GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genome-wide identification, classification, and analysis of heat shock transcription factor family in Chinese cabbage (*Brassica rapa pekinensis*)

X.Y. Huang[1,2], P. Tao[1], B.Y. Li[1], W.H. Wang[1], Z.C. Yue[1], J.L. Lei[1] and X.M. Zhong[1]

[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China
[2]College of Horticulture, Nanjing Agricultural University, Nanjing, China

Corresponding author: X.M. Zhong
E-mail: zxmlly@hotmail.com

Genet. Mol. Res. 14 (1): 2189-2204 (2015)
Received June 26, 2014
Accepted November 7, 2014
Published March 27, 2015
DOI http://dx.doi.org/10.4238/2015.March.27.5

ABSTRACT. Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) is one of the most important vegetable crops grown worldwide, and various methods exist for selection, propagation, and cultivation. The entire Chinese cabbage genome has been sequenced, and the heat shock transcription factor family (Hsfs) has been found to play a central role in plant growth and development and in the response to biotic and abiotic stress conditions, particularly in acquired thermotolerance. We analyzed heat tolerance mechanisms in Chinese cabbage. In this study, 30 Hsfs were identified from the Chinese cabbage genome database. The classification, phylogenetic reconstruction, chromosome distribution, conserved motifs, expression analysis, and interaction networks of the Hsfs were predicted and analyzed. Thirty BrHsfs were classified into 3 major classes (class A, B, and C) according to their structural characteristics and phylogenetic comparisons, and class A was further

subdivided into 8 subclasses. Distribution mapping results showed that Hsf genes were located on 10 Chinese cabbage chromosomes. The expression profile indicated that Hsfs play differential roles in 5 organs in Chinese cabbage, and likely participate in the development of underground parts and regulation of reproductive growth. An orthologous gene interaction network was constructed, and included MBF1C, ROF1, TBP2, CDC2, and HSP70 5 genes, which are closely related to heat stress. Our results contribute to the understanding of the complexity of Hsfs in Chinese cabbage and provide a basis for further functional gene research.

Key words: Chinese cabbage; Gene expression; Hsf; Interaction network; Phylogenetic analysis

INTRODUCTION

Heat stress is a major abiotic stress that limits plant production, particularly during the summer months in warm and temperate climates, which may have a greater impact resulting from global climate change (Wardlaw and Willenbrink, 1994; Ahuja et al., 2010). Heat shock proteins (Hsps) and heat stress transcription factors (Hsfs) are involved in the cellular response to various forms of heat shock and other stress (Kotak et al., 2007a). Hsps serve as molecular chaperones regulating protein folding, localization, accumulation, and degradation in both plant and animal species (Feder and Hofmann, 1999), and are considered to play a generalized role in tolerance to multiple environment stress apart from heat stress. However, Hsfs are transcriptional activators of heat shock genes and play a central role in regulating Hsp expression (Nover et al., 2001).

Hsf is typically composed of a DNA-binding domain (DBD) located near the N-terminus, an oligomerization domain (or HR-A/B) connected to the DBD, a nuclear localization signal (NLS), a nuclear export signal (NES), and activator motifs (AHA motifs) located near the C-terminal domains (Nover et al., 2001). DBD is the most conserved component of Hsfs, consisting of an antiparallel 4-stranded β-sheet (β1, β2, β3, β4) packed against a bundle of 3 α-helices (H1, H2, H3) (Damberger et al., 1994). The hydrophobic core of this domain ensures the specific recognition of heat stress promoter elements (Littlefield and Nelson, 1999; Sakurai and Enoki, 2010). Based on differences in their oligomerization domains, plant Hsf protein families fall into 3 classes (A, B, and C). All class A and class C Hsfs possess an extended HR-A/B region resulting from an insertion of 21 (class A) or 7 (class C) amino acid residues between the A and B regions, while the HR-A/B regions of class B Hsfs are compact without insertion. Furthermore, a cluster of basic amino acid residues at the C-terminal from OD domain act as an NLS, in which class B is joined with the highly conserved repressor motif LFGV. Leucine-rich NES at the C-terminal end of many plants is required to maintain the balance of import and export in cooperation with the NLS. The AHA motif is characteristic of class A Hsfs as transcription activators, which are rich in aromatic (W, Y, F), hydrophobioc (L, I, V), and acidic amino acid residues (E, D). In contrast, class B and class C Hsfs have no activator function because they possess no AHA motif (Kotak et al., 2004).

To date, a variety of heat stress transcription factors have been successfully identified and investigated in some plants, including tomato (Scharf et al., 1990), *Arabidopsis* (Nover

et al., 2001), rice (Guo et al., 2008), maize (Lin et al., 2011), *Malus domestica* (Giorno et al., 2012), soybean (*Glycine max*) (Chung et al., 2012), *Medicago truncatula*, and polar (*Populus tricocarpa*) (Wang et al., 2012). For example, *Arabidopsis*, which served as the prototype for the Hsf family, contains a set of 21 Hsf-encoding genes with 15 members belonging to class A, 5 members to class B, and 1 to class C, which are the smallest families observed thus far. The maximum number of 52 Hsf genes was identified in soybean. Tomato HsfA1a appears to have an unique function as master regulator of acquired thermotolerance, and cannot be replaced by any other Hsfs (Mishra et al., 2002), while tomato HsfB1 acts as synergistic co-activator of HsfA1a (Czarnecka-Verner et al., 2000). HsfA2, which is structurally and functionally similar to HsfA1, is one of the most strongly induced proteins in tomato, accumulating to high levels when plants are exposed to heat stress (Scharf et al., 1998), but it is only expressed in stressed plants. Notably, the expression of HsfA2 together with chaperones Hsp90 and Hsp70 was found to have an integral function in anther development in tomato, indicating that preformed chaperones may be important for protecting maturing and germinating pollen from heat damage (Giorno et al., 2010). In contrast to class A Hsfs, a considerable number of Hsfs assigned to class B and class C have no evident function as transcription activators on their own (Kotak et al., 2004), and class B was shown to have repressor functions (Czarnecka-Verner et al., 2004). Furthermore, HsfA5 acts as a specific repressor of the antiapoptotic HsfA4 in *Arabidopsis* and tomato (Baniwal et al., 2007). Additionally, HsfA3 was found to be involved in drought stress signaling in *Arabidopsis* (Sakuma et al., 2006). HsfA9 is associated with both embryogenesis and seed maturation in sunflower and *Arabidopsis* (Kotak et al., 2007b).

Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) originated in China and is an important vegetable crop in the extratropical, subtropical, and tropical belts worldwide. The Chinese cabbage genome has been sequenced and assembled (Wang et al., 2011), providing the basis for determining the Chinese cabbage Hsf family and its evolutionary history, as well as adaptations to heat and chemical stress response mechanisms at the molecular level. In this study, we identified and characterized Hsf family members in the Chinese cabbage genome and analyzed the phylogenetic relationships and orthologous genes between the Chinese cabbage and *Arabidopsis*. Moreover, we constructed Hsf interaction networks and analyzed their expression patterns. Our results provide a foundation to improve the understanding of the functional structure and genomic organization of the Hsf family in Chinese cabbage and will be useful for gene cloning and functional studies.

MATERIAL AND METHODS

Identification and classification of Hsfs in Chinese cabbage

The Chinese cabbage genome sequence is known and filtered protein and CDS sequences are available. Whole genome proteins of 2 species were downloaded, including Chinese cabbage (http://brassicadb.org/brad/geneFamily.php) and *Arabidopsis* (http://datf.cbi.pku.edu.cn/). The isoelectric point (PI) and molecular weight (MW) were computed using the Expasy tools (http://web.expasy.org/compute_pi/). The following strategy was used to isolate Hsfs from the whole genome of Chinese cabbage. First, the key word "heat shock transcription factor" was used to search directly in NCBI. Subsequently, annotated Hsf members in the Chinese cabbage genome database were selected. Third, the amino acid sequences of *Arabidopsis* Hsfs were used as standard sequences to isolate all possible homologs in Chinese cabbage

using BLASTP searches (P value = 0.001), and repetitive Hsfs were removed manually. All candidate Hsf genes meeting these standards were detected using SMART and Pfam (http://pfam.janelia.org/) to eliminate any sequences not containing the signature DBD domain of Hsfs (Bateman et al., 2004). As a final quality check, the remaining sequences were evaluated using the MARCOIL programs to identify coiled-coil structures. Sequences not containing a coiled-coil structure were removed. Numbers of BrHsfs were assigned randomly. Based on these results, we obtained chromosome locations of these genes. The chromosome location image of Hsf genes was generated using the Mapdraw V2.1 software (Liu and Meng, 2003).

Analysis of phylogenetic relationships

To understand the evolutionary relationships between the Chinese cabbage Hsf proteins and the variations in Hsf sequences, AtHsfs and BrHsfs were selected for phylogenetic tree analysis using MEGA (version 5.0) (Tamura et al., 2011). Initially, the retrieved Chinese cabbage and *Arabidopsis* Hsfs nucleotide sequences were translated into amino acid sequences using BioXM2.6 in the Fasta format, and protein sequences were then aligned using ClustalX (version 1.83) (Chenna et al., 2003). MEGA analysis was conducted after these steps. The neighbor joining (NJ) method was performed with Poisson correction and the pairwise deletion option. For statistical reliability, bootstrap analysis was conducted with 1000 replicates to assess statistical support for each mode.

Multiple sequence alignment and domain prediction

To identify signature domains, ClustalX was used to align amino acid sequences of Hsf proteins. The domain analysis programs MARCOIL (Delorenzi and Speed, 2002), SMART (Letunic et al., 2009), PredictNLS, and NetNES1.1 (La Cour et al., 2004) were used to check DBD domains and coiled-coil structures and NLS and NES domains. Conserved motifs were analyzed using MEME tools (http://meme.nbcr.net/meme/cgi-bin/meme.cgi), with the parameters were set as follows: minimum width of 6, maximum width of 50, and maximum number of motifs to identify was 20; default values were used for other parameters (Bailey et al., 2006).

BrHsfs expression and interaction networks

Chinese cabbage tissue expression information from raw RNA-seq data were downloaded from NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession No. GSE43245 (Tong et al., 2013). Expression profile cluster analysis of the Chinese cabbage Hsf family proteins was constructed using the MEV Software (Saeed et al., 2003). The interaction network associated with *Arabidopsis* orthologs of Hsf genes in Chinese cabbage was constructed using the STRING software (http://string-db.org/) (Szklarczyk et al., 2011).

RESULTS

Identification and physical location of Hsf genes in the chinese cabbage

All candidate BrHsfs were surveyed, and those containing incomplete sequences for

the DBD domain and the remaining functional domains were removed. Thirty genes encoding for BrHsfs proteins were identified (Table 1). All non-redundant BrHsf genes were distributed on every chromosome of Chinese cabbage. According to multiple sequence alignment results of the DBD and HR-A/B region, 19 genes were identified as class A, 9 genes were class B, and 2 genes were class C. There were more BrHsfs of class A than those of classes B and C. The length of BrHsfs proteins ranged from 239 (Bra000235) to 487 (Bra023258) amino acids (aa), the PIs of the BrHsfs proteins were between 4.66 (Bra023258, Bra040968) and 9.16 (Bra014054), with a distribution from acidic to alkaline; the MWs of these proteins were between 27.93 kDa (Bra000235) and 54.49 kDa (Bra023258).

Table 1. List of Hsfs genes in the Chinese cabbage genome.

Gene name	Chromosome location	Size (aa)	PI	MW(kDa)
Bra011735	A01:827942..828942	285	6.05	31.2
Bra040179	A01:4514352..4515753	432	5.05	47.98
Bra023800	A01:20020642..20022478	398	5.13	45.88
Bra021381	A01:25471297..25473657	454	5.16	50.24
Bra023584	A02:4067772..4069228	442	4.81	49.1
Bra029292	A02:26078013..26078880	255	7.87	28.87
Bra032023	A02:26626811..26629251	456	5.07	50.65
Bra000557	A03:11762357..11763740	348	4.99	39.11
Bra000235	A03:10071648..10072839	239	5.17	27.93
Bra000749	A03:12856733..12857764	316	4.86	34.39
Bra001071	A03:14600881..14602812	453	5.54	50.14
Bra001885	A03:19137642..19139463	391	4.9	44.56
Bra013253	A03:19602332..19603177	281	6.05	32.09
Bra012829	A03:22040320..22041629	248	6.73	28.8
Bra017800	A03:30557774..30558682	271	6.04	30.12
Bra032752	A04:4884874..4886387	477	5.93	53
Bra033913	A05:15136897..15138822	407	4.9	46.79
Bra010049	A06:18874693..18875694	302	6.56	33.89
Bra015050	A07:3955634..3956652	309	5.7	35.24
Bra004272	A07:17735611..17736956	384	4.82	43.68
Bra014054	A08:4045288..4046383	335	9.16	38.17
Bra035507	A08:7946929..7948855	427	5.05	47.17
Bra017595	A09:16771187..16772272	340	5.75	39.04
Bra023258	A09:20143853..20146639	487	4.66	54.49
Bra007739	A09:30668787..30670395	285	5.79	33.3
Bra040634	A10:4910958..4912016	326	8.41	37.07
Bra008593	A10:12961020..12963244	483	4.95	53.23
Bra009515	A10:15837127..15838793	363	5.03	41.16
Bra040968	Scaffold000344:11295-12467	357	4.66	38
Bra035993	Scaffold000111:359206-360216	281	5.32	33.22

MW = molecular weight; PI = Isoelectric point.

Except for 2 genes (*Bra040968*, *Bra035993*) on the scaffold that could not be mapped to a specific chromosome, the remaining 28 Hsf genes were distributed in every chromosome of the Chinese cabbage genome (Figure 1). The number of Hsf genes on each chromosome varied widely. The largest number of Hsf genes was detected on chromosome A03 (8 Hsf genes), while the lowest number was on chromosomes A04, A05, and A06 (1 Hsf gene each). Chromosomes A02, A09, and A10 had the same number of Hsf genes (3 genes), as well as chromosomes A07 and A08 (2 genes). Four Hsf genes were located on chromosome A01 (Figure 1).

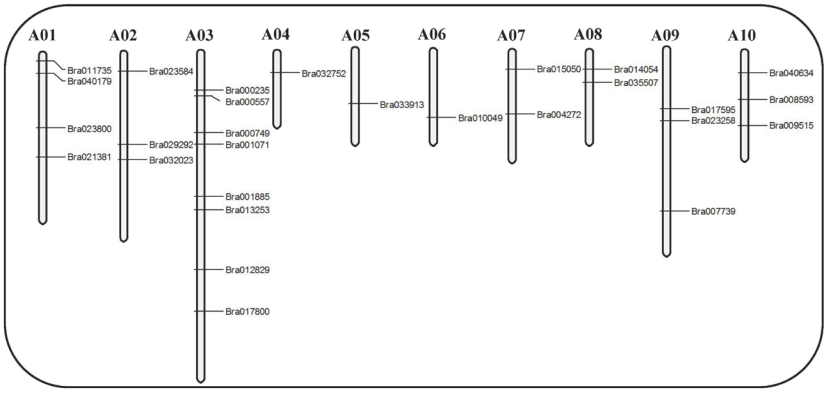

Figure 1. Distribution of 30 BrHsfs on the 10 Chinese cabbage chromosomes. Two genes on the scaffold (*Bra040968*, *Bra035993*) could not be anchored to a specific chromosome.

Analysis of conserved domains of Hsf proteins in Chinese cabbage

The modular structure of the Hsfs has been examined in some model plants (Nover et al., 2001). The details of various functional motifs/domains, such as DBD, HR-A/B, NLS, NES, and AHA motifs are shown in Table 2. Multiple alignment analysis clearly showed that

Table 2. Functional motifs of Chinese cabbage Hsfs.

Gene name	Group	DBD	OD	NLS	NES	AHA1	AHA2
Bra040179	A1a	36-129	171-221	(246)SKKRR	(373) LPQDLDI	(387)DIFEEYLAQS	(412)GGHVDKLIEEL
Bra008593	A1b	25-118	151-194	(220)NKKRR	(354) LGGSLEM	(386)DPFWEQFFAD	(428)VNHITEQTGF
Bra023584	A1b	24-117	149-199	(223)NKKRR	(428) LTEQMEL	(381)DSFWEQFFAD	(424)EMNHLTEQMEL
Bra023258	A1d	36-129	167-217	(243)SKKRR	(471) LTQQMGL	(446)PDDMEVTPVD	(457)TTKDGETEQKQM
Bra035507	A1d	41-106	183-233	(257)KDSA	(407) LAVPLAF	(380)GNFMEPDMED	
Bra032023	A1e	15-108	141-193	(218)NKKRR	(370) LAGSMGL	(390)DPLWEQFFGE	(439)MNHLTEQMGF
Bra021381	A1e	22-115	150-200	(225)NKKRR	(371) LDGSMML	(388)DSLWEQFFGE	(435)QQMNHLTEQMGL
Bra001071	A1e	24-117	152-202	(226) NKKRR	(416) IMEQLGL	(387)DSLWEQFFGE	(434)QQMNHLTEQMGL
Bra000557	A2	43-136	173-222	(248) GRKRR	(327) LDWGSEDL	(269)DHMEFDRMKD	(289)DDEASEDEQCLE
Bra009515	A3	11-104	135-185	(208) KARKK	(346)ETGFNW	(333)DVCWEQFAAG	
Bra017595	A4c	11-104	135-185	(199) RRKRK	(224) LESSLTF	(289)DVFWEQCLTE	(298)ENPGSIEQQEV
Bra032752	A5	24-117	144-194	(218) NKKRR	(326) LLLNDKT	(425)DVFWEQFLTE	(449)RENPCEEQEEK
Bra035993	A6a	17-110	158-180	(189) KKRK	(269) LSDEMCI	(242)IASMEDQRQD	
Bra033913	A6b	58-151	189-238	(263) SKKRQR	(318) LSGFEM	(370)EGFWEDMLNE	(389)EENVDVLIEQLGY
Bra023800	A6b	55-148	185-236	(251) KRK7KKR	(314) LDRLAM	(360)EGFWEDLLNE	(380)NVDVLIEQLGY
Bra001885	A6b	58-151	189-238	(251) KDK9KKR	(314) LDGLAM	(354)EDFLEGLFKE	(373)GENVDVLIEQLGY
Bra012829	A7a	27-120	135-185	(197) KQRDMRVK	(222) LEALAL	(200)DMRVKELEDE	
Bra007739	A7b	24-134	153-203	(214) EQQRKE	(239) LEALAL	(263)DGFWEELLMN	
Bra004272	A8	18-112	148-197	(215) RKAEGGGAK	(336) LDKSLAL	(331)GGRMELDKSL	
Bra000749	B1	35-128	191-221	(228) VKPLD	(258) LFGVSI		
Bra040968	B1	45-139	216-252	(260) PEGRAL	(298) LFGVSI		
Bra010049	B2a	21-114	174-210	(226) HERMK	(255) FGVSIGL		
Bra029292	B2a	13-106	155-192	(220) LKRTR	(225)EGVHVKTV		
Bra000235	B2b	38-131	183-212	(207)LVERYK	(231) LKLFGVKL		
Bra011735	B3	7-101	156-187	(245) KGERKKRGR	(268) IKNVDF		
Bra017800	B3	9-102	145-176	(234) RKKRGRDEK	(254) IKNVDF		
Bra014054	B4	32-125	202-231	(290) RKTK9KKR	(294) LFGVSL		
Bra040634	B4	26-119	195-224	(293)SKKRS	(311) LDKSDL		
Bra015050	C1	14-107	127-163	(184)KKKRR	(255) LTSTLSL		
Bra013253	C1	0-84	104-140	(161)KKKRR	(230) LTSTLSL		

Numbers in brackets indicate the position of the first amino present in the putative nuclear location signal (NLS), nuclear export signal (NES), and activator (AHA) motifs in the C-terminal domains.

the highly structured DBD domain was located in the N-terminal region (Figure 2), which was the most conserved section of Chinese cabbage Hsfs, consisting of a 3-helical bundle (α1, α2, and α3) and a 4-stranded antiparallel β-sheet. The length of the DBD domain was quite variable as a result of an insertion/deletion event with the longest motif (111 aa) in Bra007739, the smallest (84 aa) in Bra013253, while most of the other motifs were 94 aa. The presence of the coiled-coil structure that is characteristic of all Hsf proteins was predicted by using the MARCOIL tool (Table 2). Putative HR-A/B regions were consistently characterized by the predicted coiled-coil structure. Based on the peculiarities of their HR-A/B regions, 3 classes of Hsfs were identified in plants. Class A and class C Hsfs possess extended HR-A/B regions resulting from an insertion of 21 (class A) and 7 (class C) amino acid residues between the A and B regions (Figure 3). PredictNLS and NetNES were used to gather information regarding the existence of potential NLS and NES and their locations in the Hsf protein sequences. NLS of Hsfs was formed by monopartite or bippartite clusters of basic amino acid residues at the C-terminal end of HR-A/B regions. NES served as a part of a type-specific signature region at the C-terminus of class A Hsfs, which are essential for maintaining the balance between nuclear import and export. Additional sequence comparisons revealed AHA motifs in the center of the C-terminal activation domains as expected for type-A Hsfs. In contrast, these domains were not identified in the B and C type Hsfs.

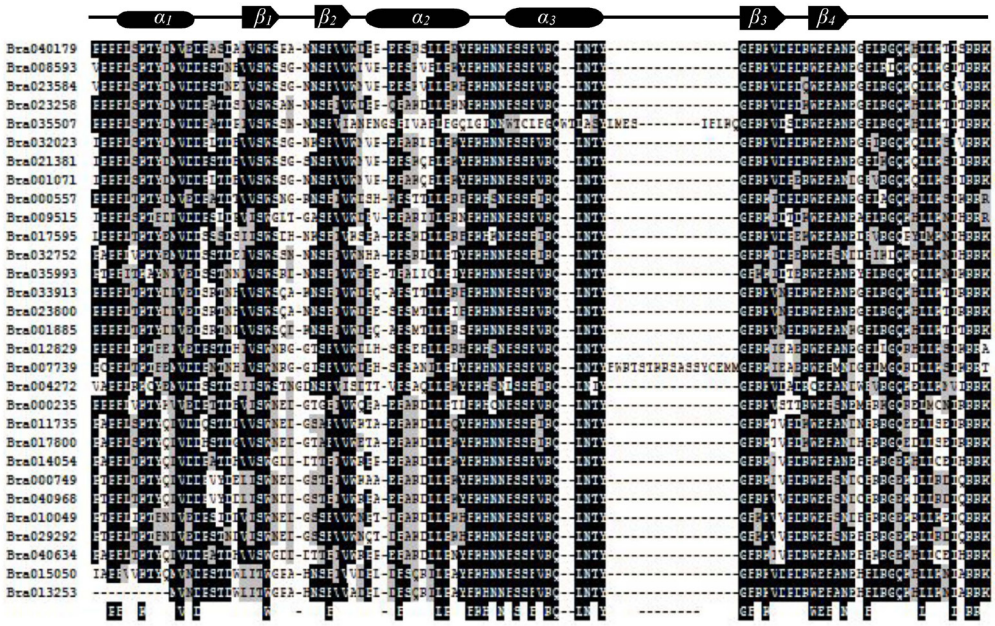

Figure 2. Multiple sequence alignment of the DBD domains of the Hsf protein family in Chinese cabbage. Multiple alignment results clearly revealed highly conserved DBD domains among Chinese cabbage Hsf genes. The secondary structure elements of DBD (α1-β1-β2-α2-α3-β3-β4) are shown above the alignment. Cylindrical tubes represent α-helices and black arrows represent β-sheets. Consensus amino acids showing complete conservation are shown at the bottom of the alignment.

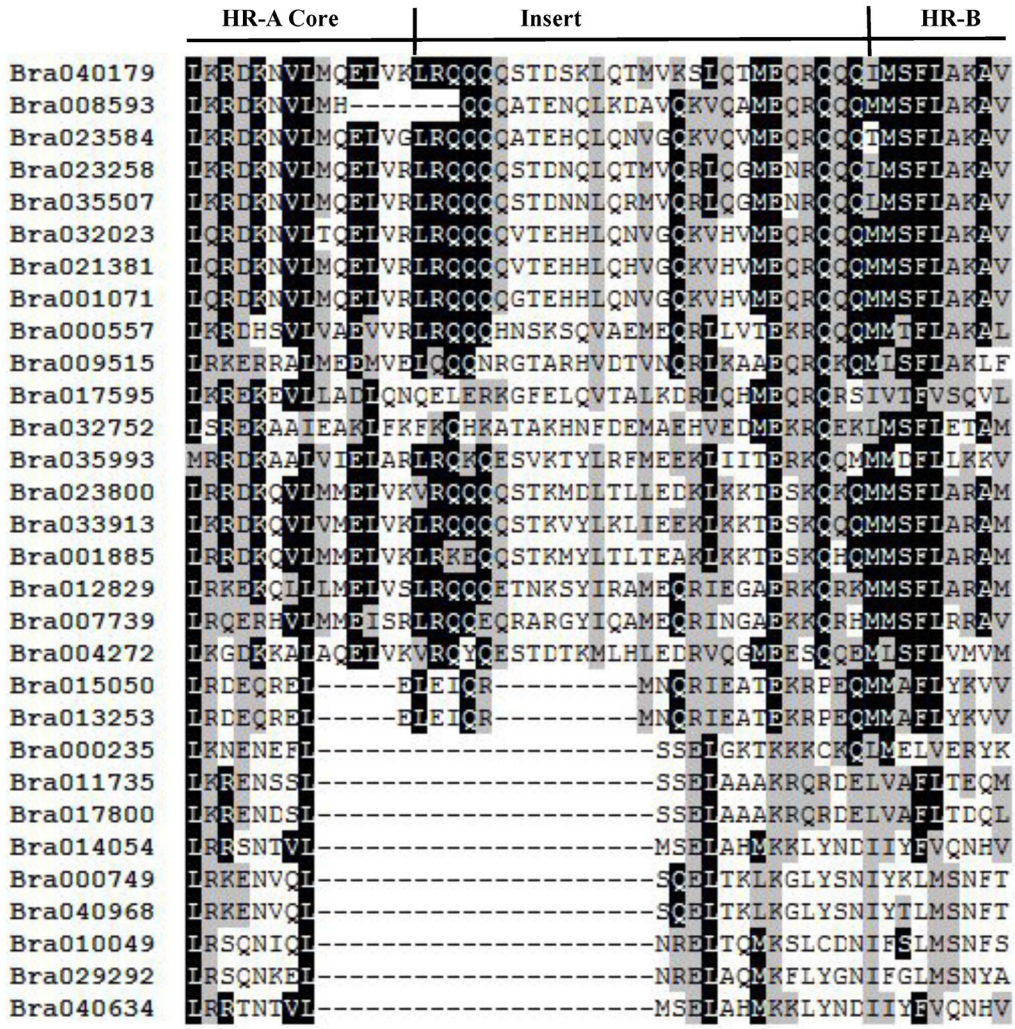

Figure 3. Multiple sequence alignment of the HR-A/B regions of the Hsf protein family in Chinese cabbage. The scheme at the top depicts the locations and boundaries of the HR-A core, insert, and HR-B regions within the HR-A/B regions.

Phylogenetic analysis of Hsf proteins of chinese cabbage

To analyze the relationships between Hsf family members, a phylogenetic tree of 30 BrHsfs and 21 *Arabidopsis* Hsfs (AtHsfs) was generated using amino acid sequences with a bootstrap of 1000 replicates to ensure statistical reliability (Figure 4). Hsfs of Chinese cabbage and *Arabidopsis* were grouped into 3 different classes corresponding to the main Hsf classes A, B, and C. In this study, class A was further divided into 9 subclasses according to the phylogenetic relationships designated as A1-A9. Eight (A1-A8) of these groups comprised the Chinese cabbage Hsf sequences, while A9 (At5g54070) appeared as a single branch of AtHsfs.

In addition, C-type Hsfs constituted 1 distinct class, which clustered more closely with class A. Moreover, 12 groups of orthologous genes (A1a, A1b, A2a, A3a, A4c, A5a, A6a, A7a, A7b, A8a, B2b, and B3a) and 7 groups of paraologous genes (A1a, A1d, A6b, B1a, B2a, B4a, and C) were identified in the tree.

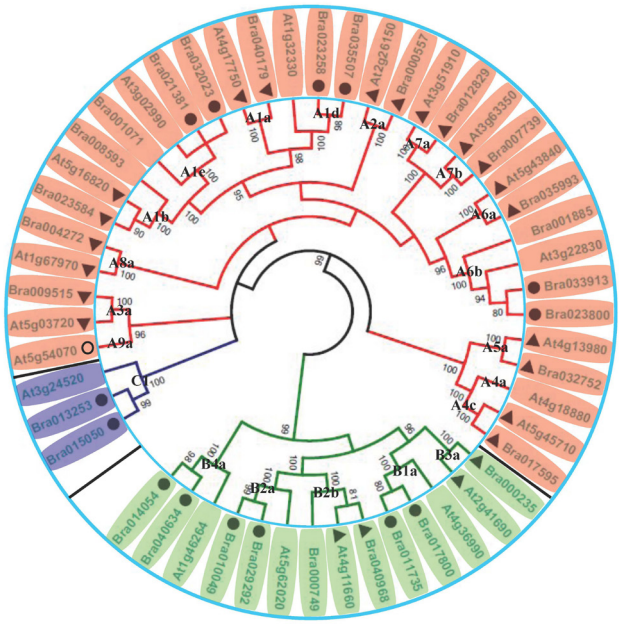

Figure 4. Neighbor-joining phylogenetic tree of Hsfs from Chinese cabbage and *Arabidopsis*. The phylogenetic tree was constructed using the Mega5.0 software for full-length amino acid sequences from Chinese cabbage and *Arabidopsis.* The tree was divided into 3 classes with a total of 51 genes, including class A (A1a, A1b, A1d, A1e, A2a, A3a, A4c, A5a, A6a, A6b, A7a, A7b, and A8), class B (B1a, B2a, B3a, B4a, and B2b), and class C (C1). Class A proteins are represented with red branches, class B are represented with green branches, and class C with blue branches. The black solid circles represent paralogous genes of Chinese cabbage, and the black solid triangles represent orthologous genes from Chinese cabbage and *Arabidopsis*, and black hollow circle represents the single branch A9 of AtHsfs. The numbers are bootstrap values based on 1000 replicates. Only bootstrap values larger than 80% support are indicated. AtHsfs, *Arabidopsis* Hsfs.

Conserved motifs of chinese cabbage Hsf proteins

Motif distribution was analyzed using MEME; the results are shown in Figure 5. The corresponding 20 consensus motifs were detected, lengths ranging from 11-50 aa (Table 3). Most BrHsfs displayed motifs 1, 2, and 3, which corresponded to the conserved DBD domain. In the HR-A/B domain, motifs 4 and 6 were detected in all members of BrHsfs. All class B Hsfs exhibited the motif 6-type HR-A/B region, whereas the motif 4-type HR-A/B region was only detected in classes A and C. Motifs 12 and 13 represented NLS, which were detected in the Chinese cabbage Hsf family. Motifs 12 was characteristic of class B, and the NLS domain was represented by motif 13 in classes A and C. Furthermore, motif 8 was a representative of an NES close to the Hsfs C-terminus. Similarly, motifs 10 containing the AHA motifs were detected in the C-terminus of many BrHsfs. Moreover, some unknown motifs were identified

by MEME motif analysis. Overall, the predicted Hsf DBD, HR-A/B region, NLS, and NES domains were conserved across each Chinese cabbage Hsf.

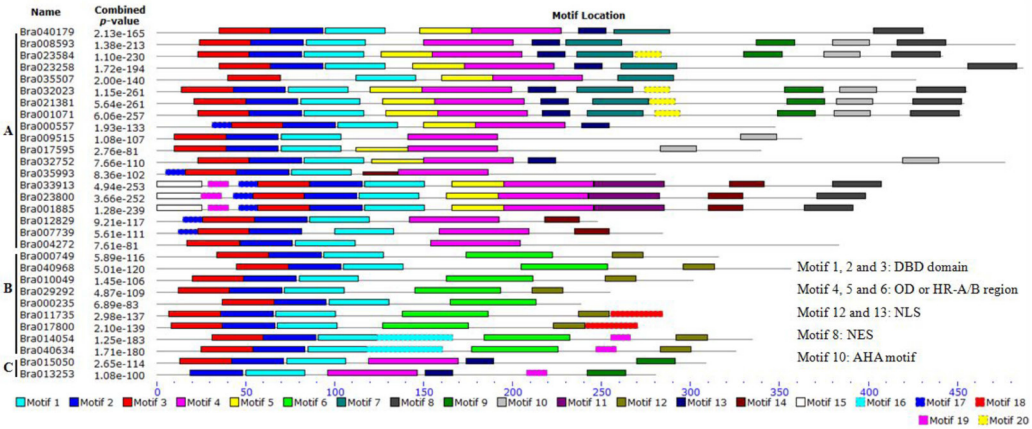

Figure 5. Distribution of conserved motifs in the Hsf in the Hsf family members. The names of all members of the gene clusters defined and the combined P values are shown on the left side of the figure, while motif sizes are indicated at the bottom of the figure. Different motifs are indicated by different color numbered 1-20. For details of motifs, refer to Table 3.

Table 3. Motif sequences identified using MEME tools in Chinese cabbage Hsfs.

Motif	Width	Best possible match
1	33	TYGFRKVDPDRWEFANEGFLRGQKHLLKNIHRR
2	29	FIVWDPPEFARDLLPKYFKHNNFSSFVRQ
3	29	PPPFLTKTYDMVDDPSTDHIVSWSEDNNS
4	50	LMMELVRLRQQQQSTKHYLQHMEQRIHAMEQRQQQMMSFLAKAVQNPHFL
5	29	QHSSVNCCVEVGKYGLEEEVERLKRDKNV
6	48	PTTSCAMVAELTEENECLRRENTQLSSELAHMKKLYDNIYYFMSNYTK
7	31	NGLDRQIVRYQPSMNEAAKTMLRQIHNMSNY
8	27	NLRDWEWNNQQMDHLTEQMGYLTSEAH
9	21	GCETDNGECLDPIMAVLGGSM
10	20	LPGVQDVFWEQFFAERPGIG
11	39	QLIEQKEKRKDMEEAIDKKRQRPIDQGTRHVVCVEDYDD
12	17	KLFGVWIGSKRKRRHHV
13	15	GNQHISEYNKKRRLP
14	19	MSELDALAMHIQGYGDQCT
15	25	MDPSYRFIKEEFPTGFNDSPSPPSS
16	42	KTSQMIPNQHSPFITHHPPPQIPFSGGASFFPLPPRADAAAA
17	11	PQPIEGLHESG
18	29	KNFVVGGSHRTDIKNVDFHAPLWKRSKVC
19	11	YYYNTAMVPNH
20	14	SNNHGSFLLGDVPN

Numbers in the first column indicate the motifs represented in Figure 5.

Expression patterns and interaction networks of Chinese cabbage Hsf proteins

Recently, a comprehensive analysis of RNA-seq data in Chinese cabbage was completed, providing a rich resource for genome annotation and gene expression. For Hsf gene expression analysis in Chinese cabbage, gene expression data was downloaded and then applied for expression profiling of BrHsfs (Figure 6). In general, the expression level in each

organ was as follows: siliques > roots > stems > flowers > leaves. Particularly, *Bra011735* and *Bra004272* showed higher expression levels in all organs, but at least 9 genes displayed lower expression in all organs. Neither *Bra023800* nor *Bra0012829* was detected in roots, while *Bra007739* was not detected in leaves. *Bra009515* was only expressed highly in the leaves, and *Bra15050* was expressed in roots. *Bra017595*, *Bra032752*, *Bra040179*, and *Bra023258*, which were clustered in Hsfs of class A, displayed higher expression in all organs. *Bra040968* and *Bra010049*, members of Hsf class B2, also showed high expression in all organs. Detailed expression values and clusters of each Hsf gene were analyzed using cluster analysis based on RNA-seq datasets (Figure 6).

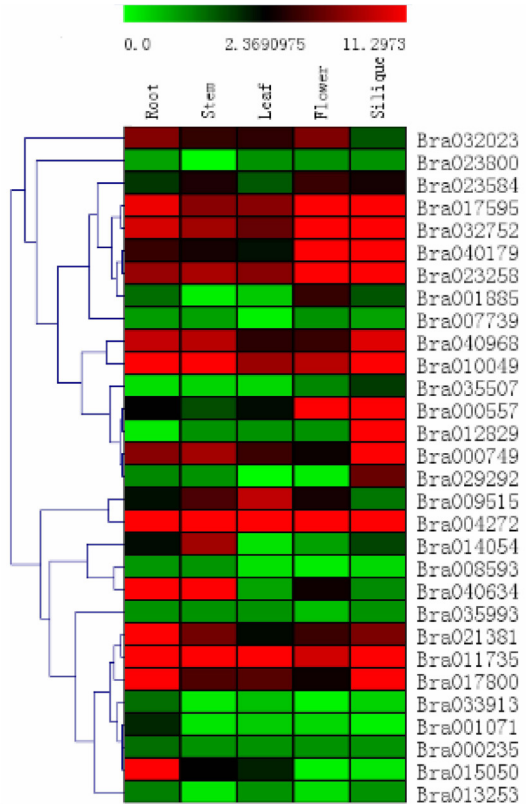

Figure 6. Expression profile cluster analysis of the Chinese cabbage HSF family proteins. Expression values of each HSF gene identified in the study were downloaded from RNA-seq data, including 5 organs, i.e. roots, stems, leaves, flowers, and siliques.

To confirm Chinese cabbage Hsfs expression patterns, 12 orthologous genes from Chinese cabbage and *Arabidopsis* were chosen to construct protein interactions using the STRING software (Figure 7). Subsequently, 5 proteins, including MBF1C, ROF1, TBP2, CDC2 and HSP70, exhibiting sequence similarity with Hsfs between Chinese cabbage and *Arabidopsis*, were involved in 1 interaction network. MBF1C was related to HsfA7A (*Bra012829*) and AtHsfA2 (*Bra00557*), HsfA1 (*Brao40179*) interacted with HSP70, TBP2, and CDC2, while HSP70 was related to the largest number of Hsfs.

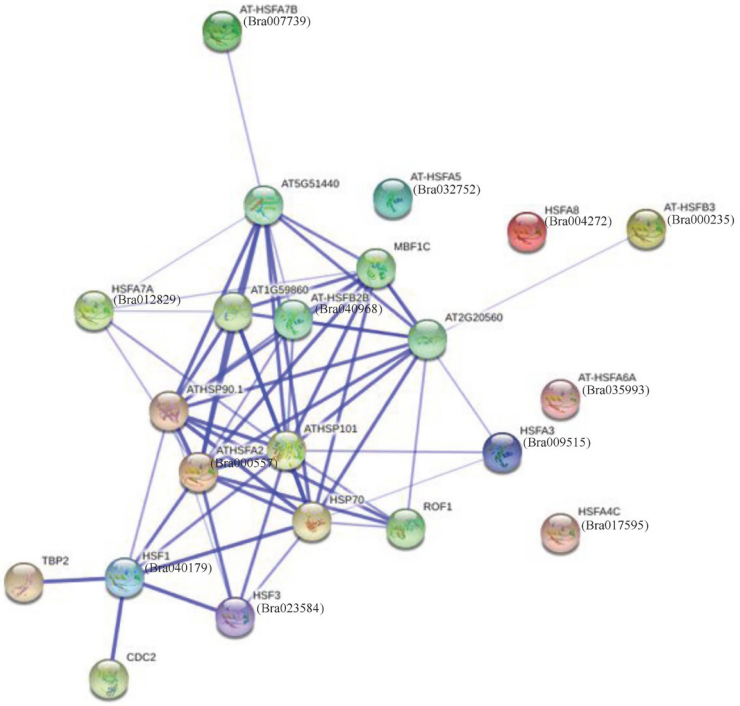

Figure 7. Interaction network of 12 BrHsfs identified in Chinese cabbage and related genes in *Arabidopsis*.

DISCUSSION

Chinese cabbage is an important vegetable and oilseed crop grown worldwide. The Chinese cabbage genome was recently sequenced and assembled, which is useful for genomic analyses. However, the Hsf genes in Chinese cabbage have not been identified. Therefore, it is essential to identify and annotate new Hsf genes in Chinese cabbage. In the present study, we identified 30 Hsf genes in the whole Chinese cabbage genome, and analyzed their phylogenetic relationships, conserved motifs, and expression.

Hsfs exist in all living organisms, but their numbers vary in other plants. There are 21 Hsfs in *Arabidopsis* and 25 Hsfs in rice and apple, 30 Hsfs in maize, and 52 Hsfs in soybean, generated through gene duplication and whole-genome duplications (Proost et al., 2011). Nearly all Hsfs belong to 3 classes, and the numbers in each class are similar in various plant species. *Arabidopsis*, rice, maize, apple, and soybean possessed 15, 13, 16, 16, 28 class A Hsfs, respectively. However, the numbers of class B Hsfs in soybean (22), maize (9), rice (8), and apple (7) are higher than that in *Arabidopsis* (5). Class C Hsfs were present as a single gene in *Arabidopsis*, but 2 Hsfs in class C were observed in apple and soybean, while multiple Hsfs of class C were observed in rice and maize. Chinese cabbage contained a set of 30 Hsfs, with 19 members belonging to class A, 9 to class B, and 2 to class C, which is similar to the plants described above. Additionally, these results indicate that Hsfs were conserved in the plant species, whereas Hsfs of class A were more conserved than those of class B and C in

the plant species overall, with some changes during the evolutionary process. A deletion event may have occurred for Hsfs in classes B and C.

As shown in Figure 1, Hsf family members of Chinese cabbage were mapped to chromosomes to determine their genomic distribution. Hsfs existed in all 10 chromosomes of the Chinese cabbage genome. This suggests that Hsf genes may have been widely distributed in the genome of the common ancestor.

A phylogenetic tree was constructed using the MEGA software using multiple alignments of 30 Chinese cabbage Hsfs and 21 *Arabidopsis* Hsfs with a bootstrap analysis of 1000 replicates to ensure statistical reliability (Figure 4). This method can be used to identify putative paraologous and orthologous Hsf genes. Orthologs are genes in different genomes that have been created through speciation events, while paralogs are genes in the same genome created through gene duplication events (Thornton and DeSalle et al., 2000). Within each Hsf protein class, 7 pairs of paraologous genes and 12 pairs of orthologous genes were identified, indicating ancestral duplication. All paraologous genes appeared between chromosomes, providing information regarding the evolutionary process of the Chinese cabbage Hsf family and indicating that genome duplication likely occurred. In contrast, in maize and Populus, segmental Hsf gene duplications and tandem duplications coexisted, with the former more prevalent than the latter (Lin et al., 2011; Wang et al., 2012). Gene duplication is a major mechanism through which genomic rearrangement and expansion occur; however, diversification of gene function is also generated during molecular evolution. It is thought that tandem duplicates are generally involved in stress responses, suggesting that these tandem duplicates are important for adaptive evolution to rapidly changing environments (Rizzon et al., 2006). In contrast, transcription factor genes encoding for nucleic acid binding proteins originated mostly through segmental duplication (Hanada et al., 2008). Therefore, segmental duplication events in BrHsf expansion may be related to the roles of these genes act as transcriptional regulators. Moreover, no Hsf member clustered in class A9 (*At5g54070*) in Chinese cabbage, which appeared as a single branch of AtHsfs, suggesting a possible gene loss event during the evolution process.

For the modular structure of Hsfs, 5 conserved domains were observed in most Hsf proteins. Near the N-terminus, the multiple alignment results clearly indicated highly conserved DBD domains with 2 insertion events and 1 deletion event in Chinese cabbage (Figure 2). Nine (*Bra035507*) and 17 (*Bra007739*) amino acid residues were increased between the α3 and β3, and 10 amino acid residues lacked *Bra013253*, contributing to biological and functional diversity. Consistent with other plants, class A and C Hsfs had insertions of 21 and 7 amino acid residues between the A and B parts of the HR-A/B regions, respectively (Figure 3). The MEME web server was used to analyze the motif distribution and verify the results of domain prediction (Figure 5 and Table 3), providing additional clues to the evolutionary relationships in BrHsfs. Most members in the same groups possessed similar gene motif structures. For example, motif 4 was only present in class A and C and all B Hsfs members exhibited motif 6; these group-specific motifs are expected to be involved in group-specific functionalities.

The expression pattern of a gene is typically closely related to its function. The available data indicated that the most identified Hsfs were expressed in 5 organs of Chinese cabbage. Furthermore, in the 7 pairs of duplicated genes in Chinese cabbage, a significant divergence in expression levels were observed between the 2 members of each pair. Interestingly, *Bra023258* showed high expression levels in all tissues, particularly in flowers and siliques, while *Bra035507* expression was lower or absent. This may be the result of insertion or dele-

tion events for the 3 genes (*Bra035507*, *Bra007739*, and *Bra013253*). In contrast, the expression of *Bra011735* and *Bra017800* were mainly detected in roots and siliques, indicating that duplicated genes had various functions in the response to heat stress in evolutionary history. We detected an HsfA8 type BrHsf (*Bra004272*) and HsfB3 type BrHsf (*Bra0011735*) that were significantly highly expressed. Among the class A Hsfs, the C-terminal domain of HsfA8 lacked any detectable AHA motifs; additionally, AtHsfA8 was shown to be inactive in a yeast monohybrid assay (Kotak et al., 2004). Hsfs in class B serve as transcriptional repressors or coactivators that cooperate with class A Hsfs (Czarnecka-Verner et al., 2000). *Bra0011735* is likely coactivated by class A Hsfs, but the details of this interaction remain unknown. Indeed, BrHsfs were expressed at higher levels in roots and flowers, indicating that Hsfs participate in the development of underground plant parts and regulation of reproductive growth in Chinese cabbage. Moreover, HsfA1 interacts with HSP70, TBP2, and CDC2 according to the interaction network (Figure 7), which may be because HsfA1 has a unique role as a master regulator in the Hsf family. HSP70, which is related to the largest number Hsfs in Chinese cabbage and *Arabidopsis*, assists a wide range of folding processes and controls the activity of regulatory proteins (Mayer and Bukau, 2005). ROF1 prolongs thermotolerance by sustaining the levels of small HSPs. We also found that CDC2 may be involved in leaf development.

CONCLUSIONS

In this study, a comprehensive set of 30 heat shock factors were identified and characterized from the Chinese cabbage genome. Based on the comparison with homologs from *Arabidopsis* and on the protein structural characteristics, the 30 BrHsfs were grouped into 3 classes (class A, B, and C), and class A was organized into 8 subclasses. Phylogenetic analysis and segmental duplications were examined and found to have contributed to expansion of the Hsf family in the Chinese cabbage genome. The expression profile in 5 organs suggested that most BrHsfs participate in the development of underground plant parts and the regulation of reproductive growth in Chinese cabbage, but significant divergence of expression levels was observed between Hsf genes. Our results increase the understanding of the molecular genetics basis for Chinese cabbage genetic improvement and provide the functional gene resources for further transgenic studies.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31301788, #31372058), the China Postdoctoral Science Foundation (#2013M540500), and Zhejiang Provincial Postdoctoral Science Foundation of China (#Bsh1202084).

REFERENCES

Ahuja I, de Vos RC, Bones AM and Hall RD (2010). Plant molecular stress responses face climate change. *Trends Plant Sci.* 15: 664-674.

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids. Res.* 34: 369-373.

Baniwal SK, Chan KY, Scharf KD and Nover L (2007). Role of heat stress transcription factor HsfA5 as specific repressor of HsfA4. *J. Biol. Chem.* 282: 3605-3613.

Bateman A, Coin L, Durbin R, Finn RD, et al. (2004). The Pfam protein families database. *Nucleic Acids. Res*. 32: 138-141.

Chenna R, Sugawara H, Koike T, Lopez R, et al. (2003). Multiple sequences alignment with the Clustal series of programs. *Nucleic Acids Res.* 31: 3497-3500.

Chung E, Kim KM and Lee JH (2013). Genome-wide analysis and molecular characterization of heat shock transcription factor gene family in *Glycine max*. *J. Genet. Genomics* 40: 127-135.

Czarnecka-Verner E, Yuan CX, Scharf KD, Englich G, et al. (2000). Plants contain a novel multi-member class of heat shock factors without transcriptional activator potential. *Plant Mol. Biol.* 43: 459-471.

Czarnecka-Verner E, Pan S, Salem T and Gurley WB (2004). Plant class B HSFs inhibit transcription and exhibit affinity for TFIIB and TBP. *Plant Mol. Biol.* 56: 57-75.

Damberger FF, Pelton JG, Harrison CJ, Nelson HC, et al. (1994). Solution structure of the DNA-binding domain of the heat shock transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Protein Sci.* 3: 1806-1821.

Delorenzi M and Speed T (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18: 617-625.

Feder ME and Hofmann GE (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* 61: 243-282.

Giorno F, Wolters-Arts M, Grillo S, Scharf KD, et al. (2010). Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. *J. Exp. Bot.* 61: 453-462.

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

Guo J, Wu J, Ji Q, Wang C, et al. (2008). Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *J. Genet. Genomics* 35: 105-118.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, et al. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148: 993-1003.

Kotak S, Port M, Ganguli A, Bicker F, et al. (2004). Characterization of C-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant J.* 39: 98-112.

Kotak S, Larkindale J, Lee U, von Koskull-Döring P, et al. (2007a). Complexity of the heat stress response in plants. *Curr. Opin. Plant Biol*. 10: 310-316.

Kotak S, Vierling E, Bäumlein H and von Koskull-Döring P (2007b). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of *Arabidopsis*. *Plant Cell* 19: 182-195.

La Cour T, Kiemer L, Molgaard A, Gupta R, et al. (2004). Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* 17: 527-536.

Letunic I, Doerks T and Bork P (2009). SMART 6: recent updates and new developments. *Nucleic. Acids. Res.* 37: D229-D232.

Lin YX, Jiang HY, Chu ZX, Tang XL, et al. (2011). Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genomics* 12: 76-89.

Littlefield O and Nelson HC (1999). A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. *Nat. Struct. Mol. Biol.* 6: 464-470.

Liu RH and Meng JL (2003). MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Hereditas* 25: 317-321.

Mayer MP and Bukau B (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *CMLS-Cell Mol. Life S.* 62: 670-684.

Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, et al. (2002). In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes Dev.* 16: 1555-1567.

Nover L, Bharti K, Döring P, Mishra SK, et al. (2001). *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperon* 6: 177-189.

Proost S, Pattyn P, Gerats T and Van de Peer Y (2011). Journey through the past: 150 million years of plant genome evolution. *Plant J.* 66: 58-65.

Rizzon C, Ponger L and Gaut BS (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLOS Comput. Biol.* 2: e115.

Saeed AI, Sharov V, White J, Li J, et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34: 374-378.

Sakurai H and Enoki Y (2010). Novel aspects of heat shock factors: DNA recognition, chromatin modulation and gene expression. *FEBS J.* 277: 4140-4149.

Sakuma Y, Maruyama K, Qin F, Osakabe Y, et al. (2006). Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc. Natl. Acad. Sci. USA* 103: 18822-18827.

Scharf KD, Rose S, Zott W, Schöffl F, et al. (1990). Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J.* 9: 4495-4501.

Scharf KD, Heider H, Höhfeld I, Lyck R, et al. (1998). 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.* 18: 2240-2251.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic. Acids Res.* 39: 561-568.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

Thornton JW and DeSalle R (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genom. Hum. G.* 1: 41-73.

Tong C, Wang X, Yu J, Wu J, et al. (2013). Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* 14: 689-698.

Wang F, Dong Q, Jiang H, Zhu S, et al. (2012). Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. *Mol. Biol. Rep.* 39: 1877-1886.

Wang X, Wang H, Wang J, Sun R, et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43:1035-1039.

Wardlaw IF and Willenbrink J (1994). Carbohydrate storage and mobilisation by the culm of wheat between heading and grain maturity: the relation to sucrose synthase and sucrose-phosphate synthase. *Aust. J. Plant Physiol.* 21: 255-271.