

Molecular Cloning and Differential Expression of Cytosolic Class I Small Hsp Gene Family in *Pennisetum glaucum* (L.)

Palakolanu Sudhakar Reddy^{1,2} · Kiran K. Sharma¹ ·
Vincent Vadez¹ · Malireddy K. Reddy²

Received: 28 November 2014 / Accepted: 26 March 2015
© Springer Science+Business Media New York 2015

Abstract Small heat shock protein (Hsp) family genes have been reported in several plant species that function as molecular chaperones to protect proteins from being denatured in extreme conditions. As a first step towards the isolation and characterization of genes that contribute to combating abiotic stresses particularly heat stress, construction and screening of the subtracted complementary DNA (cDNA) library is reported here. In this study, a subtractive heat stress cDNA library was constructed that was used to isolate members of small Hsps (sHsps) using *PgsHsp17.9A* gene as a probe. As a result, a total of 150 cDNA clones were isolated from the subtracted cDNA library screening, leading to 121 high-quality expressed sequence tags (ESTs), with an average size of 450 bp, comprising of 15 contigs, and majority of these isolated sHsp genes belong to cytosolic class I (CI) family. In silico sequence analysis of CI-sHsp family genes revealed that the length of sHsp proteins varied from 151 to 159 amino acids and showed large variation in isoelectric point value (5.03 to 10.05) and a narrow range of molecular weight (16.09 to 17.94 kDa). The real-time PCR results demonstrated that CI-sHsp genes are differentially expressed in *Pennisetum* leaves under different abiotic stress conditions particularly at high temperature. The results presented in this study provide basic information on PgCI-sHsp family genes and form the foundation for future functional studies of these genes.

Keywords *Pennisetum glaucum* · High temperature stress · CI-sHsp gene family · Library screening · Real-time PCR

✉ Palakolanu Sudhakar Reddy
p.sudhakarreddy@cgiar.org

¹ International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad 502324 Telangana, India

² Plant Molecular Biology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110 067, India

Abbreviations

sHsp	Small heat shock protein
ACD	Alpha crystallin domain
Pg	<i>Pennisetum glaucum</i>
CI	Cytosolic class I

Introduction

Understanding plant responses to environmental stress are complex and appear to be a difficult task due to several technical limitations. Current knowledge of the governing regulatory network during development and environmental stress responses is fragmentary and an understanding of the damage caused by these factors or the plant's tolerance mechanisms to deal with stress-induced damages are far from complete [1]. An understanding of the molecular mechanism of these responses to adaptation are therefore essential to characterize different components involved in stress tolerance and is therefore of great practical and basic importance. Plants activate a series of reactions in response to different stresses or combination of stresses and many other chemical stressors, including rapid accumulation of heat shock proteins. Heat shock genes are the primary products of the heat shock proteins which act as molecular chaperones that play an essential role in protecting cellular functions under stressful conditions [2, 3]. Most likely, all cellular proteins have to interact with molecular chaperones at least once in their lifetime, such as during synthesis, maturation, subcellular targeting or degradation in an ATP-independent manner [2]. Even though it is essential for plant survival to maintain the proper folding of its proteins, relatively little information on the molecular and biochemical roles of molecular chaperons is available in the millets.

Heat shock proteins (Hsps) are the members of multi-gene families in which not all members are regulated by high temperature, but many of the Hsps are also found in the unstressed cells. Therefore, Hsp proteins are necessary components of both stressed and unstressed cells and crucial for the responses of the cells to stress. Based on their size, Hsps are classified into high molecular weight Hsps comprising Hsp100, Hsp90, Hsp70/DnaK, Hsp60/GroE and low molecular weight Hsps or small heat shock proteins (sHsps) of 16–42 kDa [4]. Compared to other classes, sHsps in plants have a relatively complex diversity at DNA sequence, structure, function and cellular localization [5, 6]. This diverse and important family includes a conserved alpha-crystallin domain (ACD) flanked by variable N-terminal region and a conserved C-terminal region. These three distinct regions are believed to contribute to different aspects of sHsp function. All sHsps share a conserved 90-amino acid C-terminal domain called the ACD as the region delimited from the $\beta 2$ strand to the $\beta 9$ strand [7]. The ACD is the core structure of the protein and is formed by two conserved regions that form a sandwich of two β -pleated sheets which can dimerize forming the building block of the large oligomers [8]. With the availability of information on genome sequences of more organisms, the numbers of sHsp genes identified have increased lately. In silico sequence analysis of sHsp encoding genes predicted 24 members in sugarcane, 19 in *Arabidopsis*, 23 in *Oryza sativa* and 36 in *Populus* [9, 10]. Based on their site of action and localization, they have been identified as 11 conserved subfamilies in the angiosperm genomes [8, 10, 11]. These include six subfamilies such as cytoplasmic/nuclear localized (CI–CVI) and five sHsp subfamilies, which are localized in different organelles. In addition, sequence analysis of cytosolic Hsps in plants reveals the presence of specific amino acid motifs [12]. Cytosolic class I (CI)

sHsps are the most abundant proteins in the genome and shown to be responsible for up to 1 % of all expressed proteins in heat-stressed soybean cells [13].

Recent transcriptomic studies have greatly helped in elucidating the functional role of sHsps in response to different environmental stresses, including drought, heat, cold, salt, osmotic stress and oxidative stresses [11, 14–19]. sHsps are also highly expressed in developmental stages like embryogenic tissues, pollen maturation, zygotic embryogenesis, seed maturation and seed development and growing fruits [11, 20, 21]. In vivo heterologous expression studies of sHsps can also enhance the stress tolerance [22–26]. *Arabidopsis* plants expressing the chloroplast *sHsp21* were more resistant to heat and high light stresses than the wild types [27]. In addition, introduction of the carrot *Hsp17.7* into potato enhanced the stability of cellular membrane and tuberization in vitro [28]. This revealed that the function of sHsps in plants is more diversified. Higher diversification of sHsp classes and members of each class may reflect the adaptation to stress conditions which is a unique feature of plants, since the plants being sessile cannot escape from the stress environment and hence may have evolved unique mechanisms to overcome stress. However, the functional differences between these members and how they relate with function are not yet clear. Therefore, the expression of sHsp may reflect an increased requirement of sHsp chaperones for folding/refolding of other proteins during the plant development and different abiotic stress conditions. Although sHsps have been widely studied in different plant systems particularly in the cereals, it is not completely known how these sHsps play a role in different abiotic stress environments in the hardy, drought tolerant, orphan crop-like *Pennisetum glaucum*.

Pennisetum glaucum (L.), a monocot, belongs to the Poaceae family and has been used as a model system. As *Pennisetum* grows in arid climate, there is a high probability that this plant contains a large number of stress alleviating heat shock genes particularly Hsps that provide stress tolerance to this crop. Therefore, to isolate the genes that are important for abiotic stress tolerance particularly heat stress, we isolated the complementary DNAs (cDNAs) encoding for different isoforms of CI-sHsp family in *Pennisetum glaucum*, by constructing and screening the subtractive cDNA library. Subtractive cDNA library screening leading to cloning of 121 high-quality expressed sequence tags (ESTs), comprising of 15 contigs and majority of these isolated sHsp genes belong to cytosolic class I (CI) family. By analyzing the stress responsiveness of these sHsp genes, we confirmed abiotic stress inducible nature of these by qRT-PCR analysis. These results imply that the combination of subtractive cDNA library construction, screening and qRT-PCR can be a useful tool for identification of stress-inducible transcripts from orphan crop plants in which availability of tools is minimum. The results of this work would provide a foundation to better understand functional role of the sHsp gene family in millets and further their role in stress adaptation. Genes from this plant will be of great advantage to transfer to other cereals and crop plants to make them stress tolerant.

Materials and Methods

Plant Growth Conditions and Abiotic Stress Treatments

Pennisetum seedlings were grown on wet germination paper, which was soaked in a beaker containing nutrient solution, under a 14/10 h light/dark cycle at 28/20 °C day/night temperature for 14 days. A large number of seedlings than needed for sampling were grown so that homogenous seedlings were used for treatment and sampling. Chosen seedlings were

harvested and frozen in liquid nitrogen. For the subtractive cDNA library construction, high temperature stress treatment was applied by exposing 14-day-old seedling to 45 °C for different time intervals, i.e. 0.5, 1, 2, 4, 6, 8 and 10 h. For the qRT-PCR, two sets of 14-day-old plants were grown: One set was used as a fully irrigated control (plants were irrigated with water) and the other set was used for different types of abiotic stress treatments. Salt stress treatment was imposed by dipping the seedling in an aerated nutrient solution containing 250 mM salt solution with different time intervals, i.e. 1, 4, 8, 12, 24 and 36 h. Drought stress was imposed by removing the beaker containing the nutrient solution and on which the germination paper rolls were held. The germination paper and the seedlings were kept on a greenhouse table for 24 h prior to collecting leaf samples at different time intervals, i.e. 12, 24, 48 and 72 h. For cold and high temperature stress, two separate sets of seedlings were kept at 4 and 45 °C, respectively, in an incubator, followed by sample harvesting with different time intervals, i.e. 0.5, 1, 2, 4, 6, 8 and 10 h. Plants were kept under control conditions for the same duration as the heat, cold, salt and drought treatment at 28±1 °C. In all conditions, three biological replicates were collected for each sample and immediately frozen in liquid nitrogen and then stored in -80 °C for RNA isolation.

Preparation of Probe by PCR Labelling Method

The full-length *PgHsp17.9A* fragment was labelled with α -[32P]-deoxycytidine triphosphate (dCTP) to a specific activity of $>1 \times 10^8$ cpm/ μ g DNA by PCR using the specific primers. Approximately 10 ng of the cloned *PgHsp17.9A* cDNA in plasmid as a template and 200 μ M each of dATP, dTTP, dGTP and 20 μ M dCTP were taken and supplemented with 100 μ Ci of [α 32P] dCTP and 200 nM of each gene-specific primer and 2.5 U of Taq DNA Polymerase (Invitrogen) in a final volume of 50 μ l reaction. The following cycling conditions were used for the PCR reaction: 94 °C for 4 min, 30 cycles of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min with a final extension of 10 min at 72 °C. The cDNA library was screened by the plaque hybridization method. Many independent plaques produced strong hybridization signals.

Total RNA Isolation and cDNA Synthesis

Total RNA was extracted from the control and abiotic stress-treated samples according to the manufacturer's instructions (Qiagen, Germany). The total RNA concentration and purity were measured on a NanoVue plus spectrophotometer (GE Healthcare, USA). The integrity of RNA was checked on a 1.4 % agarose gel. Two micrograms of total RNA was used for cDNA synthesis using SuperScript III (Invitrogen), according to the manufacturer's instructions. The cDNA was then diluted to 1:12 dilution with nuclease-free water as template for qRT-PCR.

Subtractive cDNA Library Construction

Total RNA from control and high temperature stress-treated seedlings was isolated according to Chomczynski and Sacchi [30]. Purification of poly(A)+RNA from the control and heat-stressed (equal amount of different time intervals of RNA) samples was carried out from total RNA according to Mishra et al. [29]. For subtractive hybridization, 5 μ g of the first-strand control (driver) cDNA was mixed with 1 μ g of stressed (tester) tissue messenger RNA

(mRNA) and hybridization performed for 3 h at 65 °C in hybridization buffer (10 mM Tris and 0.5 M KCl). The unhybridized (differentially expressed) poly(A)+RNA was used to synthesize double-stranded cDNA and unidirectionally ligated to a lambda-ZAP vector, in vitro packaged and further allowed to infect XL1blue MRF *Escherichia coli* cells according to the manufacturer's instructions (Stratagene, California).

Screening of cDNA Library

The plaques to be screened were first replica-plated on to a nitrocellulose filter disc that had been placed on the surface of an agar plate prior to inoculation. A reference set of these plaques on the master plate was retained. The filter bearing the plaques was removed and treated with alkali to lyse the plaques and denature the DNA. The filter was then treated with proteinase K to remove protein and denatured DNA bound to the nitrocellulose since it has a high affinity in the form of a 'DNA print' of the plaques. The DNA was fixed firmly by baking the filter at 80 °C. Membranes were hybridized with the radiolabelled probe at 65 °C for 24 h in 6× SSC, 0.1 % sodium dodecyl sulfate (SDS) and 1 % Denhardt's solution. After hybridization, the filters were washed for 15 min at 65 °C with 2.0× SSC, 0.1 % SDS, and for 15 min at 65 °C with 0.1× SSC, 0.1 % SDS. Filters were then exposed to X-ray film for 24 h. A plaque whose DNA print gave a positive autoradiographic result could be picked from the reference plate. A total of 150 positive clones were identified and purified their DNA according to Reddy et al. [31], and samples were sequenced by the commercial Macrogen facility (Macrogen, Seoul, Korea).

Vector Trimming and Assembling

The input raw files generated for 150 clones from automated sequencing facility were converted to FASTA format using MacVector Assembler (V13.0.5) with results in 15 contigs. The Phred and Phrap programmes [32, 33] were used to remove vector and poor quality sequences. Only the sequences longer than 150 bp that were generated through trimming process as mentioned above were included in the dataset. Functional annotations of these sequences were carried out using Blast X, Blast N and Blast P programmes of NCBI. All the processed Hsp sequences were submitted to GenBank at NCBI.

Sequence and Phylogenetic Analysis of the CI-sHsp Family Genes

In silico sequence analysis like number of amino acids, molecular weight (M.Wt) and theoretical isoelectric point (pI) of the all *Pennisetum* small Hsps were predicted by using MacVector software. Subcellular localization of these proteins was predicted by using several algorithms like pSORT and TargetP. Information regarding ORF length and intron numbers was confirmed by sequencing of respective cDNA and genomic clones. The motifs and conserved domain analysis was performed using Pfam and InterProScan search program (<http://www.ebi.ac.uk/Tools/es/cgi-bin/iprscan>). Out of the 15 sHsp genes, 12 full-length sHsp genes were included in phylogenetic analysis. CI-sHsp sequences were aligned using ClustalW MacVector program. An unrooted neighbour-joining phylogenetic tree was constructed in ClustalW based on the full-length sequence of the proteins from *Pennisetum*, rice and *Arabidopsis*.

Isolation and Cloning of the CI-sHsp Genomic Clones

Genomic DNA of pearl millet plants was isolated using the NucleoSpin Plant II DNA isolation kit (Macherey-Nagel, Duren, Germany), following the manufacture's recommended procedure. The genomic fragments of CI-sHsp genes were amplified using respective full-length gene specific primers. PCR was carried out using 200 nM of each primer along with 200 μ M of each dNTPs, 2.0 U of Taq DNA polymerase and 150 ng of genomic DNA as a template in a 25- μ l reaction. PCR samples were amplified in an Eppendorf Thermal Cycler with an initial denaturation at 95 °C for 5 min followed by 30 cycles of 95 °C for 1 min, 62–64 °C for 1 min and 72 °C for 2 min, followed by a final extension at 72 °C for 10 min, and PCR products were tested by using 1 % agarose gel electrophoresis. The PCR amplified genomic fragments were cloned into the pCR4-TOPO TA vector (Invitrogen). The cloned genomic fragments were completely sequenced. Sequence alignments were performed with respective cDNA sequences using the MacVector programme.

PCR Primer Design and Quantitative Real-Time PCR Analysis

Gene-specific primers were designed by using the Beacon Designer software with the following criteria: T_m 60 °C, 18–23 bp length, designed one primer in the untranslated region (UTR) and other one in the coding region of the gene because of their highly conserved nature in the coding region, GC content between 40 to 60 % which generates unique PCR products between 175 and 250 bp. The sequence of each primer pair is given in the Table 1. PCR reactions contained 5 μ l of 2 \times SYBR Green master mix reagent (KAPA SYBR FAST Bio-Rad I Cycler 2X), 1 μ l cDNA and 200 nM of each gene-specific primer and reference gene in a final volume of 20 μ l. To reduce pipetting errors, an electronic Gilson multi-channel pipette was used to dispense the cDNA and the master mix (SYBR Green mix and primers). The following standard thermal profile was used for all PCR reactions: 95 °C for 3 min, 40 cycles of 95 °C for 15 s and 60 °C for 15 s. Amplicon dissociation curves, i.e. melting curves were recorded after cycle 40 by heating from 60 to 95 °C in 20 min with a ramp speed of 1.75 °C/min. At the end of the PCR, a melt curve analysis of the products was performed to check the specificity of PCR amplification. The experiments were repeated three times independently, and the data from these experiments were averaged. Relative expression profiling of the selected genes was carried out according to the REST software [34]. The data were normalized using the genes encoding for β -tubulin and 40 s ribosomal protein (S24) as reference genes for internal controls based on 'geNorm' software [35].

Results and Discussion

Cloning and Sequence Analysis of the sHsp Family Genes

The mRNA populations from the seedlings exposed to heat stress were mixed with approximately 5-folds of complementary first-strand control cDNA were heat denatured and separated as reported earlier [29]. The unbound leftover mRNA was used to synthesize the cDNA library in the UniZAP vector using cDNA synthesis kit (Stratagene, USA). Further, we performed the library screening to isolate more number of sHsp isoforms using homologous *PgCI-sHsp17.9A* gene as a probe. Plaques with positive signals were purified as reported

Table 1 Primers used in the real-time PCR

Name of the gene	Primer sequence
PgsHsp17.7	5' ACCCCTCAGTTCAATCACCGATAAG 3' 5' GGAGAAGGGGTCTGAACACGT 3'
PgsHsp16.9A	5' GCTGCGAACAACACTTCAGCACA 3' 5' GCACTACGGAGCGGAAGATG3'
PgsHsp17.6	5' CTTTCAGAGCTCAGCAGTTCGCTCAA 3' 5' CAGAGGTCGAGGGAGAAGGG3'
PgsHsp16.9B	5' ACTCAGTGGCTTGACACAGTAATT 3' 5' GGAACGGAAGATGTCGTTGAAGG 3'
PgsHsp17.1	5' GCTGCAGGAATTTTCATTAGGCA 3' 5' CCGGACTAGGGAACGGAAG3'
PgsHsp17.0A	5' AGAGTCTTCTACCAATCCACAC 3' 5' GACGGACGGGATCAGGGAG3'
PgsHsp17.0B	5' CACCGTGAAGTTCAGTTCATC 3' 5' GGGGTCTCCTCCAGTCGATG3'
PgsHsp 17.9A	5' CCTCCAAATCCCACCGACCA3' 5' GGGAACGAGGGAAAGAGGCT3'
PgsHsp17.0C	5' GCTGCGAACAACGCTTCAGCAAA 3' 5' CGGAAGATGTCGTCGAAGGG 3'
PgsHsp17.8	5' CCCAATCTCCCGCGTTACTCAGA 3' 5' ACGAGGGGAAGAGGCTGTTG3'
β -tubulin	5' GTGCTCTGAATGTGGATGTGAATG 3' 5' ACCAACCTCCTCATAGTCTTCTC 3'
S24	5' ACGAGGAGGGTAAGACGGTG 3' 5' CAAATGGTCTGGAGACATGAA 3'

earlier [31] and sequenced. Approximately 150 redundant small Hsps were obtained from library screening, and all the sequences were assembled using a MacVector *assembly* programme (V13.05) that resulted in non-redundant 15 contigs. The accession numbers for these genes are available in the GenBank from KF421827 to KF421840. Each sHsp was named according to its molecular weight in kDa followed by a Roman numerical (I or II) to classify them as members of the cytosolic I or II sHsp gene family (Table 2). Pfam analysis suggested that protein sequences of the PgCI-sHsp family had a typical single major domain. Sequence analysis of PgsHsp protein sequences showed large variations in length of amino acids (151 to 159), pI value (ranging from 5.03 to 10.05) and molecular weight (ranging from 16.09 to 17.94 kDa). Subcellular predictions were also based on the phylogenetic affinity or relationship of the proteins to other proteins with experimentally determined cellular locations [36]. The above predictions revealed that 14 PgsHsp genes belonged to cytosolic class (Table 2). The translation initiation region of the PgsHsp open reading frame is situated within a sequence, CCATGTC, which resembles the plant consensus initiation sequence [37], and the consensus polyadenylation signal (AATAAA) is found in the 3'UTR sequences. Comparative sequence alignment of the genomic and cDNA sequences of PgCI-sHsps revealed an absence of introns in the CI-sHsps. The options of intron absence or presence of an intron with smaller size may be correlated to rapid induction needs of sHsp genes [38]. Incidentally, all PgCI-sHsp genes are rapidly expressed within 10 min of high temperature stress (Fig. 3).

Table 2 Details of pearl millet sHsp genes cloned from the library screening

S. no	Gene name	GenBank Acc. No.	Full/partial	Class	Protein (AA)	Molecular weight (kDa)	pI	Subcellular localization
1	PgsHsp17.7	KF421827	Full	CI	156	17.67	5.88	Cytosol
2	PgsHsp16.9A	KF421828	Full	CI	150	16.94	6.56	Cytosol
3	PgsHsp17.6	KF421829	Full	CI	156	17.56	5.88	Cytosol
4	PgsHsp16.9B	KF421830	Full	CI	151	16.91	10.26	Cytosol
5	PgsHsp17.1	KF421831	Full	CI	151	17.08	7.70	Cytosol
6	PgsHsp17.0A	KF421832	Full	CI	151	17.02	9.04	Cytosol
7	PgsHsp17.0B	KF421833	Full	CI	152	16.97	5.86	Cytosol
8	PgsHsp 17.9A	KF421834	Full	CI	159	17.93	5.88	Cytosol
9	PgsHsp17.0C	KF421835	Full	CI	151	17.05	7.70	Cytosol
10	PgsHsp17.8	KF421836	Full	CI	158	17.79	5.87	Cytosol
11	PgsHsp17.9B	KF421837	Full	CI	158	17.94	5.88	Cytosol
12	PgsHsp17.0D	KF421838	Full	CI	151	17.05	5.87	Cytosol
13	PgHsp 11	KF421839	Partial	CP	199	–	–	Chloroplast
14	PgHsp 12	–	Partial	CII	55	–	–	Cytosol
15	PgHsp 13	KF421840	Partial	CI	71	–	–	Cytosol

The table shows the following details: putative sHsp gene name, NCBI accession number, full-length or partial, amino acid length, predicted molecular mass for the deduced proteins, isoelectric point (pI) and predicted subcellular localization

Sequence Conservation and Divergence Among Small Hsps

Unlike the high molecular weight Hsps, sHsps are highly divergent at the sequence level and in oligomerization [39, 40]. The sHsps are highly conserved, across the protein families, at the carboxyl-terminal (C-terminal) domain than at the amino-terminal (N-terminal) domain. The significant feature of the sHsp is the presence of an evolutionarily conserved ACD domain comprised of 80–90 amino acids at the C-terminus and probably forms a distinct structural and functional unit. The plant small Hsps have a specific ACD and divided into conserved consensus I and consensus II domains separated by a hydrophilic domain of variable lengths [20]. All PgCI-sHsp proteins shared a eukaryotic specific consensus region I and also shared a consensus region II, which is not present in other eukaryotic sHsp proteins (Fig. 1). The N-terminal region preceding the ACD had variable length and amino acid composition that contributed to structural diversity among different members of sHsps. The PgsHsp alignment data clearly showed higher conservation of C-terminal domain in comparison to N-terminus region (Fig. 1). Motif ‘I-P-I/V’ in the C-terminal extension (β 10 strand) was thought to be structurally and functionally important for the solubility required for chaperone activity [41]. The residues Pro-X (14)-Gly-Val-Leu in consensus region I is a conserved signature motif in almost all sHsp proteins whereas a similar motif Pro-X (14)-X-Val/Leu/Ile-Val/Leu/Ile also appeared in the consensus region II [42]. Arginine (R) at 159 position in sHsps represented the most conserved site. This Arg in *Pennisetum* is located at a similar position of α -crystalline structure in wheat. Apart from this, in the N-terminal domain (amino acids 1–62), there is a family-specific conserved region. The class I cytosolic proteins have a consensus region at the N-terminus region (amino acids 7–22), and class II cytosolic proteins have a small conserved region (amino acids 53–64), which do not present in

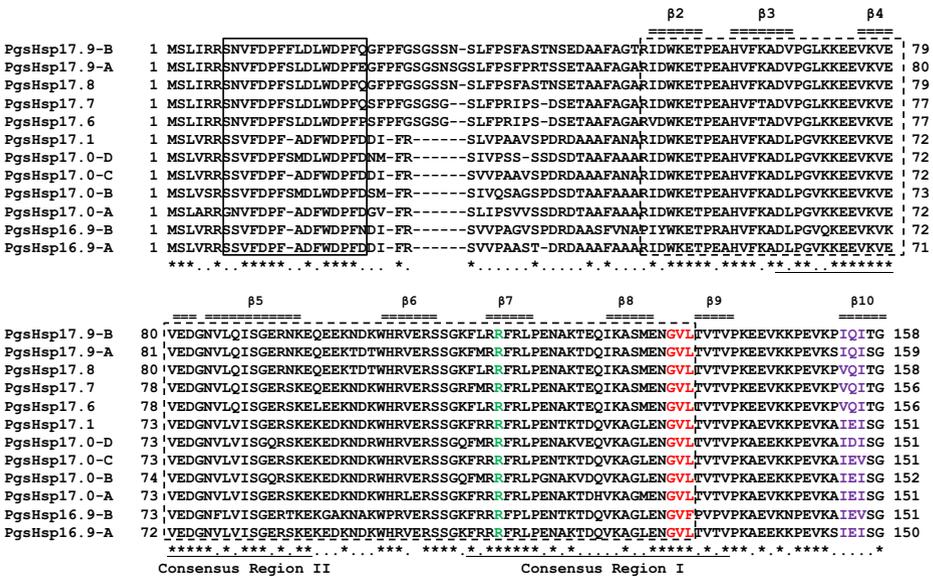


Fig. 1 Multiple alignment of cytosolic CI-sHsps. The conserved α -crystalline domain was labelled with *dotted box*. The defined consensus regions I and II are marked with *underline* below the sequences. Highly conserved and semi-conserved regions are shown in ‘asterisks’ and ‘dot’, respectively. Small Hsp region specific to CI subclass is labelled with *thick box*. The conserved Arg is displayed in green colour in the $\beta 7$ strand. The secondary structure assignments for all CI-sHsps were labelled above the sequences. The predicted β -strands depicted by *thick lines* above the alignment are based on their position in known secondary structure of Hsp16.9 from *Triticum aestivum* and PSIPRED software. The IXI/V motif in the C-terminal extension is shaded with *grey colour*. The SXXFD motif and interacting residues in the conserved alpha crystalline domain is in *italics* and shaded with *violet colour*. Alignment was performed by using MacVector programme

the other protein classes (data not shown). Similarity, between individual PgsHsps belonging to different classes ranged from 62 to 80 %, whereas the similarity between individual PgsHsps within the classes ranged from 85 to 99 %. Sequence alignment of PgsHsp proteins showed that members of different subfamilies do not share high sequence similarity; yet, the secondary structure is conserved across subfamilies. The structural features $\beta 3$, $\beta 4$, $\beta 5$, $\beta 8$ and $\beta 9$ are highly conserved (Fig. 1). The GVL sequence motif in $\beta 9$ is conserved across sHsps from archaea to eukaryotes, but it does not exist in any of the cytosolic V family members. However, the corresponding structure $\beta 9$ is conserved in these proteins.

Phylogenetic Analysis of PgCI-sHsp Gene Family Members

The CI-sHsp genes were distributed across the genomes, and sequence similarity among these genes is maintained by strong selection [43, 44]. In all genomes, the CI-sHsp is the most numerous, but the evolutionary forces that dominated varied across the genomes [10]. The phylogenetic relationships among the CI sequences were not always congruent with organismal relationships. The phylogenetic relationships within the CI subfamily deserve particular attention. Most notable finding is a high number of isoforms of CI-sHsps present in *Populus trichocarpa* genome (18 genes) and *P. glaucum* (13 genes) in comparison to *Arabidopsis thaliana* (6 genes) and *O. sativa* (7 genes). The CI genes in *P. glaucum* are more closely related to *O. sativa* in comparison to *A. thaliana*. In *P. glaucum*, here are two groups of CI genes (Fig. 2). One group

◀ **Fig. 2** Phylogenetic tree of cytosolic class I sHsp amino acid sequences. The analysis shows the relative closeness of 12 cytosolic full-length PgCI-sHsps with rice and *Arabidopsis* sHsp families. This phylogenetic tree was constructed using the deduced amino acid sequences of 19 sHsps from *Arabidopsis* and 23 sHsps from rice using Clustal W program of MacVector software. Protein sequence accession numbers for each gene were given under brackets on the right of the figure. Most of the PgsHsps fall into the cytosolic subgroup and is closer to rice sHsp proteins

comprised of *PgsHsp16.9A*, *16.9B*, *17.1*, *17.0A*, *17.0B*, *17.0C* and *17.0D* are closely related to *OsHsp16.9A-CI*, *16.9B-CI* and *16.9C-CI* and the other group consists of *PgsHsp17.7*, *17.6*, *17.9-A*, *17.8* and *17.9-B* are more closely related to *O. sativa OsHsp17.4-CI*, *17.7-CI*, *17.9A-CI* in comparison to *AtHsp18.1-CI*, *17.6-CI* and *17.4-CI* (Fig. 2). Similar types of findings were observed in barley [45]. Cytosol has large number of sHsp family genes in comparison with other cell organelles, implying that cytosol might be the primary site of action for the function of several of these sHsps. The phylogenetic analyses indicates that major cytosolic gene families form monophyletic groups and are most likely the result of gene duplications that occurred before the diversification of the angiosperms. Most likely, the expansion of the CI-sHsps is driven by local events or a number of single duplication events followed by strong selection to maintain the duplicate CI copies. Additionally, the CII and III subfamilies are found to be more closely related to each other than they are to the CI subfamily (data not shown).

Expression Profile of PgCI-sHsp Genes Under Different Abiotic Stress Conditions

In plants, a multi-gene family encodes sHsps with a very high sequence homology among its family members. Each member responds differentially to environmental and developmental cues. Therefore, it is very difficult to monitor the expression of these genes individually by Northern analysis due to cross hybridization. In *Arabidopsis* and *O. sativa*, where the complete genome sequences are available, transcript analysis has been shown during heat stress and other developmental stages for different Hsps encoding gene families [11, 46, 47]. However, detailed transcript analysis of the sHsp family from *Pennisetum* during abiotic stress conditions have not been carried out so far. To monitor the comparative expression profiling of all the sHsps under various abiotic stress conditions, we performed real-time PCR analysis to detect the accurate transcript expression level of CI-sHsp family genes in *Pennisetum*. Since sHsp family genes are highly conserved during evolution, it is difficult to design primers in their coding region. Hence, we designed forward primers in the coding region and reverse primers in the untranslated regions of each sHsp without affecting the cross hybridization among different isoforms and/or families. PgCI-sHsp genes exhibited diverse expression patterns in response to different abiotic stress conditions.

Pennisetum CI-sHsp genes have relatively faster response kinetics than the other species Hsp gene family members [48]. The regulation of Hsp gene expression has been shown to occur at the transcriptional level and in an autoregulatory manner [49]. Thus, after reaching a peak induction during prolonged exposure to heat shock, the levels of Hsp gene transcripts were reduced [48]. The relative upregulation of PgCI-sHsp genes have higher and very rapid expression in response to high temperature stress in *Pennisetum* seedlings. *PgsHsp17.7* gene is highly upregulated (up to 2311-fold) throughout at all the time points at 45 °C in comparison to other Hsps and other stress conditions (Fig. 3). However, some genes (*PgsHsp17.9A*, *17.0C* and *17.8*) have been found to be less upregulated by heat (45 °C) treatment in comparison to other sHsp genes like *PgsHsp17.7*, *16.9A*, *17.6*, *16.9B* and *17.1* (Fig. 3). Overall expression of *PgsHsps* under heat stress followed three kinetic expressions: type I, where the expression was up by 500-folds (represented by *PgsHsp17.9A*, *17.0C* and *17.8*); type II, where the expression

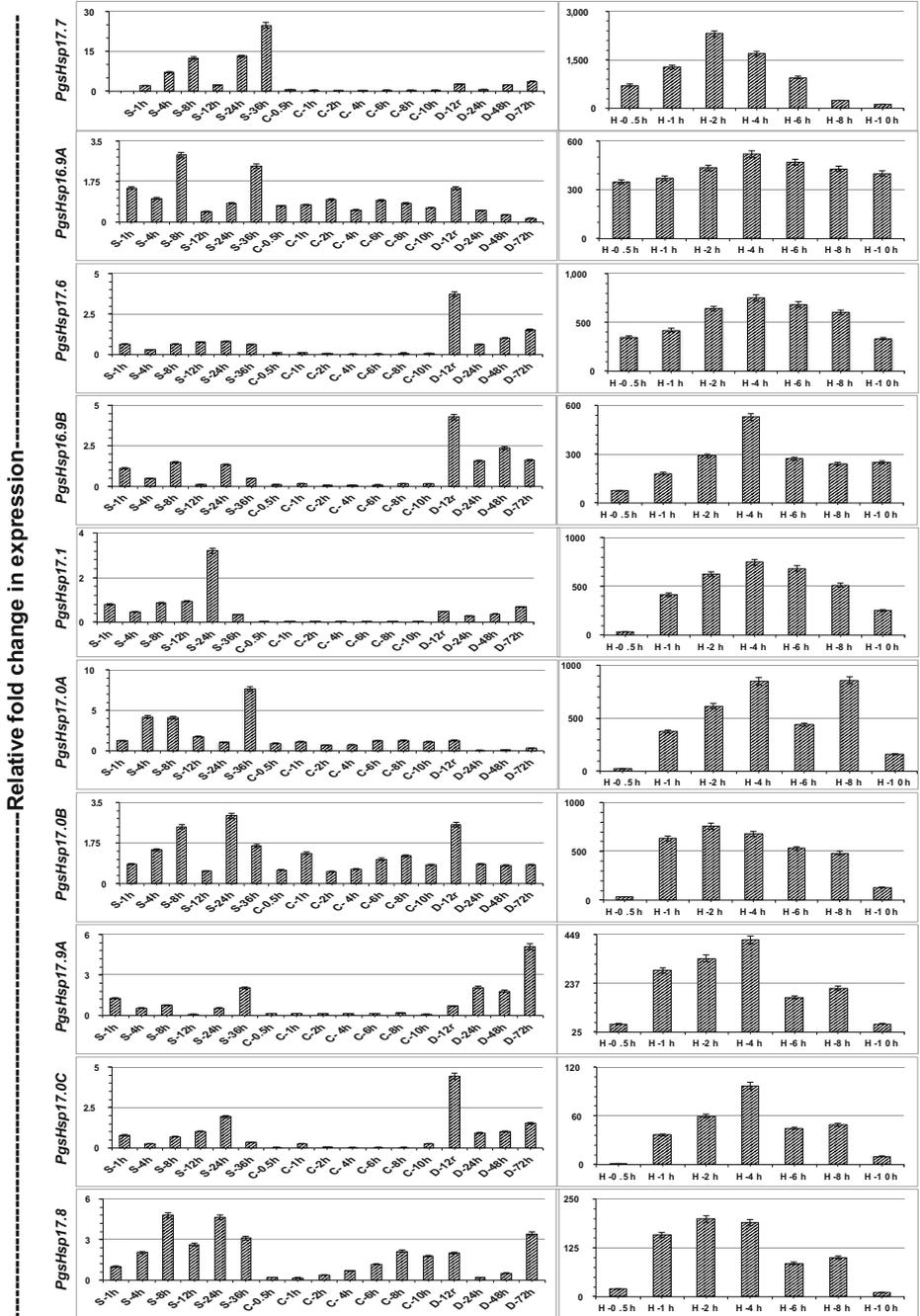


Fig. 3 Relative quantification of PgCl-sHsps under different abiotic stresses. Relative expression of PgCl-sHsp transcripts under different abiotic stress conditions in comparison to its control as revealed by quantitative RT-PCR analysis. Values represent the expressions values obtained after normalizing against control value. *Error bars* represent standard deviations from three biological replicates. *S* salt stress, *C* cold stress, *D* drought stress, *H* heat stress

was registered up to 1000-folds (represented by *PgsHsp16.9A*, *16.9B*, *17.1*, *17.0A* and *17.0B* and *17.6*) and type III, where the expression level was up to 2500-folds (represented by *PgsHsp17.7*). The differential expressions of *PgCI-sHsp* genes were also observed in response to dehydration, salinity and low temperature (cold) stress. In response to cold stress (4 °C), most of the Hsps are downregulated significantly except *PgsHsp17.0A*, *17.0B* and *17.8*, which remained unchanged or downregulated up to 6-h cold treatment and later upregulated to a lower extent (Fig. 3). Salinity stress led to the upregulation of all the Hsps except *PgsHsp17.6*, *17.1* and *17.9A*. The sHsp expression levels progressively upregulated after 4 h of post-salt stress treatment. *PgsHsp17.7* was upregulated significantly in comparison to *PgsHsp16.9A*, *17.0A* and *17.8*. *PgsHsps17.7*, *16.9A*, *17.0A* and *17.8* significantly upregulated at 4, 8, 24 and 36 h time points in comparison to others. Interestingly, many PgCI-sHsps, which are highly expressed during heat stress, were also found to be preferentially upregulated under salt stress (Fig. 3). During dehydration stress, transcript level of *PgsHsp17.7* was highly upregulated whereas the expression levels of the *PgsHsp16.9A*, *16.9B*, *17.0B* and *17.0C* dropped after 12 h. *PgsHsp17.1* and *17.0A* were significantly downregulated under drought stress (Fig. 3). In summary, the expression profiles revealed that there were differences in the transcription regulation and transcript stability among the *PgCI-sHsp* family genes. Although they belonged to the same family, they have different expression patterns from gene to gene. While some sHsps showed low expression, others expressed predominantly high under different abiotic stress conditions. Transcript profiling of PgCI-sHsp family transcripts in response to different environmental stresses positively correlates with the duration and intensity of a stress condition. This implies that CI-sHsp gene subfamily reflects its functional diversity. These results broaden our knowledge of the molecular mechanisms behind the action of CI-sHsp gene subfamily in pearl millet plant stress acclimatization. This study provides the basis towards the future studies of CI-sHsp genes under diverse environmental conditions. Hence, it can be speculated that this differential regulation of PgCI-sHsp family genes might have a positive role in protecting plants against a wide range of environmental stresses.

Conclusions

Differential heat stress-regulated EST database for *Pennisetum* under heat stress conditions was generated where 15 PgsHsps were obtained through library screening. Phylogenetic analysis revealed that majority of the *PgsHsps* belong to CI-sHsp family. Phylogenetic analysis with other angiosperm genomes revealed that *PgCI-sHsp* genes are highly evolutionarily conserved and also have a high homology to monocot species. The transcript data of 10 full-length CI-sHsps clearly showed the complex expression patterns of these genes under different abiotic stress conditions. To our knowledge, this is the largest Hsp EST database from *Pennisetum* to date with one of the most comprehensive transcript analyses of a large number of stress inducible Hsps under different abiotic stress conditions in addition to heat stress in *Pennisetum glaucum*.

Acknowledgments We thank Prof. Kavi Kishor for critical reading of the manuscript. This work was supported in part by a research grant from the Department of Biotechnology (Ministry of Science and Technology, Government of India) to MKR. PSR acknowledges the Department of Science and Technology, Govt. of India, New Delhi, for financial support through the INSPIRE Faculty Fellowship Award Grant No. IFA-LSPA-06. This work was undertaken as part of the CGIAR Research Program on Dryland Cereals.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Battisti, D. S., & Naylor, R. L. (2009). Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, *323*, 240–244.
2. Tyedmers, J., Mogk, A., & Bukau, B. (2010). Cellular strategies for controlling protein aggregation. *Nature Reviews Molecular Cell Biology*, *11*, 777–788.
3. Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*, 324–332.
4. Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, *9*, 244–252.
5. de Jong, W. W., Caspers, G. J., & Leunissen, J. A. (1998). Genealogy of the alpha-crystallin-small heat-shock protein super family. *International Journal of Biological Macromolecules*, *22*, 151–162.
6. Franck, E., Madsen, O., van Rheede, T., Ricard, G., Huynen, M. A., & de Jong, W. W. (2004). Evolutionary diversity of vertebrate small heat shock proteins. *Journal of Molecular Evolution*, *59*, 792–805.
7. Poulain, P., Gelly, J. C., & Flatters, D. (2010). Detection and architecture of small heat shock protein monomers. *PLoS One*, *5*, e9990.
8. Bondino, H. G., Valle, E. M., & Ten Have, A. (2012). Evolution and functional diversification of the small heat shock protein/ α -crystallin family in higher plants. *Planta*, *235*, 1299–1313.
9. Borges, J. C., Cagliari, T. C., & Ramos, C. H. (2007). Expression and variability of molecular chaperones in the sugarcane expressome. *Journal of Plant Physiology*, *164*, 505–513.
10. Waters, E. R., Aevermann, B. D., & Sanders-Reed, Z. (2008). Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress & Chaperones*, *13*, 127–142.
11. Sarkar, N. K., Kim, Y. K., & Grover, A. (2009). Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics*, *10*, 393.
12. Waters, E. R. (1995). The molecular evolution of the small-heat-shock proteins in plants. *Genetics*, *141*, 785–795.
13. Hsieh, M. H., Chen, J. T., Jinn, T. L., Chen, Y. M., & Lin, C. Y. (1992). A class of soybean low molecular weight heat shock proteins: immunological study and quantitation. *Plant Physiology*, *99*, 1279–1284.
14. Chang, Y. Y., Liu, H. C., Liu, N. Y., Hsu, F. C., & Ko, S. S. (2006). *Arabidopsis* Hsa32, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. *Plant Physiology*, *140*, 1297–1305.
15. Hamilton, E. W., & Heckathorn, S. A. (2001). Mitochondrial adaptations to NaCl Complex I is protected by anti-oxidants and small heat shock proteins, whereas complex II is protected by proline and betaine. *Plant Physiology*, *126*, 1266–1274.
16. Sato, Y., & Yokoya, S. (2008). Enhanced tolerance to drought stress in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7. *Plant Cell Reports*, *27*, 329–334.
17. Sun, W., Van Montagu, M., & Verbruggen, N. (2002). Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta*, *1577*, 1–9.
18. Weston, D. J., Karve, A. A., Gunter, L. E., Jawdy, S. A., Yang, X., Allen, S. M., & Wulschleger, S. D. (2011). Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. *Plant, Cell and Environment*, *34*, 1488–1506.
19. Volkov, R. A., Panchuk, I. I., Mullineaux, P. M., & Schöfl, F. (2006). Heat stress-induced H₂O₂ is required for effective expression of heat shock genes in *Arabidopsis*. *Plant Molecular Biology*, *61*, 733–746.
20. Waters, E. R., Lee, G. J., & Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants. *Journal of Experimental Botany*, *47*, 325–338.
21. Wehmeyer, N., & Vierling, E. (2000). The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests general protective role in desiccation tolerance. *Plant Physiology*, *122*, 1099–1108.
22. Kotak, S., Vierling, E., Baumlein, H., & von Koskull-Doring, P. (2007). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of *Arabidopsis*. *Plant Cell*, *19*, 182–195.
23. Yeh, C. H., Chang, P. F. L., Yeh, K. W., Lin, W. C., Chen, Y. M., & Lin, C. Y. (1997). Expression of a gene encoding a 16.9-kDa heat-shock protein, Oshsp16.9, in *Escherichia coli* enhances thermotolerance. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 10967–10972.
24. Soto, A., Allona, I., Collada, C., Guevara, M. A., Casado, R., Rodriguez-Cerezo, E., Aragoncillo, C., & Gomez, L. (1999). Heterologous expression of a plant small heat-shock protein enhances *Escherichia coli* viability under heat and cold stress. *Plant Physiology*, *120*, 521–528.
25. Song, N. H., & Ahn, Y. J. (2011). DcHsp17.7, a small heat shock protein in carrot, is tissue-specifically expressed under salt stress and confers tolerance to salinity. *Nature Biotechnology*, *28*, 698–704.

26. Jiang, C., Xu, J., Zhang, H., Zhang, X., Shi, J., Li, M., & Ming, F. (2009). A cytosolic class I small heat shock protein, RcHSP17.8, of *Rosa chinensis* confers resistance to a variety of stresses to *Escherichia coli*, yeast and *Arabidopsis thaliana*. *Plant, Cell and Environment*, *32*, 1046–1059.
27. Hamdahl, U., Hall, R. B., Osteryoung, K. W., Vierling, E., Bornman, J. F., & Sundby, C. (1999). The chloroplast small heat shock protein undergoes oxidation-dependent conformational changes and may protect plants from oxidative stress. *Cell Stress & Chaperones*, *4*, 129–138.
28. Ahn, Y. J., & Zimmerman, J. L. (2006). Introduction of the carrot HSP17.7 into potato (*Solanum tuberosum* L.) enhances cellular membrane stability and tuberization in vitro. *Plant, Cell and Environment*, *29*, 95–104.
29. Mishra, R. N., Reddy, P. S., Nair, S., Markandeya, G., Reddy, A. R., Sopory, S. K., & Reddy, M. K. (2007). Isolation and characterization of expressed sequence tags (ESTs) from subtracted cDNA libraries of *Pennisetumglaucaum* seedlings. *Plant Molecular Biology*, *64*, 713–732.
30. Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, *162*, 156–159.
31. Reddy, P. S., Nair, S., Mallikarjuna, G., Kaul, T., Markandeya, G., Sopory, S. K., & Reddy, M. K. (2008). A high-throughput, low-cost method for the preparation of “sequencing-ready” phage DNA template. *Analytical Biochemistry*, *376*, 258–261.
32. Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*, 186–194.
33. Gordon, D., Abajian, C., & Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Research*, *8*, 195–202.
34. Pfaffl, M. W., Horgan, G. W., & Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, *30*, e36.
35. Roy, N. V., Paepe, A. D., & Speleman, F. (2003). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*, 0034.1–0034.11.
36. Heazlewood, J. L., Tonti-Filippini, J. S., Gout, A. M., Day, D. A., Whelan, J., & Millar, A. H. (2004). Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell*, *16*, 241–256.
37. Joshi, C. P. (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Research*, *15*, 6643–6653.
38. Jeffares, D. C., Penkett, C. J., & Bahler, J. (2008). Rapidly regulated genes are intron poor. *Trends in Genetics*, *24*, 375–378.
39. Vierling, E. (1991). The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *42*, 579–620.
40. Kappe, G., Leunissen, J. A., & de Jong, W. W. (2002). Evolution and diversity of prokaryotic small heat shock proteins. *Progress in Molecular and Subcellular Biology*, *28*, 1–17.
41. van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C., & Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature Structural and Molecular Biology*, *8*, 1025–1030.
42. Caspers, G. J., Leunissen, J. A., & de Jong, W. W. (1995). The expanding small heat-shock protein family, and structure predictions of the conserved “alpha-crystallin domain”. *Journal of Molecular Evolution*, *40*, 238–248.
43. Tranbarger, T. J., & Misra, S. (1996). The molecular characterization of a set of cDNAs differentially expressed during Douglas-fir germination and early seedling development. *Physiologia Plantarum*, *95*, 456–464.
44. Ohta, T. (1991). Multigene families and the evolution of complexity. *Journal of Molecular Evolution*, *33*, 34–41.
45. Reddy, P.S., Kavi Kishor, P.B., Seiler, C., Kuhlmann, M., Eschen-Lippold, L., Lee, J., Reddy, M.K., Sreenivasulu, N. (2014). Unraveling regulation of the small heat shock proteins by the heat shock factor HvHsfB2c in barley: its implications in drought stress response and seed development. *PLoS One*, *4*, e89125.
46. Krishna, P., & Gloor, G. (2001). The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress & Chaperones*, *6*, 238–246.
47. Nover, L., Bharti, K., Doring, P., Mishra, S. K., Ganguli, A., & Scharf, K. D. (2001). *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress & Chaperones*, *6*, 177–189.
48. Guan, J. C., Jinn, T. L., Yeh, C. H., Feng, S. P., Chen, Y. M., & Lin, C. Y. (2004). Characterization of the genomic structures and selective expression profiles of nine class I small heat shock protein genes clustered on two chromosomes in rice (*Oryza sativa* L.). *Plant Molecular Biology*, *56*, 795–809.
49. Schoffl, F., Prandl, R., & Reindl, A. (1998). Regulation of the heat-shock response. *Plant Physiology*, *117*, 1135–1141.