramachandran plot. Each of the protein models displayed 90% accuracy. Overall, the homology model of the PgHsc70 have 94.6% of the residues occurring in most favored region, 4.4 % in allowed regions, and only 1.1 % of the residues in disallowed regions. In comparison, the OsHsp70 homology model have 94.6% residues in favored region, 4.9 % residues in allowed region and 0.5% residues in outlier region.

Molecular docking of PgHsc70 and OsHsp70 with brassinolide was studied in order to identify the critical interactions and their variation. Using LibDock the docking results for brassinolide on PgHsc70, showed high binding affinity with score of 115.231 and binding energy of 0.0019 kcal/mol, in comparison target OsHsp70 showed a LibDock score of 146.59 and binding energy of -26.586 kcal/mol. In the PgHSC70-Brassinolide complex, the electrostatic and =87.3 and -835.7 kcal/mol of van der Waals energies respectively and for OsHsp70-Brassinolide complex -83.32 and 4.139 kcal/mol were found to be higher. Docking analysis revealed both H-bonds and close interactions within the docked site of PgHSC70 and OsHsp70 (Figure 11). The PgHSC70-Brassinolide complex formed five hydrogen bonds, 3 with residue THR271, each one with LYS59 and LYS277, and the closest interactions are also found with the amino acid residue GLY236. Whereas the OsHsp70-Brassinolide complex formed two hydrogen bonds with two residues ASP35 and LYS129 and found to interact with the amino acid residue ILE133. The docking studies clearly indicated that the

The sub cellular localization of proteins PgHsc70 and OsHsp70 predicted by CELLO was found to be cytosolic in nature. The SignalP analysis revealed that none of the proteins have any of signal peptide. Using CYS_REC tool the cysteine residues in the proteins determined revealed that the protein PgHsc70 contain cysteine residues in the positions 20, 272, 319, 326, 366, 483 and 609. But all are not involved in disulfide bonding; Cys326 is probably SS-bond with a score of 1.9. On the other hand, OsHsp70 revealed that cysteine residues were present in the position 20, 272, 318, 325, 365, 482 and 608 and probably Cys326 is SS-bonded with a score of 1.9. At the molecular level the disulfide bridges presence is a positive factor for stability. Table 7 shows the results obtained using CYS_REC tool.

NetPhos 3.1 & NetNglyc servers identify post-translational modification sites. NetPhos retrieved the information related to kinases PKC, Unsp, CK1, cdk5, CKII, PKA, RSK, DNAPK, CK1, PKG, EGFR, SRC and cdc2 involved in the phosphorylation of both proteins and an extra one kinase INSR unique to OsHsp70. In both the proteins high score of 0.691 predicted for site having threonine and the percentage amino acids composition obtained is shown in Figures 6a and 6b. Using NetNglyc server, the N-glycosylation sites (38 NRTT, 155 NDSQ, 423 NTTI and 493 NVSA) of the PgHsc70 protein were predicted with a score of 0.66, 0.57, 0.58 and 0.74 respectively. The N-glycosylation sites of OsHsp70 protein predicted were (38 NRTT, 155 NDSQ, 422 NTTI and 492 NVSA) with scores of 0.6378, 0.5774, 0.5874 and 0.7403. The plot of predicted N-glycosylation sites is shown in Figures 7a and 7b.

width 70 is given in Table 6. From the computed percentage of each conformation, α-helix predominated, followed by extended strand random coil and random coil in both the proteins. The high percentage of random coils indicates protein flexibility and more interactions. Also, high coiled structural content might be because of flexible glycine and proline amino acids in the proteins.

Table 6: Secondary structure analysis of PgHsc70 and Os-Hsp70

| Parameters | PgHsc70 | Os-Hsp70 |
|-----------|---------|----------|
| Alpha helix (Hh) | N | % | N | % |
| 206 | 286 | 44.07 | 278 | 42.90 |
| 3- helix (Gg) | 0 | 0.00 | 0 | 0.00 |
| Pi helix (Ii) | 0 | 0.00 | 0 | 0.00 |
| Beta bridge (Bb) | 0 | 0.00 | 0 | 0.00 |
| Extended strand (Ee) | 119 | 18.34 | 117 | 18.06 |
| Beta turn (Tt) | 43 | 6.63 | 45 | 6.94 |
| Bend region (Ss) | 0 | 0.00 | 0 | 0.00 |
| Random coil (Cc) | 201 | 30.97 | 208 | 32.10 |
| Ambiguous states | 0 | 0.00 | 0 | 0.00 |
| Other states | 0 | 0.00 | 0 | 0.00 |

N represents total number and % the numeric percentage of each amino acid.
ligand and receptor were bound together closely to stabilize complex structure in OsHsp70 and PgHsc 70 as done in our previous study [29].

Conclusions:
We have documented the characterization of Hsp70 gene family members, PgHSC70 and OSHsp70 genes, and their proteins sequences in pearl millet and rice respectively. The results indicated conserved relationships and distinct functions of PgHSC70 and OSHsp70 highlighting the wide participation of these family members in environmental adaptation. Data from docking analysis of the homology models with brassinolide is also reported.

Conflicts of Interest:
Authors declare no conflict of interest.

Acknowledgements:
We greatly thank financial support of the funding from DST-PURSE-II, UPE-FAR, CAS-II, and RUSA to A.S and UGC-SRF to G.B. We thank Bioinformatics Division of Averin Biotech for the technical support rendered towards carrying out the work.

References:
[1] Pandey P et al. Frontiers Plant Science. 2017 18 537. [PMID: 28458674]
[2] Raza A et al. Plants (Basel). 2019 8:34. [PMID: 30704089]
[3] Lamalakshmi Devi E. et al. In: Ghurbanpour M., Varma A. (eds) Medicinal Plants and Environmental Challenges. Springer, USA 2017: 359-413. [DOI:10.1007/978-3-319-68717-9_20]
[4] Verma et al. In Stress Signaling in Plants: Genomics and Proteomics Perspective (eds. Maryam Sarwat, Altaf Ahmad, MZ Abdin) Springer 2013 1: 25.
[5] Rejeb IB, Plants (Basel), 2014 3:458. [PMID: 27135514]
[6] Mishra D et al. In: Regulation of Heat Shock Protein Responses. Heat Shock Proteins (eds. Asea A., Kaur P.) Springer, 2018 vol. 13; 41-69. [DOI:10.1007/978-3-319-74715-6_3]
[7] Yamaguchi-Shinozaki K, Shinozaki K. Annual Review of Plant Biology. 2006 57:781. [PMID: 16669782]
[8] Morozova O, Marra MA. Genomics. 2008 92: 255. [PMID: 18703132].
[9] Nicky J et al. Journal of Experimental Botany. 2012 63:3523. [PMID: 22467407].
[10] Nakajima Y, Suzuki S. International Journal of Molecular Science, 2013 14:7771. [PMID: 23574938].
[11] Saibil H Nature Reviews Molecular Cell Biology. 2013 14: 630. [PMID: 2402605]
[12] Yu A et al. Bio Med Research International. 2015 5859:1.
[13] Murthy VS, Ravishankar KV. In: Abiotic Stress Physiology of Horticultural Crops (eds. Rao N., Shivashankara K, Laxman R). Springer, New Delhi 2016 71.
[14] Efeoglu B. Gazi University Journal of Science; 2009 22:67. [https://dergipark.org.tr/en/pub/gujs/issue/7389/96762]
[15] Ahuja I et al. Trends Plant Science. 2010 15:664. PMID: 20846898.
[16] Timperio AM et al. Journal of Proteomics. 2008 71:391. [PMID: 18718564]
[17] Sugino M, et al. Plant Science. 1999 146:81.
[18] Alvim FC, et al. Plant Physiology. 2001 126:1042. [PMID: 11457955]
[19] Ono K, et al. Plant Science. 2001 160: 455.
[20] Waters, Elizabeth R. The Biochemist. 2014 36:9. [https://doi.org/10.1042/BIO0601009]
[21] Daugaard M, et al. FEBS Letters. 2007 581:3702. [PMID: 17544402]
[22] Sung DY, et al. Physiology Plantarum. 2001113:443.
[23] Tran NT, et al. Journal of Animal Science Technology. 2015 57:44. [PMID: 26672478]
[24] Fietto Luciano G, et al. Genetics Molecular Biology. 2007 30:881. [DOI:10.1590/S1415-47572007000500017]
[25] Usman MG, et al. Biotechnology and Genetic Engineering Reviews. 2017 33:26. [PMID: 28649918].
[26] Griffith AA, Holmes W. International Journal Molecular Science. 2019 20:4207. [PMID: 31466231]
[27] Sarwar, H. Journal of Cereals and Oilseeds. 2013 4:32. [https://doi.org/10.5897/JCO12.023]
[28] Awika, JM. In Advances in Cereal Science: Implications to Food Processing and Health Promotion (eds Josephph M. Awika, Vieno Piironen, Scott bean). American Chemical Society, 2011 1089:1. [DOI: 10.12691/jfmr-8-10-9]
[29] Baloji G et al. (2019) Bioinformation. 2019 15:131. [PMID: 31435159]
