International Journal of Biological Macromolecules 142 (2020) 822–834

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Genome-wide identification and transcriptional profiling of small heat shock protein gene family under diverse abiotic stress conditions in *Sorghum bicolor* (L.)

M. Nagaraju [a,b], Palakolanu Sudhakar Reddy [c], S. Anil Kumar [d], Anuj Kumar [e], G. Rajasheker [a], D. Manohar Rao [a,*], P.B. Kavi Kishor [a,*]

[a] Department of Genetics, Osmania University, Hyderabad 500 007, India
[b] Biochemistry Division, ICMR-National Institute of Nutrition, Hyderabad 500 007, India
[c] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad 502 324, India
[d] Department of Biotechnology, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, Andhra Pradesh 522 213, India
[e] Advance Center for Computational & Applied Biotechnology, Uttarakhand Council for Biotechnology (UCB), Dehradun 248 007, India

ARTICLE INFO

Article history:
Received 15 July 2019
Received in revised form 16 August 2019
Accepted 2 October 2019
Available online 14 October 2019

Keywords:
Chloroplast
Microarray
qRT-PCR

ABSTRACT

The small heat shock proteins (sHsps/Hsp20s) are the molecular chaperones that maintain proper folding, trafficking and disaggregation of proteins under diverse abiotic stress conditions. In the present investigation, a genome-wide scan revealed the presence of a total of 47 sHsps in *Sorghum bicolor* (*SbsHsps*), distributed across 10 subfamilies, the major subfamily being P (plastid) group with 17 genes. Chromosomes 1 and 3 appear as the hot spot regions for *SbsHsps*, and majority of them were found acidic, hydrophilic, unstable and intron less. Interestingly, promoter analysis indicated that they are associated with both biotic and abiotic stresses, as well as plant development. *Sorghum sHsps* exhibited 15 paralogous and 20 orthologous duplications. Expression analysis of 15 genes selected from different subfamilies showed high transcript levels in roots and leaves implying that they are likely to participate in the developmental processes. *SbsHsp* genes were highly induced by diverse abiotic stresses inferring their critical role in mediating the environmental stress responses. Gene expression data revealed that *SbsHsp-02* is a candidate gene expressed in all the tissues under varied stress conditions tested. Our results contribute to the understanding of the complexity of *SbsHsp* genes and help to analyse them further for functional validation.

1. Introduction

Plant vigour and crop yields are markedly influenced by abiotic stresses [1] throughout the world. Abiotic stresses pose a challenge for sustainable food production, as they reduce the yield potential by ~70% in crop plants [2]. Environmental stresses cause changes in gene regulation, and biochemical activities leading to considerable reduction in the final productivity [3]. Plants have evolved a wide-spectrum of molecular programs that help them to cope with the changing environments. Mittler et al. [4] pointed out that heat stress response is highly conserved among plants and associates with many pathways and regulatory networks. Different types of abiotic stresses induce small heat shock proteins (sHsps/Hsp-20s) which are implicated in preventing or in repairing the misfolded and truncated proteins due to changing environments and thus confer tolerance as pointed out by Chang et al. [5], and Guo et al. [6]. The 'small', in small heat shock proteins refers to their monomeric size, from 12 to 42 kDa. But, McHaourab et al. [7] showed that many sHsps form large oligomers of 40–50 subunits, compared to some intensively studied chaperones like that of Hsp60 and Hsp70 families. At the same time, the mechanism of these sHsps in protecting the proteins is still poorly understood [8]. This is mainly attributed to their enormous diversity and heterogeneity, not only in their oligomeric structure but also substrate binding, thus, making their study extremely challenging [9].

Based on their cellular localization, sequence homology and function, Vierling [10], Waters et al. [11] divided plant Hsp20s into various subgroups. Evolutionarily, sHsps are a super family of proteins of an ancient period [12], which is reflected by their universal presence [13]. They are strongly induced under heat

* Corresponding authors.
E-mail addresses: dmanoharrao27@gmail.com (D.M. Rao), pbkavi@yahoo.com (P.B. Kavi Kishor).

https://doi.org/10.1016/j.ijbiomac.2019.10.023
0141-8130/

stress, hence their name, but many sHsps are constitutively expressed in all kinds of cell types and abiotic stresses [14]. Hsp20s share a conserved modular structure and the most striking feature is the presence of a core conserved α-crystallin domain (ACD) containing 90 amino acids flanked by a variable N-terminal domain and a C-terminal extension [15]. The ACD is involved in substrate interactions, the N-terminus is responsible for substrate binding while the C-terminus participates in homo-oligomerization and results in the formation of heat shock (HS) granule [16,17]. Bondino et al. [18], Haslbeck and Vierling [19] showed that ACD is composed of two anti-parallel sheets of four and three β-strands; the conserved region I (CRI, β2-β3-β4-β5) in N-terminal domain and conserved region II (CRII, β7-β8-β9) at C-terminal domain, connected by hydrophobic β6-loop. Besides, Basha et al. [20] showed that the N-terminus contains the transit, target or signal sequences, while the C-terminus contains organelle-specific retention of amino acid motifs. Based on sequence homology, cellular compartmentation and immunological properties, sHsps are divided into 11 different classes; 6 cytoplasmic from I to VI and the remaining 5 located in organelles like mitochondria, chloroplasts, peroxisomes and endoplasmic reticulum [21].

Overexpression of heat and osmotic stress induced *AtHSP17.6A* gene during seedling stage conferred increased tolerance against salt and drought stresses in *Arabidopsis* [22]. Similarly, overexpression of chloroplast *sHSP21* gene exhibited enhanced tolerance against heat and high light stresses in transgenic *Arabidopsis* [23]. Overexpression of heterologous rice CI-*sHSP16.9* gene evinced increased thermotolerance in *E. coli* cells [24]; and CI-*sHsp17.7* gene displayed resistance against heat, drought and UV-B radiations [25,26]. Similarly, *DcHsp-17.7* gene overexpression exhibited tolerance against salt stress in *E. coli* [27]. The expression of *RcHSP17.8* gene conferred the resistance against a variety of stresses in *E. coli*, yeast and *Arabidopsis* [28]. Genome-wide analysis may help us in finding out the total number of *sHsps* that exist in *S. bicolor* and their expression under varied abiotic stress conditions in different tissues. Such an analysis revealed the presence of 19 *Hsp20s* in *Arabidopsis thaliana* [15], 23 in *Oryza sativa*, and 36 in *Populus trichocarpa* [29,30], 51 in *Glycine max* [31]. Reddy et al., reported 20 in *Hordeum vulgare* but recently Li et al., reported 36 [32,33] in the same plant. In *Capsicum* 35 [34], 37 in *Setaria italica* [35], 42 in *Lycopersicon esculentum* [36], 63 in *Panicum virgatum* [37], 163 in wheat [38], and 48 in potato [39].

S. bicolor, the fifth most important and dry-land cereal, the best C_4 model crop (African origin), provides food, fodder, feed and also fuel. Taxonomically, *Sorghum* belongs to the family *Poaceae* (grasses) and the tribe *Andropogoneae*. Due to its genetic diversity, smaller genome size with lesser complexity, bio-energy traits, and remarkable tolerance to drought and high temperature stress, *S. bicolor* is considered as a model crop for studying the abiotic stress tolerance compared to several other important cereals. No reports are available on the number and tissue specific expressions of *sHsps* under different abiotic stress conditions in *S. bicolor*. Hence, the present investigation is initiated in *Sorghum* with an aim to find out the number, class, distribution, motifs, promoters, phosphorylation sites and the structure besides their tissue specific expression profiles under diverse abiotic stresses and their evolutionary relationships.

2. Materials and methods

2.1. In silico identification of sHsps in S. bicolor

In silico search was made for *sHsp* gene sequences of *Arabidopsis*, and *Oryza* in NCBI database followed by blasting against *S. bicolor* genome in Gramene database (http://www.gramene.org/) [40]. Later, their respective coding sequences (CDS) and corresponding protein sequences were retrieved by employing Edit plus (http://www.editplus.com/) and Genscan (http://genes.mit.edu/GENSCAN.html) [41] programs. Based on their homology, nucleotide and protein sequences were identified and these putative protein sequences were subjected to SMART program (http://smart.embl-heidelberg.de/) and Pfam domain using HMMSCAN (http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) to identify their conserved domains [42], and validated to check the reliability by employing MOTIF search tool (http://www.genome.jp/tools/motif/). Proteins which failed to exhibit the reliability were eliminated.

2.2. Structure and sequence analysis of sHsps

Physical chromosomal map of identified *sHsp* genes was generated based on the information provided by the Gramene Genome Database (http://www.gramene.org/) [40]. The structure of the identified *sHsp* genes; exons, introns and untranslated sequence regions (UTRs) were identified based on the alignments of coding sequences with their corresponding gene sequences using Gene Structure Display Server (http://gsds.cbi.pku.edu.cn) software [43]. The GC content of *sHsps* was calculated by employing ENDMEMO software (http://www.endmemo.com/bio/gcdraw.php). The sHsp protein molecular weights (MW), isoelectric points (pI), instability and aromatic indexes, GRAVY (grand average of hydropathy), glycine, proline, and threonine contents were recorded by ProtParam software of Expasy tools (http://web.expasy.org/protparam/) [44]. The net charges of peptides were calculated by employing pepcalc.com of INNOVAGEN software (https://pepcalc.com/). The phosphorylation sites were predicted by employing NetPhosK3 software of Expasy tools [45]. The putative transmembrane helices were predicted by TMHMM server [46] and their sub-cellular localizations by employing Wolfpsort program (http://wolfpsort.org/) [47]. The Multiple Em for Motif Elicitation (MEME) software (http://meme-suite.org/) was used by setting different default parameters, and it helped to identify the new sequence patterns and their significance in sHsp proteins, by 1–10 motifs with a width of 5–50 amino acids, each containing 5–10 conserved motif sites [48].

2.3. In silico prediction of potential cis-regulatory elements, miRNA targets and phylogenetic analysis

After extracting 2000 bp upstream genomic sequences of start codons of *sHsp* genes from *Sorghum* genome sequence, the putative *cis*-regulatory elements in the promoter regions were identified by employing PLACE [49] and PLANTCARE [50] software programs. The psRNATarget server was used to screen the putative miRNAs targeting the sHsps [51]. Protein-protein interaction of sHsps was generated by employing the STRING database (https://string-db.org/) [52]. The Maximum Likelihood (ML) phylogenetic tree for sHsp sequences were constructed for *Sorghum* separately and the second one for *Sorghum* with *Oryza* (a monocot ancestor) and *Arabidopsis* (a dicot ancestor), to know the gene duplications and evolutionary relationships by employing MEGA 6.2 software with defined parameters; using Jones-Taylor-Thomton (JTT) model, Nearest-Neighbor-Interchange (NNI), ML Heuristic Method, with a bootstrap value of 1000 replicates for statistical reliability [53].

2.4. Estimation of non-synonymous and synonymous substitution rates of sHsp genes

The PAL2NAL (http://www.bork.embl.de/pal2nal/) [54] software was used to calculate the substitution rates for non-synonymous and synonymous sites of each of the identified

paralogs (*Sorghum*) and orthologous gene pairs (between *Sorghum/Oryza*, and *Sorghum/Arabidopsis*) from phylogeny.

2.5. Digital expression profiling of SbsHsp genes

In order to perform the digital expression profiling of identified *SbsHsp* genes, Affymetrix whole-transcriptome sorghum array data was accessed through SorghumFDB (http://structuralbiology.cau.edu.cn/sorghum/index.html) [55]. Genevestigator (https://genevestigator.com/gv/) [56] is an integrated platform containing a great variety of precise and defined validated experiments that allow to decipher the expression profiles of genes subjected to environmental stress conditions. In the present study, we utilized this tool to perform the microarray analysis for *SbsHsp* genes under drought, salt, heat and cold with different samples stored in the platform. The expression profiles of *SbsHsp* genes extracted from sorghum array was used for cluster analysis. Further, a heat map of expression profiling was generated by utilizing hierarchical clustering tool [57] of Genevestigator platform.

2.6. Plant material

Forty-day-old seedlings of *Sorghum bicolor* variety BTx623, grown in pots under glasshouse conditions were subjected to drought stress by subjecting them to 150 mM mannitol for 4 h; salinity stress by saturating the seedlings with 1 L of 150 mM NaCl solution for 4 h, heat stress by exposing to a high temperature of 40 °C in a growth chamber for 4 h, cold stress by subjecting them in a refrigerator at 4 °C temperature for 4 h. Triplicates were maintained for each treatment alongside the controls. After subjecting the seedlings to different types of abiotic stresses, different parts like roots, stems, and leaves along with respective controls, were collected separately, frozen immediately in liquid nitrogen and stored at −80 °C refrigerator until further use.

2.7. Isolation of total RNA and qRT-PCR analysis

Total RNA was extracted from the tissues by using nucleo spin plant RNA isolation kit (MACHEREY-NAGEL) following the manufacturer's instructions. The first strand cDNA was synthesized from total RNA (3 μg concentration) using first strand synthesis kit (Thermo Scientific). By using NCBI PRIMER Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/) [58] with the default parameters: 57–60 °C annealing temperature, 18–22 bp primer length, 50–55% GC contents, and 80–140 bp amplicon length, the gene specific primers were designed (Table S1). The SYBR Green Master Mix (2X) (Takara) was used according to the manufacturer's recommendations. The qRT-PCR analysis was carried out in Mx3000p (Agilent) with the following thermal cycles: 1 cycle at 95 °C for 10 min, followed by 40 cycles alternatively at 95 °C for 15 sec and 60 °C for 1 min. The amplicon dissociation curves were recorded with a fluorescence lamp after 40th cycle by heating from 58 to 95 °C within 20 min. The *SbAc-p2* (Acyl Carrier Protein 2) and *SbEF-P* (Elongation Factor P) genes were used as internal controls [59]. All the experiments were carried out thrice with three biological replicates. The average values are represented and relative gene expressions calculated by employing Rest software [60].

3. Results

3.1. In silico identification of SbHsp20 genes

The systematic genome-wide analysis of sorghum showed a total of 51 *SbHsp20* genes. After checking their reliability for α-crystalline domains (ACD) by employing SMART, 4 of them were excluded because of the absence of ACD, thus resulting in a total of 47 confirmed sHsps. In the *Sbs*Hsp members, the number of conserved domains (pfam) ranged from a minimum of one to the highest number of 9 in *Sbs*Hsp-01 (Table S2). The top domains appearing in sHsps were HSP20, ArsA_HSP20, BON, CS, and DUF domains which play important roles in diverse stress tolerance mechanisms.

Based on their class and chromosomal locations, the genes were named as *SbsHsp-01* to *SbsHsp-47* for convenience. The name of the gene, common name, chromosomal location, number of amino acids, class, DBDs, number of exons, molecular weight, isoelectric point (pI), subcellular localization, GRAVY, instability index and aliphatic index are listed in Table 1. Based on the subcellular localizations, they were classified into 10 subgroups. In one of the subfamilies, 17 genes were noticed, and this major subgroup localized in the chloroplast (Table 1).

3.2. Chromosomal localization and structure of SbHsp20 genes

The predicted *SbHsp20* genes were randomly distributed across 10 chromosomes. A maximum number of 12 genes were located on chromosome 1, 11 on 3rd chromosome, 7 on 4th chromosome, 6 on 10th chromosome, 3 on 2nd chromosome, 2 each on 5th and 9th chromosomes respectively, 1 each on 6, 7, and 8th chromosomes (Fig. 1 and Table 1). The structure of introns/exons along with UTRs of *SbHsp20* genes were determined by aligning genomic and full-length cDNA sequences employing GSDS software (Fig. 2). Among the 47 genes, 22 (46.80%) were found intronless, while 19 (40.42%) contained one intron. Four genes (8.51%) showed two, *SbsHsp-34* contained 3 and *SbsHsp-43* contained the highest number of 4 introns. Majority of the class-I and II groups were found to be intronless while the dominating plastid subgroup contained at least one intron (Table 1 and Fig. 2).

3.3. Analysis of SbHsp20 proteins

The lengths of the *Sbs*Hsp proteins varied from 131 (*Sbs*Hsp-08) to 546 (*Sbs*Hsp-19) amino acids. The molecular weights ranged from 13465 (*Sbs*Hsp-08) to 59839.2 Da (*Sbs*Hsp-19). A total of 31 proteins (65%) were found acidic in nature and the isoelectric points varied between 4.47 (*Sbs*Hsp-08) and 11.65 (*Sbs*Hsp-43). Majority of the identified *Sbs*Hsp20 proteins, except *Sbs*Hsp-07, 08 and 26 were hydrophilic in nature, while the GRAVY values ranged between −0.925 (*Sbs*Hsp-15) and 0.151 (*Sbs*Hsp-08). The aliphatic indices varied between 93.04 (*Sbs*Hsp-27) and 61.5 (*Sbs*Hsp-20). A total of 74.46% (35 out of 47) proteins were unstable, while the instability index varied from 30.83 (*Sbs*Hsp-38) to 67.89 (*Sbs*Hsp-34). Out of the 12 stable proteins, the major plastid subgroup contained 5, followed by class-II with 3, class-VI with 2 and one each in class-III and M (Table 1). The content of glycine ranged between 3.8% (*Sb*01g025960) and 15.5% (*Sb*10g0711300), while proline 2.3% (*Sb*10g007580) and 17.3% (*Sb*08g022220). On the other hand, threonine content ranged from 1.4% (*Sb*04g027330) to 11.5% (*Sb*03g026330). No significant difference was observed in percentages of glycine, proline and threonine. The net charges of *Sb*Hsps varied from −14.9 (*Sb*10g007570) to 26.2 (*Sb*10g0711300) (Table S3). The gene sequence analysis for GC content has not exhibited any significant differences, all of them contain high GC content, ranging from 54.1 (*Sb*03g006020) – 75.8 (*Sb*02g004080) (Table S3). Of the 47 sHsps, only 8 (*Sb*Hsp-09, 10, 28, 34, 41, 42, 44, and 47) contain the single transmembrane helices.

The *Sbs*Hsps were phosphorylated at serine and threonine residues, with the major class-IV phosphorylated at tyrosine also. The PKC, CKII, RSK, unsp, CKI, PKG, and cdc2 are the common kinases

Table 1
Hsp20 genes identified in *Sorghum bicolor* along with their chromosomal location, physico-chemical properties, number of exons, and sub cellular localizations.

Gene	Common name	Class	No. of a. a.	Chr loc.	Pi/MW	DBD	Exon	Local.	GRAVITY	Instability index	Aliphatic index
Sb01g039990	SbsHsp-01	I	158	1	6.19/17830.00	54–157	1	C	−0.716	62.58	64.75
Sb01g040000	SbsHsp-02	I	157	1	5.53/17421.4	53–156	1	C	−0.620	51.96	64.59
Sb01g040030	SbsHsp-03	I	158	1	5.82/17853.00	54–157	1	C	−0.672	60.69	67.22
Sb03g006870	SbsHsp-04	I	152	3	6.18/17131.3	48–151	1	C	−0.667	53.74	70.46
Sb03g006880	SbsHsp-05	I	172	3	5.39/19187.5	68–171	2	C	−0.620	68.37	68.55
Sb03g006900	SbsHsp-06	I	150	3	6.20/17043.2	46–149	1	C	−0.657	46.99	70.73
Sb03g006910	SbsHsp-07	I	165	3	4.72/17386.5	46–153	1	C	0.121	35.07*	88.18
Sb01g002570	SbsHsp-08	II	131	1	4.47/13465.00	106–125	1	C	0.151	31.48*	91.60
Sb01g012930	SbsHsp-09	II	294	1	8.90/32135.6	15–110	2	C	−0.671	67.73	69.80
Sb01g021170	SbsHsp-10	II	189	1	6.23/20841.4	19–113	2	C	−0.428	45.75	82.01
Sb03g006890	SbsHsp-11	II	152	3	6.76/17116.4	48–151	1	C	−0.561	53.33	74.28
Sb04g002330	SbsHsp-12	II	183	4	6.93/19888.4	79–182	1	C	−0.504	46.51	78.42
Sb04g007585	SbsHsp-13	II	201	4	5.82/22060.0	48–193	1	C	−0.681	46.48	71.34
Sb04g007600	SbsHsp-14	II	183	4	5.72/19994.7	52–175	1	C	−0.569	43.98	70.82
Sb04g030135	SbsHsp-15	II	187	4	9.54/20727.3	50–103	1	C	−0.925	43.49	51.18
Sb04g035130	SbsHsp-16	III	174	4	6.07/18637.9	59–172	2	C	−0.499	39.13*	78.39
Sb10g009090	SbsHsp-17	II	148	10	9.09/16017.3	35–145	1	C	−0.217	41.40	79.05
Sb03g003530	SbsHsp-18	II	165	3	5.93/17798.5	60–163	1	N	−0.313	32.81*	79.82
Sb03g006020	SbsHsp-19	VI	546	3	5.95/59839.2	329–349	1	N	−0.522	39.47*	65.20
Sb09g030540	SbsHsp-20	III	317	9	9.20/33986.0	34–127	2	N	−0.794	48.58	61.51
Sb10g007590	SbsHsp-21	VI	222	10	5.89/24193.3	124–222	1	N	−0.553	35.27*	74.3
Sb01g025600	SbsHsp-22	P	230	1	7.88/25326.6	170–227	3	Chl	−0.542	60.72	77.52
Sb01g025610	SbsHsp-23	P	231	1	9.69/25136.60	83–179	2	Chl	−0.683	61.19	67.97
Sb01g025960	SbsHsp-24	P	186	1	8.82/20198.2	83–186	2	Chl	−0.181	58.73	85.97
Sb01g041180	SbsHsp-25	P	215	1	5.14/23921.00	110–215	3	Chl	−0.613	49.64	72.51
Sb02g004080	SbsHsp-26	P	177	2	6.19/18168.66	88–154	1	Chl	0.102	35.07*	88.36
Sb03g006920	SbsHsp-27	P	181	3	9.94/20345.4	43–150	2	P	−0.412	41.26	93.04
Sb03g026330	SbsHsp-28	P	156	3	11.08/16141.2	36–102	2	Chl	−0.128	37.34*	69.04
Sb03g039370	SbsHsp-29	P	489	3	6.50/54544.1	360–465	1	Chl	−0.448	47.79	71.98
Sb02g10710.1	SbsHsp-30	P	251	4	8.77/27427.8	152–251	3	Chl	−0.359	42.50	83.63
Sb04g006890	SbsHsp-31	P	218	4	7.88/23711.7	119–218	2	Chl	−0.332	44.79	91.38
Sb04g027330	SbsHsp-32	P	220	4	6.78/24166.3	122–220	2	Chl	−0.537	54.02	78.45
Sb07g028370	SbsHsp-33	P	200	7	5.73/22247.0	86–200	2	Chl	−0.618	43.58	71.20
Sb08g022220	SbsHsp-34	P	306	8	9.64/32731.33	92–177	4	Chl	−0.681	67.89	63.53
Sb09g024680	SbsHsp-35	P	241	9	6.32/27040.5	153–240	2	Chl	−0.342	48.45	74.98
Sb10g007570	SbsHsp-36	P	227	10	4.65/24725.9	128–226	2	Chl	−0.320	54.61	77.05
Sb10g007580	SbsHsp-37	P	217	10	5.84/23327.5	121–217	2	Chl	−0.275	31.87*	90.83
Sb10g007600	SbsHsp-38	P	213	10	5.00/22633.7	115–213	2	Chl	−0.323	30.83*	81.08
Sb1058s002010	SbsHsp-39	P	223	10	6.76/23749.3	113–223	1	Chl	−0.319	38.03*	83.63
Sb03g026340	SbsHsp-40	M	240	3	9.36/26099.3	1–79	1	M	−0.816	37.47*	62.42
Sb01g021180	SbsHsp-41	G	263	1	9.21/28477.9	35–121	2	G	−0.626	56.63	69.32
Sb06g017856	SbsHsp-42	ER	216	6	5.62/23547.4	80–184	1	ER	−0.433	44.18	82.22
Sb10g0711300	SbsHsp-43	M	355	10	11.65/37466.3	100–156	5	M	−0.127	63.54	76.37
Sb01g012950	SbsHsp-44	VIII	252	1	6.46/27742.2	26–133	2	N	−0.642	43.89	76.19
Sb02g034760	SbsHsp-45	V	220	2	6.66/24871.0	122–206	3	N	−0.517	54.38	75.45
Sb05g002570	SbsHsp-46	M	257	5	8.67/27796.3	228–251	2	M	−0.398	47.19	75.68
Sb05g007030	SbsHsp-47	P	207	5	6.00/22149.2	83–188	1	Chl	−0.164	49.87	87.20

a.a.: amino acids, Chrom.: Chromosome, pI: iso electric point; MW: Molecular weight, Chl.: Chloroplast, C: cytoplasm, N: Nucleus, P: plastid, M: mitochondria, Extr: Extracellulr, GRAVY: Grand average hydropathy.
* Stable.

detected for the phosphorylation of *Sb*sHsp proteins (Table S4). By employing the MEME software, 10 conserved motifs were identified in *Sb*Hsp20 protein family. The lengths of these conserved motifs varied from 2 to 47 amino acids. Among the 10 motifs, motifs 1 and 2 appeared in all *Sb*Hsp20 proteins, except in *Sb*sHsp35. The motif 1 was detected at the C-terminal region, while motif 4 appeared in the N-terminus. Based on the analysis, the motifs 1, 2, 3, 4, 5, and 9 appeared to code for conserved ACD (α-crystallin domain) /HSP20/ArsA_HSP20 domain which is important for structural and biological function of sHsps. The motif 7, lysine-rich motif located at the C-terminus in some proteins, is important as a nuclear localization signal. Motifs 2, 4, 5, and 10 code for conserved region I (CR-I), while motifs 1 and 3 represent the conserved region II (CR-II). Similar motif distributions were observed in the same subgroups. The number, type and distribution of motifs have been found different in different subgroups of sHsps in *Sorghum* (Figs. 3 and S1).

3.4. Analysis of stress-responsive cis-acting elements in SbHsp20 promoters

The analysis of stress-responsive *cis*-elements in *SbHsp20* genes by PLANTCARE revealed their response to different stress conditions; both abiotic and biotic stresses besides the developmental stage specific response. Eight abiotic stress-response elements; DRE, DPBF, MYB, MYC, HSE, LTRE, GT1GM and etiolation-responsive, were detected in majority of *SbsHsp* gene promoter regions. Further, these promoter regions also exhibited the presence of abscisic acid (ABA), gibberellic acid, ethylene and methyl-jasmonic acid-responsive elements, ABA being the most dominating one. Majority of these genes showed the presence of at least one stress-responsive *cis*-element. Besides abiotic, the *SbsHsp* genes also showed the biotic stress-responsive elements such as WBOX, the pollen and endosperm-responsive, light-responsive, I-Box, and GATA BOX elements. The plastid group

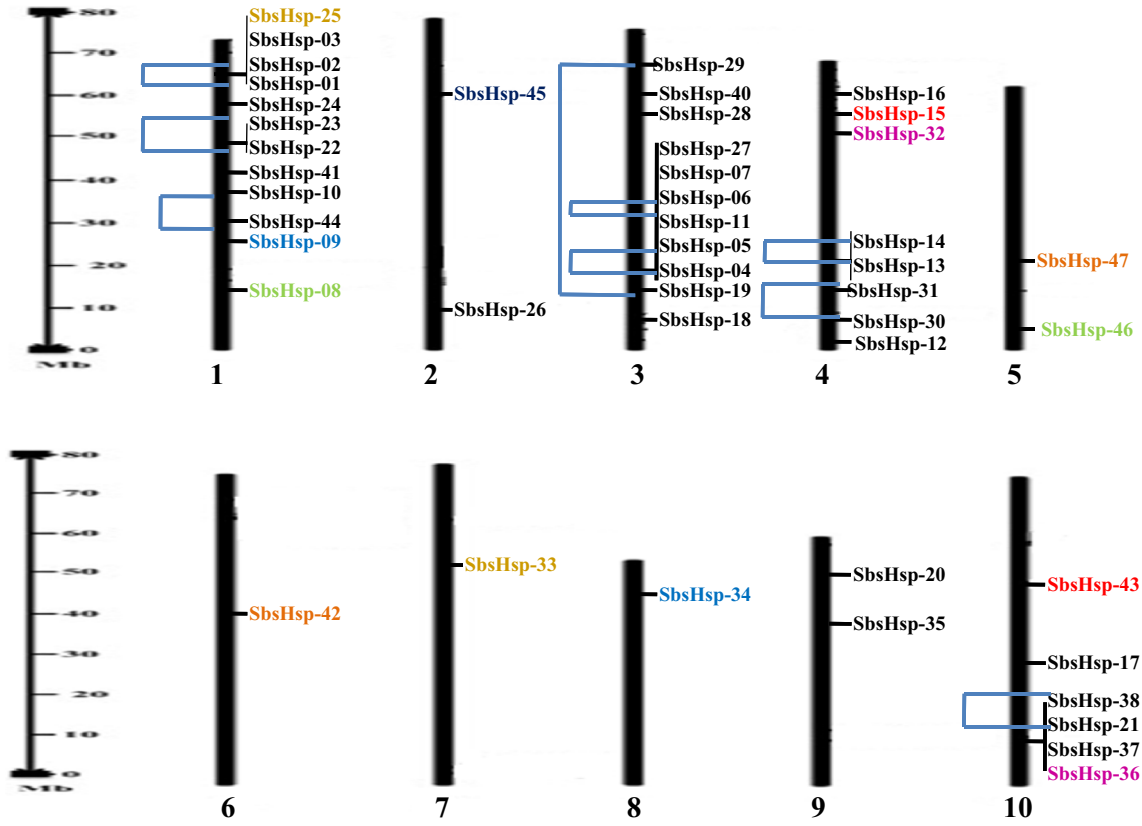

Fig. 1. Chromosomal distribution of *Sbs*Hsp genes. Duplications are illustrated by colours (Segmental duplications in same colour) and regional duplications are linked with line.

SbsHsps contained a maximum number HSEs, indicating their potential role in response to heat/temperature stress while the remaining have at least one HSE with an exception of *Sbs*Hsp-24 that does not contain any HSEs (Fig. S2 and Table S5).

3.5. Protein-protein interactions and in silico prediction of miRNA targeting sites

The protein-protein interaction network map of *Sbs*Hsps was constructed using STRING software (Fig. S3). The network illustrates that the Hsp20s are involved in protein processing in the endoplasmic reticulum, spliceosomal and endocytosis pathways. They are also involved in biosynthesis of secondary metabolites, carotenoids, steroids and participate in pyruvate, glyoxalate, cysteine and methionine metabolism and oxidative phosphorylation. The map exhibited interaction between the Hsp20 family of proteins, Clp, Hsp70, Dna-J, AAA domains, histidine kinases, lactate/-malate dehydrogenase, late embryogenesis abundant proteins, FAD binding proteins, methyltransferases and nucleoside triphosphate hydrolase proteins, indicating that they might participate in stress tolerance mechanism by interacting with other stress proteins.

The miRNAs and their target sites in *Hsp20s* were identified by employing psRNA Target Server. The analyses revealed a total of 25 different miRNAs, which are targeting 36 of the 47 *Hsp20s*. The highest number of miRNAs (23) are targeting *SbsHsp45*, while 15 *SbsHsp43*, and 14 *SbsHsp1*. The *sbi-miR5565*, *sbi-miR167*, *sbi-miR171*, *sbi-miR6230*, *sbi-miR164*, *sbi-miR6230* and *sbi-miR5568* are the most common miRNAs which are targeting *SbHsp20* genes and may participate in cleavage and translational events (Table S6).

3.6. Phylogenetic analysis of sHsp-20s

The ML phylogenetic tree of sHsp20 proteins of *S. bicolor* constructed by employing MEGA 6.2 software, showed 2 clades, the first with 11 proteins, and the second with 36 proteins. These clades were subdivided into 11 subgroups. Out of them, 10 subclasses with 47 *Sbs*Hsp proteins were grouped into 1 clade (Fig. 4a), but, 8, 9, 10, 11, 12, 15, 16, 20, 21, 34, 40, 41, 43, 44 and 46 proteins of the 10 classified subgroups were not clustered with their respective classes. Interestingly, class-II group is clustered with M (*Sbs*HSP 08/46 and 15/43), class-VIII (*Sbs*Hsp-10/44), and *Sbs*Hsp41 (G subclass), while class VI (*Sbs*HSP 19/29 and 21/38) clustered with plastid subgroup. These results indicate the evolution of M, G and VIII subgroups from class-II and VI from plastid. In all, *Sorghum* showed 15 paralogous duplications, out of which 9 were found regional duplications and 6 segmental duplications. While chromosomes 1 and 3 showed 3 regional duplications each, chromosome 4 showed 2, and chromosome 10 only 1 (Figs. 1, 4a and Table 2). A total of 19 sequences from *Arabidopsis*, 39 from *Oryza*, and 47 from *Sorghum* were used in the construction of phylogenetic tree and to analyse the evolutionary relationships of *sHsp20* genes in *Sorghum*, *Arabidopsis*, and rice which resulted into 2 clades and 13 subclades. *Sorghum sHsp20* genes exhibited a total of 20 orthologous relationships with *Oryza*, while no orthology was noticed between *Sorghum* and *Arabidopsis*. Only 1 event between *Oryza* (*Os*05g42120) and *Arabidopsis* (*At*5g54660) was observed (Fig. 4b and Table S7), inferring *Oryza* as the common ancestor. Majority of the *Arabidopsis* sHSPs showed independent branches, but few of them like *At*1G54050, *At*5G3670, *At*4G10250, *At*4G21870 and *At*4G27670 were clustered with *Oryza*, and *Sorghum* indicating the differentiation and divergence of sHSPs from the common ancestor.

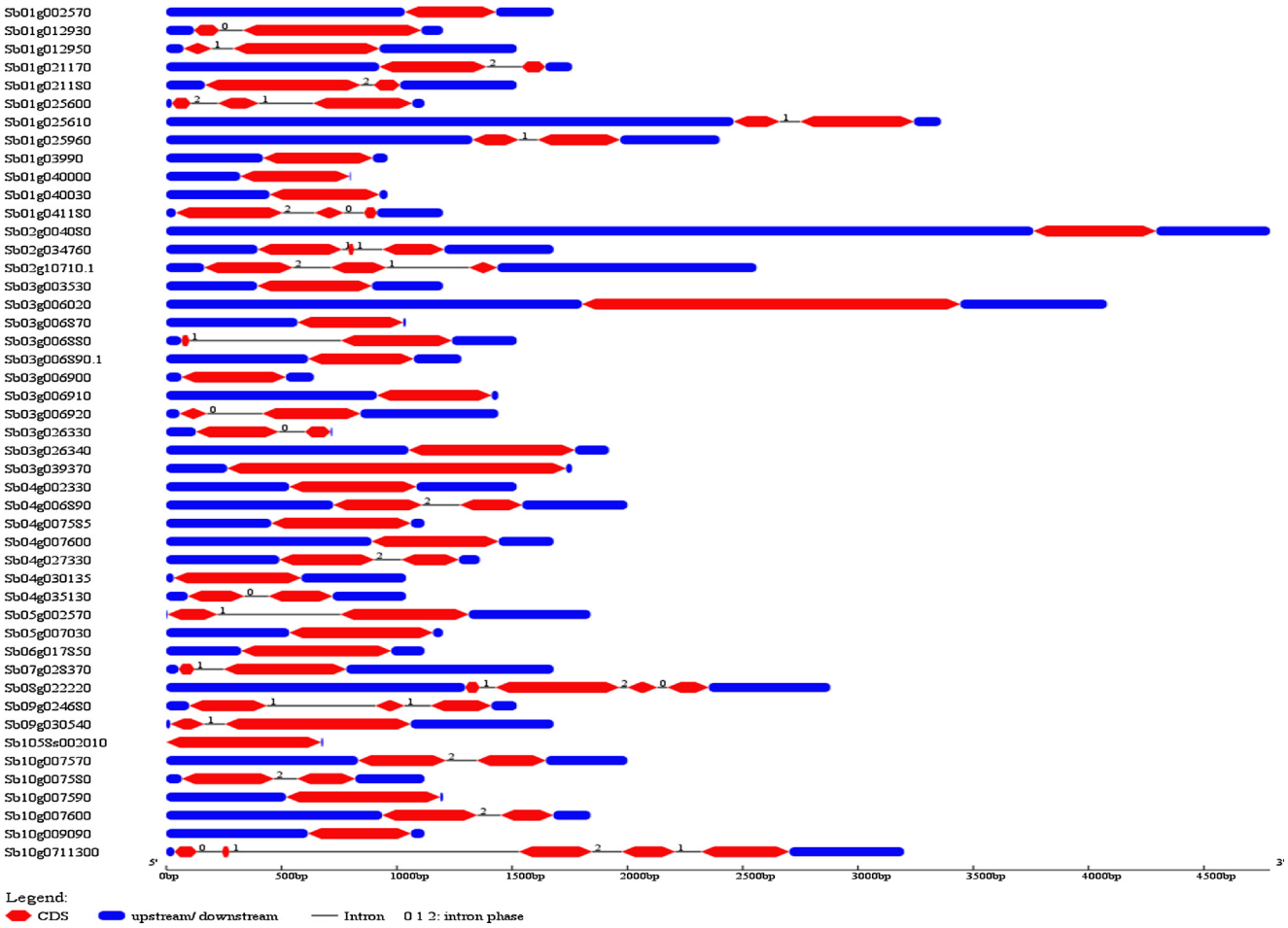

Fig. 2. Gene structures of the *SbsHsp* family genes.

3.7. Calculation of synonymous and non-synonymous substitution rates (dN/dS) of SbsHsps

The non-synonymous substitution (d_N) to synonymous (d_S) ratios of paralogous and orthologous gene-pairs were estimated to find out the Darwinian selection in duplication events. Among the 15 paralogs, only 7 events showed dN/dS ratios below 1, implying purifying/stabilizing selection, while 8 indicates a positive/Darwinian selection (Table 2). Out of 20 orthologous events, 4 events exhibited d_N/d_S ratios less than 1, while the remaining 16 showed d_N/d_S ratios > 1, following Darwian/positive selection (Table S7 and Fig. 4b).

3.8. Digital gene expression analysis of SbHSP20s in different tissues and developmental stages

In present study, it has been noticed that microarray data were available only for 43 of the 47 *SbsHSP* genes on SorghumFDB and Genevestigator platform. These 43 *SbsHSP* genes were further utilized for expression profiling. Expression of these 43 *SbsHsp* genes in six different tissues such as roots, pith, rind, internode, shoot, and leaf were investigated under abiotic stress conditions (Fig. 5a). It has been observed that the expression level was higher in the root, shoot, and leaf, indicating that abiotic stresses and/or high metabolic activity generally lead to up-regulation of *SbsHsp* genes in a tissue-specific manner. As compared with other *SbsHsps*, 10 members of the *SbsHsps* such as *SbsHsp-03*, *SbsHsp-07*, *SbsHsp-11*, *SbsHsp-12*, *SbsHsp-22*, *SbsHsp-23*, *SbsHsp-24*, *SbsHsp-29*, *SbsHsp-35*, and *SbsHsp-46* were found highly expressed in different tissues (Fig. 5a). The expression profiles of *SbsHsp* genes were also analyzed at five different developmental stages (stem elongation, booting, flowering, dough, and seedling) (Fig. 5b). Interestingly, *SbsHsps* genes were found expressed in all developmental stages (either up-regulated or down-regulated) as shown in Fig. 5b. However, the expression of *SbsHsp* genes in the flowering, dough and booting stages demonstrated a slightly different pattern; genes like *SbsHsp-03*, *SbsHsp-16*, and *SbsHsp-46* displayed a dominant expression profile compared to other developmental stages. High expression of *SbsHsp* genes in flowering, dough and booting stages might have been caused by their related cellular deteriorations, leading to substantial metabolic or physiological changes that significantly affect the overall regulation under environmental conditions (Fig. 5b).

3.9. Digital gene expression analysis of SbHSP20s under abiotic stress conditions

Hierarchical clustering based on the above expression analysis of individual *SbsHsp* genes under different environmental conditions allowed grouping of these 43 *SbsHsp* genes into two major clusters. One of these clusters contained highly upregulated *SbsHsp-04* and *SbsHsp-18* genes (only two) under different stress conditions. The remaining *SbsHsp* genes were classified among other subclusters of the second major cluster (Fig. 5c). The heat map of different *SbsHsp* genes showed significantly altered expression (either upregulated or down-regulated) up to 2.5-folds under

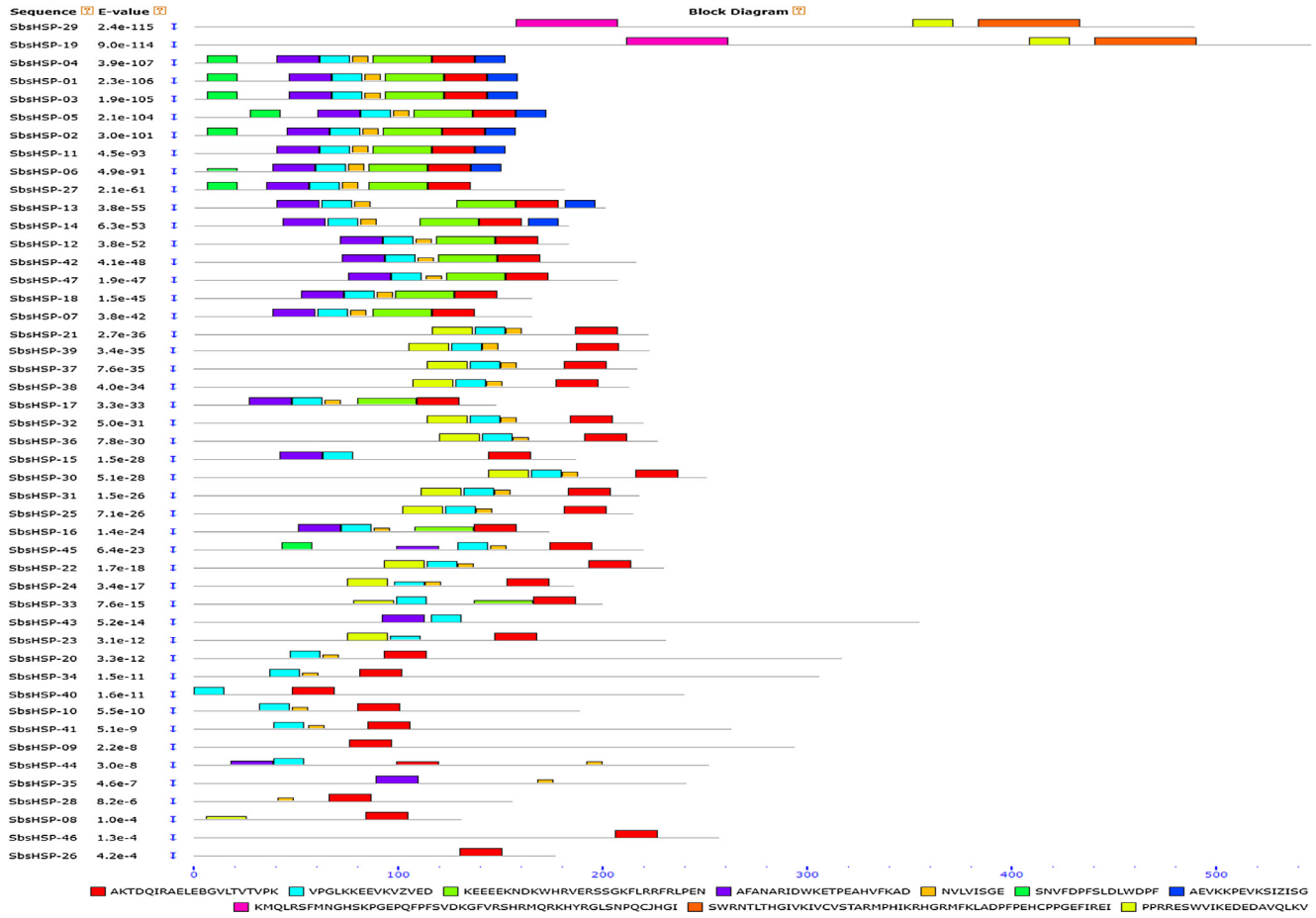

Fig. 3. Conserved motif distribution in *Sbs*Hsp sub families. The scale represents the lengths of the proteins and motifs.

different abiotic stress conditions (Fig. 5c). While 12 members of the *SbsHsp* genes such as *SbsHsp-01*, *SbsHsp-02*, *SbsHsp-03*, *SbsHsp-04*, *SbsHsp-05*, *SbsHsp-11*, *SbsHsp-12*, *SbsHsp-16*, *SbsHsp-18*, *SbsHsp-25*, *SbsHsp-32*, and *SbsHsp-33* were found upregulated, three *SbsHsps* (*SbsHsp-22*, *SbsHsp-23*, *SbsHsp-24*) divulged down-regulation under salt, cold, and drought stresses (Fig. 5c). These results are in confirmation of the quantitative real-time expression analysis carried out for a set of *SbsHsp* genes in the present study.

4. qRT-PCR expression analysis of *SbHsp20* genes

Expressions of selected *SbHsp20* genes in root, stem, and leaf tissues of *S. bicolor* were carried out by qRT-PCR. The results revealed that the expression of majority of the genes at least in one or the other tissues. All the selected genes expressed in root tissues with the exception of *SbsHsp-15* and *SbsHsp-47*. But, the highest expression values were recorded in the leaf tissues. Most of the selected genes showed down-regulation, while some did not exhibit any expression in stems. In roots, the highest expression (1.2-folds) was noted in *SbsHsp-02*, but in leaves it was *SbsHsp-47* (1.13-folds) and *SbsHsp-15* (1.05-folds). Majority of them displayed a constitutive expression in the root and leaf tissues (Fig. 6a and Table S8). To understand the role of transcript changes in the stress regulatory mechanism, analysis of selected *SbsHsp* genes in root, stem and leaf tissues under various abiotic stresses like drought, salt, high temperature and cold were carried out (Fig. 6b). The relative expression of selected *SbsHsp* genes revealed varied levels in different tissues under diverse stresses. Among the selected *SbsHsp* genes, *SbsHsp-02* was found as the candidate gene with the highest expression values in majority of the tissues under all abiotic stresses tested. All the genes in leaf tissues exhibited constitutive expression under drought stress conditions, except *SbsHsp-15*, *SbsHsp-19* and *SbsHsp-47* which evinced down-regulation. In stem and root tissues, *SbsHsp-02* and *SbsHsp-43* exhibited upregulation under all the stress conditions. The *SbsHsp-02* exhibited 56.88-folds increase in expression under high temperature stress in root tissues, 39.39-folds under cold stress, 14.42-folds under salt and 12.61-folds under drought stress. On the other hand, *SbsHsp-43* displayed moderate expressions. Under salt stress, majority of them were upregulated in leaf. Interestingly, *SbsHsp-43* being the candidate gene displayed the highest expression of 145-folds. In stems, *SbsHsp-02* gene showed activation, while *SbsHsp-46* and *SbsHsp-47* exhibited moderate expressions. Majority of the selected genes were induced under high temperature stress, and also recorded high constitutive expression in the leaf in comparison with root and stem tissues. The *SbsHsp-30*, *35* and *40* displayed significantly higher expression levels in the root tissues under high temperature stress. In stem tissues, under the high temperature stress, 6 of the 15 genes exhibited upregulation. Under cold stress conditions, genes such as *SbsHsp-02*, *30*, *35*, *41*, *43*, *45* and *47* expressed constitutively in all the tissues (Fig. 6b and Table S8).

5. Discussion

In the present study, a total of 47 *sHsp* genes were identified in *S. bicolor*. This number appears slightly higher in comparison with the number of *sHsp* genes identified in other species; 31 in *Arabidopsis* [15,29], 39 in rice [30], 36 in barley [33], 35 in pepper

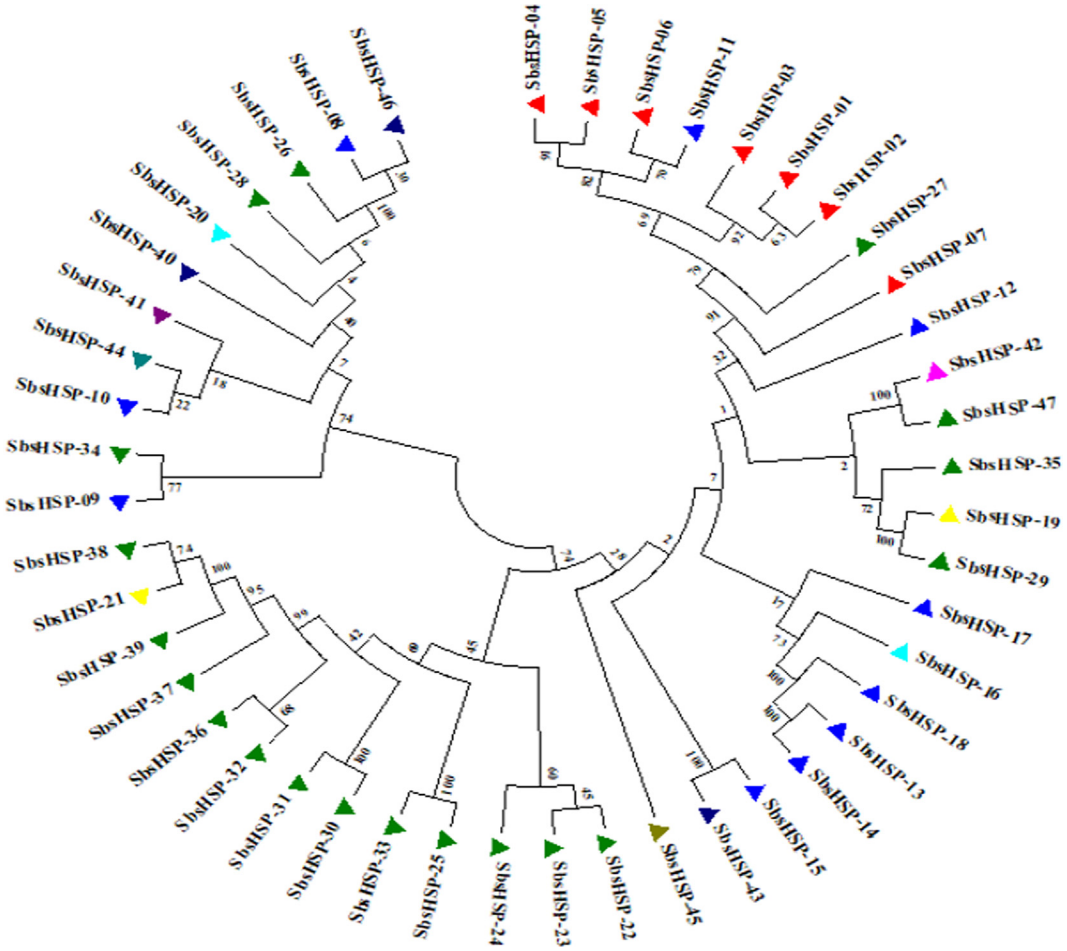

Fig. 4. (a) Phylogenetic analysis of 47 *SbsHsp* genes. The gene sub groups were classified based on their homology and are distinguished by different colors. (b) ML phylogenetic tree showing the relationship between sHsp proteins in rice, *Arabidopsis* and *Sorghum*.

[34], 37 in *Setaria* [35] and 42 in tomato [36]. Even though the molecular chaperone families exhibited the highest diversity, the sHsps shared a conserved ACD structural domain, flanked by a non-conserved N-terminal sequence of variable length with a short conserved C-terminal sequence [18,61]. The MEME characterization of all the selected *Sbs*Hsps exhibited varied ACD sequences. The gene structure and organization play an important role in the evolution and their stress response of multiple gene families [62]. Majority of the *sHsp* genes have been found intronless, a characteristic feature of *Hsps*. The results obtained in the present study are consistent with the previous reports indicating that plants retain the more number of intronless genes or with short introns perhaps to adjust themselves to the stress conditions [63]. Generally, the number and lengths of the introns negatively regulate the gene expression [64,65]. Previous reports suggest that the introns slow down the regulation of stress-responsive genes [66]. Compared to rice [30] and *Setaria* [35], the members of *Poaceae* family, *S. bicolor* sHsps showed varied molecular weights, isoelectric points, and number of introns. The instability index of the most of the *Sbs*Hsp proteins was equal or greater than 40, which is considered as an important feature of stress proteins and thus shed light on the rapid induction of *Hsp20* genes [67]. In the present study, the sequence analysis revealed that the sHsps are endowed with high content of glycine, proline and threonine. High content of these amino acids play an important role in the abiotic stress tolerance as pointed out by Chowdhury et al. [68]. Sorghum *Hsp20s* revealed more GC content than *Arabidopsis* and rice. High GC base pairs are responsible for higher thermal stability compared to the AT regions. It occurs due to the stronger stacking interaction between GC bases and triple hydrogen bond [69].

In the present study, the ML phylogenetic tree constructed to know the evolutionary relationship between the *Hsp20* genes of *Sorghum*, *Arabidopsis*, and rice, resulted into 13 subfamilies, similar to the previous reports [70,30,36]. In *Arabidopsis*, *Hsp20* genes were divided into 12 groups (CI - CVII, MI, MII, P, ER, and Px) [15,70]. In the ML phylogenetic tree of *S. bicolor*, 47 *sHsps* are grouped into 10 subclasses distributed into 13 subclades, similar to 42 *Hsp20* genes identified in tomato, and grouped into 13 subgroups of the 17 subfamilies [36]. Out of the 10 subgroups, group G is localized in the Golgi. Surprisingly, four novel nucleo-cytoplasmic subfamilies (CVIII, CIX, CX, and CXI) reported in rice [30] are absent in *S. bicolor*, inferring the divergence of Hsp20 family proteins from their ancestors. In the present investigation, the CIV, and MII subgroups did not exhibit their presence, similar to that of tomato [34,36], but in contrast to that of *Arabidopsis* [15]. In rice, the C-I is reported as the largest subfamily with 7 genes [30], whereas in the present study, plastid subgroup has been noticed as the largest group with 17 genes localized in the chloroplast, followed by subgroup II with 9 genes localized in the cytoplasm. It is reported that majority of the *Hsp20* genes are located in the cytoplasm, also believed to be the site of protein synthesis. This is crucial to interact with denatured proteins so as to prevent aggregation and denaturation of them under stress conditions [31]. Contrary to this notion, the plastid subgroup with the largest number of 17 genes was found localized

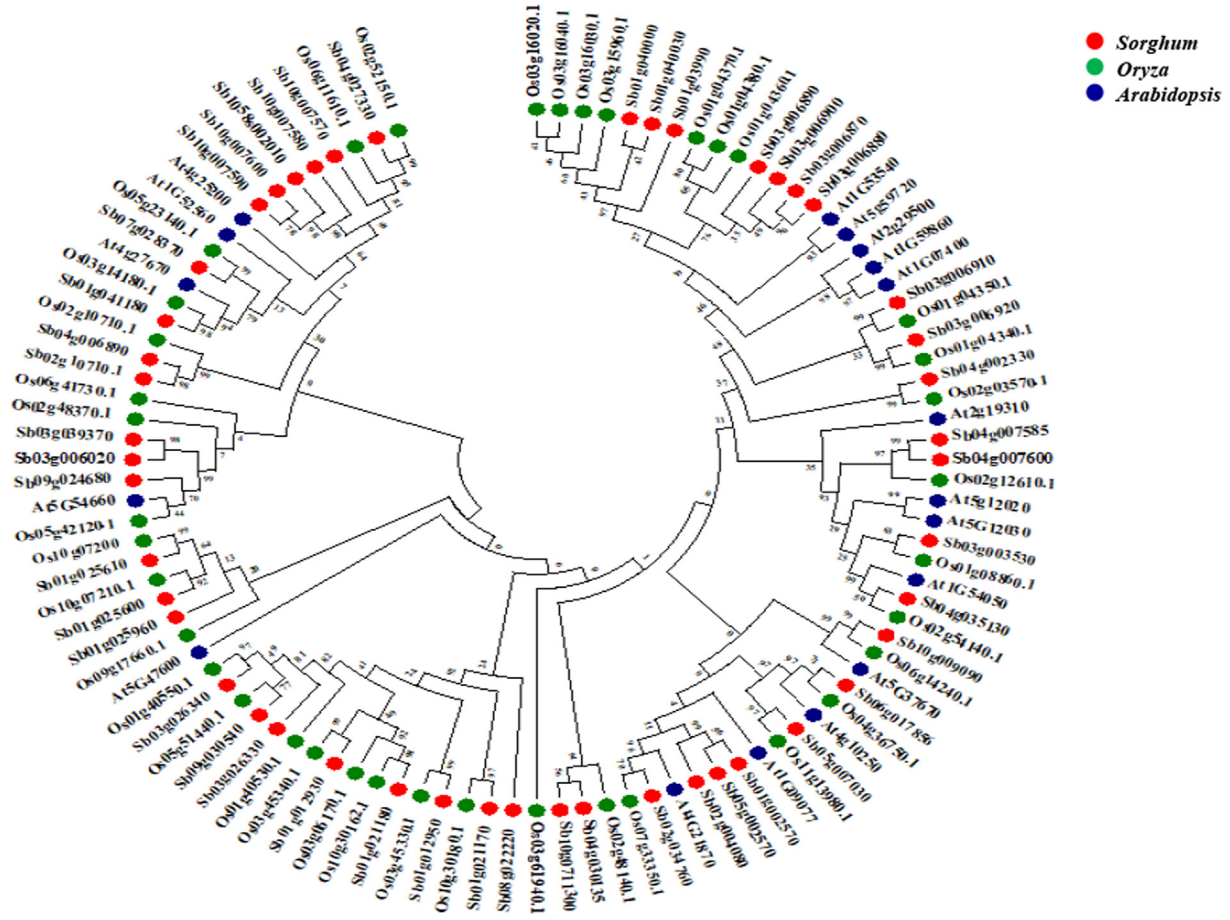

Fig. 4 (*continued*)

Table 2
Non synonymous to synonymous substitution ratios of SbHsp-20 paralogs.

SbHsp20 Gene 1	Chr	SbHsp20 Gene 2	Chr	No. non Synonymous sites (N)	No. Synonymous sites (S)	Non Synonymous substitution rate (d_N)	Synonymous substitution rate (d_S)	d_N/d_S
SbsHsp-01	1	SbsHsp-02	1	360.4	110.6	1.2504	1.7258	0.7246
SbsHsp-04	3	SbsHsp-05	3	370.7	85.3	4.9334	0.0498	99.0000
SbsHsp-06	3	SbsHsp-11	3	343.0	107.0	2.1997	0.7301	3.0128
SbsHsp-08	1	SbsHsp-46	5	302.0	91.0	16.6901	0.7616	21.9156
SbsHsp-09	1	SbsHsp-34	8	691.1	190.9	12.8366	13.5914	0.9445
SbsHsp-10	1	SbsHsp-44	1	453.1	113.9	6.4396	0.0650	99.0000
SbsHsp-13	4	SbsHsp-14	4	450.0	99.0	4.2600	1.2895	3.3035
SbsHsp-15	4	SbsHsp-43	10	415.9	145.1	3.2333	7.9415	0.4185
SbsHsp-19	3	SbsHsp-29	3	1106.4	360.6	4.8496	5.2728	0.9197
SbsHsp-21	10	SbsHsp-38	10	503.8	135.2	5.6543	0.0571	99.0000
SbsHsp-22	1	SbsHsp-23	1	542.5	147.5	5.1444	10.0994	0.5094
SbsHsp-25	1	SbsHsp-33	7	475.4	124.6	7.5404	0.0762	99.0000
SbsHsp-30	4	SbsHsp-31	4	484.0	170.0	13.0560	12.8405	1.0168
SbsHsp-32	4	SbsHsp-36	10	511.3	148.7	3.2895	3.4562	0.9518
SbsHsp-42	6	SbsHsp-47	5	449.9	171.1	2.5223	40.5448	0.0622

$d_N/d_S > 1$ = Positive or Darwinian Selection (Driving Change); $d_N/d_S < 1$ = Purifying or Stabilizing Selection (Acting against change); $d_N/d_S = 1$ Neutral Selection.

in the chloroplast of *S. bicolor*. Chloroplast proteins are also subjected to stress and undergo denaturation, and perhaps to protect such denaturations, such a large number of Hsps are inevitable.

From the evolutionary point of view, no Hsp20 genes were reported in green algae, while only cytosolic sHsps like CI, CII, and P subfamilies were reported in mosses [71,72]. This infers that evolution of *Hsps* is of a later origin. The emergence of gene families through gene duplication, gene loss, conversions and recombination events were reported [73,74]. The evolution of *sHsps* suggests that their duplications emanated in the formation of groups in higher plants [21], and such duplication events were described as the key factors ensuing in the evolution of genetic systems and genomes [75]. In the present study, of the 15 identified paralogous events, 9 were regarded as regional duplications and 6 segmental. The results suggest the remarkable contributions of both the regional/tandem and segmental duplications in the expansion of *sHsp* genes in *S. bicolor*. Identical results were reported earlier in *Oryza*, pepper and tomato [30,34,36].

The miRNAs are small RNAs, the endogenous regulators with 21–24 nucleotides. The miRNAs regulate many developmental pro-

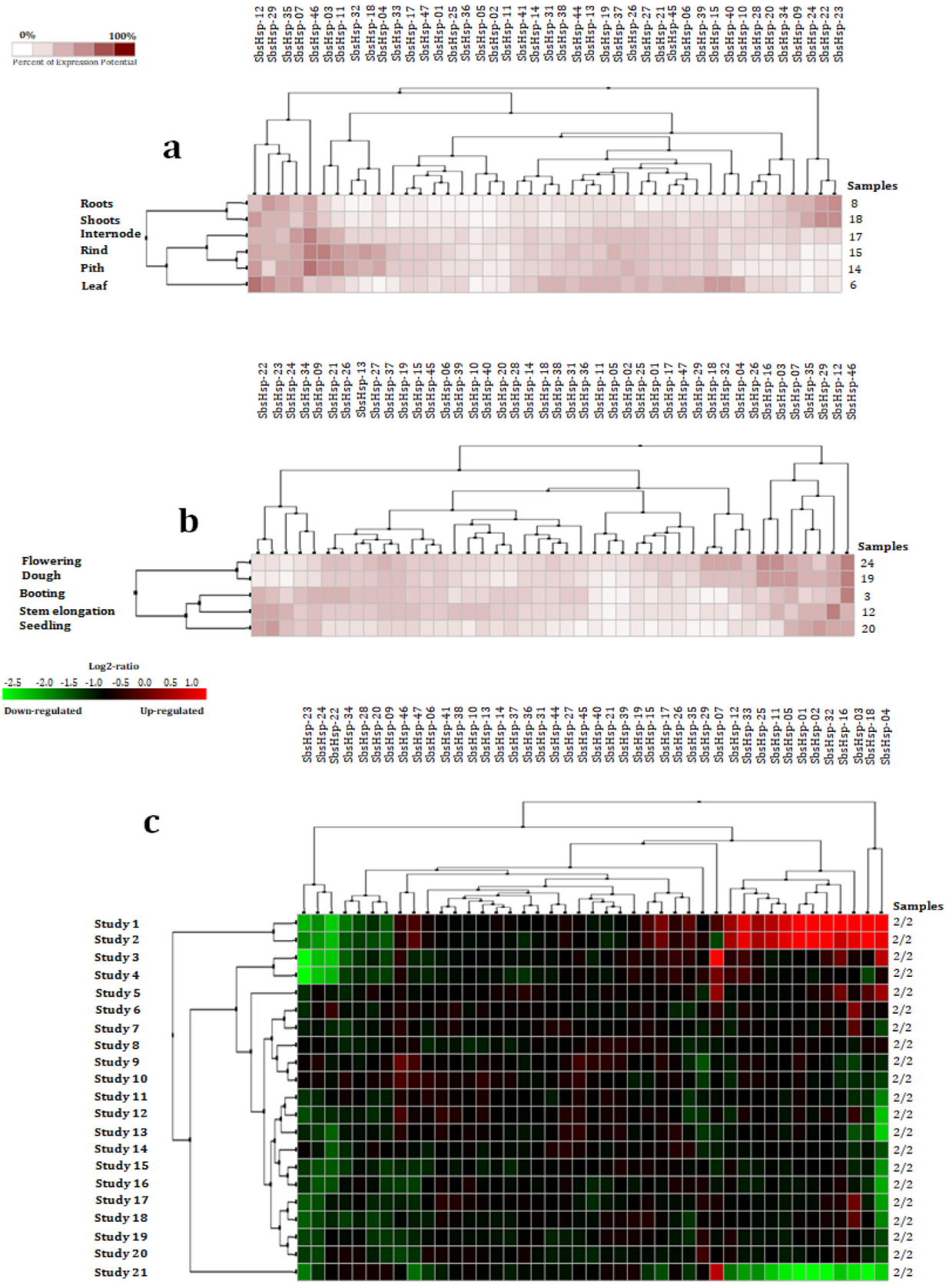

Fig. 5. Digital expression analysis of *SbsHsp* genes in (a): 6 different tissues; (b) 5 different development stages; and (c) under different abiotic stress conditions (Hierarchical clustering of *SbsHsps* genes based on their expression).

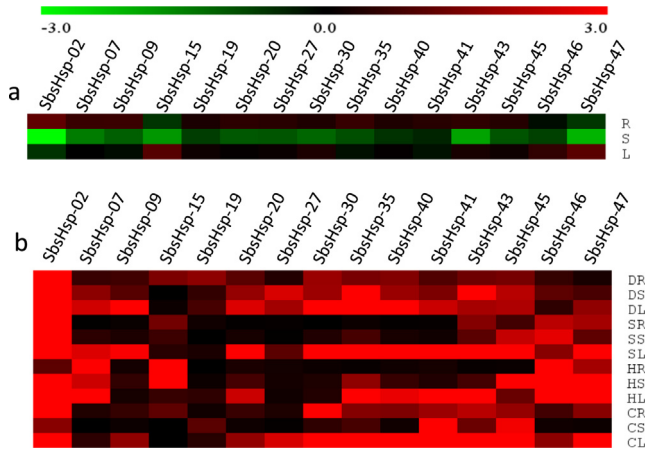

Fig 6. qRT-PCR expression analysis of *SbsHsp* genes in (a) root, stem, and leaf tissues (b) root, stem, and leaf tissues under drought, salt, heat and cold stress. (DR: Drought root, DS: Drought stem, DL: Drought Leaf, SR: Salt root, SS: Salt stem, SL: Salt leaf, HR: Heat root, HS: Heat stem, HL: Heat leaf, CR: Cold root, CS: Cold stem, CL: Cold leaf).

cesses in plants by negatively regulating the target genes by cleavage of target mRNAs [76,77]. They target the transcription factors which control developmental processes in *Arabidopsis*, maize, many woody species and also stress responses [78,79]. In the present investigation, miR167 was found to be the common miRNA and targets *SbHsp20s*. In *Arabidopsis*, miR167 participates in plant development by regulating *ARF6* and *ARF8* [80]. Interestingly, the *Sbi*miR5565 identified in the present investigation is a novel miRNA present only in *S. bicolor*, and targets E3 ubiquitin protein ligase, SNARE protein syntaxin, SAM decarboxylase, putative receptor kinases, oxidoreductase and arabinogalactan proteins. While miR164 regulates NAC transcription factors, miR171 targets Scarecrow-like proteins, and participates in diverse developmental and stress-responsive pathways [81].

It is reported that the downstream *Hsps* are regulated by the activation of *Hsfs* under heat stress conditions by binding to the HSE elements of the promoters of the *Hsp* genes. In the present study, promoter analysis of *SbHsp20* genes revealed the presence of various *cis*-elements in them which interact with many functional genes. Several abiotic stress-responsive elements like DRE, DPBF, LTRE, GT1GM, MYB, and MYC were noticed that may participate in diverse stress regulatory mechanisms as also noted by Park et al. [82] and Li et al. [83]. The biotic stress-responsive elements like WBOXNTERF3, WBOXATNPR1, and CGTCA, respond to wounds, pathogens and salicylic acid [84,85] indicating that *SbHsp20* family genes also play a pivotal role in biotic stress tolerance. The promoter analysis of *SbHsp20* revealed their participation in plant developmental processes such as pollen architecture [86], and guard cell movement [87]. Yanagisawa [88] reported that *Sb*Hsp20 family proteins along with Dof transcription factors participate in DNA binding and carbon metabolism. They are rich in CACT, a key component of mesophyll expression module, similar to the promoter of *phosphoenolpyruvate carboxylase* gene [89]. The ARR1 elements regulate the non-symbiotic haemoglobin2 in plants [90]. The *SbHsp20* genes have a maximum number of GATA box elements, and molecular light switches play a vital role in the control of light and nitrate-dependent transcription of genes [91]. These results suggest that the expression of *SbHsp20* genes are under the complex signal transduction and participate in diverse cellular functions by playing a key role in plant development as well as stress tolerance.

The diverse tissue specific expression of *SbHsp20* genes observed in the present study is similar to several other studies reported earlier in *Arabidopsis*, rice, tomato, pepper, and soybean [21,31]. The *SbHsp20* genes displayed both up- and/or down-regulations in root, stem, and leaf tissues, similar to the earlier reports in tomato and pepper [34,36]. He and Yang [92] pointed out that generally sHsps were absent in vegetative tissues but their expressions were observed during developmental stages. This infers their role in plant developmental processes and they may act as house-keeping genes [93,31]. Most of the genes were down-regulated in stems, but recorded an upregulation under stress conditions, inferring their crucial role in stress-tolerance mechanisms. Thus, the *sHsp20* family is not only involved in the plant developmental processes but also in environmental stress tolerance [94]. In the present investigation, majority of the *SbHsp20* genes exhibited upregulation in roots and leaves under heat stress, while they showed negative expression in stems, indicating their role in thermotolerance, with functional redundancy [95,21]. In the present investigation, expression analysis of *SbsHsp-02* gene, localized in the cytoplasm has been found upregulated under diverse stresses supporting the earlier view of Lopes-Caitar et al. [31]. Next to *SbsHsp-02* the *sHsp-43*, *45*, *46*, and *47* were found highly expressed under diverse abiotic stress conditions. Thus, these genes appear as potential candidates in *S. bicolor* for imparting tolerance to multiple abiotic stresses. Several *in vivo* studies demonstrate the multiple stress tolerance nature of *sHsps*; for example, expression of mitochondrial *sHsp* in tobacco showed higher thermotolerance [96], and transgenic *Arabidopsis* lines overexpressing *Hsp26* exhibited tolerance against heat stress [97]. The *RcHSP17.8*, PtHSP17.8, and ZmHSP16.9C displayed tolerance against several abiotic stresses [21,98,99], indicating their multifaceted nature and important implications in stress tolerance mechanisms.

6. Conclusions

Genome-wide screening of *S. bicolor* genome for the identification of *SbsHsps* or *SbHsp20* revealed the presence of 47 genes. The detailed analysis disclosed their structural organization, subcellular localizations, physico-chemical properties, *cis*-elements, phylogenetic and evolutionary relations, and expressions under diverse abiotic stress conditions in different tissues. The analysis further revealed a large sub-family of subgroup with 17 genes localized in the chloroplast for the first time reported in *S. bicolor*. Expression analysis indicated their role not only in plant development but also in stress tolerance. This study lays a foundation for functional characterization of *Hsp20s*, and help to understand the mechanisms of abiotic stress tolerance under diverse stress conditions.

Authors contribution

PBK and DMR designed the experiments. MN implemented and collected the data. MN, PSR, SAK, AK and RG performed the experiments. MN, PBK, DMR, PSR, SAK, and AK analyzed the results. MN, PBK and DMR prepared the manuscript and revised. All authors have read and approved the manuscript.

Declaration of Competing Interest

The authors declare that they have no competing interests. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

MN is thankful to the UGC, New Delhi, for providing fellowship. PBK is thankful to the CSIR, New Delhi, for providing CSIR-Emeritus Scientist Fellowship.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2019.10.023.

References

[1] Michael V. Mickelbart, Paul M. Hasegawa, Julia Bailey-Serres, Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability, Nat. Rev. Genet. 16 (4) (2015) 237–251, https://doi.org/10.1038/nrg3901.

[2] S. Katiyar-Agarwal, R. Morgan, D. Dahlbeck, O. Borsani, A. Villegas, J.K. Zhu, B.J. Staskawicz, H. Jin, A pathogen-inducible endogenous siRNA in plant immunity, PNAS 103 (2006) 18002–18007.

[3] R.C. de Ahuja, A.M. Vos, R.D. Hall Bones, Plant molecular stress responses face climate change, Trends Plant Sci. 15 (2010) 664–674.

[4] R. Mittler, A. Finka, P. Goloubinoff, How do plants feel the heat?, Trends Biochem. Sci. 37 (2012) 118–125.

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

[6] M. Guo, J.P. Lu, Y.F. Zhai, W.G. Chai, Z.H. Gong, M.H. Lu, Genome-wide analysis, expression profile of heatshock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum* L.), BMC Plant Biol. 15 (2015) 151.

[7] H.S. Mchaourab, J.A. Godar, P.L. Stewart, Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins, Biochemistry 48 (18) (2009) 3828–3837.

[8] M.P. Mayer, Gymnastics of molecular chaperones, Mol. Cell 39 (2010) 321–331.

[9] S.J. Eyles, L.M. Gierasch, Nature's molecular sponges: small heat shock proteins grow into their chaperone roles, Proc. Nat. Acad. Sci. 107 (2011) (2010) 2727–2728.

[10] E. Vierling, The roles of heat shock proteins in plants, Ann. Rev. Plant Biol. 42 (1991) 579–620.

[11] E.R. Waters, G.J. Lee, E. Vierling, Evolution, structure and function of the small heat shock proteins in plants, J. Exp. Bot. 47 (3) (1996) 325–338.

[12] B. Kokke, M.R. Leroux, E.P.M. Candido, W.C. Boelens, W.W. de Jong, *Caenorhabditis elegans* small heat shock proteins Hsp12.2 and Hsp12.3 form tetramers and have no chaperone-like activity, FEBS Lett. 433 (1998) 228–232.

[13] M. Haslbeck, T. Franzmann, D. Weinfurtner, J. Buchner, Some like it hot: the structure and function of small heat-shock proteins, Nat. Struct. Mol. Biol. 12 (2005) 842–846.

[14] F. Narberhaus, α-Crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network, Microbiol. Mol. Biol. Rev. 66 (2002) 64–93.

[15] K.D. Scharf, M. Siddique, E. Vierling, The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins), Cell Stress Chaperones 6 (2001) 225–237.

[16] E. Basha, K.L. Friedrich, E. Vierling, The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity, J. Biol. Chem. 281 (2006) 39943–39952.

[17] N. Jaya, V. Garcia, E. Vierling, Substrate binding site flexibility of the small heat shock protein molecular chaperones, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 15604–15609.

[18] H.G. Bondino, E.M. Valle, A. Ten Have, Evolution and functional diversification of the small heat shock protein/α-crystallin family in higher plants, Planta 235 (2012) 1299–1313.

[19] M. Haslbeck, E. Vierling, A first line of stress defense: small heat shock proteins and their function in protein homeostasis, J. Mol. Biol. 427 (2015) 1537–1548.

[20] E. Basha, C. Jones, A.E. Blackwell, G. Cheng, E.R. Waters, K.A. Samsel, M. Siddique, V. Pett, V. Wysocki, E. Vierling, An unusual dimeric small heat shock protein provides insight into the mechanism of this class of chaperones, J. Mol. Boil. 425 (2013) 1683–1696.

[21] E.R. Waters, The evolution, function, structure, and expression of the plant sHSPs, J. Exp. Bot. 64 (2013) 391–403.

[22] W. Sun, C. Bernard, B.V. Cotte, M.V. Montagu, N. Verbruggen, At-HSP17.6A, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression, Plant J. 27 (2001) 407–415.

[23] U. Harndahl, R.B. Hall, K.W. Osteryoung, E. Vierling, J.F. Bornman, C. Sundby, The chloroplast small heat shock protein undergoes oxidation-dependent conformational changes and may protect plants from oxidative stress, Cell Stress Chaperones 4 (1999) 129–138.

[24] C.H. Yeh, P.F.L. Chang, K.W. Yeh, W.C. Lin, Y.M. Chen, C.Y. Lin, Expression of a gene encoding a 16.9-kDa heat-shock protein, Oshsp16.9, in *Escherichia coli* enhances thermotolerance, Proc. Nat. Acad. Sci. 94 (1997) 10967–10972.

[25] Y. Murakami, M. Tsuyama, Y. Kobayashi, H. Kodama, K. Iba, Trienoic fatty acids and plant tolerance of high temperature, Science 287 (2000) 476–479.

[26] Y. Sato, S. Yokoya, Enhanced tolerance to drought stress in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7, Plant Cell Rep. 27 (2008) 329–334.

[27] Na-Hyun Song, Yeh-Jin Ahn, DcHsp17.7, a small heat shock protein in carrot, is tissue-specifically expressed under salt stress and confers tolerance to salinity, New Biotechnol. 28 (6) (2011) 698–704, https://doi.org/10.1016/j.nbt.2011.04.002.

[28] C. Jiang, J. Xu, H. Zhang, X. Zhang, J. Shi, M. Li, F. Ming, 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 Environ. 32 (2009) 1046–1059.

[29] E.R. Waters, B.D. Aevermann, Z. Sanders-Reed, Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns, Cell Stress Chaperones 13 (2008) 127–142.

[30] N.K. Sarkar, Y.K. Kim, A. Grover, Rice sHsp genes: genomic organization and expression profiling under stress and development, BMC Genomics 10 (1) (2009) 393.

[31] V.S. Lopes-Caitar, M.C. de Carvalho, L.M. Darben, M.K. Kuwahara, A.L. Nepomuceno, W.P. Dias, et al., Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses, BMC Genomics 14 (2013) 577.

[32] P.S. Reddy, P.B. Kavi Kishor, C. Seiler, M. Kuhlmann, L. Eschen-Lippold, et al., Unravelling 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 9 (3) (2014) e89125.

[33] J. Li, X. Liu, Genome-wide identification and expression profile analysis of the *Hsp20* gene family in Barley (*Hordeum vulgare* L.), PeerJ 7 (2019) e6832.

[34] M. Guo, J.H. Liu, J.P. Lu, Y.F. Zhai, H. Wang, Z.H. Gong, S.B. Wang, M.H. Lu, Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress, Front. Plant Sci. 6 (2015) 806.

[35] R.K. Singh, J. Jaishankar, M. Muthamilarasan, S. Shweta, A. Dangi, M. Prasad, Genome-wide analysis of heat shock proteins in C4 model, foxtail millet identifies potential candidates for crop improvement under abiotic stress, Sci. Rep. 6 (2016) 32641.

[36] J. Yu, Y. Cheng, K. Feng, M. Ruan, Q. Ye, R. Wang, Z. Li, G. Zhou, Z. Yao, Y. Yang, H. Wan, Genome-wide identification and expression profiling of tomato Hsp20 gene family in response to biotic and abiotic stresses, Front. Plant Sci. 7 (2016) 1215.

[37] H. Yan, A. Zhang, J. Chen, X. He, B. Xu, G. Xie, Z. Miao, X. Zhang, L. Huang, Genome-wide analysis of the PvHsp20 family in switchgrass: Motif, genomic organization, and identification of stress or developmental-related Hsp20s, Front. Plant Sci. 8 (2017) 1024.

[38] S.K. Muthusamy, M. Dalal, V. Chinnusamy, K.C. Bansal, Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in bread wheat, J. Plant Physiol. 211 (2017) 100–113.

[39] Peng Zhao, Dongdong Wang, Ruoqiu Wang, Nana Kong, Chao Zhang, Chenghui Yang, Wu. Wentao, Haoli Ma, Qin Chen, Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress, BMC Genomics 19 (2018) 61.

[40] M.K. Monaco, J. Stein, S. Naithani, S. Wei, P. Dharmawardhana, S. Kumari, et al., Gramene 2013: comparative plant genomics resources, Nucleic Acids Res. 42 (D1) (2014) D1193–D1199.

[41] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, J. Mol. Biol. 268 (1997) 78–94.

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

[43] A.Y. Guo, Q.H. Zhu, X. Chen, J.C. Luo, Y.I. Chuan, GSDS: a gene structure display server, Yi Chuan 29 (2007) 1023–1026.

[44] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy Server, in: John M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, 2005, pp. 571–607.

[45] N. Blom, T. Sicheritz-Ponten, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post translational glycosylation and phosphorylation of proteins from the amino acid sequence, Proteomics 4 (2004) 1633–1649.

[46] S. Moller, M.D.R. Croning, R. Apweller, Evaluation of methods for the prediction of membrane spanning regions, Bioinformatics 17 (2001) 646–653.

[47] P. Horton, P. Keun-Joon, O. Takeshi, F. Naoya, C.J.H. Harada, C. Adams, N. Kenta, WoLF PSORT: protein localization predictor, Nucleic Acids Res. 35 (2007) W585–W587.

[48] T.L. Bailey, N. Williams, C. Misleh, W.W. Li, MEME: discovering and analyzing DNA and protein sequence motifs, Nucleic Acids Res. 34 (suppl. 2) (2006) W369–W373.

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

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

[51] X. Dai, P.X. Zhao, psRNATarget: a plant small RNA target analysis server, Nucleic Acids Res. 39 (2011) 155–159.

[52] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, STRING v10: protein–protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (D1) (2014) D447–52.

[53] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.

[54] M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments, Nucleic Acids Res. 34 (2006) W609–W612.
[55] T. Tian, Q. You, L. Zhnag, X. Yi, H. Yan, W. Xu, Z. Su, SorghumFDB: sorghum functional genomics database with multidimensional network analysis, Database (Oxford) 2016 (2016), baw099.
[56] T. Hruz, O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, P. Zimmermann, Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes, Adv. Bioinformatics 2008 (2008) 420747.
[57] A.K. Grennan, Genevestigator. Facilitating web-based gene expression analysis, Plant Physiol. 141 (2006) 1164–1166.
[58] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.
[59] P.S. Reddy, D.S. Reddy, K. Sivasakthi, P. Bhatnagar-Mathur, V. Vadez, K.K. Sharma, Evaluation of sorghum [*Sorghum bicolor* (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization, Front. Plant Sci. 7 (2016) 529.
[60] W.M. Pfaffl, W.G. Horganl, D. Leo, Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR, Nucleic Acids Res. 3 (2002) 900.
[61] T. Kriehuber, T. Rattei, T. Weinmaier, A. Bepperling, M. Haslbeck, J. Buchner, Independent evolution of the core domain and its flanking sequences in small heat shock proteins, FASEB J. 24 (2010) 3633–3642.
[62] G. Xu, C. Guo, H. Shan, H. Kong, Divergence of duplicate genes in exon-intron structure, PNAS 109 (2012) 1187–1192.
[63] J.S. Mattick, M.J. Gagen, The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms, Mol. Biol. Evol. 18 (2001) 1611–1630.
[64] C.I. Castillo-Davis, S.L. Mekhedov, D.L. Hartl, E.V. Koonin, F.A. Kondrashov, Selection for short introns in highly expressed genes, Nat. Genet. 31 (2002) 415–418.
[65] X.Y. Ren, O. Vorst, M.W. Fiers, W.J. Stiekema, J.P. Nap, In plants, highly expressed genes are the least compact, Trends Genet. 22 (2006) 528–532.
[66] D.C. Jeffares, C.J. Penkett, J. Bahler, Rapidly regulated genes are intron poor, Trends Genet. 24 (2008) 375–378.
[67] P.K. Rao, B.A. Roxas, Q. Li, Determination of global protein turnover in stressed mycobacterium cells using hybrid-linear ion trap fourier transform mass spectrometry, Anal. Chem. 80 (2008) 396–406.
[68] S. Chowdhury, A. Basu, S. Kundu, Overexpression of a new osmotin-like protein gene 628 (SindOLP) confers tolerance against biotic and abiotic stresses in sesame, Front. Plant Sci. 8 (2017) 410.
[69] P. Yakovchuk, E. Protozanova, M.D. Frank-Kamenetskii, Base-stacking and base-pairing contributions into thermal stability of the DNA double helix, Nucleic Acids Res. 34 (2) (2006) 564–574.
[70] M. Siddique, S. Gernhard, P. von Koskull-Doring, E. Vierling, K.D. Scharf, The plant sHSP superfamily: five new members in Arabidopsis thaliana with unexpected properties, Cell Stress Chaperones 13 (2008) 183–197.
[71] E.R. Waters, E. Vierling, The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses, Mol. Biol. Evol. 16 (1999) 127–139.
[72] E.R. Waters, E. Vierling, Chloroplast small heat shock proteins: evidence for atypical evolution of an organelle-localized protein, PNAS 96 (1999) 14394–14399.
[73] M. Nei, A.P. Rooney, Concerted and birth-and-death evolution of multigene families, Annu. Rev. Genet. 39 (2005) 121–152.
[74] L.E. Flagel, J.F. Wendel, Gene duplication and evolutionary novelty in plants, New Phytol. 183 (2009) 557–564.
[75] R.C. Moore, M.D. Purugganan, The early stages of duplicate gene evolution, PNAS 100 (2003) 15682–15687.
[76] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[77] C. Lu et al., Elucidation of the small RNA component of the transcriptome, Science 309 (2005) 1567–1569.
[78] N. Lauter, A. Kampani, S. Carlson, M. Goebel, S.P. Moose, microRNA172 down-regulates glossy15 to promote vegetative phase change in maize, PNAS 102 (2005) 9412–9417.
[79] M.W. Jones-Rhoades, D.P. Bartel, B. Bartel, MicroRNAS and their regulatory roles in plants, Annu. Rev. Plant Biol. 57 (2006) 19–53.
[80] S. Barik, A. Kumar, S.S. Das, S. Yadav, V. Gautam, A. Singh, S. Singh, A.K. Sarkar, Co- evolution pattern and functional conservation or divergence of miR167s and their targets across diverse plant species, Sci. Rep. 5 (2015) 14611.
[81] L. Zhang, Y. Zheng, G. Jagadeeswaran, Y. Li, K. Gowdu, R. Sunkar, Identification and temporal expression analysis of conserved and novel microRNAs in *Sorghum*, Genomics 98 (2011) 460–468.
[82] H.C. Park, M.L. Kim, Y.H. Kang, J.M. Jeon, J.H. Yoo, M.C. Kim, et al., Pathogen- and NaCl-induced expression of the *SCaM-4* promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor, Plant Physiol. 135 (2004) 2150–2161.
[83] C.H. Li, Y. Sun, The receptor-like kinase *SIT1* mediates salt sensitivity by activating MAPK3/6 and regulating ethylene homeostasis in rice, Plant Cell 26 (2014) 2538–2553.
[84] C. Chen, Z. Chen, Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced *Arabidopsis* transcription factor, Plant Physiol. 129 (2002) 706–716.
[85] T. Nishiuchi, H. Shinshi, K. Suzuki, Rapid and transient activation of transcription of the *ERF3* gene by wounding in tobacco leaves, J. Biol. Chem. 279 (2004) 55355–55361.
[86] N. Bate, D. Twell, Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements, Plant Mol. Biol. 37 (1998) 859–869.
[87] G. Plesch, T. Ehrhardt, B. Mueller-Roeber, Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression, Plant J. 28 (2001) 455–464.
[88] S. Yanagisawa, Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize, Plant J. 21 (2000) 281–288.
[89] U. Gowik, J. Burscheidt, M. Akyildiz, U. Schlue, M. Koczor, M. Streubel, P. Westhoff, cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveriatrinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene, Plant Cell 16 (2004) 1077–1090.
[90] E.J. Ross, J.M. Stone, C.G. Elowsky, R. Arredondo-Peter, R.V. Klucas, G. Sarath, Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor, ARR1, J. Exp. Bot. 55 (2004) 1721–1731.
[91] P.M. Gilmartin, L. Sarokin, J. Memelink, N.H. Chua, Molecular light switches for plant genes, Plant Cell 2 (1990) 369–378.
[92] D. He, P. Yang, Proteomics of rice seed germination, Front. Plant Sci. 4 (2013) 246.
[93] Y. Ouyang, J. Chen, W. Xie, L. Wang, Q. Zhang, Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice, Plant Mol. Biol. 70 (2009) 341–357.
[94] H. Chauhan, N. Khurana, A. Nijhavan, J.P. Khurana, P. Khurana, The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress, Plant Cell Environ. 35 (2012) 1912–1931.
[95] Y. Sun, T.H. MacRae, Small heat shock proteins: molecular structure and chaperone function, Cell. Mol. Life Sci. 62 (2005) 2460–2476.
[96] K. Sanmiya, K. Suzuki, Y. Egawa, M. Shono, Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants, FEBS Lett. 557 (2004) 265–268.
[97] Y. Xue, R. Peng, A. Xiong, X. Li, D. Zha, Q. Yao, Over-expression of heat shock protein gene hsp26 in *Arabidopsis thaliana* enhances heat tolerance, Biol. Plant. 54 (2010) 105–111.
[98] J. Li, J. Zhang, H. Jia, Y. Li, X. Xu, L. Wang, M. Lu, The *Populus trichocarpa PtHSP17.8* involved in heat and salt stress tolerances, Plant Cell Rep. 35 (2016) 1587–1599.
[99] L. Sun, Y. Liu, X. Kong, D. Zhang, J. Pan, Y. Zhou, X. Yang, *ZmHSP16. 9*, a cytosolic class I small heat shock protein in maize (Zea mays), confers heat tolerance in transgenic tobacco, Plant Cell Rep. 31 (2012) 1473–1484.