Send Orders for Reprints to reprints@benthamscience.ae

Current Genomics, 2015, *16*, 279-291

Genome-wide Scanning and Characterization of *Sorghum bicolor* L. Heat Shock Transcription Factors

M. Nagaraju[1,#], Palakolanu Sudhakar Reddy[2,#], S. Anil Kumar[1], Rakesh K. Srivastava[2], P. B. Kavi Kishor[1,*] and D. Manohar Rao[1]

[1]*Department of Genetics, Osmania University, Hyderabad 500 007, India;* [2]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502324, Hyderabad, India*

P. B. Kavi Kishor

Abstract: A genome-wide scanning of *Sorghum bicolor* resulted in the identification of 25 *SbHsf* genes. Phylogenetic analysis shows the ortholog genes that are clustered with only rice, representing a common ancestor. Promoter analysis revealed the identification of different *cis*-acting elements that are responsible for abiotic as well as biotic stresses. Hsf domains like DBD, NLS, NES, and AHA have been analyzed for their sequence similarity and functional characterization. Tissue specific expression patterns of *Hsfs* in different tissues like mature embryo, seedling, root, and panicle were studied using real-time PCR. While *Hsfs*4 and 22 are highly expressed in panicle, 4 and 9 are expressed in seedlings. *Sorghum* plants were exposed to different abiotic stress treatments but no expression of any *Hsf* was observed when seedlings were treated with ABA. High level expression of Hsf1 was noticed during high temperature as well as cold stresses, 4 and 6 during salt and 5, 6, 10, 13, 19, 23 and 25 during drought stress. This comprehensive analysis of *SbHsf* genes will provide an insight on how these genes are regulated in different tissues and also under different abiotic stresses and help to determine the functions of *Hsfs* during drought and temperature stress tolerance.

Keywords: Heat shock transcription factors, Phylogenetic analysis, *Cis*-acting elements, Paralogs, Molecular chaperones, Abiotic stress.

1. INTRODUCTION

High temperature and drought have adverse effects on water relations, photosynthesis and results in 50% crop reduction [1]. In response to heat stress, rapid accumulation of small heat shock proteins (Hsps) was observed in all eukaryotes and plants. Hsps act as molecular chaperones and prevent the aggregation and denaturation of proteins [2]. Heat shock transcription factors (*Hsfs*) transcriptionally regulate the *Hsp* genes. Plant *Hsfs* play a central role in the heat stress response. Tomato *HsfA1*, *A2*, and *A3* confer heat stress tolerance when overexpressed [3-5]. *LpHsfA1a* and *AtHsfA2* enhance thermotolerance upon overexpression but abolished when knocked-out or interfered [6, 7]. Transcription factor *A2* has been found as a key regulator in response to many environmental stresses [8]. In *Arabidopsis*, overexpression of HsfA4a leads to decreased production of cytosolic H_2O_2 scavenging ascorbate peroxidase (APX) and it was hypothesized that *Hsfs* may act as H_2O_2 sensors in the plants [9]. HSFA1D, HSFA2, and HSFA3 act as key factors in regulating *APX2* expression during diverse stress conditions [9]. Overexpression of *AtHsfA1b*-gusA in transgenic tomato plants led to the constitutive expression of Hsps, elevated levels of APX activity, with enhanced heat and chilling tolerance. *Hsfs* are also induced by other abiotic stresses like salinity, temperature, cold, and metal [10]. Overexpression of *OsHsfA2e* and *AtHsfA3* showed tolerance to salt stress [11, 12] but *HsfA3* conferred enhanced thermotolerance and salt hypersensitivity in germination in *Arabidopsis* [13]. While *HsfA1b* (*AtHsfA1b*) gene is involved in chilling tolerance in tomato [14, 15], *OsHsfA4a* is involved in cadmium tolerance in rice and wheat [16]. Besides imparting abiotic stress tolerance, several heat shock factors are also involved with disease resistance and developmental activities. *HsfB1* and *HsfB2b* are associated with pathogen resistance in *Arabidopsis* [17]. Further, HsfA9 was reported to be essential for embryogenesis and seed maturation in sunflower and *Arabidopsis* [18, 19]. *Hsfs* bind to the conserved *cis*-acting (5'-nGAAn-3') heat shock elements (HSE) of the promoters. At least 3 HSE are required for better interaction with Hsf. Based on homology and conservation of domains, plant *Hsfs* are classified into three classes. When compared with fungi and animals, plants have many *Hsf* genes [20, 21]. Genome-wide screening of many plants resulted in the identification of 16 to 35 *Hsfs* depending on the species [22-24].

The *Hsf* gene family has not been characterized in *Sorghum bicolor*. But, functional and evolutionary relationship between organisms can be studied only when multiple sequences of these families are available for alignment and phylogenetic analysis. Therefore, an attempt has been made in the present study to identify, classify and to characterize *Sorghum Hsf* genes and predict their evolutionary relationship with *Arabidopsis* and *Oryza*. Further, it is also not known where and when these *Hsf* genes are expressed in

*Address correspondence to this author at the Department of Genetics, Osmania University, Hyderabad 500 007, India; Tel: 91-40-2768-2335; Fax: 91-40-2709-5178; E-mail: pbkavi@yahoo.com

 # MN and PSR contributed equally.

1875-5488/15 $58.00+.00

Sorghum. Therefore, in the present investigation, tissue specific expression profiles of these *Hsfs* have been studied by carrying out quantitative real-time PCR under different abiotic stress treatments (by with-holding water for for 5-days for drought, by keeping at 4^0C for 4 h for cold, by exposing to 40^0C in a growth chamber for 4 h, by saturating potted plants with 150 mM NaCl and by collecting the tissue samples after 4 h treatment and by spraying 100 μM ABA and incubating the plants for 4 h for tissue collection). These results will be useful not only for studying the structure and function of *SbHsfs* but also for enhancing abiotic stress tolerance in this crop plant.

2. MATERIALS AND METHODS

2.1. Plant Materials and Stress Treatments

Sorghum bicolor variety cultivar Parbhani Moti, an improved desi variety was used for gene expression related experiments. *Sorghum* plants were grown in earthen pots containing 4.5 kg of black clay soil (Vertisol) under glass house conditions with $28/20^0$C day/night temperatures. Plants were maintained up to 28 days under well watered conditions and then used for different abiotic stress treatments. Drought stress was imposed by with holding the water supply for 5-days followed by leaf sample collection. For cold stress (low temperature) treatment, the plants were kept at 4^0C in a refrigerator for 4 h and was used for tissue collection. For heat stress (high temperature) treatment, plants were kept at 40^0C in a growth chamber and tissues were collected after 4 h of treatment. Salinity stress was induced by saturating the potted plants with 150 mM NaCl solution and leaf samples were collected after 24 h of treatment. For ABA stress, plant leaves were sprayed with 100 μM ABA solution and leaf sample was collected after 4 h. Different tissue samples like seedlings, leaf, flower, mature embryos, and roots were collected from different growth stages of Sorghum plants grown under normal growth conditions. For each sample, tissues were collected from three different plants grown under the same experimental condition ($28/20^0$C day/night temperature), to provide biological replicates. Tissues were snap frozen immediately in liquid nitrogen and stored at -80^0C until RNA extraction.

2.2. Identification and Localization of *Hsfs* in *Sorghum* Genome

Non-redundant nucleotide and amino acid sequences of *Arabidopsis* and rice *Hsfs* [25] were collected from TIGR and NCBI data bases. A total of 47 sequences were collected and each *Hsf* coding sequence (cds) was blasted against *Sorghum bicolor* genome in Gramene database by default settings. Gene sequences from the genome were retrieved using Edit plus (http://www.editplus.com/) and the sequences are subjected to Genscan (http://genes.mit.edu/GENSCAN.html) for coding sequences (cds) and amino acids. The redundant sequences which share the same chromosome location were eliminated and the remaining candidate genes were checked for *Hsf* DBD (DNA binding domain) in the Pfam database by employing SMART program [26], to identify coiled - coil structure and core of HR - A/B region. Sequences without the presence of DBD and coiled - coil regions have been eliminated.

2.3. Multiple Sequence Alignment

ClustalX2 [27] was used for multiple sequence alignment and domain prediction with default parameters. Bioedit (http://bioedit.software.informer.com/7.1/) and Genedoc (Free Software Foundation Inc.) were used for manually editing. For subcellular localization, WoLFPSORT [28], for finding out transmembrane helices TMHMM [29] and for gene characterization GSDS [30, Gene Structure Display Server http://gsds.cbi.pku.edu.cn] were used. NLS and NES were predicted with the help of NLStradamus [31], Nucleo [32], and Net NES [33]. Conserved motif analysis was carried out using MEME [34].

2.4. Promoter Analysis

In silico promoter analysis was carried out using 1 kb sequence upstream to all the *Sorghum Hsfs*. Promoter sequences were retrieved from the genome using Edit plus. PLACE [35] and Plant Care [36] softwares were used to identify the *cis*-acting elements in the promoter sequences. The distribution of *cis*-elements in promoter regions were further identified using MEME tool [34].

2.5. Phylogenetic Tree

Phylogenetic tree was constructed by MEGA 5.1 using the N-J method with 1000 boot strap replicates [37] on the basis of amino acid sequences of *Oryza sativa*, *Arabidopsis*, and *Sorghum*. Gene duplication events were also investigated using phylogenetic tree based on the 70% similarity and 80% coverage of aligned sequences [38, 39].

2.6. RNA Isolation and qRT-PCR

The list of primers used for the qRT-PCR analysis is shown in the supplementary (Table **1**). Total RNA was extracted from control and treated tissues using MACHEREY-NAGEL kit according to the manufacturer's instructions. A total of 2.5 μg RNA was reverse-transcribed into cDNA using SuperScript III First-Strand Synthesis Kit (Invitrogen) for qRT-PCR analysis. The cDNA was diluted into 1:12 with nuclease free water as template for qRT-PCR. The Bioline Master Mix (2X) was used to detect gene expression profile according to the manufacturer's recommendations on the RealPlex (Eppendrof). qRT-PCR was carried out in 96-well optical PCR plates, and the reaction was performed in a total volume of 10 μL containing 0.4 μM of each primer (1.5 μL), cDNA (1.0 μL) and Bioline Master Mix (2X) and nuclease free water was added upto 2.7 μL. qRT-PCR primers were designed using Primer3 software with GC content of 40-60%, Tm $>50^0$C, primer length 20-25 nucleotides, with expected product size of 90-180 bp (Table **1**). The thermal cycles performed were as follows: 95^0C for 10 min, followed by 40 cycles at 95^0C for 15 s and 62^0C for 1 min. Amplicon dissociation curves were recorded after cycle 40 by heating from 58 to 95^0C with fluorescence measured within 20 min. Three technical replicates were used for each gene. Expression levels of the *SbACP2*, *EIF4A*, and *S/T-PP* genes were used as internal controls. The experiments were independently repeated three times, and the data from these experiments were averaged. Relative gene expression calculations were carried using Rest software [40].

Table 1. List of *Sorghum* Hsf proteins. The identified Hsf proteins are listed according to their chromosome location. Hsf proteins are designated according to their locus id, protein sequence (AA) length, annotations, chromosomal locations their molecular weight (Mw), isoelectric point (pi), protein localization, and introns.

S. No.	Gene Name	Locus	AA Length	Annotation	Chromosome Location	Mw (Da)	pI	Localization	Intron
1	*SbHsf01*	Sb03g06630	467	RHsf 7	1	51515.94	5.48	Cytoplasm	2
2	*SbHsf02*	Sb03g12370	371	RHsf 8/Hsf 3	1	42186.32	4.96	Cytoplasm	2
3	*SbHsf03*	Sb03g53340	371	RHsf 4	1	40678.33	4.94	Nucleus	1
4	*SbHsf04*	Sb03g63750	477	-	1	52598.68	4.92	Nucleus	1
5	*SbHsf05*	Sb10g28340	328	RHsf 6	1	37535.13	5.00	Nucleus	1
6	*SbHsf06*	Sb3g02990	383	Putative Hsf sp 17	1	43217.74	5.58	Nucleus	2
7	*SbHsf07*	Sb3g63350	313	-	1	35158.51	7.23	Nucleus	1
8	*SbHsf08*	Sb01g042370	415	RHsf 8/ Hsf 3	1	46456.79	4.91	Nucleus	0
9	*SbHsf09*	Sb03g25120	302	RHsf 12 / Hsf 5	2	33727.63	6.78	Nucleus	1
10	*SbHsf10*	Sb08g36700	334	-	2	34544.70	9.71	Nucleus	0
11	*SbHsf11*	Sb09g28200	482	-	2	51240.71	10.10	Chloroplast	2
12	*SbHsf12*	Sb01g35790	561	-	2	59489.99	7.61	Chloroplast	1
13	*SbHsf13*	Sb02g004370	372	RHsf 5	2	41766.57	4.70	Nucleus	0
14	*SbHsf14*	Sb01g39020	456	Putative Hsf 8	3	49714.46	6.73	Chloroplast	4
15	*SbHsf15*	Sb01g53220	421	RHsf 11/Hsf 8	3	46415.83	9.60	Chloroplast	1
16	*SbHsf16*	Sb01g54550	434	RHsf 9	3	48351.37	5.13	Nucleus	1
17	*SbHsf17*	Sb03g028470	365	RHsf 13/Put. Hsf 1	3	39232.33	6.05	Lysosome	0
18	*SbHsf18*	Sb02g13800	347	-	4	37301.03	9.63	Chloroplast	2
19	*SbHsf19*	Sb02g29340	143	-	4	15257.53	8.07	Chloroplast	2
20	*SbHsf20*	Sb02g32590	176	-	4	19217.10	4.78	Chloroplast	1
21	*SbHsf21*	Sb4g13980	404	Putative Hsf sp 17	4	44957.02	5.34	Nucleus	1
22	*SbHsf22*	Sb04g48030	439	RHsf 1	6	46314.56	5.52	Chloroplast	1
23	*SbHsf23*	Sb09g026440	476	RHsf 10/ Hsf sp 17	9	52621.30	5.05	Nucleus	2
24	*SbHsf24*	Sb06g35960	279	-	10	29070.51	6.98	Cytoplasm	1
25	*SbHsf25*	Sb06g36930	439	-	10	47365.84	4.85	Cytoplasm	0

3. RESULTS

3.1. Identification and Localization of *Hsfs*

Screening of *Sorghum* genome resulted in the identification of 25 *SbHsfs* and are named according to their chromosomal locations (Table **1**). *Hsfs* are distributed on chromosomes 1, 2, 3, 4, 6, 9, and 10 and the number of *Hsfs* varied from chromosome to chromosome. Eight *Hsfs* were identified on chromosome 1, five on chromosome 2, four on 3 and 4, two on chromosome 10 and one on chromosomes 6 and 9 (Fig. **1**). WoLFPSORT was employed to identify subcellular localization of *Hsfs* and 12 of them are located in nucleus, 4 in cytoplasm, 8 in chloroplast and 1 in lysosome (Table **1**). Transmembrane helices were not observed in the *Hsfs* identified.

3.2. Sequence Analysis of *SbHsfs*

The length of the Hsf proteins varied from 143 to 561 amino acids, the molecular weights between 15.25 to 59.48 KDa and the pI from 4.7 to 10.10. Most of the *SbHsf* contain only 1 intron, 4 introns were noticed in *SbHsf14*, but no introns in *SbHsf8*, 10, 13, 17, and 25 (Table **1**). The multiple sequence alignment shows highly conserved DBD domains in *Sorghum bicolor Hsfs* (Fig. **2**). The N terminal DBD of *Hsfs* contains 3α and 4β folds, which is the specific location of HSE. The DBD is approximately 100 amino acids in length, but *SbHsf2*, 9 and 18 contain only 30 residues. HR-A/B domains in *Hsfs* are characterized by coiled - coil structures, which is the key feature containing Leu-Zipper protein interaction domains (Fig. **S1**). SMART program was used to

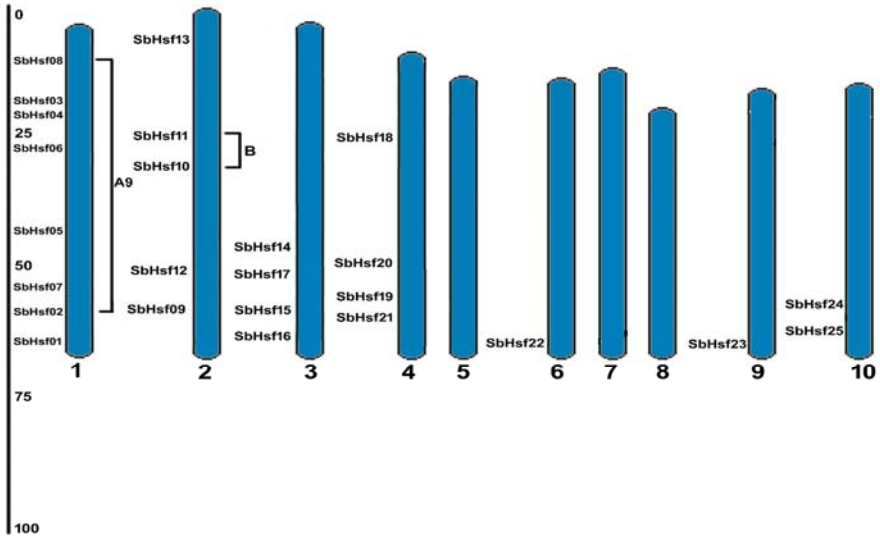

Fig. (1). Locations and duplications of *Sorghum* Hsf paralogs are shown on chromosomes 1-10. The scale represents megabases. The chromosome numbers are indicated at the top of each bar.

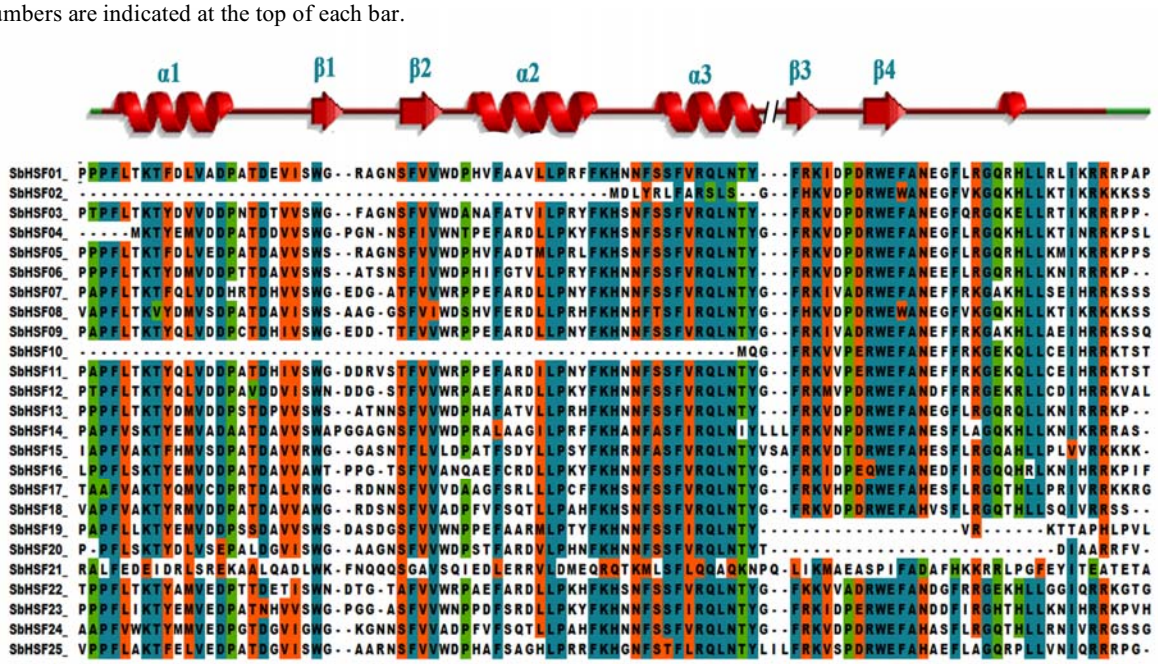

Fig. (2). Multiple sequence alignment of the DBD of the 25 members of *Sorghum Hsf* family is shown. The definition of *Hsf* number corresponds to order of alignment. The results clearly show the highly conserved DBD domains among all *Hsf* genes. The secondary structure elements of DBD (α1-β1-β2-α2-α3-β3-β4) are shown.

predict the DBD characteristic features of HR-A/B regions of *Hsfs* (Table **2**). NLS and NES are important for intracellular distribution of *Hsfs* between the nucleus and cytoplasm and was predicted by using cNLS, NLstradmus, and NET NES 1.1 tools. Most of the *Hsfs* contain two motifs of basic amino acids K/R. Previous comparisons from *Arabidopsis*, *Oryza*, and *Zea mays* show a wide range of NLS monopartite and bipartites found near C terminal of HR A/B regions of *Hsfs*. Only *SbHsfs* 2, 9, and 16 contain bipartite NLS (Table **2**). MEME tool was employed to explore motif distribution both in gene and promoter sequences. It supports the phylogenetic analysis and helps to determine conserved motifs which are species specific, class specific and group specific (Fig. **3**). The *Sorghum Hsfs* contain 30 highly conserved motifs with 5 to 43 residues in length (Fig. **S2**) and the number of motifs vary from 4 to 12. The SbHsf 18, 19, 20, and 24 contain 4 conserved motifs. Out of these, 2 and 3 are DBD, 15, 16, and 19 are coiled coil structure, 21, 24, and 25 are NLS, 12, 13, and 14 are AHA and 23 is NES motif. MEME finds the NLS motifs in SbHsf 11, 12, 17, 18, 22, 23, and 24 which could not be detected by NLS software.

3.3. Promoter Analysis

Table **3** shows the conserved *cis*-acting element motifs present in promoter regions. Motifs 12, 14, and 16 have ABA responsive elements; 16 and 24 have TATA box 2; 16, 26, and 29 have TATA box 3; 21 has LTRE which are low temperature and cold responsive elements, 26 have Myb and

Table 2. Functional domains and motifs of *Sorghum bicolor* Hsfs.

Gene	Group	DBD	NLS	NES
SbHsf01	A2	119-212	327 (ASRKRRRPIG)	384 (LENLALNI)
SbHsf02	A9	3-43	155 (DGNRKRRFQAL)	94 (LLMQQLLV)
SbHsf03	A2	55-148	143 (RTIKRRRPPS)	333 (VELLSLGL)
SbHsf04	A1	1-88	199 (ANKKRRLPKQ)	-
SbHsf05	A2	8-101	207 (ISKKRRRPID)	-
SbHsf06	A2	36-129	232 (ISKKRRRRIV)	165 (LLMTEVVKL)
SbHsf07	B4	44-137	280 (DGKKRRAQQV)	-
SbHsf08	A9	1-83	252 (DGNRKRRFQAL)	191 (LLMQQLVDL)
SbHsf09	B4	3-33	280 (GKKKKRAHQD)	-
SbHsf10	B4	87-181	-	313 (LALEGADLSLTV)
SbHsf11	B4	200-293	-	461 (LALEGADLSLTV)
SbHsf12	B2	107-232	-	90 (FFLVLLLLL)
SbHsf13	A2	155-288	246 (ISKKRRRRID)	-
SbHsf14	A10	10-103	227 (KNIKRRRASK)	-
SbHsf15	C	107-200	382 (PAPGKRRRIG)	366 (VVLRAML)
SbHsf16	A4	23-115	199 (HGKKRRLPIP)	166 (LEDKLIFL)
SbHsf17	C	63-135	-	11 (LHTELALGLL)
SbHsf18	C	2-36	-	-
SbHsf19	A4	140-233	-	113 (LVYDALLVL)
SbHsf20	A3	9-102	-	23 (MLLEPKLEDEDV)
SbHsf21	A5	88-203	137 (FHKKRRLPG)	97 (VSQIEDLERRV)
SbHsf22	B3	47-140	-	422 (LDVLTLSV)
SbHsf23	A4	30-123	-	279 (MELALVSL
SbHsf24	C	49-142	-	179 (MLAFLLKVV)
SbHsf25	A10	24-117	307 (AGRKRRLLD	336 (VLAFEELAL)

Number in brackets indicates the position of the putative localization signal (NLS), nuclear export signal (NES) and DNA Binding Domains (DBD).

28 have Myc waterstress responsive elements (Fig. **4** and Fig. **S3**). The promoter elements like ABRE, ANAERO, ARF, DPBF, DRE, LTRE, MYB, and MYC responsive to ABA, drought, low temperature, and cold are commonly present in all the 25 *Hsfs* along with high temperature responsive elements. The *Hsfs* also contained pathogenesis and salt stress responsive *cis*-elements GT1GMSCAM4 and WBOXNTERF3 for wound response and WBOXANTNPR1 for salicylic acid signal response. The CGCGBOX *cis*-elements present in *Hsfs* are involved in multiple signal transduction and KST1 is involved in guard cell-specific gene expression and pollen specific elements associated with pollen and anther development in different stress conditions. *SbHsfs* 9 and 13 contain a maximum of 15 ABRE *cis*-elements and *SbHsfs*2, 4, and 21 contain a minimum of one ABRE *cis*-elements (Table **4**).

3.4. Phylogenetic Analysis

Phylogenetic tree was constructed by using MEGA 5.1, and neighbour joining method was employed for multiple sequence alignment of 22 *Arabidopsis*, 25 rice, and 25 *Sorghum Hsfs*. Based on the bootstrap values and phylogenetic relationship, they were classified into 3 major *Hsf* classes A, B, and C. Phylogenetic analysis of rice, *Arabidopsis*, and *Sorghum* depicts a close relationship of rice and *Sorghum*, both being members of poaceae. While 10 subgroups are present in class A, 4 are seen in B and the least in C. The contrasting feature of the phylogenetic analysis is in the number of *Hsfs* that varied among the subclass A in rice, *Sorghum* and *Arabidopsis*. For example A2 (five) subgroup is present in the species rice and *Sorghum*, it is absent in *Arabidopsis*. While A6, A7, and A8 subgroups could not be found in monocot species like rice and *Sorghum*, 2, 2, 3

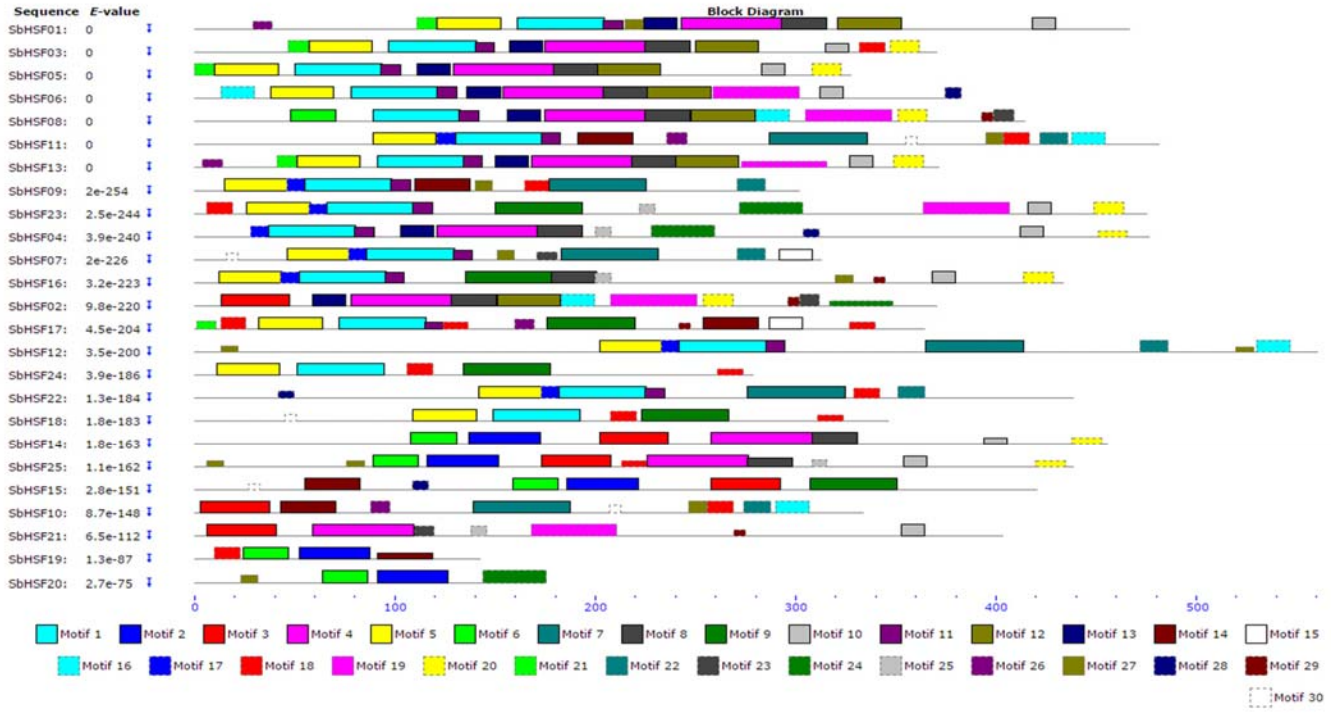

Fig. (3). Distribution of conserved motifs in the Hsf family members is shown. All motifs were identified by MEME using the 25 complete amino acid sequences of *Sorghum Hsf* genes. Names of all the members among the defined gene clusters and combined p values are shown on the left side of the figure, motif sizes are indicated at the bottom. Different motifs represented by different colours are numbered 1-30.

Table 3. Conserved *cis*-acting elements of *Sorghum bicolor* Hsfs. MEME motifs, *cis* elements, signal sequence and their functional roles.

S. No.	Motif	*Cis* Elements	Seq (signal)	Functions
1	12, 14, 16	ABA	ACGTG	Etiolation-induced expression (erd1)
2	17	Anaero 2	AGCAGC	Fermentative pathway
3	2,3,23, 24	ARR	NGATT	Response regulator
4	3, 25	CAAT	CAAT	Promoter of legumin gene
5	27,29,30	CACTT	CACT	Promote phosphoenolpyruvate carboxylase
6	6,17	CGC Box	VCGCGB	Ca++/calmodulin binding
7	1,3,21,24,25,26,27,28,29,30	DOF	AAAG	DNA binding proteins and carbon metabolism
8	2,12,28	DPBF	ACACNNG	ABA and embryo-specification
9	2,26,29	GATA	GATA	Chlorophyll a/b binding
10	6,13	GCC CORE	GCCGCC	G box high level constitution expression
11	9,25,27	GT1	GRWAAW	SA inducible
12	20,23,24,26	GTGA	GTGA	Late pollen gene g10, pectate lyase
13	5	HEXA	CCGTCG	Histone H4
14	3	I BOX CORE	GATAAG	Light regulated
15	21	LTRE	CCGACA	Low temperature and Cold
16	26	MYB	CNGTTR	Water stress
17	28	MYCONSES	CANNTG	erd1 (etiolation responsive to dehydration)
18	5,6	PAL BOX	CCGTCC	Phenylalanine ammonia-lyase
19	8,16,23	POLASI GI	AATAAA	Poly adenylation

(Table 3) contd….

S. No.	Motif	*Cis* Elements	Seq (signal)	Functions
20	4,20	POLASI G2	AATTAAA	Poly adenylation rice amylase
21	16,23,29	POLASI G3	AATAAT	Poly adenylation
22	25,27,28,29	POLLEN	AGAAA	Endo beta mannose, anther and pollen Development
23	21	PRE CONSES	SCGAYNRNN	Plastid responsive and light
24	3,16,29	ROOT MOTIF	ATATT	Promotes rol D
25	2	RYREPEATLE	CATGCAT	GLYCININE, ABA res., embryogenesis
26	11,13,14	SORLIP1	GCCAC	Phytochrome A, root development
27	5,12,24	SORLIP2AT	GGGCC	Light inducible
28	15	SORLREPSAT	TGTATATAT	Phytochrome A
29	2	SPH CORE	TCCATGCAT	Viviporous 1, seed specific development
30	1,12,14,20,23	SURE	GAGAC	Sulfur transporter
31	27	TAAAGSTKSTK1	TAAAG	Controlling guard cells and K+ influx
32	16,29	TATA2	TATAAAT	Accurate initiation for phaseolin
33	16,26,29	TATA 3	TATTAAT	Accurate initiation
34	8,16	TATA 4	TATATAA	Accurate inhibition G
35	8,16,23,26,29	TATA 5	TTATTT	lutamine synthase (non photo syn) ?
36	30	WBOXATNPR1	TTGAC	Response to SA signal
37	2, 11, 30	WBOXNTERF3	TGACY	Response to wound signal
38	2, 12, 30	WRKY	TGAC	Repressor for gibberellin signaling

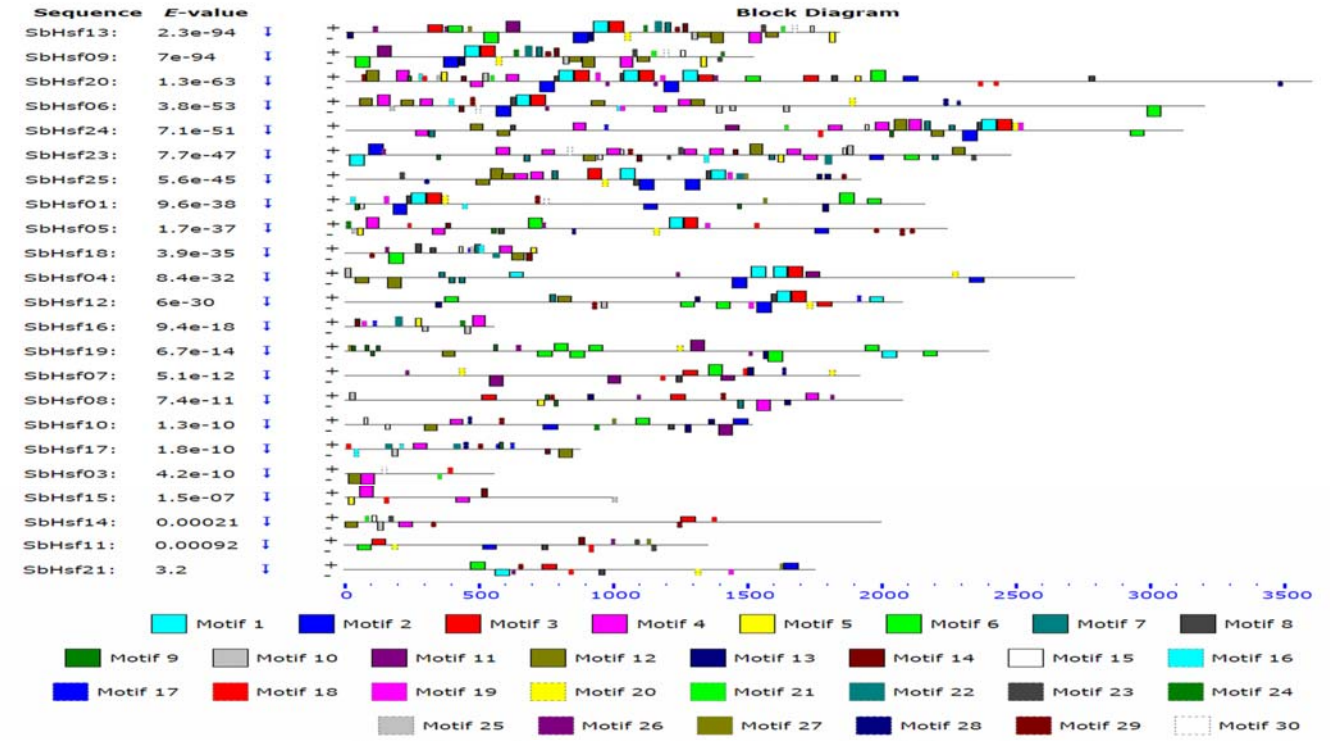

Fig. (4). Distribution of conserved motifs in promoter regions of *Hsf* family is shown. All motifs were identified by MEME using the promoter sequences of *Sorghum* Hsf family analyzed by PLACE and PlantCARE software. Different motifs are indicated by different colours and numbered 1-30, which represent the conserved *cis*-acting elements. For details of motifs refer to table **3**.

Table 4. Conserved *cis*-acting elements present in promoter of *Sorghum* Hsfs.

Gene	Cis Acting Elements														
	ABRECTAL (MACGYGB)	AN-AERO (AAACAAA)	ARF (TGTCTC)	CGCGBOXAT (VCGCGB)	CURE (GTAC)	DPBF (ACACNNG)	DRE (RCCGAC)	GT1GMSCAM4 (GAAAAA)	LTRE (CCGAAA)	MYB (WAACCA/YAACKG/CNGTTR)	MYC (CATGTG/CANNTG)	POL-LEN1LELAT52 (AGAAA)	TAAAGSTKST1 (TAAAG)	WBOXN-T ERF3 (TGACY)	WBOX-AT NPR1 (TTGAC)
SbHsf01	5	1	2	6	10	8	1	4	2	15	38	9	3	7	6
SbHsf02	1	2	1	0	8	0	0	0	2	16	8	4	3	5	3
SbHsf03	3	0	0	2	6	1	1	1	2	4	4	2	0	0	0
SbHsf04	1	2	1	10	20	6	4	3	4	23	24	8	7	8	3
SbHsf05	8	0	3	4	16	5	3	3	4	27	40	6	3	7	6
SbHsf06	9	9	0	26	10	4	2	4	7	30	38	9	10	5	4
SbHsf07	4	5	1	0	22	3	1	4	0	18	22	4	7	13	6
SbHsf08	4	7	0	6	6	3	2	3	5	16	24	4	5	3	6
SbHsf09	15	2	0	24	6	7	5	4	7	18	16	9	3	4	3
SbHsf10	3	3	2	0	12	4	0	1	0	7	14	7	6	3	0
SbHsf11	9	2	0	0	8	2	1	3	0	9	18	11	3	3	2
SbHsf12	7	4	2	4	14	1	0	5	0	23	10	5	1	3	2
SbHsf13	15	6	0	24	6	5	4	5	7	21	18	10	6	7	2
SbHsf14	3	3	2	4	12	3	0	0	1	29	22	11	4	7	1
SbHsf15	9	2	1	12	4	2	0	3	0	12	8	3	5	2	1
SbHsf16	2	1	0	10	0	0	0	0	0	10	0	2	0	2	2
SbHsf17	6	3	0	12	0	4	0	1	1	3	12	1	1	2	1
SbHsf18	12	3	0	14	4	1	3	1	3	10	4	0	0	3	1
SbHsf19	6	0	0	0	14	2	2	7	3	30	8	5	5	12	16
SbHsf20	2	6	1	10	8	4	4	10	6	25	36	16	6	9	7
SbHsf21	1	2	0	0	4	3	2	7	2	22	12	12	2	4	8
SbHsf22	4	0	0	4	8	3	0	5	4	14	10	12	2	6	2
SbHsf23	9	6	1	44	6	4	12	6	12	23	16	7	1	8	3
SbHsf24	8	4	3	42	12	4	3	5	7	35	30	13	3	9	6
SbHsf25	6	6	1	8	6	2	1	1	2	15	14	7	5	2	2

ABRECTAL: Response to ABA, ANAERO: Anaerobic conditions, ARF: ABA and auxin responsive, CGCGBOX: Multiple signal transduction, CURE: Cu and oxygen responsive, , DPBF: ABA, DRE: Dehydration responsive elements, GT1GMSAM4: Salt and pathogenesis related, LTRE: Low temperature and cold responsive, MYB: responsive to drought and ABA, MYC: Response to drought, cold and ABA, POLLEN: pollen and anther development, TKST1: Guard cell-specific gene expression, WBOXNTERF3: Wound signal and WBOXATNPR1: Salicylic acid responsive.

have been detected respectively in *Arabidopsis* (Figs. **5** and **6**). Among the four subclasses of B, B1 are absent in Sorghum, but one is detected in rice. Further, in class C, the genome of *Arabidosis* revealed only one *Hsf*, but four each could be identified in rice and *Sorghum*.

3.5. Gene Duplication Events

Two paralogs participated out of 25 *Sorghum Hsfs* in regional duplications within the chromosomes. These paralogs evolved from their common ancestral genes through gene duplication events. While no segmental duplication events were observed in *Sorghum* 8 and 7 were recorded in maize and rice respectively out of nine paralogs. Maize and rice Hsf family is expanding with large number of segmental duplications (Fig. **5**).

3.6. Transcript Profiling of *SbHsfs* in Different Tissues

SbHsf genes displayed differential expression in different tissues (Fig. **7a**). Out of four major tissues (mature embryo, seedling, root, and panicle), panicle showed higher levels of *Hsf* abundance than the mature embryos. No *Hsfs* were up- or down-regulated in the case of mature embryos (Fig. **7a**). While in seedling *Hsfs*4, 9 are highly expressed, 13 and 22 are moderately expressed. In the case of roots, only 4 and 13 are well expressed. Moderate expression levels were also recorded in *Hsfs*5, 6, 21, 23, and 25 in roots. On the other-hand, *Hsfs*4 and 22 are highly expressed, *Hsf1*, 3, 5, 9, 10, 13, 16, 19, 23, and 25 recorded moderate transcript levels in panicle tissues (Fig. **7a**).

3.7. Abiotic Stress Induced Expression of *Hsfs*

All *Hsfs* displayed a differential expression in response to various abiotic stresses (Fig. **7b**). Among the five treatments (ABA, cold, heat, salt, and drought), drought stress induced higher transcript abundance than the other treatments. ABA, did not enhance the levels of *Hsfs* except in Hsf23, where only minor increase was noticed. Expression was significantly upregulated in HSf1, 15, 19, and 25 under cold stress (Fig. **7b**). Moderate levels of expression was observed in *Hsfs* 2, 3, 4, 5, 6, 8, 10, 13, 16, 21, 23, and 24. During heat stress, Hsf1 was highly expressed, and moderate expression were displayed in *Hsfs* 6, 9, 13, 21 etc. During salt stress, *Hsfs*4, 6, 13, 16, 21, and 23 were up-regulated. In contrast, many *Hsfs* like 1, 5, 6, 10, 13, 18, 22, 23, and 25 were upregulated during drought stress (Fig. **7b**).

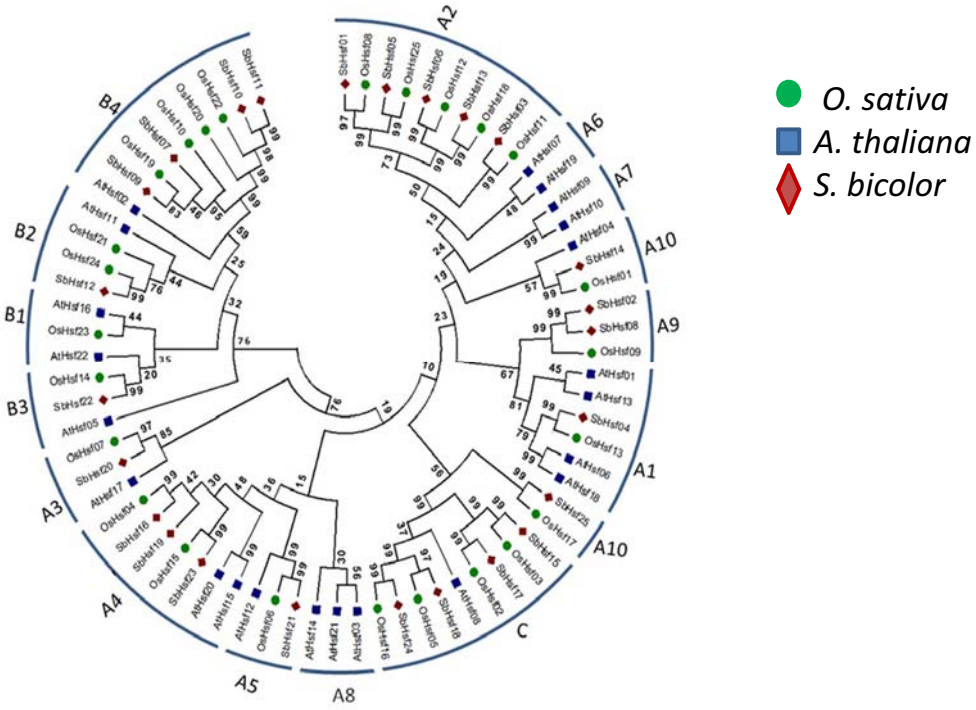

Fig. (5). The *Hsf* phylogenetic tree is constructed using neighbour joining method. The phylogenetic tree constructed with MEGA 5.1, has been generated on the basis of amino acid sequences of *Oryza sativa*, *Arabidopsis thaliana* and *Sorghum bicolor*. The Hsf proteins are classified into 3 major groups A, B and C, in which group A is subdivided into 10 groups, A1 to A10, and B is subdivided into 4 groups, B1 to B4. The abbreviations: Os = *Oryza sativa*, At = *Arabidopsis thaliana*, Sb = *Sorghum bicolor*.

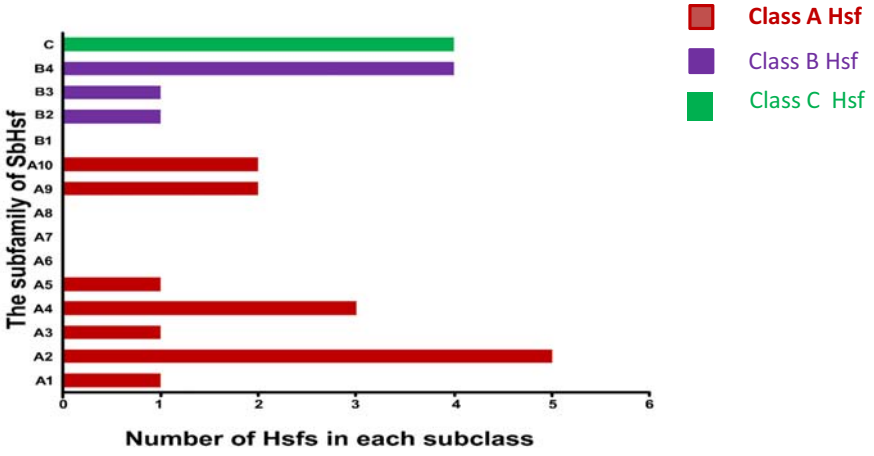

Fig. (6). The number of *SbHsf* subgroups in three classes are shown.

4. DISCUSSION

4.1. Sequence Analysis

Hsfs have been identified in several plants [21-24, 41-43] but not in *S. bicolor* which is often exposed to salt, drought, and temperature stresses. Genetic variability for drought tolerance exists in *Sorghum* [44] but the effects of high temperature and water stresses on reproductive biology and seed-set needs further investigations and identification of candidate genes for breeding programs aimed at crop improvement. While eight *Hsfs* are distributed on chromosome 1, no *Hsfs* could be detected on 5, 7 and 8. In the case of *Arabidopsis*, maize and rice, *Hsfs* are spread all over the chromosomes but chromosomes 11 and 12 lack them [47, 25]. Like rice and maize, *S. bicolor* has also the same number of *Hsfs*, which reflects that *Hsfs* are conserved during the process of evolution [47, 25]. The theoretical pI of *Hsfs* range between 4.7 to 10.10, which indicates that they contain both acidic and basic proteins. *Hsfs* 2, 9 and 18 contain 30 residue-length DBD, which may occur due to deletions in DBD regions of α and 4 β-helices and due to genetic diversity in *SbHsfs*. Class A requires AHA motifs for their functioning, but *SbHsf14* and 20 lack such motifs. *SbHsfs*18 and 24 belong to class C but do not contain AHA motifs. They

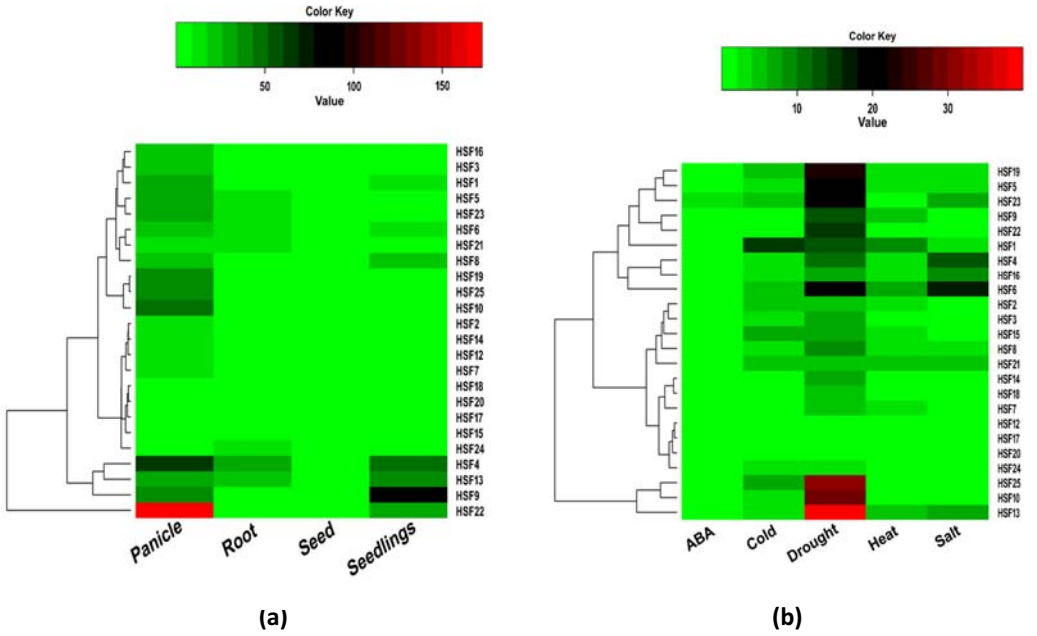

Fig. (7). (**a**). Relative expression of *SbHsfs* at the transcript level is shown in different tissues. Relative expression of *SbHsf* transcripts are shown during different abiotic stress conditions in comparison to its control as revealed by quantitative RT-PCR analysis. Values represent the expression values obtained after normalizing against control value. All samples were analyzed in triplicates, in three independent experiments. Names on the horizontal axis indicate the identified *SbHsfs*, and the vertical axis represents the various tissues, i.e., mature embryos, panicle, seedlings and root. Each color represents the relative expression levels. (**b**). Relative quantification of *SbHsfs* under diverse abiotic stress treatments is shown. Relative expression of *SbHsf* transcripts is shown during different abiotic stress conditions in comparison to its control as revealed by quantitative RT-PCR analysis. Values represent the expression values obtained after normalizing against control value. All samples were analyzed in triplicates, in three independent experiments. Names on the horizontal axis indicate the identified *SbHsfs*, and the vertical axis represent various treatments such as ABA, cold, heat, salt and drought. Each color represents the relative expression levels.

may bind to other classes of A and C *Hsf* types and form hetero oligomers and start their function [25]. *In silico* survey of the putative *cis*-elements of the *Sorghum Hsfs* showed the presence of HSE, ABA responsive elements, ARR, Anaero, CACTT, low temperature responsive elements (LTRE), pollen specific *cis*-regulatory (AGAAA) and desiccation responsive elements. This indicates that *Hsfs* are not only expressed during high temperature but also during other environmental stresses. The presence of HSE *cis*-elements in the promoter regions is correlated with the expression of *Hsf* genes under high temperature stress in *Arabidopsis*, rice, maize, and wheat [47-50]. Bate and Twell [51] observed that transcriptional activation of late pollen gene (*lat52*) is controlled by a pollen-specific *cis*-regulatory elements AGAAA and TCCACCATA to attain high gene expression levels associated with pollen maturation. Promoter analysis of the endo-β-mannanase gene demonstrated pollen-specific *cis*-acting elements POLLEN1LELAT52 (AGAAA) which are associated with anther and pollen development [52]. In the present study also, such AGAAA elements were detected in the promoter regions of *Hsfs* indicating that these *Hsfs* may be involved in anther and pollen development in *Sorghum*. Promoter analysis of the *KST1* gene, (an inward rectifying potassium channel) revealed a sequence motif TAAAG and the involvement of these elements suggests a role for Dof transcription factors in guard cell-specific gene expression and stomatal conductivity [53]. Such TAAAG elements have been observed in our promoter analysis, raising scope for speculation of Hsf promoters in K^+ influx and guard cell movement. *Hsfs* are not only expressed during abiotic stress, but also biotic stress since their promoter regions contain potential *cis*-elements such as WBOXNTERF3 and WBOXATNPR1 which are responsive to biotic stresses like wound, pathogen, and salicylic acid [54, 55]. While *ERF3* gene is activated by wounding in tobacco [55], the disease resistance regulatory protein NPR1 has been found to be required to activate AtWRKY18 [56]. Detecting ABA and salicylic acid response elements in the promoter regions of *Hsfs* provide valuable clues on the underlying regulatory mechanisms of *Hsfs* that may further lead to development of plants with biotic and abiotic resistance.

4.2. Phylogenetic Analysis

The phylogenetic tree revealed that proportion of the three *Hsf* classes differed considerably among the three species. While class A contained the large number of *Hsfs*, class B accounted for small number, and class C the minimum. *Hsfs* with three distinct classes A, B, and C appeared to be more in number in majority of angiosperms except in *Medicago truncatula* (class C absent), when compared to lower plants that contain classes like A and B as in the case of *Picea abies*, *Selaginella moellendorffii*, *Physcomitrella patens*, *Chlorella* sp. NC64 etc. [21]. Differences in different subgroups of A4, A9, B1 and B2 were observed between rice and a relatively temperature and drought tolerant *S. bicolor*, which is a C_4 plant. Subgroup B1 is absent in *Sorghum* while it is present in rice. Perhaps these differences in different

subclasses of A and B play critical roles during various types of abiotic stresses and developmental activities in these two contrasting plants. However, such an assumption needs to be validated experimentally. In plants, gene duplication events play an important role in evolution [57]. In polyploidization, gene duplicates accumulate [58] and these processes involve several transcription factors [59]. Recently, Song *et al.* [24] observed duplication events in the expansion of *Hsf* genes in Chinese cabbage. These observations clearly indicate that Hsf transcription factor family contributed to polyploidy [24, 59]. In the present study, segmental gene duplication events could not be observed in *Sorghum* unlike that of maize and rice [47, 25].

4.3. Transcript Analysis in Different Tissues and During Different Abiotic Stress Conditions

The expression patterns of different *Hsf* genes may differ depending on the plant species [21]. Yamaguchi-Shinozaki and Shinozaki [60] have shown that transcription factors interact with each other. It appears that each of the *Hsf* genes respond differentially to different abiotic stresses and developmental stages. Several transcriptome studies show that *Hsf* transcription plays significant roles in response to abiotic stress [23, 24, 48]. This type of unique expression patterns of *Hsf* transcripts were observed in response to both abiotic stresses and developmental stages also [9, 21, 43]. The varied patterns of *Hsf* expressions in different tissues may relate to the differences in *cis*-acting elements present in different promoter sequences. In the present study, *Hsfs* that are expressed during one type of abiotic stress, did not up-regulate when exposed to the other type of stresses, the exception being *Hsf1* for cold and drought, and *Hsf6* for salt and drought. Cross-talk exists between abiotic stress signal and plant growth and the expression of different transcription factor gene families [24, 41, 45, 60] indicating that these *Hsfs* play critical roles in maintaining drought and temperature stress tolerance and also play a vital role during development [21, 41, 46].

Six out of 21 *Hsfs* in Arabidopsis and 8 and 9 out of 25 in *Oryza* and *Sorghum* were induced by heat stress respectively [61, 62]. In many plants, intron-mediated enhancement (IME) of gene expression was noticed as in the case of *Alcohol dehydrogenase 1*, and *Bronze 1* as reported by Callis *et al.* [63], *Shrunken 1* in maize [64] and *Phosphoribosylanthranilate transferase 1* in *Arabidopsis* [65]. Introns increased the transcription initiation and mRNA levels in these cases [66]. While in rice, intron mediated enhanced gene expression was observed, in *Sorghum*, exceptions were noticed in *SbHsf08*, 10, 13 and 25. These *Hsfs* in *S. bicolor* showed elevated expression levels without any intron. Intriguingly, *SbHsf14* contains 4 introns but displayed lower expressions during stress. This infers that IME gene expression may vary depending upon the *Hsf* present in a specific species. *OsHsfA2d*, which is duplicated with *OsHsfA2c*, has two introns in place of one in the original gene *A2c* and *OsHsfB2b*/*OsHsfB2c*. This *OsHsfB2b*/*OsHsfB2c* has 2 introns and exhibited more expression during heat stress and considerably higher expression in almost all the other abiotic stresses and during seed development [62]. In *S. bicolor*, regional duplicated gene pair *SbHsf02*/*SbHsf08* has no introns instead of 2 in the original gene *SbHsf02*, *Hsf08* expressed abundantly in all the tissues and during all stress treatments. On the other hand, *SbHsf10*/*SbHsf11* has 2 introns, but not expressed during all stresses.

Class A HSFs have been characterized in more detail than class B and C HSFs in plants. In *Arabidopsis*, expression of *HsfA2* was high among the class A *HSFs* under high temperature and light stresses [8]. In rice, the expression of all *OsHsfA2* genes increased by heat stress except for *A2b*, which is actually a duplicated gene with *A2e* [62]. In *Sorghum*, 5 members of *HsfA2* genes have been noticed in contrast to 6 in rice, and are also highly induced during drought, salt, heat, and cold stresses. *HsfB1* is absent in *Sorghum* and *Oryza* but present in *Arabidopsis*. Though *HsfB1* is heat inducible, its overexpression did not lead to thermotolerance in *Arabidopsis* [14, 61]. On the other hand, in tomato, *HsfB1* is a transcription co-activator functioning along with *HsfA1* and hypothesized as a heat shock induced factor essential for maintenance and restoration of house keeping gene transcription during stress [67]. *OsHsfB2a*, *B2b* and *B2c* were induced by heat stress but expressed in developing seeds. In *Sorghum*, *Hsf B2* was not induced under any stress but observed in panicles. Double knock-out mutants for *AtHsfB1* and *B2* displayed normal fertility and thermotolerance as compared with single knock-out mutants in *Arabidopsis* [17]. In *S. bicolor*, *B3* was highly expressed in panicle and early seedling stage during droght but not in rice. On the other hand, *B4* and class C *Hsfs* are moderately induced under all stress conditions. Thus, several differences exist among different classes of *Hsfs* between water loving rice and relatively drought tolerant *S. bicolor*.

In conclusion, 25 *SbHsfs* genes were identified in the genome of *S. bicolor*. Such a systematic analysis of *Hsfs* help us in finding out the functions of *Hsf* signaling pathways associated with different abiotic stress conditions and also growth and development. The diverse expression patterns suggest that these genes may perform different physiological functions depending on the type of tissue and its needs. Some *SbHsfs* were constitutively expressed, while others exhibited a distinct expression pattern in different tissues and under diverse abiotic stress treatments, implying that *SbHsfs* genes have functional diversity. This study provides the first step towards the future studies of Hsf protein functions and enhancing drought or thermotolerance stress and also the association of *SbHsf* genes under diverse environmental conditions.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

MN is thankful to the UGC, New Delhi, for providing fellowship. PSR is thankful to the Department of Science and Technology, Govt. of India, New Delhi, for research funding through the INSPIRE Faculty Award Grant No. IFA-11LSPA-06. PBK is thankful to the CSIR, New Delhi, for providing Emeritus Fellowship. DMR is thankful to the UGC, New Delhi, for financial support through UGC-MRP.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Wang, W.; Vinocur Altman, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, **2003**, *218*(1), 1-14.

[2] Feder, M.E.; Hofmann, G.E. Heat-shock proteins, molecular chaperones, and stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol*., **1999**, *61*, 243-282.

[3] Scharf, K.D.; Heider, H.; Hohfeld, I.; Lyck, R.; Schmidt, E.; Nover, L. The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules. *Mol. Cell Biol*., **1998a**, *18*(4), 2240-2251.

[4] Bharti, K.; Schmidt, E.; Lyck, R.; Bublak, D.; Scharf, K.D. Isolation and characterization of HsfA3 a new heat stress transcription factor of *Lycopersicon peruvianum. Plant J.*, **2000**, *22*(4), 355-365.

[5] Mishra, S.K.; Tripp, J.; Winkelhaus, S.; Tschiersch, B.; Theres, K.; Nover, L.; Scharf, K.D. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermo tolerance in tomato. *Genes Dev*., **2002**, *16*(12), 1555-1567.

[6] Charng, Y.Y.; Liu, H.C.; Liu, N.Y.; Chi, W.T.; Wang, C.N.; Chang, S.H.; Wang, T.T. A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermo tolerance in *Arabidopsis. Plant Physiol*., **2007**, *143*(1), 251-262.

[7] Ogawa, D.; Kazuo, Y.; Takumi, N. High-level over expression of the *Arabidopsis* HsfA2 gene confers not only increased themo tolerance but also salt/osmotic stress tolerance and enhanced callus growth. *J. Exp. Bot*., **2007**, *58*(12), 3373-3383.

[8] Nishizawa, A.; Yabuta, Y.; Yoshida, E.; Maruta, T.; Yoshimura, K.; Shigeoka, S. *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant J*., **2006**, *48*(4), 535-547.

[9] Jung, H.S.; Crisp, P.A.; Estavillo, G.M.; Cole, B.; Hong, F.; Mockler, T.C.; Pogson, B.J.; Chory, J. Subset of heat- shock transcription factors required for the early response of *Arabidopsis* to excess light. *Pr. Nat. Acad. Sci.*, **2013**, *110*(35), 14474-14479.

[10] Swindell, W.R.; Huebner, M.; Weber, A.P. Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*, **2007**, *8*, 125.

[11] Yokotani, N.; Ichikawa, T.; Kondou, Y.; Matsui, M.; Hirochika, H.; Iwabuchi, M.; Oda, K. Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic Arabidopsis. *Planta*, **2008**, *227*(5), 957-967.

[12] Yoshida, T.; Sakuma, Y.; Todaka, D.; Maruyama, K.; Qin, F.; Mizoi, J. Functional analysis of an Arabidopsis heat-shock transcription factor HsfA3 in the transcriptional cascade downstream of the DREB2A stress- regulatory system. *Biochem. Biophys. Res. Commun*., **2008**, *368*(3), 515-521.

[13] Li, M.Y.; Wang, F.; Jiang, Q.; Li, R.; Ma, J.; Xiong, A.S. Genome-wide analysis of the distribution of AP2/ERF transcription factors reveals duplication and elucidates their potential function in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *Plant Molecular Biology Reports*, **2013**, *31*(4), 1002-1011.

[14] Prandl, R.; Hinderhofer, K.; Eggers-Schumacher, G.; Schoffl, F. HSF3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermo tolerance when over expressed in transgenic plants. *Mol. Gen. Genet*., **1998**, *258*(3), 269-278.

[15] Nover, L.; Bharti, K.; Doring, P.; Mishra, S.K.; Ganguli, A.; Scharf, K.D. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones*, **2001**, *6*(3), 177-189.

[16] Shim, D.; Hwang, J.U.; Lee, J.; Lee, S.; Choi, Y.; An, G.; Martinoia, E.; Lee, Y. Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. *Plant Cell*, **2009**, *21*(12), 4031-4043.

[17] Kumar, M.; Busch, W.; Birke, H.; Kemmerling, B.; Nurnberger, T.; Schoffl, F. Heat shock factors HsfB1 and HsfB2b are involved in the regulation of Pdf1.2 expression of Pdf1.2 expression and pathogen resistance in *Arabidopsis. Molecular Plant*, **2009**, *2*(1), 152-165.

[18] Kotak, S.; Vierling, E.; Baumlein, H.; Von Koskull-Doring, P. A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *The Plant Cell*, **2007**, *19*, 182-195.

[19] Prieto-Dapena, P.; Castano, R.; Almoguera, C.; Jordano, J. The ectopic overexpression of a seed-specific transcription factor, HaHSFA9, confers tolerance to severe dehydration in vegetative organs. *The Plant J.*, **2008**, *54*(6), 1004-1014.

[20] Pirkkala, L.; Nykanen, P.; Sistonen, L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J.*, **2001**, *15*(7), 1118-1131.

[21] Huang, Y.; Li, M.Y.; Wang, F.; Xu, Z.S.; Huang, W.; Wang, G.L.; Ma, J.; Xiong, A.S. Heat shock factors in carrot: genome-wide identification, classification and expression profiles response to abiotic stress. *Mol. Biol. Reports*, **2014**, DOI 10.1007/ s11033-014-3826-x.

[22] Wang, F.; Dong, Q.; Jiang, H.; Zhu, S.; Chen, B.; Xiang, Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula. Mol. Biol. Reports,* **2012***, 39*(2), 1877-1886.

[23] Chung, E.; Kim, K.M.; Lee, J.H. Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max. J. Genet. Genomics*, **2013**, *40*(3), 127-135.

[24] Song, X.; Liu, G.; Duan, W.; Liu, T.; Huang, Z.; Ren, J.; Li, Y.; Hou, X. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Mol. Genet. Genomics*, **2014**, *289*(4), 541-551.

[25] Guo, J.; Wu, J.; Ji, Q.; Wang, C.; Luo, L.; Yuan, Y.; Wang, Y.; Wang, J. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *J. Genet. Genomics*, **2008**, *35*(2), 105-118.

[26] Letunic, I.; Copley, R.R.; Schmidt, S.; Ciccarelli, F.D.; Doerks, T.; Schultz, J.; Ponting, C.P.; Bork, P. SMART 4.0: towards genomic data integration. *Nucleic Acids Res*., **2004**, *32D*, 142-144.

[27] Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*, **2007**, *23*(21), 2947-2948.

[28] Horton, P.; Keun-Joon Park.; Takeshi Obayashi.; Naoya Fujita.; Hajime Harada, C.J.; Adams-Collier,; Kenta Nekai. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*., **2007**, *35*, W585-W587.

[29] Moller, S.; Croning, M.D.R.; Apweller, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **2001**, *17*(7), 646-653.

[30] Guo, A.Y.; Zhu, Q.H.; Chen, X.; Luo, J.C.; Chuan, Y.I. GSDS a gene structure display server. *Yi Chuan*, **2007**, *29*(8), 1023-1026.

[31] Nguyen Ba, A.N.; Pogoutse, A.; Provart, N.; Moses, A.M. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, **2009**, *10*, 202.

[32] Hawkins, J.; Davis, L.; Boden, M. Predicting nuclear localization. *J. Proteome Res.*, **2007**, *6*, 1402-1409.

[33] la Cour, T.; Klemer, L.; Molgaard, A.; Gupta, R.; Skriver, K.; Brunak, S. Analysis and prediction of leucine- rich nuclear export signals. *Protein Eng. Des. Sel.*, **2004**, *17*(6), 527-536.

[34] Bailey, T.L.; Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **1998**, *14*(1), 48-54.

[35] Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.*, **1999**, *27*(1), 297-300.

[36] Lescot, M.P.; Dehais, G.; Thijs, K.; Marchal, Y.; Moreau, Y.; Van de Peer, P.; Rombauts, R.S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*., **2002**, *30*(1), 325-327.

[37] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA 5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony Methods. *Mol. Biol. Evol.*, **2011**, *28*(10), 2731-2739.

[38] Gu, Z.; Cavalcanti, A.; Chen, F.C.; Bouman, P.; Li, W.H. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.*, **2002**, *19*(3), 256-262.

[39] Yang, S.; Zhang, X.; Yue, J.X.; Tian, D.; Chen, J.Q. Recent duplications dominate NBS-encoding gene expansion in two woody

species. *Mol. Genet. Genomics*, **2008**, *280*(3), 187-198.

[40] Pfaffl, W.M.; Horgan1, W.G.; Dempfle Leo. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, **2002**, *30*(9), 900.

[41] Reddy, P.S.; Kishor, P.B.K; Seiler, C.; Kuhlmann, M.; Eschen-Lippold, L.; Lee, J.; Reddy, M.K.; Sreenivasulu, N. 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*, **2014**, *9*(3), e89125.

[42] Hahn, A.; Bublak, D.; Schleiff, E.; Scharf, K.D. Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *The Plant Cell*, **2011**, *23*(2), 741–755.

[43] Giorno, F.; Guerriero, G.; Baric, S.; Mariani, C. Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. *BMC Genomics*, **2012**, *13*, 639.

[44] Ali, M.A.; Niaz, S.; Abbas, A.; Sabir, W.; Jabran, K. Genetic diversity and assessment of drought tolerant *Sorghum* landraces based on morpho-physiological traits at different growth stages. *Plant Omics J.*, **2009**, *2*(5), 214-227.

[45] Durbin, R.S.E.; Anders, K.; Graeme, M. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. *Cambridge University Press*, **1998**.

[46] McGinnis, S.; Madden, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*., **2004**, *32W*, 20−25.

[47] Lin, Y.X.; HaiYang Jiang.; Zhang-Xin Chu,; Xiu-Li Tang,; Su-Wen Zhu,; Bei-Jiu Cheng. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genomics*, **2011**, *12*, 76.

[48] Mittal, D.; Chakrabarti, S.; Sarkar, A.; Singh, A.; Grover, A. Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol. Biochem.*, **2009**, *47*(9), 785-795.

[49] Sarkar, N.K.; Kim, Y.K.; Grover, A. Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics*, **2009**, *10*, 393.

[50] Chauhan, H.; Khurana, N.; Nijhavan, A.; Khurana, J.P.; Khurana, P. The wheat chloroplastic small heat shock protein (*sHSP26*) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant Cell Environ.*, **2012**, *35*(11), 1912-1931.

[51] Bate, N.; Twell, D. Functional architecture of a late pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Mol. Biol.*, **1998**, *37*(5), 859-869.

[52] Flicchkin, S.A.; Leonard, J.M.; Monteros, A.; Liu, P.P.; Nonogaki, H. A novel endo-beta mannanase gene in tomato *LeMAN5* is associated with anther and pollen development. *Plant Physiol.*, **2004**, *134*(3), 1080-1087.

[53] Plesch, G.; Ehrhardt, T.; Mueller-Roeber, B. Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. *Plant J.*, **2001**, *28*(4), 455-64.

[54] Chen, W.; Provart, N.J.; Glazebrook, J.; Katagiri, F. Chang, H.S.; Eulgem, T.; Mauch, F.; Luan, S.; Zou, G.; Whitham, S.A.; Budworth, P.R.; Tao, Y.; Xie. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell*, **2002**, *14*(3), 559-574.

[55] Nishiuchi, T.; Shinshi, H.; Suzuki, K. Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. *J. Biol. Chem.*, **2004**, *279*(53), 55355-55361.

[56] Chen, C.; Chen, Z. Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen- induced *Arabidopsis* transcription factor. *Plant Physiol.*, **2002**, *129*(2), 706-716.

[57] Wendel, J.F. Genome evolution in polyploids. *Plant Mol. Biol*., **2000**, *42*, 225-249.

[58] Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, P.; Hu, Y.; Liang, H.; Soltis, P.S.; Soltis, D.E.; Clifton, S.W.; Schlarbaum, S.E.; Schuster, S.C.; Ma, H.; Leebens-Mack, J.; dePamphilis, C.W. Ancestral polyploidy in seed plants and angiosperms. *Nature*, **2011**, *473*(7345), 97-100.

[59] Ma, J.; Wang, F.; Li, M.Y.; Jiang, Q.; Tan, G.F.; Tan, G.F.; Xiong, A.S. Genome-wide analysis of the NAC transcription factor family in Chinese cabbage to elucidate responses to temperature stress. *Scientia Horticultrae*, **2014**, *165*, 82-90.

[60] Yamaguchi-Shinozaki, K.; Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.*, **2006**, *57*, 781-803.

[61] Busch, W,; Wunderlich, M,; Schoffl, F. Identification of novel heat shock factor-dependent genes and Bio chemical pathways in *Arabidopsis Thaliana. Plant J*., **2005**, *41*(1), 1-14.

[62] Chauhan, H,; Khurana, N,; Agarwal, P,; Khurana, P. Heat shock factors in rice (*Oryza sativa* L.): genome wide expression analysis during reproductive development and abiotic stress. *Mol. Genet. Genomics*., **2011**, *286*(2), 171-187.

[63] Callis, J,; Fromm, M,; Walbot, V. Introns increase gene expression in cultured maize cells. *Genes Dev*., **1987**, *1*(10), 1183–1200.

[64] Clancy, M,; Hannah, L.C. Splicing of maize Sh1first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing. *Plant Physiol.*, **2002**, *130*(2), 918–929.

[65] Rose, A.B.; Last, R.L. Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1. *Plant J*., **1997**, *11*(3), 455–464.

[66] LeHir, H,; Nott, A,; Moore, M.J. How introns influence and enhance gene expression? *Trends Biochem. Sci*., **2003**, *28*(4), 215-220.

[67] Bharti, K,; Von Koskull-Doring, P,; Bharti, S,; Kumar, P,; Tintschl-Korbitzer, A,; Treuter, E,; Nover, L. Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription co activator with a histone-like motif interacting with the plant CREB binding protein ortholog HAC1. *Plant Cell*., **2004**, *16*(6), 1521-1535.

Received on: January 28, 2015 **Revised on: March 07, 2015** **Accepted on: March 12, 2015**